Rejection of Top Background in the WW Control Region for the $H \to WW \to \ell\nu\ell\nu$ Analysis

Author:
Cecilia Nyström (890219-8509)
cecnys@kth.se

Department of Physics
Royal Institute of Technology (KTH)

Supervisor: Jonas Strandberg
Examiner: Bengt Lund-Jensen

16 september 2014

TRITA-FYS 2014:58
ISSN 0280-316X
ISRN KTH/FYS/–14:58—SE

Abstract

One important task when the ATLAS collaboration studies the Higgs boson is to have good background rejection, since it is vital for finding a small signal amongst a large background. Equally important is to find pure control regions in data where a certain background process can be estimated. The goal of this study is to find possible discriminating variables between top and WW processes in the WW control region containing events with one jet. To accomplish this, the multivariate machine learning method Boosted Decision Trees (BDT) is used. Different BDT methods are trained to find the one that is best at separating signal from background. Starting from a set of 15 variables, an optimization is performed to find the optimal subset containing fewer variables. It is found that, using a set of four variables, it is possible to reject 43% of the top background while keeping 90% of the WW signal in the WW control region.

Sammanfattning

När ATLAS-kollaborationen studerar Higgsbosonen är en av deras viktigaste uppgifter att uppskatta mängden bakgrund. Detta är avgörande för att kunna urskilja en liten signal i data dominerad av bakgrund. För att kunna uppskatta mängden av en viss bakgrund är det viktigt att hitta kontrollregioner som inte innehåller andra bakgrunder. Målet med denna studie är att hitta variabler som kan särskilja topp- och WW-processer i WWs kontrollregion. För att åstadkomma detta används maskininlärningsmetoden Boosted Decision Trees (BDT). Olika BDT-konfigurationer provas för att se vilken som är bäst på att separera signal och bakgrund. En mängd med 15 variabler provas för att hitta den delmängd av färre variabler som ger bäst resultat. Det konstaterades att med en uppsättning av fyra variabler så är det möjligt att avvisa 43% av topp-bakgrunden samtidigt som 90% av WW-signalen behålls i WWs kontrollregion.

Contents

Chapter 1

Introduction

On the fourth of July 2012, CERN (European Organization for Nuclear Research) announced the discovery of a previously unknown boson, later confirmed to be the Higgs boson. The discovery of the Higgs boson marked the end of a more than 40 year long search for what was the last missing piece of the Standard Model, but also marked the beginning of new research to see if all the predictions regarding this particle would fit with reality. The discovery was made using the Large Hadron Collider (LHC), the world's largest and most powerful particle accelerator, situated beneath the earth's surface on the border between France and Switzerland. The discovery was made independently by the two teams working on the two general purpose detectors at the LHC; ATLAS (A Toroidal LHC Apparatus) and CMS (Compact Muon Solenoid). This thesis is limited to the work done by the ATLAS collaboration studying the decay channel $H \to WW^* \to \ell\nu\ell\nu$, where $\ell = e, \mu$.

The ATLAS collaboration generally uses a cut based analysis. This means that cuts are placed on one variable at a time. These cuts have been optimized in such a way that as much background as possible is rejected while as much signal as possible is kept. In the analysis it is important to have good background rejection and one method to estimate the amount of background is to count the number of background events in a control region (CR) and then using a factor to extrapolate to the signal region (SR). The CRs are chosen so that they are dominated by a particular background but is often contaminated with other backgrounds. These other backgrounds need to be reduced in order to get a pure CR which is ideal to get a good background estimation. Two of the more difficult backgrounds to separate are from top quarks and W bosons, as they can have the same final state particles. It is possible that the cut based analysis could be improved using machine learning to separate signal from background. When using machine learning, a computer is trained with a set of inputs and known outputs, so that it can later predict the output using only the input. In this case, the computer is trained to separate signal and background (output) when given an event (input). One of the most common machine learning methods is the Boosted Decision Tree (BDT). The Toolkit for Multivariate Analysis (TMVA) is a framework based on ROOT [1] that gives the user access to machine learning methods to differentiate between signal and background. It incorporates various multivariate methods, among them several BDT methods which will be used in this study. The goal of this study is to find possible discriminating variables between the top and WW processes in the WW 1-jet CR using BDTs. Different BDT

methods will be tested to find which one works best. It is based on a similar study performed by Jelena Jovićević in October 2013.

This thesis is structured as follows: in chapter two an overview of the Standard Model is presented, followed by a presentation of the Large Hadron Collider and the ATLAS detector in chapter three. In chapter four Higgs phenomenology is presented and in chapter five the analysis done by the ATLAS $H \rightarrow WW$ group is presented. Chapter six begins with an introduction to BDTs and TMVA and the rest of the chapter is dedicated to a description of the variable optimization performed and the results from this study. Lastly chapter seven contains the conclusions from the study.

Chapter 2

The Standard Model

The Standard Model (SM) of particle physics is a quantum field theory (QFT) and describes almost everything we know about fundamental particles and forces. It has made several predictions that have been proven to be correct, the latest of which is the discovery of the Higgs boson. It contains all the known fundamental particles and describes the electromagnetic, the weak, and the strong interactions.

The history of the SM started in 1961 when Glashow wrote down a theory that was the basis of the unification of the electromagnetic and weak interaction into the electroweak interaction [2]. A problem was that in this theory the gauge bosons were massless and we knew that the gauge bosons of the weak interaction are massive. The story continued in 1964, when three very important papers were published that developed the BEH-mechanism [3–5]. All three papers have different approaches but all show a mechanism through which gauge bosons can acquire mass. In 1967 Weinberg and Salam incorporated this into the theory of electroweak interaction [6, 7]. The moment when people accepted the SM was when t'Hooft and Veltman showed that the SM is renormalizable [8]. A theory needs to be renormalizable, since a theory that describes physical processes needs to do this with well defined and finite mathematical expressions. In the 1970s the strong interaction was formulated and with that the SM started to look like its modern form. For their discoveries, Glashow, Salam, and Weinberg were awarded the Nobel Prize in Physics 1979, t'Hooft and Veltman in 1999, and Englert and Higgs in 2012.

The predictions of the SM have been verified experimentally to great precision [9]. The last piece of the puzzle was the discovery of the Higgs boson [10, 11]. While this is one of the biggest discoveries in the history of physics we know that the SM is not a complete theory. Firstly, it does not incorporate a theory of gravity. Secondly, we have very strong evidence from rotational curves of galaxies, galaxy clusters, and gravitational lensing that there exists dark matter in the universe. The only SM particle that could be a dark matter candidate is the neutrino, but it could only consititute a very small fraction of the dark matter in the universe [12]. Finally, the SM does not explain neutrino oscillations, see sec 2.1.1. But even with these deficits, the SM is an extremely successful theory.

2.1 Particles and Interaction fields

The first building block of the SM is its particles. In QFT, particles can be considered to be excitations of fields. The particles are divided into two categories: fermions and gauge bosons.

2.1.1 Fermions

Fermions are spin 1/2 particles that respect the Pauli exclusion principle [13]. There are 12 fermions in the SM as shown in table 2.1. Every fermion (f) has a corresponding anti particle denoted with $\bar{f}$. The anti particle has the same mass as the particle but opposite electric charge. The fermions form three generations and there are strong experimental evidence that the number of generations are exactly three [14, 15].

Generation	1	2	3	Charge
Leptons	ν_{e} (e-neutrino)	ν_μ (μ-neutrino)	ν_τ (τ-neutrino)	$q = 0$
	e^- (electron)	μ^- (muon)	τ^- (tau)	$q = -e$
Quarks	u (up)	c (charm)	t (top)	$q = \frac{2}{3}e$
	d (down)	s (strange)	b (bottom)	$q = \frac{-1}{3}e$

Table 2.1: The Standard Model fermions with their associated quantum numbers: charge and weak isospin.

As can be seen in table 2.1, the fermions are divided into two categories: leptons and quarks, and the two members of each generation is said to have different *flavour*. The charged leptons interact via the electromagnetic and weak interaction and the neutrinos only via the weak interaction. Because of this the neutrinos are difficult to detect. In the SM the neutrinos are treated as massless, but we know that this is not true [16], from the fact that we can observe neutrino oscillations where one type of neutrino oscillates to another and this is only possible if neutrinos have mass. The defining property of the quarks is that they carry colour charge (red, green, and blue) and thus interact via the strong interaction, but they can also interact via the electromagnetic and the weak interaction. Quarks cannot exist freely, instead they bind together to form colourless, or colour neutral, particles. This is called colour confinement. All observable particles are colourless and the particles composed of quarks come in two forms: baryons (qqq) and mesons ($q\bar{q}$).

A higher generation of quarks and leptons, with the possible exception of neutrinos where we have not been able to determine their masses, corresponds to more massive particles. The higher order particles are unstable and will decay into the first generation. Ordinary, stable matter is made up of particles from the first generation: u, d, and e^-. The charged proton (uud), the neutral neutron (udd), and the electron bind together to form the atoms that make up ordinary matter.

Spin and Chirality

In particle physics, spin is an intrinsic angular momentum measured in units of $\hbar$ that all particles except the Higgs boson carry. The *helicity* is the projection of a particle's spin onto the direction of its momentum and is thus frame dependent if the particle is massive. Helicity and *chirality* are two very closely related ideas and they are the same for massless particles such as the photon, but the chirality is more abstract. The chirality of a particle is a frame independent property. The SM is a *chiral theory*, which means that left- and right-handed chiral particles behave differently. In the SM the fermions are assigned to be left handed doublets and right handed singlets. The charged leptons (the electron, the muon, and the tau) as well as the quarks all have both left handed and right handed chiral components, whereas the neutrinos only have left handed components. The quark doublets and singlets are:

$$\begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L, u_R, d_R, c_R, s_R, t_R, b_R. \tag{2.1}$$

The lepton doublets and singlets are:

$$\begin{pmatrix} e^- \\ \nu_e \end{pmatrix}_L, \begin{pmatrix} \mu^- \\ \nu_\mu \end{pmatrix}_L, \begin{pmatrix} \tau^- \\ \nu_\tau \end{pmatrix}_L, e_R^-, \mu_R^-, \tau_R^-. \tag{2.2}$$

2.1.2 Gauge Bosons and Interaction Fields

There are four fundamental forces in nature: the strong, the weak, the electromagnetic, and the gravitational force. The first three are a part of the SM but the gravitational force is not. This is because gravity is very weak on a quantum level and therefore does not play a role in elementary particle physics. At high energies the electromagnetic and the weak interaction can be combined into one electroweak interaction. The three interactions in the standard model are all mediated by gauge bosons. In contrast to the fermions, the gauge bosons have integer spin, and thus do not follow the Pauli exclusion principle. The different types of gauge bosons are summarized in table 2.2.

Boson	Interaction	Mass [GeV]
γ	Electromagnetic	0
$W^{\pm}$	Weak, charged	80.4
Z^0	Weak, neutral	91.19
8 gluons	Strong	0

Table 2.2: The Standard Model gauge bosons and their masses.

When mediating an interaction, the gauge bosons usually exist as virtual particles, where they temporarily do not satisfy the conservation of energy and momentum. We can see how this is possible from the Heisenberg uncertainty principle, $\Delta E \Delta t > \hbar/2$. For a short time we can get their range as:

$$R \leq c\Delta t \sim \frac{\hbar}{mc} \tag{2.3}$$

and the range is therefore inversely proportional to the mediating particle's mass. The electromagnetic interaction thus has infinite range and the weak interaction very short range. The gluons are massless but the strong interaction is short ranged because of colour confinement.

Electromagnetic Interaction

Photons mediate the electromagnetic interaction and couple to all charged particles. Since the photon is uncharged it does not couple to itself. The electromagnetic interaction is described by quantum electrodynamics (QED), a gauge theory based on the symmetry group $U(1)$.

Weak Interaction

The $\mathrm{W}^{\pm}$ and Z^0 bosons mediate the charged and neutral weak interactions respectively. In contrast to the photon and the gluons the weak interaction vector bosons are massive. The weak interaction affects all the fermions, the electroweak gauge bosons themselves, and the Higgs boson and is the only interaction that can transform particles. The charged weak interaction is described by a gauge theory based on the symmetry group $SU(2)_L$, where the subscript L represents the fact that it only couples to fermions of left-handed chirality. The leptons couple to $W^{\pm}$ within the generations. This means that, for example, $W^{\pm}$ can take an electron to an electron neutrino but not to a neutrino in any other generation. With quarks, the same thing is generally true but it is not the whole story; the $W^{\pm}$ actually couples to the quark weak eigenstates and these weak eigenstates are related to the mass eigenstate by a mixing matrix called the CKM matrix [17]. In contrast to the charged weak bosons, the Z^0 couples to both left and right handed fermions and conserves flavour.

Strong interaction

The eight electrically neutral gluons mediate the strong interaction. They carry colour charge and thus couple to quarks and to themselves. The strong interaction is described by quantum chromodynamics (QCD), a gauge theory based on the symmetry group $SU(3)$. The strong interaction does not affect colourless particles such as the leptons.

Quarks cannot exist freely. If two quarks are pulled apart the strong interaction between them increases in strength, until the point where the potential between them contains enough energy to create another pair of quarks. This is known as confinement and it is because of this jets are created in accelerator experiments. Another aspect of QCD is that the strong coupling constant becomes small on small length scales. This means that at high energies the quarks act like free particles. This is known as asymptotic freedom.

2.2 Electroweak Unification and the BEH-mechanism

The electromagnetic and the weak interactions look very different at low energies, but as mentioned before, they are just two aspects of the same phenomenon. The electroweak theory is a gauge theory based on the symmetry group $SU(2) \times U(1)$. The electroweak Lagrangian contains four gauge fields, W^i $(i = 1, 2, 3)$ and B_A. The fields of the experimentally observable weak bosons $W^{\pm}$, Z, and the photon are given by combinations of the electroweak gauge fields. The charged weak bosons are given by a combination of the first two components of W^i:

$$W^{\pm} = \frac{1}{\sqrt{2}}(W^1 \mp W^2) \tag{2.4}$$

The neutral weak boson Z^0 and the photon are formed from the mixing of the third component of W^i and the B_A via the weak mixing angle θ_W [18]:

$$\begin{pmatrix} \gamma \\ Z^0 \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_A \\ W^3 \end{pmatrix} \tag{2.5}$$

The electroweak Lagrangian describes massless gauge bosons, and we know that this is not an accurate description of nature. It is not possible to explicitly add mass terms for the gauge bosons, because then the Lagrangian is not gauge invariant and hence the theory is not renormalizable. Instead we need to get masses from a mechanism that preserves gauge invariance: the BEH-mechanism. The BEH mechanism was developed by three independent teams of physicists [3–5]. The idea is to introduce a scalar field with a non-vanishing vacuum expectation value, called the Higgs field. The Higgs field is a complex $SU(2)$ scalar doublet field with four degrees of freedom:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{2.6}$$

To understand the BEH-mechanism we need to understand *spontaneous symmetry breaking* and *local gauge invariance* [19, 20]. The idea behind spontaneous symmetry breaking is that a physical system that is initially symmetric may naturally go into a non-symmetric vacuum state. This means that spontaneous symmetry breaking is only relevant when the lowest energy state (the vacuum) is degenerate. The potential of the Higgs field is that of the "Mexican hat" depicted in fig. 2.1.

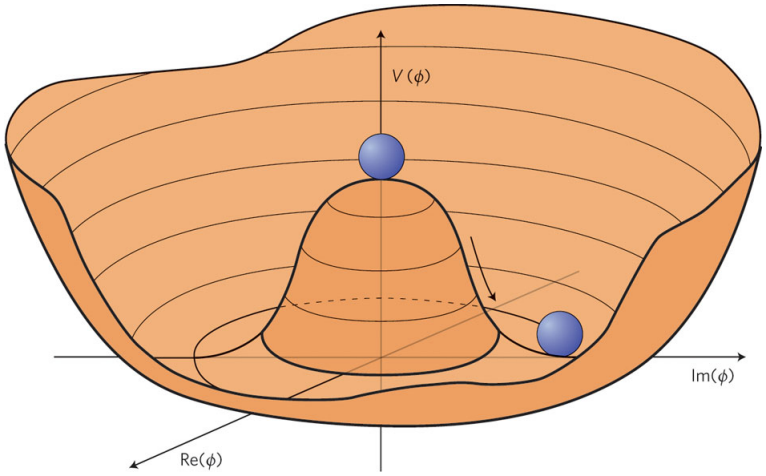

Figure 2.1: Illustration of the "Mexican hat" potential. The vacuum is an arbitrarily chosen point around the bottom of the potential.

This potential is given by:

$$U(\phi) = \frac{-\mu^2}{2}\phi^2 + \frac{\lambda}{4}\phi^4 \tag{2.7}$$

where μ and λ are constants and ϕ are the fields. It has a local maximum that lies on the symmetry axis, and degenerate minima on a circle. This potential is unstable at $\phi = 0$, and any small fluctuation will spontaneously break the symmetry. We can arbitrarily choose a point on the circle to expand the Lagrangian around. This new Lagrangian will represent the same physical system as before but the symmetry is "broken". Using this spontaneous symmetry breaking on the electroweak Lagrangian will create three unwanted massless scalar bosons in accordance with the Goldstone theorem [21], called the Goldstone bosons.

Local gauge invariance means that we can do a local gauge transformation to the fields involved in the theory without the physical laws changing as a result. By combining this fact with spontaneous symmetry breaking we can eliminate the Goldstone bosons by a "smart" choice of gauge. Instead we are left with massive weak gauge bosons ($W^{\pm}$ and Z^0), the photon and a new massive scalar boson, the Higgs boson. We can look at this more intuitively by considering the degrees of freedom of the system. Before the spontaneous symmetry breaking we had one scalar doublet with four degrees of freedom and four massless gauge fields, that will have two degrees of freedom each. After the spontaneous symmetry breaking we end up with three massive interaction fields that each have three degrees of freedom, the massless photon with two degrees of freedom and the Higgs boson with one degree of freedom. Thus, the degrees of freedom of the system is the same as before the spontaneous symmetry breaking. An important note is that this is the same physical system we started with, we have just made the content more clear. A similar procedure can be done to acquire fermion masses, by introducing a Yukawa coupling between the fermion fields and the Higgs field.

The Higgs boson is thus a massive scalar particle, predicted by the BEH-mechanism and discovered by CERN [10, 11]. Since the Higgs field is a scalar field the Higgs boson has spin 0, which has been verified experimentally [22]. The SM does not predict the mass of the Higgs boson, but from the discovery we know that the mass is $m_H = 125$ GeV. The Higgs boson couples to all massive particles, including itself.

Chapter 3

The LHC and the ATLAS Experiment

CERN (European Organization for Nuclear Research) is the world's largest particle physics laboratory. The organization is mostly European and has 21 member states. CERN had made many important scientific achievements before the discovery of the Higgs boson, these include the discovery of the electroweak gauge bosons and the creation of anti-hydrogen. In this chapter, the LHC (Large Hadron Collider) and the ATLAS (A Toroidal LHC ApparatuS) experiment, will be presented.

3.1 LHC

The LHC [23] at CERN is the world's largest and most powerful particle accelerator, a schematic view of the layout of the LHC can be seen in fig. 3.1. It is situated in a tunnel that originally housed another CERN particle accelerator, the LEP (Large Electron–Positron Collider). The circular tunnel is 26.7 km long and lies between 45 m and 170 m beneath the earth's surface on the border between France and Switzerland. The goal of the LHC is to probe physics with centre of mass collision energies of up to 14 TeV using colliding bunches of protons.

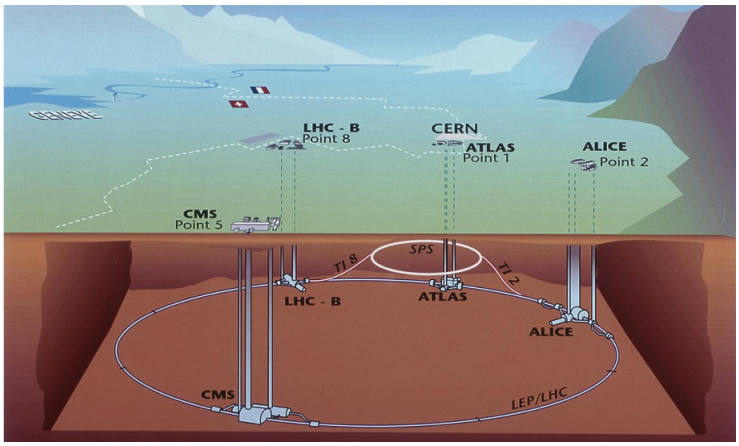

Figure 3.1: Schematic view of the layout of the LHC, including the four main experiments (Image courtesy of CERN).

Before being injected into the main accelerator tunnel through two transfer tunnels, the protons are accelerated by a series of accelerators at the CERN accelerator complex that successively increase their energy. In the tunnel, the two protons beams are accelerated in opposite directions in two parallel beam pipes until they reach the desired energy. To make the protons go in the two different directions, the LHC needs opposite magnetic dipole fields in the two beam pipes. The protons are kept in their circular paths by the magnetic fields from 1231 superconducting dipole magnets that need to be kept at 1.9 K. To be able to keep the magnets that cold, the accelerator is connected to a cooling system that uses liquid helium. In addition, 392 quadrupole magnets are used to focus the protons closer together.

A very central concept at LHC is *luminosity*. Luminosity, $\mathcal{L}$, has unit $\mathrm{cm}^{-2}\mathrm{s}^{-1}$ and its magnitude depends on the beam. From the luminosity we can get the integrated luminosity as $L = \int \mathcal{L} dt$. The integrated luminosity is related to the number of events (N) generated for a process with cross section σ:

$$N = L \cdot \sigma. \tag{3.1}$$

Higher integrated luminosity means that there is more data to analyze and that rare processes can be observed. One way to increase the luminosity is to focus the beam even more, while another is to increase the number of particles in the beam. ATLAS is designed to have peak luminosity of $\mathcal{L}_{ATLAS} = 10^{34}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$. This is a very high luminosity and leads to more data than can be recorded, because of this ATLAS need a good trigger system to reduce the amount of data.

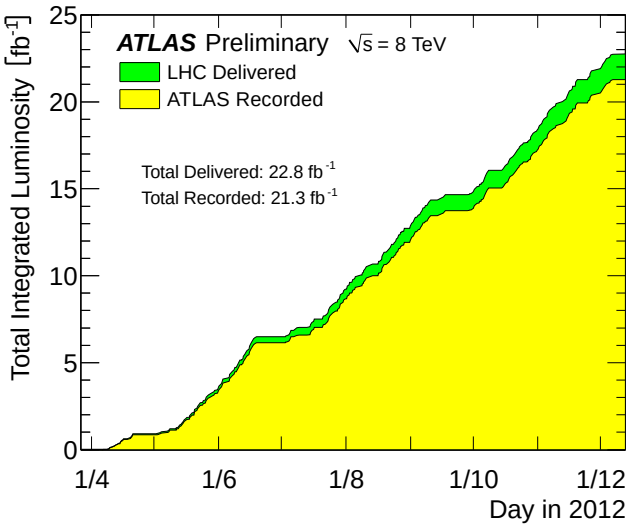

Figure 3.2: Integrated luminosity delivered by the LHC (green) and recorded by ATLAS (yellow) during 2012 (Image courtesy of CERN).

3.2 ATLAS Experiment

The ATLAS experiment [24] is one of two general purpose detectors at the LHC, a schematic view of the ATLAS detector can be seen in fig. 3.3. ATLAS is about 45 m long and more than 25 m high and is situated in a cavern 92.5 m underground to be shielded from cosmic rays. The ATLAS detector is built up as an onion with layers upon layers of detectors surrounding the interaction point where the proton beams from the LHC collide. The innermost detector is a tracking detector, which is then surrounded by calorimetry and muon chambers. All of these three systems are independent but complementary. The only particles ATLAS cannot detect are neutrinos, they are the reason for missing momentum. For this reason it is important that ATLAS detector is hermetic, this means that it detects all other particles.

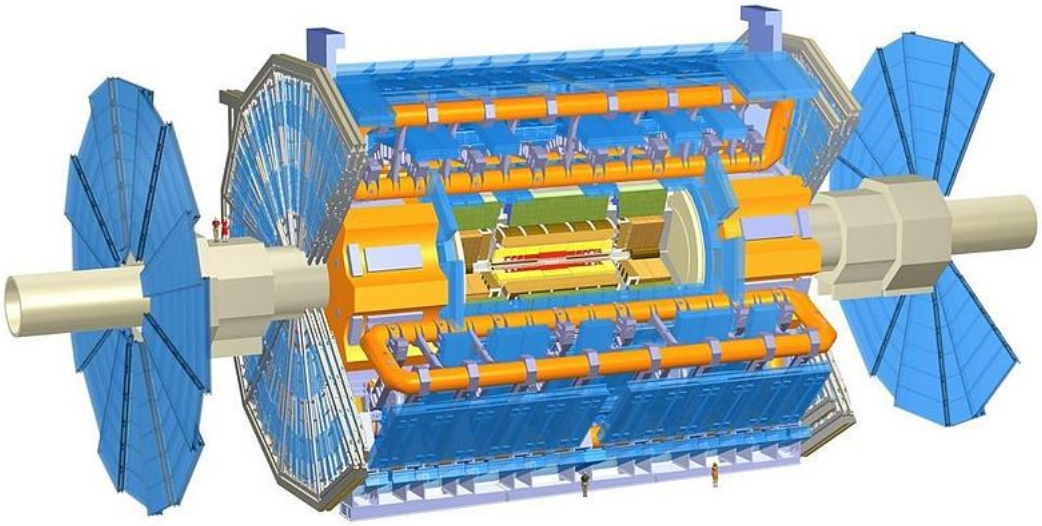

Figure 3.3: Schematic view of the ATLAS detector (Image courtesy of CERN).

Before presenting the different subsystems of the detector we want to present the coordinate system used. It is a right-handed cylindrical coordinate system where the origin is placed at the interaction point. The beam direction defines the z-axis, the positive x-axis is pointing from the beam towards the center of the LHC ring and the positive y-axis is pointing upwards towards the earth's surface. From this we get the azimuthal angle as $\phi = arctan(x/y)$ that runs in the transverse plane around beam, and the polar angle θ is measured from the beam axis. Instead of the polar angle a spatial coordinate called the *pseudorapidity* is often used. The pseudorapidity is defined as:

$$\eta = -ln\left[tan\left(\frac{\theta}{2}\right)\right]. \tag{3.2}$$

A small pseudorapidity corresponds to directions close to orthogonal from the beam direction, and a large pseudorapidity corresponds to directions close to the beam pipe.

3.2.1 Inner detector

The purpose of the inner detector (ID) is to track the particles created in the collision by detecting the particle's interactions in the detector at different points along the particle's track. Because approximately 1000 particles will be created in each collision, there will be a high track density in the ID. Very good momentum resolution and pattern recognition is therefore essential. The layout of the ID can be seen in fig. 3.4. The whole ID is contained in a cylindrical case that is 7.024 m long and has a radius of 1.150 m and covers $|\eta| < 2.5$. It is enclosed in the 2 T magnetic field from the central solenoid. This magnetic field makes the charged particles bend and from the curvature we can get the charge and the momentum of the particle.

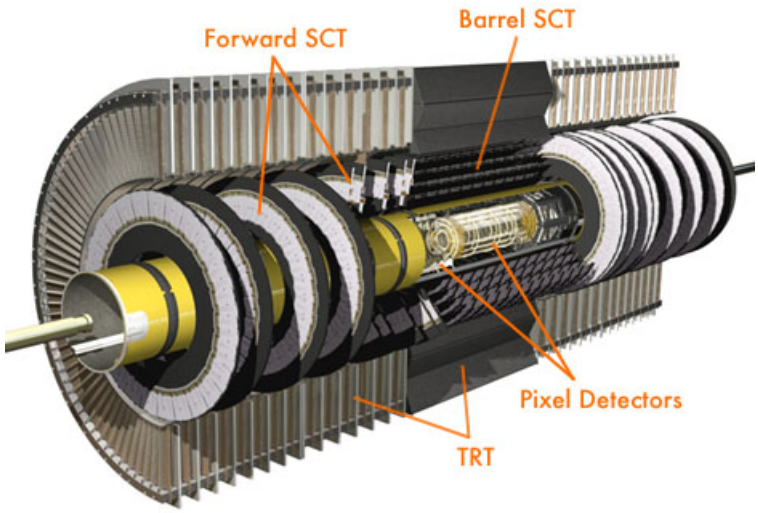

Figure 3.4: Cut away view of the ATLAS Inner Detector, with its various parts indicated (Image courtesy of CERN).

As can be seen in fig. 3.4, the ID is made up of three independent subdetectors. The innermost detector is the pixel detector, which extends from radius $r = 5.05$ cm to $r = 12.25$ cm [25]. A pixel detector uses many isolated pixel cells to get two-dimensional position information. In ATLAS, both the cylindrical central part and the end-caps of the pixel detector consists of three layers of modules with silicon sensors. Each sensor is approximately 6 cm $\times$ 2 cm and is divided into 47232 pixel cells. The small pixel size gives very precise tracking close to the interaction point. The whole pixel detector add up to a total of 1744 modules giving an active area of 1.7 m^2 of silicon. Outside the pixel detector, the Semi-Conductor Tracker (SCT), which extends from radius $r = 29.9$ cm to $r = 51.4$ cm, continues the high precision tracking [26]. The function of the SCT is similar to the pixel detector, but in the SCT the modules consist of thin silicon strips, 80 μ m $\times$ 0.12 m, instead of pixels. The cylinder area is made up of four layers of modules, and each endcap consists of nine layers. The whole SCT adds up to 4088 modules giving an active area of 63 m^2 of silicon. The outermost part of the tracking system consists of several layers of gaseous straw tubes (TRT), which extends from radius $r = 55.4$ cm to $r = 108.2$ cm [27]. The TRT consists of 370000 straws that are 4 mm in diameter. The

straws are arranged parallel to the beam axis in the cylindrical region and radially in the end caps. The TRT is not as precise as the two inner detector in the ID, but it would not be cost effective to have silicon detectors to cover such a large volume.

3.2.2 Calorimetry

Outside of the solenoidal magnet that surrounds the ID there are the calorimetry [28]. The purpose of a calorimeter is to measure the energy of a particle by completely stopping it. There are two calorimeters in the ATLAS detector, as can be seen in fig. 3.5. The first calorimeter, which is closer to the beam pipe, is the electromagnetic (EM) calorimeter and the second is the hadronic calorimeter. They are both so called sampling calorimeters. In a sampling calorimeter there are alternating layers of absorber and sampling material. When the particle enters the calorimeter it will interact in the absorber, generating a particle shower. Then the energy will be collected in the sampling material. The EM calorimeter stops and measures most of the energy of electrons, positrons, and photons. Jets produced by hadrons will lose some of their energy in the EM calorimeter and the hadronic calorimeter stops and measure the remaining energy of the jets. For electrons and photons, the main mechanism for energy loss is radiation losses due to bremsstrahlung and pair production. The hadronic shower is similar to the EM one but more complex because a lot of processes contribute to the shower.

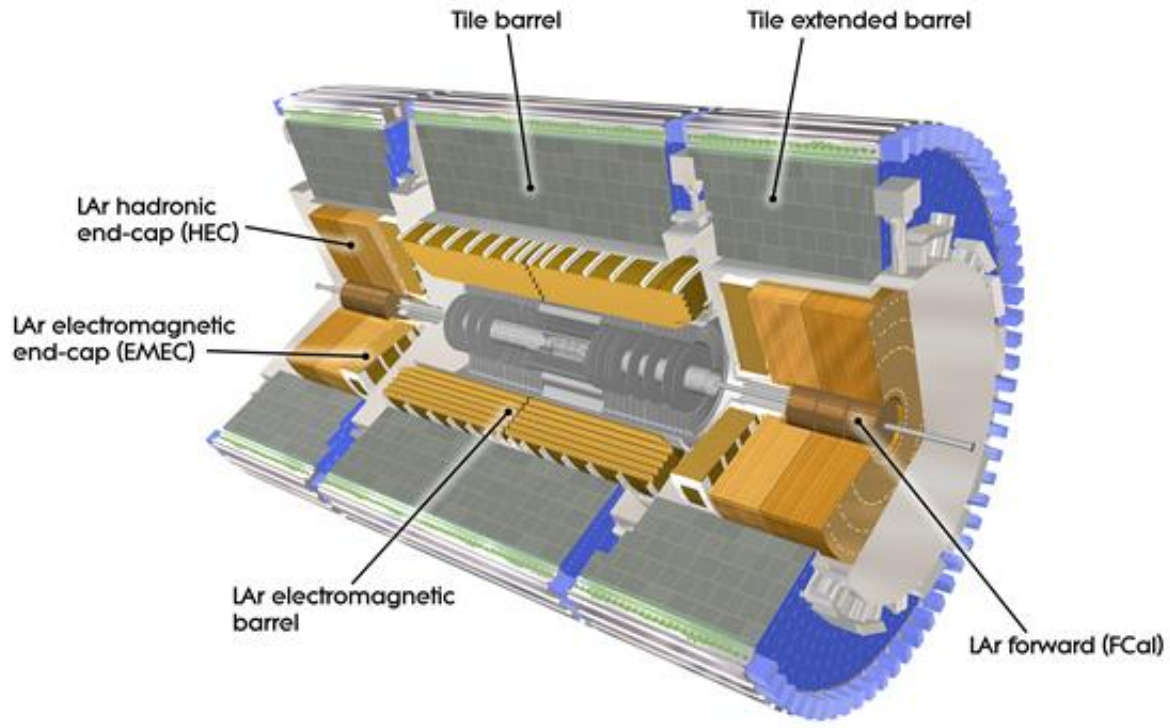

Figure 3.5: The ATLAS Calorimeter system with its various parts indicated (Image courtesy of CERN).

The EM calorimeter is made up of lead absorbers that are immersed in liquid argon that acts as the sampling material. It is divided into a barrel part and two endcap parts and covers the pseudorapidity range $|\eta| < 3.2$. The hadronic calorimeter is made up of different materials in different parts of the calorimeter. In the central region of the hadronic calorimeter, that covers the range $|\eta| < 1.7$, iron is used as absorber and scintillating plates are used as sampling material. The hadronic endcap calorimeter, that covers the range $1.5 < |\eta| < 3.2$, is made up of copper plates as absorbers and liquid argon is used as the sampling material. The forward calorimeter, situated close to the beam pipe that covers the range $|\eta| < 4.9$, consists of tungsten rods as absorber that are immersed in liquid argon that acts as the sampling material. The forward calorimeter is used to provide energy measurements of forward jets.

3.2.3 Muon System

The outermost layer of the ATLAS detector is the muon system, see fig. 3.6. The muon system consists of two subsystems and covers $|\eta| < 2.7$ [29]. Three large toroids generate the magnetic field. The first toroid is placed around the calorimeters and provides a magnetic field of 0.5 T. The other two, the endcap toroids, are placed at each end and are lined up with the central solenoid and provides a magnetic field of 1 T. Each toroid is made up of eight coils that are assembled radially around the beam pipe.

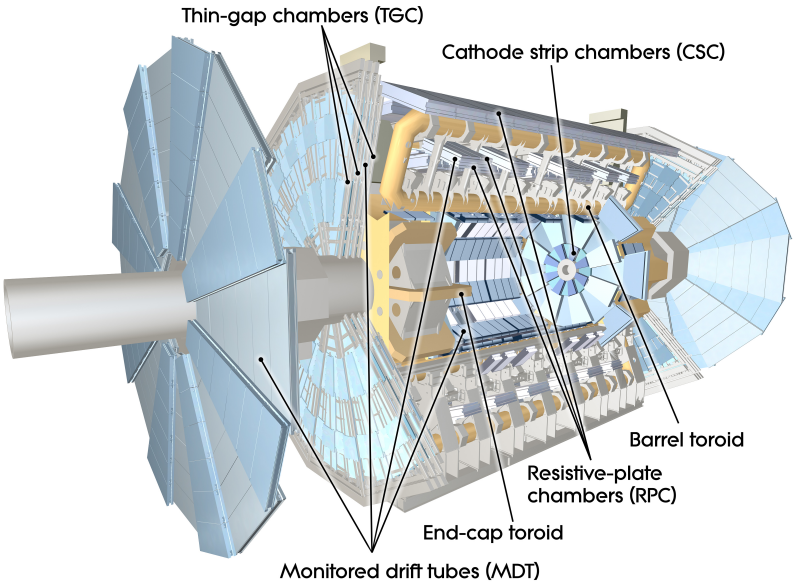

Figure 3.6: The ATLAS muon system, with its various parts indicated (Image courtesy of CERN).

The muon spectrometer is made up of trigger and tracking chambers. ATLAS has two types of trigger chambers, the Resistive Plate Chambers (RPC) for the barrel region and the Thin Gap Chambers (TGC) in the endcap region. The tracking chambers are, as the name suggests, used for momentum measurements and precision tracking. The Monitored Drift Tube (MDT) chambers are used everywhere except for the innermost layer of the endcap, where ATLAS has Cathode Strip Chambers (CSC). The MDTs are

gas filled tubes with an anode wire in the middle that collects the ionization as muons pass. The CSCs are multiwire proportional chambers placed on two disks.

3.2.4 Trigger System

Massive amounts of data are generated by the ATLAS detector (50 TB/s) and a good trigger system is needed to reduce the amount by more than five orders of magnitude. The reduction is made in three steps [30,31] and after each step the event may be deleted or sent to the next step depending on predefined criteria. The first level (L1) is hardware based and uses only some of the information from the calorimeter and the muon system. The L1 has 2.5 μs to make a decision and reduces the event rate to 75 kHz. The second level (L2) is software based and uses information from all subdetectors. It has 40 ms to make a decision and reduces the event rate to about 3 kHz. The L2 makes a decision before the event is fully assembled. The last level is the event filter (EF). It has about 4 s to make a decision. It reduces the rate to 200 Hz. For the trigger system to work it needs to be perfectly timed. To ensure this, tests with cosmic muons have been performed [32].

3.3 Particle Identification

The purpose of the structure of the ATLAS detector is the ability to distinguish different particles from each other by exploiting their unique features. Typical signals, from common stable particles, in the ATLAS detector can be seen in fig. 3.7.

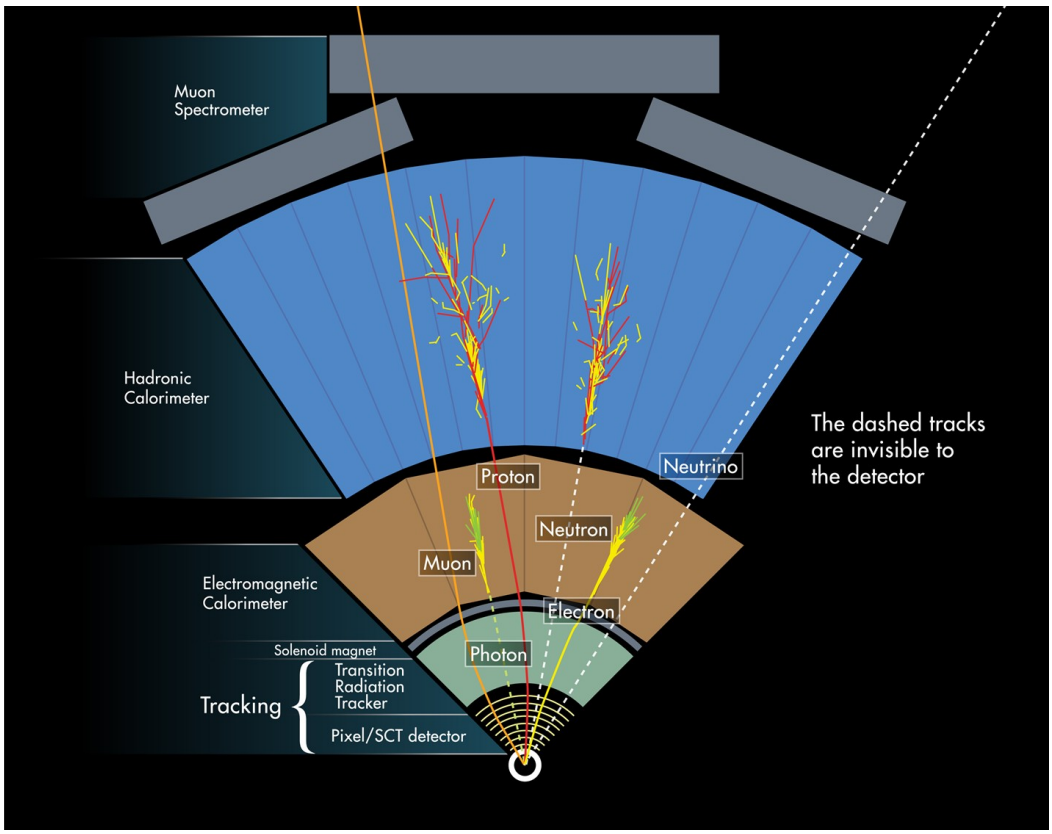

Figure 3.7: Typical particle signals in the ATLAS detector (Image courtesy of CERN).

- Photons, electrons, and positrons will all have similar signature in the EM calorimeter. They can be distinguished from the fact that the uncharged photon will not leave any track in the ID but the charged electrons and positrons will.
- Jets and hadrons are identified from the fact that they deposit energy in both the EM and the hadronic calorimeters. Charged and neutral hadrons, such as the proton and the neutron, can be distinguished from the fact that the uncharged hadrons will not leave a track in the ID.
- Muons leave tracks in all of the detector until they reaches the Muon Spectrometer where they will be bent in the magnetic field and measured before they escape the detector. They are identified from information from the ID and the Muon Spectrometer.
- Neutrinos are the only particles that ATLAS cannot detect: they will traverse out of the detector without leaving a signal, and are identified as missing transverse momentum. This missing momentum can be reconstructed using the fact that the protons only have momenta in the z-direction and by taking a negative vector sum of the momenta of all reconstructed objects in the event.

Chapter 4

Higgs Boson Phenomenology

Once the mass of the Higgs boson (m_H) is known, the SM gives us the information we need to calculate the cross section of different production channels for the Higgs boson. Since it is unstable it will quickly decay into other particles, and these final state particles are what we can detect. The problem is that the same set of particles can often be created by other processes, and all of these make up the background to the Higgs boson signal. In this chapter V denotes a vector boson that can either be a W boson or a Z boson.

4.1 Higgs Boson Production

The Higgs boson can be produced at the LHC through various processes [33]. There are four main production channels of the Higgs boson and they can be seen in table 4.1 together with their cross sections with $m_H = 125$ GeV and $\sqrt{s} = 8$ TeV [34]. In fig. 4.1 we can see the cross section of the different production channels plots as a function of the Higgs boson mass. It should be read at the Higgs boson mass, $m_H = 125$ GeV.

Process	Denotation	Description	σ (in pb) at $m_H = 125$ GeV
$gg \to H$	ggF	Fusion of gluons to produce a quark loop, which then produces a real Higgs boson	$19.3^{+15\%}_{-15\%}$
$qq \to V^*V^* \to qq + H$	VBF	Creation and fusion of oppositely charged W, or of Z bosons to a Higgs boson	$1.58^{+3\%}_{-2\%}$
$q\bar{q} \to V + H$	WH/ZH	Quark and anti-quark merge to form a virtual vector boson that then decays to a Higgs boson and a vector boson	$0.7^{+4\%}_{-5\%}/0.41^{+6\%}_{-6\%}$
$gg, q\bar{q} \to q\bar{q} + H$	ttH	Two gluons produces a quark–antiquark pair and through the exchange of a quark, Higgs boson	$0.13^{+12\%}_{-18\%}$

Table 4.1: The four main production channels of the Higgs boson and their cross sections (in pb) at $m_H = 125$ GeV and at center of mass energy $\sqrt{s} = 8$ TeV.

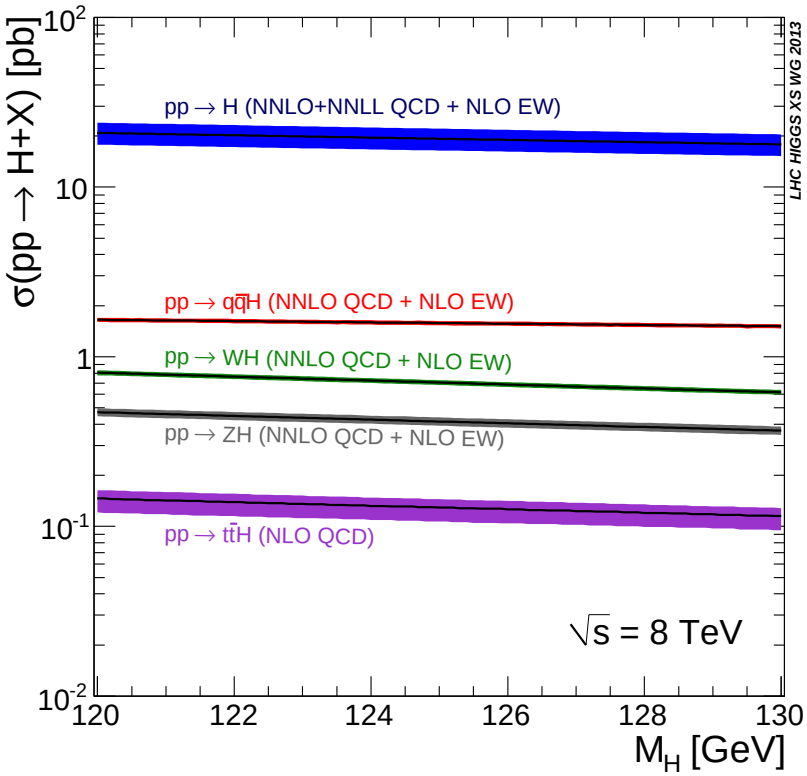

Figure 4.1: Higgs boson production cross sections, as a function of mass around the measured value of $m_H = 125$ GeV (Image courtesy of CERN).

As can be seen in fig. 4.1, gluon-gluon fusion (ggF) is the dominant process. The Feynman diagram for this process can be seen fig. 4.2. Since gluons are massless they cannot couple to the Higgs boson directly, instead the production proceeds through triangular quark loops. The loop is dominated by the top quark because the Higgs boson couples more strongly to heavy particles. The second production channel is the Vector Boson Fusion (VBF) process. The Feynman diagram for this process can be seen fig. 4.3. Two quarks each radiate a virtual vector bosons, which in turn annihilate and produce a Higgs boson. The vector bosons can either be a pair of WW or ZZ. It has a lower cross section than the ggF mechanism. The third production process is the Higgs-Strahlung, shown in fig. 4.4. In Higgs-Strahlung a quark and an anti-quark annihilate into a virtual vector boson. This virtual vector boson radiates a Higgs boson and turn into an on-shell vector boson. The fourth main production process is the associated production with two quarks. The Feynman diagram for the process can be seen fig. 4.5. In this process two gluons creates a heavy quark and anti-quark pair and a Higgs boson through the exchange of a heavy quark.

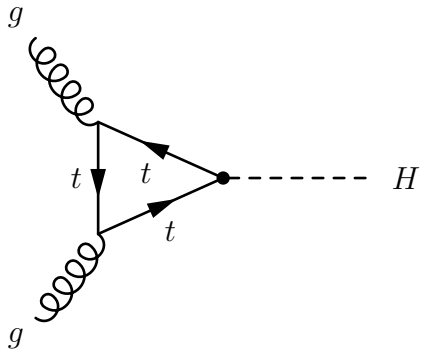

Figure 4.2: A Feynman diagram of the production of a Higgs boson via the gluon-gluon fusion process. In the loop, the dominant contribution is from the t quark.

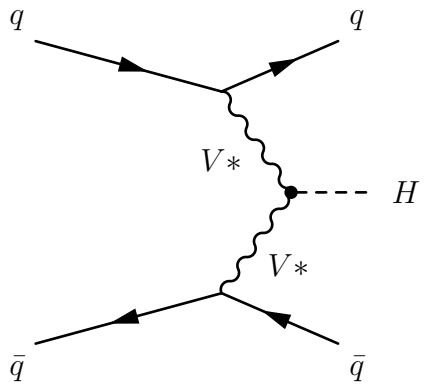

Figure 4.3: A Feynman diagram of the production of a Higgs boson via the vector boson fusion process.

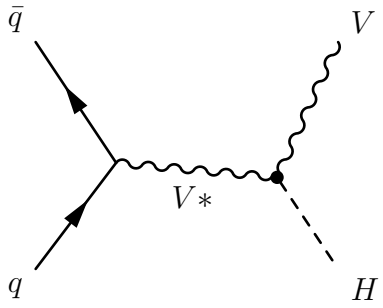

Figure 4.4: A Feynman diagram of the production of a Higgs boson via the Higgs-Strahlung process.

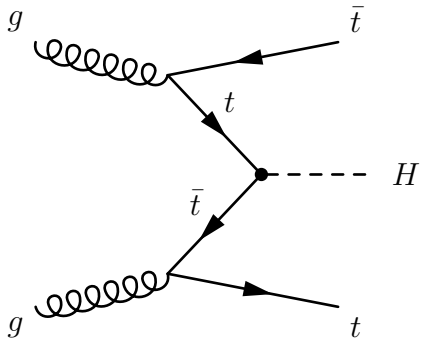

Figure 4.5: A Feynman diagram of the production of a Higgs boson via the associated production with two quarks process.

4.2 Higgs Boson Decay

The Higgs boson decay branching ratios can be seen in fig. 4.6, for mass values around $m_H = 125$ GeV. The Higgs boson can decay to most types of particles but since the strength of the coupling between the Higgs boson and other particles is proportional to their masses it is more likely to decay into heavy particles. This favours heavy quark anti-quark pairs such as $b\bar{b}$, however since they will hadronize and create jets in the detector, and since di-jet production is such a common process, it is very difficult to distinguish the signal from the background. The Higgs boson does not have high enough mass to decay into two real weak vector bosons, but it can decay into one real and one virtual, V^*. As can be seen in the figure, the WW^* decay is more dominant than the ZZ^* decay. This is because of the mass difference and because WW has two degrees of freedom (W^+W^- and W^-W^+). One very important decay channel is $H \to \gamma\gamma$. This decay channel is possible because the decay goes through a vector boson or heavy quarks loop. It has a low branching ratio but gives a clear signal and it is possible to completely reconstruct the mass.

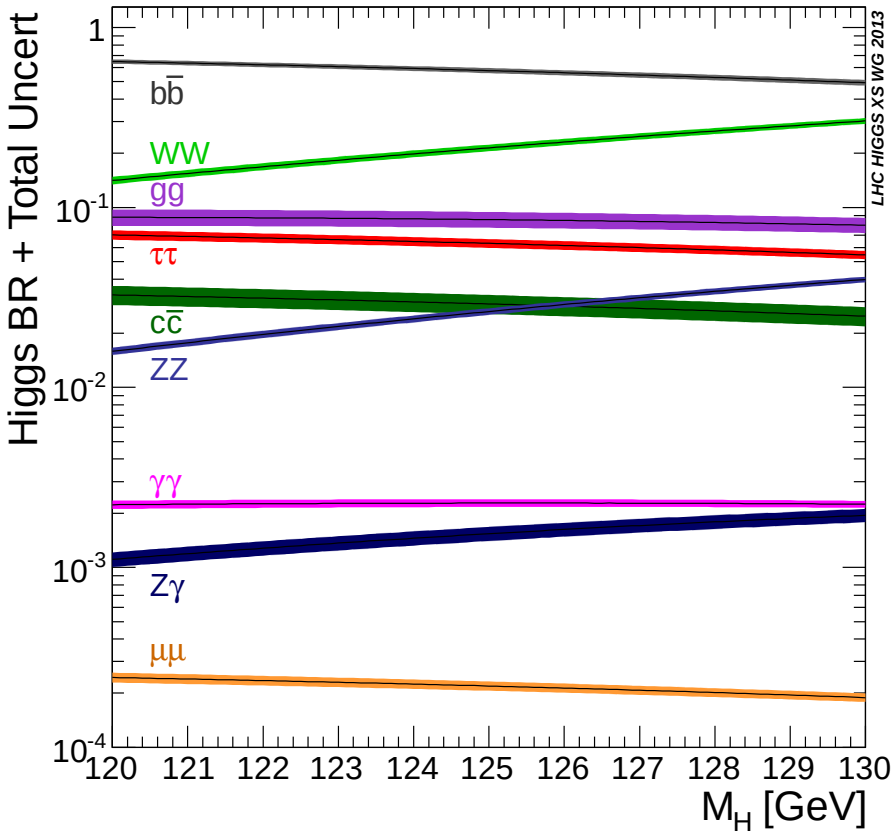

Figure 4.6: Higgs boson decay branching ratios, as a function of mass around the measured value of $m_H = 125$ GeV (Image courtesy of CERN).

The most studied decay channels at ATLAS are the ZZ, WW, and the $\gamma\gamma$ decay modes. The weak vector bosons will decay almost immediately into other particles. The final state particles of the ZZ that are considered are $\ell^+\ell^- q\bar{q}$, $\ell^+\ell^-\nu\bar{\nu}$, and $\ell^+\ell^-\ell^+\ell^-$. The final state particles of the WW is either $\ell^\pm\nu q\bar{q}$, or $\ell^+\nu\ell^-\nu$. It is difficult to study the decays that have quarks as the final state particles because they will create jets and it is very difficult to distinguish the events from the background. Instead, the decays with leptonic final state particles are used in the search for the Higgs boson. They have the nice feature of low background and in the case of $ZZ \to \ell^+\ell^-\ell^+\ell^-$ it is possible to reconstruct the Higgs boson mass completely.

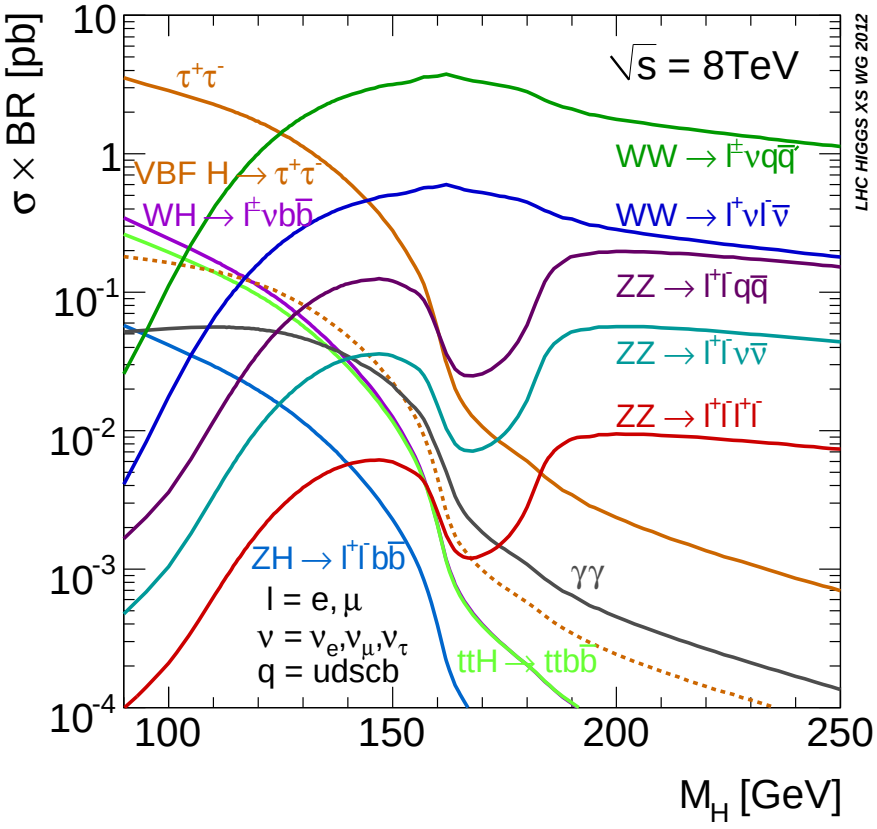

Figure 4.7: Cross section times branching ratio, as a function of mass around the measured value of $m_H = 125$ GeV (Image courtesy of CERN).

Chapter 5

The Higgs Boson Search in the $H \to WW^* \to \ell\nu\ell\nu$ Decay Channel

This chapter presents the analysis done by the ATLAS group that investigates the $H \to WW^* \to \ell\nu\ell\nu$ decay channel, where $\ell = e, \mu$. Details are given about how the background estimation and the cut based $H \to WW^*$ analysis is done using data from the first run of the LHC. The data was collected by ATLAS during 2011 and 2012 and corresponds to an integrated luminosity of 4.6 fb^{-1} collected at a center-of-mass energy $\sqrt{s} = 7$ TeV and 20 fb^{-1} collected at $\sqrt{s} = 8$ TeV. The analysis presented here is described in a note [35], which has not been published but has been internally reviewed, and is an update from the previous version which was described in a conference note for Moriond in 2013 [36].

The Higgs boson at mass $m_H = 125$ GeV is mainly created via the gluon-gluon fusion (ggF) and vector boson fusion (VBF) processes described in section 4.1. The Feynman diagrams of these processes are depicted in fig. 5.1 and fig. 5.2. The mass of the Higgs boson is too low to create two real W bosons in the decay, which means that one of them will be a virtual boson. The W bosons will quickly decay and the final state in this channel is characterized by two isolated leptons with opposite signs of their electric charge. When the final leptons are ee or $\mu\mu$ it is called the same flavour (SF) channel and when they are $e\mu$ or μe it is called the different flavour (DF) channel. These two leptons will be accompanied by missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) that comes from the two neutrinos that escape the detector without leaving a signal. However, events passing this basic selection are dominated backgrounds and one of the most important challenges in the Higgs boson reasearch is to estimate these backgrounds.

To reduce the risk of bias, based on the observed data, the analysis is first done *blinded*. This means that the data in the signal region, the region with high expected signal to background, have been removed from the analysis. When the analysis has been optimized using only the other regions, the data in the signal region is *unblinded* and thus avaliable to the analysis. All plots in this chapter are created using the HWWAnalysisCode, a part of the Common Analysis Framework that is specific to the $H \to WW^*$ analysis. The HWWAnalysisCode takes ntuples with data and Monte Carlo (MC) simulations as inputs and gives cut flows and histograms as output.

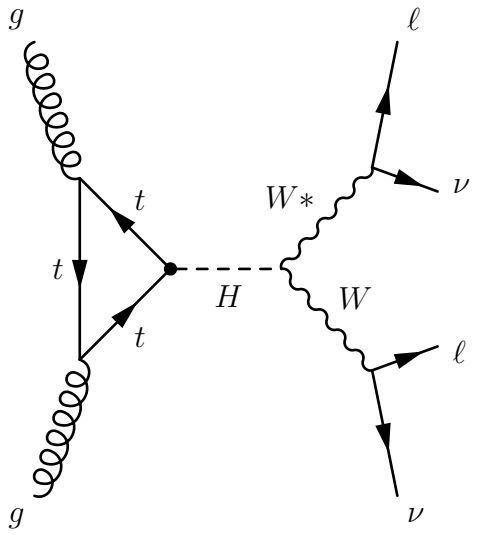

Figure 5.1: A Feynman diagram of the decay of a Higgs boson to two leptons and two neutrinos, where the Higgs boson was created via the gluon-gluon fusion process.

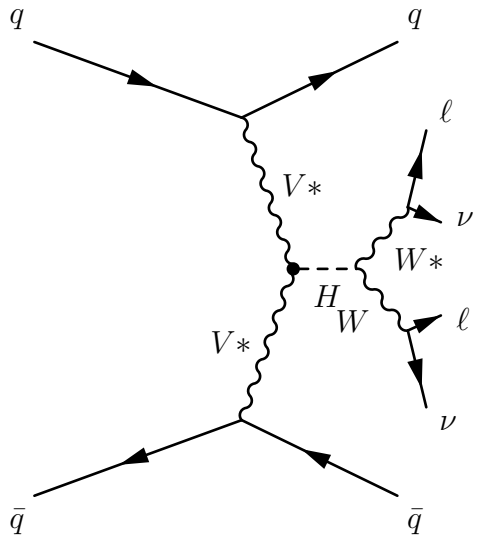

Figure 5.2: A Feynman diagram of the decay of a Higgs boson to two leptons and two neutrinos, where the Higgs boson was created via the vector boson fusion process.

5.1 $H \to WW^*$ Signature

Signal candidates in the $H \to WW^* \to \ell\nu\ell\nu$ decay channel, where $\ell = e, \mu$, will in their final state have two oppositely charged leptons (ee, $\mu\mu$, $e\mu$, or μe), where the one that is written first is the leading lepton and will have higher transverse momentum than the sub-leading lepton. Only electrons and muons are considered because taus decay very quickly into other particles. In the final state there will also be a high missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$) from the two neutrinos that leave the detector without leaving a signal. Because of the two neutrinos it is not possible to reconstruct the invariant mass of the Higgs boson. Instead an alternative variable is used, the transverse mass defined by:

$$m_{\mathrm{T}}^2 \equiv (E_{\mathrm{T}}^{ll} + E_{\mathrm{T}}^{miss})^2 + |\boldsymbol{p}_{\mathrm{T}}^{ll} + \boldsymbol{E}_{\mathrm{T}}^{\mathrm{miss}}|^2, \tag{5.1}$$

where $E_{\mathrm{T}}^{ll} \equiv \sqrt{|\mathbf{p}_{\mathrm{T}}^{ll}|^2 + m_{ll}^2}$.

Since the Higgs boson is a spin zero particle and the W bosons have spin one, the spin of the two W bosons in the decay $H \to WW^*$ must antialign to conserve angular momentum. This fact will make the two leptons emerge in the same direction in the decay, with the two neutrinos going in the opposite direction.

5.1.1 Backgrounds to the $H \to WW^*$ Signal

This section describes the five major backgrounds: top, WW, other diboson processes, Z+jets, and W+jets and discusses how they are reduced. The dominant background depends on jet multiplicity and whether the leptons are SF or DF. Some of the backgrounds, namely the top, WW, and other diboson have real leptons and real $E_{\mathrm{T}}^{\mathrm{miss}}$ in the final state while the others require either *fake* leptons or *fake* $E_{\mathrm{T}}^{\mathrm{miss}}$ to resemble a signal event. In this context, fake means that the object has been misidentified.

Top Background

The top background includes both $t\bar{t}$ and single top (tW, tb, and tqb) production. The top quark decays through the electroweak interaction to a W-boson and predominantly a bottom quark. This is because decays into down or strange quarks are suppressed by their respective CKM elements [37]. The top quark is the only quark that has enough mass to decay into a real W boson. The experimental signature is the decay products of the W boson, $\ell\nu$ or jets, together with a jet from the bottom quark. Since this background has W bosons in the final state, it is very difficult to distinguish it from the Higgs signal and it forms a significant background, especially in the case of Higgs events that contain jets. The top background is suppressed by b-tagging of the jets. It is a technique that is used to separate jets originating from bottom quarks from jets originating from lighter quarks or gluons. Bottom quarks have a relatively long lifetime which means that they will transverse a small distance in the detector before it decays, creating a second vertex. This second vertex can be measured by the ID.

The Standard Model WW Background

The SM WW background is large and since it has the same decay products as the $H \to WW^*$ it is difficult to distinguish the signal from the background. It is reduced using kinematical properties of the decay of the Higgs boson.

Other Dibosons Backgrounds

Other dibosons, $WZ/ZZ/W\gamma$, also constitute a background to Higgs events. Background from WZ arises when the Z boson decays into two leptons. Then the final state will have three leptons and real $E_{\mathrm{T}}^{\mathrm{miss}}$ and can resemble a Higgs event when one lepton is not identified. ZZ constitutes a background when one Z boson decays into two SF leptons and the other into two neutrinos. The $W\gamma$ constitutes a background when the photon is misidentified as an electron.

Z+jets Background

In this background a Z boson or virtual photon, that may be accompanied by any number of jets, decays into two SF leptons with opposite electric charges, therefore this background is mostly important in the SF channel. These events do not contain real $E_{\mathrm{T}}^{\mathrm{miss}}$ but can resemble a Higgs event when the energy of the event is not measured correctly, and thus fake $E_{\mathrm{T}}^{\mathrm{miss}}$ is detected. The Z+jets background also contributes to the DF channel if the Z boson decays into two tau leptons that in turn decay into electrons and muons and real $E_{\mathrm{T}}^{\mathrm{miss}}$. This background is suppressed by requiring a large $E_{\mathrm{T}}^{\mathrm{miss}}$ and the Z veto, the requirement that the invariant mass of the two leptons does not fall close to the mass of the Z boson.

W+jets and QCD multijets Background

When a W boson is produced in combination with jets it can resemble a Higgs event if one of the jets produces or is misidentified as a lepton: a fake lepton. These events contain one real lepton and real $E_{\mathrm{T}}^{\mathrm{miss}}$ from the decay of the W boson. QCD multijets are a similar, but smaller, background to Higgs events. In this case there must be two fake leptons and fake $E_{\mathrm{T}}^{\mathrm{miss}}$.

5.2 Event Selection

The event selection is divided into two parts: first the common preselection and the branch division, where the analysis is split into different branches depending on the jet multiplicity, and then background rejection and topological cuts. The cuts are sometimes different depending on the flavour of the final leptons. This is because the composition of backgrounds differs if the final state leptons are SF or DF. The composition of backgrounds also depends on the number of jets (N_{jets}).

5.2.1 Common Preselection and Branch Division

At first, several preselection cuts are applied to select possible signal events from the large amount of events in the pp collisions. A summary of the preselection cuts can be seen in table 5.1. The first cut is the requirement that the event should have exactly two isolated leptons whose primary vertex must be at the collision point. The leptons also need to be oppositely charged since the decay of the Higgs boson will have a zero net charge. Since the Higgs boson has high mass, the final leptons in a signal event will have high transverse momentum, and a cut is made to ensure this. But because the Higgs boson will decay into one real and one virtual W boson the leading lepton will have higher transverse momentum than the sub-leading lepton and thus the restriction on the sub-leading lepton is lower than that of the leading lepton. The cut is a lower limit and requires the leading lepton to have $p_{\mathrm{T}} > 22$ GeV and the sub-leading lepton to have $p_{\mathrm{T,sub\text{-}lead}} > 10$ GeV. This reduces the risk of a jet being misidentified as a lepton. Then a lower limit on the two leptons invariant mass is set at $m_{ll} > 12$ GeV in the SF channel and $m_{ll} > 10$ GeV in the DF channel. This suppresses the contribution from leptons originating from virtual photons. In the SF channel a Z veto is applied that requires that the invariant mass of the leptons is not close to the mass of the Z

boson, $|m_{ll} - m_Z| > 15$ GeV. This suppresses the contributions from $Z \to \ell\ell$. The last preselection cut is to require high missing transverse energy from the two neutrinos that escape the detector without leaving a signal. Since the Higgs boson is spin zero, the two leptons in a signal event will emerge with a small angle between them and the two neutrinos will traverse in the opposite direction. Because of this a more refined definition of missing transverse energy is used:

$$E_{\mathrm{T,rel}}^{\mathrm{miss}} = \begin{cases} E_{\mathrm{T}}^{\mathrm{miss}} & \text{if } \Delta\Phi \geq \pi/2 \\ E_{\mathrm{T}}^{\mathrm{miss}} sin\Delta\Phi & \text{if } \Delta\Phi < \pi/2 \end{cases} \tag{5.2}$$

where $\Delta\Phi$ is the angle between the direction of the missing transverse momentum and the closest lepton or jet. This will suppress the events where the missing transverse momentum is in the same direction as the leptons. The limits are set at $E_{\mathrm{T,rel}}^{\mathrm{miss}} > 40$ GeV in the SF channel and $E_{\mathrm{T,rel}}^{\mathrm{miss}} > 20$ GeV in the DF channel.

Cuts	Same Flavour (ee, $\mu\mu$)	Different Flavour ($e\mu$, μe)
Number and sign of leptons	Exactly two oppositely signed leptons	
Lepton p_{T}	$p_{\mathrm{T,lead}} > 22$ GeV, $p_{\mathrm{T,sub\text{-}lead}} > 10$ GeV	
m_{ll}	> 12 GeV	> 10 GeV
Z veto	$\lvert m_{ll} - m_Z \rvert > 15$ GeV	-
$E_{\mathrm{T,rel}}^{\mathrm{miss}}$	> 40 GeV	> 20 GeV

Table 5.1: A summary of the preselection cuts used in the $H \to WW^*$ analysis. Divided into the same flavour and different flavour channels.

After the common preselection, the analysis is split into different branches depending on jet multiplicity. The 0-jet branch, the 1-jet branch and the $\geq$ 2-jet branch, where the 0-jet branch is the most sensitive in the search for the Higgs boson. The likelihood of different number of jets depends on the production process. In the 0-jet and 1-jet branch the predominant production process is the ggF. Most of the Higgs events produced by ggF will be in the 0-jet branch, but the ggF production channel can also produce jets by radiating quarks or gluons and these events usually fall in the 1-jet branch. In the $\geq$ 2-jet branch the predominant production process is the VBF, where two highly energetic jets from the two final state quarks appear naturally.

5.2.2 Background Rejection and Topological Cuts

The background composition depends on the jet multiplicity, as can be seen in fig. 5.3. We will not describe all of the cuts that are done at this stage in the analysis, instead we have chosen a selection that is very central to the analysis. A summary of those can be found in table 5.2.

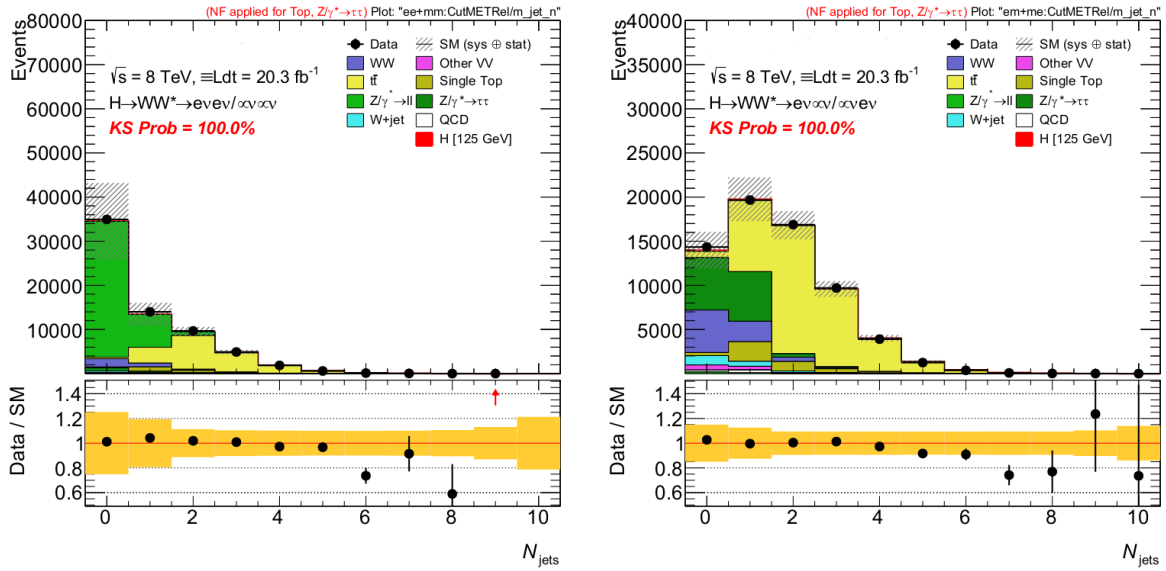

Figure 5.3: Jet multiplicity after the preselection cuts, divided into the combined same flavour (left) and different flavour (right).

0-Jet Branch

The first cut is the jet veto that selects only 0-jet events, which removes a lot of the top background. The next cut is the requirement that the dileptons transverse momentum (p_{T}^{ll}) is big, which reduces the amount of Z+jets background. This is because events with fake $E_{\mathrm{T}}^{\mathrm{miss}}$ usually have small p_{T}^{ll}. After that a cut is made that ensures that the angle between the leptons and the direction of the E_T^{miss} ($\Delta\Phi_{ll,MET}$) is big. This removes events where this angle is small which cannot be the case in a Higgs decay. There is also a cut on a variable called the soft hadronic recoil (f_{recoil}). This variable is a measure on the soft hadronic activity opposite the leptons that gives rise to the fake $E_{\mathrm{T}}^{\mathrm{miss}}$ in Z+jets events, and a cut on this variable further reduces this background. After that kinematic properties are exploited to separate $H \to WW^*$ from the WW background. Since the Higgs boson is spin zero, when it decays into two W boson their spins will be anti-aligned. This causes the two final leptons to favour traversing in the same direction with a small angle between them, while the direction of the neutrinos will be in the opposite direction. Two leptons from background processes favours traversing back to back. By exploiting this, it is possible to cut out a lot of background. To use this a number of cuts are made. One cut is the requirement that the angle between the two leptons ($\Delta\Phi_{ll}$) is small. Another cut is that the dilepton invariant mass (m_{ll}) is small.

1-Jet Branch

To reduce the amount of top background, which will have a jet originating from bottom quarks, a b-jet veto is applied. This means that all events that contain any b-tagged jets are rejected. To reduce the amount of background from $Z \to \tau\tau$, a veto is applied that removes events with invariant mass close to the Z boson. A cut on the f_{recoil} variable is made to further reduce the amount of background from Z+jets. The same way as the 0-jet branch cuts are then made on the m_{ll} and $\Delta\Phi_{ll}$ to separate $H \to WW^*$ from the WW background.

$\geq$ 2-Jet Branch

Events with two or more jets have less signal from the ggF production, so the analysis is therefore more focused to look for Higgs events from VBF production. In the same way as in the 1-jet branch a b-jet and $Z \to \tau\tau$ veto is applied. A VBF cut veto is applied that uses the special characteristics from a VBF event, that there are two jets with high p_{T} that have high separation in η (Y_{jj}) and high invariant mass (M_{jj}). The VBF cut veto also rejects events that have a third jet in the middle of the two, which is called central jet veto (CJV). It also rejects events where the leptons are outside the jets, and this is called the outside lepton veto (OLV). In the same way as for the 0-jet branch cuts are then made on the m_{ll} and $\Delta\Phi_{ll}$ is made to separate $H \to WW^*$ from the WW background.

Selection	$N_{\mathrm{jet}} = 0$	$N_{\mathrm{jet}} = 1$	$N_{\mathrm{jet}} \geq 2$
p_{T}^{ll}	> 30 GeV	-	-
$\Delta\Phi_{ll,MET}$	$> \pi/2$ radians	-	-
b-Jet Veto	-	$N_{\text{b-jet}} = 0$	$N_{\text{b-jet}} = 0$
$Z_{\tau\tau}$ Veto	-	(DF) $m_{\tau\tau} < m_Z - 25$ GeV	(DF) $m_{\tau\tau} < m_Z - 25$ GeV
VBF cut Veto	-	-	$\Delta Y_{jj} > 3.6$ & $M_{jj} > 600$ GeV & CJV & OLV
f_{recoil}	< 0.1	< 0.1	-
m_{ll}	< 55 GeV	< 55 GeV	< 55 GeV
$\Delta\Phi_{ll}$	< 1.8 radians	< 1.8 radians	< 1.8 radians

Table 5.2: A summary of a selection of background rejection and topological cuts used in the $H \to WW^*$ analysis, divided into three categories depending on jet multiplicity.

5.3 Background Estimation and Control Regions

One common method to estimate the amount of background is to correct the MC predictions by the number of background events in a control region (CR). The CRs are ideally a region that is dominated by a particular background. The CR is selected using similar methods as the event selection but with other criteria specific to the target background. Using the CR, a factor is then used to extrapolate to the signal region (SR). Even though the CR is chosen in a place where the desired background is dominating, the CR is often contaminated with other backgrounds that have to be reduced.

5.3.1 Top Background

The majority of the top background events include jets, but because the top background has a very high cross section it is a large background in all jet branches. To get an estimate of the top background in the 0-jet branch, a CR where top is dominating is

used. To get the estimate of the number of top events in this CR other backgrounds and also the Higgs signal is subtracted from the number of events that passes the preselection cuts, up to but not including the jet veto cut. This number is then multiplied by the probability that a top event is reconstructed with zero jets in simulation.

To get an estimate of the top background in the 1- and $\geq$ 2-jet branches, CRs are defined by dropping the requirements on the $\Delta\Phi_{ll}$ and m_{ll} and reversing the b-jet veto. These CRs are dominated by top events, but as can be seen in fig. 5.4, they are contaminated by other backgrounds.

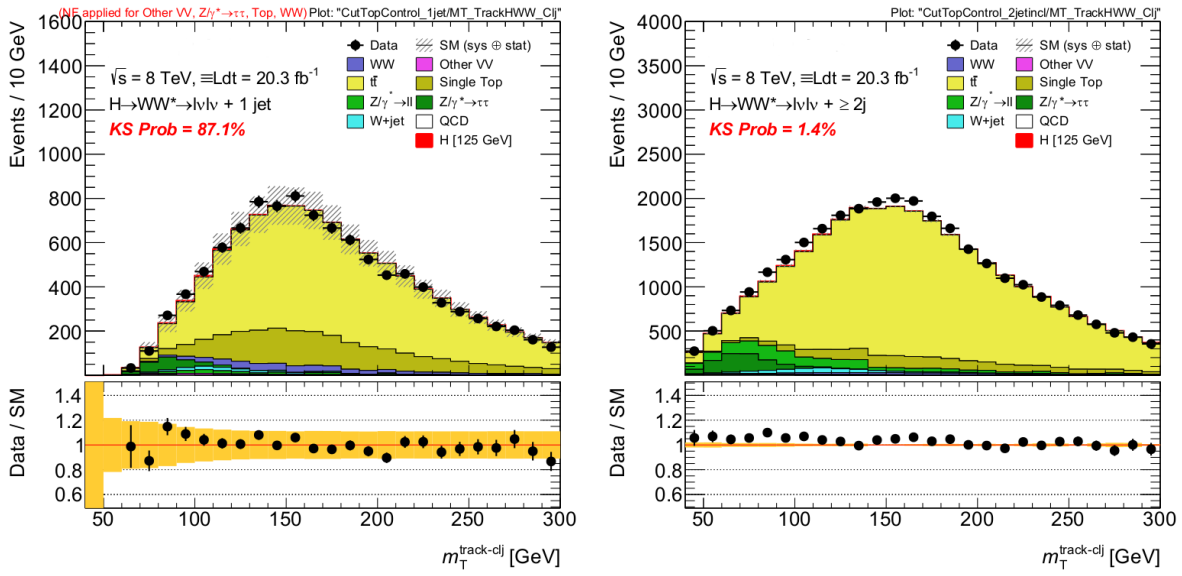

Figure 5.4: Kinematic distributions of the transverse mass in the top CRs, divided into $N_{\text{jets}} = 1$ (left) and $N_{\text{jets}} = 2$ (right).

5.3.2 The Standard Model WW Background

The SM WW is the dominant background in the 0-jet branch, close in magnitude to the top background in the 1-jet branch, and a lot smaller but still a significant background in the $\geq$ 2-jet branch. The estimate in the 0- and 1-jet branch is taken using CRs, these CRs are dominated by WW but are contaminated by other backgrounds, as can be seen in fig. 5.5. The degree of contamination differs a lot between the 0- and 1-jet branches, where the latter CR contains a lot of other backgrounds, predominantly from top. The background estimate in the $\geq$ 2-jet branch is taken from MC simulations only since it is not possible to find a CR that is not very contaminated with other backgrounds.

The SM WW CR is defined by applying the preselection cuts, as well as cuts that are specific to this CR. A summary with some of the cuts specific to the WW CR can be seen in table 5.3. One of the differences between the SR and CR is the range of the invariant mass of the dileptons. It is set as close as possible to the SR to minimize the systematic error that comes from the extrapolation to the signal region. To reduce the W+jets contamination of the CR, a cut is made on the transverse momentum on the subleading lepton. In the 0-jet branch a cut is made on the angle between the two leptons to reduce the contamination from $Z \to \tau\tau$.

Cut	$N_{\text{jet}} = 0$	$N_{\text{jet}} = 1$
m_{ll}	$55 < m_{ll} < 110$ GeV	$80 < m_{ll}$ GeV
$p_{\text{T}}^{\text{sublead}}$	> 15 GeV	
$\Delta\Phi_{ll}$	< 2.6 radians	-

Table 5.3: A summary of the cuts specific to the WW control region, divided into two categories depending on jet multiplicity.

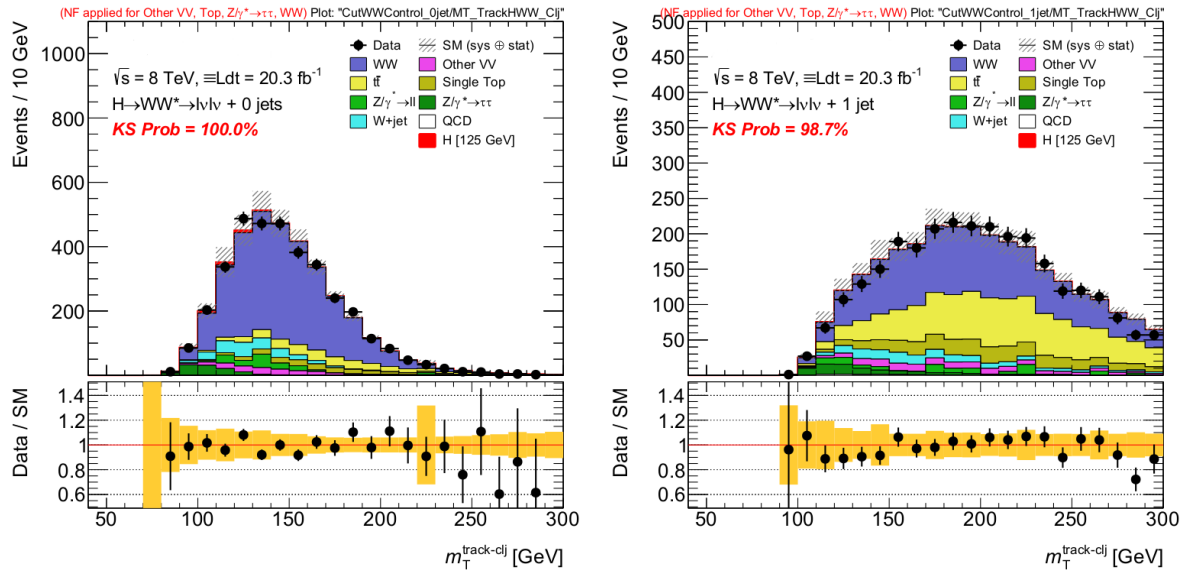

Figure 5.5: Kinematic distributions of the transverse mass in the SM WW control region, divided into $N_{\text{jets}} = 0$ (left) and $N_{\text{jets}} = 1$ (right).

5.3.3 Z+jets Background

The Z+jets backgrounds is estimated using control regions that depends on the jet multiplicity. How they are estimated depends on the final state lepton flavours. When the final state leptons have the same flavour the amount of background is difficult to estimate because fake $E_{\text{T}}^{\text{miss}}$ may not be accurately described in simulations. Instead the background is estimated from data using a specific method, called the *Pacman* method which uses the variable f_{recoil} described in sec. 5.2.2. The method exploits the fact that the distribution of this variable looks different in the $Z/\gamma^* \to ll$ events than events with real $E_{\text{T}}^{\text{miss}}$ to estimate the amount of Z+jets background. When the final state leptons have different flavour the Z+jets background originates almost entirely from $Z \to \tau\tau$ events. To estimate the contribution from this background a CR is used. This CR is defined by applying the preselection cuts and the jet veto, as well as cuts use the fact that the leptons in this process are usually emitted back-to-back so the requirement of a large angle between the leptons will reduce other backgrounds.

5.3.4 W+jets

The W+jet background is not well modeled by MC and is instead estimated from data for all jet multiplicities. The W+jets CR is defined by events that fulfill the same criteria as the signal event selection except that one or both of the leptons fail the ordinary lepton selection. Instead they pass a looser identification criteria and are called *anti identified*. The W+jets background is then obtained from a scaling of the events in the CR by a so called *fake factor*. This fake factor is derived from Z+jets data with corrections for expected differences between the jets in the two samples.

5.4 Results of the $H \to WW^* \to \ell\nu\ell\nu$ Analysis

Since the result from this analysis is not yet released the result from the previous version, which is described in a conference note for Moriond in 2013 [36], will be presented in this section.

A summary of the expected (N_{sig} and N_{bkg}) and observed (N_{obs}) yields for the signal and the background processes can be found in fig. 5.6 for the 7 TeV data and in fig. 5.7 for the 8 TeV data. These yields are given in different ranges of the transverse mass depending on jet multiplicity. For $N_{jet} \leq 1$, the range is $0.75m_H < m_T < m_H$ and for $N_{jet} \geq 2$ the range is $m_T < 1.2m_H$. In these tables, the SF and DF channels are combined. In the table to the right the composition of the background is given.

N_{jet}	N_{obs}	N_{bkg}	N_{sig}	N_{WW}	N_{VV}	$N_{t\bar{t}}$	N_t	N_{Z/γ^*}	$N_{W+\text{jets}}$
= 0	154	161 ± 11	25 ± 5	113 ± 10	12 ± 2	5 ± 1	4 ± 1	6 ± 2	21 ± 5
= 1	62	47 ± 6	7 ± 2	16 ± 6	5 ± 1	10 ± 3	6 ± 2	5 ± 2	5 ± 1
≥ 2	2	4.6 ± 0.8	1.4 ± 0.2	0.7 ± 0.2	-	0.7 ± 0.5	0.1 ± 0.1	2.4 ± 0.6	0.3 ± 0.1

Figure 5.6: Summary of the expected and observed yields for the signal and the background processes for the 7 TeV data with the SF and DF channels combined. In the table to the right the composition of the background is given.

N_{jet}	N_{obs}	N_{bkg}	N_{sig}	N_{WW}	N_{VV}	$N_{t\bar{t}}$	N_t	N_{Z/γ^*}	$N_{W+\text{jets}}$
= 0	831	739 ± 39	97 ± 20	551 ± 41	58 ± 8	23 ± 3	16 ± 2	30 ± 10	61 ± 21
= 1	309	261 ± 28	40 ± 13	108 ± 40	27 ± 6	68 ± 18	27 ± 10	12 ± 6	20 ± 5
≥ 2	55	36 ± 4	10.6 ± 1.4	4.1 ± 1.5	1.9 ± 0.4	4.6 ± 1.7	0.8 ± 0.4	22 ± 3	0.7 ± 0.2

Figure 5.7: Summary of the expected and observed yields for the signal and the background processes for the 8 TeV data with the SF and DF channels combined. In the table to the right the composition of the background is given.

The results are given from the combined data of both the SF and the DF channel, for all jet multiplicities, from the 7 TeV and the 8 TeV data. For the signal with $m_H = 125$ GeV the expected significance is 3.7σ and an observed significance is 3.8σ. The excess strength for this signal is calculated to be:

$$\mu_{obs} = 1.01 \pm 0.21(\text{stat.}) \pm 0.19(\text{theo. syst.}) \pm 0.12(\text{expt. syst.}) \pm 0.04(\text{lumi.}) = 1.01 \pm 0.31, \tag{5.3}$$

where $\mu = 1$ corresponds to the Standard Model hypothesis and $\mu = 0$ corresponds to no signal.

Chapter 6

Rejection of Top Background in the WW 1-jet Control Region

In the previous chapters we have described the cut based analysis currently done by the ATLAS collaboration. It is possible that this could be improved by using multivariate machine learning to separate signal from background. One of the most common machine learning methods is the Boosted Desicion Tree (BDT). BDTs have been used in high energy physics analysis before with good success [38, 39].

To estimate a certain background process, control regions in data are used. These control regions are dominated by the background process in question but they are often contaminated by other backgrounds. These other backgrounds need to be separated from the background in question in order to get a pure CR, which gives a better background estimation. Two of the more difficult backgrounds to separate are from top quarks and W bosons, since they can have the same final state particles. They will have real missing transverse energy and two leptons. The goal of this study is to find possible discriminating variables between top and WW processes in the WW 1-jet CR using TMVA. As discussed in sec. 5.3.2, this CR is very contaminated by top background. Different BDT methods will be tested to find which one works best. In this study we limit ourselves to the DF channel because the SF channel is much more contaminated with backgrounds from Z bosons. This study is based on one performed by Jelena Jovićević in October 2013.

6.1 Boosted Decision Trees

A Decision Tree (DT) [40, 41] has a two dimensional tree structure, see fig. 6.1. Starting with a root node, it is split into two secondary nodes that correspond to the events that pass or fail the root node criterion, and are therefore either classified as signal or background. This process of splitting the nodes in two continues until a stop criterion is fulfilled. This stop criterion can for example be a maximum depth of the tree. This means that an event that in one node is classified as background will not be discarded, instead it will continue to be analyzed. The last nodes are called "leafs" and will contain many events. A majority vote decides if the leaf is classified as signal or as background and the leaf is given a weight between -1 (all background) and 1 (all signal) depending on the composition of the events in the leaf. Every split uses the variable that best

separates signal from background to determine the node criterion, which means that the same variable can be used multiple times whereas some may never be used. This makes DTs unaffected by the inclusion of variables that are poor at separating signal from background, unlike many other machine learning methods. When using DTs it is necessary to have MC samples of both background and signal. These samples are then used to train the DTs to differentiate between signal and background. Training a tree is the process where the tree is built from the top down, and the criterion for every node is decided. This is done by a check at every node to see which variable and at what cut value gives the most differentiation of signal and background. A problem DTs almost always suffer from is overtraining. Overtraining means that the classifier makes decisions based on statistical fluctuations rather than learning the underlying principles in order to separate signal from background. This leads to a decrease in performance when the classifier is used on an independent test sample. The preferred method of decreasing the risk of overtraining is to not use deep DTs and to demand that there are a lot of events in the leaf nodes. This makes the DT less sensitive to statistical fluctuations in the training sample.

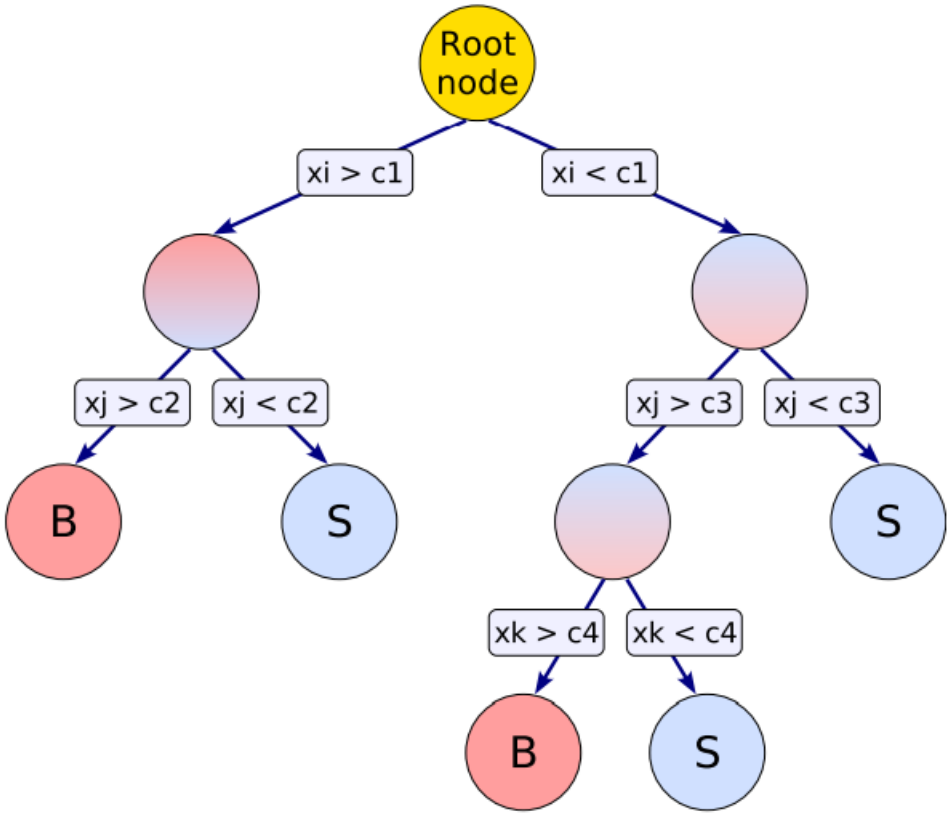

Figure 6.1: Schematic view of a decision tree. Starting with a root node, it is split into two secondary nodes that correspond to the events that pass or fail the root node criterion. This process of splitting the nodes in two continues until a stop criterion is fulfilled [41].

Boosting refers to when instead of one, many DTs are used. The output of the BDT is then simply a weighted average of the individual decision tree outputs, the weight of which is determined by their classification ability. There are several different boosting algorithms, the two we will present here are Adaptive boost (AdaBoost) and Gradient boost. Both of these work similarly. The same training set is used multiple times, and a loss function is minimised to determine the optimal weight parameters in the BDT. The boosting can be applied multiple times, and is performed on one tree at a time. Bagging is in the strict sense not a boosting method, but is also used to improve the performance of the DTs.

- Adaptive boost uses an exponential loss function, which leads to the following reweighting algorithm. At first, all events are given the same weight. If a signal event lands on a background leaf, and vice versa, its original weight is multiplied by a boost weight which is given by:

$$\alpha = \frac{1 - err}{err}, \tag{6.1}$$

  where err is the misclassification rate. After all events are given their weights the weights are renormalized so that the sum of the weights is constant. Then the next tree uses these new weights.

- Gradient boost allows for other loss functions which can improve on the exponential loss function's weaknesses. As a result, gradient boost is less susceptible to overtraining and is less vulnerable to noisy data.

- When using Bagging, the classifier is trained using resampled events, this means that each tree is trained on a subset of the original sample. The resampled events are picked at random, which means that individual events can be picked several times.

6.2 TMVA

The Toolkit for Multivariate Analysis (TMVA) [41] is a package based on ROOT [1] which is a framework written in C++. ROOT was originally developed by CERN to be able to handle and analyze the large amounts of data from the different experiments. TMVA uses multivariate machine learning to differentiate between signal and background. It includes various methods for multivariate classification, including several BDT methods.

TMVA produces multiple tables and plots as a result of training. One of the outputs from TMVA is a correlation matrix displaying the linear correlation coefficients for the signal and background training samples. TMVA also gives a variable ranking as output, both a non-method specific and a method specific ranking of how good the variable is at separating signal from background. This variable ranking is usually not very good when the variables are correlated, because then the classifier will always pick the one that is slightly better, which will result in a low ranking for one of the variables. Another output is a plot of the so-called Receiver Operating Characteristics (ROC) curve: a curve

of background rejection versus signal efficiency. It is used to evaluate the performance of the classifier, and the larger the area below the curve, the better separation of signal and background can be achieved. From the area under the ROC curve, TMVA gives a ranking of the classifier methods that indicates which ones are best at classifying the events. Another output that is used to evaluate the performance of a method is the significance. It is defined as the difference between the methods means for signal and background divided by their quadratic sum of their root-mean-squares and a larger significance means a better classifier. The last output that will be used in this thesis is an overtraining check.

6.3 Rejection of top background in the WW 1 jet control region

This study is divided into two parts. In the first part an ntuple containing the training data set is created and in the second part TMVA is used to train different BDT methods to differentiate between signal and background. The first part was done using a program called "Top_bkg_work.C". It was originally written for older versions of the ntuples and had to be updated to the newest version of the ntuples. As input, this program takes lists of top or WW 1-jet ntuples containing MC simulated events, and outputs a smaller ntuple containing only the branches of interest. This step is necessary since the original ntuples are large and this makes the training slow. The code is run separately for the $e\mu$ and μe channels. These ntuples are then merged into two new ntuples that will be used as input in the TMVA training: one containing WW events (signal) and one containing top events (background).

In the second part a modified version of the program TMVAClassification.C is used. The original program is included in the TMVA analysis package. It splits the input data set into two, one subset for training and one for testing. The events that go into the different sets are chosen by giving each event a random number with mean 0, and all events with a number > 0 goes to the training set. In this case 55377 of 110814 signal events, and 35285 of 70450 background events were used in the training.

The training begins with the creation of a Factory object. The Factory needs at least two input variables and can train multiple multivariate methods in one training. The booked methods are trained on the training subset. The output is a .root file containing all evaluation plots and a weight file in XML format that contains configuration options such as the variables used and the training results for the method. Some of the simulated events have negative event weight, since this is unphysical they are ignored in the training but in order to not bias the results they are used in the testing.

6.3.1 Variables and Methods

Seven different BDT methods were chosen to be tested. They are listed in table 6.1, together with the boost type and some of the tree parameters used. TMVA gives the user the opportunity to set many different tree parameters. One of the parameters is NTrees which corresponds to the number of DTs. Another is MaxDepth which is the

maximum depth of the DTs. It is important that the depth is not too big, as this leads to few events in each level and increases the risk of overtraining. Another criterion to reduce the risk of overtraining BDTs is a stop criterion requiring at least 5% of all events in each leaf.

Method name	Boost type	NTrees	MaxDepth
BDT	AdaBoost	850	3
BDT1	AdaBoost	2000	3
BDT2	AdaBoost	850	5
BDT3	AdaBoost	850	2
BDTB	Bagging	400	-
BDTG	Grad	1000	2
BDTD	AdaBoost + Decorrelation	400	3

Table 6.1: The seven methods used in the first BDT training together with different tree parameters. NTrees is the number of decision trees and MaxDepth maximum allowed depth of the decision tree.

To train the TMVA classifier we start with a set of variables that could be important to differentiate between background and signal. The 15 variables used in the training of the BDTs can be seen in table 6.2. This list includes the transverse mass, m_T, to be able to see which variables are correlated to it. We want to choose variables that are not strongly correlated to m_T since that is the variable used to fit distributions to in the $H \to WW^* \to \ell\nu\ell\nu$ analysis, and a variable strongly correlated to it might already have been accounted for in the analysis. The goal is to use this set of variables to find a subset of four to five variables that best differentiates between signal and background, this is because a set with fewer variables reduces the risk of overtraining and also makes it easier to estimate the uncertainties.

Variable	Description
leadJetPt	The transverse momentum of the leading jet.
leadLepPt	The transverse momentum of the leading lepton.
subleadLepPt	The transverse momentum of the sub-leading lepton.
DPhiLeadLepMET	The angle between the leading lepton and the missing transverse energy.
DPhiSubleadLepMET	The angle between the sub-leading lepton and the missing transverse energy.
DPhill	The angle between the leading and sub-leading lepton.
DRll	The difference in radius between the two leptons.
MT	The transverse mass.
MTlep0	The transverse mass of the leading lepton.
MTlep1	The transverse mass of the sub-leading lepton.
HT	The scalar sum of pTs of all objects in the event.
Mll	The dilepton invariant mass.
nJets_Pt20	The number of jets with a transverse momentum larger than 20 GeV.
MET	The missing transverse energy.
METRel	The relative missing transverse energy.

Table 6.2: The names of the 15 variables used in the BDT training together with their definition.

6.3.2 Variable optimization

At first a training with all seven methods and all 15 variables was conducted. As can be seen in fig. 6.2 and fig. 6.3, several of the variables are highly correlated to m_T. Therefore these variables will not be used in the next sets of variables. From the ROC curve, seen in fig. 6.4, we can see that there is a possibility of rejecting 44% of the background while keeping 90% of the signal. We can also see that the difference in method performance is large. In table 6.3 the non-method specific variable ranking given by TMVA is listed along with the separation. This gives an indication of the impact of the different variables.

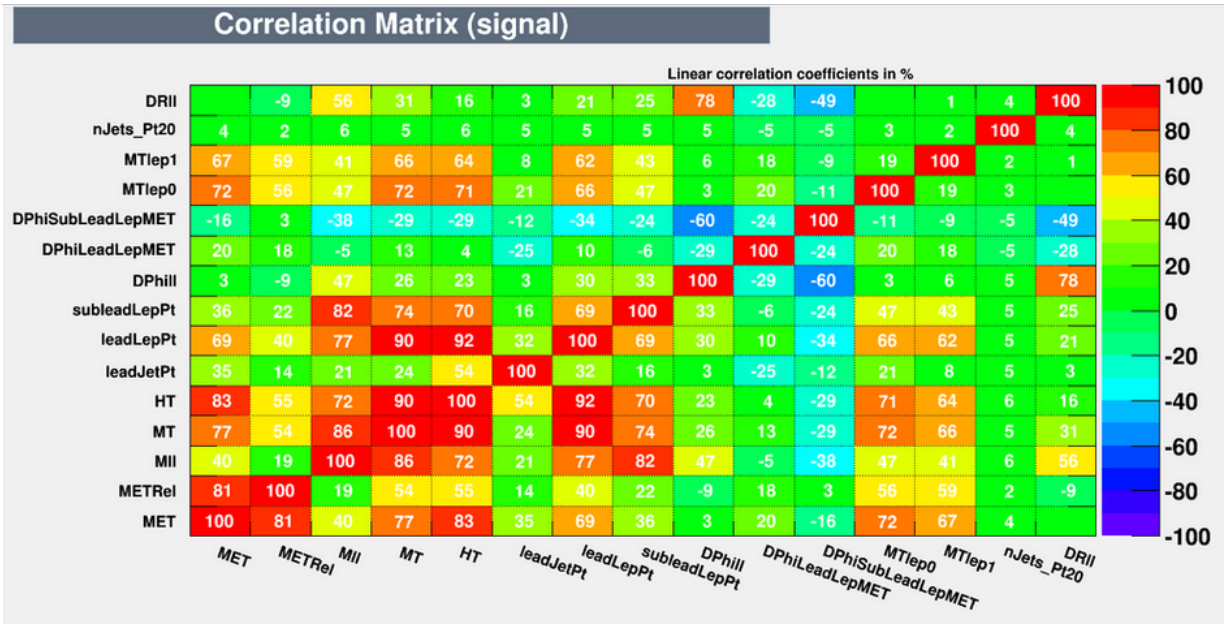

Figure 6.2: Linear correlations in the signal set of all 15 of the discriminating variables. Uncorrelated variables are shown as green.

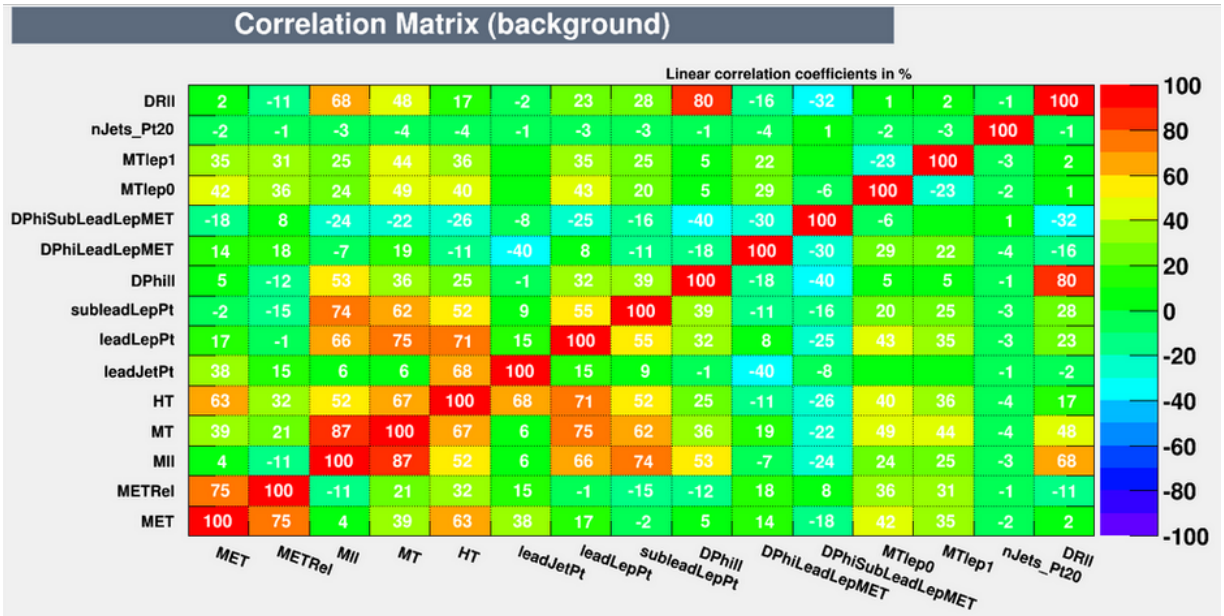

Figure 6.3: Linear correlations in the background set of all 15 of the discriminating variables. Uncorrelated variables are shown green.

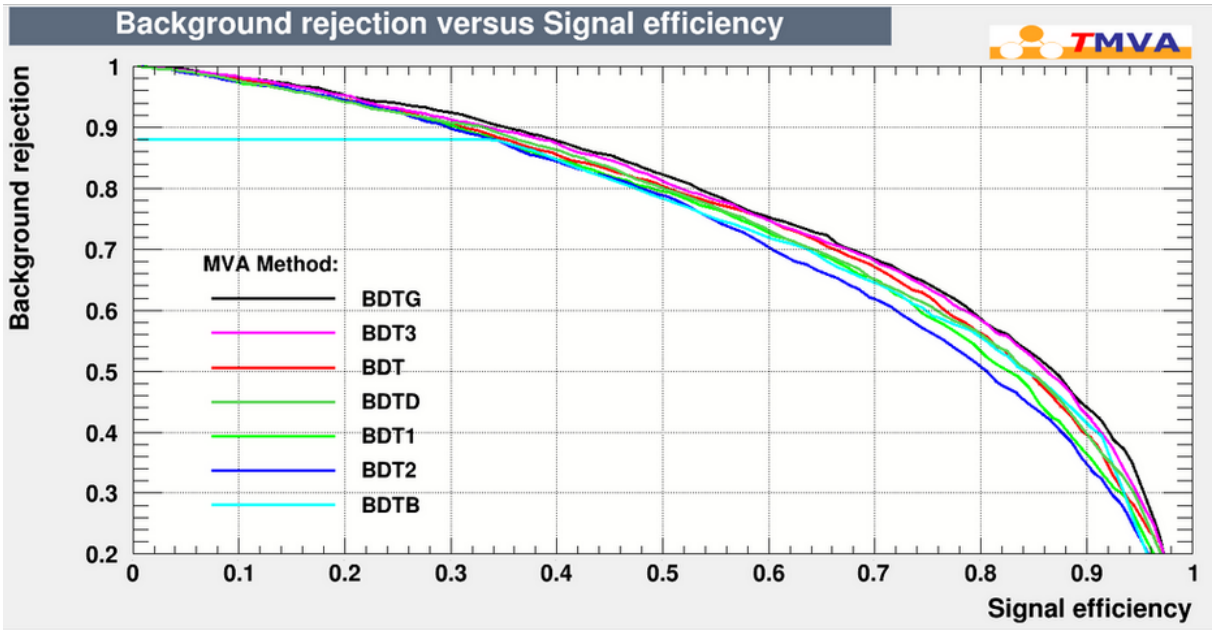

Figure 6.4: ROC curve showing the performance of the different BDT methods listed on the plot, using all 15 of the discriminating variables.

Rank	Variable	Separation
1	leadJetPt	$1.594 \cdot 10^{-1}$
2	HT	$7.297 \cdot 10^{-2}$
3	DPhiLeadLepMET	$4.574 \cdot 10^{-2}$
4	leadLepPt	$2.222 \cdot 10^{-2}$
5	nJets_Pt20	$2.004 \cdot 10^{-2}$
6	MT	$1.839 \cdot 10^{-2}$
7	DPhiSubLeadLepMET	$1.616 \cdot 10^{-2}$
8	Mll	$1.585 \cdot 10^{-2}$
9	MET	$1.543 \cdot 10^{-2}$
10	MTlep1	$1.506 \cdot 10^{-2}$
11	subleadLepPt	$1.363 \cdot 10^{-2}$
12	DPhill	$1.2687 \cdot 10^{-2}$
13	DRll	$1.162 \cdot 10^{-2}$
14	MTlep0	$1.018 \cdot 10^{-2}$
15	METRel	$8.900 \cdot 10^{-3}$

Table 6.3: The non-method specific ranking given by TMVA of the 15 discriminating variables, together with their separation. Higher ranking indicates a more discriminating variable.

A new training was made with the set of variables that are not correlated with m_T (leadJetPt, DPhiLeadLepMET, DPhiSubLeadLepMET, DPhill, DRll, nJets_Pt20). From the results, using the ranking of methods by TMVA, the three most promising methods were chosen (BDTG, BDT3, BDTD), see table 6.4. As can be seen in fig. 6.5 and fig. 6.6 some of the variables are strongly correlated. As mentioned before, this may make the TMVA ranking perform poorly. To solve this and get a correct variable ranking the following $N-1$ iterative procedure, proposed by A. Hoecker, is conducted to optimize the variables: one variable is removed at a time, a new training is performed and the variable that caused the worst result is the best ranked variable. This variable is permanently removed from the set. The procedure is repeated until all variables are ranked. This procedure can have its own problems if one variable alone has a low ranking but gives a good performance together with another one. A comparison between the result of this procedure, the non-method specific ranking given by TMVA, and the method specific ranking given by TMVA can be found in table 6.5. From this we can see that in this case the non-method specific variable ranking from TMVA agrees well with the result from the $N-1$ iterative procedure, only two variables changed ranking. The method specific rankings differ a lot, but they all have in common that the variable nJets_Pt20 performs the worst.

In the output of TMVA, a table with an overtraining check is included. In this table, the efficiency using the training set and the testing set of events is compared at three different background efficiencies (1− background rejection). If the training efficiency is significantly better than the testing efficiency the method is overtrained. From the result, as can be seen in table 6.6, it can be seen that this is a clear case of overtraining where the methods perform better with the training set than the testing set in all methods and in all cases. It is clear that the BDTG method is the least overtrained and the BDTD method is the most overtrained. From the ROC curve seen in fig. 6.7 it looks like it is possible to reject 42% of the background while keeping 90% of the signal.

Rank	Method	ROC Area
1	BDTG	0.737
2	BDTD	0.730
3	BDT3	0.729
4	BDT	0.724
5	BDT1	0.715
6	BDT2	0.707
7	BDTB	0.706

Table 6.4: The ranking of the 7 methods classification abilities, using the 6 discriminating variables not correlated to m_T, given together with the area under each methods ROC curve. Higher ranking indicates a better classifier.

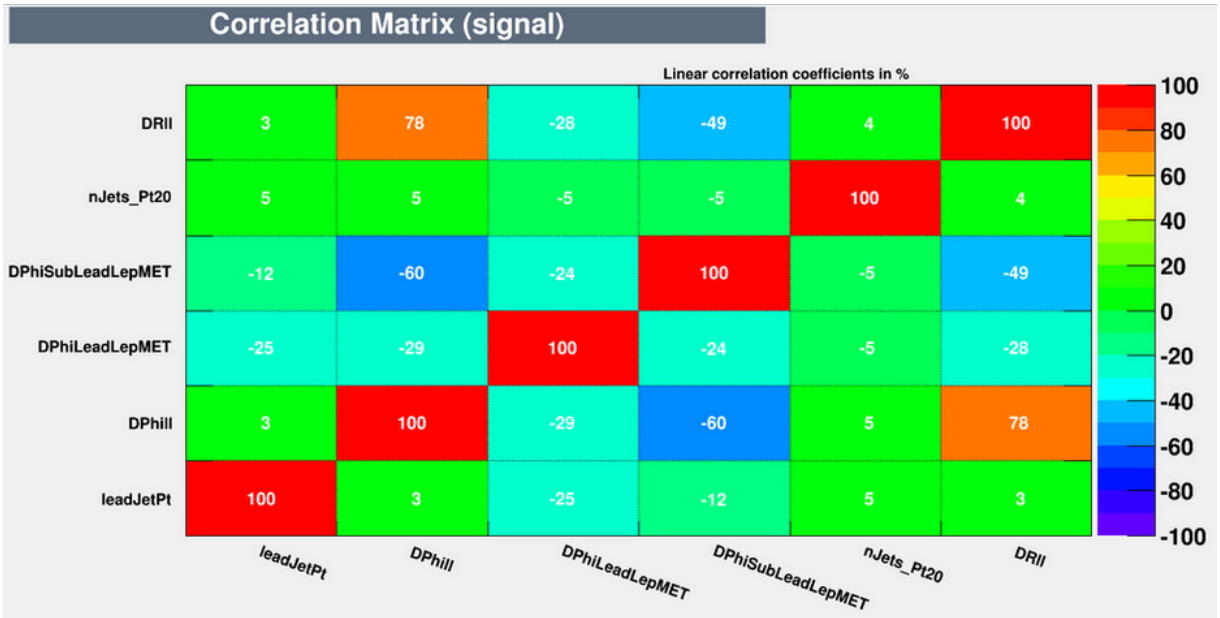

Figure 6.5: Linear correlations in the signal set of the 6 discriminating variables not correlated to m_T. Uncorrelated variables are shown as green.

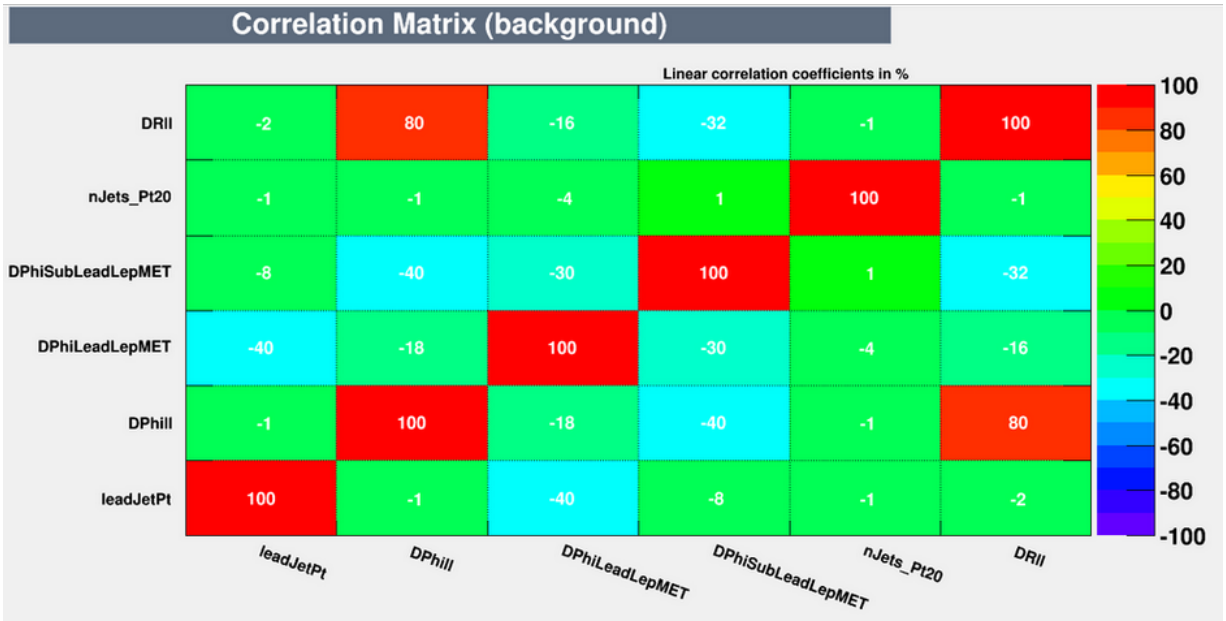

Figure 6.6: Linear correlations in the background set of the 6 discriminating variables not correlated to m_T. Uncorrelated variables are shown as green.

Variable	$N-1$ iterative Procedure	TMVA	BDTG	BDT3	BDTD
leadJetPt	1	1	5	5	1
DPhiLeadLepMET	2	2	3	2	2
DPhiSubLeadLepMET	3	4	2	1	4
nJets_Pt20	4	3	6	6	6
DPhill	5	5	1	3	3
DRll	6	6	4	4	5

Table 6.5: The ranking of the 6 variables not correlated with m_T, from the $N-1$ iterative procedure described in the text, non-method specific ranking from TMVA and method specific ranking from TMVA.

Method:	B=0.01	B=0.10	B=0.30
BDTG	0.047 (0.053)	0.315 (0.323)	0.631 (0.644)
BDTD	0.041 (0.070)	0.285 (0.336)	0.625 (0.651)
BDT3	0.043 (0.053)	0.298 (0.322)	0.621 (0.632)

Table 6.6: Overtraining check via testing efficiency compared to, in parentheses, training efficiency at three different background efficiencies. The method is overtrained if the training efficiency is significantly better than the testing efficiency.

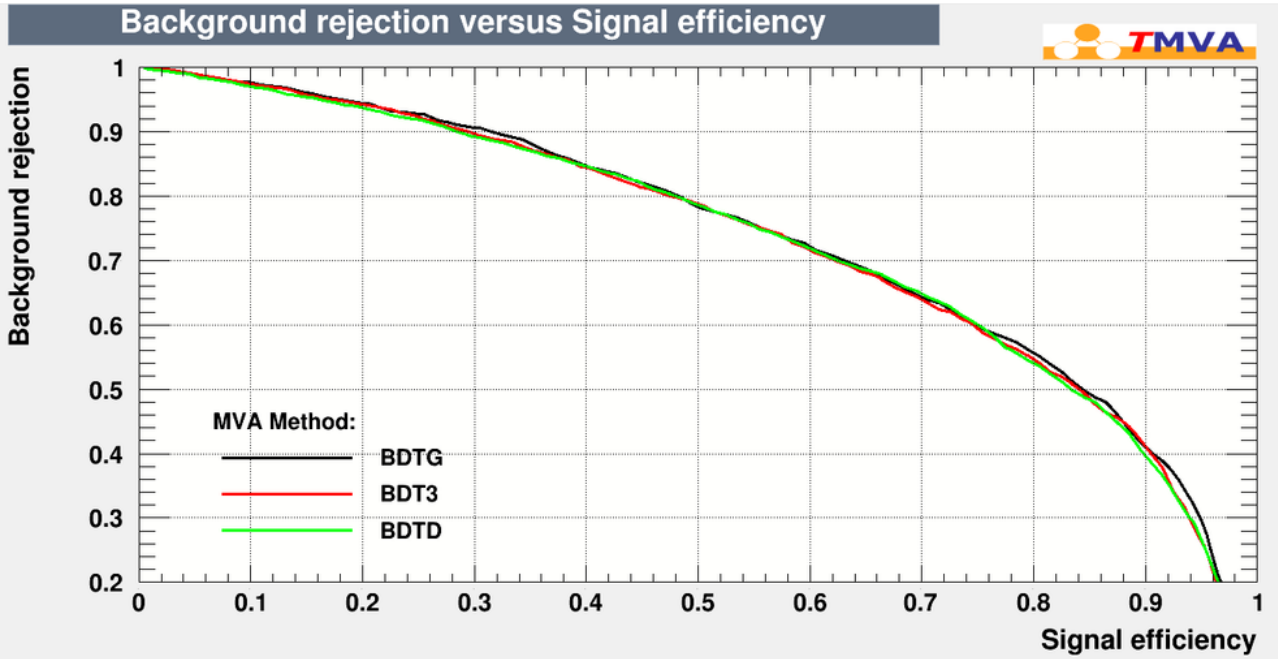

Figure 6.7: ROC curve showing the performance of the three different BDT methods listed on the plot, for the set of 6 variables not correlated to m_T.

New trainings are performed with two sets of five variables. One set was chosen from the top ranked variables from the $N-1$ iterative procedure (leadJetPt, DPhiLeadLepMET, DPhiSubLeadLepMET, DPhill, and nJets_Pt20), while the other set was chosen from the method specific ranking from TMVA (leadJetPt, DPhiLeadLepMET, DPhiSubLeadLepMET, DPhill, and DRll). The variables chosen from the $N-1$ iterative ranking procedure show better performance in both the area under the ROC curve and in significance. This can be seen in table 6.7. From the ROC curve, seen in fig. 6.8 it seems there is a possibility of rejecting 43% of the background while keeping 90% of the signal.

Method	ROC Area	Significance
BDTG	0.737 (0.722)	0.653 (0.618)
BDT3	0.733 (0.715)	0.633 (0.587)
BDTD	0.729 (0.712)	0.625 (0.583)

Table 6.7: A comparison of the area under the ROC curves and the significance for the two sets of 5 variables. The one from the $N-1$ iterative procedure compared to, in parentheses, the one from the method specific variable ranking given by TMVA.

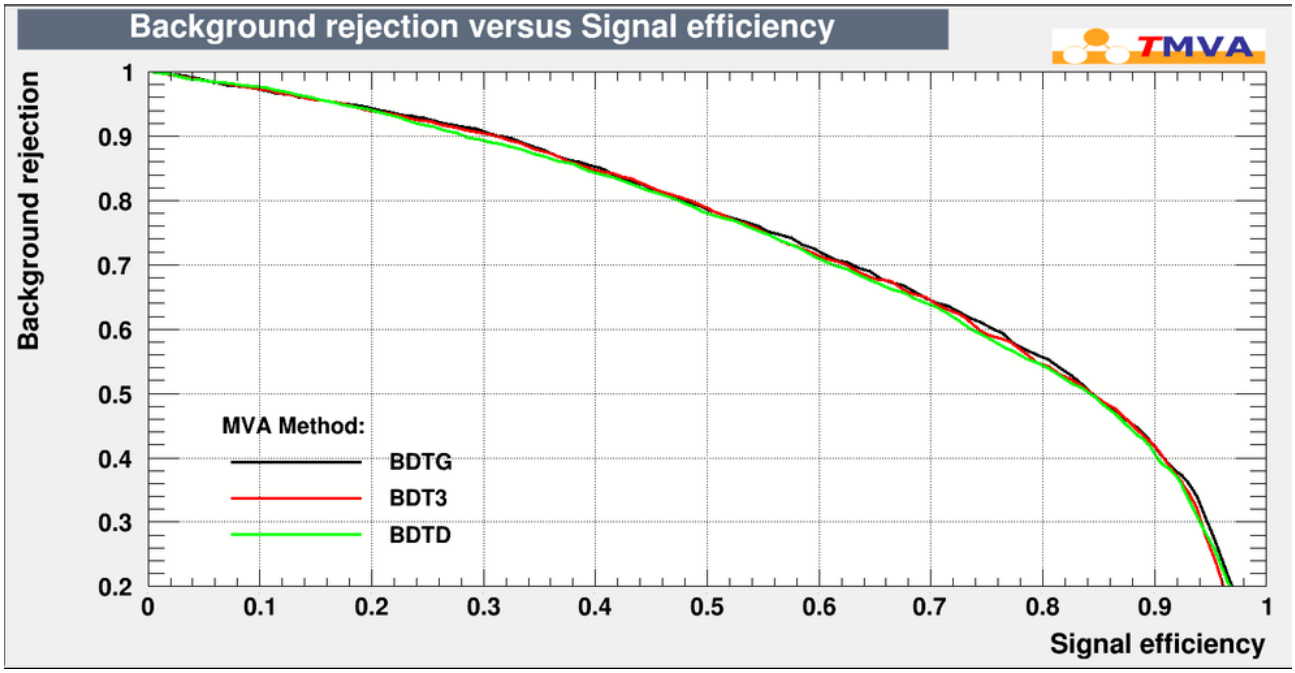

Figure 6.8: ROC curve showing the performance of the three different BDT methods listed on the plot, using the 5 top ranked variables from the $N-1$ iterative procedure.

A new training was made with the top four variables from the $N-1$ iterative procedure (leadJetPt, DPhiLeadLepMET, DPhiSubLeadLepMET, and nJets_Pt20). The result from the training showed no difference in background rejection seen in the ROC curve compared to the 5 variable case, meaning that it seems there is a possibility of rejecting 43% of the background while keeping 90% of the signal, see fig. 6.9. However, the set of 4 variables shows slightly better performance than 5 variable case for all methods when comparing significance, and slightly better performance for all methods except the BDTD method when comparing area under ROC curve, as can be seen in table 6.8.

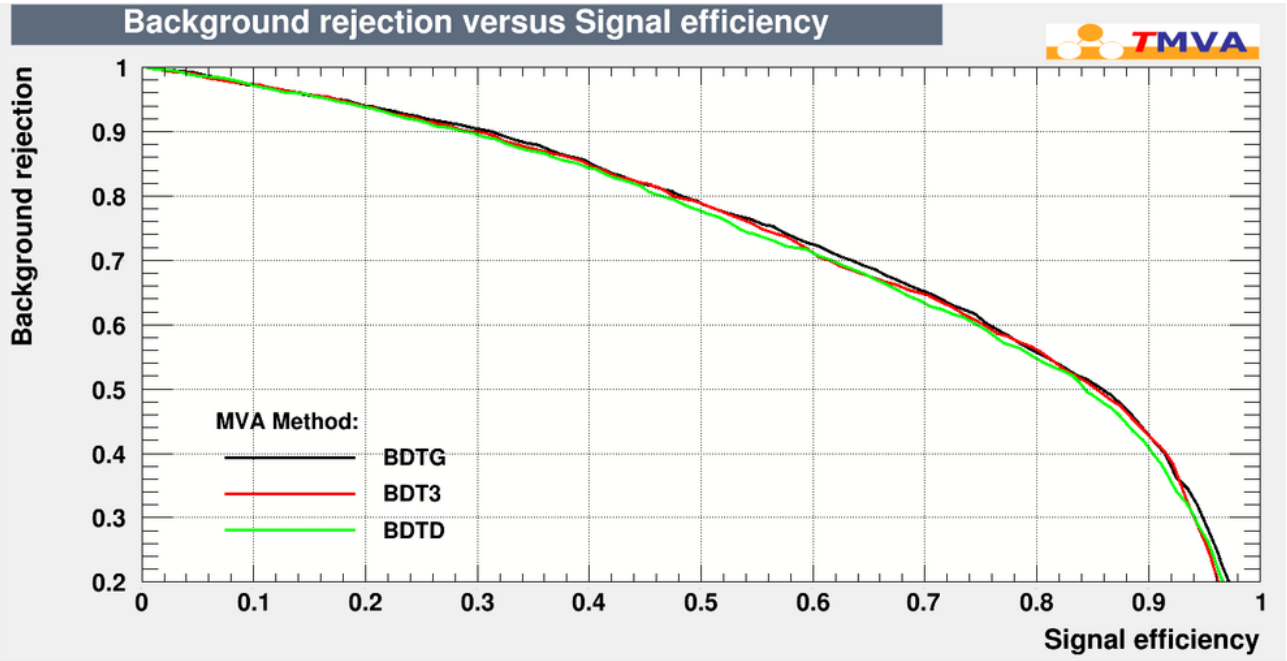

Figure 6.9: ROC curve showing the performance of the three different BDT methods listed on the plot using the 4 top ranked variables from the procedure.

Method	ROC Area	Significance
BDTG	0.737 (0.739)	0.653 (0.662)
BDT3	0.733 (0.734)	0.633 (0.640)
BDTD	0.729 (0.728)	0.625 (0.626)

Table 6.8: A comparison of the area under the ROC curves and the significance for the sets of the top 5 variables compared to, in parenthesis, the top 4 variables from the $N-1$ iterative procedure.

6.3.3 Discussion and Conclusions

The result from the study showed that, in this case, the non-method specific variable ranking worked well with only two variables changing rank from the set found in the $N-1$ iterative procedure that lead to the set of variables that gave the best result. It was shown that the set of variables could be optimized to four variables without a large loss of possible background rejection, from 44% with the whole set of 15 variables to 43% with the best set of 4 variables, as can be seen in fig. 6.9. In addition, with fewer variables the BDTs are less overtrained which may suggest that the result with fewer variables is more accurate. From this it can be concluded that the most promising set of of variables was leadJetPt, DPhiLeadLepMET, DPhiSubLeadLepMET, and nJets_Pt20. In this set, leadJetPt is the most important variable. During the variable ranking, in the runs where it was excluded, there was only a possibility of rejecting 24% of the background while keeping 90% of the signal. Two surprising results were found. One was that the ROC curves indicates that the background rejection increased from about 42% to about 43%

while keeping 90% of the signal when going from 6 variables to a subset of 5 variables. The other was that the area under the ROC curves and the significance increased slightly for some of the methods when going from 5 to 4 variables. Since BDTs are not sensitive to the inclusion of variables that are bad at separating signal from background, it should be the other way around, a reduction of variables should only be a loss of information. These improvements, when using fewer variables, are probably in the margin of error.

From fig. 6.9, it is clear that the method BDTD performs worse than the other two, which are difficult to distinguish in the 4 variable case. But since BDTG has slightly higher area under the ROC curve, together with the fact that it is the least overtrained and that TMVA ranks it as the best classifier of the three, it can be concluded that the BDTG method is to be preferred.

These results are promising, and it is possible that the inclusion of BDTs could lead to a more pure WW 1-jet control region, which in turn could lead to a better background estimation. However, further investigation of the optimal set of variables and an evaluation of the uncertainties are needed to determine if the inclusion of BDTs could contribute to the $H \to WW^*$ analysis.

Chapter 7

Conclusions

This thesis is limited to the work done by the ATLAS collaboration studying the decay channel $H \to WW \to l\nu l\nu$, where $l = e, \mu$. Along with an introduction to the Standard Model, the ATLAS detector, and Higgs phenomenology, the cut based analysis performed by this group has been described. In that analysis, cuts that have been optimized to separate signal from background are placed on one variable at a time. Two of the more difficult backgrounds to separate are from top quarks and W bosons and a possible way to improve this cut-based analysis is by using multivariate machine learning.

A study was conducted with the goal of finding possible discriminating variables between top and WW in the 1-jet control region using the multivariate machine learning method Boosted Decision Trees (BDT). The study was limited to the different flavour channel. Using Monte Carlo simulated events of top and WW one jet events, and 15 possible discriminating variables, 7 different BDTs were trained using TMVA. Starting from the result of that training, a variable optimization procedure was performed to get the number of variables down. From the result the most promising BDT methods were selected. Sets of variables not correlated with m_T were tested to find the best one. The result from the study showed that the most promising method was BDTG, and the best set of variables was a set of four variables, consisting of leadJetPt, DPhiLeadLepMET, DPhiSubLeadLepMET, and nJets_Pt20. When using this set it seems there is a possibility of rejecting 43% of the top background while keeping 90% of the WW signal in the WW 1-jet control region.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Jonas Strandberg for all the support during my work with this master thesis. From teaching me about research, to offering thoughtful suggestions and correcting my many spelling mistakes. I also owe much thanks to Jelena Jovićević and Edvin Sidebo for patiently helping me with the HWWAnalysisCode and TMVA, as well as answering my questions. I also want to thank Nabila Shaikh for all our fruitful discussions. Lastly I would like to thank my family and friends for their love and endless support.

List of Figures

List of Tables

Bibliography

[1] Brun, R. and Rademaker, F.: "ROOT — an Object Oriented Data Analysis Framework", *Nucl. Instr. Meth. Phys. Res.*, Vol. 389, April 1997

[2] Glashow, S.L.: "Partial-Symmetries of Weak Interactions", *Nucl. Phys.*, Vol. 22, February 1961

[3] Brout, R. and Englert, F.: "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.*, Vol. 13, August 1964

[4] Higgs, P.: "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.*, Vol. 13, October 1964

[5] Guralnik, G.S., Hagen, C.R., and Kibble, T.W.B.: "Global Conservation Laws and Massless Particles", *Phys. Rev. Lett.*, Vol. 13, November 1964

[6] Weinberg, S.: "A Model of Leptons", *Phys. Rev. Lett.*, Vol. 19, November 1967

[7] Salam, A. in: *"Elementary Particle Physics: Relativistic Groups and Analyticity"*, Eighth Nobel Symposium, Almqvist & Wiksell, 1968

[8] t' Hooft, G. and Veltman, M.: "Regularization and Renormalization of Gauge Fields", *Nucl. Phys. B*, Vol. 44, July 1972

[9] edited by Hoddeson, L. *et al.*: *"The Rise of the Standard Model"*, Cambridge University Presss, 1997

[10] The ATLAS collabortation: "Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC.", *Phys. Lett. B*, Vol. 716, September 2012

[11] The CMS collaboration: "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC", *Phys. Rev. Lett.*, Vol. 716, September 2012

[12] Masiero, A. and Pascoli, S.: "Neutrinos as Dark Matter Candidates", *Int. J. Mod. Phys.*, Vol. 17, May 2002

[13] Griffiths, D.: *"Introduction to Elementary Particles"*, Wiley-VCH, 2004

[14] The ALEPH Collaboration: "Determination of the Number of Light Neutrino Species", *Phys. Lett. B*, Vol. 231, November 1989

[15] The ALEPH Collaboration: "Precision Electroweak Measurements on the Z Resonance", *Phys. Rep.*, Vol. 427, May 2006

[16] MINOS Collaboration: "Measurement of the Neutrino Mass Splitting and Flavor Mixing by MINOS", *Phys. Rev. Lett.*, Vol. 106, May 2011

[17] Beringer, J. *et al.*: "Review of Particle Physics *", *Phys. Rev. D*, Vol. 86, July 2012

[18] Beringer, J. *et al.* (Particle Data Group): "Electroweak Model and Constraints on New Physics", *Phys. Rev. D.*, Vol. 86, 2012

[19] Peskin, M.E. and Shroeder, D.V.: *"An Introduction to Quantum Field Theory"*, Addison-Wesly, 1995

[20] Mandl, F. and Shaw, G: *"Quantum Field Theory"*, John Wiley and Sons, 1984

[21] Goldstone, J. *et al.*: "Broken Symmetries", *Phys. Rev.*, Vol. 127, August 1962

[22] The ATLAS Collaboration: "Evidence for the Spin-0 Nature of the Higgs Boson using ATLAS Data", *Phys. Lett. B*, Vol. 726, October 2013

[23] Evans, L. and Bryant, P.: "LHC Machine", *J. Instrum.*, Vol. 3, August 2008

[24] The ATLAS Collaboration: "The ATLAS Experiment at the CERN Large Hadron Collider", *J. Instrum.*, Vol. 3, August 2008

[25] Klingenberg, R. on behalf of the ATLAS Pixel Collaboration: "The ATLAS Pixel Detector", *Nucl. Instrum. Meth. A*, Vol. 579, September 2007

[26] Pater, J.R. on behalf of the ATLAS Collaboration: "The ATLAS Semiconductor Tracker Operation and Performance", *J. Instrum.*, Vol. 7, April 2012

[27] Bingul, A. on behalf of the ATLAS TRT Collaboration: "The ATLAS TRT and its Performance at LHC", *J. Phys.: Conf. Ser.*, Vol. 347, 2012

[28] Damazio, D.O. on behalf of the ATLAS Collaboration: "Performance of the ATLAS Calorimeter Trigger with 7 TeV Collision Data", *J. Phys.: Conf. Ser.*, Vol. 331, 2011

[29] Polini, A. on behalf of the ATLAS Muon Collaboration: "Design and Performance of the ATLAS Muon Detector Control System", *J. Phys.: Conf. Ser.*, Vol. 331, 2011

[30] Hauser, R. on behalf of the ATLAS Collaboration: "The ATLAS Trigger system", *Eur. Phys. J. C*, Vol. 34, 2004

[31] Padilla, C. on behalf of the ATLAS Collaboration: "The ATLAS Trigger System", *IEEE Trans. Nucl. Sci.*, Vol. 57, April 2010

[32] The ATLAS Collaboration: "Studies of the Performance of the ATLAS Detector Using Cosmic-Ray Muons", *Eur. Phys. J. C*, Vol. 71, March 2011

[33] Baglio, J. and Djouadi, A.: "Higgs Production at the LHC", *J. High Energ. Phys.*, Vol. 2011, 2011

[34] Beringer, J. *et al.* (Particle Data Group): "Status of Higgs Boson Physics", *Phys. Rev. D.*, Vol. 86, 2012

[35] ATLAS Collaboration: "Analysis of $H \to WW^* \to l\nu l\nu$ ggF and VBF Production Modes with 20 fb^{-1} and 4.7 fb^{-1} of Data Collected with the ATLAS Detector at $\sqrt{s} = 8$ and $\sqrt{s} = 7$ TeV", ATL-COM-PHYS-2014-466, 2014

[36] ATLAS Collaboration: "Measurements of the Properties of the Higgs-Like Boson in the $WW^* \to l\nu l\nu$ Decay Channel with the ATLAS Detector Using 25 fb^{-1} of Proton-Proton Collision Data", ATLAS-CONF-2013-030, 2013

[37] Beringer, J. *et al.* (Particle Data Group): "The Top Quark", *Phys. Rev. D.*, Vol. 86, 2013

[38] Yang, HJ. *et al.*: "Studies of Boosted Decision Trees for MiniBooNE Particle Identification", *arXiv:physics/0508045v1*, August 2005

[39] V.M. Abazov *et al.* (DØ Collaboration): "Evidence for Production of Single Top Quarks and First Direct Measurement of $|V_{tb}|$", *Phys. Rev. Lett.*, Vol. 98, May 2007

[40] Alpaydin, E.: *"Introduction to Machine Learning"*, The MIT Press, 2010

[41] Hoecker, A. *et al.*: "TMVA 4 - Toolkit for Multivariate Data Analysis with ROOT", *arXiv:physics/0703039v5*, July 2009