Search for a Neutral Higgs Boson decaying to Tau Leptons using CMS at the LHC

by

Joshua Swanson

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Physics)

at the

University of Wisconsin – Madison

2013

Defended on 30 Apr 2013

Dissertation approved by the following members of the Final Oral Committee:

Wesley Smith · Bjorn Wiik Professor of Physics

Sridhara Dasu · Professor of Physics

Matt Herndon · Professor of Physics

Lisa Everett · Associate Professor of Physics

Terry Millar · Professor of Mathematics

Abstract

This thesis presents a search for the standard model Higgs boson decaying to a $\tau\tau$ pair using CMS data recorded at the LHC with 7 TeV and 8 TeV center of mass energy during 2011 and 2012 respectively. The search is performed using two channels $\tau_\mu\tau_h$ and $\tau_e\tau_h$ where τ_e, τ_μ represent a τ decay to an electron or muon and τ_h represents a τ decay to hadron(s). The search is performed in the range 100 GeV $< M_H <$ 145 GeV. An excess of events is observed with a maximum local significance of 2.56 σ at $M_H = 125$ GeV. A best fit to the standard model Higgs boson yields signal strength of $\mu = 1.20^{+0.51}_{-0.48}$ at $M_H = 125$ GeV.

Acknowledgements

There are many people that I would like to thank, without them I wouldn't be where I am today. I would say that for many years I found education to be boring and a waste of my time. There was a time that I could not envisage going to college, there was even a chance that I wouldn't finish high school. From a D in 8th grade science to where I am today is quite a journey. First and foremost I would like to thank my mother Patricia Swanson for always believing in me in spite of the hell I put her through. Also I would like to thank my father Samuel Swanson for always supporting me and teaching me many useful skills that I use in my daily life.

Outside of my parents I would first like to thank the people who employed me in my years after high school that helped me to realize that there was more to life. These early jobs and how unhappy I was doing them made me realize that I wasn't challenging myself and led me to enroll in college. This first college experience, enrolled at Wisconsin Indianhead Technical College, allowed me to gain confidence in myself and lead me to believe that I could do whatever I wanted. This first college experience didn't contribute much towards my future in physics but gave me the desire to further my education. While taking courses at Metropolitan State University in Saint Paul, MN I took a class in General Chemistry. This class first peaked my interest in science and especially how particles could combine together to form elements.

While pursuing a degree in engineering at Saint Cloud State University I took my first physics class, taught by my eventual advisor Dr. John Harlander. John not only introduced me to the joy I found in solving physics problems but really helped me realize that I could do anything. I soon switched to a major in Physics. In addition I have to thank Dr. Kevin Haglin for many wonderful classes and really cementing my

interest in particle physics. He mentored me in one of my first research projects on pressure in a quark gluon plasma. Not to mention his door was always open to discuss anything. Other professors at SCSU deserve acknowledgment as well including Dr. Juan Cabenala for his support in another very early research project and Dr. Maria Womack for many very interesting Astronomy courses

Finally during my graduate career there is also many people to thank. I have to give thanks to my advisor Dr. Wesley Smith for his constant support and encouragement. His enthusiasm for particle physics and triggers has surely left a mark on me. Additionally though not my advisor Dr. Sridhara Dasu has given my constant support during my PhD career. He was always at meetings where I presented my research and always available to give me advice and direction. I also have to give a thanks to the computing support staff at Wisconsin, Dan Bradley, Ajit Mohapatra, Tapis Sarangi and Will Maier for keeping the computers running smoothly. Additionally I would like to give thanks to my fellow graduate students Ian Ross, Jeff Klukas, Lyndsey Gray, Marc Weinberg, Kira Grogg, Mike Anderson and Michalis Bachtis for answering my many questions early on and helping countless times along the way. Especially Michalis earns a special thanks, though we often argued, which encouraged me to be a better physicist. Also I would like to thank the Postdocs and scientists Evan Friis, Maria Cepeda, Pam Klabbers and Sascha Savin for their considerable support.

Lastly I would like to thank the members of the CERN basketball team for giving me an escape from physics without this I surely would have lost my sanity.

Contents

List of Figures

List of Tables

Chapter 1

The Standard Model of Particle Physics

The standard model of particle physics (SM) is a theory that explains the nature of all known particles, quarks and leptons, that make up the world we live in and their interactions. It is a remarkably successful theory that came from collaboration of the entire particle physics community and was formulated in its current form in 1974 [1] describing electromagnetic, weak interactions and strong interactions amongst quarks and leptons. The one undiscovered piece of the SM is the mechanism for consistently providing masses for the $W^{\pm}$ and Z^0 particles that moderate weak interactions while keeping the photon massless [2, 3, 4, 5, 6, 7], which in the SM comes from the Higgs mechanism. The Higgs mechanism predicts the presence of a neutral particle, the Higgs Boson.

The Higgs boson is believed to be responsible for the mass of the fundamental particles observed in nature. Particles such as electrons which by the SM exist as point particles obtain mass through interaction with the Higgs field. Discovery of the Higgs boson would validate a large part of our understanding of the world at the

smallest scales. It can also give direction to further theories attempting to understand the universe and its origin.

In the summer of 2012 the Compact Muon Solenoid (CMS) and ATLAS experiments at the Large Hadron Collider (LHC) independently confirmed the presence of a previously undiscovered neutral boson [8, 9]. The newly discovered boson was found decaying to ZZ and $\gamma\gamma$ final states and is consistent with the SM Higgs expectations in those channels. Confirming that this new boson is the SM Higgs boson means measuring the properties of this new particle. A major part of claiming this discovery would mean confirming the coupling of this new boson to fermions. Confirming that this new boson decays to τ leptons is an important piece of this puzzle and the topic of this thesis. Furthermore confirming that this newly discovered particle doesn't couple to fermions (τ leptons) would hint at new physics beyond the standard model.

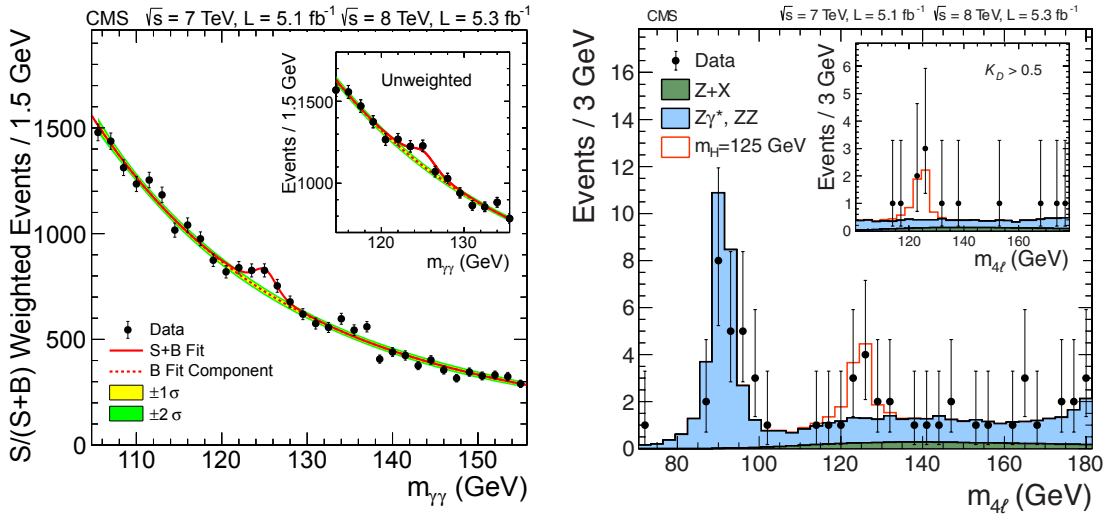

Figure 1.1: Observation of a new boson by the CMS experiment. (Left) Di-Photon invariant mass distribution the bump in the distribution at 125 GeV indicates the presence of a new particle, (Right) 4 lepton invariant mass from the ZZ search channel the peak of events at 125 GeV represents a new particle.

1.1 Particles and Interactions

The SM consists of a total of 61 distinct elementary particles/anti-particles. There are 12 leptons, 36 quarks, 12 force mediators, and the theorized Higgs boson. The leptons come in three generations and consist of the familiar electron, its heavier relatives, the muon and the tau along with corresponding neutrinos and all of their anti-particles. The quarks also come in three generations and a total of six varieties: up, down, charm, strange, top and bottom along with their antiparticles coming each in one of three *colors*. Exchange of force mediator particles is responsible for the forces between particles: Electromagnetic force with the photon, weak force with the $W^{\pm}$ and Z^0 bosons and the strong force with 8 gluons one with each color permutation. A summary of the discovered particles and their properties can be found in Figure 1.2.

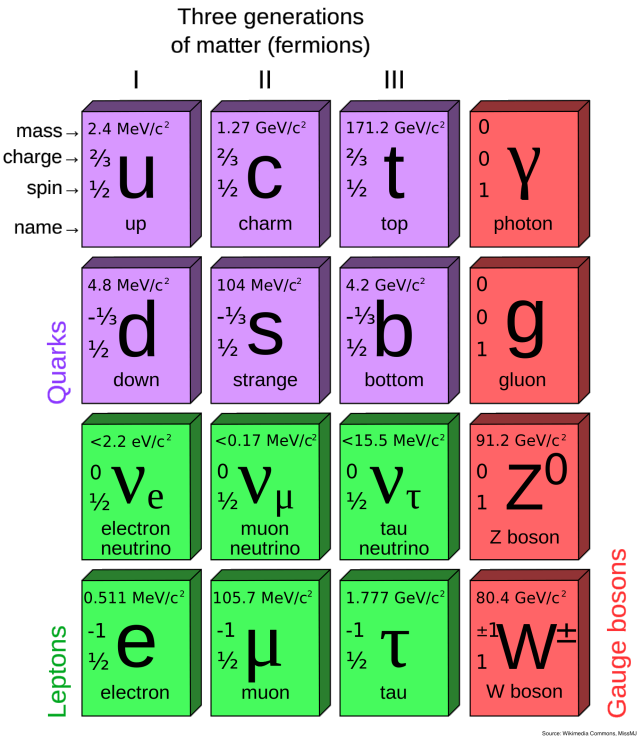

Figure 1.2: Standard model particles and their properties.

There is a total of 3 flavors of leptons: electrons, muons and taus which each have an associated neutrino. In addition to those six particles each has an anti-particle to make a total of 12 leptons. All leptons are fermions and as such have half integer spin. Electrons, muons, and taus all have charge of -1 in units of electron charge ($1.60217646 \times 10^{-19}$ coulombs) while the neutrinos are neutral particles. In addition to charge, all leptons carry a unit of lepton flavor quantum number: electron number, muon number or tau number. The anti-particles have opposite units of charge and lepton flavor number. The heaviest lepton is the tau lepton at 1.777 GeV which quickly decays to an electron, muon, or hadrons plus one or two neutrinos. The τ lepton will be discussed in more detail in Section 1.2. Muons have a mass of 105.7 MeV and decay to an electron and two neutrinos. Lastly is the electron at 0.511 MeV which is stable. Neutrinos negligible mass.

The quarks come in 3 generations each with an up type and down type quark. The up, charm and top quarks are said to be up type quarks and all have $\frac{2}{3}$ units of electron charge, while the down, strange and bottom quarks are down type quarks and carry $-\frac{1}{3}$ units of electron charge. Like the leptons the quarks are fermions and have half integer spin but unlike the leptons the quarks have units of color charge; red, green, or blue. Free quarks have not been observed, instead the theory of the strong interaction states that all particles in nature should be 'colorless'. Quarks then can combine in one of two ways to produce particles. The first is by combining 3 quarks together, one of each color charge, to produce a baryon such as a proton or a neutron. The other is in quark-antiquark pairs ($q\bar{q}$) combined in a color anti-color combination to create a meson such as a pion. In addition to color charge quarks, have flavor quantum numbers: isospin, charm, strangeness, topness, and bottomness.

The SM governs the interactions produced by 3 of the 4 forces of nature; electromagnetic, weak and strong. Gravity is not included in the SM since its influence

on the sub-atomic level is negligible. The electromagnetic force is mediated by the photon and is explained through the theory of Quantum Electrodynamics [11]. The weak force is mediated by the $W^{\pm}$ and Z^0 bosons. The electromagnetic and weak forces are unified in the standard model under the electroweak theory [12, 13, 14]. In the electroweak theory the unification requires an additional boson that is explained through the Higgs mechanism. The Higgs boson will be discussed in Chapter 2. Lastly the gluon mediates the strong force. The strong force is explained by the theory of Quantum Chromodynamics. Gluons are massless and electrically neutral. They carry one unit of color and anti-color charge each. Quarks and gluons are collectively known as partons. A summary of the force-carrying bosons can be found in Table 1.1.

All fermions are influenced by the weak force, where as only quarks carry color charge and interact through the strong force. Only particles that carry electrical charge can interact through the electromagnetic force. Lastly since neutrinos don't have electric or color charge they only interact through the weak force.

Boson	Force	Mass [GeV]	Charge	Force Strength	Force Range [m]
γ	Electromagnetic	0	0	10^{36}	∞
Z^0	Weak	91.2	0	10^{25}	10^{-18}
$W^{\pm}$	Weak	80.4	± 1	10^{25}	10^{-18}
gluon	Strong	0	0	10^{38}	$< 10^{-15}$

Table 1.1: Table summarizing the force carrying bosons. The strength of the force is relative to the force a gravity.

1.2 Tau Leptons

Tau leptons are the heaviest member of the lepton family. They decay via electroweak processes with a lifetime on average of 2.9 $\times 10^{-13}$ s [15]. The first evidence for the

tau was observed in the period between 1974 and 1977 [16]. Anomalous events were observed of the type $e^+ + e^- \rightarrow e^{\pm} + \mu^{\mp} +$ at least two undetected particles. This was explained by the presence of two tau leptons produced in an electroweak interaction and their subsequent decays to an electron, a muon and four undetected neutrinos.

Due to the tau lepton's short lifetime it is not directly detectable in current collider experiments, instead it is detected by reconstructing its decay products. Due to the mass of the τ it is heavier then some hadrons, this means that it can decay to a hadron(s) and a τ neutrino. The majority of tau decays are to hadrons with approximately 65% decaying to various combinations of hadrons plus a neutrino. In addition to decaying to hadrons a large fraction of taus decay to either an electron plus two neutrinos or a muon plus two neutrinos at about 17% each. About 90% of the tau decays involve either a lepton decay or the decay to one to three pions. The final 10% of decays are split between 25 more exotic decay modes including kaons or other hadrons. A Feynman diagram representing the decay of a τ lepton can be seen in Figure 1.3. A Feynman diagram is a graphical representation of a particle interaction. The diagrams have well defined rules governing which interactions are allowed and are used to simplify complex quantum field theory calculations. A summary of known tau decay modes and their branching ratios [15] can be found in Figure 1.4.

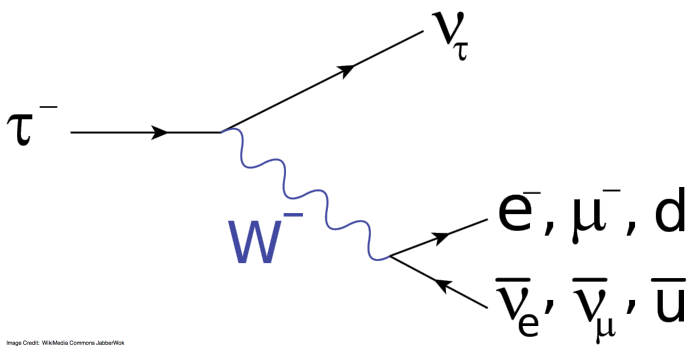

Figure 1.3: Feynman diagram showing the decay of the τ.

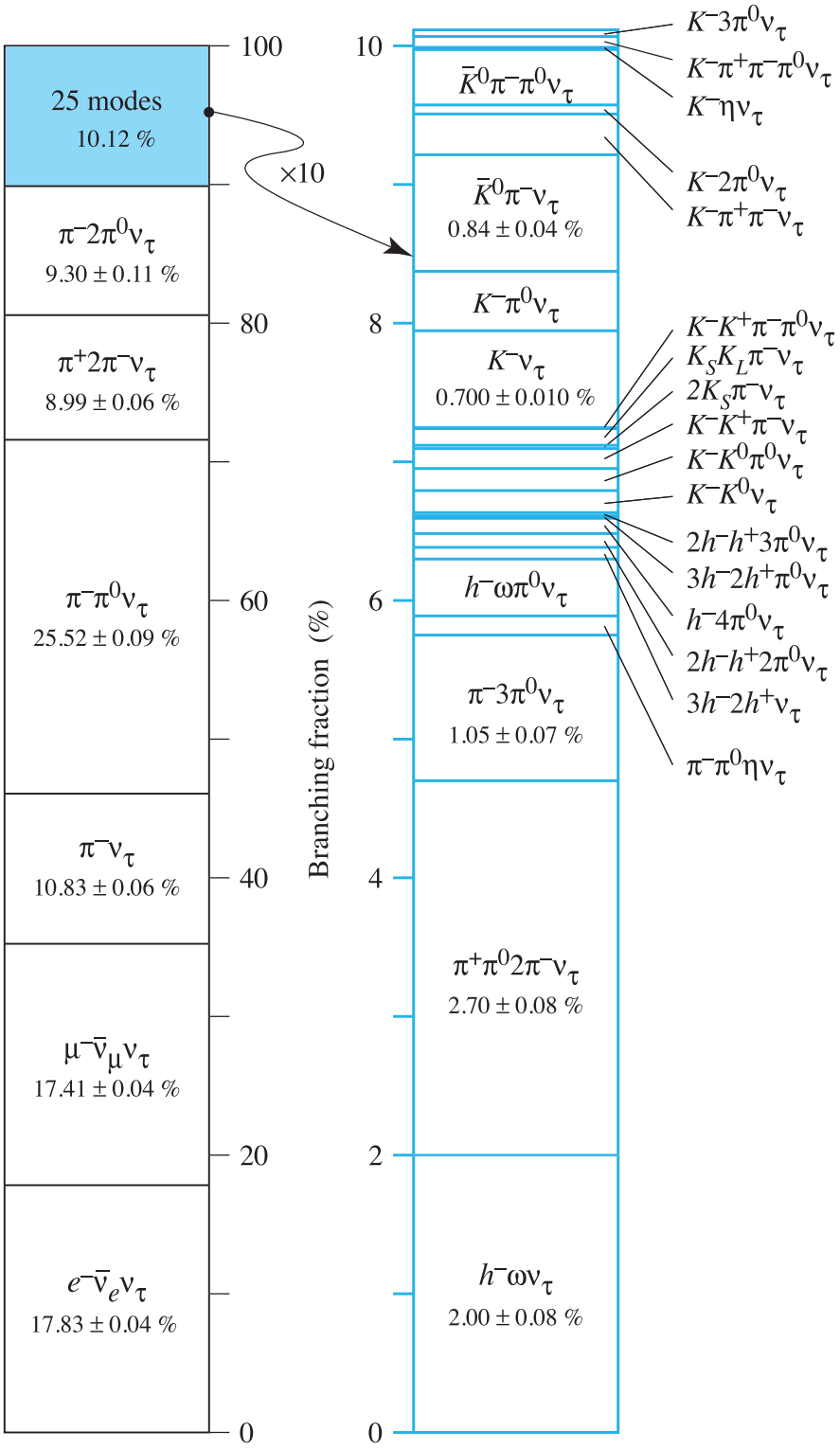

Figure 1.4: Branching fractions for the τ to the final state decay products.

1.3 Electroweak Theory

Electroweak (EWK) unifies the description of the weak force with that of the electromagnetic force. At energies in our everyday lives these are very different forces due to the relatively high mass of the $W^{\pm}$ and Z^0 bosons. This means the the weak force is short ranged but at energies on the order of 100 GeV, overcoming the mass barrier, the two forces will behave the same. The force carrying bosons in the EWK theory are the photon for electromagnetism and the $W^{\pm}$ and Z^0 bosons for the weak interaction. The short range nature of the weak force led to the prediction of the masses of the $W^{\pm}$ and Z^0 bosons. The $W^{\pm}$ and Z^0 bosons were later discovered at CERN by the UA1 and UA2 experiments. A broken EWK symmetry [2, 3, 4, 5, 6, 7] results in the $W^{\pm}$ and Z^0 bosons having considerable mass when compared to the massless photon. This is explained via the Higgs Mechanism.

The $W^{\pm}$ and Z^0 bosons are very massive particles when compared to the fermions, with the exception of the top quark. They are produced and decay into pairs of leptons or quarks. Interactions involving $W^{\pm}$ and Z^0 bosons can be seen in Figure 1.5. The interactions include the Z^0 boson coupling to all fermions, the $W^{\pm}$ boson coupling to fermions when charge is conserved, or the $W^{\pm}$ coupling to the other bosons including self coupling. Since Z^0 bosons can decay directly to a pair of tau leptons and have a cross section that is two orders of magnitude higher then the SM Higgs boson, they form the most significant irreducible background for the Higgs to tau pair search described in this thesis.

1.4 Quantum Chromodynamics

Quantum Chromodynamics (QCD) is the theory that governs the strong force. The strong force is responsible for binding the nucleus of the atom and explains the nature

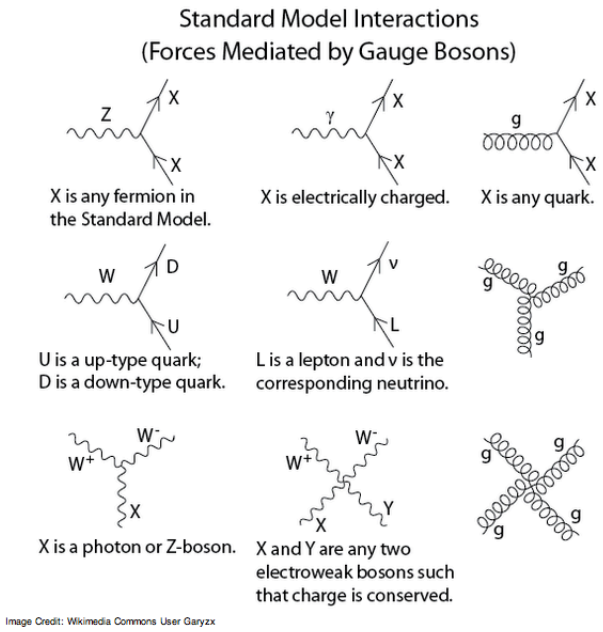

Figure 1.5: Feynman diagrams showing the allowed vertices in the Standard Model.

of interactions between quarks and gluons. The strong force is relatively large when compared to the other forces but acts only on a very short range, $\approx 10^{-15}$ m or the approximate size of atomic nuclei.

The force carrying particle in QCD is the gluon. Unlike the photon the gluon can also self interact because it carries color charge. The result of this is two of the three allowed QCD vertices, shown in Figure 1.5. The third allowed interaction is that of a quark with a gluon. In total these interactions produce two types of particles, baryons and mesons. Baryons are combinations of three quarks, one of each color. Baryons include for example the proton and neutron. Mesons are the combination of two quarks in a color anti-color pair. Common mesons include pions and kaons.

Asymptotic freedom is a property that allows QCD interactions to be weaker at higher energies or shorter distances. This is important for two reasons. It allows treatment of QCD interactions in high energy physics using perturbative methods to

calculate cross sections etc. It also provides an important physical feature preventing the unbinding of protons and neutrons that make up all the matter around us.

The standard model of particle physics is a theory that explains the interactions of all known fundamental particles. The confirmation of a standard model Higgs boson would complete the standard model. One important feature to be tested is the coupling of this new boson to fermions, and specifically its decay to τ leptons. This thesis will describe a search for a SM Higgs boson decaying to a pair of τ leptons using data from the CMS detector taken in 2011 and 2012 at 7 and 8 TeV center of mass energy respectively. This study can help to confirm this new boson is consistent with the SM Higgs boson or confirm the necessity for new physics beyond the standard model.

Chapter 2

The Higgs Boson

The Higgs boson is the last undiscovered particle in the SM. In the SM the $W^{\pm}$ and Z^0 bosons acquire mass via a broken symmetry [2, 3, 4, 5, 6, 7], called the Higgs Mechanism which as a consequence predicts an additional scalar boson. In addition to giving mass to the $W^{\pm}$ and Z^0 bosons the mass of the fermions can be explained by their coupling to the Higgs field. Higgs bosons in high energy collisions are produced in 3 basic modes. These modes include direct gluon-gluon fusion, in association with an additional boson, or through vector boson fusion. In the SM couplings to the Higgs field are proportional to the mass of the particle. This means that the decay of the Higgs boson is dependent on its mass. In extensions to the SM the Higgs boson theory can change considerably. These different theories have predictions, varying from the way they couple to fermions and even to multiple Higgs fields and associated Higgs bosons.

2.1 The Higgs Mechanism

The theory of the electroweak force unifies the electromagnetic and the weak forces. Explaining the unification of these forces requires an explanation of why the carrier of the electromagnetic force the photon is massless compared to the $W^{\pm}$ and Z^0 bosons which have considerable mass. This is accomplished through the Higgs field. The breaking of the symmetry of the Higgs field gives masses to the $W^{\pm}$ and Z^0 bosons while the electric charge part of the group remains unbroken and the photon remains massless. Another remarkable property of the Higgs field is that other particles can acquire their mass through coupling to the Higgs field. Particles that have a stronger coupling to the field in turn have a higher mass.

The electroweak sector of the SM is a gauge group of type $SU(2) \times U(1)$. A doublet of complex fields is added that we refer to as the Higgs field. The Higgs field acquires a non-zero vacuum expectation value that spontaneously breaks the symmetry of the group. An unbroken group remains that is consistent with electromagnetism and the massless photon. Three out of the four degrees of freedom introduced by the gauge group mix with the $W^{\pm}$ and Z^0 bosons to give them mass, the final degree of freedom becomes the Higgs boson. Finally the SM Higgs boson obtains a mass given by Equation 2.1 where λ is the Higgs self-coupling parameter and v is the vacuum expectation value and is a function of the fermi coupling (G_F). v is measured [19] to be about 246 GeV to ppm precision using muon decay measurements.

$$m_H = \sqrt{\frac{\lambda}{2}} v \qquad\qquad v = (\sqrt{2} G_F)^{-\frac{1}{2}} \tag{2.1}$$

Since λ is unknown the mass of the Higgs boson cannot be predicted by theory. Theoretical constraints can be made on the possible value, for instance if m_H is too high then the Higgs self coupling term will diverge below the Planck scale.

In addition to the explanation of the masses for the $W^{\pm}$ and Z^0 bosons the masses of all fermionic particles can be explained through direct coupling to the Higgs field. This is achieved using a Yukawa term in the SM Lagrangian, see Equation 2.2. A Yukawa interaction is an interaction between a scalar field (Φ) and a Dirac field (Ψ). The scalar field is the Higgs field. A Dirac field is a fermionic field which in quantum field theory (QFT) define the spin and momentum of fermions.

$$V = g\bar{\Psi}\Phi\Psi \tag{2.2}$$

Then we see, in Equation 2.3, that the mass of the fermions is proportional to the coupling constant (g) and the vacuum expectation value.

$$m_f = gv \tag{2.3}$$

Since the value of the coupling is unknown the masses of the fermions cannot be predicted by the theory. Alternatively since the mass of the fermions are known to a high precision the theory can be tested by measuring the ratio of the couplings to the different fermions that should be proportional to m_f.

2.2 Higgs Production Modes

The dominating Higgs production mode at high energy colliders is through gluon-gluon fusion, $gg \to H + X$. Following gluon-gluon fusion the other production modes include vector boson fusion, $qq \to qqH + X$, and associated production, $qq \to VH + X$ where V is a $W^{\pm}$ or Z^0 boson. Lastly with a greatly reduced production rate you have a Higgs produced in association with a top quark pair. Feynman diagrams showing the production modes can be seen in Figure 2.1.

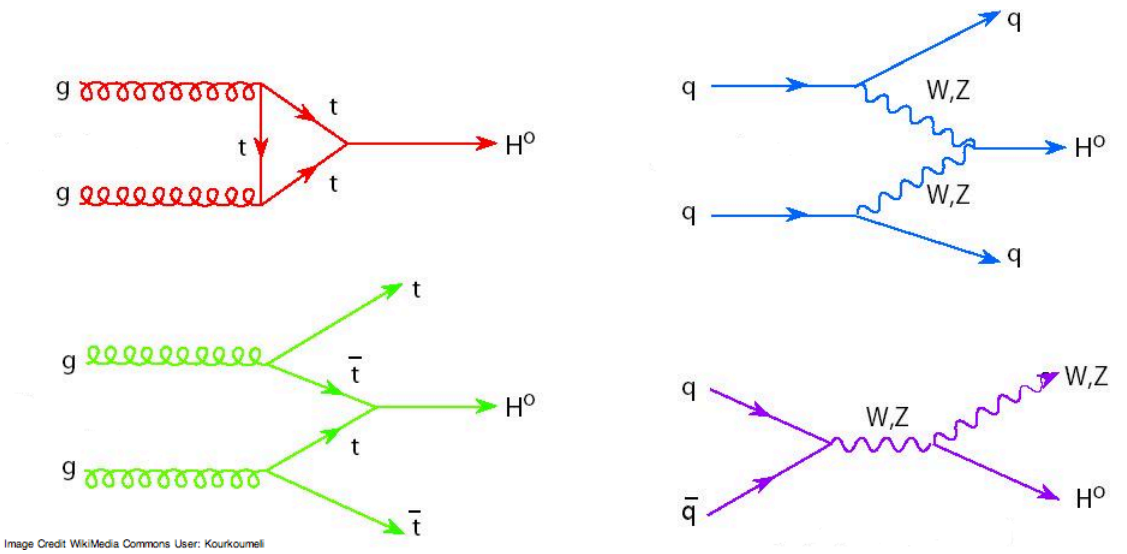

Figure 2.1: Feynman diagrams showing the Higgs production modes in high energy collisions: Gluon-Gluon Fusion (Top Left), Vector Boson Fusion (Top Right), $t\bar{t}$ Associated Production (Bottom Left), $W^{\pm}/Z^0$ Associated Production (Bottom Right)

Gluon-Gluon fusion (ggF) production is achieved through a top quark loop. It is the dominating Higgs production mode at the LHC. The cross sections are calculated to next-to-next-to-leading-logarithmic (NNLL) accuracy in QCD and next-to-leading-order (NLO) accuracy in EWK. The Higgs production cross sections as a function of mass at the LHC are shown in Figure 2.2. The predicted cross sections for a 125 GeV Higgs via ggF are about 20 pb with an uncertainty of about 15% at 8 TeV center-of-mass energy at the LHC.

Terms like NNLL, NLO, or NNLO refer to the accuracy of the calculation. Since the Feynman diagrams for any process have an infinite amount of diagrams with increasing number of loops to be calculated, processes in QFT are approximated to a given accuracy. For example an NLO calculation includes all diagrams up to one loop accuracy.

Vector boson fusion (VBF) production is the second most common production mode at the LHC. It is achieved through a pair of quarks radiating a $W^{\pm}$ or Z^0 boson

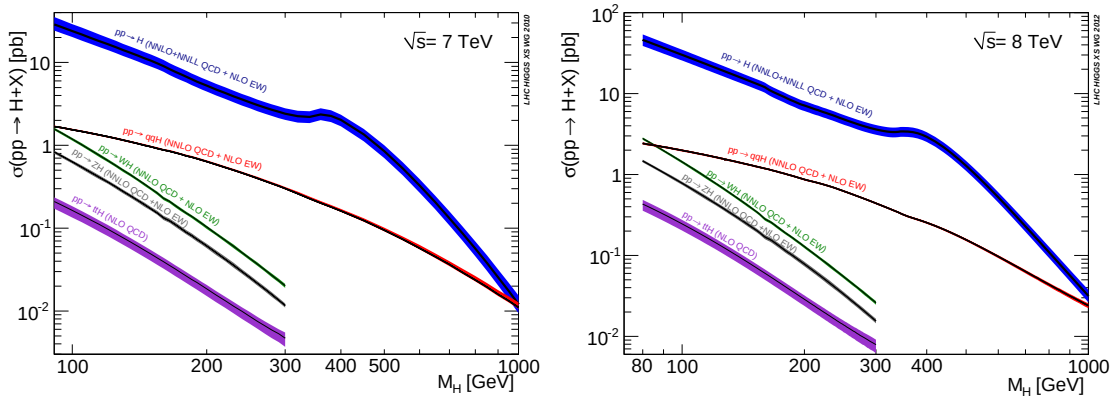

Figure 2.2: Standard Model Higgs production cross sections at the LHC for 7 TeV (right) and 8 TeV (left) center of mass energy.

which combine to produce a Higgs Boson. VBF events have a unique signature of two remaining hadronic jets in the event. These jets are generally in the forward direction with a large psuedorapidity separation. The cross sections are calculated to NNLO accuracy in QCD and NLO accuracy in EWK. Predicted cross sections for a 125 GeV Higgs via VBF are about 1.6 pb with an uncertainty of about 3% at 8 TeV center-of-mass energy at the LHC.

Last there is the associated production (VH) mode, it is the smallest production mode at the LHC. The Higgs is produced when it is radiated from a $W^{\pm}/Z^0$ boson or directly from a $t\bar{t}$ pair. VH events have a remaining $W^{\pm}/Z^0$ boson in the event while ttH have a remaining top quark pair. The WH/ZH cross sections are calculated to NNLO accuracy in QCD and NLO accuracy in EWK. While ttH cross sections are calculated to NLO in QCD. Predicted cross sections for a 125 GeV Higgs via WH(ZH) are about 0.7(0.4) pb with an uncertainty of about 4%(5%) at 8 TeV center-of-mass energy at the LHC. A summary of cross sections and uncertainties for some representative Higgs mass points can be found in Table 2.1.

$m_H[GeV]$	ggF		VBF		WH		ZH		ttH	
-	7 TeV	8 TeV	7 TeV	8 TeV	7 TeV	8 TeV	7 TeV	8 TeV	7 TeV	8 TeV
100	24.0	30.1	1.6	2.0	1.2	1.4	0.6	0.8	0.2	0.2
125	15.3	19.5	1.2	1.6	0.6	0.7	0.3	0.4	<0.1	0.1
150	10.6	13.7	1.0	1.3	0.3	0.4	0.2	0.2	<0.1	<0.1
200	5.4	7.1	0.7	0.9	0.1	0.1	<0.1	<0.1	<0.1	<0.1
400	2.0	2.9	0.2	0.3	-	-	-	-	-	-

Table 2.1: Higgs production cross sections (pb) at 7 TeV and 8 TeV at the LHC

2.3 Higgs Decay Modes

The decay of the Higgs boson depends strongly on its mass. In the low mass region below 135 GeV decays to fermions dominate, specifically to $b\bar{b}$ and $\tau\tau$. At higher mass, above 135 GeV, the Higgs decays to WW and ZZ dominate. The branching ratio for the Higgs boson as a function of its mass is displayed in Figure 2.3.

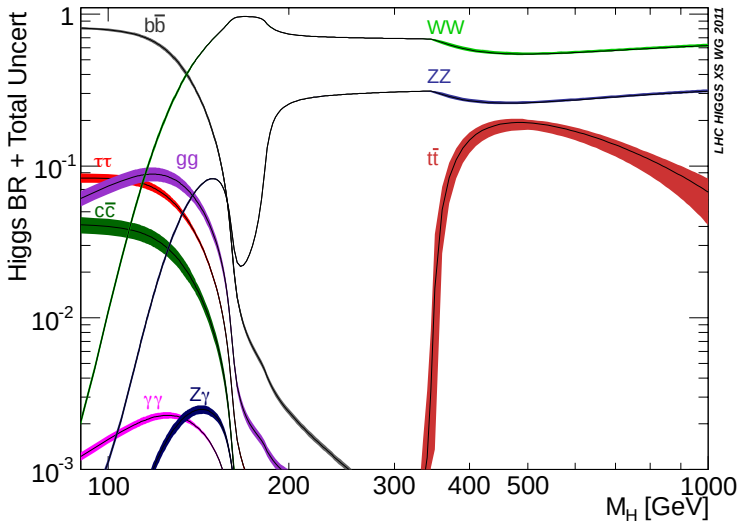

Figure 2.3: Branching ratio of the Higgs boson as a function of m_H with bands for the theoretical uncertainties.

In the low mass region the decays to leptons and quarks are involved in the majority of Higgs decays. Firstly $b\bar{b}$ comes with a branching ratio starting at 80%

at 90 GeV and then slowly going down as mass increases and WW and ZZ decays are increasing. After $b\bar{b}$ you have the $\tau\tau$ decays at about 1 order of magnitude lower. Finally in the low mass region you have $\gamma\gamma$ and Zγ decay modes that are introduced through a WW or $t\bar{t}$ loop. Searches at the LHC are particularly reliant on ZZ and $\gamma\gamma$ due to the good separation of the signal from the irreducible backgrounds in these channels. In these channels the full energy of the decay products is reconstructible resulting in a high resolution mass peak compared to backgrounds that are non-resonant. The $b\bar{b}$ and $\tau\tau$ modes are important for verifying coupling to leptons. They have much larger backgrounds but have a considerably higher cross section times branching ratio. For $\tau\tau$ this is particularly evident if you look at Figure 2.4 where you can see that it has the highest reconstructable cross section times branching ratio in the low mass region.

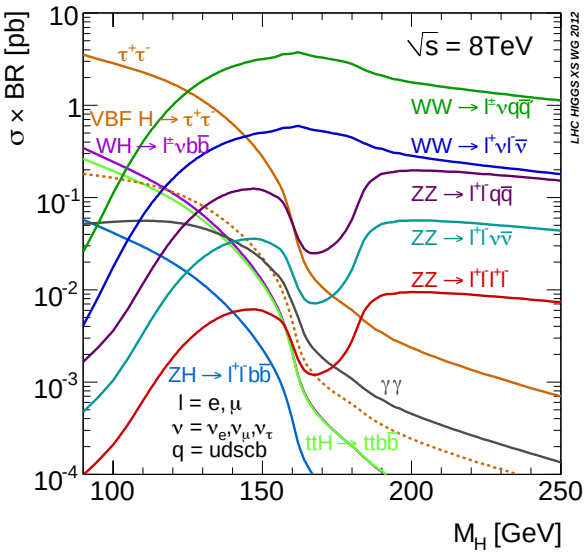

Figure 2.4: Cross Section × Branching ratio of the Higgs boson as a function of m_H for the Higgs boson below 200 GeV. Modes that are not traditionally accessible in collider experiments are not shown.

At high mass the landscape changes quite considerably. As m_H increases decays

to WW and ZZ start to dominate as they become kinematically allowed. At even higher masses decays to $t\bar{t}$ also become possible. A summary of the Higgs branching ratios for a few representative points in m_H can be found in Table 2.2

$m_H[GeV]$	$\tau\tau$	bb	$\gamma\gamma$	Zγ	WW	ZZ	$t\bar{t}$
100	0.08	0.79	0.002	0.00005	0.01	0.001	0.0
125	0.06	0.57	0.002	0.002	0.22	0.03	0.0
150	0.02	0.16	0.001	0.002	0.70	0.08	0.0
200	0.0003	0.002	0.00006	0.0002	0.74	0.26	0.0
400	0.00003	0.0002	0.000003	0.00002	0.58	0.27	0.15

Table 2.2: Higgs decay branching ratios for a few representative mass points.

In addition to the branching ratio the Higgs decay width varies considerably with m_H. For low mass the decay width is as low as 10 MeV. As mass increases it rises quickly, reaching 1 GeV at m_H=200 GeV and up to 100 GeV for m_H=500 GeV. The Higgs decay width as a function of mass is shown in Figure 2.5.

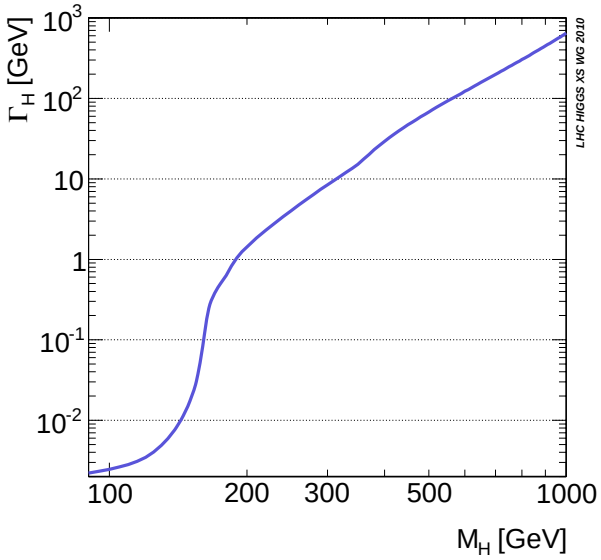

Figure 2.5: The decay width of the Higgs boson as a function of m_H.

2.4 Physics Beyond the Standard Model

The Standard Model of particle physics is a wonderful theory that not only explains the interactions of particles but also predicted several other particles and effects that were later confirmed. While this is a remarkable success it also has short comings which need to be covered by a more comprehensive theory. Problems with the standard model include the 'hierarchy problem' and the fact that it doesn't include the gravitational force.

Simply stated the 'hierarchy problem' asks why the Higgs mass should be so small when compared to the Planck scale. The Planck scale is the energy scale where quantum field theory breaks due to the non-renormalizability of gravity. It occurs at $\approx 10^{19} GeV$. The standard model predicts large radiative corrections to the Higgs mass that make it divergent without some large fine tuning. This problem can be solved by supersymmetry (SUSY). In SUSY all fermions and bosons have super-partners. The symmetry cannot be exact, or the SUSY particles would have been observed. Instead it is said that SUSY is a broken symmetry with particle masses on the TeV scale. Searches for SUSY at the LHC have been very active with current upper limits on squarks of 1.1 TeV and gluinos of up to 500 GeV [17].

As discussed previously, one of the marvelous achievements of the standard model was the unification of the weak force with the electromagnetic force. It is natural to want to extend this unification further to unify the strong force with the EWK force and ultimately gravity. The idea is that at high enough energy all forces are just different manifestations of one fundamental theory. This topic is referred to as Grand Unification Theories (GUTs), see for instance [18], or in combination with gravity, a Theory of Everything (TOE).

2.5 Previous Standard Model Higgs Boson Search Results

Higgs boson decaying to a pair of τ leptons has been searched for by a number of experiments. The first searches for Higgs to $\tau\tau$ came from the LEP electron-positron collider. These first searches excluded a large range of standard model Higgs masses. At the Tevatron searches for Higgs to $\tau\tau$ were performed by the CDF and D0 experiments. More recently Higgs to $\tau\tau$ searches are also being performed by the ATLAS and CMS experiments at the LHC.

At the LEP collider searches for the Higgs boson were done by the ALEPH, DELPHI, L3 and OPAL collaborations. The primary production mode came with a Higgs boson in association with a Z boson. The search channels included; $H \rightarrow b\bar{b} + Z \rightarrow q\bar{q}$, $H \rightarrow \tau\tau + Z \rightarrow q\bar{q}$, $H \rightarrow b\bar{b} + Z \rightarrow \tau\tau$, $H \rightarrow b\bar{b} + Z \rightarrow \nu\bar{\nu}$ and $H \rightarrow b\bar{b} + Z \rightarrow ee, \mu\mu$. A combination was done from all experiments resulting in a Standard Model Higgs excluded [20] for masses less then 114.4 GeV with 95% confidence.

At the Tevatron the CDF and D0 experiments performed a wide range of Higgs boson searches including $\tau\tau$. The dominant production modes at the Tevatron are the ggF process and the $W^{\pm}/Z^0$ associated production. The VBF production has a much smaller cross section but is exploited in some search channels. For a Higgs mass below 135 GeV the dominant search channel is Higgs decaying to $b\bar{b}$. Above 135 GeV the $W^{\pm}W^{\mp}$ is dominant. The combination search from the two Tevatron experiments excludes [21] a Higgs boson in two mass ranges $100 < M_H < 103$ GeV and $147 < M_H < 180$ GeV at 95% confidence. They also observe a mild excess with a global significance of 2.9 standard deviations with a local significance of 3.0 σ at 120 GeV.

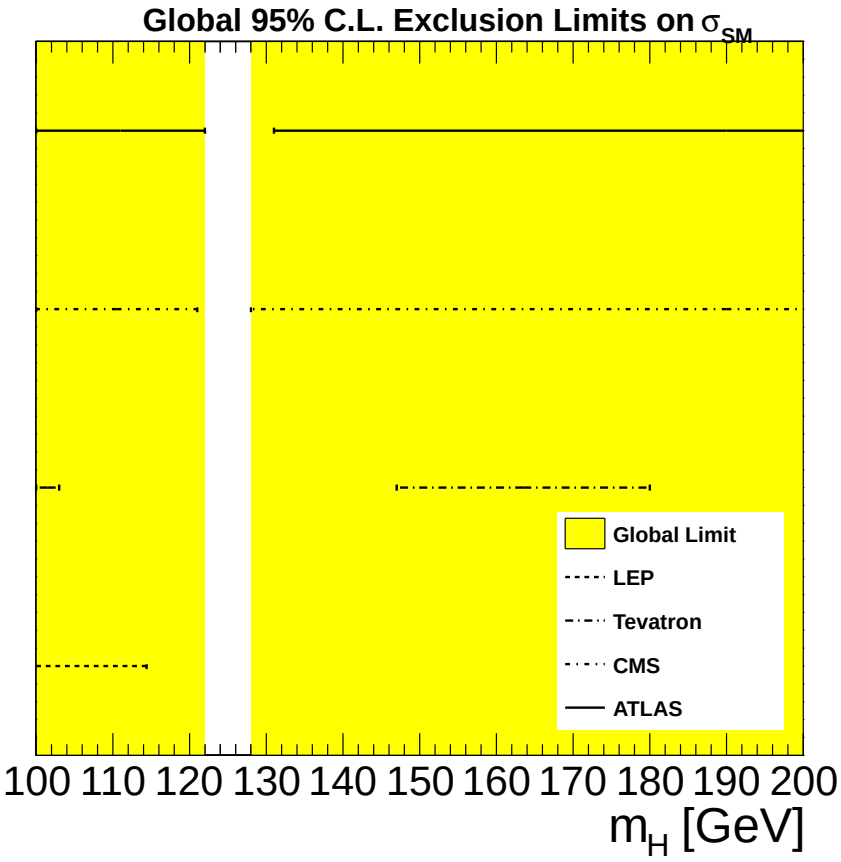

Figure 2.6: 95% C.L. Exclusions on the standard model Higgs cross section in m_H. Exclusion limits are shown separately for LEP [20], Tevatron [21], ATLAS [8] and CMS[9]. The allowed range is 122 GeV to 128 GeV. The ATLAS and CMS limits extend to higher values of m_H. The LEP limit extends to lower values of m_H

The $\tau\tau$ search channel at the Tevatron is done in association with 1 or more jets to enhance the VBF and associated Higgs production modes. The CDF search [22] corresponds to 8.3 fb^{-1} of CDF Run II data. CDF has an expected upper limit exclusion of 12.6 times the standard model Higgs cross section at $M_H = 115$ GeV. The D0 experiment also performed a search [23] with a luminosity up to 7.3 fb^{-1} which is sensitive at 21.8 times the standard model cross section for $M_H = 115$ GeV.

The ATLAS experiment at the LHC has a number of Higgs boson search channels. The dominant search channels are the Z^0Z^0 and the $\gamma\gamma$ modes. The latest search has

a total integrated luminosity of $4.8 fb^{-1}$ at $\sqrt{s} = 7$ TeV and $20.7 fb^{-1}$ at $\sqrt{s} = 8$ TeV. The results from $\gamma\gamma$ [24] and Z^0Z^0 [25] both show an excess greater than 5 σ for a new boson with mass [26] of 125.5 GeV.

The latest $\tau\tau$ results [27] from ATLAS have an integrated luminosity of $4.6 fb^{-1}$ at $\sqrt{s} = 7$ TeV and $13.0 fb^{-1}$ at $\sqrt{s} = 8$ TeV. The search is done in 6 decay channels; $\tau_h\tau_h$, $\tau_\mu\tau_h$, $\tau_e\tau_h$, $\tau_e\tau_\mu$, $\tau_\mu\tau_\mu$ and $\tau_e\tau_e$. It is done separating events in jet bins to accentuate Higgs production modes. They observe expected limits on the order of 1 times the standard model cross section with an observation that fluctuates upwards by slightly more then 1 σ.

CMS also has a wide range of Higgs boson search channels with the dominant channels having Higgs decays into Z^0Z^0 and $\gamma\gamma$. The latest results from CMS [28, 29] have a total integrated luminosity of $5.1 fb^{-1}$ at $\sqrt{s} = 7$ TeV and $19.6 fb^{-1}$ at $\sqrt{s} = 8$ TeV. The results are consistent with a standard model Higgs boson with a mass of 125.8 ± 0.5 (stat) $\pm$ 0.2 (syst).

Chapter 3

Experimental Setup

The Large Hadron Collider (LHC) was built to study particle interactions at a new energy frontier. To do this a proton-proton collider was chosen with design energy of 14 TeV and luminosity of $10^{34} cm^2 s^{-1}$. The ring was equipped with four experimental interaction points. The Compact Muon Solenoid (CMS), the ATLAS experiment, LHCB and ALICE. CMS and ATLAS are all purpose detectors designed for full luminosity. LHCB was built primarily for B-Physics aiming for a reduced peak luminosity. Lastly the LHC is also designed for operation with ion beams, ALICE was designed primarily for this LHC operating mode. ATLAS, LHCB and ALICE will not be discussed further in this thesis, see [32] for more information. The CMS experiment [33] was designed with a 3.8 Tesla magnetic field. Inside the large magnetic field the detector elements include an all silicon pixel and strip tracking system, a lead-tungstate scintillating-crystal electromagnetic calorimeter and a brass-scintillator sampling hadron calorimeter. In addition there is a iron yoke with a return field that contains a muon detection system.

3.1 The Large Hadron Collider

The LHC is a proton-proton collider built at the European Organization for Nuclear Research (CERN) near Geneva Switzerland. The LHC is built in an underground tunnel with a circumference of 27 km. The tunnel was previously used for the electron-positron collider (LEP) [34]. The LHC is designed to achieve center of mass energies of $\sqrt{s} = 14$ TeV and luminosities of $10^{34} cm^2 s$. The maximum center of mass energy was constrained by the available magnet technology and the circumference of the existing tunnel. In the early running period of the LHC from 2010-2012 the energy and luminosity were reduced. In 2010 and 2011 the energy consisted of 3.5 TeV per beam and luminosities up to 3.5 $\times 10^{33} cm^{-2} s^{-1}$. In the 2012 run the center of mass energy was increased to $\sqrt{s} = 8$ TeV, 4 TeV per beam with luminosities of up to 7 $\times 10^{33} cm^{-2} s^{-1}$.

Luminosity in high energy particle physics is a measure of the number of particles per unit area per unit time available for collisions. It is directly a function of the number of protons in each bunch (N_b), the number of bunches per beam (n_b), the frequency of the revolutions (f), and the effective area of the beams (A_{eff}). It is calculated approximately using Equation 3.1.

$$L = \frac{N_b^2 n_b f}{A_{eff}} \tag{3.1}$$

The LHC protons are obtained by stripping the electrons off from Hydrogen atoms. The protons are then accelerated in several steps before reaching the main LHC ring. They start in a linear particle accelerator called Linac2, the Linac2 uses radio frequency (RF) cavities. The charged protons leave the Linac2 accelerator with 50 MeV of energy and are injected into the Proton Synchrotron Booster (PSB). The PSB also uses RF cavities and ramps the protons up to an energy of 1.4 GeV and also

start to squeeze the beams to a smaller size. The protons are then sent to the Proton Synchrotron (PS) to be accelerated further to 24 GeV. The bunches are shortened using an 80 MHz RF system. This is done so that they can fit into the 200 MHz brackets of the Super Proton Synchrotron. The SPS is the final stage before the protons are injected into the main ring. Here the protons are accelerated to 450 GeV. At design specifications there would be a total of 2,808 bunches per beam each with about 10^{11} protons. The LHC is designed with an RF frequency of 40 MHz so the the bunches will be spaced at 25 ns apart. During the first LHC run period operation was done with 50 ns bunch spacing so that the maximum allowed number of bunches per beam was reduced by a factor of 2. A diagram of the LHC accelerator complex can be found in Figure 3.1.

Once the beams are in the main LHC ring they are accelerated to the final target energy. While being accelerated the beams travel in special superconducting dipole magnets. This is necessary since both beams have the same charge and need to travel in opposite directions. The magnets needed to fit in the existing 3.7 m tunnel width as such they have a compact twin-bore design. They include two sets of coils and two beam channels in the same structure and cooling system. They are mechanically and magnetically coupled. A total of 1232 dipole magnets are situated around the ring. They produce a magnetic field of 8.3 T. The magnets are made out of NbTi and cooled to a temperature of 1.9K using liquid helium. In addition to the dipole magnets there are approximately 400 superconducting quadruple magnets situated around the ring to focus the beam.

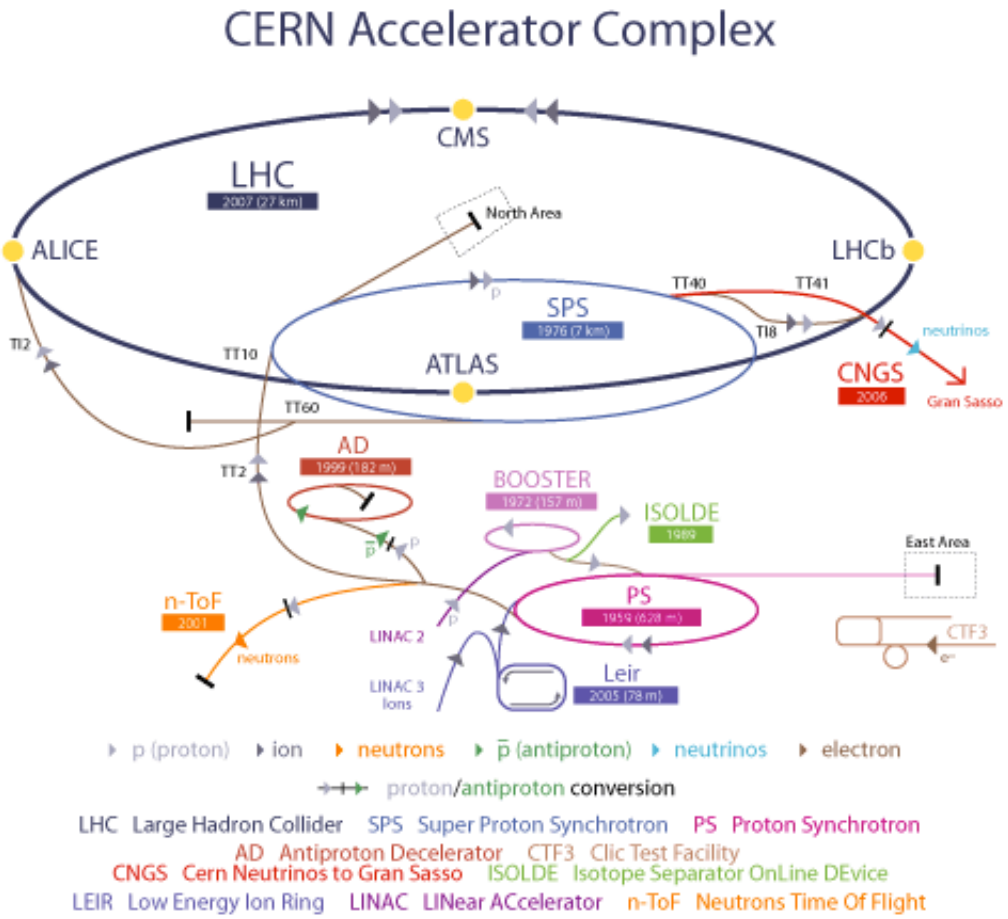

Figure 3.1: A diagram showing the complete LHC accelerator complex from the Linac2 all the way to the main beam. Also displayed are the locations of the 4 major experimental interaction points.

3.2 The Compact Muon Solenoid Experiment

The CMS detector [33] is a 12,500 ton multi-purpose detector built on the French side of the LHC ring. The primary feature of the CMS detector is a superconducting solenoid. The solenoid has a 6 m internal diameter and produces a magnetic field of 3.8 T. Inside the magnetic field there is a silicon tracking system, lead tungstate crystal based electromagnetic calorimeter and a brass and scintillator based hadron calorimeter. Outside of the solenoid there is a steel-flux return yoke which houses

gas-ionization muon detectors. Lastly there is a set of forward calorimeters to extend the rapidity coverage offered by the barrel/endcap colorimeters. The building of all calorimeters inside the solenoid was a novel approach to a high energy physics experiment. This allows for compact calorimetry using $PbWO_4$. Additionally particles don't have to traverse the materiel of the magnet before reaching the calorimeters. A diagram of the complete CMS detector can be seen in Figure 3.2.

Geometry

The CMS detector has a cylindrical shape. It has a diameter of 14.6 m and a total length of 21.6 m. The detector is divided into five regions: the barrel region, two end cap regions on each end of the barrel and two forward regions outside each end cap region. A right handed Cartesian coordinate system is used by the CMS detector. The origin is at the interaction point with the x direction towards the center of the LHC ring, the y direction pointing towards the surface and the z direction along the beam line. It can also be described using a polar coordinate system with ϕ in the transverse direction and the angle in the direction of the beam path defined by θ. In general instead of θ the variable pseudorapidity η is used, defined in Equation 3.2. The advantage of η is that the distribution of particles produced in collisions is roughly uniform from 0, in the vertical direction, to infinity, along the beam path.

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

An image showing a cross section of the CMS detector can be found in Figure 3.3. The barrel region for the calorimeters is defined as approximately $|\eta| < 1.5$ and the end cap regions run from $1.5 < |\eta| < 3.0$. The forward hadron calorimeters extend the coverage in η out to 5.0. The CMS tracking system has coverage to $|\eta|$ of 2.5.

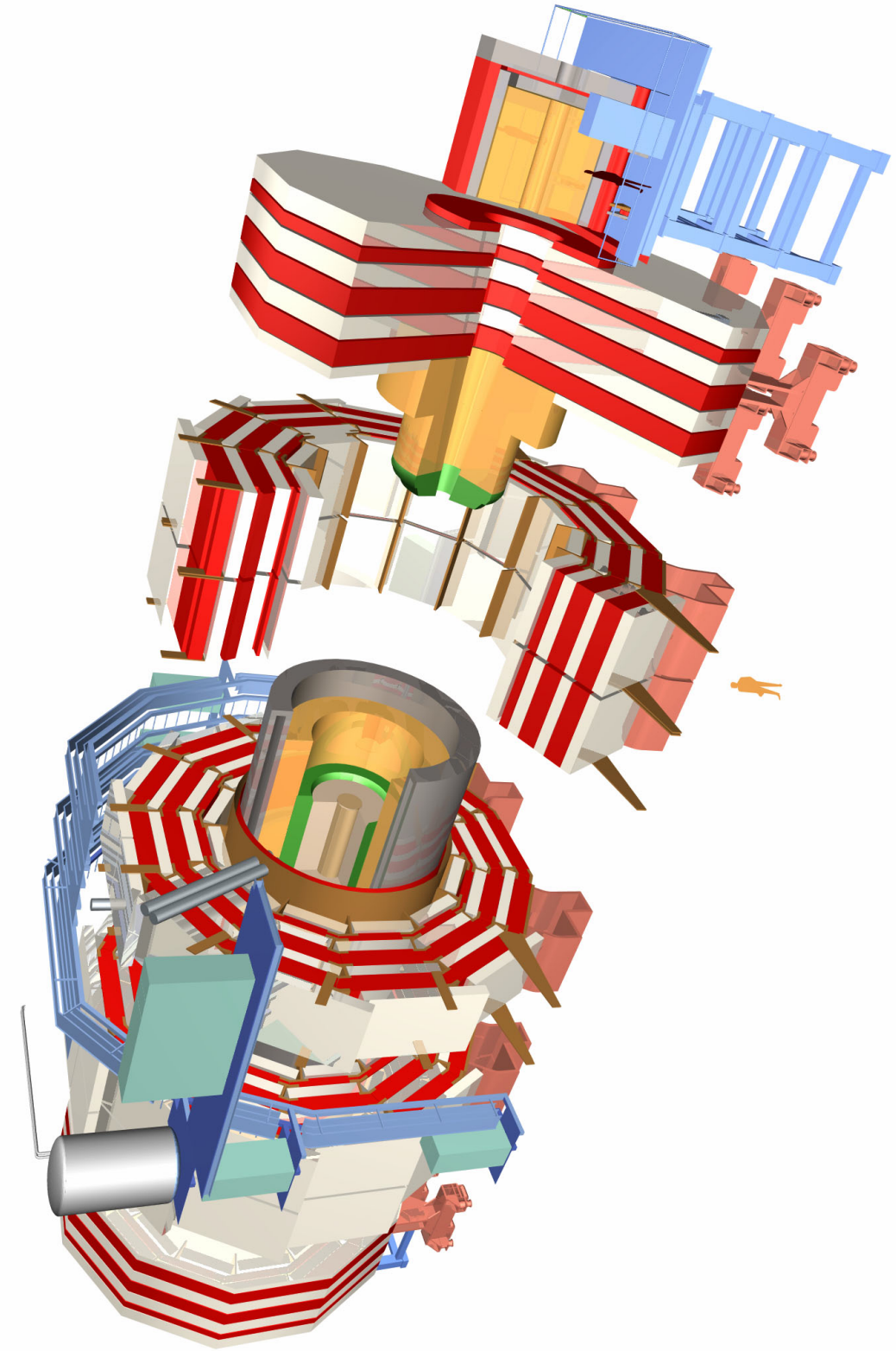

Figure 3.2: A diagram of the CMS detector starting from the inside with the Tracker (tan), Electromagnetic Calorimeter (green), Hadron Calorimeter (Gold), Solenoid (grey) and the Muon system (red). People placed in the picture give a reference for scale.

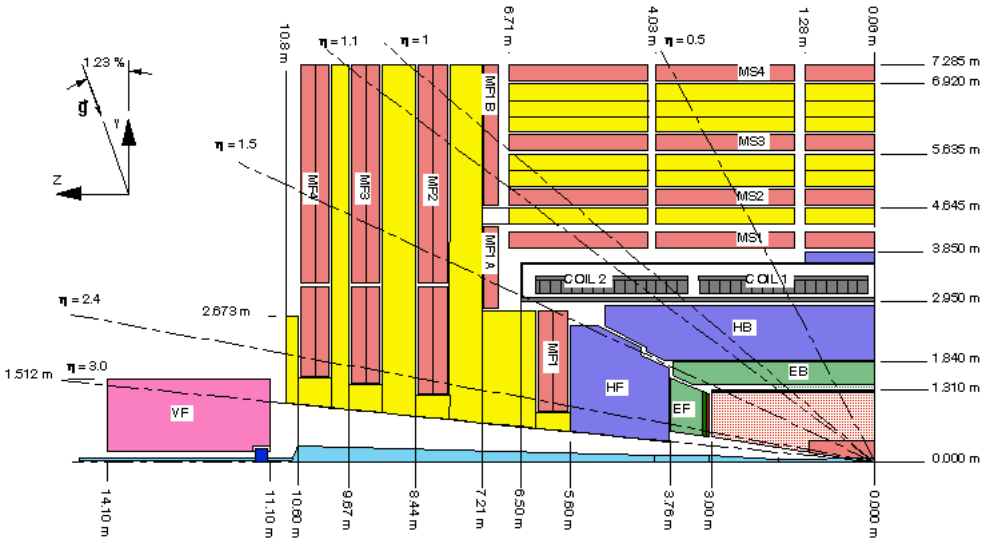

Figure 3.3: A side view diagram of the CMS detector. The image shows the η coverage of the various CMS detector components and regions.

Tracking System

The CMS tracking system consists of two distinct systems built inside the 3.8T B-field produced by the solenoid. The inner part is a silicon pixel detector covering radii between 4.4 cm to 10.2 cm from the interaction point and the outer part a silicon strip tracking system extending the radius of the tracking system out to 1.1 m. The decision to use an all silicon tracking system was a compromise between a material that could handle the intense particle flux expected at the LHC and keeping minimum amount of material, to limit multiple scattering, bremsstrahlung, photon conversions and nuclear interactions. The CMS tracker has coverage in $|\eta|$ to 2.5 with reliable tracking to $|\eta| < 2.4$. The total tracking system has about 200 m^2 of active silicon area making it the largest silicon tracker ever built. A schematic of the CMS tracker geometry is displayed in Figure 3.4.

The pixel tracking system consists of 3 cylindrical layers (BPix) of hybrid pixel detector modules at radii of 4.4, 7.3 and 10.2 cm and 2 disks (FPix) extending from

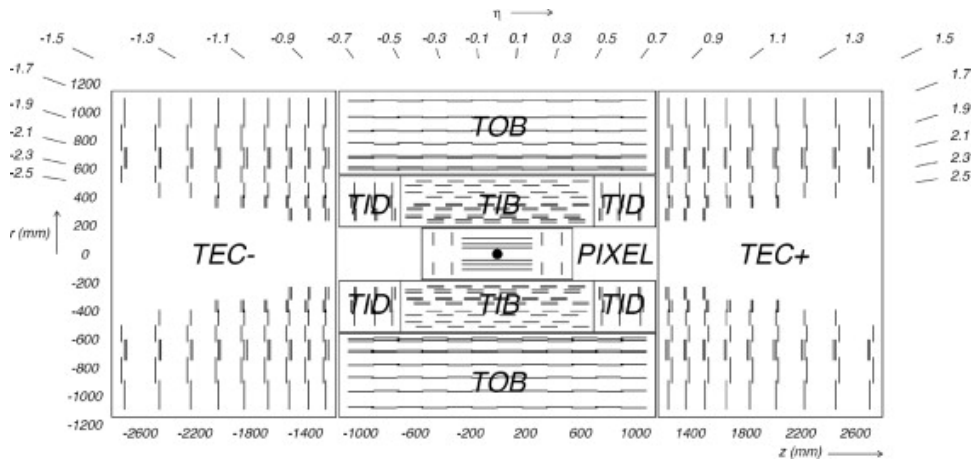

Figure 3.4: A side view diagram of the CMS tracking detector.

about 6 to 15 cm placed at z = ±34.5 and ±46.5 cm. The combination of the BPix and FPix gives 3 tracking points over almost the whole η coverage of the system. The system contributes to precise tracking points in r-ϕ and z and is mostly responsible for small impact parameter resolution [33]. The pixels consist of n-implants on high resistance n-substrate with a pn-junction placed on the back side [33]. The complete system consists of 65.9 million pixels covering an area of 1 m^2.

Outside the pixel detector in the radial region from 20 cm to 116 cm lies the silicon strip tracking system. The system is composed of three separate subsystems: the tracker inner barrel and disks (TIB/TID), the tracker outer barrel (TOB) and the tracker endcaps (TEC). The TIB/TID systems extend the tracker radius out to 55 cm. They are composed of 4 barrel layers and 3 disks at each end respectively. Outside the TIB/TID system is the TOB with a total of 6 layers and an outer radius of 116 cm and z=±118cm. Lastly the TEC system has 9 layers going from 124 cm $< |z| <$ 282 cm and a radius between 22.5 cm and 113.5 cm. The layout of the tracking system ensures any charged particle will encounter approximately 9 layers in the region $|\eta| < 2.4$. The sensor elements are made of p-on-n type silicon micro-strip

sensors. The total silicon strip system has a total of 9.3 million strips with 198 m^2 of active silicon area.

Tracking systems work by measuring the curvature of charged particle tracks in the magnetic field. The transverse momentum (p_T) of a charged particle is given by Equation 3.3 where q is the charge, r is the track radius and B is the magnetic field strength.

$$p_T = qrB \tag{3.3}$$

The large radius, magnetic field and fine granularity of the detector elements give the tracking system a high resolution. The approximate resolutions are given by Equation 3.4 and 3.5 for particles with $|\eta| < 1.6$ and $|\eta| > 1.6$ respectively.

$$\frac{\sigma_{p_T}}{p_T} = (15 p_T \oplus 0.5)\%(TeV) \qquad |\eta| < 1.6 \tag{3.4}$$

$$\frac{\sigma_{p_T}}{p_T} = (60 p_T \oplus 0.5)\%(TeV) \qquad 1.6 < |\eta| < 2.4 \tag{3.5}$$

The resolution of measuring higher p_T tracks is reduced due to reduced track curvature. Tracks with $p_T = 40$ GeV have resolution on the order of 0.6% compared to 100 GeV tracks with resolution near 1.5%.

Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) is responsible for measuring the energy of electromagnetic particles. The system is built out of lead tungstate crystals ($PbWO_4$) and is placed inside the magnetic field produced by the CMS solenoid. In total there are 61,200 crystals in the barrel region and 7,324 crystals in each end cap region. The barrel (EB for ECAL Barrel) is defined as the region of the ECAL with $|\eta| <$

1.479 and the end cap (EE for ECAL end cap) covering the region from $1.479 < |\eta| < 2.5$. Specific care has to be made for electromagnetic particles in the region between $1.4442 < |\eta| < 1.56$ due to the crack between the EB and EE systems. In addition to the EE there is also a preshower detector that was placed in front of the end cap crystals primarily for the purpose of detecting neutral pions. A diagram of the ECAL system can be seen in Figure 3.5.

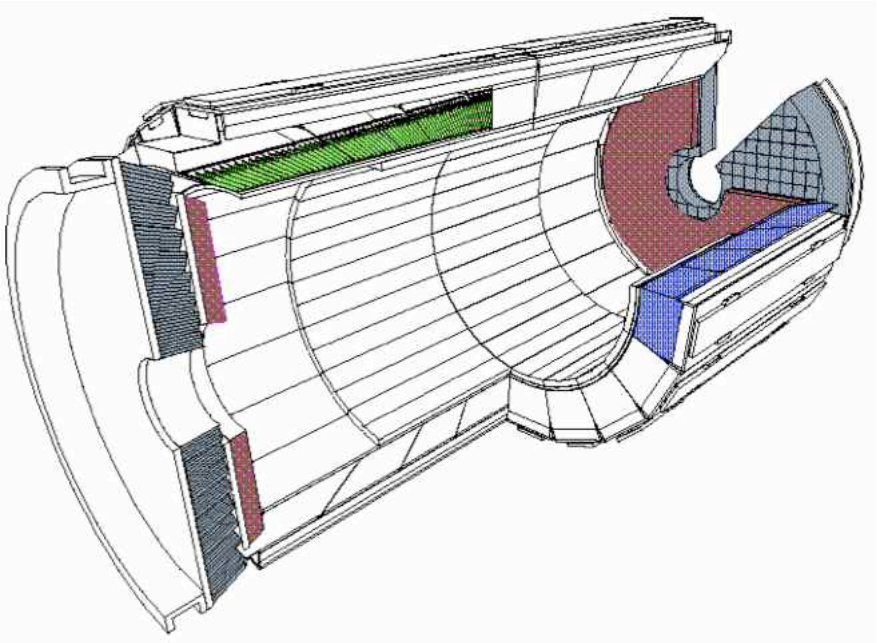

Figure 3.5: A diagram of the CMS ECAL system. The green elements represent the EB crystals, the red the preshower system and the grey to the EE crystals.

The choice to put the CMS ECAL system inside of the solenoid was novel. The choice allows measurement of the energy of electromagnetic objects prior to traversing the material of the solenoid. This led to the use of the lead tungstate crystals. Lead tungstate crystals are optically clear, have a high density (8.28 g/cm^3), a short radiation length (0.89 cm) and a small Molière radius (2.2 cm). These features result in a fine granularity and a compact design. The Molière radius is the radius of the cylinder containing 90% of an electromagnetic particles average shower energy.

The crystals have dimensions of 22 mm $\times$ 22 mm $\times$ 230 mm and are grouped in supermodules of 1700 crystals, 20 in ϕ and 85 in η. In general due to the density of the crystals, particles are likely to interact when passing through. At the same time since they are optically clear the light from the interaction is transmitted through the crystal to the photodetectors. In the EB region Avalanche photodiodes are used as photodetectors while vacuum photo triodes are used in the EE region.

The preshower detector is a sampling calorimeter built in two layers. The first layer is a lead radiator to create an electromagnetic shower from the incoming electrons or photons while the second layer is a silicon strip sensor to measure the deposited energy. The preshower has a thickness of 20 cm. The purpose of the preshower is to distinguish neutral pions from real photons with the secondary advantage that it helps identification of minimum ionizing particles and improves the position measurement of electrons and photons with high granularity.

The energy resolution of the ECAL system is given by Equation 3.6.

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.12}{E}\right)^2 + (0.30\%)^2 \tag{3.6}$$

The first term is the stochastic term that is dominated by fluctuations in lateral shower containment, photostatistics, and fluctuations in the energy measured by the preshower. The second term is a noise term which is dominated by electronics noise, digitization noise and event pileup noise. Pileup is the effect of multiple interactions incident at the same bunch crossing or out of time collisions from nearby bunch crossings. Lastly there is the constant term that comes from non-uniformity of the longitudinal light collection, inter calibration errors, and leakage of energy from the back of the crystal. The energy resolution of the ECAL system is very good at better then 1% for a 40 GeV energy deposit.

Hadron Calorimeter

Outside the Tracker and ECAL systems lies the Hadron Calorimeter (HCAL). Like the ECAL and Tracker systems it lies within the CMS solenoid and the 3.8 T magnetic field it produces. The HCAL system is responsible for measuring energy from hadronic showers (jets). It is also very important for measuring missing transverse energy (MET), energy that is carried away by neutrinos which are not detectable in CMS. The HCAL system is split into three subdetectors: the barrel (HB) which extends to $|\eta| < 1.3$, the end cap (HE) which cover $1.3 < |\eta| < 3.0$ and the forward hadron calorimeters which cover the region from $3.0 < |\eta| < 5.2$. A diagram of the HCAL system can be found in Figure 3.6.

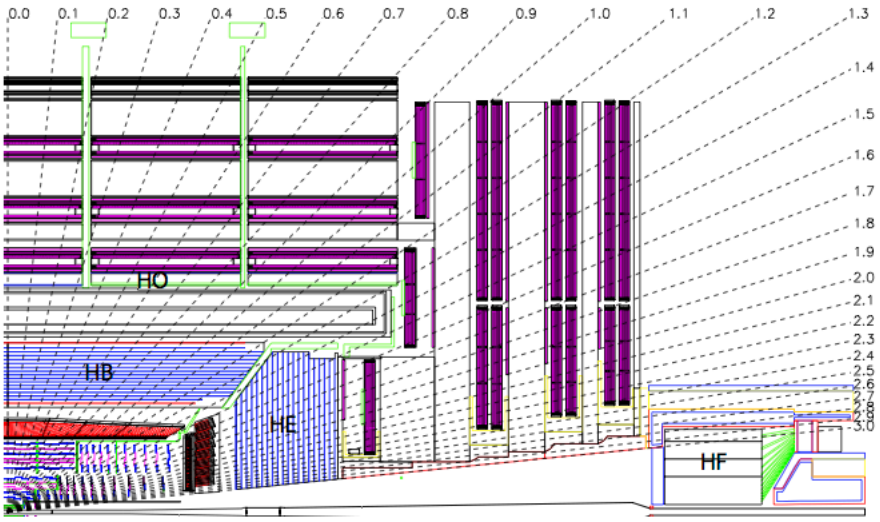

Figure 3.6: A longitudinal diagram of the CMS HCAL system. Dashed lines are at fixed η values.

The HB system is a sampling calorimeter consisting of 36 identical azimuthal wedges made of brass and plastic scintillator each weighing 26 tons. The brass was chosen because it is non-magnetic and it has a short interaction length, the average

length traveled before a particle loses 1/e of its energy. The HE system also a sampling calorimeter using brass and plastic scintillator with 36 end cap wedges. Lastly the HF system, built from steel and quartz fibers, is a Cherenkov-based detector. Cherenkov radiation is emitted when charged particles travel through a medium with a phase velocity faster then the speed of light in that medium. Steel and quartz were chosen for radiation hardness due to the large amount of radiation expected in that forward region.

Hadrons interact through the strong force. As such when they pass through the HB/HE system they interact with the nucleons in the brass producing light which is absorbed by the scintillating fibers connected to hybrid photodiodes which record the energy information. Similarly the light from the Cherenkov radiation is produced when charged particles pass through the HF.

The energy resolution of the HCAL is given by Equations 3.7 and 3.8.

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{90\%}{\sqrt{E}}\right)^2 + (4.5\%)^2 \qquad (HB/HE) \tag{3.7}$$

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{172\%}{\sqrt{E}}\right)^2 + (9.0\%)^2 \qquad (HF) \tag{3.8}$$

The first term is the stochastic term which covers statistical fluctuations and the intrinsic shower fluctuations. The second term is a constant term to cover non-uniformity and calibration uncertainties.

Muon System

Muons are are usually unaffected as they travel through ordinary matter. As such they leave charge tracks in the tracking system and then travel out of the detector. In order to properly identify muons, measure their momentum and for event triggering;

a dedicated muon system is needed. The system is placed outside of the CMS solenoid where it is greater then 10 interaction lengths from the beam line. This is done to distinguish punch through events from real muons. Punch through refers to particles that traverse the detector without interacting. The CMS muons system uses 3 types of gaseous detectors that operate inside a 1.8T return field. In the barrel region covering the range $|\eta| < 1.2$ the system uses drift tube (DT) chambers. In the end cap region there is a system of cathode strip chambers (CSC) that cover the range $0.9 < |\eta| < 2.4$. In addition to the DT and CSC systems there is a system of resistive plate chambers (RPC) that provide fast, independent and highly segmented transverse momentum measurement. The RPC system covers the region of $|\eta| < 1.6$. A side view schematic of the CMS muon system can be seen in Figure 3.7.

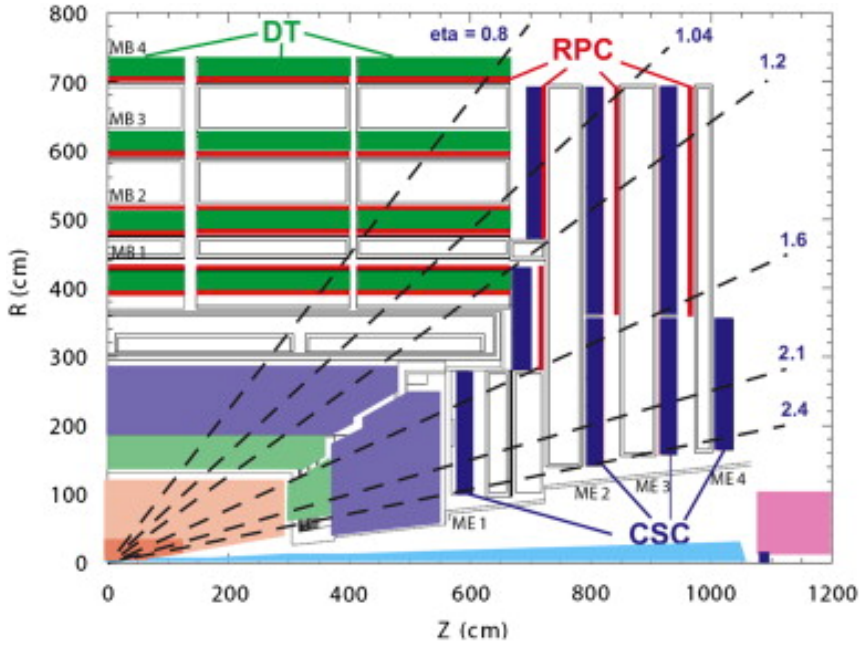

Figure 3.7: A longitudinal diagram of the CMS Muon system. Dashed lines are at fixed η values and show the coverage of the three systems.

The DT system has a total of 4 stations forming concentric cylinders around the beam line which are interspersed among the layers of the flux return plates. The

3 inner cylinders have a total of 60 drift chambers while the outer cylinder has 70 chambers. A drift chamber works by detecting charged particles using a gas that is ionized as the particle traverses the detector. The gas used in the DT is a 85/15 split of Ar/CO_2. The CSC system consists of a total of 486 cathode strip chambers providing 3 or 4 layers of measurement in the CSC range from $1.2 < |\eta| < 2.4$, during the first LHC long shut down in 2013-2014 an additional 72 chambers will be installed providing 4 measurement layers in the whole region. The CSC chambers are also ionizing gas detectors but they are packed with anode wire and cathode planes which provide fast and accurate position information. The CSC system was chosen due to the different particle flux and uneven magnetic fields in the forward region. The RPC system is a complementary system that was designed to tag the time of an ionizing event faster than the 25 ns, the time between 2 consecutive LHC bunch crossings. This provides a dedicated device for triggering muons in the relevant bunch crossing in the high rate of background produced at the LHC. The RPC chambers are built with two thin sheets, an anode and a cathode, separated by a gaseous region. The RPC provides a fast and high resolution measurement of the transverse momentum but doesn't have the spatial resolution of the DT and CSC systems.

The Muon transverse momentum resolution is influenced by several effects including alignment of the tracker and muon systems, composition of material in the tracking region and uncertainty on the magnetic field both inside and outside the CMS solenoid. The final resolution is dependent on $|\eta|$ with resolutions between 1.3% and 2.0% in the barrel and up to 6% in the end cap [35].

Trigger

At the LHC design luminosity it provides proton-proton collisions at a crossing frequency of 40 MHz with approximately 20 pileup interactions. Recording all crossing

would produce a staggering amount of data which cannot possibly all be stored. CMS uses a 2 stage trigger process to reduce the data rate to a reasonable size. The first stage is the Level-1 trigger (L1) which is made of custom designed programmable electronics. The L1 is designed to reduce the 40 MHz input rate to a manageable 100 kHz. The second stage is the High Level Trigger (HLT) which feeds the 100 kHz output from the L1 into a software based system in a filter farm of commercial processors. The HLT has access to complete readout of the collision event and reduces the data rate to what is stored, which has to be less then 1 kHz.

The L1 trigger system starts with Trigger Primitive Generators (TPG) which are based on calorimeter trigger towers and track segments or hit patterns in the muon chambers. The TPG data from the calorimeters are sent to the Regional Calorimeter Trigger (RCT) which is a system with 18 regional VME crates. Each crate consists of 7 receiver cards each covering a region in $\Delta\eta \times \Delta\phi = 5.0 \times 0.7$. The receiver cards in combination with Electron Identification Cards (EIC) take the TPG information and using look up tables (LUT) and ASICs classify and choose the four highest E_T isolated and non-isolated electron/photon candidates to send to the Global Calorimeter Trigger (GCT). In addition to the electron/photon candidates the RCT calculates the four most energetic jets each from the HCAL and the HF regions and forwards them to the GCT. The GCT then performs complex calculations on FPGA cards determining jets, taus, total transverse energy, missing transverse energy and the scalar transverse energy sum of all jets above a programmable threshold (H_T). All three muons systems take part in the L1 trigger. TPG data from the DT and CSC are fed to regional track finding triggers. Information from the regional track finding triggers and the RPC are then forwarded to the Global Muon Trigger (GMT). The GMT then combines all of the information and selects to muons for triggering. An outline of the L1 trigger architecture can be found in Figure 3.8.

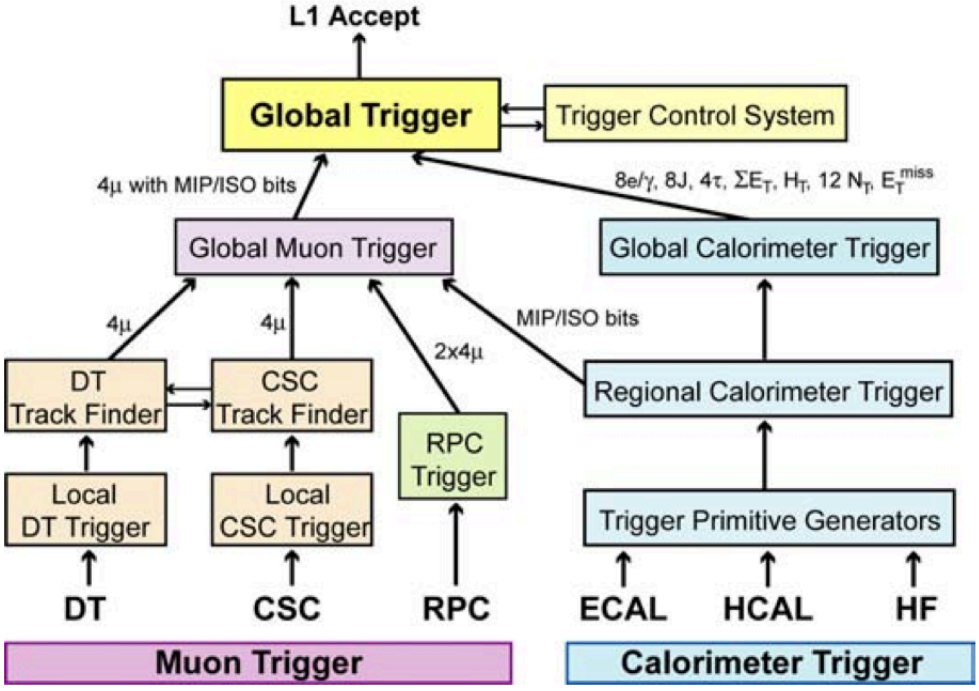

Figure 3.8: Overview of the architecture of the CMS L1 Trigger.

The HLT takes the information from the L1 system for further processing. Significant physics decisions need to be made at the HLT level. The HLT is a software trigger that is run on over 1000 commercial processors. A complete readout of the detector information is available at this step. An HLT menu is composed of multiple trigger algorithms each built out of a set of modules. Each trigger is built on multiple steps, with each step performing a very specific task filtering events as each subsequent module adds in complexity. Reducing the number of events in steps before running more complex algorithms significantly reduces CPU usage. More detailed information about the algorithms used for this analysis will be presented in Section 5.9.

Chapter 4

Event Simulation

Accurate simulation of CMS events is essential in order to characterize details of the events observed in the data and to identify interesting events. The simulation is done is several steps of a chain. It starts with generating the physics event from a proton collision and proceeds to the final particles observed in the CMS detector. The event generation step includes hadronization, underlying event, and event pileup. The next step involves simulating the detector response to the final state particles or detector simulation. The last step, digitization, emulates the detector electronics response to the physics event.

Event generation is done using Monte Carlo (MC) [36] event generators. MC generators used in this thesis include Pythia [37], Madgraph [38], Powheg [39], and TAUOLA [40]. The MC generators use Parton Distribution Functions (PDFs) and Feynman calculus to designate the momentum and decay of various particles produced in the hard scattering process. The hard scattering process is the interaction between the constituents of the proton that produce the physics interaction of interest. The rest of the physics event is characterized with detector simulation and digitization. Final state particles from the generated event are modeled for their interaction with

the CMS geometry and materials.

4.1 Monte Carlo Event Generation

Event generation starts with a matrix element calculation to give a differential cross section for the hard scattering process, also taking into account interference due to processes that have the same initial and final states. In addition to the hard scatter interaction the decay of short lived particles is also handled at this stage. Decays of particles with a finite lifetime are handled by separate programs, one example is the tau lepton which is interfaced with TAUOLA to handle its decay. A further step deals with the colored partons produced in the previous step and describes the hadronization process in which they are turned into colorless particles. This is referred to as parton showering. In a last step a program deals with the underlying event, or the soft interactions of the partons present in the proton that are not involved in the hard scatter process.

In a proton-proton collision each proton carries a known amount of momentum but inside the proton the distribution of the parton momentum is constantly in flux. This means that every hard scatter interaction is unique. Event generators use PDFs which describe the relative probability of each parton to be carrying a particular momentum. Current theories are unable to calculate these distributions accurately. As a result PDFs are measured experimentally using fits to deep inelastic scattering, electroweak boson production studies, and using high energy jet events.

MC event generators randomly sample from the PDFs and then again from the differential cross section calculated from the matrix element to define the momentum of the final state particles. For this analysis most samples were produced with NLO generators from Madgraph and Powheg. Where possible cross sections from other

CMS measurements are used. In other cases detailed theoretical calculations using programs such as MC@NLO [41] or from the LHC Higgs XS WG [42, 43] are used. A list of the various MC samples used for this thesis is given in Table 4.1. The event generator is further interfaced with TAUOLA to handle the decay of the tau leptons. The decay branching ratios come from experimental observations.

Signal Processes			
Dataset Description	MC Generator	7 TeV Cross-Section [pb]	8 TeV Cross-Section [pb]
$gg \to H$	Powheg	15.2	19.6
$qq \to qqH$	Powheg	1.2	1.55
$gg \to t\bar{t}/VH$	Pythia	0.97	1.14
Background Processes			
Dataset Description	MC Generator	7 TeV Cross-Section [pb]	8 TeV Cross-Section [pb]
$t\bar{t}$	Madgraph	165.8	225.2
$t \to X\ (tW)$	Powheg	7.9	11.1
$\bar{t} \to X\ (tW)$	Powheg	7.9	11.1
$Z \to ll$	Madgraph	3048.0	3503.7
W+jets	Madgraph	31314	36257
W+1jet	Madgraph	5042	6381
W+2jets	Madgraph	1628	2030
W+3jets	Madgraph	344	616
W+4jets	Madgraph	188	254
WW $\to$ 2l2ν	Madgraph	4.8	5.82
WZ $\to$ 3lν	Madgraph	0.86	1.06
WZ $\to$ 2l2q	Madgraph	1.79	2.21
ZZ $\to$ 2l2q	Madgraph	0.78	1.25
ZZ $\to$ 4l	Madgraph	0.06	0.2

Table 4.1: Samples of simulated events used for the analysis. The production cross sections for the SM Higgs boson are given for $m_H = 125$ GeV.

In addition to the hard scatter event further simulation is needed to account for initial and final state radiation (ISR/FSR). ISR/FSR refer to photons radiated from charged particles or gluons radiated from quarks or other gluons before/after the primary hard scatter interaction. Additionally quarks produced in the hard scatter interaction or in the underlying event need to be made into colorless objects as they are observed in nature. These processes are handled by interfacing the Madgraph/Powheg event generators with Pythia. Pythia handles the hadronization of the quarks using

the Lund String Model [44]. Strings are used to simulate quark/antiquark pairs. The strings break to form new pairs. Additionally, new quarks and mesons can be formed by combining with adjacent strings. A schematic example of a parton shower can be seen in Figure 4.1. The resulting streams of particles from the parton shower generally form in collimated groups that we refer to as jets.

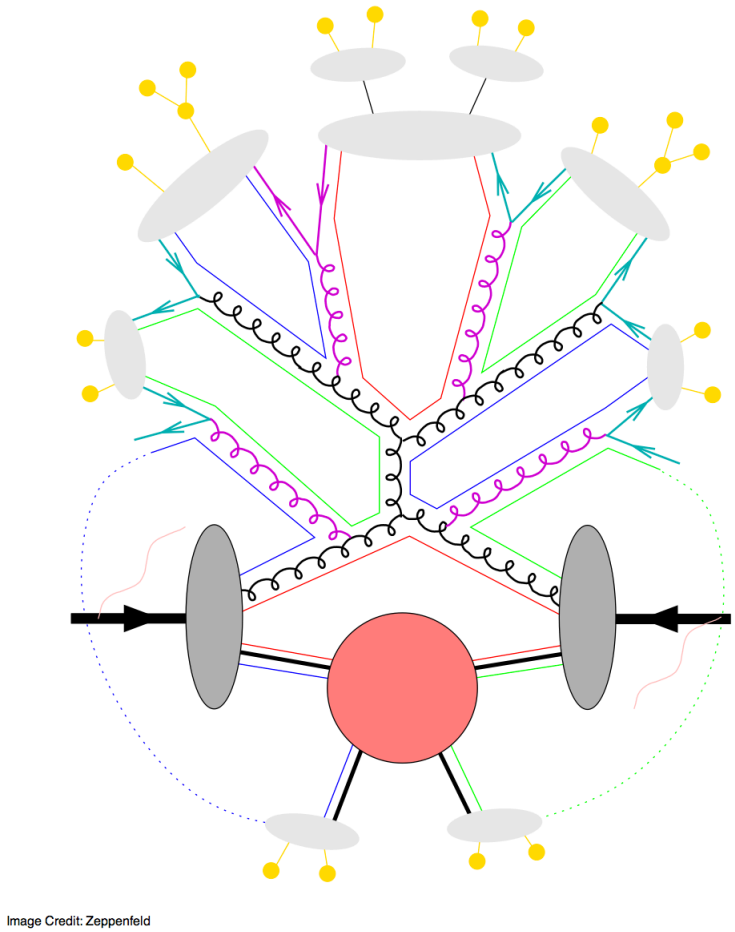

Figure 4.1: A schematic example of a parton shower in a gluon-gluon fusion to di-jet event.

In addition to the hard scatter interaction there can be significant energy coming from the underlying event. The underlying event refers to interactions between spectator partons in the proton. The additional particles from the underlying event are

generally soft but can effect the hadronization from the hard scatter, they are added to the event at the same time.

Due to the large number of protons in a collision bunch and also the possibility of collisions from a bunch before or after the incident bunch crossing, additional event pileup must also be considered. For simulated events a number of soft interactions are injected on top of the primary physics event. Simulated events are additionally weighted to the same pileup fraction observed in the data. The data pileup is derived by using the per bunch-crossing-per-luminosity section instantaneous luminosity with the total pp inelastic cross-section to generate an expected pileup distribution. The distribution of the number of reconstructed vertices, after pileup reweighing, showing the agreement between data and MC is shown in Figure 4.2.

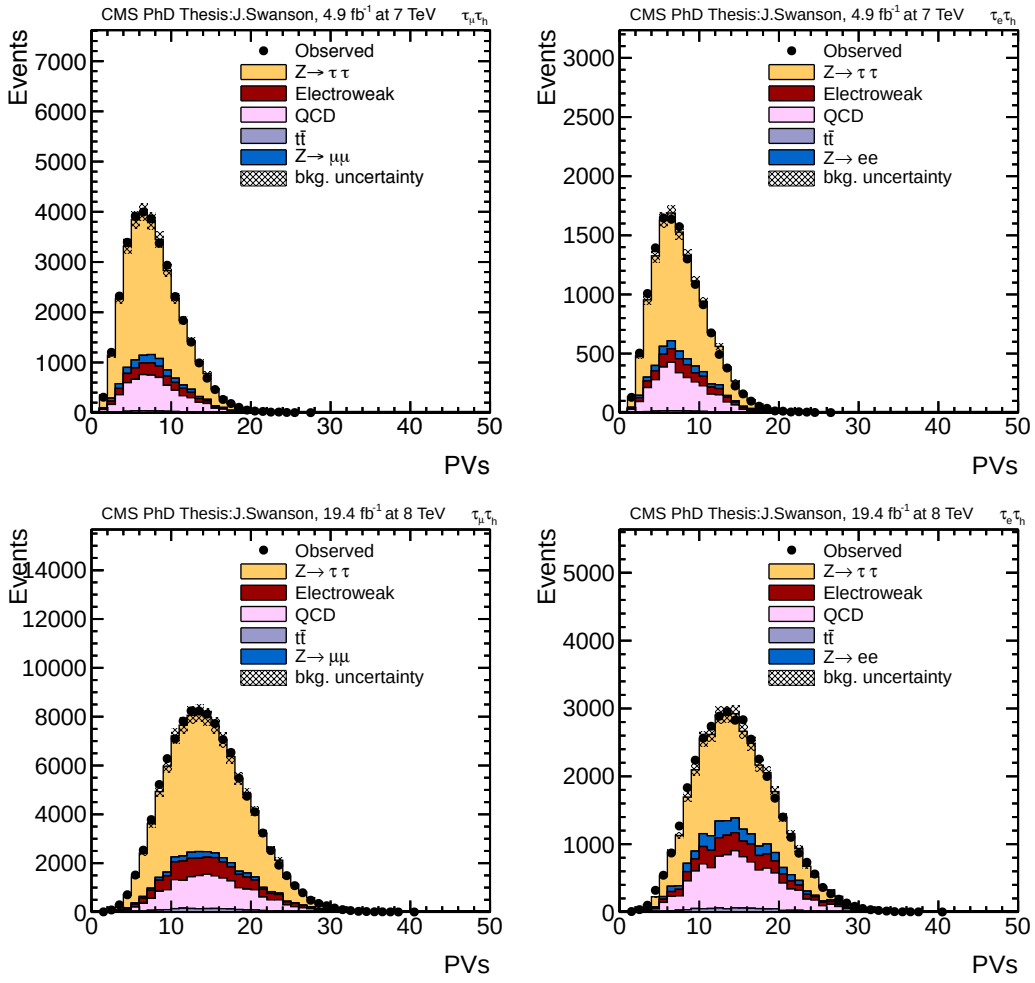

Figure 4.2: Expected and observed number of reconstructed primary vertices (upper row) for the data collected in 2011 and (lower row) for the data collected in 2012 (left) $\tau_h\tau_\mu$ and (right) $\tau_h\tau_e$. The background components are scaled to the results of a fit to the inclusive mass distribution. The uncertainty on the background is represented in the shaded region and comes from the fitted nuisances.

4.2 Detector Simulation

Additional simulation of the event is needed to model the interaction with the detector and the response of the electronics to the final state particles. This simulation is done using the GEANT4 [45] toolkit. It takes a detailed description of the detector materials including sensitive detector, parts of the detector with sensor readout, and dead materials such as cabling and cooling components. The particle interaction with the material is accurately modeled taking into account energy loss and secondary particles produced in the interaction. Additionally accurate information about the magnetic field is used to calculate particle trajectories. The final set of simulated particles are then input to emulators to estimate the response of the detector to their presence. Detector noise and other factors are included at this stage. Finally the information is output in the same format as that produced by the actual detector data to be input to the CMS software framework (CMSSW) in such a way the processing of data and simulation is nearly identical during the subsequent stages of analysis.

Chapter 5

Event Reconstruction

Collision event information is stored in a raw format containing the information output from the various sub detectors. This information can be then used in the online filter farm for making HLT decisions and offline to reconstruct the final state particles used in offline analysis. The analysis techniques include various algorithms for distinguishing the different particles including electrons, muons, charged hadrons, neutral hadrons and photons. These final state particles can then be input to more advanced algorithms to reconstruct jets, taus and other general event quantities such as missing transverse energy (MET).

5.1 Track and Vertex Reconstruction

Efficient charged track reconstruction is very important for reconstructing electrons, muons, taus and jets. Additionally distinguishing tracks coming from a primary event vertex from those due to pileup interactions is important for isolating those particles from jets efficiently.

Track reconstructions starts with local reconstruction. In local reconstruction

zero-suppressed pixels or strips are clustered into hits measuring their position and uncertainty. The hits are then input to the CMS tracking software which uses a combinatorial track-finding algorithm. The final tracks are produced through multiple iterations of the software. The software has a total of 6 iterations, where each iteration seeks to reconstruct tracks of increasing complexity. The first iteration produces most of the tracks used in analyses. They are prompt and have $p_T > 0.8$ GeV and 3 pixel hits. The second iteration can recover prompt tracks that have only 2 pixel hits. The third iteration handles prompt tracks of lower p_T. The rest of the iterations aim to find tracks that originate outside of the beamspot and recover any tracks not found in previous iterations. The beamspot is the luminous region produced by the proton-proton collisions.

Each iteration consists of four steps: seed generation, track finding, track fitting, and track selection. The seed generation provides the initial tracks and gives an estimate of the trajectory and its uncertainty. The track finding extrapolates the seed trajectories along the track's expected path, searching for additional hits that can be associated with the track candidate. The track finding is based on a global Kalman filter [46]. The track fitting is used to provide the final trajectory of the track candidate using a Kalman filter and smoother. In the last step, the track are selected, the quality of the track quantified and tracks not meeting the appropriate quality for a given iteration are discarded. More details about the tracking algorithm and its performance can be found in [47].

The assignment of a vertex is done in 3 steps: track selection, clustering of the tracks, and fitting the position. The track selection uses tracks of a given quality produced during the track reconstruction. The clustering refers to assigning a set of tracks that originate from the same interaction vertex. The clustering uses the deterministic annealing (DA) algorithm [48]. The DA algorithm attempts to find a

global minimum in a problem with many degrees of freedom. The mathematics used is similar to a thermodynamic system reaching its minimal energy. The finial step uses the cluster to fit the final position of the vertex.

The primary vertex used in this analysis is required to be within 24 cm of the center of the detector, with 2 cm in ρ with respect to the beam spot and have more then 4 degrees of freedom.

5.2 Electron Reconstruction

Electrons are reconstructed using track candidates produced as they traverse the tracking system along with their energy deposits in the ECAL system. The reconstruction is done using two distinct algorithms: ECAL-seeded electrons and track-seeded electrons. The track-seeded electrons start with a track candidate and then extrapolate, looking for bremsstrahlung radiation in the tracking system, to the ECAL system. The track-based reconstruction is mostly for low p_T electrons and those produced inside jets and are not important for this analysis. ECAL seeded electron reconstruction starts from an ECAL energy deposits (clusters) and extrapolates into the tracking system.

ECAL clusters are selected with energy of at least 4 GeV which then are associated back to track seeds from the inner tracker layers. The clusters are wider in ϕ to collect energy lost to bremsstrahlung radiation as the electrons bend through the magnetic field and due to the presence of tracking material. A graphical representation of the electron reconstruction can be seen in Figure 5.1. The electron trajectories are reconstructed using a model to account for the energy lost in the tracker material and then fit using a Gaussian Sum Filter (GSF) algorithm [49]. The transverse momentum assignment comes by combining information from the ECAL cluster energy and the

momentum from the GSF track.

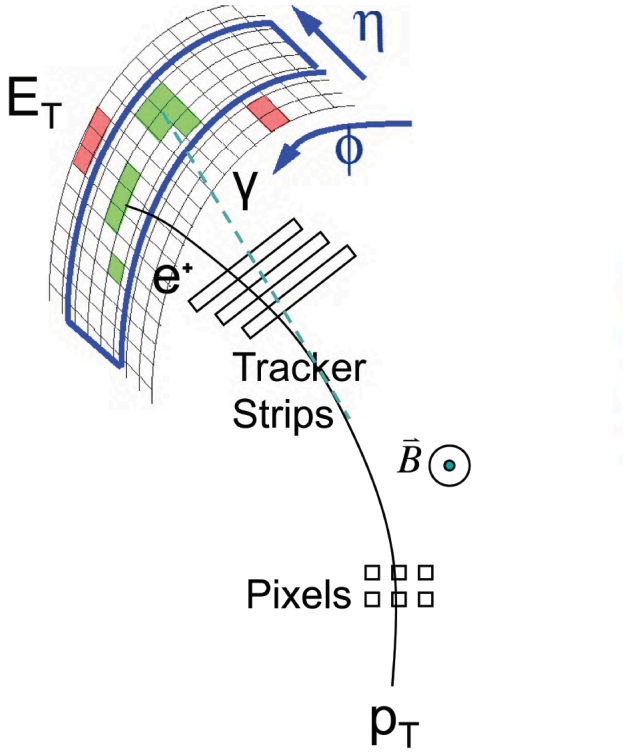

Figure 5.1: A graphical representation of an electron produced in the CMS detector. Notice the bending of the electron inside the magnetic field and the subsequent energy lost due to bremsstrahlung radiation which spreads the ECAL supercluster in the ϕ direction.

After reconstruction, real electrons have to be distinguished from photons and charged hadrons. This identification is done using a boosted decision tree (BDT) discriminator that uses track quality, cluster shape, and kinematic variables. The BDT was trained using real electrons in $Z \rightarrow ee$ events and misidentified electrons from the jets in $Z \rightarrow ee + Jet$ events. The BDT uses a total of 19 variables:

- The normalized χ^2 of the common track fit, the number of valid hits in the track fit, the normalized χ^2 of the *GSFTrack* fit.

- The distance in η ($\Delta\eta_{SC}(\text{Track}_{vtx})$) and ϕ ($\Delta\phi_{SC}(\text{Track}_{vtx})$) between the reconstructed super cluster in the calorimeter and the track evaluated at the primary vertex position

- The distance in η between the super cluster seed and the track evaluated at the calorimeter surface.

- The cluster shape variables $\sigma_{i\eta,i\eta}$ and $\sigma_{i\phi,i\phi}$, where $i\eta$ ($i\phi$) indicate the integer label of the electromagnetic calorimeter cell in η (ϕ), the cluster shape variable $f_e = 1 - e1X5/e5X5$, where $e1X5$ ($e5X5$) indicate the energy deposition in an array of 1×5 (5×5) cells in the vicinity of the super cluster seed, the cluster shape variable $R9 = e3x3/E_{SC}$, where $e3x3$ and E_{SC} indicate the energy in an array of 3×3 cells in the vicinity of the super cluster seed and the raw energy of the reconstructed super cluster.

- The ratio of the hadronic energy over electromagnetic energy of the super cluster (H/E), the ratio of the super cluster energy over the momentum of the associated track evaluated at the selected primary vertex (E/P), the variable $1/E_e - 1/p_e$, where E_e and p_e indicate the reconstructed energy and momentum of the electron candidate, the ratio of the electron cluster over the momentum of the associated track and the ratio of the seed cluster over the associated track, where each time the track momentum has been evaluated at the surface of the calorimeter.

- The ratio of the energy that has been reconstructed in the pre-shower detector over the raw energy of the reconstructed super cluster. The momentum and η of the reconstructed electron candidate.

The electron candidate is considered to be identified if the BDT is above a threshold depending on its p_T and η. The values used are shown in Table 5.1. In addition to the BDT some additional criteria is applied to reject electrons coming from pileup or photon conversions. To reduce the number of electrons coming from pileup, candidate electrons must be within $d_z < 0.1$ cm from the selected primary vertex and have a $d_0 < 0.045$ cm in the plane perpendicular to the z direction. To remove electrons from photon conversions there is an additional requirement that there are no missing hits in the inner layers of the pixel detector and a vertex fit probability of more then $P > 10^{-6}$.

	BDT Discriminator Value (>)		
	$\|\eta\| < 0.8$	$0.8 \leq \|\eta\| < 1.479$	$1.479 \leq \|\eta\|$
$p_T > 20$ GeV	0.925	0.975	0.985

Table 5.1: Thresholds for the BDT discriminator to identify electrons. For an identified electron the discriminator value has to fall above the indicated threshold.

5.3 Muon Reconstruction

Muon reconstruction in CMS is achieved using one of four algorithms: stand-alone muons, tracker muons, global muons or calo-muons. The stand-alone muons use information from the muon systems, while tracker muons rely on information from the inner tracking system. Global muons use a combination of information from the muon systems and the tracking system. Lastly calo-muons use information from the tracking system and then looking for calorimetric signatures compatible with a minimum ionizing particle. Tracker muons and calo-muons are used mostly for low p_T muons that are unlikely to leave significant information in the muon system.

The muon reconstruction starts with a local reconstruction. In this step, hits are

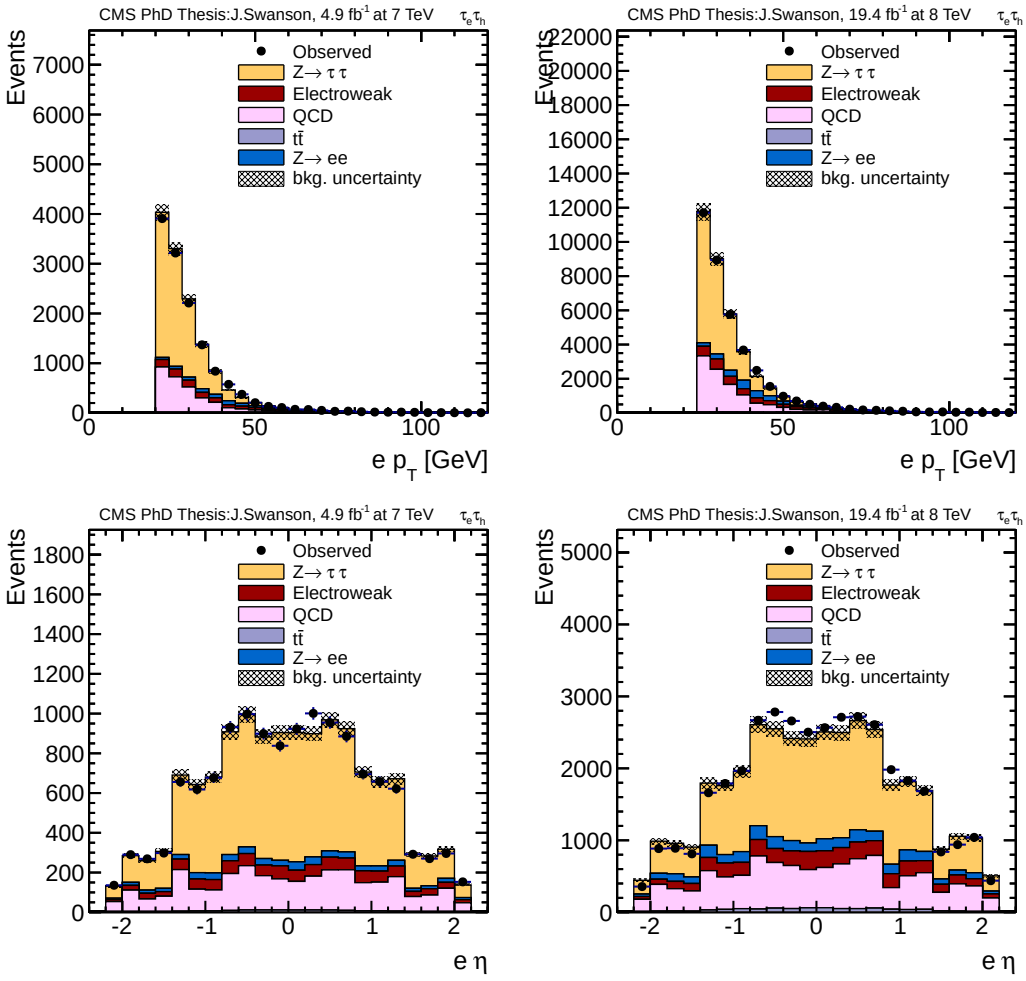

Figure 5.2: Distribution of the electron transverse momentum and psuedorapidity from the $\tau_e\tau_h$ channel. The background components are scaled relative to the results of a fit to the inclusive mass distribution. The uncertainty on the background is represented in the shaded region and comes from the fitted nuisances.

produced in the DT, CSC, and RPC systems then the hits in the DT and CSC systems are matched to form segments. The stand-alone muon reconstruction uses the segments and hits as seeds for muon candidates. The position, direction and momentum information from these seeds are then fit using a Kalman filter technique [46]. In addition to the muon system information the stand-alone muons can use the beam spot location in the fit to improve the muon resolution. The global muons are then reconstructed starting with the collection of stand-alone muons and looking for compatible tracks in the tracking system. The best matching track is selected considering the compatibility of the position and momentum in the tracker and muon systems. For global muons the tracker hits are combined with the muon segments to perform a fit again using a Kalman filter technique [46]. This combination gives the best measurement of the muon momentum.

Prompt muons need to be distinguished from substantial background, mostly from heavy flavor quark decays. The muons produced in heavy flavor decays generally are produced away from the beam spot due to the lifetime of the b quark. Also the muon will usually be accompanied by soft hadrons from the associated jet's fragmentation. This can be rejected using isolation to be discussed in Section 5.5. A tight requirement is used to identify muons in this analysis. First muons are required to pass both the global and tracker muon reconstruction and reconstructed by the particle flow algorithm, to be discussed in the next section. Additionally the muon candidate is required to have $d_z < 0.1$ cm, $d_0 < 0.045$ cm, to have a $\chi^2/ndof < 10$ for the global track fit, at least one segment in the muon detector included in the global track fit, muon track segments in at least 2 stations of the muon system, at least one hit in the pixel detector, and hits in more then 5 layers of the inner track detector.

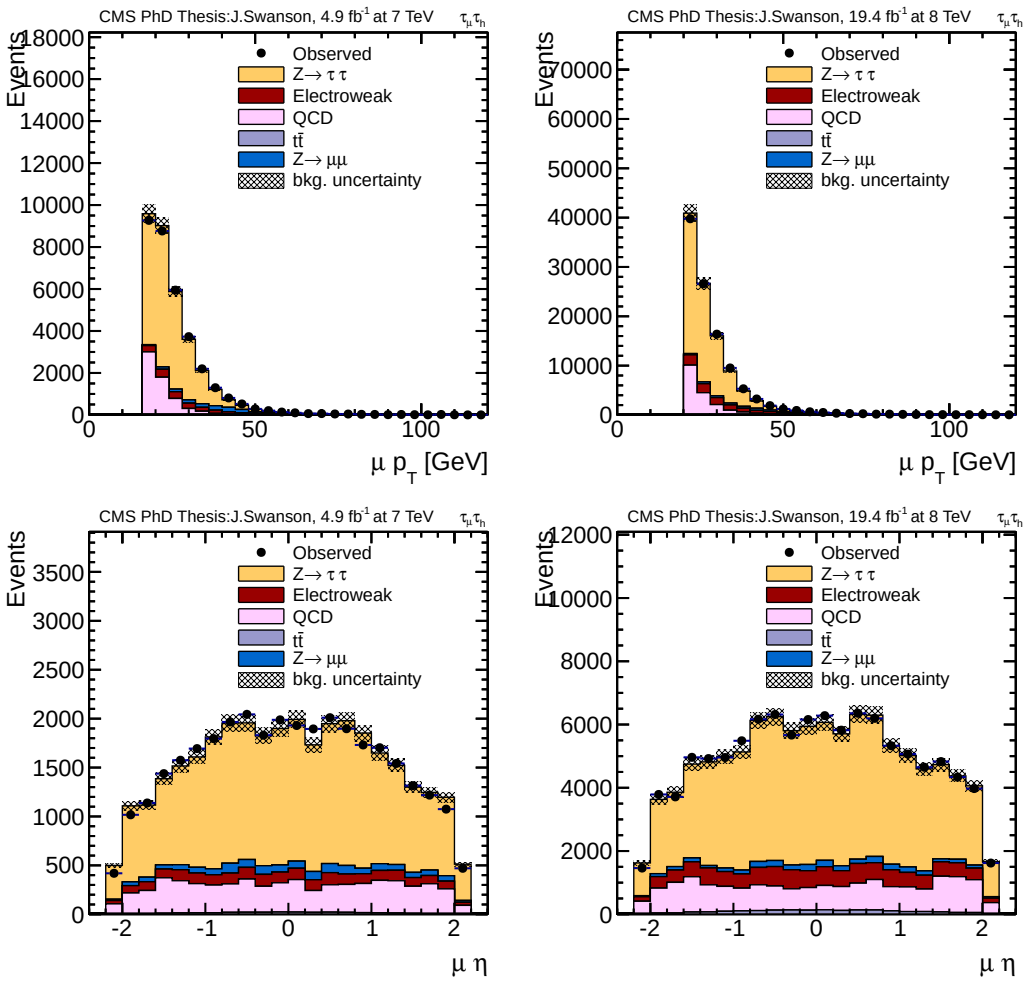

Figure 5.3: Distribution of the muon transverse momentum and psuedorapidity from the $\tau_\mu\tau_h$ channel. The background components are scaled relative to the results of a fit to the inclusive mass distribution. The uncertainty on the background is represented in the shaded region and comes from the fitted nuisances.

5.4 Particle Flow

The CMS particle flow (PF) algorithm attempts to give a global description of a collision event. The PF algorithm uses information from all sub detectors in order to reconstruct individual stable particles, including muons, electrons, photons, charged hadrons and neutral hadrons. The individual identified particles then can be combined together to build composite objects including jets, hadronic tau leptons, and the events missing transverse energy. The particles from the PF algorithm can also be used for the isolation of other objects.

The PF algorithm starts by identifying electrons and muons. The tracks and calorimeter deposits determined to be compatible with the reconstructed electrons and muons are removed from the particle lists for identification of the subsequent photons and hadrons. In the reconstruction of the hadrons and photons calorimeter deposits are linked to tracks in the inner tracking system to identify charged hadrons. If the energy in the ECAL/HCAL systems is larger then that associated with the measured track additional photons/neutral hadrons are produced from outlying clusters near the energy deposit. All remaining tracks that are not linked to calorimeter deposits are identified as additional charged hadrons. A illustration showing the reconstruction of PF vs. traditional detector objects is shown in Figure 5.4. A more detailed description of the PF algorithm and its performance can be found in [50].

5.5 Lepton Isolation

Electrons and muons used in this analysis are required to be isolated from other soft particles to reduce backgrounds from multi jet and heavy flavor events. The isolation is done using particle flow objects; charged particles including other leptons, neutral hadrons and photons. The particles are summed up in a cone of $\Delta R < 0.4$ around

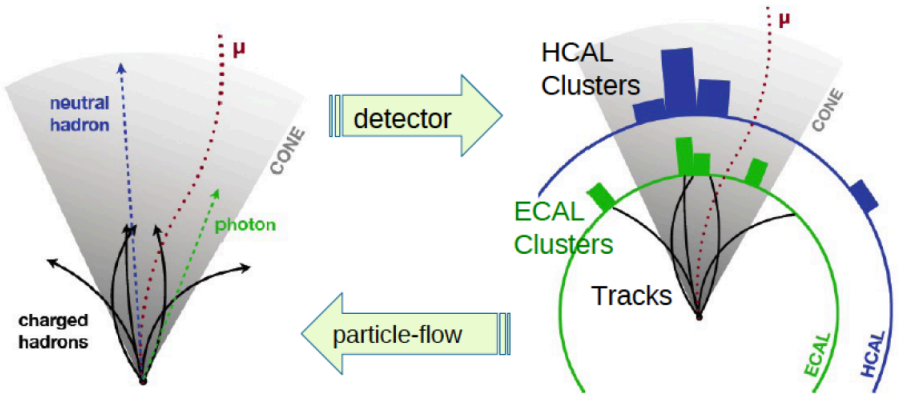

Figure 5.4: An illustration showing how PF objects are identified from calorimeter information (left) compared to traditional detector based reconstruction (right).

the lepton axis excluding an inner veto cone to remove the lepton candidate. A summary of the veto cones used is given in Table 5.2. The photon and neutral hadron candidates are required to have $E_T > 0.5$ GeV for the muon isolation and no minimum for electron isolation.

	Charged Part. Cand.	Phot. Cand.	Neutr. Had. Cand.	Charged Part. (PU)
μ	0.0001	0.01	0.01	0.01
e (EB)	0.01	0.08	none	none
e (EE)	0.015	0.08	none	none

Table 5.2: Summary of the veto cones, in ΔR, used for the calculation of relative isolation for electrons and muons.

The charged particles used for the isolation calculation are required to come from a vertex compatible with the lepton in order to not count particles from pileup interactions. This is done by checking the dz of the track candidate compared to the primary vertex. In addition since one cannot determine the origin of photon and neutral hadrons a correction is done to remove energy coming from pileup interac-

tions. This is done by summing up the track energy not associated with the lepton candidate vertex in a cone of $\Delta R < 0.4$ around the lepton axis. Assuming a 2:1 ratio for charged:neutral particles the pileup is subtracted from the sum of the neutral isolation. This is referred to as a delta beta ($\Delta\beta$) pileup correction.

$$I_{\text{rel}} = \frac{\sum p_T(\text{q}) + \max\left(\sum E_T(\text{neut}) + \sum E_T(\gamma) - 0.5\sum p_T(q_{pu}), 0\right)}{p_T(\mu \text{ or } e)} \tag{5.1}$$

5.6 Jet Reconstruction

The reconstruction of jets is important for this analysis. Jets are used to distinguish different production modes of the Higgs boson in order to increase signal significance. The jets are reconstructed using particle flow objects and then clustered using the anti-k_t algorithm [51]. An additional ID is applied to reduce jets coming from pileup interactions. Lastly jet energy scale corrections (JEC) are applied to account for any residual effects due to pileup. The selection of heavy flavor jets is also preformed using so called b-tagging algorithms.

The anti-k_t algorithm takes hard hadrons and then clusters other soft hadrons within a conical region to form jets. For this analysis a cone of $\Delta R < 0.5$ was used. The anti-k_t algorithm will always cluster soft particles with nearby hard objects instead of clustering a set of soft objects together to form their own jet. In the case there are hard objects within 2 conical regions that overlap with each other the algorithm will do one of three things. If $k_{t1} > k_{t2}$ jet 1 will be conical and jet 2 will be conical except subtracting of the region encompassed by the conical region of jet 1. If $k_{t1} \approx k_{t2}$ the neither jet will be conical with part of each cone belonging to the other jet. Lastly if the two hard hadrons are within the same conical region they will be combined into one jet.

Further identification is done to distinguish jets from the hard interaction and those coming from pileup. This is done using a BDT. The BDT uses compatibility of the tracks with the primary vertex, jet shape variables, and the multiplicity of the neutral and charged components within the jet. In this analysis jets are required to have $p_T > 30$ GeV, $|\eta| < 4.7$ and not be overlapping with any selected lepton within the jets cone. This selection gives an efficiency >99% in the tracker region and ≈95% in the forward region.

A number of corrections are then applied in order to adjust the energy of the reconstructed jets. L1 Fast jet corrections account for energy coming from underlying event or from pileup. These corrections are done by taking an average energy density in the detector and subtracting it from the jet. Additional corrections are applied to correct for non-uniformity and non-linear response in the detector and finally residual corrections to account for small differences in p_T and η between the data and simulation. More details about jet energy corrections can be found in [52].

Jets from heavy flavor decays are further identified using a combined secondary vertex (CSV) algorithm [53]. The algorithm combines information about impact parameter significance, the secondary vertex and jet kinematics to identify b quark jets. Jets are required to have an output value of the CSV discriminator > 0.679 and have a $p_T > 20$ GeV and $|\eta| < 2.4$. These jets are used in this analysis to reject events coming from $t\bar{t}$ events that often include b jets in their decay. These type of jets are referred to as b-tagged jets.

5.7 Missing Transverse Energy

There are many different particles that are produced in proton collisions which decay in part to neutrinos, tau leptons and W bosons for example. Neutrinos are not

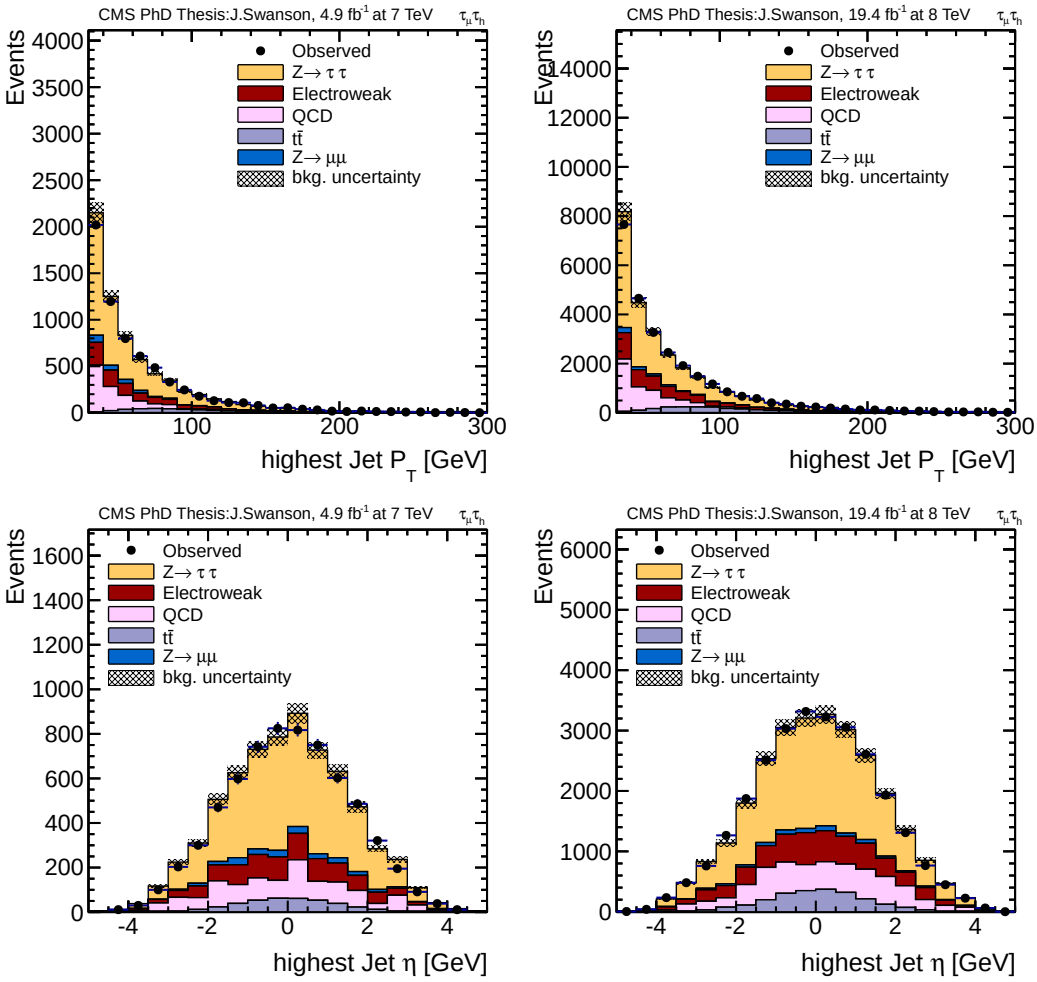

Figure 5.5: Distribution of the leading jet transverse momentum and psuedorapidity from the $\tau_\mu\tau_h$ channel for events containing at least one jet. The background components are scaled relative to the results of a fit to the 1 jet mass distribution. The uncertainty on the background is represented in the shaded region and comes from the fitted nuisances.

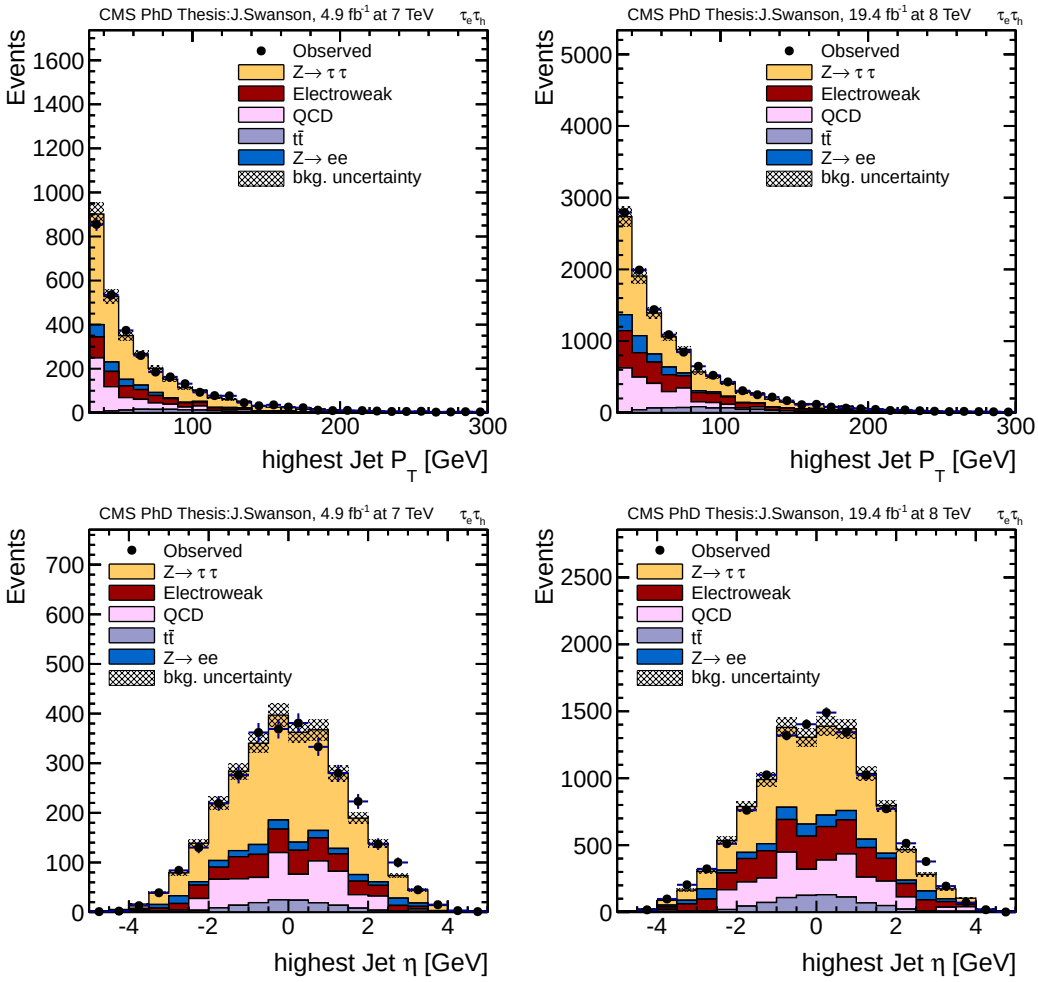

Figure 5.6: Distribution of the leading jet transverse momentum and psuedorapidity from the $\tau_e\tau_h$ channel for events containing at least one jet. The background components are scaled relative to the results of a fit to the 1 jet mass distribution. The uncertainty on the background is represented in the shaded region and comes from the fitted nuisances.

detectable with the CMS detector but their signature can be detected by looking for missing transverse energy. Basically even though the neutrinos are not detectable they carry momentum away from the collision. The initial transverse momentum in the collision is nearly zero so a missing transverse energy (MET) can be calculated using the total visible transverse momentum.

The MET is used to discriminate events containing Higgs bosons from background events with W bosons in association with jets or $t\bar{t}$ events. Also the MET is used to reconstruct the full mass of the tau pair system which is used to separate Higgs from Drell Yan tau pair production. A total of five different missing transverse energy (MET) variables are calculated from particle-flow candidates using the information of the jet identification as described above. These MET variables are:

- The negative vectorial sum of all particle-flow candidates (particle-flow MET).
- The negative vectorial sum of all tracks that have been associated to the selected primary vertex.
- The negative vectorial sum of all tracks that have been associated to the selected primary vertex and all neutral particle-flow candidates within those jets that have passed the jet identification as described above.
- The negative vectorial sum of all tracks that have not been associated to the selected primary vertex and all neutral particle-flow candidates within those jets that have failed the jet identification as described above.
- The negative vectorial sum of all tracks that have been associated to the selected primary vertex and all neutral particle-flow candidates (also those that have not been clustered into jets) plus the *positive* vectorial sum of all neutral particle-

flow candidates within jets that have failed the jet identification as described above.

For each MET variable the vectorial recoil is calculated as defined by:

$$\vec{u} = \text{MET} \cdot \hat{\phi} - \sum_i \vec{p_T}^{\,\text{lep}} \tag{5.2}$$

where $\hat{\phi}$ corresponds to the direction of MET in the transverse plane of the detector and $\vec{p_T}^{\,\text{lep}}$ to the p_T vector of the leptons originating from the hard interaction (which in this case corresponds to the respective e, μ or τ_h lepton pair).

The corresponding magnitude, the azimuthal angle ϕ of the recoil and the scalar $\sum E_T$ of each respective MET variable, the momentum vectors of the two leading jets and the number of primary vertices are added to a multivariate boosted decision tree (BDT) regression, which is used to compute a correction to both the angle and magnitude of the particle flow recoil to match the true recoil.

The training on the recoil has been performed on a sample of simulated $Z \to \mu\mu$ events and validated on a selection of $Z \to \mu\mu$ events in data. The final corrected recoil is added to the vector sum of the leptons following equation 5.7 to give the corrected MET used further in the analysis. More details about CMS MET determination and performance can be found in [54].

5.8 Tau Reconstruction

The performance of the hadronic tau identification algorithm is of prime importance for analysis involving decays to tau leptons. Since about two thirds of the tau leptons decay to hadrons and due to the high rate of jet production at the LHC, managing the tau misidentification rate while maintaining a high reconstruction efficiency

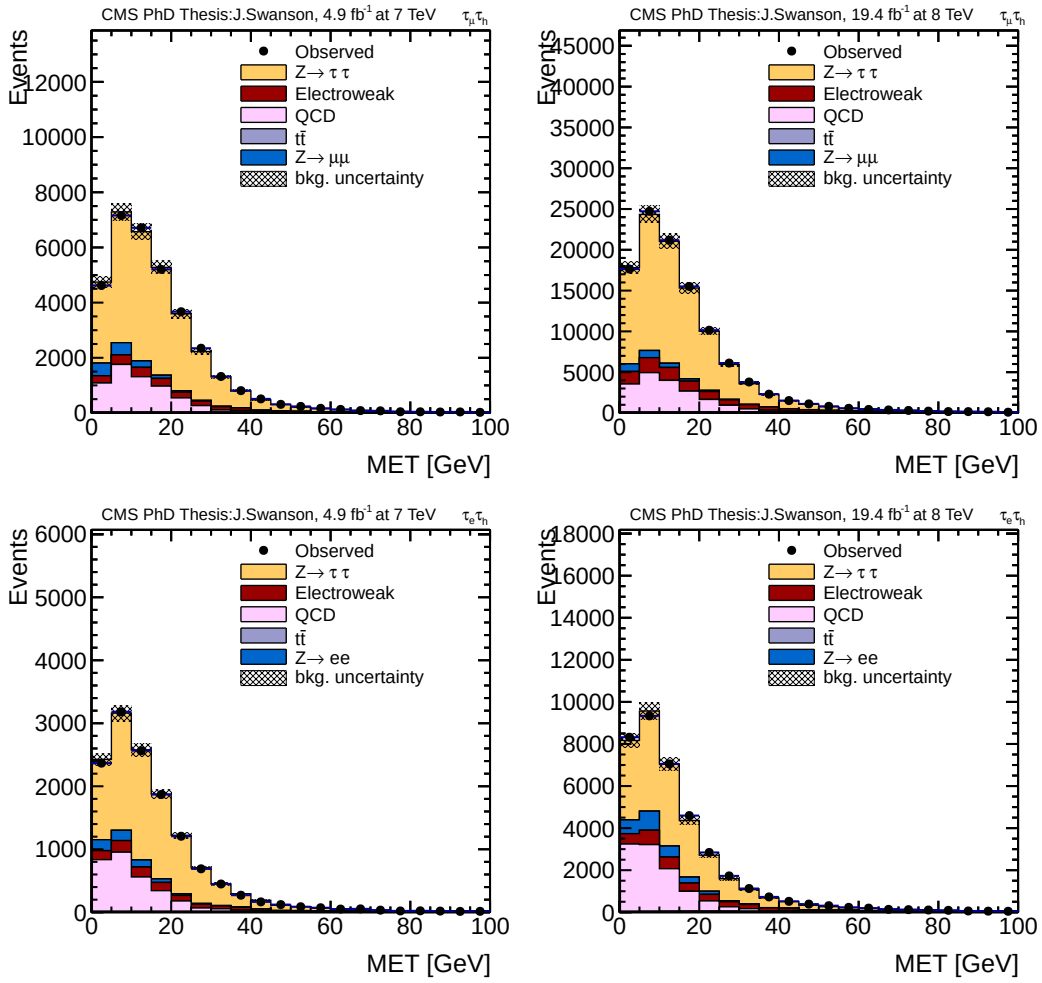

Figure 5.7: Distribution of the missing transverse momentum from the $\tau_\mu\tau_h$ (top) and $\tau_e\tau_h$ (bottom) channels. The background components are scaled relative to the results of a fit to the inclusive mass distribution. The uncertainty on the background is represented in the shaded region and comes from the fitted nuisances.

is challenging. In CMS the tau identification algorithm is called hadron plus strips (HPS) [55]. HPS attempts to identify hadronic tau decays by searching for the resonances of the intermediate decay products. As discussed in Section 1.2 the majority of tau decays involve either a lepton, a charged hadron, a charged hadron plus 1 or more neutral hadrons, or three charged hadrons. The two resonant decays of the tau include the majority of these hadronically decaying events; an a_1 meson with a mass of 1260 MeV or a ρ meson with a mass of 770 MeV.

The HPS tau algorithm uses particle flow objects, charged hadrons and neutral clusters from the calorimeters. Electromagnetic clusters are used to reconstruct strips representing π^0 decays. The strips are regions elongated in ϕ to collect energy spread from photon conversions. The strip has a size of $\Delta\eta = 0.05 \times \Delta\phi = 0.20$. The charged hadrons and the strips are reconstructed in a narrow cone, $\Delta R = 2.8/p_T\tau$ with a maximum/minimum size of 0.1/0.05, into one of four decay modes:

- Single Hadron: Reconstruction of a single hadron where the tau candidate decayed directly to a charged hadron or in the presence of soft neutral pions.
- Hadron plus One Strip: Reconstruction of a single charged hadron plus a neutral pion. The charged hadron and the strip are required to have an invariant mass compatible with the ρ meson with 0.3 GeV $< M_\tau <$ 1.3 GeV.
- Hadron plus Two Strips: Reconstruction of a single charged hadron plus a neutral pion where the pion energy is separated on the calorimeter surface. The charged hadron and the strip are required to have an invariant mass compatible with the ρ meson with 0.4 GeV $< M_\tau <$ 1.2 GeV. Additionally the mass of the two strips is required to be between 50 MeV and 200 MeV.
- Three Hadrons: Reconstruction of a tau to three charged hadrons. The three charged hadrons are required to have a sum of one unit of electron charge and

have an invariant mass compatible with the a_1 meson with 0.8 GeV $< M_\tau <$ 1.5 GeV.

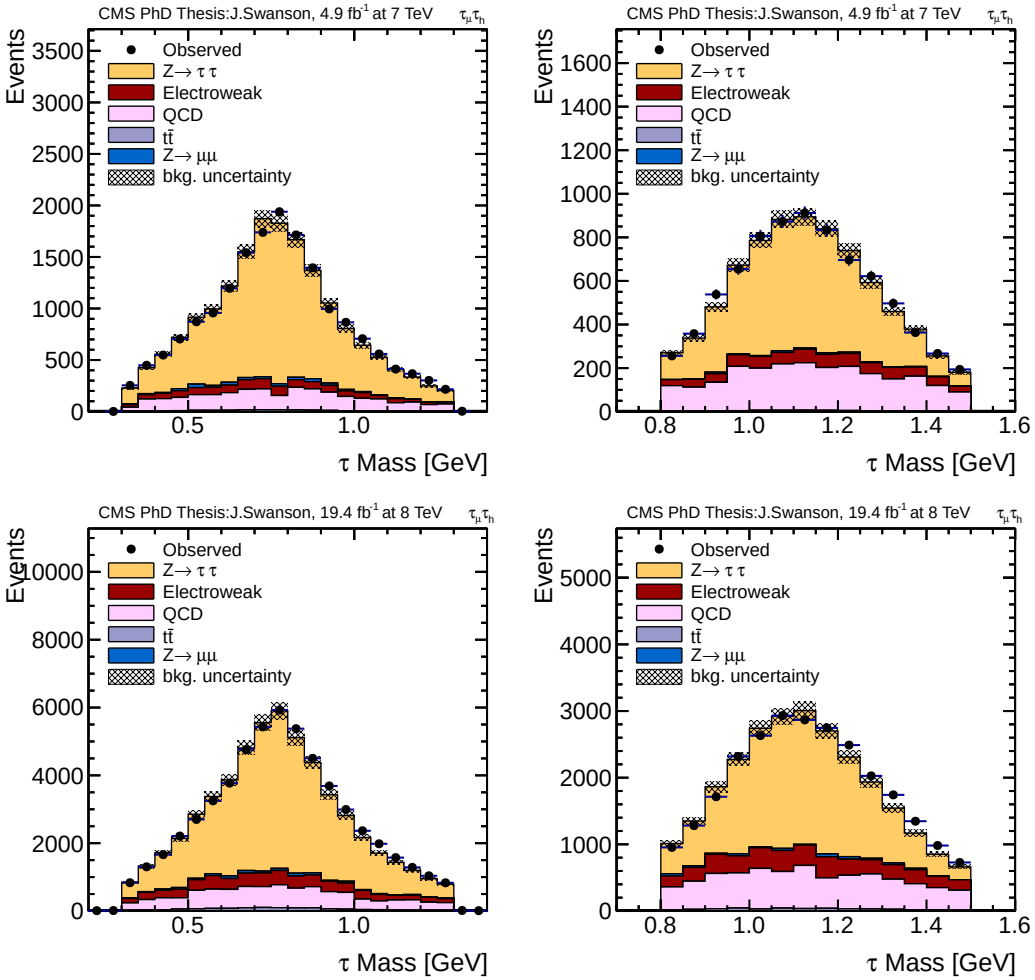

Figure 5.8: Distribution of the τ_h mass for Hadron plus One Strip events (Left) and Three Hadron events (Right) for 2011 (Top) and 2012 (Bottom) data. The plots come from 7 TeV data in the $\tau_\mu\tau_h$ channel. The background components are scaled relative to the results of a fit to the inclusive mass distribution. The uncertainty on the background is represented in the shaded region and comes from the fitted nuisances.

After the decay mode reconstruction the tau candidates are required to be isolated. The isolation is done in a cone of $\Delta R = 0.5$ around the tau candidate axis. All charged

hadrons and photons in the cone contribute to the isolation sum after subtracting the specific objects used in the decay mode that was reconstructed. Much like the lepton isolation discussed in Section 5.5 the neutral electromagnetic isolation needs to be corrected for pileup. Also like the electron/muon isolation a $\Delta\beta$ correction is applied but with some small differences. The pileup is estimated in a larger cone of 0.8 to increase statistics. The final isolation is given by Equation 5.3 where $f^{\Delta\beta} = 0.458$. Taus selected in this analysis are required to have isolation less then 2.0 GeV.

$$I = \sum p_T(\mathrm{q}) + \max\left(\sum E_T(\gamma) - f^{\Delta\beta}\sum p_T(q_{pu}), 0\right) \tag{5.3}$$

Another issue affecting the reconstruction of hadronic taus is fakes from electrons and muons. The presence of a charged track and some deposit of energy in the calorimeters is the perfect signature of a tau. Fortunately, further discrimination can be done to reject electrons/muons which pass the tau identification.

The muon discrimination relies on rejecting any tau that is matched to a reconstructed muon. For the $\tau_e\tau_h$ channel a loose muon discriminator is used that rejects taus with the leading charged hadron matched to any muon segment. For the $\tau_\mu\tau_h$ channel a tight muon discriminator is used that requires that the leading charged hadron is not matched to any muon chamber hit. For 1-prong taus the sum of the taus ECAL and HCAL energy is required to be at least 20% of the track energy.

The electron discrimination is more difficult than muon discrimination. For the $\tau_\mu\tau_h$ where the electron discrimination is less important a loose discriminator is used that rejects any tau where the leading charged hadron is loosely identified as an electron. For the $\tau_e\tau_h$ channel, where the Drell-Yan to di-electron background is considerable, a more aggressive approach is needed. For this a multivariate analysis (MVA) is performed. The MVA is trained using leading track, tau energy, and tau PF gamma energy variables.

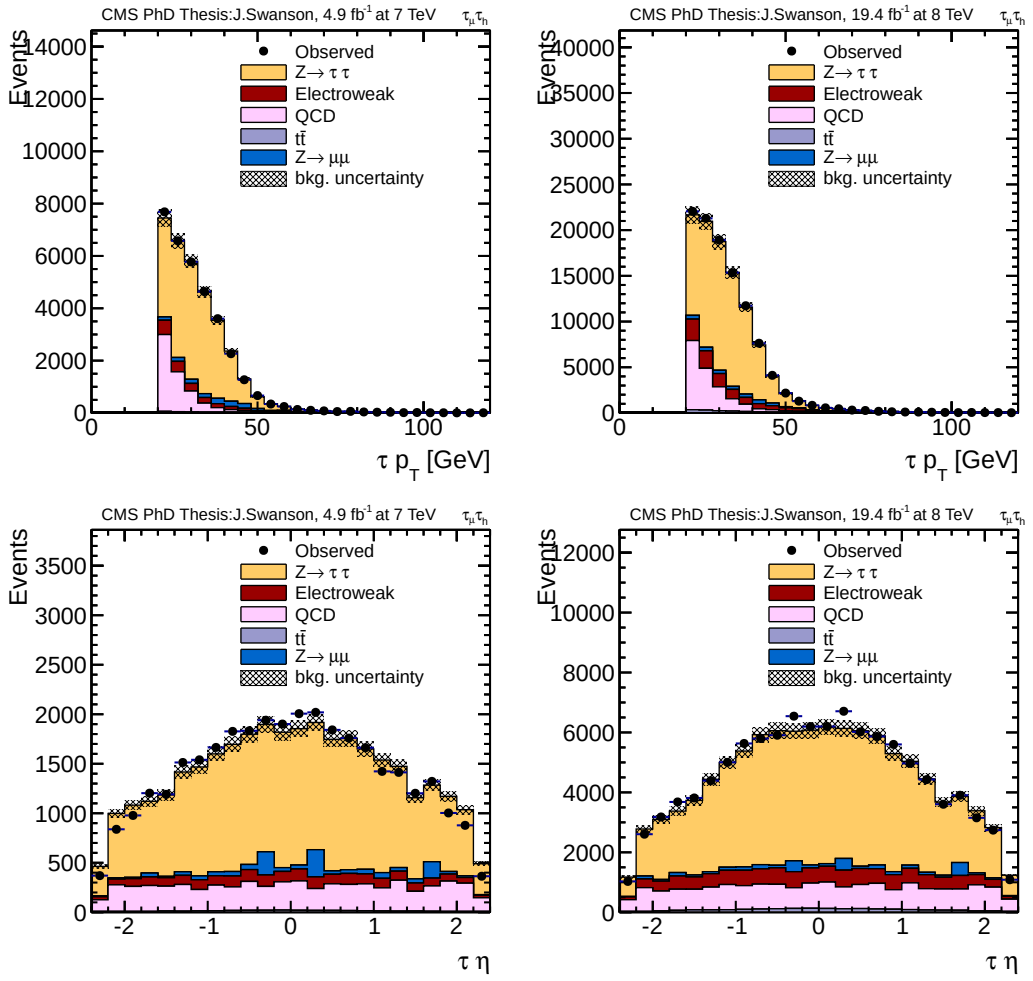

Figure 5.9: Distribution of the τ_h transverse momentum and psuedorapidity from the $\tau_\mu\tau_h$ channel. The background components are scaled relative to the results of a fit to the inclusive mass distribution. The uncertainty on the background is represented in the shaded region and comes from the fitted nuisances.

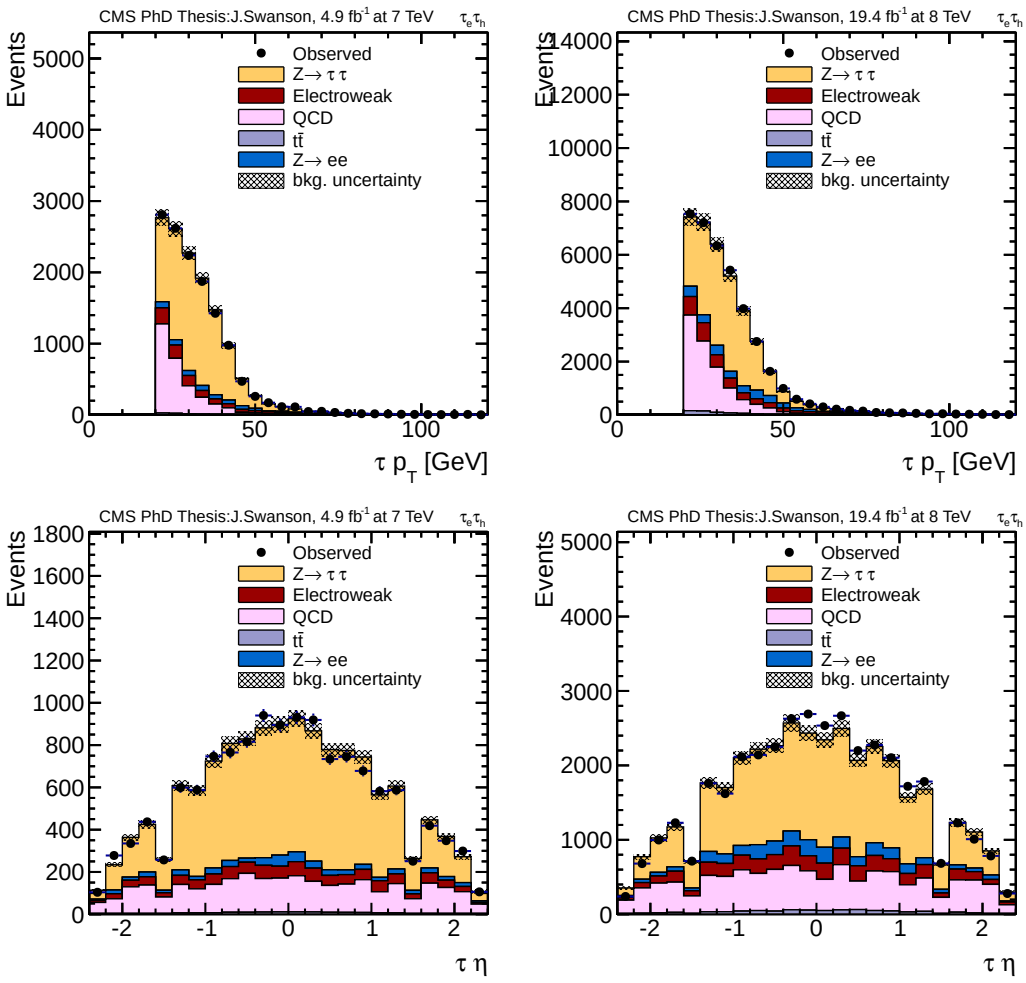

Figure 5.10: Distribution of the τ_h transverse momentum and psuedorapidity from the $\tau_e\tau_h$ channel. The background components are scaled relative to the results of a fit to the inclusive mass distribution. The uncertainty on the background is represented in the shaded region and comes from the fitted nuisances.

5.9 Trigger Selection

This analysis brings many challenges at the trigger level. On the electron/muon side the loss of a neutrino means that a large fraction of the energy is lost, leaving a soft spectrum for the particles selected for this analysis. On the hadronic tau side the large multi-jet background and relative complexity of tau algorithms present limits on how frequently the trigger algorithm can be run. To solve both of these problems a cross trigger object is used. An electron/muon is first selected by the L1 trigger system. The L1 seeds then are sent to HLT where complex electron/muon algorithms are used to reduce the rate. Finally for events with high quality electrons or muons a L3 tau algorithm is run.

At L1 the electrons are required to have an E_T ranging from 12 GeV to 20 GeV depending on the instantaneous luminosity. Additionally a loose isolation criteria is applied, which is relaxed with increasing energy to maintain a flat efficiency. At the HLT level a more detailed electron algorithm is applied with identification and isolation criteria similar to the offline electron reconstruction. Specific cuts on the electron track quality and the ratio of hadronic and electromagnetic energy are applied. Finally a tight isolation with a ρ pileup correction is applied. The ρ pileup correction works by subtracting the average amount of neutral and γ energy, measured over the whole detector, from the isolation cones. A final HLT transverse momentum is required ranging from 15 GeV to 22 GeV depending on the instantaneous luminosity.

The muon trigger also starts with a L1 bit with p_T requirements ranging from 8 GeV to 17 GeV depending on luminosity. The HLT algorithm uses full reconstruction similar to what is done offline with track quality cuts and various requirements on the segment matches. Finally a tight isolation with a ρ energy correction is required. The HLT transverse momentum requirements range from 10 GeV to 18 GeV.

Finally for events with a high quality electron or muon the tau HLT algorithm is run. The τ trigger uses a light version of particle flow event description algorithm. The trigger is a so called "fixed cone" algorithm with a signal cone of $\Delta R = 0.18$ and an isolation annulus of $0.18 < \Delta R < 0.4$. The trigger uses tracker only isolation built from a vertex compatible with the primary vertex from the reconstructed electron or muon to minimize pileup dependence. A τ is considered isolated if no tracks with transverse momentum greater than 1.5 GeV are found in the the isolation annulus.

5.10 Efficiencies

A proper measurement of the efficiencies for triggers and objects used in an analysis are essential. In this analysis all efficiencies are measured in data and MC The MC is then corrected to match the efficiency in the data. Most efficiencies are measured using Drell-Yan events with a *Tag & Probe* technique [56]. Events are scanned looking for a good tag lepton and then probed for a second lepton, the two leptons should have an invariant mass consistent with a Drell-Yan event. Usually a fit to the mass distribution is performed to subtract contributions from fake backgrounds.

Efficiencies for electrons/muons are measured using *Tag & Probe* with Drell-Yan to $ee/\mu\mu$ events. A fit is performed to the mass distribution. The MC Z shape is used for the signal and an exponential is used for the background. The efficiencies are then corrected in bins of p_T and η. Measured efficiencies in the data and MC can be found in Table 5.3.

Scale Factors for muon in the $\mu\tau_h$-channel

	Id			Isolation		
	$\lvert\eta\rvert < 0.8$	$0.8 < \lvert\eta\rvert < 1.2$	$1.2 < \lvert\eta\rvert < 2.1$	$\lvert\eta\rvert < 0.8$	$0.8 < \lvert\eta\rvert < 1.2$	$1.2 < \lvert\eta\rvert < 2.1$
$20 < p_T \leq 30$	0.9853 ± 0.0006	0.9818 ± 0.0009	0.9899 ± 0.0006	0.9685 ± 0.0012	0.9808 ± 0.0020	0.9972 ± 0.0012
$30 < p_T$	0.9857 ± 0.0001	0.9805 ± 0.0001	0.9900 ± 0.0001	0.9872 ± 0.0003	0.9924 ± 0.0002	1.0012 ± 0.0005

Scale Factors for electron in the $e\tau_h$-channel

	Id		Isolation	
	$\lvert\eta\rvert < 1.479$	$1.479 < \lvert\eta\rvert < 2.1$	$\lvert\eta\rvert < 1.479$	$1.479 < \lvert\eta\rvert < 2.1$
$24 < p_T \leq 30$	0.9100 ± 0.0020	0.8244 ± 0.0039	0.9468 ± 0.0018	0.9586 ± 0.0002
$30 < p_T$	0.9493 ± 0.0002	0.9260 ± 0.0001	0.9820 ± 0.0001	0.9948 ± 0.0001

Table 5.3: Correction factors for the identification and isolation efficiencies for the electron and muon in the $\mu\tau_h$-channel and $e\tau_h$-channel.

The efficiencies for the electron and muon legs of the triggers are also measured using the same *Tag & Probe* technique but then are fit with an error function of a crystal ball and corrected as a function of p_T in different η bins. There are a total of 5 parameters in the fit: m_0, σ, α, n, norm. m_0 is the mean value of the Crystal Ball function and defines the 50% value for the efficiency turn-on. σ is the resolution of the Crystal Ball function and defines the slope of the turn-on curve. α is the gaussian tail from the Crystal Ball function and defines the concavity at the turning points. n is the normalization of the crystal ball function. Lastly the norm is the efficiency on the plateau. The fit results can be found in Table 5.4.

Trigger	m_0	σ	α	n	norm
IsoMu17/18 run 2012ABCD 0< η <0.8	15.9802	0.0548775	0.020313	1.79791	0.968398
IsoMu17/18 run 2012ABCD -0.8< η <0	15.9828	0.0412999	0.0177441	1.66934	0.970097
IsoMu17 MC 0< η <0.8	15.9289	0.0271317	0.00448573	1.92101	0.978625
IsoMu17 MC -0.8< η <0	15.9556	0.0236127	0.00589832	1.75409	0.981338
IsoMu17/18 run 2012ABCD 0.8< η <1.2	16.8396	0.458636	0.633185	1.5706	0.8848
IsoMu17/18 run 2012ABCD -1.2< η <-0.8	17.3283	0.707103	1.2047	1.3732	0.900519
IsoMu17 MC 0.8< η <1.2	16.5678	0.328333	0.354533	1.67085	0.916992
IsoMu17 MC -1.2< *eta* <-0.8	17.3135	0.747636	1.21803	1.40611	0.934983
IsoMu17/18 run 2012ABCD endcap plus (η >1.2)	15.9825	7.90724e-05	5.49275e-08	1.6403	0.858285
IsoMu17/18 run 2012ABCD endcap minus (η <-1.2)	15.9987	8.94398e-05	5.18549e-08	1.8342	0.854625
IsoMu17 MC plus (η >1.2)	16.0051	2.45144e-05	4.3335e-09	1.66134	0.87045
IsoMu17 MC minus (η <-1.2)	15.997	7.90069e-05	4.40036e-08	1.66272	0.884502

EleX EB run	22.9041	1.04728	1.38544	1.22576	1.13019
Ele20 EB MC	21.7243	0.619015	0.739301	1.34903	1.02594
EleX EE	21.9941	1.43419	1.01152	2.28622	0.939872
Ele20 EE MC	22.1217	1.34054	1.8885	1.01855	4.7241

Table 5.4: Fit parameters for the electron, muon leg of the trigger in the $\mu\tau_h$-channel and the $e\tau_h$-channel.

The efficiency of the hadronic tau leg of the trigger is also measured with a *Tag & Probe* technique. The measurement is done using Drell-Yan to $\tau\tau$ events with the probe being either an electron or a muon. Additionally, no fit is performed to isolate the Drell-Yan events. Instead the fake background is estimated separately and in the MC measurement mixed with the real $\tau\tau$ events in the proper proportion. As is done for the electron and muon trigger legs, a fit is performed with an error function of a crystal ball and corrected as a function of p_T in different η bins.

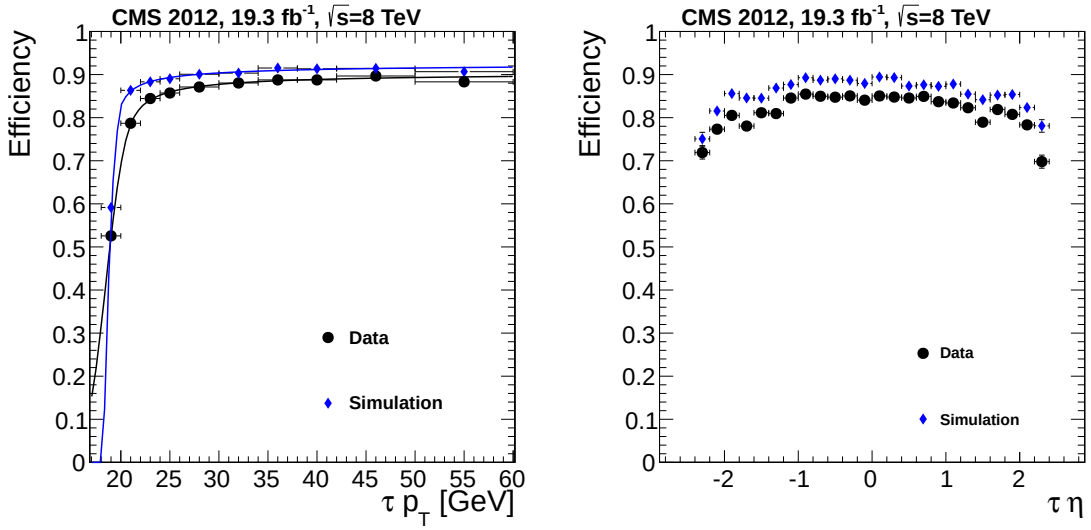

Figure 5.11: Efficiency of the L3 τ algorithm from the $\mu\tau$ cross object trigger plotted as a function of transverse momentum (left) and psuedorapidity (right).

Trigger	m_0	σ	α	n	norm
Loose Tau 20 Barrel run ABCD	18.52036251	1.47760312	2.53574445	1.71202550	0.93019930
Loose Tau 20 Barrel 53X MC	18.88740627	0.10718873	0.12277723	1.60581265	0.95041892
Loose Tau 20 Endcap run ABCD	18.41225333	0.76598912	0.60544260	5.38350881	0.85870108
Loose Tau 20 Endcap 52X MC	18.30439676	1.44360240	3.79358997	1.07560564	0.93103925
Loose Tau 20 Barrel run ABCD	18.686211	1.993524	3.202713	3.612693	0.871640
Loose Tau 20 Barrel 53X MC	18.431118	1.572877	3.301699	4.760769	0.899620
Loose Tau 20 Endcap run ABCD	18.472954	1.606388	3.468975	55.629620	0.828977
Loose Tau 20 Endcap 53X MC	18.257217	1.632443	9.283116	40.219585	0.858643

Table 5.5: Fit parameters for 2012 data of the τ_h trigger efficiencies in the $\mu\tau_h$- and the $e\tau_h$-channel. The first set of tau triggers correspond to the $\mu\tau_h$-channel, the second set correspond to the $e\tau_h$-channel.

Chapter 6

Analysis Method

The reconstruction of various objects used in this analysis, as discussed in the previous chapter, is only part of the story. The objects have to be put together in order to be used for analysis. In this chapter I will discuss the methodology used for the analysis. This includes selecting events, categorizing events to increase signal significance, reconstructing the full τ pair invariant mass, estimating the background components, systematic uncertainties and finally performing a fit to search for the presence of signal.

6.1 Event Selection

The event selection starts with selecting an appropriate di-τ candidate. This is done by selecting a $\tau_\mu\tau_h$ or $\tau_e\tau_h$ pair passing the kinematic requirements and the identification and isolation requirements discussed in Chapter 5. The individual object kinematic requirements are summarized in Table 6.1. The selected di-τ pair is required to be separated in ΔR by more than 0.5 to remove objects that directly overlap or are reconstructed by both reconstruction algorithms. In the case that more than

one di-τ candidate is found, the candidate with the highest linear sum p_T is selected.

Lepton	Year	$> p_T[GeV]$	$< \|\eta\|$
τ_μ	2011	17	2.1
τ_μ	2012	20	2.1
τ_e	2011	20	2.1
τ_e	2012	24	2.1
τ_h	2011	20	2.3
τ_h	2012	20	2.3

Table 6.1: Minimum kinematic thresholds for individual leptons. The differences between 2011 and 2012 are due to trigger threshold differences.

An additional selection is done on the transverse mass (M_T), shown in Figure 6.1, of the τ_μ or τ_e and the missing transverse momentum. In real di-τ events the neutrino is produced nearly collinear with the visible part of the τ, because the τ energy is much larger then its mass. This results in real di-τ events having low M_T. The backgrounds, specifically W+Jets, are generally produced with the lepton and the neutrino back to back in $\Delta\phi$ resulting in a distribution that peaks at the $W^\pm$ mass. Events selected for this analysis are required to have $M_T < 20$ GeV.

Finally, in addition to the event selection above there is an additional cut to reject events with an additional lepton. This is done to ensure that various decay channels do not overlap in the final $\tau\tau$ combination. In this analysis any event with an additional muon or electron passing the identification criteria discussed in Chapter 5 will be rejected.

$$M_T = \sqrt{2 p_T E_T^{miss} (1 - cos(\Delta\phi))} \tag{6.1}$$

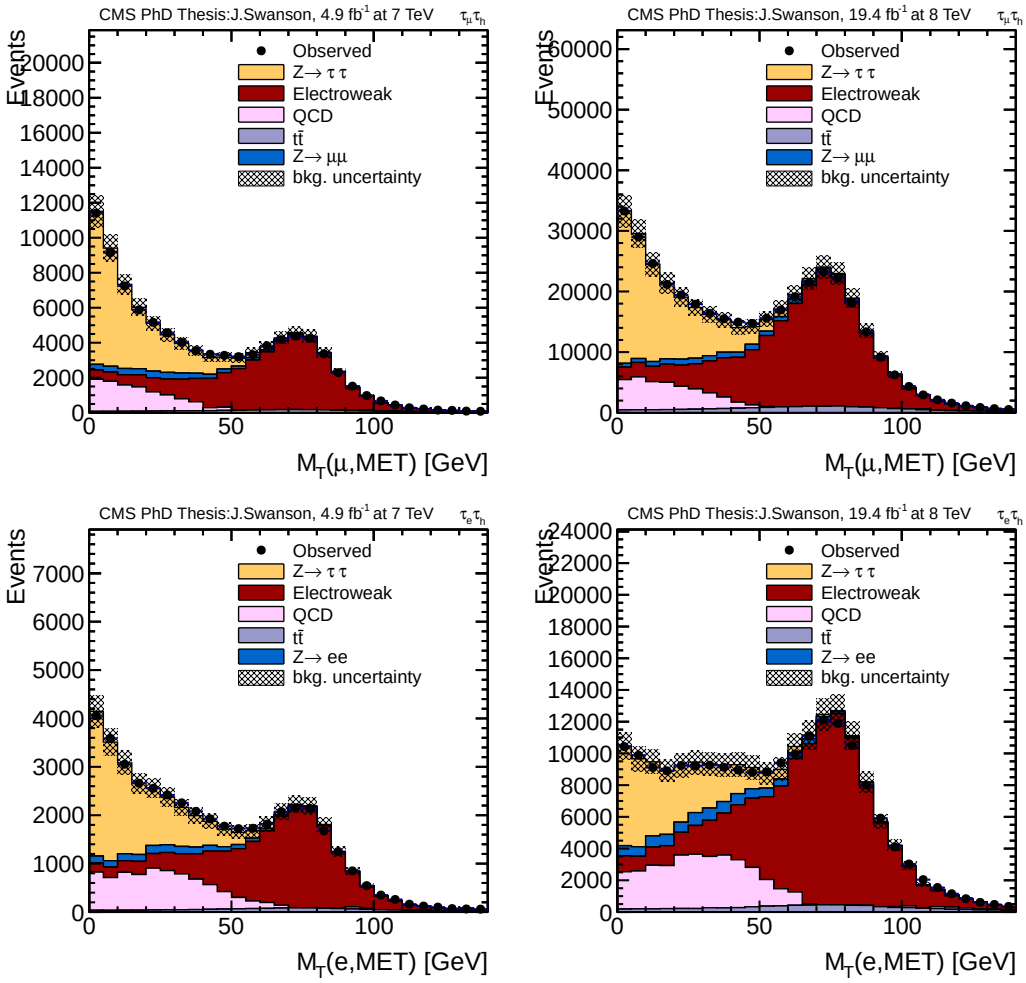

Figure 6.1: Distribution of the transverse mass from the $\tau_\mu\tau_h$ (top) and $\tau_e\tau_h$ (bottom) channels. The plots on the left and right correspond to the 7 TeV and 8 TeV periods respectively. The background components (except $W^{\pm}$) are scaled relative to the results of a fit to the inclusive mass distribution. The uncertainty on the background is represented in the shaded region and comes from the fitted nuisances.

6.2 Category Selection

Selected events are categorized to increase the significance in the Higgs Boson production modes of GGF and VBF. Details of the Higgs Boson production modes were discussed in Chapter 2. The categorization starts with looking for the number of jets in the event. As discussed in Chapter 5 jets selected in this analysis have a $p_T >$ 30 GeV and a $|\eta| < 4.7$. VBF categories have the highest significance and always require the presence of at least two jets. Next are categories with 1 or more jets that don't pass the VBF category selections. These categories are dominated by GGF events. The initial state gluons are more likely to radiate jets than the initial state quarks present in the dominant Drell-Yan to $\tau\tau$ production mode. This increases the signal significance for events with 1 or more jets. All categories have the additional requirement of no b-tagged jets. A b-tagged jet refers to the identification of jets as containing b quarks, the identification criteria are discussed in Chapter 5. Real Higgs events produced through VBG or ggF rarely include real b quark jets. The categories have been selected by optimizing the cuts for the expected significance for a 125 GeV standard model Higgs boson.

The signature of VBF Higgs Boson production is two jets from the initial state quarks in the event. These jets are generally at high η, at opposite ends of the detector. These events can be selected by requiring a large separation in eta between the jets and a large jet-jet invariant mass. Additionally the two forward jets cause the Higgs to come out largely boosted in the transverse direction. For a Higgs decaying to a pair of τ leptons this boost means a large missing transverse momentum due to the loss of the neutrinos from the τ decays.

To select these VBF events, in the 8 TeV analysis, two categories are defined; VBF loose and VBF tight. The VBF tight category requires at least two jets with

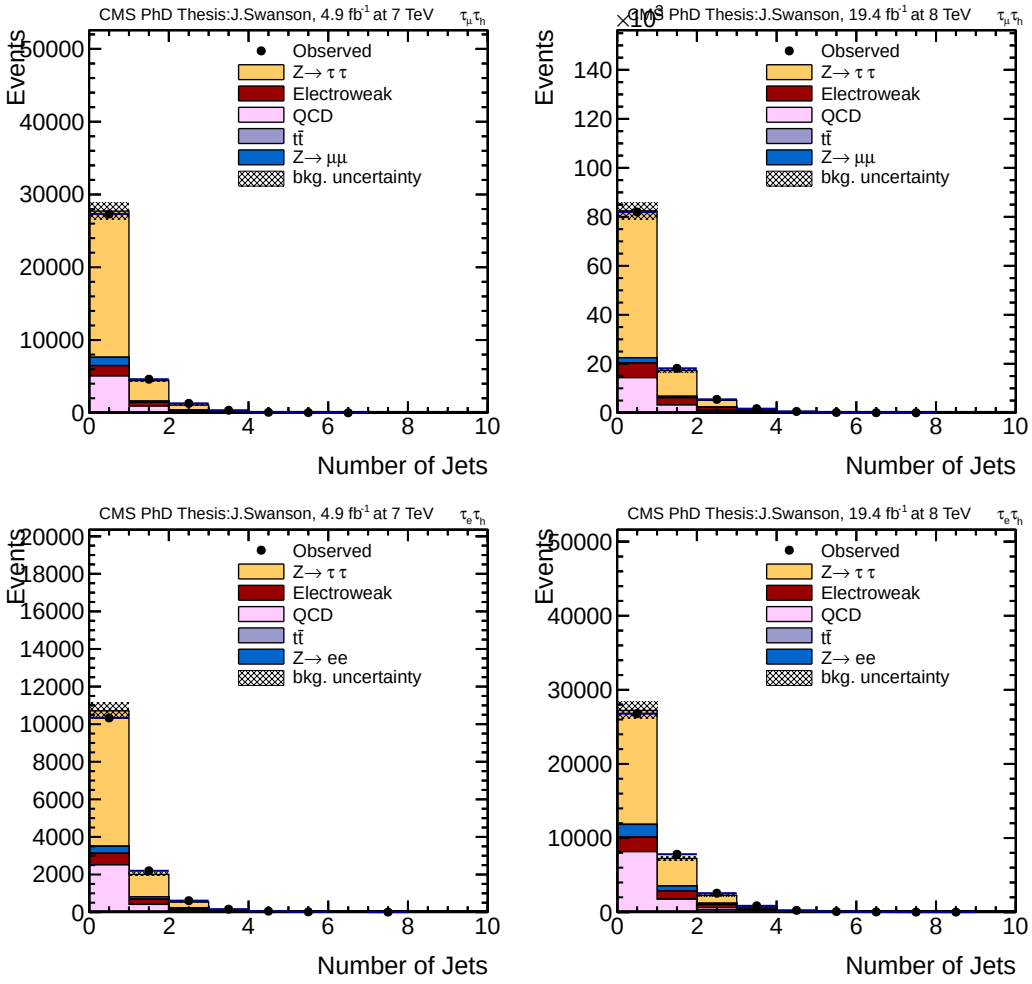

Figure 6.2: Distribution of the number of identified jets in the sample of selected events from $\tau_\mu\tau_h$ (top) and $\tau_e\tau_h$ (bottom) channels. The background components are scaled relative to the results of a fit to the inclusive mass distribution. The uncertainty on the background is represented in the shaded region and comes from the fitted nuisances.

the jets having a separation in $\Delta\eta$ larger than 3.5, and invariant mass of the di-jet system greater than 800 GeV, a rapidity gap veto and a missing transverse energy of 50 GeV for $\tau_\mu\tau_h$ and 30 GeV for $\tau_e\tau_h$. A rapidity gap veto helps reject background with additional jets in the event. It is defined by requiring that no identified jet is found in the η separation between the two highest p_T jets in the event. The VBF loose category also requires at least two jets having a separation in $\Delta\eta$ larger than 3.5, and invariant mass of the di-jet system greater than 500 GeV, a rapidity gap veto and should not be selected by the VBF tight category.

In the 7 TeV analysis where the integrated luminosity is approximately a factor of 4 less then in 8 TeV only one VBF category is defined. This category is just a combination of events selected by either of the two 8 TeV categories and selects events with at least two jets having a separation in $\Delta\eta$ larger than 3.5, an invariant mass of the di-jet system greater than 500 GeV and a rapidity gap veto.

Finally for GGF two additional categories are defined. These categories require at least 1 identified jet and should not be selected by either of the VBF categories. The split of the two categories is defined by the p_T of the τ_h candidate in the event. The 1-jet high category is all events with τ_h $p_T > 40$ GeV and the 1-jet low category is all remaining events with τ_h $p_T < 40$ GeV. For the $\tau_e\tau_h$ channel an additional missing transverse energy cut of 30 GeV is applied to reject the significant Drell-Yan to ee background. A summary of all of the categories can be found in Table 6.2.

6.3 Mass Reconstruction

To reconstruct the mass of a di-τ system many options are available. The easiest approach is to use the visible invariant mass (m_{vis}) or the invariant mass of the visible portion to the two τ decays.

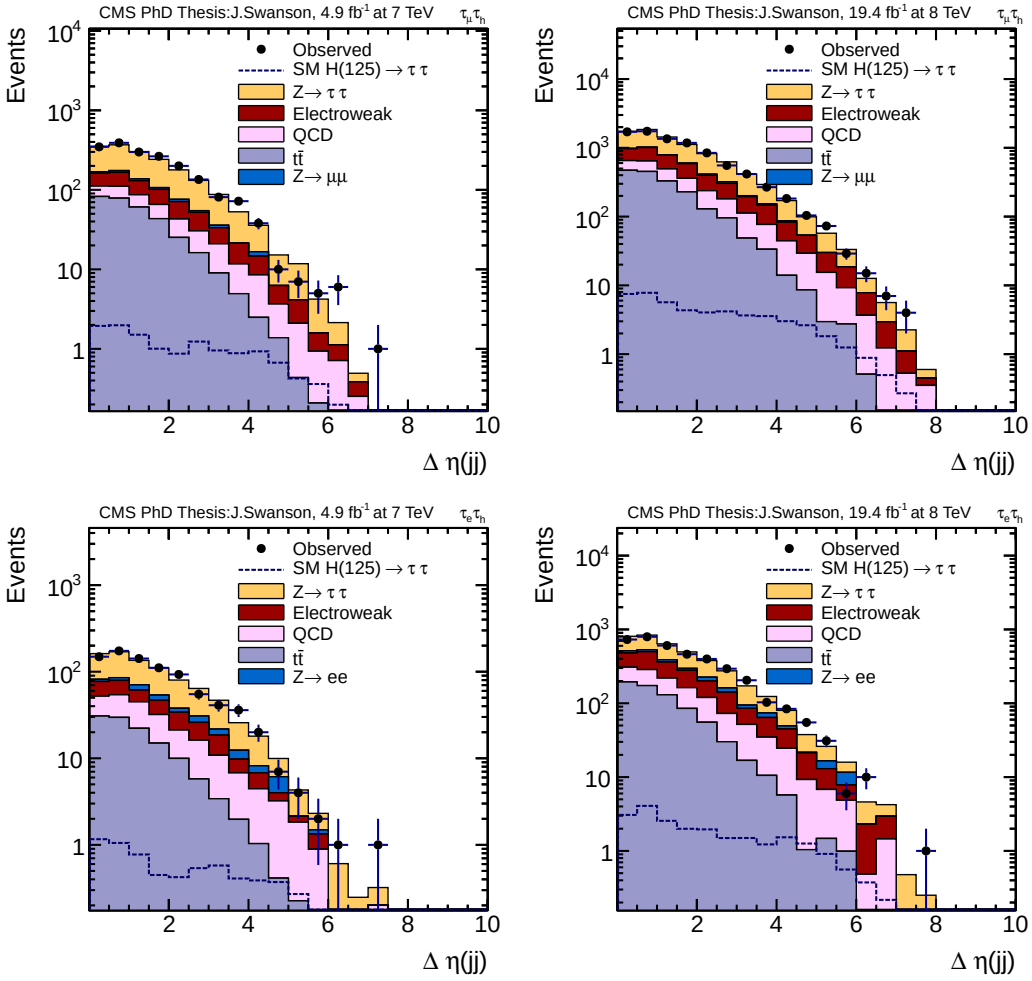

Figure 6.3: Distribution of the $\Delta\eta$ separation for di-jet events selected from $\tau_\mu\tau_h$ (top) and $\tau_e\tau_h$ (bottom) channels.

$$m_{vis} = \sqrt{2p_{T1}p_{T2}\left(cosh(\eta_1 - \eta_2) - cos(\phi_1 - \phi_2)\right)} \tag{6.2}$$

The disadvantage of m_{vis} is that a large fraction of the energy of the original τ is carried by the invisible neutrinos. This results in a loss of mass separation between Drell-Yan to $\tau\tau$ background and the Higgs Boson signal. Another traditional approach is the collinear approximation [57]. The disadvantage to the collinear approximation is that it results in an unphysical solution approximately 20% of the

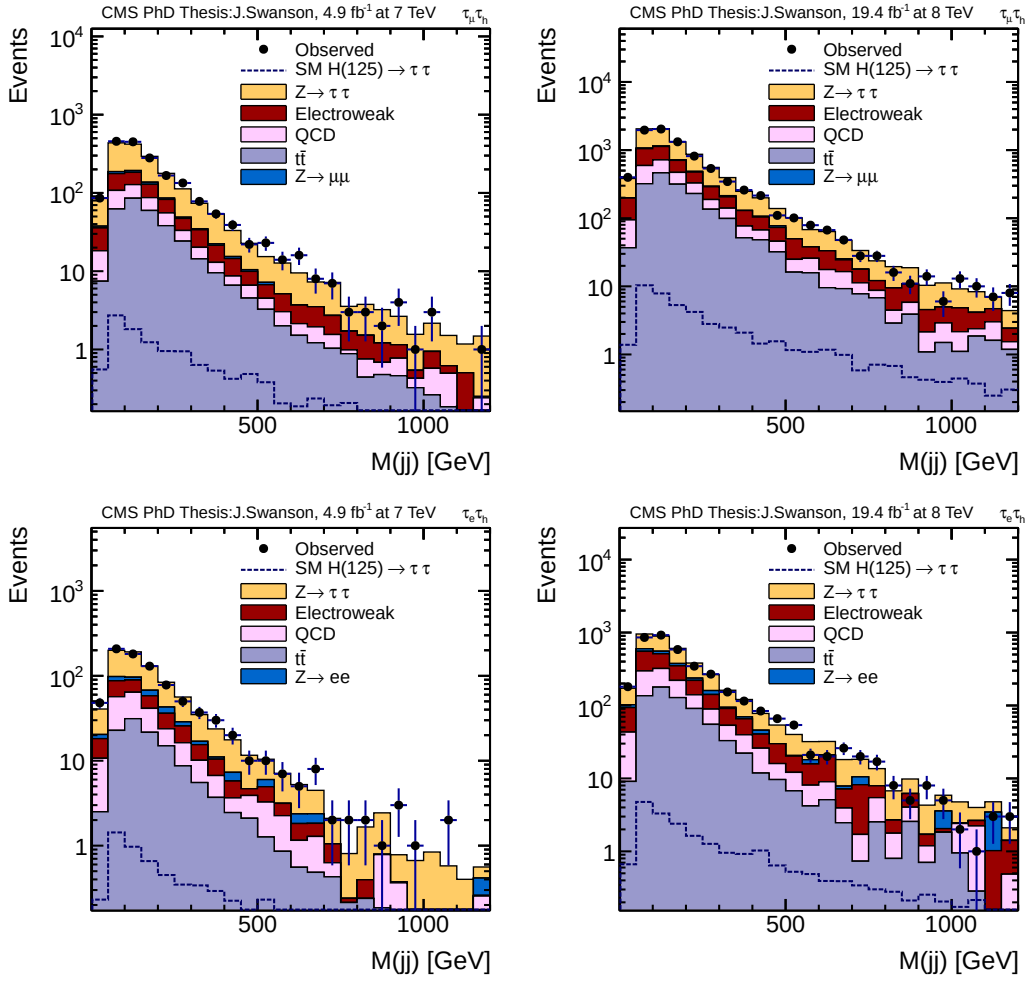

Figure 6.4: Distribution of the di-jet invariant mass for di-jet events selected from $\tau_\mu\tau_h$ (top) and $\tau_e\tau_h$ (bottom) channels.

time. In this analysis the SVFit algorithm [58] is used. SVFit is a event by event estimator of the true di-τ mass through a likelihood approach resulting in a physical solution for every event. The rest of this section is devoted to a brief explanation of the SVFit algorithm, it is a summary of the full description available in [58].

To completely classify full τ kinematics a total of 6 (7) parameters are needed for τ_h (τ_l) decays: the polar and azimuthal angles of the visible decay produced in the rest frame of the τ, the three boost parameters bringing the τ from the rest frame to

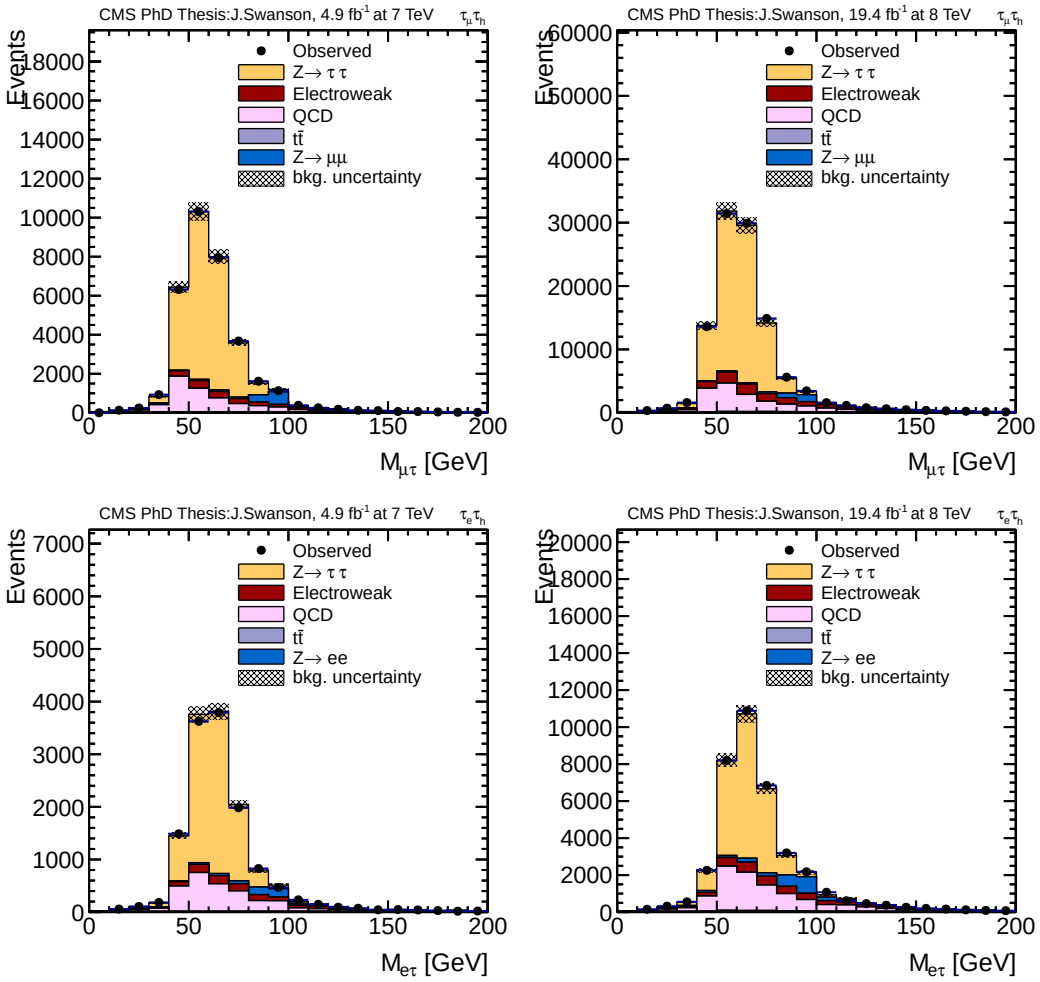

Figure 6.5: Distribution of the Visible Mass distribution for inclusive selection, before event categorization, from $\tau_\mu\tau_h$ (top) and $\tau_e\tau_h$ (bottom) channels. The background components are scaled relative to the results of a fit to the inclusive mass distribution. The uncertainty on the background is represented in the shaded region and comes from the fitted nuisances.

Category	Chn	Selection
VBF tight	$\tau_\mu\tau_h$	njets≥2, M_{jj}>800 GeV, $\Delta\eta_{jj}$>3.5, Rapidity Gap Veto, MET>50 GeV
	$\tau_e\tau_h$	M_{jj}>800 GeV, njets≥2, $\Delta\eta_{jj}$>3.5, Rapidity Gap Veto, MET>30 GeV
VBF loose	$\tau_\mu\tau_h$	njets≥2, M_{jj}>500 GeV, $\Delta\eta_{jj}$>3.5, Rapidity Gap Veto, !VBF tight
	$\tau_e\tau_h$	njets≥2, M_{jj}>500 GeV, $\Delta\eta_{jj}$>3.5, Rapidity Gap Veto, !VBF tight
VBF 7 TeV	$\tau_\mu\tau_h$	njets≥2, M_{jj}>500 GeV, $\Delta\eta_{jj}$>3.5, Rapidity Gap Veto
	$\tau_e\tau_h$	njets≥2, M_{jj}>500 GeV, $\Delta\eta_{jj}$>3.5, Rapidity Gap Veto
1-Jet High	$\tau_\mu\tau_h$	njets ≥ 1, No B-Tagged jets, τ_h p_T > 40 GeV, !VBF
	$\tau_e\tau_h$	njets ≥ 1, τ_h p_T > 40 GeV, MET > 20 GeV, !VBF
1-Jet Low	$\tau_\mu\tau_h$	njets ≥ 1, No B-Tagged jets, τ_h p_T < 40 GeV, !VBF
	$\tau_e\tau_h$	njets ≥ 1, τ_h p_T < 40 GeV, MET > 20 GeV, !VBF

Table 6.2: Summary of the criteria for the analysis categories.

the laboratory frame, m_{vis}, and in the case of τ_l decays the invariant mass of the two-neutrino system. Four parameters are constrained by the four vector of the visible τ decay products as measured in the laboratory frame. This leaves a total of 2 or 3 parameters unconstrained depending on the τ decay mode. SVFit chooses the three parameters to be left unconstrained:

- x - the fraction of the τ energy in the lab frame of the visible decay products.
- ϕ - the azimuthal angle of the τ direction in the lab frame.
- $m_{\nu\nu}$ - the invariant mass of the two-neutrino system. For τ_h decays $m_{\nu\nu} = 0$.

The two components of the missing transverse energy provide an additional constraint but have a poor experimental resolution. A maximum likelihood fit is performed using the available observables along with the missing transverse momentum resolution. The likelihood model calculates the probability

$$P(M^i\tau\tau) = \int \delta\left(m^i_{\tau\tau} - m_{\tau\tau}(\overrightarrow{y}, \overrightarrow{a_1}, \overrightarrow{a_2})\right) f(\overrightarrow{z}|\overrightarrow{y}, \overrightarrow{a_1}, \overrightarrow{a_2}) d\overrightarrow{a_1}\overrightarrow{a_2} \tag{6.3}$$

where $\overrightarrow{y}$ are the measured visible decay products, $\overrightarrow{a_1}$,$\overrightarrow{a_2}$ are the full τ four momenta and f is the likelihood function. The likelihood function f is different for τ_h and τ_l decays. For τ_l decays the full matrix element of the decay is used as described in [59] and for the τ_h decay is based on a two-body phase-space model described in [15].

The SVFit algorithm provides a mass that peaks near the true mass of the di-τ system and a resolution of about 20%, which can be observed in Figure 6.6.

6.4 Backgrounds

The main challenge with a Higgs Boson to $\tau\tau$ analysis is dealing with several large background contributions. The challenge in dealing with all these backgrounds is multifaceted. They should be estimated where possible with data driven techniques, shapes must be derived for use in the maximum likelihood fit, and the estimation of systematic uncertainties on those yields and shapes.

The largest contribution comes from the Drell-Yan to $\tau\tau$ background. This background is reduced after category selection but is always the dominant background. The best way to distinguish the Drell-Yan from the Higgs signal is due to its different mass compared to that of the Higgs boson. Additional backgrounds come from jets in an event faking a τ_h candidate. These backgrounds include W+Jets and QCD multijet backgrounds. Another background contribution comes from Drell-Yan to ee or $\mu\mu$ background where a light lepton fakes a τ_h candidate. Finally there are some other small backgrounds coming from $t\bar{t}$, di-bosons, and single top quark production.

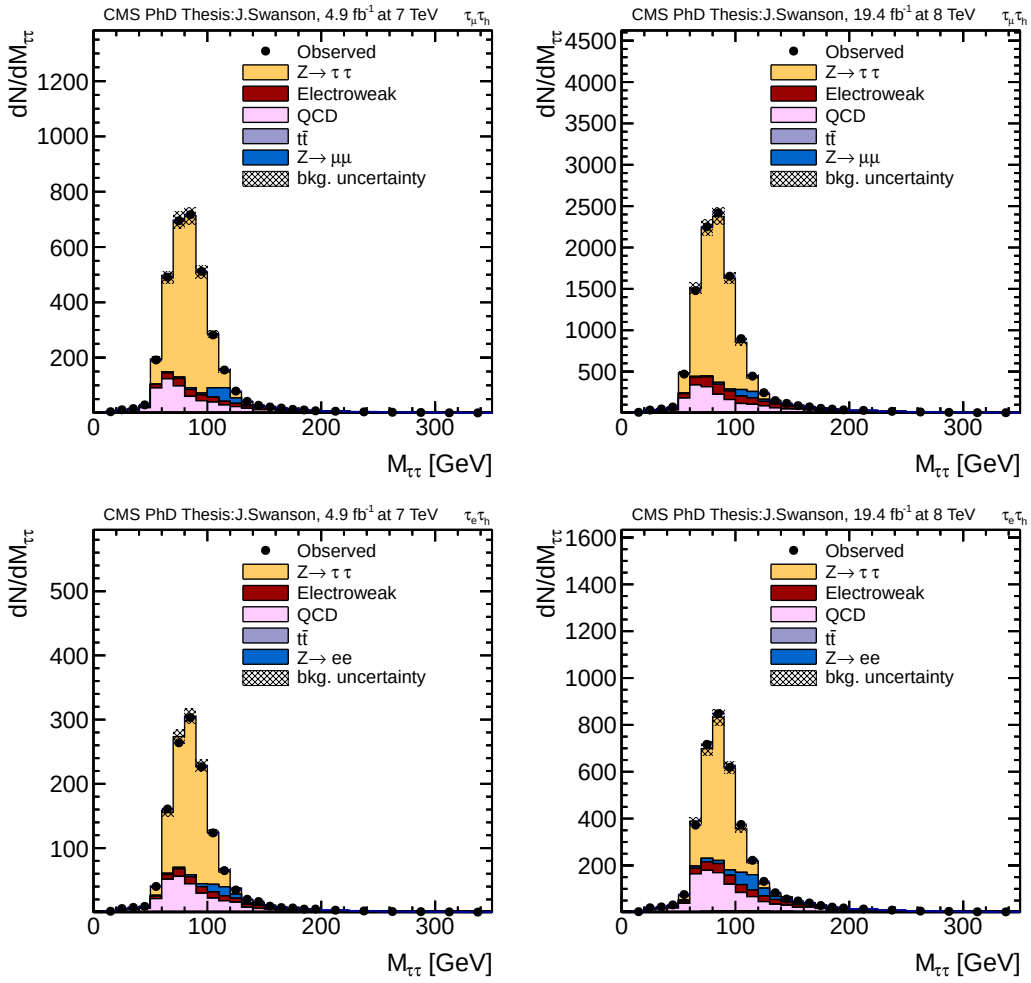

Figure 6.6: Distribution of the SVFit Mass distribution for inclusive selection, before event categorization, from $\tau_\mu\tau_h$ (top) and $\tau_e\tau_h$ (bottom) channels. The background components are scaled relative to the results of a fit. The uncertainty on the background is represented in the shaded region and comes from the fitted nuisances.

Drell-Yan to $\tau\tau$

The presence of two real τ leptons in the decay of a Drell-Yan event provides a real challenge to this analysis. The requirements on additional jets, τ_h p_T, and VBF cuts help reduce these events but the real discrimination comes from differing SVFit mass shapes.

The Drell-Yan to $\tau\tau$ background is handled by producing a Drell-Yan to $\tau\tau$ embedded sample. The embedding technique starts by selecting events in data with a high purity to be Drell-Yan to $\mu\mu$. The muons in the events are removed and replaced with MC τ candidates. The τ candidates are then allowed to decay as usual using Tauola [40]. The events are required to pass minimum kinematic thresholds for p_T and η. Events that don't pass are re-decayed, and the final distribution is reweighed by the probability for a given kinematic result. The real advantage to this method is it provides large statistics for this important background. Also since the event comes from the original Z$\rightarrow$ $\mu\mu$ data event the jets, missing transverse energy and the pileup are more accurately modeled.

The yield of the Drell-Yan to $\tau\tau$ is then estimated using a combination of the embedded Drell-Yan to $\tau\tau$ and the MC. The yield is measured before the category selection (inclusive selection), discussed earlier in this chapter, in the Drell-Yan to $\tau\tau$ MC then corrected for effects due to the kinematic cuts and efficiencies. The embedded sample is then used to calculate the efficiency to pass the category selection for a given category from the inclusive selection. This efficiency is then applied to the MC inclusive estimate to give the expected background contribution in each category. The shape for the Drell-Yan to $\tau\tau$ comes from the embedded sample after applying the full category selection in any given category.

W+Jets

The second most significant background in this analysis comes from W+Jets. It comes about when the W decays to a real electron or muon and an additional jet fakes a τ_h candidate. This background is reduced by the M_T and τ_h isolation but due to its large production cross section some events still pass the final selection.

To estimate the W+Jets background a high M_T sideband, defined as events with $M_T > 70$ GeV are used, where the M_T comes from the electron or muon and the missing transverse energy. In this region the W+Jets is the dominant contribution, this can be seen in Figure 6.1. The W+Jets MC simulation is then used to estimate the yield in the signal region, defined with events of $M_T < 20$ GeV. The expected ratio of W+Jets events between the high M_T sideband and the signal region is calculated. This ratio is applied to the data observed in the high M_T sideband after subtracting the MC expectation for other backgrounds in this region.

The shape for the W+Jets background is taken from the MC simulation. Samples produced with different jet multiplicities are merged together in the appropriate ratios to increase statistics for events with high jet multiplicity. For the VBF categories a relaxed selection is used to increase the shape template statistics. For VBF loose the relaxed selection requires $M_{jj} > 200$ GeV, $\Delta\eta_{jj} > 2.0$, and the central jet veto. For the VBF tight category the same relaxed selection is used with the addition of the MET > 50 (30) GeV for the $\tau_\mu\tau_h$ ($\tau_e\tau_h$) applied in this category.

Multijet

Multijet production at the LHC is large and is a threat for almost all analyses. Even though the probability of a multijet event passing this analysis selection is extremely low its production is so large that it needs to be carefully evaluated. Multijet events

can fake $\tau_\mu\tau_h$ events by having a real muon from a heavy flavor quark decay and a jet faking the τ_h candidate. For $\tau_e\tau_h$ the fake rate is higher due to charged pions and photons faking the electron in addition to the same heavy flavor decay.

Production of a sufficient amount of MC simulation of multijet background is computing intensive so a completely data driven approach is used. The method relies on the fact that in multijet events the amount of events produced with the two legs having the same charge (SS) and opposite charge (OS) are expected to about the same. The events in the data requiring the SS selection are used after subtracting the contribution from the W+Jets, estimated as described above, and the other backgrounds as predicted from the MC simulation. To extrapolate into the OS signal selection a OS/SS scale factor is calculated with the electron or muon isolation inverted ($l_{iso} > 0.2$). The scale factor is measured to be 1.06 for both the $\tau_\mu\tau_h$ and $\tau_e\tau_h$ channels.

In the VBF categories a slightly different method is needed due to the limited statistics in the SS region after applying the VBF selection. Instead an efficiency is measured with SS and inverted lepton isolation to go from the inclusive selection to the VBF selection. This efficiency is then applied to the multijet background estimated before with the inclusive selection using the OS/SS method.

For the multijet shape the SS data events are also used. For the 1-jet categories the full selection is applied except with SS and inverted lepton isolation. Shapes for other backgrounds in this region are small and are subtracted. In the VBF categories the shapes come from the SS inverted lepton isolation region but with a relaxed selection. For VBF loose the relaxed selection requires $M_{jj} > 200$ GeV, $\Delta\eta_{jj} > 2.0$. For the VBF tight category the same relaxed selection is used with the addition of the MET > 50 (30) GeV for the $\tau_\mu\tau_h$ ($\tau_e\tau_h$) as applied in this category.

Drell-Yan to ee/$\mu\mu$

Drell-Yan to electrons or muons brings two distinct background sources to this analysis. The first involves an electron or muon faking a τ_h candidate. The second comes when an additional jet in the event fakes the τ_h candidate. Since these backgrounds have very different shapes and yields they are treated separately.

The Drell-Yan to a lepton faking the τ is estimated using MC simulation, requiring that the τ_h candidate is matched to a generated lepton. The yield for the MC is corrected by decay mode looking in a control region, defined as events with, $85 < m_{vis} < 95$ GeV. The shapes also come from the MC simulation with relaxed cuts for the VBF selection as is done for the W and Multijet background. Additionally in the $\tau_e\tau_h$ channel the shape is relaxed by not applying the additional lepton vetoes.

The Drell-Yan to jet faking the τ background is also estimated using MC simulation but requiring that the τ_h candidate is not matched to a generated lepton. The yield and the shape comes from the MC simulation with the same relaxed selection for the VBF category shapes.

$t\bar{t}$, di-bosons and other backgrounds

Other backgrounds in this analysis include $t\bar{t}$, di-boson production and single top quark production in association with a $W^{\pm}$ Boson. These backgrounds include a combination of real $\tau\tau$ production and leptons/jets faking a τ_h. These backgrounds are generally very small and reduced significantly by b-jet vetoes, extra lepton vetoes and isolation/identification requirements.

The yields from these backgrounds is estimated from MC simulation normalized to CMS measured cross sections, where available. When no CMS measured cross section is available the NLO cross sections are used as calculated from MC@NLO [41]. The

full set of cross sections used is presented in Table 4.1. The shapes for these smaller background contributions come from the MC simulation with the common relaxed selection for the VBF categories.

6.5 Systematic Uncertainties

As with any analysis careful consideration of the systematic uncertainties is important. The dominant uncertainties in this analysis come from experimental factors such as object efficiencies and energy scale. Additional uncertainties come from the uncertainties on the background estimates, statistics on background templates and shape uncertainties due to differences in the MC simulation or due to relaxed selections used for the shapes for some backgrounds. Finally theoretical uncertainties come into play on the signal contribution.

The theoretical uncertainties on the object efficiencies and energy scale come from many sources. The measurement of the efficiencies for the electrons and muons, done by the tag and probe technique as discussed in Chapter 5, have an uncertainty of 2% including both identification, isolation and trigger. The efficiency of the τ_h candidate is also measured using a tag and probe and has an uncertainty of 7% [55], including the trigger efficiency systematic. An energy scale systematic is applied to the τ_h candidate of 3% as a shape uncertainty. The shape of the Drell-Yan to $\tau\tau$ and the signal contributions is varied by allowing the energy of the τ_h candidate to vary up and down by the 3% τ_h energy scale uncertainty. The τ_h energy scale shape uncertainty covers any energy scale uncertainty on the electrons and the muons which are expected to be small. Lastly scale uncertainties on the MET and jets are applied as normalization uncertainties. The uncertainties related to object identification and energy scale are constrained by including 0-jet events split between high and low τ_h

p_T to be fit simultaneously with the search categories.

Important uncertainties on the backgrounds include those on Drell-Yan, Multijet and W+Jets backgrounds. The uncertainty on the Drell-Yan to $\tau\tau$ background comes from the uncertainty on the embedding technique and the statistics used to measure the category efficiencies as discussed earlier in this chapter. These uncertainties vary from 3% in 1-jet low to 20% for the VBF tight category. For the W+Jets the uncertainty comes from the simulation statistics in the signal region and the statistics in the data sideband. Additional W+Jet uncertainties come from the uncertainty on the MC simulation. The uncertainties range from 10% in 1-jet to 38% in VBF. For the multijet background, uncertainties come from the measurement of the OS/SS factor and statistics in the sidebands and range from 9% to 81%.

In addition to the normalization uncertainties shape uncertainties are considered for the W+Jets and the multijet backgrounds. These shape uncertainties come by allowing bin-by-bin fluctuations on the templates. The bins are allowed to fluctuate with the statistical uncertainty and a gaussian constraint. See for instance [60]. The Drell-Yan to $\tau\tau$ also incorporates bin-by-bin shape uncertainties in addition to the τ_h energy scale shape uncertainty in the VBF Tight category.

Finally we have theoretical uncertainties coming from the parton distribution functions, underlying event and the Higgs production modes, see for instance [42, 43]. The dominant theoretical uncertainty comes from the scale of gluon-gluon fusion production entering into the VBF categories and is measured to be 30% and 50% for loose and tight respectively [61]. A summary of all systematics can be found in Table 6.3.

Experimental Uncertainties		Propagation into Limit Calculation	
Uncertainty	Uncert.	*1-Jet*	*VBF*
Electron ID & Trigger (†*)	$\pm 2\%$	$\pm 2\%$	$\pm 2\%$
Muon ID & Trigger (†*)	$\pm 2\%$	$\pm 2\%$	$\pm 2\%$
Tau ID & Trigger (†*)	$\pm 7\%$	$\pm 7\%$	$\pm 7\%$
Tau Energy Scale (†*)	$\pm 3\%$	$\pm 3\%$	$\pm 3\%$
JES (Norm.) (†*)	$\pm 2.5-5\%$	$\pm 1-6\%$	$\pm 5-20\%$
MET (Norm.) (†*)	$\pm 5\%$	$\pm 2-7\%$	$\pm 5-8\%$
b-Tag Efficiency (†*)	$\pm 10\%$	$\mp 2-3\%$	$\mp 3\%$
Mis-Tagging (†*)	$\pm 30\%$	$\mp 2\%$	$\mp 2-3\%$
Norm. Z production (†*)	$\pm 3\%$	$\pm 3\%$	$\pm 3\%$
$Z \to \tau\tau$ Category	$\pm 3\%$	$\pm 3-5\%$	$\pm 10-20\%$
Norm. $t\bar{t}$ (†*)	$\pm 10\%$	$\pm 10\%$	$\pm 12-33\%$
Norm Diboson (†*)	$\pm 15-30\%$	$\pm 15-30\%$	$\pm 15-100\%$
Norm QCD Multijet	$\pm 6-32\%$	$\pm 9-30\%$	$\pm 12-81\%$
Luminosity 2011(2012)	$\pm 2.2(4.2)\%$	$\pm 2.2(4.2)\%$	$\pm 2.2(4.2)\%$
Norm. W+jets	$\pm 10-30\%$	$\pm 10-33\%$	$\pm 12.4\%-38\%$
Norm. Z: $l \to e$ fakes τ_h (†)	$\pm 20\%$	$\pm 36\%$	$\pm 22-69\%$
Norm. Z: $l \to \mu$ fakes τ_h (†)	$\pm 30\%$	$\pm 30\%$	$\pm 30\%$
Norm. Z: jet fakes τ_h	$\pm 20\%$	$\pm 20\%$	$\pm 40\%$

Theory Uncertainties (SM)		Propagation into Limit Calculation	
Uncertainty	Uncert.	*1-Jet*	*VBF*
PDF (†*)	-	$\pm 2-8\%$	$\pm 2-8\%$
$\mu_r/\mu_f(gg \to H)$ (†*)	-	$\pm 10\%$	$\pm 30-50\%$
$\mu_r/\mu_f(qq \to H)$ (†*)	-	$\pm 4\%$	$\pm 5\%$
$\mu_r/\mu_f(qq \to VH)$ (†*)	-	$\pm 4\%$	$\pm 4\%$
UE & PS (†*)	-	$\pm 4\%$	$\pm 4\%$

Table 6.3: Uncertainties for the limit calculation. The lower part of the table corresponds to the theoretical uncertainties used for the SM search. Note that in some cases '$\mp$' indicates that the uncertainty is anti-correlated with respect to other event categories. All uncertainties indicated with (*) are correlated among the separate channels. All uncertainties indicated with a (†) are correlated between the separate categories.

6.6 Statistical Treatment

In order to constrain the background contribution observed in the data and search for the presence of a signal a maximum likelihood fit is performed. The signal is defined with a signal strength modifier μ to account for a Higgs production that is different compared to the prediction in the standard model. The expected standard model cross section then becomes $\sigma = \mu \cdot \sigma_{sm}$. Next each source of systematic uncertainty is assigned a nuisance parameter θ_i. The expected signal $s(\theta)$ and background $b(\theta)$ yields are a function of these nuisance parameters. These nuisance parameters are constrained giving a probability density function $p_i(\tilde{\theta}_i|\theta_i)$ which defines the probability of a given value of $\tilde{\theta}_i$. A likelihood function is defined given the data and the measurements $\tilde{\theta}_i$:

$$\mathcal{L}(data|\mu \cdot s(\theta) + b(\theta)) = \mathcal{P}(data|\mu \cdot s(\theta) + b(\theta)) \cdot p(\tilde{\theta}|\theta) \tag{6.4}$$

where $\mathcal{P}(data|\mu \cdot s(\theta) + b(\theta))$ is the product of the probabilities of all bins of the SVFit mass distribution. Finally to test a given Higgs hypothesis an appropriate test statistic is defined. The test statistic is a number giving information on the observed data, expected signal, expected background and the uncertainties on these expectations. The test statistic ranks whether the observation is background like or consistent with the presence of a signal.

The chosen test statistic will depend on what is being testing. When quantifying the statistical significance of an excess the test statistic q_0 is defined:

$$q_0 = -2ln\frac{\mathcal{L}(data|b(\hat{\theta}_0))}{\mathcal{L}(data|\hat{\mu} \cdot s(\hat{\theta}) + b(\hat{\theta}))}, \qquad \hat{\mu} \geq 0 \tag{6.5}$$

where $\hat{\theta}_0$, $\hat{\theta}$, and $\hat{\mu}$ are the values that maximize the likelihoods in the numerator and the denominator. From here a p-value can be defined which is the probability of

the value of q_0 to be at least as large as the observed value under a background only interpretation. The significance $\mathcal{Z}$ can be determined taking the square root of the observed value of the test statistic $\mathcal{Z} = \sqrt{q_0^{obs}}$.

In the case where no signal excess is observed an exclusion limit can be set. Here the test statistic becomes:

$$q_0 = -2ln\frac{\mathcal{L}(data|\mu \cdot s(\hat{\theta}_\mu) + b(\hat{\theta}_\mu))}{\mathcal{L}(data|\hat{\mu} \cdot s(\hat{\theta}) + b(\hat{\theta}))}, \qquad 0 \leq \hat{\mu} \leq \mu \tag{6.6}$$

where the likelihood in the numerator is maximized as a function of the signal strength μ. From here the exclusion limit is calculated using a modified frequentist construction CL_s [62] to get the probability of finding q_μ larger than the observed value:

$$CL_{s+b} = \mathcal{P}(q_\mu \geq q_\mu^{obs}|\mu \cdot s + b) \tag{6.7}$$

$$CL_b = \mathcal{P}(q_\mu \geq q_\mu^{obs}|b) \tag{6.8}$$

From here we obtain the CL_s value from the ratio:

$$CL_s = \frac{CL_{s+b}}{CL_b} \tag{6.9}$$

If the value $\mathrm{CL}_s \leq \alpha$ for $\mu = 1$, then μ is excluded for 1-α confidence level. Then the limit with 95% confidence is found adjusting the value of μ to reach $\mathrm{CL}_s = 0.05$. For more details on the statistical treatment see [63, 9].

6.7 Summary

The analysis uses advanced techniques to reconstruct and identify electrons, muons, taus, jets and missing transverse energy. Collisions at the LHC come with a large

multi-jet background and are further complicated by the presence of event pileup. The particle flow algorithm makes use of information from all sub-detectors simultaneously to efficiently overcomes these issues. The jets and missing transverse energy used for this analysis start with the particle flow objects and then use multivariate approaches to minimize the effects of the event pileup. Electrons and muons use the particle flow objects to isolate them from fake leptons inside of jets. The τ_h candidates used in this analysis are identified using the hadron plus strips algorithm (HPS) which is built on top of particle flow objects. HPS is a combinatorial approach to τ_h identification that reconstructs the specific τ_h resonant decays looking for either a ρ or a_1 meson. Additionally single $\pi^{\pm}$ decays are considered. The τ_h candidates are then isolated directly by removing the tau decay components and looking for additional particle flow objects nearby. All important backgrounds are measured using data driven techniques: Drell-Yan to $\tau\tau$ using $Z \to \mu\mu$ events with τ leptons embedded in their place, the W+Jets normalizing to the data with $M_T > 70$ GeV and the multijet background using same-sign τ pair events.

The analysis is split into categories to exploit the Higgs boson production mechanisms in an attempt to boost the significance of the Higgs over the large irreducible $\mathrm{Z}^0 \to \tau\tau$ background. Events with vector boson fusion (VBF) topology are selected looking for 2 jets with a large separation in psuedorapidity and di-jet invariant mass. Additionally a missing energy selection is used to exploit the boosted nature of VBF events. The bulk of Higgs boson events are produced through gluon-gluon fusion (ggF). In ggF events additional jets from initial state radiation are more common then in the $q\bar{q}$ production of Z^0 boson events. This is used in this analysis with event categories that include at least one jet. Finally the 1-jet categories use the p_T of the τ_h to increase event significance by reducing the backgrounds due to the higher mass of the Higgs boson being searched for and the reduced τ_h fake rate at higher p_T.

Chapter 7

Analysis Results

7.1 Standard Model Higgs Search

In order to search for the presence of a Higgs boson signal a maximum likelihood fit of the SVFit mass distribution is performed, as discussed in Chapter 6. The search categories 1-jet low, 1-jet high, VBF and VBF Tight (8 TeV Only), as described in Chapter 6, are fit simultaneously along with two 0-jet categories, split into high and low τ_h p_T. The inclusion of the 0-jet category adds a constraint on the τ_h efficiency and energy scale. The choice not to fit signal for 0-jet is due to the fact that the signal contribution in this 0-jet region is very small, so that the statistical uncertainty on the background templates is much larger then the expected signal contribution which can lead to instability in the fit. The resulting post-fit mass distribution for the 0-jet category can be seen in Figure 7.1.

The fit results from the 1-jet and VBF categories can be seen in Figures 7.2 and 7.3 respectively. The power of the VBF categories can be clearly seen due to large separation of the signal and $Z \rightarrow \tau\tau$ distributions and the greatly reduced backgrounds. The shortcoming of this category is the low VBF production cross

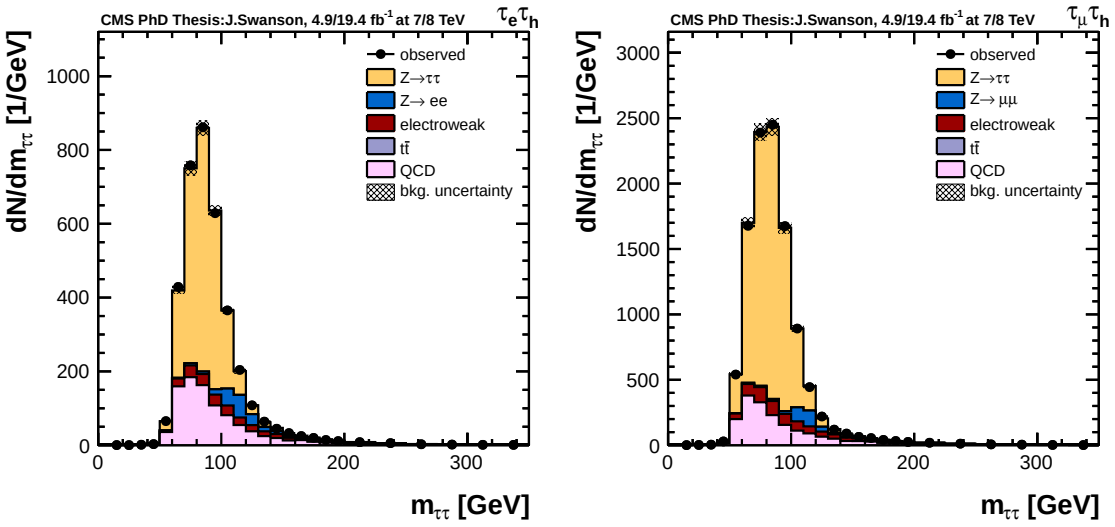

Figure 7.1: Post fit SVFit mass distributions of $\tau\tau$ events from the 0-Jet category. The low and high τ_h p_T categories and 7 TeV and 8 TeV are combined. The distribution are for the $\tau_e\tau_h$ (left) and $\tau_\mu\tau_h$ (right) channel respectively. The uncertainty on the background is represented in the shaded region and comes from the fitted nuisances.

section which makes the category a low statistics search category. The 1-jet category has a larger signal contribution but less separation between the Higgs boson signal and the backgrounds, most notably the $Z \to \tau\tau$.

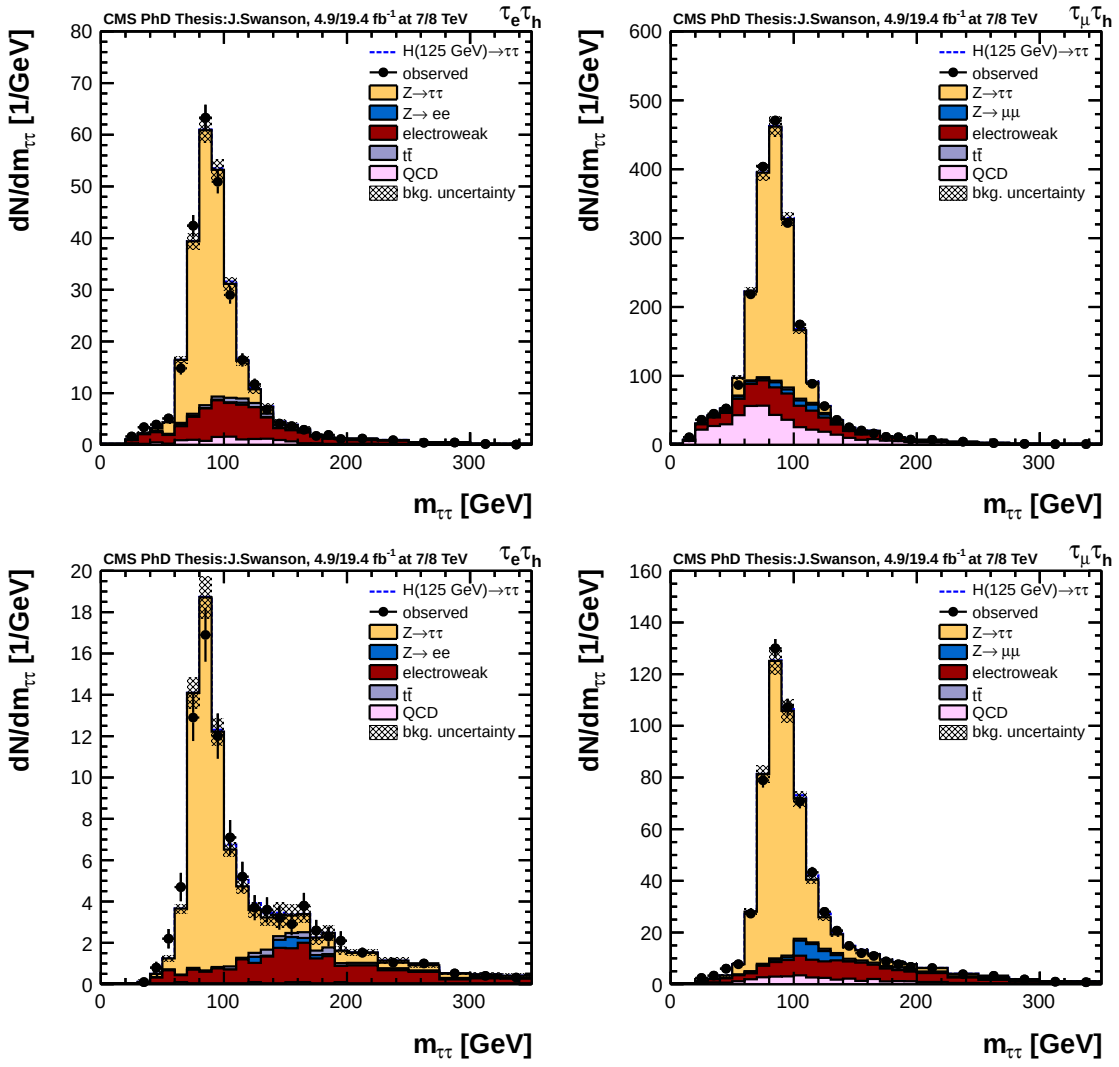

Figure 7.2: Post fit SVFit mass distributions of $\tau\tau$ events from the 1-jet categories. The 7 TeV and 8 TeV distributions are combined. The distribution are for the $\tau_e\tau_h$ (left) and $\tau_\mu\tau_h$ (right) channel and low (top) and high (bottom) τ_h p_T categories respectively. The uncertainty on the background is represented in the shaded region and comes from the fitted nuisances.

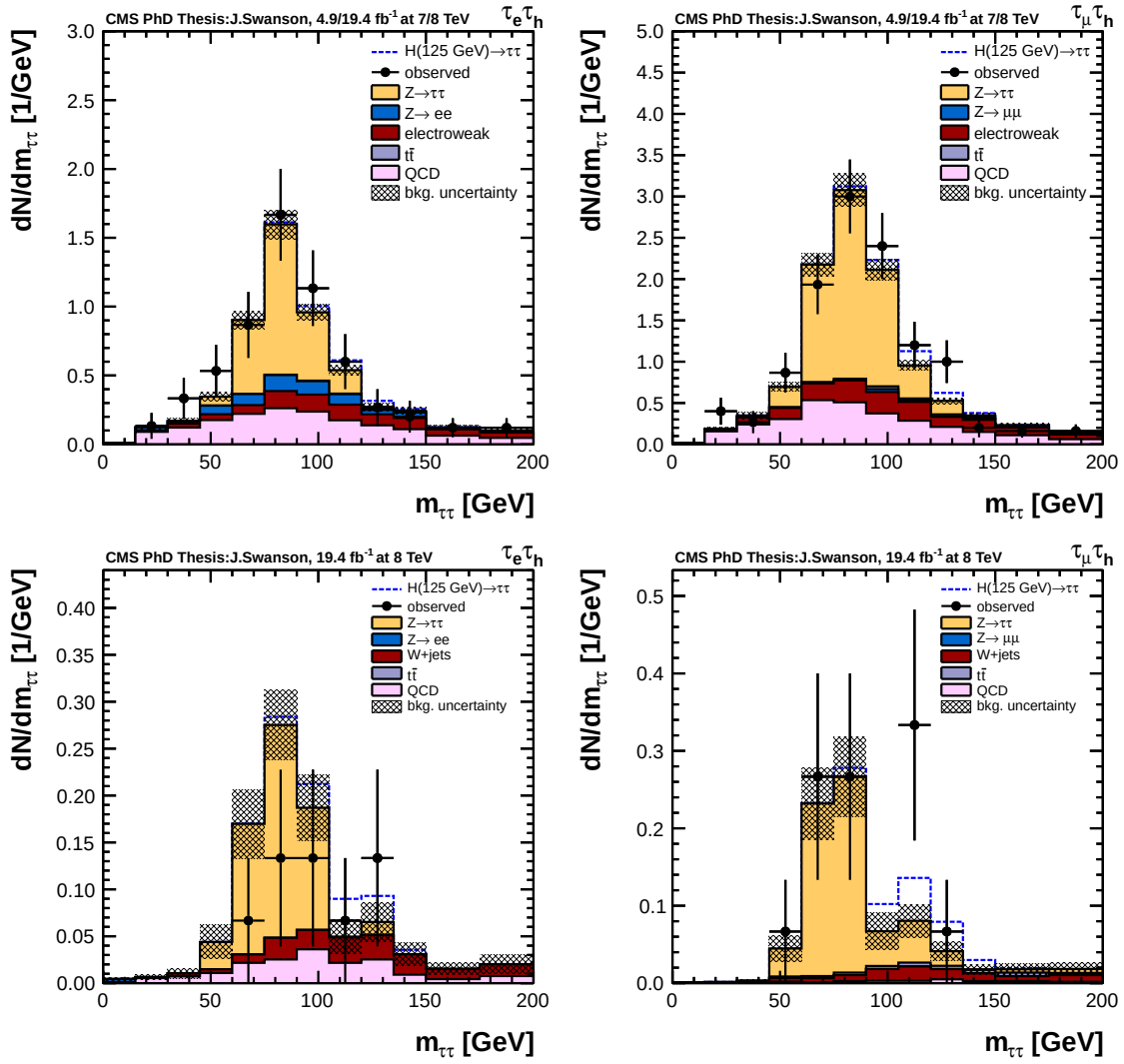

Figure 7.3: Post fit SVFit mass distributions of $\tau\tau$ events from the VBF categories. The 7 TeV and 8 TeV distributions are combined for VBF Loose and VBF tight is 8 TeV only. The distribution are for the $\tau_e\tau_h$ (left) and $\tau_\mu\tau_h$ (right) channel and loose (top) and tight (bottom) VBF categories respectively. The uncertainty on the background is represented in the shaded region and comes from the fitted nuisances.

Limits on Higgs boson Cross Section

In the absence of an excess of events a limit is set on the Higgs boson signal strength. The limits are plotted in Figures 7.4, 7.5 and 7.6 as a function of M_H at 95% confidence level (C.L.). The five plots include observed and expected exclusion limits and bands that identify 1 and 2 σ uncertainties on the expected result in the absence of signal. The limit results are plotted separately for the categories and channels in Figures 7.4 and 7.5 respectively. The combined limit is displayed in Figure 7.6 along with a plot showing the expected result with a standard model Higgs boson signal injected. The signal injected plot is done by producing toy MC for the Higgs signal contribution. This plot shows whether the observed excess is consistent with the expectation from the standard model Higgs cross section. The expected sensitivity at $m_H = 125$ GeV at 95% confidence is 0.96 times the standard model Higgs cross section with an observed exclusion of 2.12.

M_H	-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$	Obs. Limit
100 GeV	0.75	1.00	1.38	1.92	2.54	1.57
105 GeV	0.66	0.88	1.22	1.70	2.26	1.77
110 GeV	0.56	0.74	1.03	1.43	1.90	1.77
115 GeV	0.54	0.72	1.00	1.39	1.85	1.99
120 GeV	0.52	0.69	0.96	1.33	1.77	2.09
125 GeV	0.52	0.70	0.96	1.34	1.78	2.12
130 GeV	0.55	0.73	1.01	1.41	1.87	2.21
135 GeV	0.66	0.88	1.21	1.69	2.24	2.32
140 GeV	0.76	1.01	1.39	1.94	2.57	2.38
145 GeV	1.00	1.32	1.84	2.55	3.39	2.56

Table 7.1: Combined observed and expected upper limits on the standard model Higgs cross section at a 95% C.L.

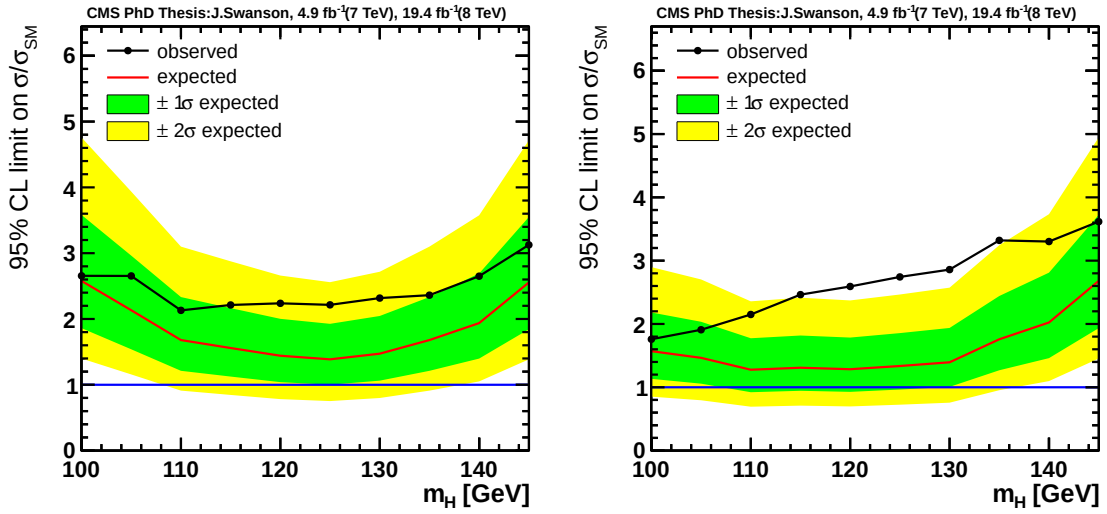

Figure 7.4: Observed and expected upper limit exclusions on the standard model Higgs cross section at at 95% C.L. as a function of Higgs boson mass. Green and yellow bands correspond to 1 and 2 sigma expected uncertainty in the absence of signal. The plots correspond to the 1-jet (left) and VBF (right) categories individually combined.

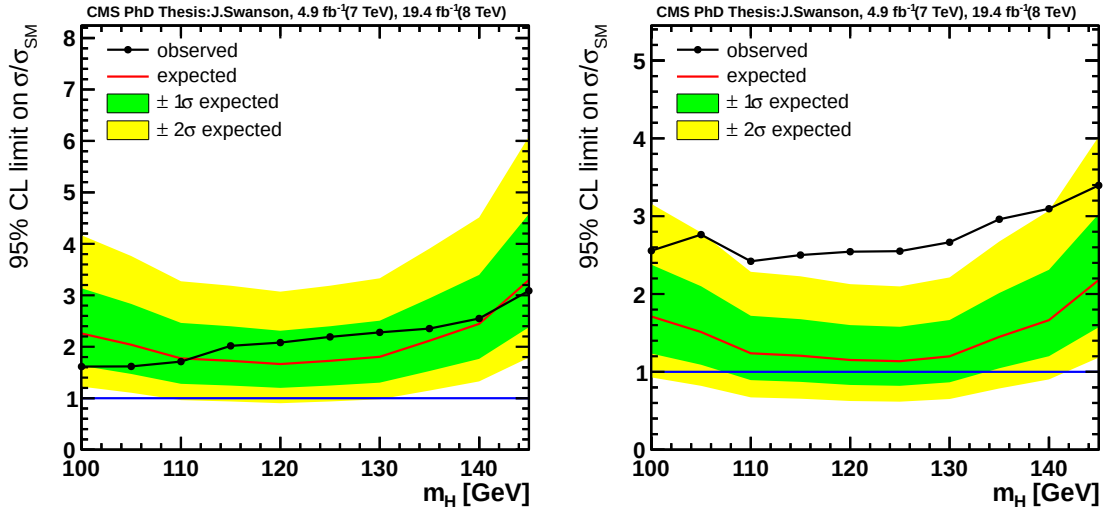

Figure 7.5: Observed and expected upper limit exclusions on the standard model Higgs cross section at at 95% C.L. as a function of Higgs boson mass. Green and yellow bands correspond to 1 and 2 sigma expected uncertainty in the absence of signal. The plots correspond to the limit set in the $\tau_e\tau_h$ (left) and $\tau_\mu\tau_h$ (right) channels separately.

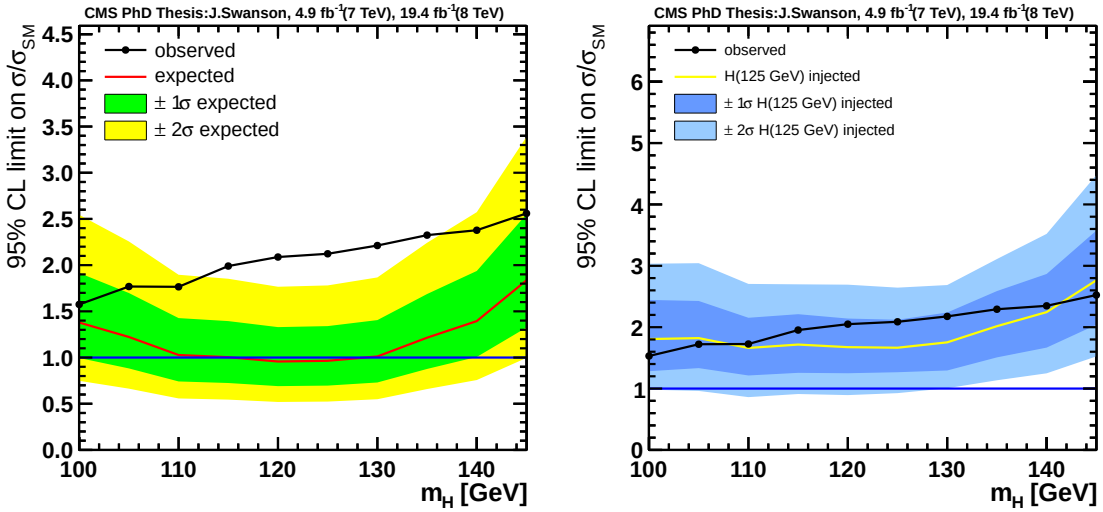

Figure 7.6: (Left) Combined observed and expected upper limit exclusions on the standard model Higgs cross section at at 95% C.L. as a function of Higgs boson mass. Green and yellow bands correspond to 1 and 2 sigma expected uncertainty in the absence of signal. (Right) Combined observed and expected, with MC signal toys injected, upper limit exclusions on the standard model Higgs cross section at at 95% C.L. as a function of Higgs boson mass. The inner and outer bands correspond to 1 σ and 2 σ uncertainties on the expected limit in the presence of a standard model Higgs signal injected.

Probability of a Background Only Hypothesis

With the presence of the observed excess of events a P_0 value is calculated. A P_0 value corresponds to the probability of the observed excess of events coming from an upward fluctuation of the background only. The expected P_0 value at $M_H = 125$ GeV is 1.5% corresponding to a needed 2.17 σ upward fluctuation of the background with an observed value of 0.5% corresponding to a needed 2.56 σ upward fluctuation of the background only. The observed and expected P_0 values as a function of M_H are shown in Figure 7.8.

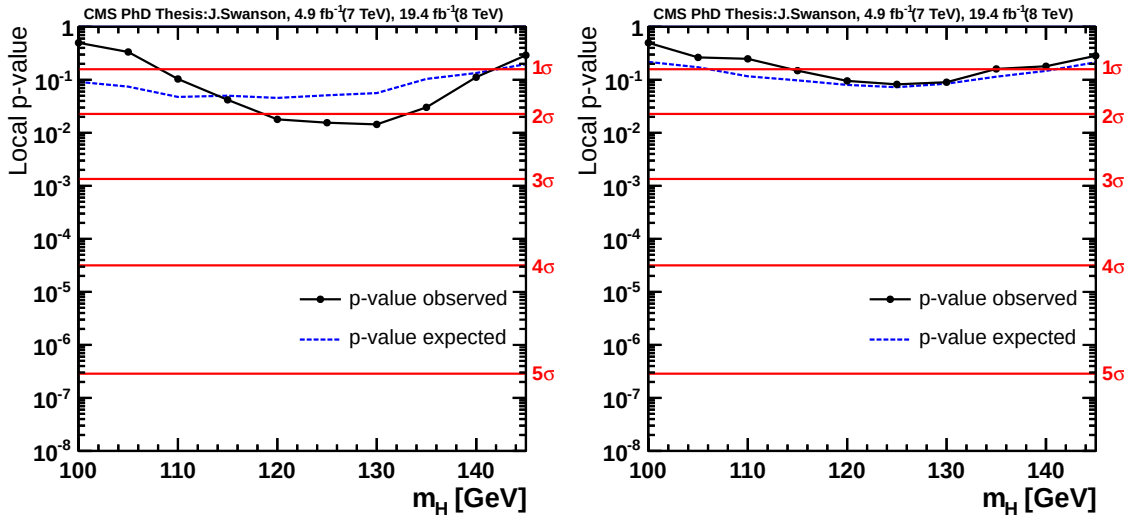

Figure 7.7: Probability of the background only producing the observed excess of events as a function of M_H. The blue dashed line represents the expected P_0 value and the black the observed. Statistical significance is represented with red horizontal lines at the appropriate P_0 values. The left plot corresponds to the VBF categories and the right to 1-Jet which is dominated by ggF Higgs production

Best fit Higgs boson Signal Strength

The maximum likelihood fit on the SVFit mass distribution also gives the best fit on the standard model Higgs signal strength. This tests the compatibility of the observed distributions with the standard model Higgs hypothesis. The fit is done separately for each category, channel and combined. The results are shown in Figure 7.9 for M_H = 125 GeV with a combined signal strength of $1.18^{+0.51}_{-0.48}$. Additionally the combined best fit signal strength as a function of M_H is shown in Figure 7.10.

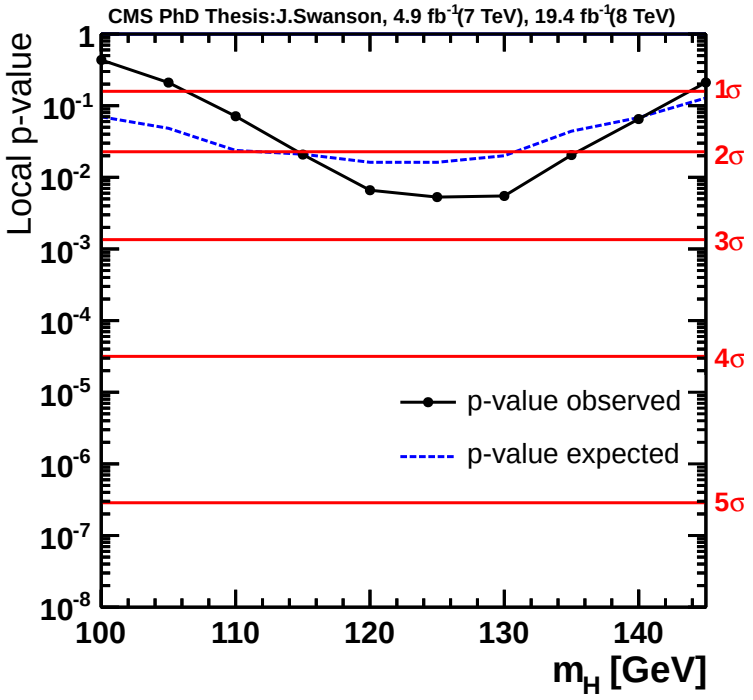

Figure 7.8: Probability of the background only producing the observed excess of events as a function of M_H. The blue dashed line represents the expected P_0 value and the black the observed. Statistical significance is represented with red horizontal lines at the appropriate P_0 values.

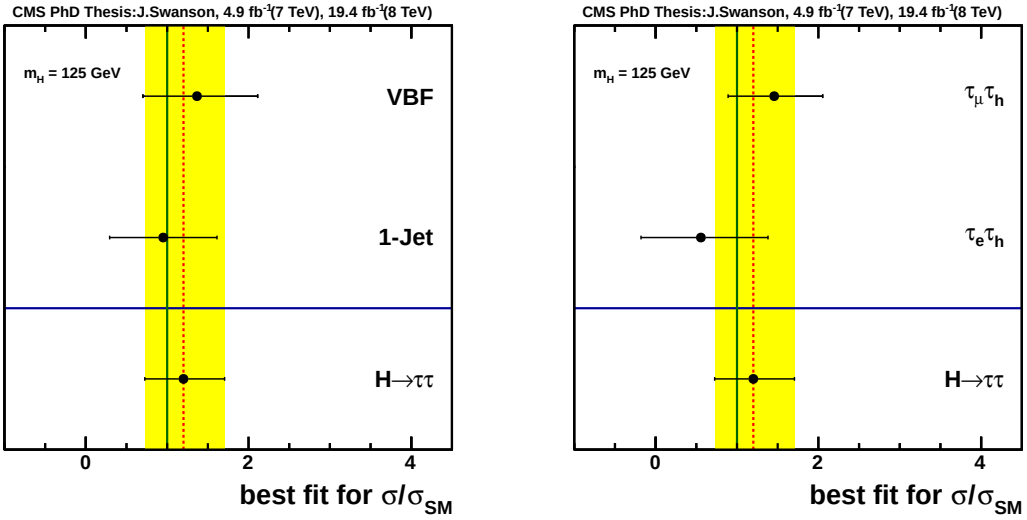

Figure 7.9: Value for the best fit standard model Higgs boson signal strength at M_H = 125 GeV split by category (left) and channel (right) with the combined value below the blue line. The yellow band corresponds to the combined $\pm$ 1 σ uncertainties and the green horizontal line representing 1 times the standard model Higgs boson cross section.

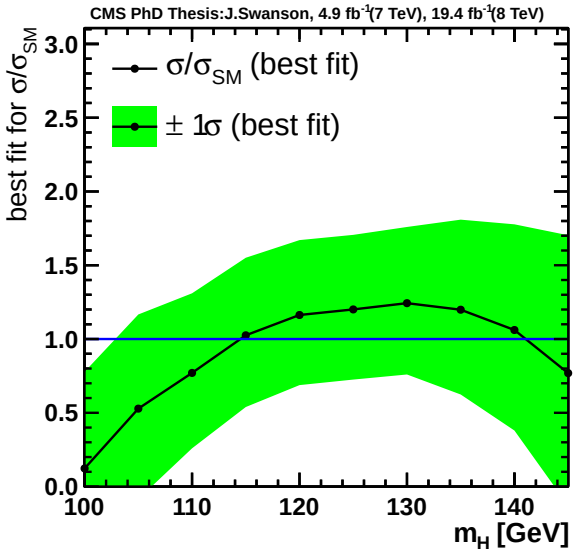

Figure 7.10: Value for the best fit standard model Higgs boson signal strength as a function of M_H. The green band corresponds to $\pm$ 1 σ uncertainties.

7.2 Conclusions

In this thesis a search for a standard model Higgs boson decaying to a pair of τ leptons was performed. The analysis used 19.4 fb^{-1}/4.9 fb^{-1} of 8 TeV/7 TeV data collected from the CMS experiment at the LHC during 2012/2011. It takes advantage of detailed detector information to efficiently trigger and reconstruct events. Events are further split into categories to fully exploit the Higgs production mechanisms to boost signal significance. Events with vector boson fusion (VBF) topology are selected looking for 2 jets with a large separation in psuedorapidity and di-jet invariant mass. Additionally a missing energy selection is used to exploit the boosted nature of VBF events. The bulk of Higgs boson events are produced through gluon-gluon fusion (ggF). To search for ggH production the analysis includes event categories that have at least one jet. The 1-jet categories use the p_T of the τ_h to increase event significance by reducing the backgrounds due to the higher mass of the Higgs boson being searched for and the reduced τ_h fake rate at higher p_T. The result shows a small excess of events over the backgrounds that hint that the new boson that was recently discovered at the LHC also decays to fermions and specifically to τ leptons.

Sensitivity to the standard model Higgs boson cross section is achieved. The expectation is to exclude the standard model cross section between 112 GeV $< m_H <$ 130 GeV at 95% C.L. instead a limit of 2.12 on the standard model Higgs signal strength is observed at $m_H = 125$ GeV. The probability that the observed excess of events is due to an upward fluctuation of the background only is 0.05% which amounts to an observed local significance of 2.56 σ. A best fit to the standard model Higgs boson signal strength of $1.20^{+0.51}_{-0.48}$ is observed at $m_H = 125$ GeV. This excess provides a hint that the new boson observed independently by the CMS and ATLAS experiments also decays to fermions and explicitly to τ leptons.

The analysis presented represents an extension of the latest CMS Higgs to $\tau\tau$ result [64]. The search utilizes 5 decay channels $\tau_h\tau_h$, $\tau_e\tau_\mu$ and $\tau_\mu\tau_\mu$ in addition to the two search channels discussed in this thesis $\tau_\mu\tau_h$ and $\tau_e\tau_h$. This latest search result has a local significance of 2.94 σ at $M_H = 120$ GeV.

Looking forward an updated version of the Higgs to $\tau\tau$ analysis from the CMS collaboration is expected for publication in Summer 2013. This analysis will incorporate some of the improvements discussed in this thesis along with improvements to the other search channels and possibly the inclusion of an additional search channel, $\tau_e\tau_e$. This updated analysis has the potential to give strong evidence for a new boson decaying to τ leptons. In 2015 the LHC plans to begin colliding protons again, at an energy closer to the design energy of 14 TeV and doubling the luminosity to up to $10^{34} cm^2 s$. During the higher energy run the search for $\tau\tau$ will look to confirm the decay of a new boson with 5 σ evidence and to measure its signal strength as accurately as possible. In addition decays to $\tau\tau$ has the potential to confirm VBF production of this new boson further validating whether the new boson is a standard model Higgs boson.

Bibliography

[1] S. Weinberg, "Recent progress in the gauge theories of the weak, electromagnetic and strong interactions," *Rev. Mod. Phys.*, vol. 46, no. 255, 1974.

[2] R. B. F. Englert, "Broken symmetry and the mass of gauge vector mesons," *Phys. Rev. Let.*, vol. 13, pp. 321–323, 1964.

[3] P. Higgs, "Broken symmetries, massless particles and gauge fields," *Phys. Let.*, vol. 12, pp. 132–133, 1964.

[4] P. Higgs, "Broken symmetries and the masses of gauge bosons," *Phys. Rev. Let.*, vol. 13, pp. 508–509, 1964.

[5] C. R. H. G. S. Guralnik and T. B. Kibble, "Global conservation laws and massless particles," *Phys. Rev. Let.*, vol. 13, pp. 585–587, 1964.

[6] P. Higgs, "Spontaneous symmetry breakdown without massless bosons," *Phys. Rev.*, vol. 145, pp. 1156–1163, 1966.

[7] T. B. Kibble, "Symmetry breaking in non-abelian gauge theories," *Phys. Rev.*, vol. 155, pp. 1554–1561, 1967.

[8] ATLAS Collaboration, "Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc," *Phys. Let. B*, vol. 716, no. 1, pp. 1–29, 2012.

[9] CMS Collaboration, "Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc," *Phys. Let. B*, vol. 716, no. 1, pp. 30–61, 2012.

[10] Q. R. A. et al. (SNO Collaboration), "Measurement of the rate of ν_e + d $\rightarrow$ p + p + e^- interactions produced by 8b solar neutrinos at the sudbury neutrino observatory," *Phys. Rev. Let.*, vol. 87, no. 7, 2001.

[11] R. P. Feynman, "Mathematical formulation of the quantum theory of electromagnetic interaction.," *Phys. Rev.*, vol. 80, pp. 440–457, 1950.

[12] S. L. Glashow, "Towards a unified theory: Threads in a tapestry.," *Rev. Mod. Phys.*, vol. 52, pp. 539–543, 1980.

[13] A. Salam, "Gauge unification of fundamental forces.," *Rev. Mod. Phys.*, vol. 52, pp. 525–538, 1980.

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

[15] J. et al. (Particle Data Group), *2012 Review of Particle Physics*. Phys. Rev. D86, 010001, 2012.

[16] M. L. P. et al., "Evidence for anomalous lepton production in $e^+e^?$ annihilation," *Phys. Rev. Let.*, vol. 35, p. 1489, 1975.

[17] CMS Collaboration, "Search for supersymmetry at the lhc in events with jets and missing transverse energy," *Phys. Rev. Let.*, vol. 107, no. 22, 2011.

[18] G. Ross, *Grand Unified Theories*. Westview Press, 1984.

[19] D. M. W. et al., "Measurement of the positive muon lifetime and determination of the fermi constant to part-per-million precision," *Phys. Rev. Lett.*, vol. 106, 2011.

[20] L. ALEPH, DELPHI and O. Collabs., "Search for the standard model higgs boson at lep," *Phys. Lett.*, vol. B565, 2003.

[21] CDF and H. W. G. D0 collabs.: Tevatron New Physics, "Updated combination of cdf and d0's searches for standard model higgs boson production with up to 10.0 fb^{-1} of data," *arXiv:1207.0449 [hep-ex]*, 2012.

[22] CDF Collaboration, "Search for the standard model higgs boson in $\tau^+\tau^-$+ jets final state with $8.3fb^{?1}$ of cdf data," *CDF Note 10625*, 2011.

[23] D0 Collaboration, "Search for the standard model higgs boson in tau lepton pair final states," *D0 Note 6305-CONF*, 2012.

[24] ATLAS Collaboration, "Measurements of the properties of the higgs-like boson in the two photon decay channel with the atlas detector using 25 fb^{-1} of proton-proton collision data," *ATLAS-CONF-2013-012*, 2013.

[25] ATLAS Collaboration, "Measurements of the properties of the higgs-like boson in the four lepton decay channel with the atlas detector using 25 fb^{-1} of proton-proton collision data," *ATLAS-CONF-2013-013*, 2013.

[26] ATLAS Collaboration, "Combined coupling measurements of the higgs-like boson with the atlas detector using up to 25 fb^{-1} of proton-proton collision data," *ATLAS-CONF-2013-034*, 2013.

[27] ATLAS Collaboration, "Search for the standard model higgs boson in h $\rightarrow$ tau tau decays in proton-proton collisions with the atlas detector," *ATLAS-CONF-2012-160*, 2012.

[28] CMS Collaboration, "Properties of the higgs-like boson in the decay h to zz to 4l in pp collisions at sqrt s =7 and 8 tev," *CMS-PAS-HIG-13-002*, 2013.

[29] CMS Collaboration, "Updated measurements of the higgs boson at 125 gev in the two photon decay channel," *CMS-PAS-HIG-13-001*, 2013.

[30] P. Fayet, "Spontaneously broken supersymmetric theories of weak, electromagnetic and strong interactions," *Phys. Lett. B*, vol. 69, 1977.

[31] CMS Collaboration, "Combination of sm, sm4, fp higgs boson searches," *CMS-PAS-HIG-12-008*, 2012.

[32] R. V. e. a. A. Breskin, *The CERN Large Hadrons Collider: Accelerator and Experiments*. CERN, 2008.

[33] CMS Collaboration, "The cms experiment at the cern lhc," *JINST*, vol. 3, 2008.

[34] S. Myers and E. Picasso, "The design, construction and commissioning of the cern large electronpositron collider," *Cont. Phys.*, vol. 31:6, 1990.

[35] CMS Collaboration, "Performance of cms muon reconstruction in pp collision events at $\sqrt{s} = 7$ tev," *CMS-PAS-MUO-10-004*, 2012.

[36] N. Metropolis and S. Ulam., "The monte carlo method," *Journal of the American Statistical Association.*, vol. 27, pp. 192–203, 1978.

[37] P. S. T. Sjostrand, S. Mrenna, "Pythia 6.4 physics and manual," *JHEP*, vol. 0605:026, 2006.

[38] J. A. et al., "Madgraph 5: Going beyond," *JHEP*, vol. 1106:128, 2011.

[39] P. Nason, "A new method for combining nlo qcd with shower monte carlo algorithms.," *JHEP*, vol. 0411, 2004.

[40] Z. Was, "Tauola the library for tau lepton decay, and kkmc/koralb/koralz status report," *Nucl. Phys. Proc. Suppl.*, vol. 98, pp. 96–102, 2001.

[41] S. Frixione and B. Webber, "Matching nlo qcd computations and parton shower simulations," *JHEP*, vol. 0206:029, 2002.

[42] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.), "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables," *CERN-2011-002*, 2011.

[43] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.), "Handbook of LHC Higgs Cross Sections: 2. Differential Distributions," *CERN-2012-002*, 2012.

[44] B. A. et al., "Partons fragmentation and string dynamics," *Phys. Reports*, vol. 97, 1983.

[45] S. A. et al., "G4- a simulation toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506:3, pp. 250–303, 2003.

[46] R. Fruhwirth, "Application of kalman filtering to track and vertex fitting," *Nucl.Instrum.Meth*, vol. A262, pp. 444–450, 1987.

[47] CMS Collaboration, "Description and performance of the cms track and primary vertex reconstruction," *CMS-PAS-TRK-11-001*, 2011.

[48] K. Rose, "Deterministic annealing for clustering, compression, classification, regression and related optimization problems," *Proceedings of IEEE*, vol. 86:11, 1998.

[49] W. A. et al, "Reconstruction of electrons with the gaussian-sum filter in the cms tracker at the lhc," *J. Phys. G: Nucl. Part. Phys.*, vol. 31 N9, 2005.

[50] CMS Collaboration, "Particle-flow event reconstruction in cms and performance for jets, taus, and met," *CMS-PAS-PFT-09-001*, 2009.

[51] M. C. et al, "The anti-k$_t$ jet clustering algorithm," *JHEP04*, vol. 063, 2008.

[52] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in cms," *JINST*, vol. 6, 2011.

[53] CMS Collaboration, "A combined secondary vertex based b-tagging algorithm in cms," *CMS-NOTE-2006-014*, 2006.

[54] CMS Collaboration, "Performance of the missing transverse energy reconstruction in the 8 tev data," *CMS-PAS-JME-12-002*, 2012.

[55] CMS Collaboration, "Performance of τ-lepton reconstruction and identification in cms," *JINST*, vol. 7, 2012.

[56] CMS Collaboration, "Measuring electron efficiencies at cms with early data," *CMS-PAS-EGM-07-001*, 2008.

[57] R. E. et al, "Higgs decay to $\tau^+\tau^-$: A possible signature of intermediate mass higgs bosons at the ssc," *Nucl. Phys. B*, vol. 297, p. 221, 1988.

[58] L. B. et al., "New svfit developments," *CMS-AN-2012-124*, 2012.

[59] K. H. B. Bullok and A. Martin, "Tau polarization and its correlations as a probe of new physics," *Nucl. Phys. B*, vol. 395, 1993.

[60] J. S. Conway, "Incorporating nuisance parameters in likelihoods for multisource spectra," *arXiv:1103.0354 [physics.data-an]*, 2011.

[61] CMS Collaboration, "Vbf task force recommendations," *CMS-AN-13-020*, 2013.

[62] A. L. Read, "Presentation of search results: the cls technique," *J. Phis. G: Nucl. Part. Phys.*, vol. 28, 2002.

[63] ATLAS and CMS Collaborations, LHC Higgs Combination Group, "Procedure for the lhc higgs boson search combination in summer 2011," *Technical Report ATL-PHYS-PUB 2011-11, CMS NOTE 2011/005*, 2011.

[64] CMS Collaboration, "Search for the standard-model higgs boson decaying to tau pairs in proton-proton collisions at sqrt(s) = 7 and 8 tev," *CMS-PAS-HIG-13-004*, 2013.