UNIVERSITY OF CALIFORNIA
RIVERSIDE

Search for the Standard Model Higgs Boson Produced in Association with a Z Boson in the Electron-Muon Final State and the Higgs Boson Decaying into Bottom Quarks

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Physics

by

Rachel A. Bartek

August 2013

Dissertation Committee:

Professor Robert Clare, Chairperson
Professor Aaron Dominguez
Professor Stephen Wimpenny

The Dissertation of Rachel A. Bartek is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I would like to thank my advisors, Prof. Bob Clare and Prof. Aaron Dominguez, for guiding me through a Ph.D. in high energy physics. I would also like to thank my honorary advisors, Prof. Ken Bloom and Prof. Stephen Wimpenny, for the many informal discussions and valued suggestions as my analysis developed.

I would also like to thank the EMU Collaboration. Their hard work and dedication inspired me to plunge into a career in high energy physics. Without their support and mentorship I would not have been at CERN for the start of the LHC run.

Looking further back in time, I would like the thank my undergraduate advisor, Prof. James Cochran who first wetted my palate for a taste in high energy physics and was a constant and continuous pillar of support from my graduate school applications through thesis writing. I look forward to having coffee with you soon, dear friend.

A special thanks also to my friends who did more than should have been asked of them, especially after my accident. Farz B., Roberta M., and Erdem T. were so supportive during this process. I would also like to thank the CERNies whose taco nights and TRT BBQs made life at CERN a true joy. We were brothers in arms, all going through the same process but facing our own challenges. I would also like to thank Josh F. and Jet S. for proof reading my thesis even though they are not physicists.

Last but certainly not least, I would like to thank my family. Mary, Sara, Anne, and Leah always kept the phone lines open for advice, for sharing my joys or just to lend a sympathetic ear. My grandma whose 92 years of wisdom always helped keep things in perspective.

To my parents.

Yes this is the default dedication from the template, but I really mean it. My parents instilled in me a sense that I could do anything - even discover a new particle. It is only because of their inspiration, continuing support and encouragement that this document exists today, so thanks Mom and Dad. This is dedicated to you!

ABSTRACT OF THE DISSERTATION

Search for the Standard Model Higgs Boson Produced in Association with a Z Boson in the Electron-Muon Final State and the Higgs Boson Decaying into Bottom Quarks

by

Rachel A. Bartek

Doctor of Philosophy, Graduate Program in Physics
University of California, Riverside, August 2013
Professor Robert Clare, Chairperson

A search for the standard model Higgs boson is presented in the associated production channel $Z(\tau\tau)H(b\bar{b})$ where each tau decays leptonically, one to an electron, the other to a muon and associated neutrinos. A data sample comprising of $5.0\,\mathrm{fb}^{-1}$ and $19.0\,\mathrm{fb}^{-1}$ from the 2011 and 2012 proton collision running periods at a center of mass of 7 and 8 TeV, respectively, has been analyzed and 95% C.L. upper limits derived for Higgs masses of 110-135 GeV.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Particle physics strives to explain the matter in our universe with a handful of elementary particles and a few simple interactions. This goal is partially achieved by the standard model (SM). This quantum field theory describes elementary particles of matter and their weak, electromagnetic and strong, interactions via gauge bosons. The constituents of matter are classified in three generations in the standard model. The standard model, developed in the 1960s and early 1970s, has been remarkably successful. Its quantitative description of the interactions of fundamental particles has been experimentally tested at the per mille level accuracy by high-precision measurement carried out at the accelerators LEP, SLC, HERA, the Tevatron, and other experiments and accelerators over the last decades.

One of the main goals of the Large Hadron Collider (LHC) physics program is the discovery of the Higgs boson. The standard model predicts the existence of a unique physical Higgs scalar [1] that generates the masses of the vector gauge bosons. The search

for the elusive Higgs boson was one of the main motivations for building the LHC which is a proton-proton collider which has been operating at a center-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$ starting March 30, 2010 and later at $\sqrt{s} = 8\,\text{TeV}$ in 2012. The discovery of a new particle at $125\,\text{GeV}$ [2] using $\sqrt{s} = 7\,\text{TeV}$ and $\sqrt{s} = 8\,\text{TeV}$ marked a great achievement for high-energy particle physics and science as a whole. While the mass of the Higgs boson is a free parameter in the SM, its coupling to massive vector bosons, Yukawa couplings to fermions, decay branching fractions, and production cross sections in proton-proton collisions are fully defined once the mass is specified. At $M_H = 125$ GeV the dominant decay mode for SM Higgs is $H \to b\bar{b}$ as seen in Figure 1.1.

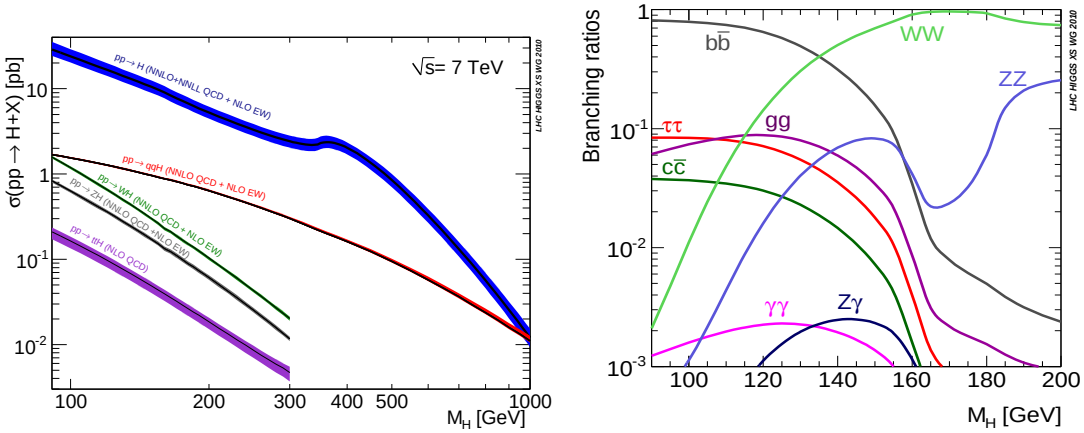

Figure 1.1: Production cross sections at $\sqrt{s} = 7\,\text{TeV}$ (left) and decay branching fractions (right) of the SM Higgs boson [3].

The search for a SM Higgs boson when the Higgs is produced in association with a Z boson, where the Higgs decays into b quarks and the Z decays into taus is presented in this transcript using all 7 and 8 TeV data accumulated by the CMS detector before Long Shutdown 1, the first of several periods without running to allow for accelerator and detector

improvements. The focus of this transcript is specifically the case where each tau decays leptonically, one into an electron, the other tau into a muon and associated neutrinos. This thesis uses natural units where $c = \hbar = 1$. This is convenient in particle physics because then spin is $1/2$ instead of $1/2\ \hbar$ and mass is GeV instead of GeV/c^2. These units elegantly simplify particle physics expressions. The need for a Higgs boson is motivated through the formalism of the SM in Chapter 2, where the importance of the Higgs boson is emphasized. The data used in this analysis is of collisions provided by the LHC accelerator recorded by the CMS detector as described in Chapter 3. A detailed description of the analysis and results are given in Chapter 4 including object selection requirements. Finally the main conclusions and lessons to be applied to future LHC data in the $e\mu b\bar{b}$ final state with associated neutrinos are summarized in the last chapter.

Chapter 2

The Higgs Boson and The Standard Model

Particle physicists strive to find the most comprehensive and simple explanation of the components of matter and their interactions. For more than fifty years the most comprehensive description has been the standard model of particle physics. This theory, while incomplete, is widely accepted and is described in brief in the next section.

2.1 Brief Discussion of The Standard Model

The standard model (SM) [4, 5] describes matter and the forces that control their interaction with each other. The SM is strikingly simple with only twelve point-like particles that have no known sub-structure. These elementary particles, called fermions, have half-integer spin and come in three generations. They come in two types: leptons and quarks. Each elementary particle has an antiparticle with same mass and characteristics,

but opposite charge. Antiparticles are denoted by a bar above their symbol or plus sign above their symbol in the charged lepton case. For example, an antineutrino is indicated by $\bar{\nu}$ and an anti-electron, also referred to as a positron, is indicated by e^+. The elementary particles and the corresponding generations of matter are listed in Table 2.1. There are three generations of matter. Each member of a higher generation has greater mass than the corresponding particle of the previous generation, with the possible exception of the neutrinos. Neutrinos are known to have small but non-zero mass [6] that have not been determined accurately, but neutrino mass is beyond the scope of the SM because the mechanism for giving neutrinos mass is unknown. Quarks are subject to both the strong and electroweak interaction, while leptons are only subject to the electroweak interaction.

The first generation consists of the particles that make up atoms: electrons, up quarks and down quarks. The nucleus of an atom consists of neutrons (n) and protons (p). These originally were thought to be elementary particles like electrons, but were found to consist of quarks [7]: the neutron (udd) and proton (uud). From chemistry, one recalls that the neutron is neutral and the proton has a +1 charge equal and opposite of the electron where the charge units are defined in such a way that the electron charge is -1. As both baryons, containing three quarks, and mesons, containing two quarks, have unit charge, it

Table 2.1: Overview of fermions in the standard model.

	First Generation	Second Generation	Third Generation
Quarks	up (u) down (d)	charm (c) strange (s)	top (t) bottom (b)
Leptons	electron neutrino (ν_e) electron (e)	muon neutrino (ν_μ) muon (μ)	tau neutrino (ν_τ) tau (τ)

becomes clear that quarks have fractional charge: $+2/3$ for u, c, t and -1/3 for d, s, b. The strong force is what holds two positively charged quarks together in a proton.

The term *baryon* is used to categorize these three quark combinations. One can imagine other three quark combinations to form other baryons. Consider a Δ^{++} consisting of three up quarks. At first glance this violates the Pauli Exclusion principle because quarks are fermions. The SM gives quarks another property called *color* charge, which unlike binary electric charge can have $2 \times 3 = 6$ values. This ternary property means that a red up quark, a blue up quark, and a green up quark can make a Δ^{++} without the three up quarks being in the same state. This characterization of the strong force leads to *quantum chromodynamics* (QCD). While this property does not have anything to do with actual color, the use of red, blue, and green is convenient. These primary colors of light combine to make white light. A key feature of the SM is that all naturally occurring particles are colorless.

Baryons and *mesons* both belong to the particle type known as *hadrons*. These particles are composed of color-carrying, fractional charged quarks. Mesons are quark-antiquark combinations. They are naturally colorless because one color combined with its anti-color is colorless. The first generation meson, the pion (π^-) consisting of an antiup quark and a down quark, is produced in copious amounts by cosmic rays in the upper atmosphere.

The SM is based on a set of conservation laws. Familiar laws like conservation of energy and charge are included. There are also several quantum numbers that must be conserved, depending on interaction. Each generation of lepton has an associated lepton flavor that must be conserved in decays. Antiparticles have opposite quantum numbers.

For example, when a tau lepton decays into a muon there are also neutrinos that conserve the lepton flavor on each side of the decay. A tau lepton has a tau lepton flavor (L_τ) of +1. A muon has a muon lepton flavor (L_μ) of +1. Therefore the decay of a tau into a muon is $\tau^- \rightarrow \nu_\tau + \mu^- + \bar{\nu_\mu}$. The charge is -1 and the lepton flavor is $L_\tau = +1$ before and after the decay. The complete list of lepton quantum numbers is listed in Table 2.2.

Similar to lepton flavor, quarks also have quantum numbers associated with each quark that must be conserved. However, in sharp contrast to lepton number, these quantum numbers must only be conserved in strong interactions not weak decays. The quantum number associated with the strange quark, strangeness, must be conserved in strong interactions, but can be violated in weak interactions. This phenomena was seen in pion beam deep inelastic scattering experiments [8]. If a pion-proton collision created a strange quark, an antistrange quark would also be created. Table 2.3 lists all baryon numbers for the quarks.

The interactions of these elementary particles is described by an exchange of force carriers. These particles are called *bosons* and have integer spin. The force carriers and the forces they mediate are listed in Table 2.4. The SM includes the well-described electroweak

Table 2.2: Quantum numbers for leptons.

Lepton	L_e	L_μ	L_τ	Antilepton	L_e	L_μ	L_τ	
ν_e	1	0	0	$\bar{\nu_e}$	-1	0	0	
e^-	1	0	0	e^+	-1	0	0	
ν_μ	0	1	0	$\bar{\nu_\mu}$	0	-1	0	
μ^-	0	1	0	μ^+	0	-1	0	
ν_τ	0	0	1	$\bar{\nu_\tau}$	0	0	-1	
τ^-	0	0	1	τ^+	0	0	-1	

Table 2.3: Quantum numbers for quarks along with their charges. There are six "flavors" of quarks, which are classified according to charge (Q), strangeness (S), charm-ness (C), bottom-ness(B) and top-ness (T). Upness (U) and downness (D) are redundant because only a quark with S=C=B=T=0 and Q= $\frac{2}{3}$ is the up quark.

Quark	Charge	D	U	S	C	B	T
u	2/3	0	1	0	0	0	0
d	-1/3	-1	0	0	0	0	0
c	2/3	0	0	0	1	0	0
s	-1/3	0	0	-1	0	0	0
t	2/3	0	0	0	0	0	1
b	-1/3	0	0	0	0	-1	0

and strong forces. Gravity and its corresponding particle, the graviton, are not part of the SM. However, there is no experimental indication of quantum gravity so it is unknown how to include it in the SM. Gravity, the force that holds together the universe, inspired the beginnings of physics. Its great disparity in strength from the other forces makes it negligible at the distance and mass scales being considered in particle physics.

The SM is built on the principle of looking for local symmetry to describe fundamental interactions between particles. Particle theorists use group theory to help restore order to what had become a zoo of particles. First they found that the $U(1)_{EM}$ group describes the electromagnetic interaction well [9]. The unification of the weak and elec-

Table 2.4: Overview of force carriers in the standard model along with their properties of mass and electric charge.

Name	Symbol	Force	Mass (GeV)	Electric Charge
gluon	g	strong	0	0
photon	γ	electromagnetic	0	0
W boson	W^+	weak	80.4	+1
W boson	W^-	weak	80.4	-1
Z boson	Z	weak	91.187	0

tromagnetic forces to the electroweak was described by Glashow, Weinberg, and Salam as the local symmetry of $SU(2)_L \times U(1)_Y$ [5, 10, 11] in the 1960s. The mathematical formalization of the color symmetry describes the strong interaction as $SU(3)_C$ [12]. Thus, the combination $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetry forms the modern standard model.

Local gauge invariance of *quantum electrodynamics* (QED) and *quantum chromodynamics* (QCD) yield massless photons and gluons. Even the force carrier of gravity, the graviton, is postulated to be massless in many models. While small variations of the SM can be made to accept small masses, in Table 2.4 we see that the weak force carriers are far from massless. Local gauge invariance in the electroweak sector yields naturally massless weak force carriers, but theses particle have been measured to have mass. If these masses are put into the SM by hand, unrenormalizable divergences appear and the theory becomes meaningless. This necessitates the Higgs Mechanism [1] to save the SM.

2.2 The Higgs Mechanism in the Standard Model

The mathematical formalism of relativistic quantum field theory [13] will be introduced in this section. Quantum field theory describes the realm of the very small and the very fast (relativistic particles). This is the realm where particle physics is done. In the previous section, it was mentioned that the masses of the weak force carriers, the W and Z bosons, break the local gauge symmetry in the SM. This begs the question: is symmetry important? The short answer is yes.

Noether's theorem [14] states that symmetries, or invariances, imply a conservation law. Translation invariance implies the conservation of momentum. Time translational

invariance implies the conservation of energy. Rotational invariance implies the conservation of angular momentum. Through the field formalism of the electromagnetic forces we will show that gauge invariance of the fermion field leads to charge conservation in the following subsection. Then by building upon the field formalism in the following sections a proof of local gauge invariance implying the existence of the Higgs boson will be presented.

2.2.1 Electromagnetic Interactions in The Standard Model

To prove gauge invariance leads to conservation of charge current, we first need to express the electromagnetic force as a Lagrangian density, or simply Lagrangian, of fields. We consider the electromagnetic Lagrangian of a Dirac field ψ, which describes fermions with spin-1/2 $s = 1/2$, and mass m:

$$\mathcal{L} = i\bar{\psi}\gamma_\mu\delta^\mu\psi - m\bar{\psi}\psi \tag{2.1}$$

where γ_μ are Dirac matrices and $\bar{\psi} \equiv \psi^\dagger\gamma^0$. The Dirac equation gives the corresponding equation of motion:

$$(i\gamma_\mu\delta^\mu - m)\psi = 0 \tag{2.2}$$

If we consider a phase transformation of the form $\psi(x) \to e^{i\alpha}\psi(x)$ where α is a real constant, it is clear that 2.1 is invariant under this transformation. This family of phase transformations $U(\alpha) \equiv e^{i\alpha}$ forms a unitary *Abelian* group, the U(1) group. An Abelian group has the property that multiplication is commutative, $U(\alpha_1)U(\alpha_2) = U(\alpha_2)U(\alpha_1)$, for all members of the group. By Noether's theorem this global gauge invariance implies the existence of a conserved current.

By extending this global gauge transformation by changing α from a constant to

be dependent on space or time, $\psi(x) \rightarrow e^{i\alpha}\psi(x)$ becomes $\psi(x) \rightarrow e^{i\alpha(x)}\psi(x)$. Now we test to see if the electromagnetic Lagrangian is invariant under local gauge transformations. At first glance:

$$\delta_\mu \psi \rightarrow e^{i\alpha}\delta_\mu \psi + ie^{i\alpha}\psi\delta_\mu\alpha \tag{2.3}$$

the second term in 2.3 ruins the invariance. We seek a modified derivative that transforms covariantly under local gauge transformations like $\psi(x) \rightarrow e^{i\alpha}\psi(x)$. We introduce D_μ, the covariant derivative, that is D_μ transforms like ψ itself. The covariant derivative is defined as:

$$D_\mu \rightarrow e^{i\alpha(x)} D_\mu \tag{2.4}$$

$$D_\mu \equiv \delta_\mu - ieA_\mu \tag{2.5}$$

where A_μ is a gauge vector boson field and is constructed such that the unwanted term in 2.3 is canceled.

$$A_\mu \rightarrow A_\mu + \frac{1}{e}\delta_\mu\alpha \tag{2.6}$$

Replacing δ_μ by D_μ, the Lagrangian becomes:

$$\begin{aligned} \mathcal{L}' &= i\bar{\psi}\gamma_\mu D^\mu\psi - m\bar{\psi}\psi \\ &= i\bar{\psi}\gamma_\mu(\delta^\mu - ieA^\mu)\psi - m\bar{\psi}\psi \\ &= \bar{\psi}(i\gamma_\mu\delta^\mu - m)\psi + e\bar{\psi}\gamma_\mu\psi A^\mu \end{aligned} \tag{2.7}$$

The $e\bar{\psi}\gamma_\mu\psi A^\mu$ term in Equation 2.7 is the interaction term. If one replaces it by $-j^\mu A_\mu$, where j^μ is the current density, conservation of electromagnetic current and therefore the conservation of charge fall directly from the Lagrangian.

The full QED Lagrangian must contain a kinetic term for the new added photon vector field. The kinetic term must be invariant under Equation 2.6. The field strength tensor, $F_{\mu\nu} = \delta_\mu A_\nu - \delta_\nu A_\mu$, is added to Equation 2.7 to form the QED Lagrangian:

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma_\mu \delta^\mu - m)\psi + e\bar{\psi}\gamma_\mu \psi A^\mu - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{2.8}$$

a massive photon yields the Proca Lagrangian:

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + m_A A^\nu A_\nu \tag{2.9}$$

$A^\nu A_\nu$ is not invariant under Equation 2.6. Therefore, the gauge field must be massless in order to preserve local gauge invariance. A massless photon is expected because the range of the electromagnetic force is infinite and photon have been observed to travel at the speed of light.

Thus through local gauge invariance we derive the QED Lagrangian in Equation 2.8 which is invariant under U(1) gauge transformations. We developed a description of a massless gauge particle, the photon, interacting with Dirac fields, electrons and positrons. While U(1) is just a number, in the next section we will derive the electroweak interaction which SU(2)$\times$ U(1), thus we invoke group theory to describe the electromagnetic interaction here.

2.2.2 Electroweak Interactions in The Standard Model

The weak force is combined with the electromagnetic force to form the electroweak interaction. $SU(2)_L$ is the weak isospin group, which acts on left-handed fermions, or particles with left chirality. $U(1)_Y$ is the weak hypercharge group where the quantum

number hypercharge is defined as:

$$Q \equiv T^3 + \frac{Y}{2} \tag{2.10}$$

where Q is the electric charge and T^3 is weak isospin. Fermions can be divided into right-handed singlets, ℓ_R, and left-handed field doublets:

$$\chi_L = \begin{pmatrix} \nu_{\ell,L} \\ \\ \ell_L \end{pmatrix} \tag{2.11}$$

Neutrinos have no right-handed component, as their mass is taken as zero. The doublet the Lagrangian then becomes:

$$\mathcal{L} = i\bar{\chi}\gamma_\mu\delta^\mu\chi - \bar{\chi}M\chi \tag{2.12}$$

where M is a 2x2 diagonal mass matrix. If the two masses happen to be equal we recover the Lagrangian 2.1. Local gauge transformation in the non-Abelian SU(2) group is written as a function of its generators $T_i = \frac{\sigma_i}{2}$ where σ_i are the Pauli matrices.

The number of gauge bosons associated with $SU(2)_L \times U(1)_Y$ is an iso triplet vector field W^i_μ with a coupling constant g, where $i = 1, 2, 3$ and the vector field B_μ with a coupling constant g'. We consider the local gauge transformations:

$$\chi_L \to e^{i\alpha_a(x)T^a} e^{i\beta(x)Y} \chi_L \tag{2.13}$$

and

$$\psi_R \to e^{i\beta(x)Y} \psi_R \tag{2.14}$$

We find the corresponding covariant derivative to be:

$$D_\mu = \delta_\mu - ig'\frac{Y}{2}B_\mu(x) - ig\frac{\tau_a}{2}W^a_\mu(x) \tag{2.15}$$

Exactly as in the QED case, the $U(1)_Y$ field strength tensor is $B_{\mu\nu} = \delta_\mu B_\nu - \delta_\nu B_\mu$. The $SU(2)_L$ field strength tensor is $W_{\mu\nu} = W^a_{\mu\nu}\frac{\tau_a}{2}$. The resulting Lagrangian is:

$$\begin{aligned}\mathcal{L} &= \bar{\chi}_L\gamma^\mu\left[i\delta_\mu - g\frac{1}{2}\tau_a W^a_\mu - g'(-\frac{1}{2})B_\mu\right]\chi_L \\ &+ \bar{e}_R\gamma^\mu\left[i\delta_\mu - g'\left(-1\right)B_\mu\right]e_R - \frac{1}{4}W^a_{\mu\nu}W^{\mu\nu}_a - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}\end{aligned} \tag{2.16}$$

Describing the weak force using relativistic quantum field theory seems simple enough: find the appropriate gauge transformation and adjust the Lagrangian accordingly. The problem arises from the massive bosons that mediate the weak force. The W and Z are known to be massive because of the low strength of the weak force. These massive particles were discovered in 1983 by the UA1 and UA2 experiments using collisions from the Super Proton Synchrotron (SPS) at CERN [15,16]. masses were QCD and QED require massless bosons to preserve local gauge invariance. The solution is to preserve the global symmetry of the potential but spontaneously break the symmetry by choosing to expand around a minimum point which is not the global symmetry point. This work was developed by Carl R. Hagen, Robert Brout, François Englert, Gerard Guralnik, Thomas W.B. Kibble, and Peter Higgs [1, 17, 18] in the 1960s. Most of these prominent particle theorists were in attendance when the discovery of their particle, referred to as the Higgs boson, was announced July 4, 2012 at CERN.

2.2.3 Spontaneous Symmetry Breaking

We approach spontaneous symmetry breaking using the same methods as in QED and the electroweak quantum field theory formalism. We consider a complex scalar field

$\phi = (\phi_1 + i\phi_2)/\sqrt{2}$. The potential of ϕ is:

$$V(\phi) = \mu^2\phi^*\phi + \lambda(\phi^*\phi)^2 \tag{2.17}$$

where μ is the particle mass and λ is a dimensionless constant, representing the coupling of the 4-boson vertex. The Lagrangian associated with ϕ is:

$$\mathcal{L} = (\delta_\mu\phi) * (\delta^\mu\phi) - \mu^2\phi^*\phi - \lambda(\phi^*\phi)^2 \tag{2.18}$$

Consider the transformation $\phi \to e^{i\alpha}\phi$. A familiar covariant derivative, $D_\mu = \delta_\mu - ieA_\mu$ is invariant under this local gauge transformation. A_μ is the gauge field that transforms as $A_\mu \to A\mu + \frac{1}{e}\delta_\mu\alpha$. The gauge invariant Lagrangian is then

$$\begin{aligned} \mathcal{L} &= (\delta^\mu + ieA^\mu)\,\phi^*\,(\delta_\mu - ieA_\mu)\,\phi \\ &- \mu^2\phi^*\phi - \lambda(\phi^*\phi)^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \end{aligned} \tag{2.19}$$

Without the λ term in Equation 2.19 is the QED Lagrangian in Equation 2.8. However, if we take $\mu^2 < 0$ the potential takes the form of the potential pictured in Figure 2.1. The circle of minimum is the ϕ_1, ϕ_2 plane has a radius v. The minimum is:

$$\phi_1^2 + \phi_2^2 = v^2 \text{ where } v^2 = -\frac{\mu^2}{\lambda} \tag{2.20}$$

Note that this potential still maintains a radial symmetry. It is not until we arbitrarily choose $\phi_1 = v$ and $\phi_2 = 0$ is this symmetry spontaneously broken. We invoke perturbation theory, expanding around this chosen minimum by replacing ϕ with:

$$\phi(x) = \sqrt{\frac{1}{2}}\,[v + \eta(x) + i\xi(x)] \tag{2.21}$$

where $\eta(x)$ represents quantum fluctuations about the minimum. $\xi(x)$ is a massless field with fluctuations in the ϕ_2 direction. We substitute this expansion around the vacuum into

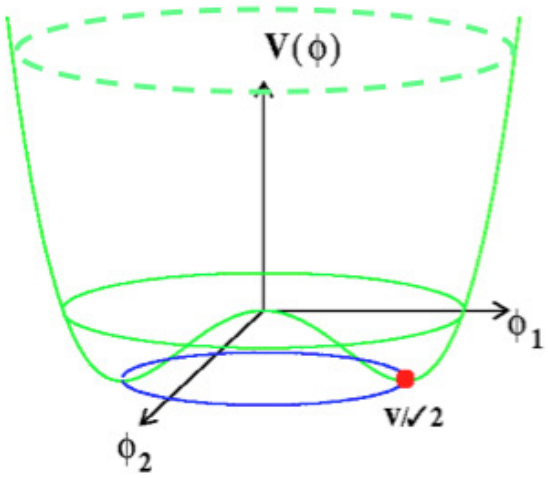

Figure 2.1: The potential $V(\phi)$ for a complex scalar field $\phi = (\phi_1 + i\phi_2)/\sqrt{2}$ where $\mu^2 < 0$ and $\lambda > 0$ with a minimum at $v^2 = -\frac{\mu^2}{\lambda}$.

Equation 2.19 to obtain:

$$\begin{aligned} \mathcal{L}' &= \frac{1}{2}(\delta_\mu \xi)^2 + \frac{1}{2}(\delta_\mu \eta)^2 + v^2\lambda\eta^2 + \frac{1}{2}e^2v^2A_\mu A^\mu - evA_\mu\delta^\mu\xi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\ &+ \text{constants} + \text{ higher order interaction terms of } \eta \text{ and } \xi. \end{aligned} \tag{2.22}$$

The third term is the mass term for the η-field. Simple algebra gives a mass of $m_\eta = \sqrt{2\lambda v^2}$. A massive scalar field has been dynamically generated, achieving the goal of creating a Lagrangian with a massive vector field with $m_A = ev$. Our theory is still re normalizable, unlike when the masses of the W and Z bosons are put in by hand. The symmetry of the Lagrangian is not lost, just hidden by our choice of the ground state. The price is a massless scalar field ξ, known as a Goldstone boson. The Goldstone theorem [19] states that massless scalars occur whenever a continuous symmetry of a physical system is spontaneously broken. How do we interpret this new field? We look to the $evA_\mu\delta^\mu\xi$ term in Equation 2.22 for

guidance. We choose a gauge with a different set of real fields:

$$\begin{aligned} \phi &\rightarrow \sqrt{\frac{1}{2}}\left(v + h(x)\right) e^{i\theta(x)/v} \\ A_\mu &\rightarrow A_\mu + \frac{1}{ev}\delta_\mu\theta \end{aligned} \tag{2.23}$$

In particular $\theta(x)$ is chosen so that h is real. With this more careful choice of gauge the Goldstone boson is "gauged away", thus does not appear in the Lagrangian. This leaves just two massive particles: a vector gauge boson A_μ and scalar h, referred to as the Higgs boson.

In $SU(2)_L$, the field ϕ is a doublet of complex scalar fields. Using the more complex $SU(2)$ gauge results in three Goldstone bosons that are gauged away. After the smoke clears, there are mass terms for all weak vector bosons, a massless photon field and a single remaining scalar field with a mass term, the Higgs boson. Keep in mind that one Higgs boson is the simplest solution to give vector bosons mass, there is nothing that prohibits a more complex Higgs sector. The SM uses the simplest case.

2.3 Searches for and Discovery of the SM Higgs Boson

The SM predicts a Higgs boson, but it does not specify its mass. From the previous section, the masses from the Lagrangian are:

$$m_H = \sqrt{2\lambda v^2};\; m_W = \frac{gv}{2};\; m_Z = \frac{\sqrt{g^2 + g'^2 v}}{2} \tag{2.24}$$

From the masses of the W and Z bosons we calculate v to be of the order of 246.6 GeV. The Higgs self-coupling constant λ is undetermined by the SM.

Theoretical constraints for the Higgs boson mass give upper and lower bounds on this free parameter in the SM. Triviality gives an upper bound. This is a cut off energy for the Higgs self-coupling λ to remain finite. We require that $\Lambda < \Lambda_C$ where the critical energy is:

$$\Lambda_c = ve^{\frac{4\pi^2 v^2}{m_H^2}} \tag{2.25}$$

Assuming the cut off energy is close to the typical Grand Unified Theory (GUT) scale, $\Lambda = 10^{16}\,\text{GeV}$, Triviality gives an upper bound of $m_H < 170\,\text{GeV}$. This limit is shown in Figure 2.2.

Unitarity gives another upper bound for the Higgs mass. In the high energy limit, introduction of a new scalar particle controls divergences in the scattering of longitudinally polarized vector bosons. This constraint is interpreted as an upper bound on the Higgs mass: $m_H < \sqrt{\frac{16\pi}{3}}v \sim 1\,\text{TeV}$. More stringent constraints give $m_H \lesssim 780\,\text{GeV}$. With this upper limit in mind, physicists were confident the Higgs boson would be discovered at the LHC if it existed as predicted by the SM.

The lower theoretical bound for the Higgs mass comes from vacuum stability. In spontaneous symmetry breaking the minimum or vacuum is asymmetric. The vacuum must still be stable under quantum corrections. This implies that:

$$m_H^2 > \frac{4m_t^4}{\pi^2 v^2} \ln\left(\frac{\Lambda}{v}\right) \tag{2.26}$$

Using $m_t = 172.5\,\text{GeV}$ the lower limit coming from vacuum stability is $m_H > 140\,\text{GeV}$. A lower constraint is found by requiring metastable bounds [20] as shown in Figure 2.3.

Indirect measurements of the Higgs boson through electroweak precision measurements favor a low mass Higgs. LEP, CDF and D0 measured contributions to loop diagrams

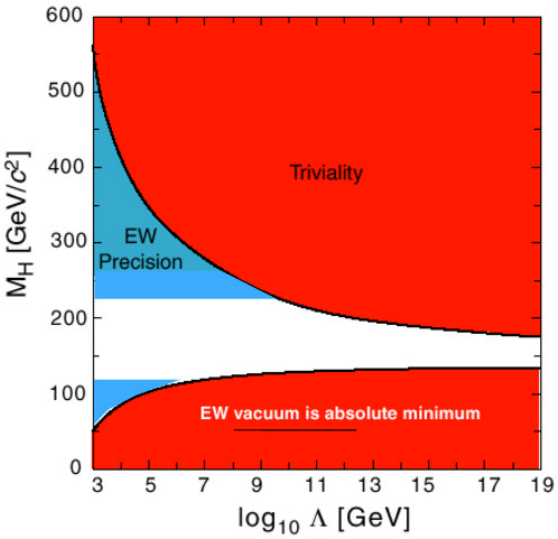

Figure 2.2: The theoretical constraints on the Higgs mass are shown in red. The bounds from electroweak precision measurements are shown in blue.

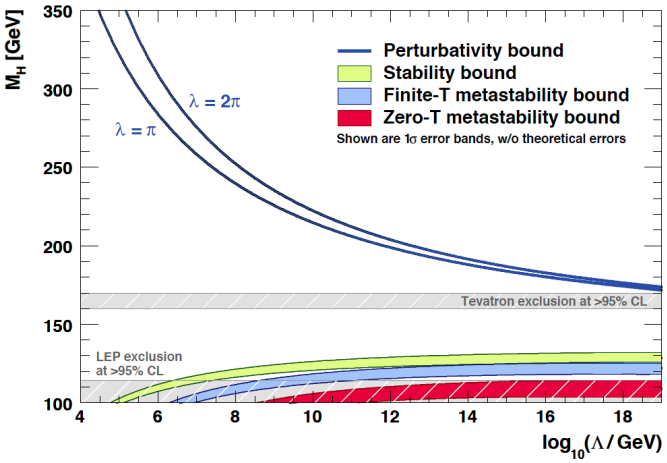

Figure 2.3: The theoretical constraints on the Higgs mass are shown with experimental excluded regions from LEP and Tevatron with 2.0-3.6 fb^{-1} of data analyzed at CDF and 0.9-4.2 fb^{-1} data analyzed at D0 in grey. The perturbativity bound, also known as the triviality bound, is given for $\lambda = \pi$ and $\lambda = 2\pi$. The difference between the two blue curves indicates the theoretical uncertainty on this bound. The vacuum stability bound is in light green, while less restrictive metastability scenarios are shown in blue and red.

involving the Higgs boson. The $\Delta\chi^2$ curve shown in Figure 2.4 was derived from precision electroweak measurements. Fits such as these were used to guide experimental design choices for the LHC.

Before the LHC was constructed the highest limits from direct searches of the Higgs boson came from the Large Electron Positron Accelerator (LEP) [21]. This was an electron-positron collider that allowed for very precise measurements of the W and Z boson masses. The Higgs boson was searched for in two production modes: associated production with a Z boson and vector boson fusion. The combined results of all LEP experiments: ALEPH, L3, DELPHI, and OPAL, gave a limit of $m_H > 114.4\,\text{GeV}$ [21]. The final reconstructed Higgs mass plot Figure 2.5 shows a few events around 120 GeV. These events lead some physicists to lobby to an extension of the LEP program. Ultimately the decision was made to shut down LEP to make way for the LHC.

After the first collisions at the LHC on March 30, 2010, the Tevatron and LHC were racing to be first to discover the Higgs boson. The Tevatron experiments had the advantages of well understood detectors and initially more data. The more mature Tevatron analyses also employed more advanced analysis techniques such as jet energy regression and background-specific boosted decision trees for their $H \to b\bar{b}$ analysis. The LHC soon delivered as many collisions in 18 months as the Tevatron had in its 10 years of running. The LHC also had a huge advantage in energy as the Tevatron ran at $\sqrt{s} = 1.8\,\text{TeV}$ and the LHC ran at $\sqrt{s} = 7\,\text{TeV}$ in 2011 and $\sqrt{s} = 8\,\text{TeV}$ in 2012. The newer LHC detectors also employ the most recent detector technologies. The final combined Tevatron [22] result is shown in Figure 2.6 shows a broad excess in the region $m_H = 125\,\text{GeV}$, but the result

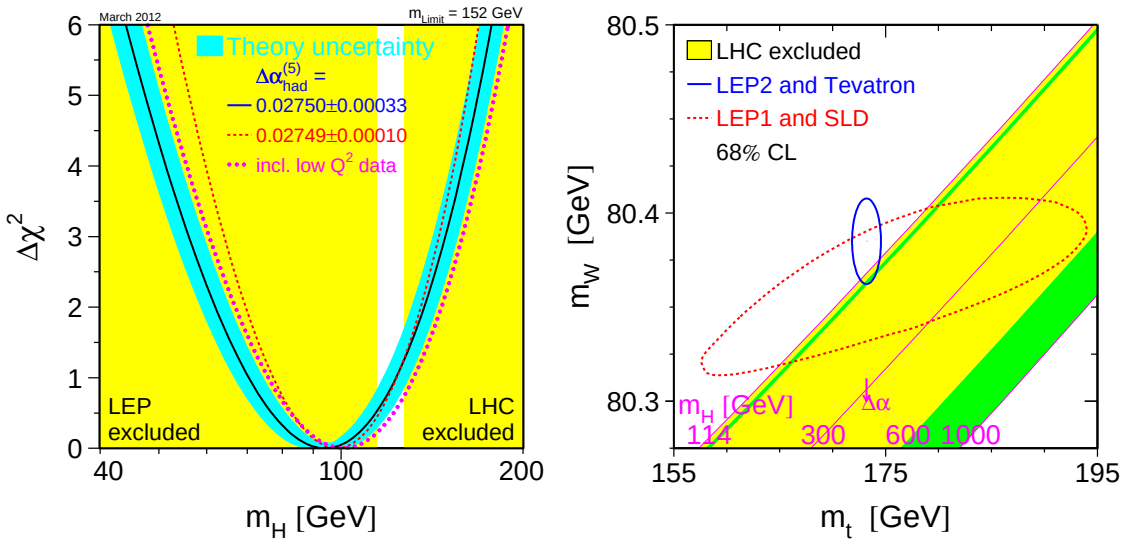

Figure 2.4: Best fit from electroweak precision measurements (left) and contours in W boson and top mass space (right) are shown. The Higgs mass regions excluded from direct measurements by the LHC experiments are shown in yellow.

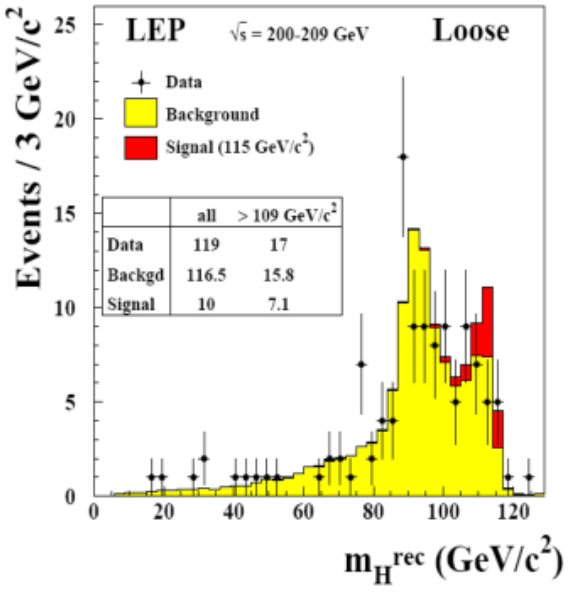

Figure 2.5: Reconstructed Higgs mass from candidate events from LEP.

is not significant enough to claim observation of the Higgs boson. However, the Tevatron does rule out the region between $145 < m_H < 178$ GeV.

The Higgs boson is produced via several processes at the LHC. The Feynman diagrams of the dominant four processes are shown in Figure 2.7 and their production rates are shown in Figure 2.8. The Higgs couples to mass, therefore, it couples more strongly to heavy particles like the vector bosons and top quarks. Note that even though searching for the Higgs boson produced via gluon fusion and decaying to the $b\bar{b}$ final state would give the highest rate, we look for $H \to b\bar{b}$ when produced in association with a vector boson due to the overwhelming background of direct $b\bar{b}$ production and the lepton decay modes of the vector boson give a more controllable trigger handle. The two cleanest Higgs boson signals at the LHC are $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$. These two decay modes gave the largest contributions to the Higgs boson discovery at CMS announced July 4, 2012 [2].

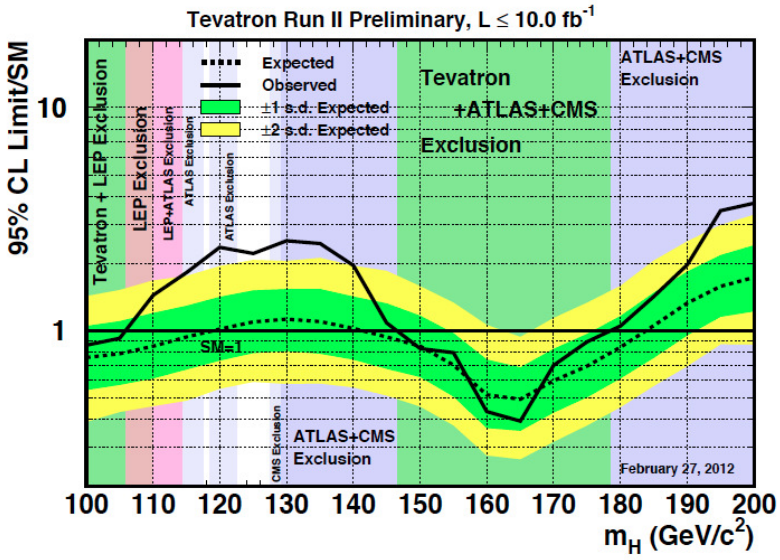

Figure 2.6: Observed and expected 95% confidence level upper limits on the ratio to the SM Higgs cross section for combined CDF and D0 analyses. The Higgs mass regions excluded from the ATLAS, CMS, LEP, and Tevatron experiments are shown.

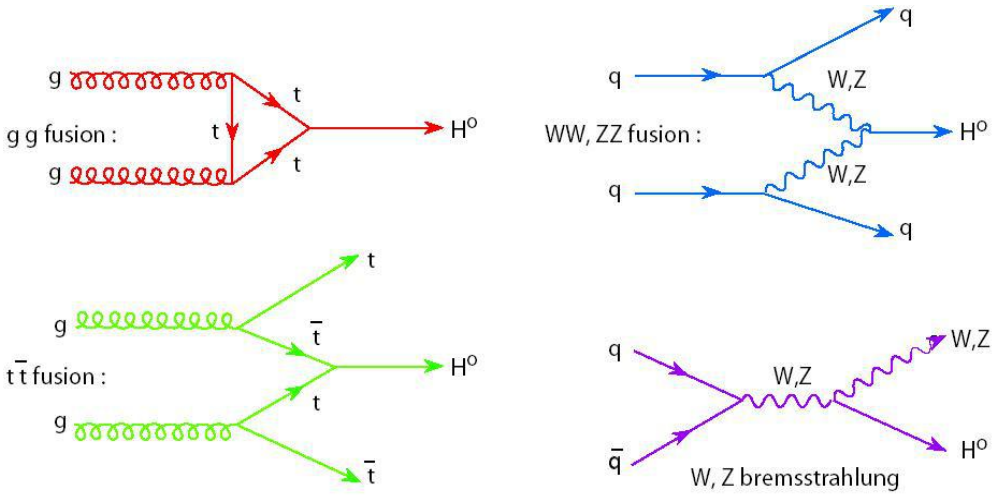

Figure 2.7: Feynman diagrams for the dominant four Higgs production processes at the LHC.

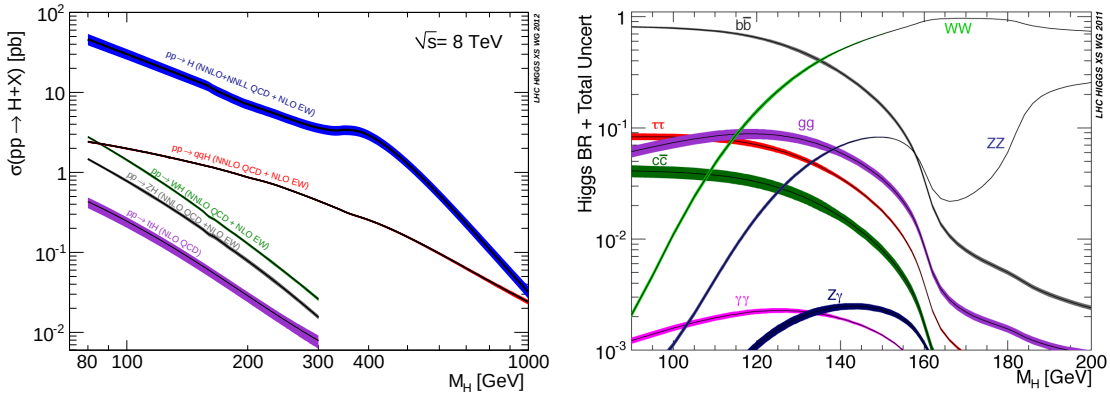

Figure 2.8: Theoretical production cross sections at $\sqrt{s} = 8\,\text{TeV}$ (left) and decay branching fractions (right) of the SM Higgs boson.

The 7 TeV and 8 TeV results for CMS Higgs analyses using 2011 and 2012 data have not yet been finalized. The combined result from Moriond 2013 [23] use up to $5.1\,\text{fb}^{-1}$ at 7 TeV and up to $19.6\,\text{fb}^{-1}$ at 8 TeV proton-proton data from the LHC. The Higgs mass is measured to be $125.7 \pm 0.4\,\text{GeV}$. A few quantum numbers of this newly discovered particle were also measured [24] and found to be consistent with the SM Higgs boson hypothesis.

The most recent public result from the $H \to b\bar{b}$ decay mode is 2.0σ local p-value with $12.2\,\text{fb}^{-1}$ of 8 TeV data. Projections for the entire 8 TeV run, $19.5\,\text{fb}^{-1}$, still leave the associated production $H \to b\bar{b}$ analysis below the traditional threshold for "evidence of", 3σ. It is the hope of this author that addition of all $Z \to \tau\tau$ decay channels with the other $ZH \to \ell\ell b\bar{b}$ decay modes will improve the sensitivity enough to reach the "evidence of" threshold. The topic of this thesis is one such $Z \to \tau\tau$ decay mode:

$ZH \to \tau\tau b\bar{b} \to \nu_\tau e \bar{\nu_e} + \bar{\nu_\tau}\mu\nu_\mu + b\bar{b}$.

Chapter 3

The CMS Experiment at the LHC

CERN is the European Organization for Nuclear Research and one of the world's largest centers for scientific research. It is an international lab that was founded in 1954 and is located near Geneva, Switzerland. CERN provides the infrastructure and scientific instruments such as accelerators to particle physicists around the world. Their newest accelerator, the LHC, is described in the next section.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [25] is a two-ring superconducting accelerator built in the already existing Large Electron-Positron (LEP) Collider tunnel which has a circumference of 26.7 km. The ring is 45 m to 170 m below the surface and straddles the French-Swiss border. It is a proton-proton collider which has been designed to operate at a center-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$. The beam energy and design luminosity were chosen in order to study physics at the TeV scale given the superconducting magnet technology

available at the time of construction. The LHC is versatile as it can accelerate ions as well, extending its physics reach. Four detectors are located at the accelerator's collision points: ATLAS, CMS, LHCb, and ALICE are shown in Figure 3.1 as the four yellow points around the LHC ring.

The LHC consists of 1232 superconducting dipole magnets which are cooled using liquid helium to a temperature of 1.9 Kelvin. These magnets provide a field of 8.3 T which is needed to keep the protons in orbit during acceleration. In order to fit in the old LEP tunnel, the LHC embraced John Blewett's twin-bore magnet design [26]. This has two beam pipes of opposite magnetic field in one cryostat. This ingenious design allows for proton-proton collisions as the like signed protons are circulate in opposite directions in the two beam pipes. Instead of having to go through the complex process of producing anti-protons like the Tevatron, the LHC accelerates protons which are hydrogen atoms with their electrons stripped off by an electric field. Over 3500 higher-order corrector magnets are installed around the LHC accelerator to focus and stabilize the beam.

The protons are injected into the LHC after traveling through the CERN accelerator complex show in Figure 3.1. The accelerator complex has been growing and adapting to serve the needs of particle physics since CERN was founded in the 1950s. Protons are pre-accelerated via this injection chain. Linac-2, a linear accelerator, produces the proton beam and boosts the protons to about 50 MeV. The Proton Synchrotron Booster, labeled "Booster" in Figure 3.1, boosts the protons to 1.4 GeV and combines the protons into a single bunch. The third largest ring in the CERN accelerator complex is the Proton Synchrotron (PS). It can hold up to seven bunches, accelerating the protons to 26 GeV. Protons

travel through the LHC injection chain to the Super Proton Synchrotron (SPS). The protons coming from the SPS have an energy of 450 GeV and are injected into the LHC. The first part of 2010 run was at the SPS injection energy of 450 GeV.

When operated at design energy and luminosity, these protons are then accelerated to an energy of 7 TeV in bunches of 1.15×10^{11} particles per bunch. The ramp time, or the time it takes to accelerated the protons from 450 GeV to 7 TeV is limited by the time it takes to increase the current in the magnets. The beams are accelerated by a radio frequency (RF) acceleration system. At design beam energy, the RF system provides up to 16 MV at 400 MHz, rather the RF system gives 16 MV to each bunch at a frequency of 400 MHz [27]. At the design luminosity of 10^{34} cm^{-2}s^{-1} the LHC will have 2808 bunches per beam.

The unprecedented intensity of the LHC beam presented a unique design challenge. If all the beam energy was dumped in one place, the beam dump would be damaged. Special kicker magnets were designed to "paint" the beam on the graphite core beam dump. These fast-pulsed magnets spread the energy of the LHC beam across a larger surface thus protecting the beam dump from damage. The LHC beam has an abort gap of at least 3 μs to ensure ability to safely dump the beam, so these kicker magnets must be able to ramp up and divert the beam in a short amount of time.

A key parameter of a particle collider is luminosity. As mentioned above the design luminosity of the LHC is 10^{34} cm^{-2}s^{-1}. The luminosity is given by:

$$L = \frac{n_b n_p^2 f \gamma}{4\pi \epsilon_n \beta^*} \times F \tag{3.1}$$

where n_b is the number of bunches in the machine, n_p is the number of protons per bunch (This number is squared because it is symmetric between clockwise and counterclockwise

beams), f is the revolution frequency, γ is the relativistic gamma factor, ϵ_n is the normalized beam emittance, F is a geometric factor due to the crossing angle of the beams, and β^* is the β function of the beam at the collision point. The β function parameterizes the spread in position and momentum space of the beam. During 2012, the LHC reached a peak instantaneous luminosity of 7.5×10^{33} cm^{-2}s^{-1} at 50 ns bunch spacing with a center of mass energy of $\sqrt{s} = 8$ TeV. The instantaneous luminosity is measured using detectors located at small angles with respect to the interaction point and the forward hadronic calorimeter. The event rate is given by the luminosity and particles' interaction cross section. Thus, the amount of data delivered by the LHC and recorded by CMS is measured in integrated luminosity or the amount of instantaneous luminosity 3.1 integrated over time LHC provided collisions at that particular instantaneous luminosity. The event rate is measured in CMS by the forward hadronic calorimeter and by pixel cluster counting in the pixel detector.

The design running parameters of the LHC have a collision happening in CMS every 25 ns, corresponding to a bunch crossing frequency of 40 MHz. At design luminosity, an average 23 inelastic interactions per bunch crossing will occur. This phenomenon is known as pileup. High pileup environments demand high granularity in detectors to maintain low occupancy. As described in the next section, CMS had to be designed to be robust enough to handle this high collision rate environment.

3.2 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) [26,29] is one of two general purpose detectors at the LHC. It is designed to investigate a wide range of physics, including the search for

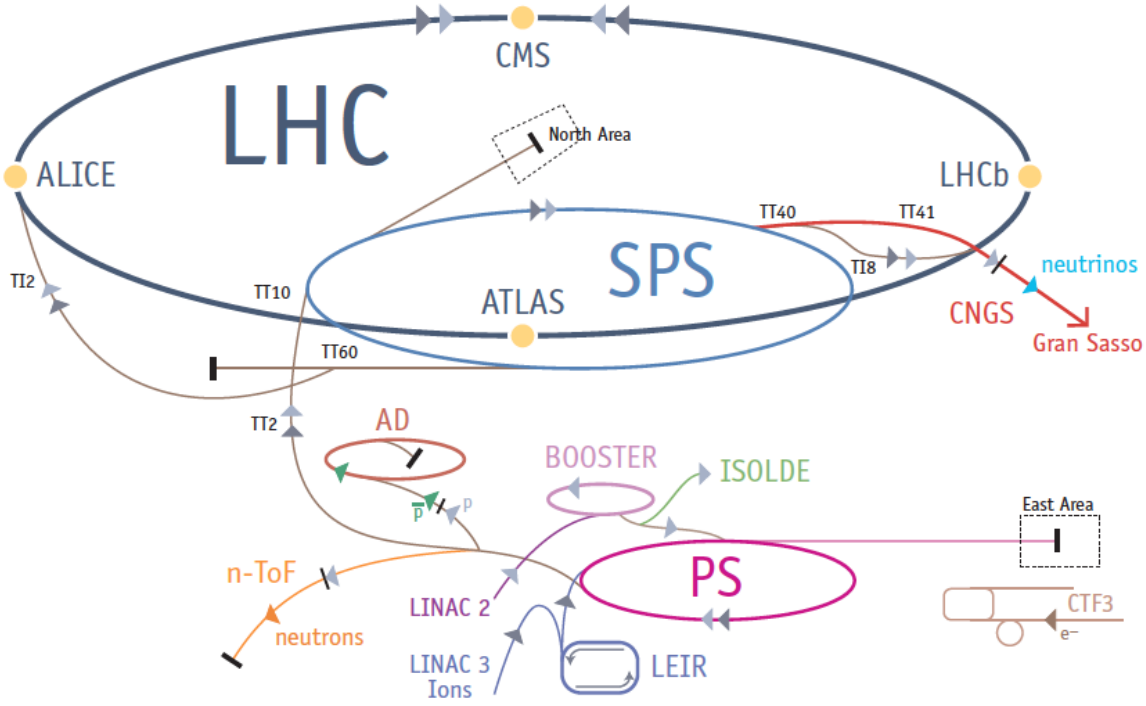

Figure 3.1: A schematic view of the CERN accelerator complex and the four main experiments: ATLAS, CMS, LHCb, and ALICE [28].

the Higgs boson. Weighing in at over 12,500 tons and measuring 15 meters tall, CMS is one of the largest scientific instruments in the world. CMS was designed to operate at a peak instantaneous luminosity of $L = 10^{34}$ cm^{-2}s^{-1} during proton proton operation. This high luminosity poses a great challenge for detector design. High granularity is needed in order to have low occupancy in each channel. The short bunch spacing of 25 ns means that the entire detector must be timed in so one bunch can be distinguished from the next. Multiple inelastic interactions for each bunch crossing requires a detector that can reconstruct events with more than one primary vertex and distinguish these primary vertices from secondary ones. All of these requirements meant that CMS used the frontier of detector technology when designing and building this state of the art detector.

General purpose detectors, like CMS, are divided into many subdetector systems. As one can see in Figure 3.2 various particles interact with each subdetector differently. Each kind of particle leaves a unique signature in the detector. Charged particles, like electrons and pions, both leave detectable hits in the silicon tracker. However only charged hadrons will leave a shower in the hadron calorimeter. Neutral particles travel in straight lines, but charged particles curve due to the magnetic field. The solenoid not only helps to identify charge, but the curvature of charged particles tracks in the muon and tracking system indicate the particle's momentum. Low momentum particle curve more than high momentum particles. By using the information from all subdetectors in CMS, physicists are able to determine which kind of particle went through the detector.

The detector geometry is cylindrical described by azimuthal angle ϕ and the pseudorapidity $\eta = -\ln(\tan(\theta/2))$, where θ is the polar angle of a particle with respect to the proton beam axis. Pseudorapidity is used because particle production is flat in η. Rapidity transforms linearly for Lorentz boosts along the beam axis, but pseudorapidity is used because it can be measured. Nearest to the beam pipe is the silicon pixel detector surrounded by a silicon strip tracking detector. A crystal electromagnetic calorimeter and hadronic calorimeter are the next detectors as you move outwards from the interaction point. The tracker and calorimeters are inside a superconducting solenoid that provides a high magnetic field in CMS. The muon system is outside of the solenoid. It is installed inside a iron yoke that provides support structure and return for the magnetic system. A longitudinal view of CMS can be seen in Figure 3.3 where one can see the coverage in pseudorapidity of the various subdetectors. These subdetectors are explained in greater detail in the following

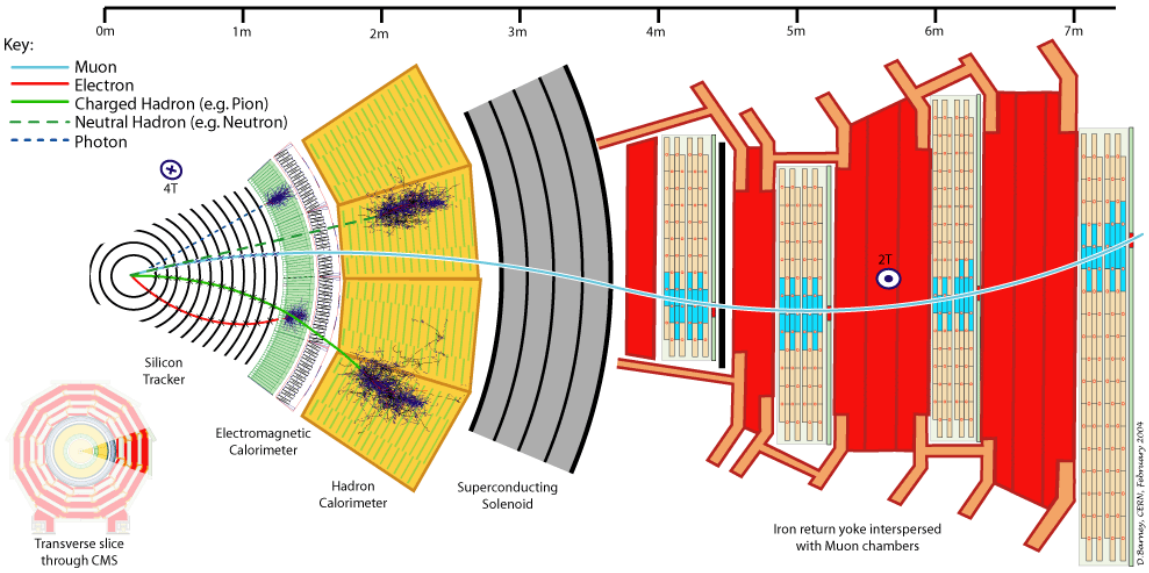

Figure 3.2: Transverse view of CMS detector showing interaction of various particles with different components of the detector [30].

sections, from the interaction point out, as is the tradition in particle physics.

3.2.1 The Tracking System

At the heart of CMS, is an all silicon tracking system consisting of a total active area of 200 m^2 which is enough silicon to tile an Olympic-sized swimming pool. These 75 million readout channels provide precision resolution out to a pseudorapidity of 2.5 for charged particles. An ideal tracker measures the path of charged particle accurately while disturbing the particles as little as possible. The tracking system consists of two major parts: the pixel detector which lies closest to the interaction point and the silicon strip tracker. Both of these detectors are silicon detectors with sensors places under a reverse bias. When a charged particle transverses the material it gives enough energy to an electron in the silicon to eject it from the silicon atom creating electron-hole pairs. These electron-

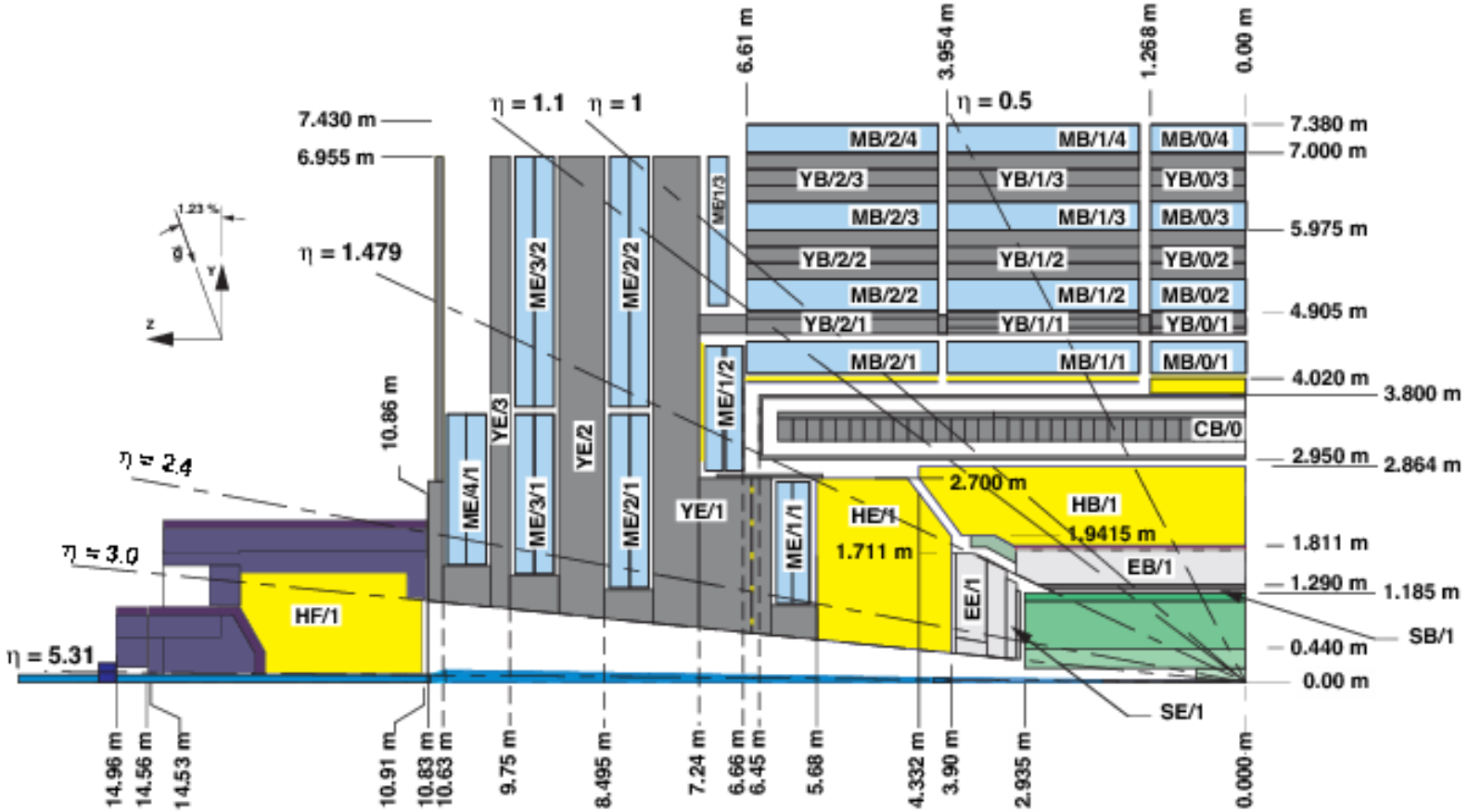

Figure 3.3: Longitudinal view of CMS detector showing coverage in pseudorapidity (η) of the various subdetectors. One should note that ME/4/2 was not included for the 2011-2012 run, but is being installed during long shutdown 1 [31].

hole pairs travel to the surface of the wafer and are collected to create a hit signal on the silicon sensor.

The pixel detector consists of 65.9 million pixels with a cell size of 100×150 μm^2 that provide a single hit resolution of 10 μm in r-ϕ. The barrel pixel has three layers at radii of 4.4 cm, 7.3 cm and 10.2 cm [32]. These modules provide three high resolution space points out to a pseudorapidity of 2.4. The pixel detector also has two forward/backward disks at 34.5 cm and 46.6 cm on each side of the interaction point that span from a radius of 6 to 15 cm. The high segmentation of the pixel detector provides high quality seeds for the track reconstruction algorithm offline and is crucial for tau identification and b-tagging.

The silicon tracker is made up of 9.3 million silicon strips which surround the pixel detector. The barrel is divided into two parts: the tracker inner barrel (TIB) and the tracker outer barrel (TOB) are shown in Figure 3.4. The tracker inner barrel has four layers of 320 μm thick silicon strips [33]. It has two double sided layers which provide stereo hits. The stereo sensors measure hits in Rz as well as r-ϕ. The TIB provides r-ϕ measurements with a resolution of $23 - 35$ μm. The track outer barrel has 6 layers of 500 μm thick silicon strips. The thickness of the silicon wafers is increased to maintain signal to noise ratio. It also has two double sided layers and provides a resolution of $10 - 40$ μm in r-ϕ.

The tracker disks also have thinner inner sensors and thicker outer sensors. The tracker inner disk provides a resolution of $23 - 35$ μm in r-ϕ. The Tracker End-Caps (TEC) are composed of 9 disks containing seven rings each. Each wedge is 22.5°. Rings 1, 2, and 5 are double sided. These double sided rings have one sensor tilted 100 milli-radians with respect to the other sensors to give a ϕ coordinate in the tracker.

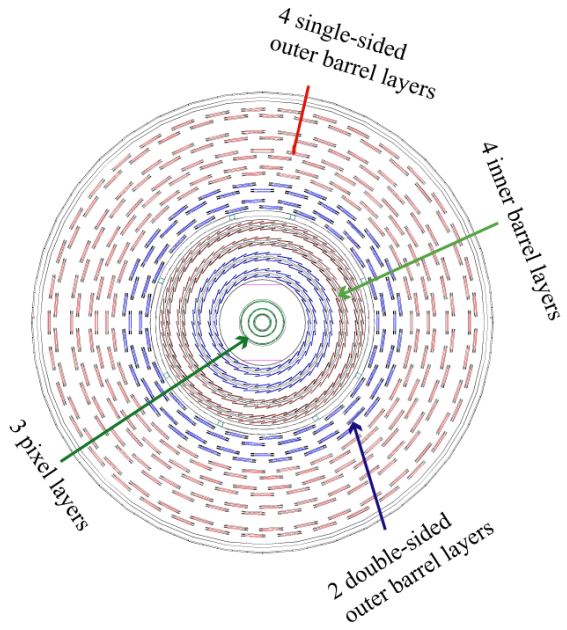

Figure 3.4: A diagram of the tracker barrel showing inner and outer layers. The three layers of the pixel barrel can also be seen [34].

The running environment of the LHC is a neutron rich environment, especially close to the interaction point. It is important, therefore, that the pixel and track electronics are radiation hard. It is expected that the tracker electronics will receive a lifetime dose of 100 kGy. Ten grays is fatal for humans. To mitigate the radiation damage effects and prolong the lifetime of the detector modules, the tracking system is designed to run at subzero temperatures. The pixel readout chip, PSI46v2 [32], is radiation hard and performs well in the LHC environment up to a pileup of about 50. The silicon strip tracker uses the APV25 ASIC which utilizes radiation hard 0.25 μm CMOS [35] technology.

The analysis described in detail in Chapter 4 uses electrons and muons down to 10 GeV and b-tagging to identify b-jets. These objects rely heavily on the tracking system. Excellent impact parameter resolution is needed for correct b-jet identification.

The transverse impact parameter resolution, essential for identifying displaced vertices from b-jets, is shown in Figure 3.5. Additionally, the tracking system must efficiently reconstruct soft tracks with $p_T < 1\,\text{GeV}$ which are used to measure the luminosity and optimize the jet energy resolution using particle flow.

3.2.2 Electromagnetic calorimeter

The CMS electromagnetic calorimeter [37] (ECAL) is made up of about 80,000 lead tungstate ($PbWO_4$) crystals. This crystal ECAL measures the energies of electron and photons with very high precision via scintillation. High energy electrons and photons create electromagnetic showers when they come close to the heavy lead tungstate nuclei. Photons shower via pair production. Electron radiate photons in heavy material. This braking radiation is called Bremsstrahlung. Incident showering particles excite valence electrons in the crystal. The excited valence electron quickly returns to the group state by emitting light which is typically less energetic than the energy absorbed from the incident particle by this mechanism electromagnetic showers are produced in the crystals. The amount of light produced is proportional to the energy of the incident particle. Lead tungstate was chosen because it has a high density thus can contain electromagnetic showers up to 500 GeV [38]. This allowed for a CMS design where both the ECAL and HCAL could fit inside the solenoid. The novel material is heavier than stainless steel but transparent so photons can be collected by the photo detectors on the back-end of the crystals. Lead tungstate scintillates when electrons and photons pass through it. The single segment design of the ECAL allows for straight forward reconstruction and high granularity. The small Moliere radius of lead tungstate reduces the effect of pileup contributions to the energy measurement by reducing

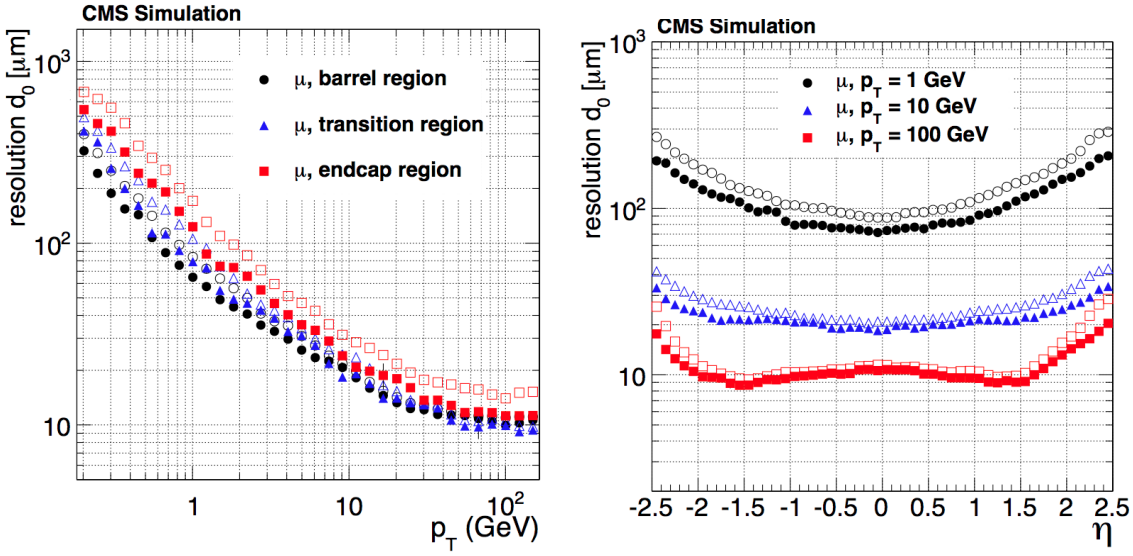

Figure 3.5: Resolution as a function of p_T (left) and pseudorapidity (right) of single isolated muons for the transverse impact parameter, d_0. The solid symbols correspond to the width of the gaussian fitted to the residuals distribution. The open symbols represent the RMS of the residuals distribution. The residual is the difference between the transverse impact parameter of the reconstructed track and the transverse impact parameter of the simulated particle [36].

the area over which the energy is summed. The Moliere radius characterizes the transverse direction of the electromagnetic shower. When CMS was designed and constructed it was not known what the mass of the Higgs boson was. As mentioned in section 2.3, electroweak fits favor a low mass Higgs boson. Therefore, CMS was designed with an ECAL that was well suited to detect $H \to \gamma\gamma$.

The barrel crystals are 1.29 m from the beam pipe. The crystals are tapered to generate a quasi-pointing geometry. There is a gap in ECAL coverage at a pseudorapidity of about 1.5. The endcap crystals also point towards the interaction point. The nominal resolution of the ECAL is $\sigma/E \sim$ 2-5%$/\sqrt{E}$, but the actual performance depends highly on shower containment and pileup. The energy resolution of the ECAL can be parameterized by the following expression:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{3.2}$$

where S is the stochastic term, N is the noise term and C is a constant. These parameters were determined by a electron test beam to be $S = 2.8\%$, $N = 0.12$, and $C = 0.3\%$ [29].

The barrel ECAL is consists of 36 super modules. Each super module contains 1,700 crystals. The barrel ECAL crystals are 0.0174×0.0174 in $\Delta\eta \times \Delta\phi$. This corresponds to a front face of 22×22 mm^2. These crystals are 25.8 radiation lengths thick, where a radiation length is the probability for photons and electrons to lose energy as they transverse a specified material. This corresponds to a lead tungstate crystal length of 23 cm in the barrel region. Photodetectors specially designed to work within CMS's high magnetic field called avalanche photodiodes (APD) are used to convert the light produced via scintillation in the crystals into an electronic signal.

The endcap ECAL is made up of 14,648 crystals. These crystals are arranged in 5×5 super-crystals. Crystals in the endcap ECAL have a front face of 28.62×28.63 mm^2 and a rear face of 30×30 mm^2. This corresponds to 0.0175×0.0175 to 0.05×0.05 in $\Delta\eta \times \Delta\phi$. They are 24.7 radiation lengths thick which corresponds to a crystal length of 22 cm. They are read out by vacuum photo-triode (VPT) photodetectors. The faceplates of the CMS VPTs are made of radiation hard, UV transmitting borosilicate glass [39]. This faceplate glass was shown to have a total transmission loss of less than 10% from the gamma radiation doses expected in the ECAL endcap after 10 years running in CMS.

Preshower

To aid in distinguishing photons from $\pi^0 \to \gamma\gamma$, a preshower detector has been installed in front of the endcap ECAL. Single high energy photons may indicate a photon from $H \to \gamma\gamma$. On the other hand $\pi^0 \to \gamma\gamma$ is typically a signature of less interesting physics. Photons from a π^0 are usually close to one another and can be identified by the preshower. The preshower has a higher granularity that the endcap ECAL, thus can distinguish the two photons from the pion. The preshower has a depth of 3 radiation lengths and a resolution of 300 μm at 50 GeV. As pictured in Figure 3.6, the preshower detects π^0 mesons showering in the lead layers. The showering charged particles are detected in the silicon detector tiles that have pitch of 1.9 mm. The preshower has an effective π^0 rejection of $60 - 40\%$ depending on the transverse energy of the incident π^0.

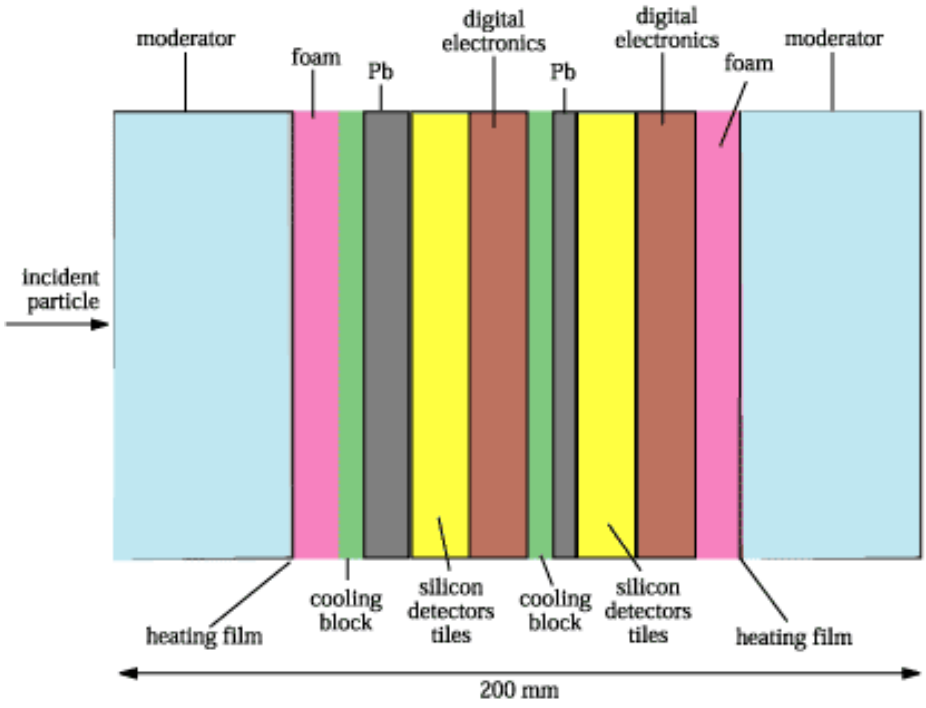

Figure 3.6: A diagram of the preshower showing the layers of the active material.

3.2.3 Hadron calorimeter

Surrounding the ECAL is the Hadronic Calorimeter (HCAL). This calorimeter measures the hadronic activity in the event such as protons and pions and is instrumental in calculating the missing transverse energy (MET). The HCAL is hermetic with staggered layers, so no energy is lost through cracks in the detector. An imbalance in transverse energy is interpreted as missing transverse energy (MET), a key measurement for the analysis presented in chapter 4. Mesons and baryons produce hadronic showers via nuclear interactions. The energy is harder to measure than in the case of electromagnetic showers because the complexity of hadronic showers and their nonlinear response in the HCAL. The quality of the energy measurements depends on the fraction of the hadronic shower detected in

the calorimeter. Thus, the thickness of the material in the HCAL layers has to be large enough to absorb a large fraction of the hadronic energy, and is supplemented by an outer hadron calorimeter (HO) or tail catcher, placed outside the solenoid in order to provide additional absorbing material and sufficient shower containment. The entire HCAL system, HB, HCAL endcap (HE), HO, and Forward HCAL (HF), can be seen in Figure 3.3.

The HB and HE are sampling calorimeters made of alternating layers of absorbing plates and plastic scintillator sheets pictured in Figure 3.7. These scintillators work on the same physical principles as the ECAL crystals, but on the assumption only hadronic energy is left to be measured after the ECAL. Brass was chosen for the material in the absorbing plates as is it non-magnetic, radiation hard, and has a short interaction length ($\lambda_I = 16.42$ cm) [40]. To quantify the interaction characteristic for hadron showers we use interaction length, λ instead of radiation length in the electromagnetic case. The HB is made of 17 sampling layers of 50 mm thick brass absorbing plates and 4 mm thick plastic scintillator tiles. 70000 and 20916 tiles are installed in the HB and HE, respectively. The HB and HE have a very fine granularity of 0.0875×0.0875 radians in $\Delta\eta \times \Delta\phi$. The HB consists of two 4.3 m barrels that fit inside the CMS solenoid. Its energy resolution is $\sigma_E/E \sim 100\%/\sqrt{E}$.

The HF is also a sampling calorimeter, but the material used in this detector has to withstand a much more intense radiation environment. Most of the 23 minimum bias interactions happening every beam crossing at the LHC continue to travel down the beam pipe. The HF will experience unprecedented particle fluxes with an average energy of 760 GeV deposited per proton-proton interaction at $\sqrt{s} = 14$ TeV. Therefore, the HF is made from steel absorber plates with inserted quartz fibers as the active medium that can

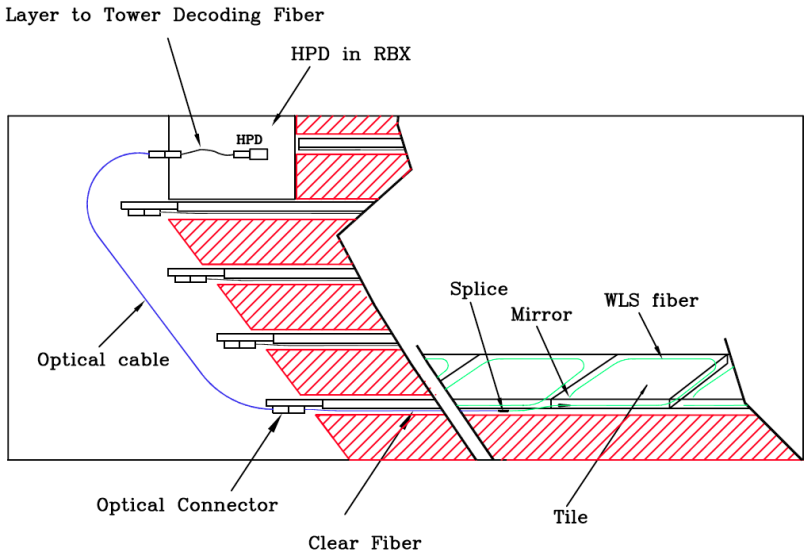

Figure 3.7: A diagram of the HCAL scintillator. The brown shaded region is brass absorbing plates. The plastic scintillator (green) is read out by wavelength shifting (WLS) fibers that shirt the blue-violet light emitted by the scintillator to green light which is then sent through transparent fibers to hybrid photodetectors (HPDs).

quickly ($\sim$ 10 ns) collect the Cherenkov radiation. These fibers are 300 μm in diameter. The plates are 0.175×0.175 radians in $\Delta\eta \times \Delta\phi$. The HF gives very far forward coverage, out to a pseudorapidity of 5. These far forward jets are very important for some analyses like the search for the Higgs boson produced in association with a single top quark as identification of a far forward jet helps to distinguish signal from background. The very far forward detectors can be lowered and put in garages in the CMS experimental cavern to allow for the CMS experiment to open for maintenance and reduces the radiation risk for personnel working in the cavern.

3.2.4 Magnet

The CMS magnet is a Helium-cooled superconducting solenoid that provides a 3.8 T magnetic field for the trackers and calorimeters. The high magnetic field provides enough bending power to precisely measure the momentum of charged particles in the tracking system. This magnet stores 2.7 giga-joules of energy and has a magnetic radial pressure of 64 atmospheres. The magnet is 13 m long and has an inner diameter of 6 m making it the largest superconducting magnet ever built. There is a 2 T field in the 10,000 ton iron return yoke, that also functions as the support structure of the CMS detector. A map of the magnetic field can be seen in Figure 3.8.

The first operation of the CMS magnet took place in the surface experimental hall in a test called the Magnetic Test Cosmic Challenge (MTCC) [41] during the summer and autumn of 2006. This was the first opportunity to map the magnetic field and test integration of many CMS systems in a cosmic run. The test was an invaluable exercise for many subdetectors, but was mainly to train and map the magnet.

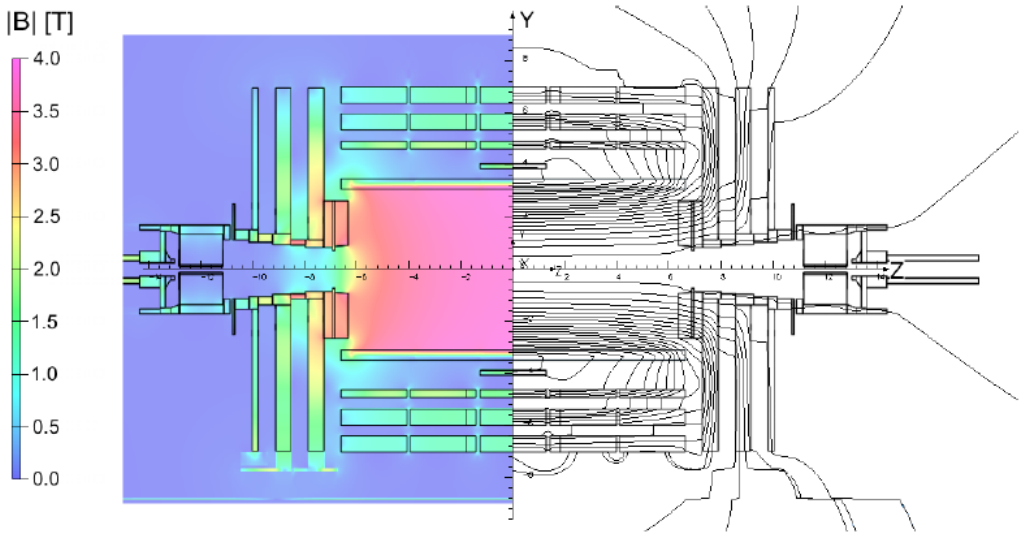

Figure 3.8: The field map of the CMS magnetic field. The magnetic field —B— is on the left and the field lines predicted on the right. Each field line represents a magnetic flux change of 6 weber [42].

3.2.5 Muon System

Interleaved in the iron return yoke of the magnet system is the muon system [43] of CMS. It employs three types of detectors: Drift Tubes (DT), Resistive Parallel Plate Chambers (RPC), and Cathode Strip Chambers (CSC) shown in Figure 3.11. The muon system provides muon identification out to a pseudorapidity of 2.4 and correct charge assignment for muons with momentum up to 7 TeV. The resolution is $\delta p_T/p_T$ = 8-10% for muons with 10 GeV and $\delta p_T/p_T$ = 20-40% for muons with 1 TeV. When the information from the muons system is combined with tracker information the resolution is 1-1.5%.

Muon Barrel

The muon barrel consists of 250 layers of DT chambers. The chambers cover 30° sections around the beam pipe. Five wheels that make up the iron return yoke are interleaved with four layers of DT chambers as pictured in Figure 3.9. Therefore most muons traveling from the interaction point within a pseudorapidity of $|\eta| < 1.2$ leave hits in four DT chambers.

When charged particles pass through a DT, they ionize the gas in the detector. This is why it is imperative that all other charged particles, like $\pi^{\pm}$ are absorbed by the HCAL. DTs are put under a bias so that the electrons knocked off the atoms in the gas drift through the chamber and collect on a positively charged wire. The drift velocity is 57 μm/ns. The gas mixture is 85% Ar and 15% CO_2 [43]. The 40 mm by 11 mm tubes are placed side by side in layers. Chambers consist of 8 ϕ layers and 4 η layers. The tubes are relatively thick, 1 mm, to decouple adjacent tubes in the presence of electromagnetic debris accompanying the muon itself. This detector technology can only be used in the barrel region of CMS where the magnetic field is small and the muon and neutron rates are small because the long tubes must have low occupancy.

Cathode Strip Chambers

The CSC [44] constitute the primary muon tracking device in the CMS endcaps. The endcap muon system started taking data in 2010 with 468 chambers which corresponds to a $6,000$ m^2 active area. The rings of the chambers are designated ME$\pm$S/R, where "ME" is the abbreviation for muon endcap, S is the station or disk, and R is the ring number.

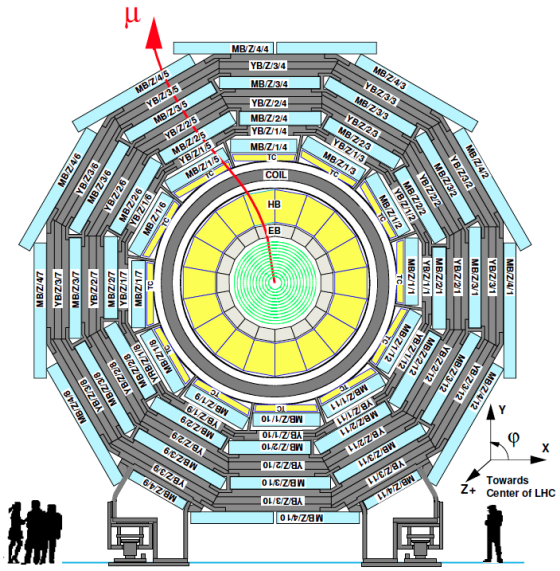

Figure 3.9: Layout of the Muon barrel chambers in one of 5 wheels. The change of curvature of a muon traveling through the CMS can be seen as it travels out of the solenoid. MB/Z/L/S indicates muon barrel chambers on barrel yoke Z, layer 1 through 4 and sector S [29].

The endcap muon system consists of 4 stations of chambers on 3 yoke endcap disks on each side of CMS. The first station has three rings installed on endcap yoke 1, closest to the center of CMS. Stations 2, 3, and 4 are designed to have two rings. The outer ring on the fourth station, was initially eliminated due to budget constraints, is being installed during long shutdown 1. Five spare ME4/2 chambers were installed during the winter shutdown in 2010 on the plus side to measure particle flux rates before the endcap yoke disk 3 shielding was installed during long shutdown 1. CSC technology was chosen for the endcap because higher flux rate of particles is expected than in the barrel region. CSCs provide fast detector response, higher granularity and better resistance to radiation. A muon traveling through the CSC system with p_T of 1 TeV will have a resolution of $\delta p_T/p_T = 5\%$.

Each trapezoidal CSC chamber has 6 measuring planes with a resolution of about

200 μm per layer. CSCs consist of cathode etched circuit board copper strips and anode wires spaced 3.2 mm apart. The strips, like the chambers themselves, are trapezoidal in shape. To compensate for the Lorentz effect of the magnetic field the ME $\pm$ 1/1 anode wires are angled 25^o with respect to the axis perpendicular to the central strip axis. Each layer has a 9.5 mm gas gap filled with 40% Ar, 50% CO_2, 10%!CF_4 [43] gas mixture. Figure 3.10 shows how the charges are deposited in the layers of a CSC chamber. With a 10 ns drift time, the time resolution is 4 ns and track resolution is 100 μm.

CSC chambers vary by size. The outer ring chambers in stations 2, 3, and 4 are longer and cover 10° segments. The inner ring is comprised of shorter chambers that cover 20° segments. The chambers are staggered in layers with a small overlap to ensure hermetic coverage. The strip run parallel to the long side of the chamber, with the exception of ME $\pm$ 1/1 where there strips run parallel to the short side of the trapezoid to reduce occupancy. Selected chamber specifications are listed in Table 3.1. One can see the smaller strip width in ME $\pm$ 1/1a to reduce the occupancy of these chambers. The wires are read out in groups which vary in width from 1.5 to 5 cm for different chambers.

The CSC electronics are split into on-chamber electronics, peripheral crate electronics boards which are in the experimental cavern, and the CSC track finder (TF) which is in the counting room. The counting room is underground next to the CMS experimental cavern. On-chamber electronics include cathode front end boards, anode front end boards and an anode local charge track board. The peripheral crate electronics include trigger mother boards, DAQ mother boards, a clock and control board and a muon port card. These crates are on the periphery of the return yoke. These boards can be easily replaced

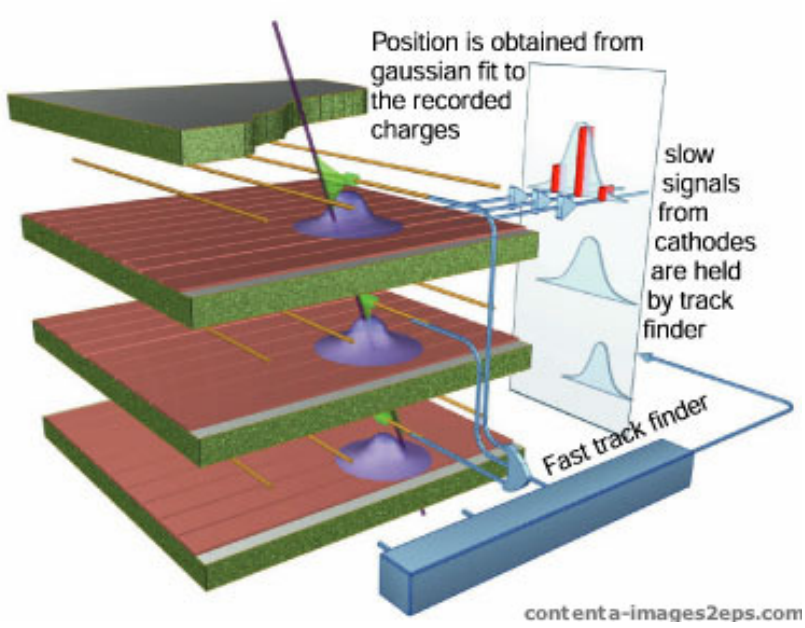

Figure 3.10: Four cathode strip and three anode wire layers. An incident muon will create charged particles in the gas mixture that drift to the anode wires and charge deposits in the strips.

Table 3.1: Selected physical specifications of the CSCs. The range of strip width is give, and the average width is in square brackets [43].

Ring	Wires/Layer	Wires/Group	Strips/Layer	Strip width (mm)	Pitch (mrad)
$ME \pm 1/1a$	600	$11-12$	2×64	$4.11-5.82[4.96]$	3.88
$ME \pm 1/1b$	600	$11-12$	2×64	$4.44-7.6[6.0]$	2.96
$ME \pm 1/2$	528	11	80	$6.6-10.4[8.5]$	2.33
$ME \pm 1/3$	560	12	64	$11.1-14.9[13.0]$	2.16
$ME \pm 2/1$	620	$5-6$	80	$6.8-15.6[11.2]$	4.65
$ME \pm 2/2$	1028	16	80	$8.5-16.0[12.1]$	2.33
$ME \pm 3/1$	550	$5-6$	80	$7.8-15.6[11.7]$	4.65
$ME \pm 3/2$	1028	16	80	$8.5-16.0[12.1]$	2.33
$ME \pm 4/1$	492	5	80	$8.66-15.6[12.1]$	4.65

during beam access time during LHC runs even when the CMS magnet is on. The peripheral crate makes first level trigger decisions, decides if there are local muon hits, before passing information via optical links to the counting room.

Resistive Plate Chambers

The triggering for the muon barrel system is largely carried out by the RPCs. These simple but fast chambers are used to tag hits in DT and CSC muons chambers as coinciding with beam crossings. In Figure 3.11, one can clearly see the RPCs have 6 stations in the barrel and 4 in the endcap region.

Their resolution $\Delta\eta \times \Delta\phi$ of $20 - 100$ cm x $1 - 4$ cm is larger, but their time resolution of 1-2 ns is crucial to matching hits in CSC and DT chambers to the correct beam crossing. They are filled with a gas mixture of 95.5% $C_2H_2F_4$ and 4.5% iso$-C_4H_{10}$. Muons passing through RPC chambers ionize the gas which is between aluminum foil placed at high voltage and a plastic resistive plate. The electrons deposited on the resistive plate are detected by strips etched into the plate.

3.2.6 Trigger and Data Acquisition

It would be an impossible feat to readout the millions of channels in the CMS detector at every bunch crossing provided by the LHC. Because only a small fraction of events can be recorded, events must be quickly searched for signatures that match the goals of CMS's physics program. Therefore, complex trigger systems have evolved to select the relative handful of potentially interesting events from the flood of all collisions. A quick decision must be made for each event. Rejected events are lost forever, so the task of

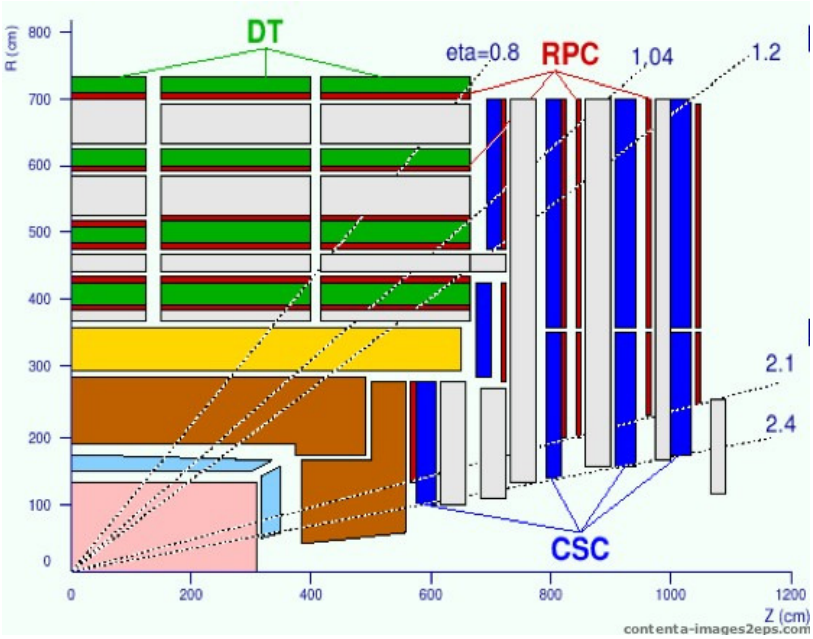

Figure 3.11: RPCs are not on both sides of all DT and CSC chambers, but are placed in the muon system to provide muon triggers.

the trigger is to keep the most important events given limitations in event processing and storage. The general purpose CMS detector has a very extensive physics agenda. Even though most of the proton-proton interactions at the LHC contain well understood physics that is not part of the CMS physics program, keeping only the potentially interesting events means yields of unprecedented scale in High Energy Physics experiments. At design luminosity, the input rate is on the order of a billion events per second. This input must be reduced by a factor of 10^7 to 100 Hz. The trigger must ensure a high data recording efficiency for a wide variety of physics topologies while reducing the online event rate to about 100 Hz.

The CMS trigger [45] is a two level trigger, instead of employing a middle step Level 2 trigger. This is a novel trigger approach for a modern physics experiment. This

was made possible by strides in processing power and network speed technology. The first step in reducing the number of events recorded by the CMS detector is called the Level 1 (L1) which reduces the event rate to a maximum 100 kHz. The next step, the High Level Trigger (HLT), takes the events from the L1 and is designed to reduce the output rate to about 100 Hz. CMS is able to send events to permanent storage at the rate of 100 Hz. The L1 system is based on custom electronics and keeps data for a mere 3.2 μs. The HLT uses an extensive farm of commercial processors that use sophisticated software algorithms to quickly select events with interesting physics.

The L1 rate is limited by the speed of the detector readout electronics and the rate at which the data can be passed to the Data Acquisition (DAQ) system. The original design of the L1 was 75 kHz, but advances in computing technology allow for a 100 kHz online rate. During the 2011-2012 run the time between beam crossings was 50 ns. However the time between beam crossings at LHC design luminosity is 25 ns. Thus the L1 has been designed with a 25 ns beam crossing frequency in mind. The design collision rate of 25 ns is too short a time to read out the approximately 700 kB of data for each raw event and provide a trigger decision. Therefore, data held in pipelined memories in the front-end electronics while the L1 decision is made. While the front end electronics can store about 3 μs of data, the time available for actually processing the data is no more than $\sim$ 1 μs after latencies in transfer time are taken into consideration. Speed is essential for the L1 so iterative algorithms like jet-finding and fetching information from database systems is not possible. Instead the L1 uses memory lookup tables to deterministically process events.

The L1 trigger is comprised of three major subsystems: the calorimeter trigger,

the muon trigger and the global trigger pictured in Figure 3.12. If an event passes the L1 trigger requirements it is said to have a L1 accept. If a L1 accept is found the event information along with the trigger information is sent to the DAQ. The trigger timing and control system tells all detector front end and readout subsystems which specific bunch crossings have passed the L1 trigger.

The calorimeter trigger begins with local triggers, called trigger primitives, based on energy deposits in calorimeter trigger towers. The trigger primitive generator reads out this information and passes it via high speed copper links to the Regional Calorimeter Trigger (RCT). The RCT identifies candidate electrons, photons, taus and jets. The global calorimeter trigger calculates the total transverse energy and missing energy of the event given the calorimeter information. The GCT also sorts candidate electrons, photons, taus, and jets by p_T, forwarding the top 4 of each type to the global trigger.

The muon system poses a unique challenge for the L1. The DTs have a drift time of 400 ns, so data in the muon system is not available at the same time as for the calorimeters. Thus information is looked at in the RCT and the Global Muon Trigger (GMT) separately first. The GMT sorts the muon tracks from the RPC, DT and CSC and tries to match RPC information to CSC or DT tracks. It uses information from the GCT to determine if the muons are isolated. The top 4 muons are sent to the global trigger.

The global trigger makes the L1 trigger decision based on information from the GCT and GMT. It synchronizes the calorimeter and muon information through the trigger timing and control system. Using a set of criteria known as the L1 trigger menu, a binary decision to keep the event is made. If there is a L1 accept, the event passes the L1 trigger

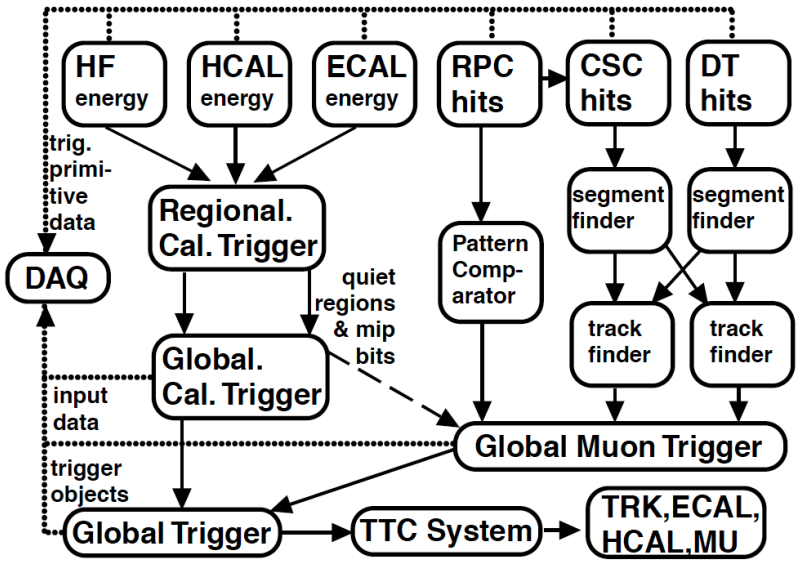

Figure 3.12: Overview of Level 1 Trigger.

criteria, it initiates the readout passing the information to the HLT [46]. Common Mezzanine Cards (CMC), called S-Link64 Sender cards, were custom developed to buffer up to 1.5 kBytes of data before generating back pressure to the Front-End Drivers (FEDs) [29]. The nominal data transfer rate, 100 MBytes/s, is twice the maximum sustained throughput of the DAQ system.

The HLT acquires the detector data, analyzes it, and archives selected events. Tracker information waits in readout buffers while the L1 decision is made. It along with the rest of the detector information is read out by the DAQ after a L1 accept. The DAQ provides the means for connecting 700 front end modules to 1,000 processors via a switching network. Ideally, the HLT should have no built-in design or architecture limitations - other than the total bandwidth and computing power. The HLT selects events based upon algorithms executed in a single processor farm. The fact that the HLT is fully programable makes it

agile in the face of unforeseen backgrounds or changes in physics priorities. As this cutting edge technology becomes industry standard, the HLT is able to use commercially available components resulting in savings of man hours, development, production and maintenance costs.

To minimize computing time needed by the HLT, algorithms that partially reconstruct detector information are key. This means that some events with objects that do not correspond to a L1 object are sacrificed inorder to preserve speed. The processing time per event is 10 μs on average. To save on time algorithms do not loop through the entire reconstruction chain, instead events are discarded as soon as the objects fail a requirement. For example, muons in the muon stream will not run the tracker road finding algorithm if they fail the muon validation requirements. The algorithms employed by the HLT must be almost as sophisticated as those used in offline reconstruction. The HLT is composed of a set of trigger paths, each path is a specific physics object selection. A set of these trigger paths is called a HLT menu. The menu can be changed while running, but is constant for each luminosity section. A luminosity section is defined by CMS as 2^{18} LHC orbits corresponding to 23 seconds. The HLT was designed to transfer events from the CMS detector site at a rate of 100 Hz, therefore sets of trigger paths are created to throttle event rates as the instantaneous luminosity increases. Numerous improvements to the HLT have made rates of $\sim$ 350 Hz and $\sim$ 600 Hz for core and parking data, respectively, possible.

Chapter 4

The Search for the SM Higgs boson in $e\mu b\bar{b}$ final state

4.1 Introduction

The search for $H \to b\bar{b}$ where the Higgs boson is produced in association with a Z boson is presented in this chapter. The presence of a vector boson decaying into charged leptons allows for tight lepton requirements which suppress the QCD background, and allows us to trigger on such events. Originally the search for $H \to b\bar{b}$ was limited to cases where the Higgs and associated boson were highly boosted as proposed in Ref. 47 and searches presented in Ref. 48 and 49, but advances in analysis techniques, b-tagging, and particle reconstruction have allowed us to open the acceptance to all ZH events. The $\tau\tau b\bar{b}$ final state is still largely unexplored at LHC energies. This channel could provide valuable information about the newly discovered particle at 125 GeV.

The search for a SM Higgs boson produced in association with a Z boson, where the Higgs decays into b quarks and the Z decays into taus is presented in this thesis. The case where each tau decays leptonically, one into an electron, the other tau into a muon is the focus of this analysis. In order to maintain orthogonality with $Z(ee)H$ and $Z(\mu\mu)H$ analyses, the same flavor lepton case is not included. The analysis strategy and cut and count analysis were presented in Ref. 50. Changes to the analysis and full results for the search for the SM Higgs boson produced in association with a Z boson in the $e\mu b\bar{b}$ final state with associated neutrinos are presented in this Chapter.

4.2 Analysis Strategy

This analysis is largely based on the boosted $VHbb$ analysis described in Ref. 51. In the $VHbb$ analysis a cut $p_T(\text{jj}) > 100\,\text{GeV}$ is made to limit the Z+jets background. This cuts is referred to as "boosting" because it selects VH events with a high Lorentz boost. However for the $ZH \to \tau\tau b\bar{b} \to e\mu + \text{neutrinos} + b\bar{b}$ channel the overall cross section is low due to the low branching ratio for $\tau\tau \to e\mu$. Taus decay leptonically 17% of the time [52], therefore a ditau decay will decay into $e\mu$ and assiciated neutrinos 6% of the time. Due to this small branching factor, no boosting is used. It was found that boosting, while effective for channels with higher branching ratio, cuts out too many signal events in this analysis.

The signal we are looking for is particularly challenging because we are searching for two b quarks and two taus in the busy LHC environment. We choose the case where both tau leptons decay into leptons. Electrons and muons have high trigger and reconstruction efficiencies in the CMS detector. Requiring two isolated leptons helps to control QCD and

W+jets background but comes at the cost of multiple neutrinos. Our signal events yield four neutrinos, but only one missing momentum vector, $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$, in the plane transverse to the beam can be reconstructed. The magnitude of this missing transverse momentum vector is referred to as missing transverse energy or E_{miss}^{T}. Final state radiation and initial state radiation increases the number of high p_T jets in the event, determining which jets came from the Higgs boson is challenging.

Only events with all 5 objects: $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$, electron, muon, and two jets, reconstructed in the final state are considered. These "$e\mu$ candidate" events are then required to pass the preselection. The preselection requires one lepton to have $p_T > 20\,\mathrm{GeV}$ and the leptons to be of opposite charge and have $\Delta R > 0.3$ where $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$. Jet identification requirements are also applied at the preselection step. The preselection is summarized in Table 4.1 for clarity. The objects listed in Table 4.1 are described in more detail in the following sections.

Backgrounds arise from production of Z bosons in association with one or more jets, pair production of top quarks, and diboson production. QCD, W+jets, and single top production in both the s- and t-channels are also included but are almost negligible.

Table 4.1: The preselection. A dash (-) indicates that no requirement is applied. In addition to the cuts listed in the table below jet identification is applied at the preselection step. Also the electron and muon are required to be of opposite charge and have $\Delta R > 0.3$.

	Minimum Number	p_T(GeV)	$\lvert\eta\rvert$
Electron	1	10	2.5
Muon	1	10	2.4
Jets	2	20	2.5
$\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$	1	0	-

Control regions for Z+light flavor jets ($udscg$) and $t\bar{t}$ are defined. Data is used to estimate the expected yields of these two dominant backgrounds in the signal region.

This analysis was optimized assuming a Higgs mass of 125 GeV using the full 8 TeV 2012 data sample. The analysis methods such as cuts, control region definitions, and variables, are then applied to the 7 TeV 2011 data and MC samples. A fit of the shape of the resulting boosted decision tree (BDT) [53] output for the background only and background plus signal hypotheses is performed to achieve the lowest possible limits of the 95% confidence level upper limit on the SM Higgs boson cross section.

4.3 Data and MC Samples

The samples used in this analysis are described in this section. First the 2011 data and MC is described in detail, then the 2012 data and MC is listed. Both 2012 and 2011 have similar contributions for backgrounds in simulation. However differences in calibration, pileup scenarios and center of mass energy merit separate simulation to model the two data samples. Simulation samples are labeled by when they were produced. All Fall11 samples were produced with the same calibration, pileup, and center of mass energy, with the bulk

Table 4.2: List of 2011 data samples used for this analysis. The sum includes approximately 5.0 fb^{-1}. An uncertainty of 2.2% is assigned for the 2011 integrated luminosity.

Dataset	$\mathcal{L}$ (fb^{-1})
/MuEG/Run2011_May10ReReco	0.22
/MuEG/Run2011A_Aug05ReReco	0.40
/MuEG/Run2011A_PromptRecoV4	0.96
/MuEG/Run2011A_PromptRecoV6	0.71
/MuEG/Run2011B_PromptRecoV1	2.7
Total Lumi	5.0

Table 4.3: List of signal Summer11 Monte Carlo samples used in this analysis.

Mode	Dataset	$\mathcal{L}\,(\mathrm{fb}^{-1})$
$Z(\ell\ell)H$	/ZH_ZToLL_HToBB_M-110_7TeV-powheg_herwigpp	31000
$Z(\ell\ell)H$	/ZH_ZToLL_HToBB_M-115_7TeV-powheg_herwigpp	37700
$Z(\ell\ell)H$	/ZH_ZToLL_HToBB_M-120_7TeV-powheg_herwigpp	46700
$Z(\ell\ell)H$	/ZH_ZToLL_HToBB_M-125_7TeV-powheg_herwigpp	59800
$Z(\ell\ell)H$	/ZH_ZToLL_HToBB_M-130_7TeV-powheg_herwigpp	79500
$Z(\ell\ell)H$	/ZH_ZToLL_HToBB_M-135_7TeV-powheg_herwigpp	110000

Table 4.4: List of diboson Fall11 Monte Carlo samples used in this analysis.

Mode	Dataset	$\mathcal{L}\,(\mathrm{fb}^{-1})$
ZZ	/ZZtoAnything_TuneZ2_7TeV-pythia6-tauola	675.912
WW	/WWtoAnything_TuneZ2_7TeV-pythia6-tauola	98.38
WZ	/WZtoAnything_TuneZ2_7TeV-pythia6-tauola	234.35

Table 4.5: List of V+jets Fall11 Monte Carlo samples used in this analysis, with the corresponding integrated luminosity.

Mode	Dataset	$\mathcal{L}\,(\mathrm{fb}^{-1})$
W+jets	/WjetsToLNu_TuneZ2_7TeV-madgraph-tauola	2.598
$Z \to \ell\ell(M_{\ell\ell} > 50)$	/DYJetsToLL_TuneZ2_M-50_7TeV-madgraph-tauola	11.51
$Z \to \ell\ell(p_{\mathrm{T}} > 100\,\mathrm{GeV})$	/DYJetsToLL_pt100_7TeV-madgraph-tauola	39.90

Table 4.6: List of $t\bar{t}$ and single top samples used in this analysis, with the corresponding integrated luminosity.

Mode	Dataset	$\mathcal{L}\,(\mathrm{fb}^{-1})$
$t\bar{t}$	/TTJets_TuneZ2_7TeV-madgraph-tauola/Fall11-PU_S4_START42_V11-v1/AODSIM	22.44
Single Top (tW)	/T_TuneZ2_tW-channel-DR_7TeV-powheg-tauola/Fall11-PU_S6_START42_V14B-v2/AODSIM	103.5
	/Tbar_TuneZ2_tW-channel-DR_7TeV-powheg-tauola/Fall11-PU_S6_START42_V14B-v2/AODSIM	102.9
Single Top (t-ch)	/T_TuneZ2_t-channel_7TeV-powheg-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM	93.0
	/Tbar_TuneZ2_t-channel_7TeV-powheg-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM	85.9
Single Top (s-ch)	/T_TuneZ2_s-channel_7TeV-powheg-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM	81.5
	/Tbar_TuneZ2_s-channel_7TeV-powheg-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM	95.8

of the simulation produced in Fall of 2011.

In signal and background samples including τ-leptons the subsequent decay of the τ-lepton is simulated using the `Tauola` software package [54]. $Z \to \tau\tau$ events are generated with Next-to-Leading-Order (NLO) event generator `Madgraph` [55] and interfaced with Pythia to account for QCD initial and final state radiation. W+jets production is simulated using the `Madgraph` event generator interfaced with `Pythia`. The $t\bar{t}$ background is simulated using `Madgraph` event generator interfaced with `Pythia`. The remaining background samples, including diboson production, were simulated using `Pythia6` [56].

All generated events have been passed through the full simulation of the detector response according to the CMS implementation of `GEANT4` [57]. The samples were analyzed with CMSSW, the software framework used to analyse data in CMS, and processed through the standard physics analysis toolkit (PAT) configuration. CMSSW consists of over a thousand subpackages which have been created to provide an extensive toolkit for physicists to carry out analyses of data while contributing only a small amount of code by themselves. Leptons, jets, and $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ are reconstructed by the particle flow (PF) algorithm [58] which performs a global event reconstruction and provides a full list of particles identified as electrons, muons, photons, and hadrons.

Table 4.2 summarizes the 2011 data samples used and the approximate integrated luminosity. The total integrated luminosity of the analyzed data is $5.0\,\mathrm{fb}^{-1}$. The data are compared to Monte Carlo simulations based on theoretical calculations for signal and background processes. Table 4.3–4.6 summarize the simulated samples and their equivalent luminosities and cross sections. Cross sections are obtained from the standard CMS

twiki [59]. All 2011 simulation and data are analyzed using CMSSW_4_2_X.

The signal samples used are described in Table 4.3. A larger Summer11 sample was used to probe the optimum phase space for this analysis. The optimization was then checked with the smaller Fall11 samples. The Summer11 samples were filtered on MCTruth for $e\mu$ events only. Contamination from other associated production ZH events, where $H \to b\bar{b}$ and the Z boson decays directly into a muon or election pair, was verified to be small, $< 10\%$ after $e\mu$ candidates are selected and $< 5\%$ after BDT selection.

The 8 TeV analysis samples are listed in 4.7– 4.11. These samples were analyzed with CMSSW_5_3_X and processed through the same reconstruction chain as the 7 TeV samples. The runs in range 207883-208307 are vetoed because of a pixel misalignment problem. This problem in the pixels impacts many observables related to b-tagging. This removes approximately 600 pb^{-1} of data and is already removed in the Run2012D luminosity reported in Table 4.7.

Tables 4.12 and 4.13 summarize the cross sections and branching fractions assumed for the signal channel and corresponding mass points. The cross sections are computed at Next-to-Next-to-Leading-Order (NNLO), as described in Ref. 3. Additional Higgs samples and cross sections are listed in Section 4.12. These samples are used to estimate the amount of bleed through Higgs events from other production and decay modes into the $ZH \to \tau\tau b\bar{b}$ analysis.

Table 4.7: List of 2012 data samples used for this analysis. The sum includes approximately 19.0 fb^{-1}. An uncertainty of 5% is assigned for the 2012 integrated luminosity.

Dataset	Run Numbers	$\mathcal{L}$ (fb^{-1})
`/MuEG/Run2012A-13Jul2012-v1`	190456-193621	0.81
`/MuEG/Run2012A-recover-06Aug2012-v1`	190782-190949	0.082
`/MuEG/Run2012B-13Jul2012-v1`	193834-196531	4.4
`/MuEG/Run2012C-24Aug2012-v1`	198049-198522	0.49
`/MuEG/Run2012C-PromptReco-v2`	198941-203002	6.4
`/MuEG/Run2012C-EcalRecover-11Dec2012`	201191	0.13
`/MuEG/Run2012D-PromptReco-v1`	203894-208686	6.7
Total Lumi		19.0

Table 4.8: List of signal Summer12 Monte Carlo samples used in this analysis.

Mode	Dataset	$\mathcal{L}\,(\mathrm{fb}^{-1})$
$Z(\ell\ell)H$	/ZH_ZToLL_HToBB_M-110_8TeV-powheg-herwigpp/Summer12_DR53X-PU_S10_START53_V7A-v1	22607
$Z(\ell\ell)H$	/ZH_ZToLL_HToBB_M-115_8TeV-powheg-herwigpp/Summer12_DR53X-PU_S10_START53_V7A-v1	27109
$Z(\ell\ell)H$	/ZH_ZToLL_HToBB_M-120_8TeV-powheg-herwigpp/Summer12_DR53X-PU_S10_START53_V7A-v1	34085
$Z(\ell\ell)H$	/ZH_ZToLL_HToBB_M-125_8TeV-powheg-herwigpp/Summer12_DR53X-PU_S10_START53_V7A-v1	38959
$Z(\ell\ell)H$	/ZH_ZToLL_HToBB_M-130_8TeV-powheg-herwigpp/Summer12_DR53X-PU_S10_START53_V7A-v1	57818
$Z(\ell\ell)H$	/ZH_ZToLL_HToBB_M-135_8TeV-powheg-herwigpp/Summer12_DR53X-PU_S10_START53_V7A-v1	79713

Table 4.9: List of diboson Summer12 Monte Carlo samples used in this analysis.

Mode	Dataset	$\mathcal{L}\,(\mathrm{fb}^{-1})$
ZZ	/ZZ_TuneZ2star_8TeV_pithier_tool/Summer12_DR53X-PU_S10_START53_V7A-v1	1181
WW	/WW_TuneZ2star_8TeV_pithier_tool/Summer12_DR53X-PU_S10_START53_V7A-v1	176
WZ	/WZ_TuneZ2star_8TeV_pithier_tool/Summer12_DR53X-PU_S10_START53_V7A-v1	295

Table 4.10: List of V+jets Summer12 Monte Carlo samples used in this analysis.

DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball
DYJetsToLL_PtZ-100_TuneZ2star_8TeV-madgraph
DYJetsToLL_PtZ-50To70_TuneZ2star_8TeV-madgraph-tarball
DYJetsToLL_PtZ-70To100_TuneZ2star_8TeV-madgraph-tarball
WJetsToLNu_PtW-100_TuneZ2star_8TeV-madgraph
WJetsToLNu_PtW-70To100_TuneZ2star_8TeV-madgraph
WJetsToLNu_TuneZ2Star_8TeV-madgraph-tarball

Table 4.11: List of $t\bar{t}$ and single top samples used in this analysis, with the assumed cross section.

Mode	Dataset	σ (pb)
$t\bar{t}$	/TT_CT10_TuneZ2star_8TeV-powheg-tauola/Summer12_DR53 X-PU_S10_START53_V7A-v*	234
Single Top (tW)	/Tbar_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/Sum mer12_DR53X-PU_S10_START53_V7A-v1	11.1
	/T_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/Summer 12_DR53X-PU_S10_START53_V7A-v1	11.1
Single Top (t-ch)	/Tbar_t-channel_TuneZ2star_8TeV-powheg-tauola/Summer1 2_DR53X-PU_S10_START53_V7A-v1	30.7
	/T_t-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_ DR53X-PU_S10_START53_V7A-v1	56.4
Single Top (s-ch)	/Tbar_s-channel_TuneZ2star_8TeV-powheg-tauola/Summer1 2_DR53X-PU_S10_START53_V7A-v1	1.76
	/T_s-channel_TuneZ2star_8TeV-powheg-tauola/Summer12_ DR53X-PU_S10_START53_V7A-v1	3.79

Table 4.12: Cross sections and branching fractions for signal events at $\sqrt{s} = 7\,\text{TeV}$ for masses between 110 and 135 GeV.

7 TeV	$Z(\ell\ell)H$		$\text{H} \to \text{b}\bar{\text{b}}$	
M_{H}	σ(pb)	$\pm$ (%)	BR	$\pm$ (%)
110	0.472	+5.3, −5.3	0.745	+2.1, −2.2
115	0.411	+5.5, −5.4	0.704	+2.4, −2.5
120	0.360	+5.0, −4.7	0.648	+2.8, −2.8
125	0.316	+4.9, −5.1	0.577	+3.2, −3.2
130	0.278	+5.2, −5.1	0.493	+3.7, −3.8
135	0.245	+5.3, −5.0	0.403	+4.2, −4.3

Table 4.13: Cross sections and branching fractions for signal events at $\sqrt{s} = 8\,\text{TeV}$ for masses between 110 and 135 GeV.

8 TeV	$Z(\ell\ell)H$		$\text{H} \to \text{b}\bar{\text{b}}$	
M_{H}	σ(pb)	$\pm$ (%)	BR	$\pm$ (%)
110	0.587	+5.4, −5.4	0.745	+2.1, −2.2
115	0.512	+5.6, −5.6	0.704	+2.4, −2.5
120	0.448	+5.0, −4.9	0.648	+2.8, −2.8
125	0.394	+5.1, −5.0	0.577	+3.2, −3.2
130	0.347	+5.4, −5.3	0.493	+3.7, −3.8
135	0.307	+5.4, −5.2	0.403	+4.2, −4.3

4.4 Triggers

This analysis uses the logical OR of HLT_Mu8_Ele17* and HLT_Mu17_Ele8* where * indicates the specific trigger version and calorimeter isolation plus electron identification. These are both cross triggers found int he MuEG primary data sample. Both require a L1 trigger of a high p_T lepton, electron or muon, respectively. The HLT requires the p_T of the lepton is higher than the value listed next to it. For example, given the HLT_Mu8_Ele17* trigger, the muon is required to have a $p_T > 8\,\text{GeV}$ and the electron is required to have a $p_T > 17\,\text{GeV}$. The trigger efficiency of the electron and muon are measured separately as a function of p_T and η. The trigger efficiency of each lepton are the applied to Monte Carlo events as event weights. The trigger efficiency was evaluated separately for each epoch of data taking, but was found to be consistent over the entire 2011 run. It was statistically advantageous to evaluate the trigger efficiency for all 2011 runs combined. 2011 and 2012 trigger efficiencies are measured separately due to different beam conditions in 2011 and 2012.

The efficiency of the upper leg, the lepton with higher p_T, was determined using standard tag-and-probe techniques using $Z \to \ell\ell$ decays. For electrons, the SingleElectron and ElectronHad dataset are compared to the Summer11-DYToEE sample. For muons, reconstructed Z boson mass for visible mass (left) and SVfit mass (left) for events used in the BDT analysis the SingleMuon dataset is compared to the Summer11-DYToMuMu sample. These samples are described in Ref. 51. In the preselection, at least one lepton is required to have $p_T > 20$ GeV, requiring that the leading lepton on the plateau of the upper leg trigger turn on curve.

The efficiency of the trailing leg, or the lepton with lower p_T, is measured from $e\mu$ candidate events that are in the singleMuon and electronHad primary data samples. SingleMuon events are known to have a muon trigger. ElectronHad events are known to have an electron trigger. If singleMuon events also have a reconstructed electron then a HLT_Mu17_Ele8* trigger flag is looked for. Similarly, if an electronHad event has a reconstructed muon then a HLT_Mu8_Ele17* trigger flag is looked for. In this way, we measure the lower leg trigger efficiency independent of the upper leg and lepton reconstruction response. As seen in Figure 4.1, the measured trigger efficiency is found to be close to one. The zero of the y-axis of Figure 4.1 is highly suppressed, magnifying fluctuations in the efficiency. These fluctuations are covered by the trigger systematics of 1% for muons and 2% for electrons.

The trigger paths and thresholds for the 8 TeV data remained the same as 7 TeV. However the efficiency was remeasured to account for the different pileup conditions, or the number of interactions per bunch crossing was different in 2011 and 2012. Standard tag-and-probe techniques were used for both upper and lower legs of the trigger. The results are consistent with 7 TeV measurements. The trigger efficiencies are applied to Monte Carlo events as event weights. The trigger efficiency was evaluated separately for each epoch of data taking. A weighted average is used as the trigger efficiencies changed due to higher and higher pileup conditions as the 2012 run progressed. The turn on curves of the lower leg trigger efficiency for 8 TeV data is show in Figure 4.2.

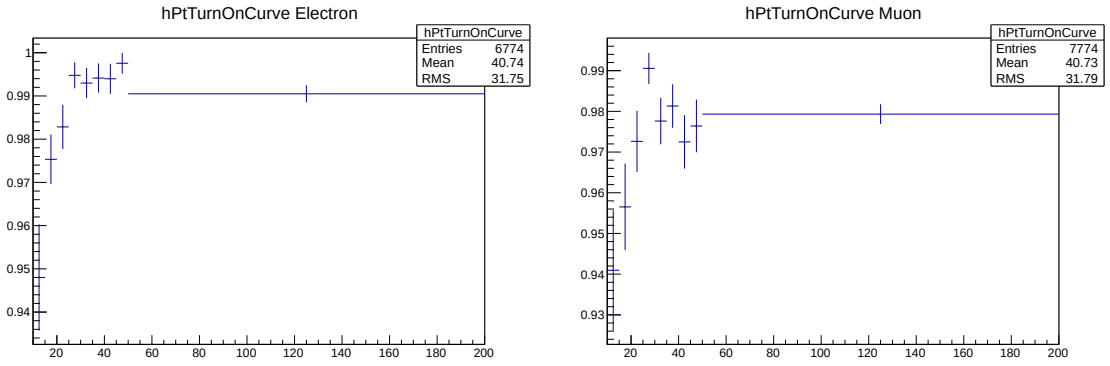

Figure 4.1: Turn on curve for Electron lower leg (left) and Muon lower leg (right) of cross triggers for 2011 data.

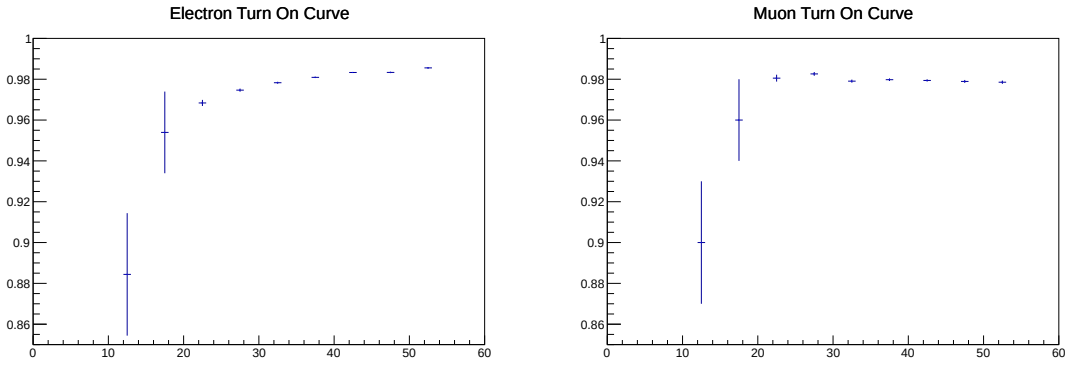

Figure 4.2: Turn on curve for Electron lower leg (left) and Muon lower leg (right) of cross triggers for 2012 data.

4.5 Physics Objects

This analysis uses standard physics objects from the PAT framework that are approved by the respective POGs. The first step of data processing from `AODSIM` to PATuples is shared in common with the main $VHbb$ analysis [48] . This section describes the reconstruction, identification, and selection of electrons, muons, jets, b-jets, and missing transverse energy. To better model the effects of pileup on these physics objects, we use methods described in Section 4.5.1.

4.5.1 Primary vertex selection and pileup treatment

Primary vertices are reconstructed using the Deterministic Annealing [60] clustering algorithm. Their distance to the nominal interaction point must be smaller than 24 cm in z and 2 cm in the transverse plane. The number of degrees of freedom of the vertex fit is required to be greater than 4. The selected vertices are then sorted by p_T. The vertex with the largest summed squared p_T of tracks associated with it is chosen as the event vertex.

The average number of primary vertices was 7.6 ± 3.7 for the 2011 run for events with $e\mu$ candidates. In order to represent the number of pileup events as observed in data the MC events are re-weighted. In Figure 4.3 the number of primary vertices is shown. The Summer11 Monte Carlo samples were generated with an assumed pileup distribution and were reweighted as a function of the simulated PU events as described in Ref. 61 with different weighting applied for 2011A and 2011B. The Fall11 samples are reweighted by the same method, but their simulated PU distribution is closer to data. Any remaining discrepancies are taken into account in the systematic uncertainties.

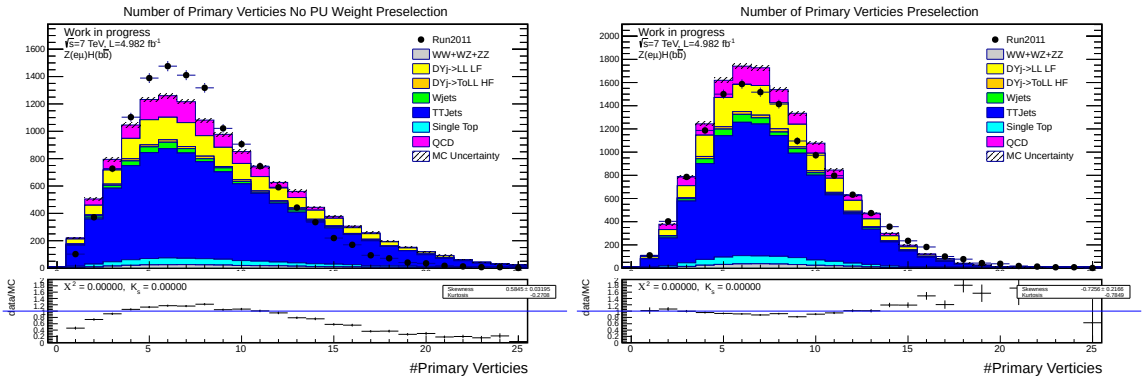

Figure 4.3: Distribution of the number of reconstructed primary vertices in 2011 data compared to simulation for $e\mu$ candidate events early in selection when the signal significance is still low. The figure on the left is before pileup reweighting is applied. Better agreement can be seen after pileup reweighting (right).

The same procedure is applied to 2012 data where the average number of primary vertices was 15.6 ± 5.4. The Summer2012 Monte Carlo samples are reweighted to a weighted sum of the pileup from 2012A, B, C, and D data taking periods. The pileup corrections for the 2012 data is shown in Figure 4.4.

4.5.2 Electrons

Particle Flow (PF) is used the identify objects consistent with the electron hypothesis. Electron selection for 7 TeV electrons begins with the standard Vector Boson Task Force 95% working point (WP95) [62] for PF electrons with $|\eta| < 2.5$, excluding the gap region $1.44 < |\eta| < 1.57$. This includes a relative combined isolation, defined as

$$R \equiv \frac{\sum_i \left[p_{\mathrm{T}i}(\texttt{chargedHadron}) + p_{\mathrm{T}i}(\texttt{neutralHadron}) + p_{\mathrm{T}i}(\texttt{Photon}) \right]}{p_{\mathrm{T}}{}^{\mu}} \tag{4.1}$$

computed in a cone around the electron direction. The isolation cut for WP95 is 0.35. A p_T threshold of 10 GeV is applied.

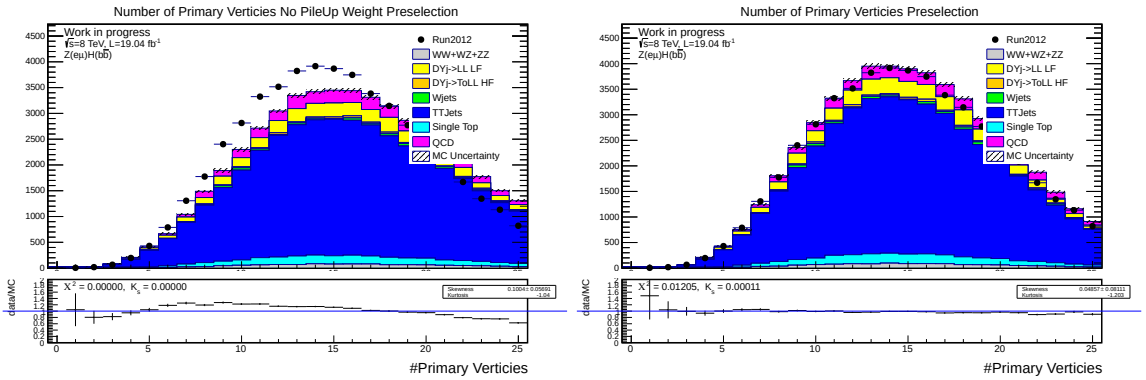

Figure 4.4: Distribution of the number of reconstructed primary vertices in 2012 data compared to simulation for $e\mu$ candidate events early in selection when the signal significance is still low. The figure on the left is before pileup reweighting is applied. Better agreement can be seen after pileup reweighting (right).

For 8 TeV electrons, the new multivariate approach suggested by the Egamma POG is used. These electrons are reconstructed using the Gaussian Sum Filter algorithm (GSF Electrons). We cut on the MVA ID discriminator at the point where it is expected to have a selection efficiency of 95%.

In order to keep as high of an acceptance as possible WP95 was kept for the $e\mu$ analysis. However, when comparing simulation to 7 TeV data, it was noticed that muons sometimes fake electrons when electron and muons are close together. Therefore, in addition to Vector Boson Task Force electron identification being applied, we require that the distance $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ between the electron and muon is more than 0.3. As seen in Figure 4.5, this effectively eliminates the discrepancy between simulation and data while preserving as much signal as possible in both 7 TeV and 8 TeV data.

Since the lepton reconstruction and selection efficiencies can be slightly different between data and simulated events, correction factors have to be applied to the MC to

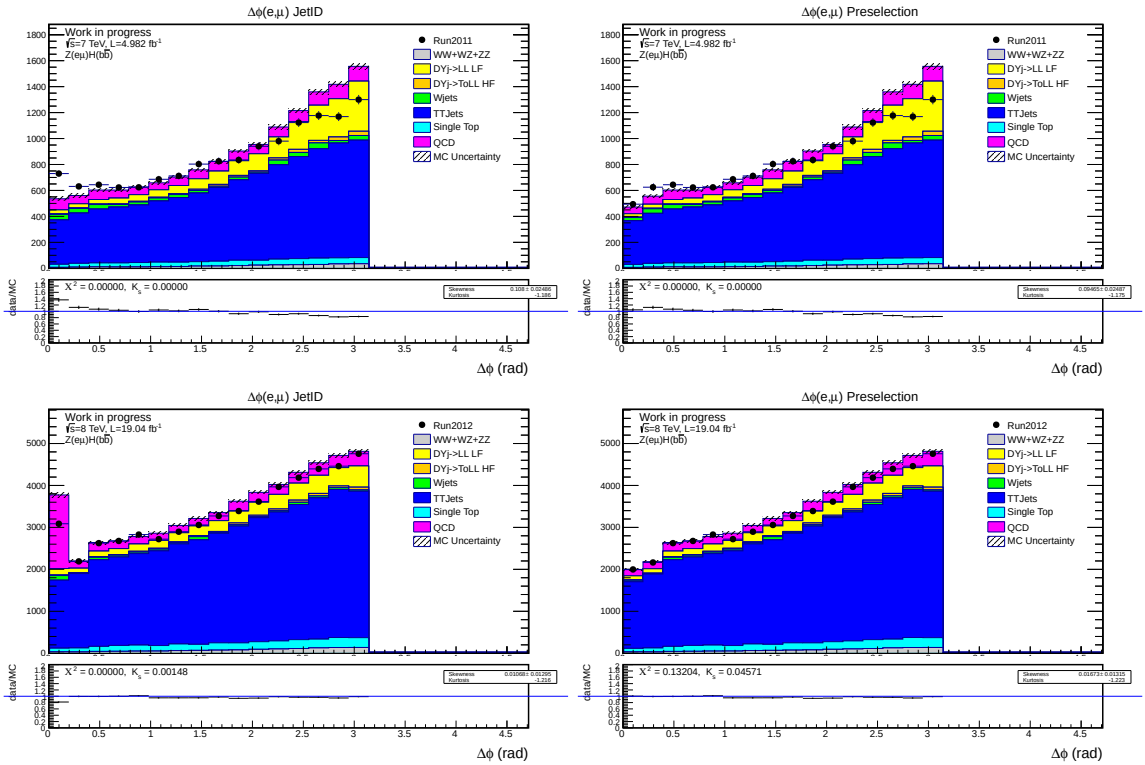

Figure 4.5: $\Delta\phi$ between the electron and muon from the $e\mu$ candidate before (left) and after (right) $\Delta R > 0.3$ requirement is applied for 7 TeV (top) and 8 TeV (bottom) data.

account for these differences. Data/MC scale factors for reconstruction and identification efficiencies, including isolation efficiencies, are listed in Tables 4.15 and 4.16 in bins of p_T and η using standard tag-and-probe [63] techniques using $Z \to \ell\ell$ decays. For electrons, the SingleElectron and ElectronHad primary datasets are compared to the Summer11-DYToEE sample using the tag-and-probe method as described in Ref. 51. Results measured from $Z \to ee$ events were found to be consistent with scale factors obtained by using $e\mu$ events and events with b-jets. Namely $e\mu$ events and events with b-jets were found to be consistent with scale factors found with $Z \to \ell\ell$ events. However there are low statistics for ZZ events where one $Z \to b\bar{b}$ and the other $Z \to ee$, so these events were only used to cross check the tag and probe result. For electrons below 20 GeV, it was found that scale factors used for WP95 are consistent with veto electrons within uncertainties [64]. Definitions of a few electron identification recipes are shown in Table 4.14. The p_T of the electron used to build the Z boson candidate is shown in 4.6 before and after the electron identification and isolation scale factors are applied. Statistical uncertainties vary bin to bin in p_T and η and there is a systematic uncertainty associated with the tag and probe method, thus an uncertainty of 2% is applied to electron reconstruction and identification efficiencies to be conservative.

Table 4.14: Selection criteria for a few electron identification definitions in the barrel region. The differences are small, thus common scale factors with generous systematic uncertainties can be used.

Cut Variable	WP95	veto	simple cuts
$\Delta\eta$	0.007	0.007	0.007
$\Delta\phi$	0.8	0.8	0.8
$\sigma_{i\eta i\eta}$	0.01	0.01	0.01
H/E	0.15	0.15	0.15
d0(vtx)	0.04	0.04	
dZ(vtx)	0.2	0.2	
isolation	0.15	0.15	0.15
missing hits	< 1	-1	< 1

Table 4.15: Electron scale factors in bins of $|\eta|$ and p_T from 2011 data applied to Fall2011 MC. Square brackets [X,Y] indicate the p_T bin $X < p_T < Y$.

η bin	$[10, 15]$	$[15, 20]$	$[20, 30]$	$[30, 40]$	$[40, 50]$	$[50, 200]$
$0 < \|\eta\| < 0.80$	1.033	0.995	0.992	0.991	0.995	0.995
$0.80 < \|\eta\| < 1.44$	1.072	0.944	0.913	0.986	0.993	0.993
$1.44 < \|\eta\| < 1.56$	1.213	0.984	0.945	0.978	0.983	0.983
$1.56 < \|\eta\| < 2.00$	1.289	1.068	0.995	0.994	0.996	0.996
$2 < \|\eta\| < 2.5$	0.999	1.122	1.018	1.003	1.004	1.004

Table 4.16: Electron scale factors in bins of $|\eta|$ and p_T from 2012 data applied to Summer2012 MC. Square brackets [X,Y] indicate the p_T bin $X < p_T < Y$.

η bin	$[10, 15]$	$[15, 20]$	$[20, 30]$	$[30, 40]$	$[40, 50]$	$[50, 200]$
$0 < \|\eta\| < 0.80$	0.853	0.949	1.017	1.023	1.018	1.008
$0.80 < \|\eta\| < 1.44$	0.841	0.971	1.001	1.009	1.004	0.995
$1.44 < \|\eta\| < 1.56$	1.089	1.146	1.144	1.018	0.988	0.999
$1.56 < \|\eta\| < 2.00$	0.817	0.924	0.997	1.007	1.002	0.99
$2 < \|\eta\| < 2.5$	1.124	1.15	1.078	1.042	1.026	1.004

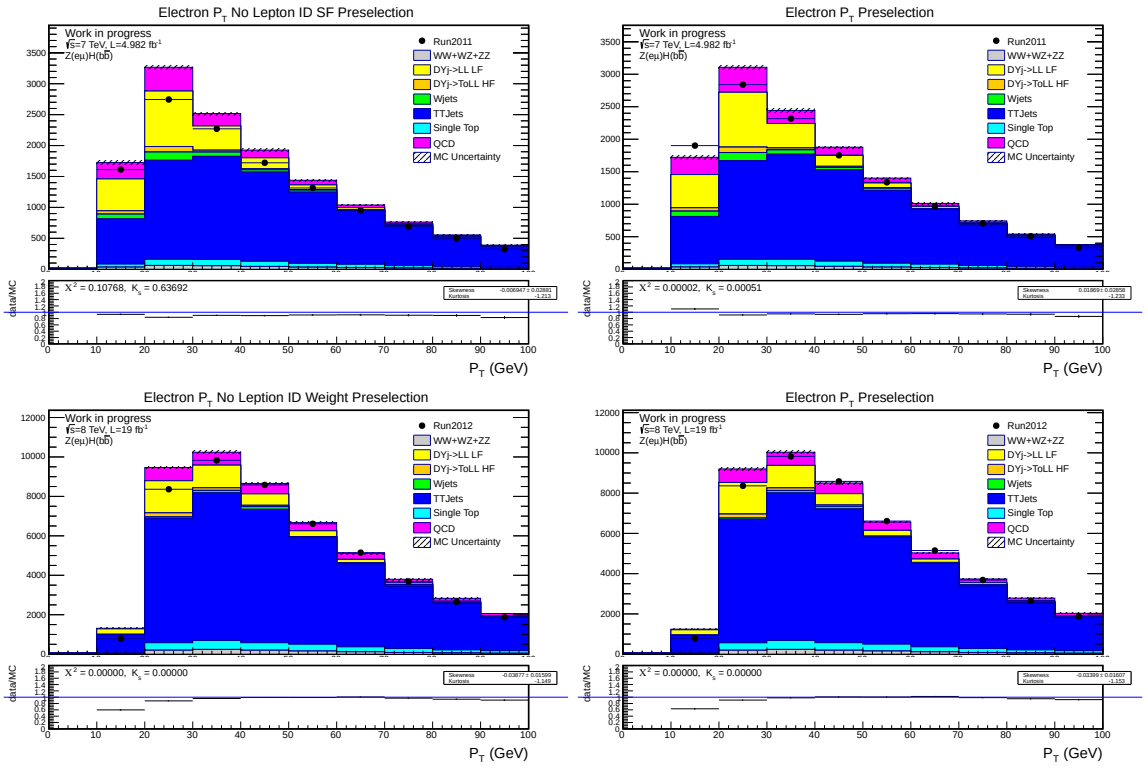

Figure 4.6: p_T of electron used to building the Z candidate before (left) and after (right) electron identification and isolation scale factors have been applied for 2011 (top) and 2012 (bottom) datasamples for events passing preselection.

4.5.3 Muons

RECO muons are selected using the criteria from the Vector Boson Task Force:

- Global and Tracker;
- $\chi^2/\texttt{ndof} < 10$ for the global muon fit;
- Tracks associated to muons must satisfy:
  - at least one pixel hit;
  - at least eleven total hits (strip + pixel);
  - at least one valid hit in the muon chambers;
  - at least two muon stations;
  - impact parameter in the transverse plane $d_{xy} < 2\,\text{mm}$;
  - $|\eta| < 2.4$;
  - Relative combined isolation < 0.15, computed in a cone of radius 0.3.
- $p_T(\mu) > 10\,\text{GeV}$.

Data/MC scale factors for reconstruction and identification efficiencies, including isolation efficiencies, are listed in Tables 4.17 and 4.18 in bins of p_T and η using standard tag-and-probe techniques [63] using $Z \to ll$ decays. For muons, the SingleMuon dataset is compared to the Summer11-DYToMuMu sample as described in Ref. 51. These results were checked against scale factors obtained by using events closer to our signal topology. The p_T of the electron used to build the Z boson candidate is shown in 4.7 before and after the electron identification and isolation scale factors are applied. Statistical uncertainties vary bin to bin in p_T and η and there is a systematic uncertainties associated with the tag and probe method, so an uncertainty of 2% is applied to muon reconstruction and identification efficiencies to be conservative.

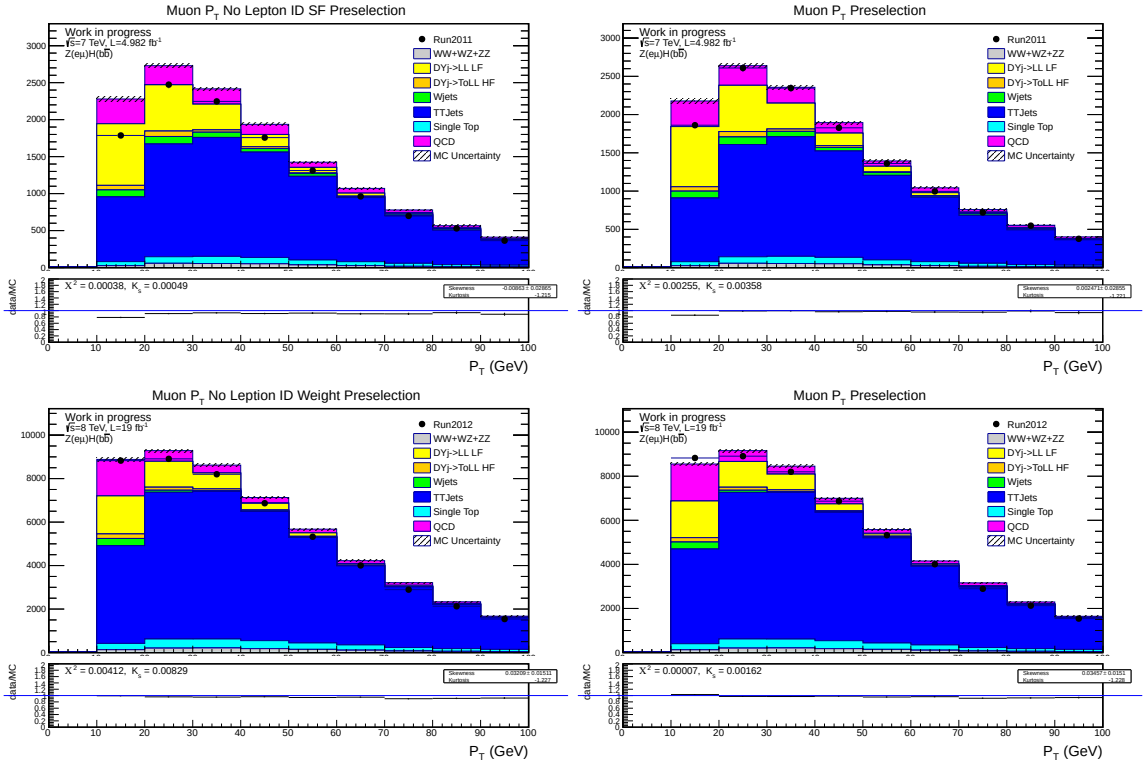

Figure 4.7: p_T of muon used to building the Z candidate before (left) and after (right) electron identification and isolation scale factors have been applied for 2011 (top) and 2012 (bottom) datasamples for events passing preselection.

Table 4.17: Muon scale factors in bins of $|\eta|$ and p_T from 2011 data applied to Fall2011 MC.

p_T bin	$0 < \|\eta\| < 1.2$	$1.2 < \|\eta\| < 2.4$
$10 < p_T < 20$	0.935	0.994
$20 < p_T < 30$	0.982	0.991
$30 < p_T < 40$	0.993	0.996
$40 < p_T < 50$	0.996	0.999
$50 < p_T < 60$	0.996	0.998
$60 < p_T < 80$	0.996	1.001
$80 < p_T < 250$	0.999	0.997

Table 4.18: Muon scale factors in bins of $|\eta|$ and p_T from 2012 data applied to Summer2012 MC.

p_T bin	$0 < \|\eta\| < 0.9$	$0.9 < \|\eta\| < 1.2$	$1.2 < \|\eta\| < 2.1$	$2.1 < \|\eta\| < 2.4$
$10 < p_T < 20$	0.985	0.987	1.012	1.014
$20 < p_T < 25$	0.989	0.987	1.002	0.994
$25 < p_T < 30$	0.994	0.994	0.999	0.994
$30 < p_T < 35$	0.994	0.991	0.998	0.994
$35 < p_T < 40$	0.994	0.990	0.997	0.994
$40 < p_T < 50$	0.992	0.990	0.997	0.994
$50 < p_T < 60$	0.991	0.991	0.998	0.994
$60 < p_T < 90$	0.990	0.986	0.994	0.994
$90 < p_T < 140$	1.004	1.012	1.019	0.994
$140 < p_T < 300$	1.028	0.956	1.016	0.994
$300 < p_T < 500$	1.000	1.000	0.609	0.994

4.5.4 Jets

Jets are clustered from all particle-flow candidates using the anti-k_T algorithm [65] with an opening parameter of 0.5. Each jet is required to lie within $|\eta| < 2.5$, have at least two tracks associated to it, and have electromagnetic and hadronic energy fractions of at least 1% of the total energy to avoid mis-reconstruction from noise clusters. Standard jet energy scale corrections (including $L2L3Residual$ corrections) are applied as described in Ref. 66. To take the effect of event pileup into account $L1Fastjet$ correction are applied. Jets that are found to be too close to leptons are assumed to be fakes and are removed from the event. A minimum threshold of $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ is applied.

We smear the MC jet energy resolution using the prescription from the JetMET POG [67]. Based on the observed data/MC difference in energy resolution, we smear jets by 5%. These smeared jet energies are then used as default throughout the analysis for the both signal and background. Jets are overall well simulated as the effect on the mass

resolution is almost negligible, a few % relative. We apply the standard scale and resolution systematic uncertainties to the BDT shape, on top of jet smearing, when calculating the 95% confidence level upper limits on the production of the Higgs boson with respect to the SM expectation for the combined 7 and 8 TeV data sets.

4.5.5 Identification of *b*-jets

b-jets are identified by the Combined Secondary Vertex (CSV) algorithm [68]. As the name suggests, this algorithm uses a combination of three categories of information: secondary vertex categories, bottom *vs* charm and bottom *vs* light flavor discrimination, and other relevant jet variables, like the impact parameter significance and decay lengths. This algorithm, in an optimal way, makes use of information about track impact parameters and identified secondary vertices within jets, even when full vertex information is not available. An additional category for jets where no vertex is found is defined. These jets are said to have a "pseudo vertex". These vertex categories are inputs to a likelihood discriminant to provide maximal separation of *b*-jets from the much larger background of jets arising from charm decay, and from the fragmentation of light quarks and gluons.

The CSV algorithm provides a continuous discriminator output that can be used to select optimal working points. The BTV POG has defined standard Loose/Medium/Tight working points described in Table 4.19 [69]. For the second jet, it was found that the lowest limit on the SM Higgs cross section is found when no *b*-tag requirement is applied. This preserves as much signal as possible. For the leading jet, ordered in CSV value, it was found that the lowest limit is found when the CSVL *b*-tag requirement is applied. For the Higgs candidate algorithm it was found that a threshold of 0.4 is optimal for both jets that

make up the Higgs candidate. This is a threshold of 0.4, not a cut. The Higgs candidate algorithm is described in detail in section 4.7. While 0.4 is not a BTV POG working point, it is similar to the optimal working b-tag cut found by the $VHbb$ analysis [51], which also found an optimal b-tag cut between CSVM and CSVL working points.

We correct for data/MC b-tag scale factor over the entire range of CSV values using the method developed by the $VHbb$ working group. This function returns a corrected CSV value based on the original CSV value and jet flavor for MC event. The function uses different correction values for 2011 and 2012 data. In Figure 4.8, better matching between MC and data is seen in the CSV discriminant value after correction is applied. As seen in Figure 4.5, this effectively eliminates the discrepancy between simulation and data while preserving as much signal as possible. Uncertainties in the b-tag scale factors are then propagated to shape uncertainty distributions that depend on the CSV value.

4.5.6 Missing transverse energy

Missing transverse momentum ($\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$) is used in the reconstruction of the Z boson. The missing transeverse energy E_{miss}^{T} is computed from the list of particle-flow particles (electrons, muons, neutral and charged hadrons, and photons) with the method described

Table 4.19: CSV working point discriminator values, b-tagging efficiency and fake rate. Misidentification probability is reported for jet p_T in the range of 80-120 GeV. The fake rate should be lower for jets typically used in this analysis.

	label	disc. value	efficiency	fake rate
Loose	CSVL	0.244	0.84	0.1020 ± 0.0004
Medium	CSVM	0.679	0.71	0.0151 ± 0.0002
Tight	CSVT	0.898	0.57	0.00120 ± 0.00005

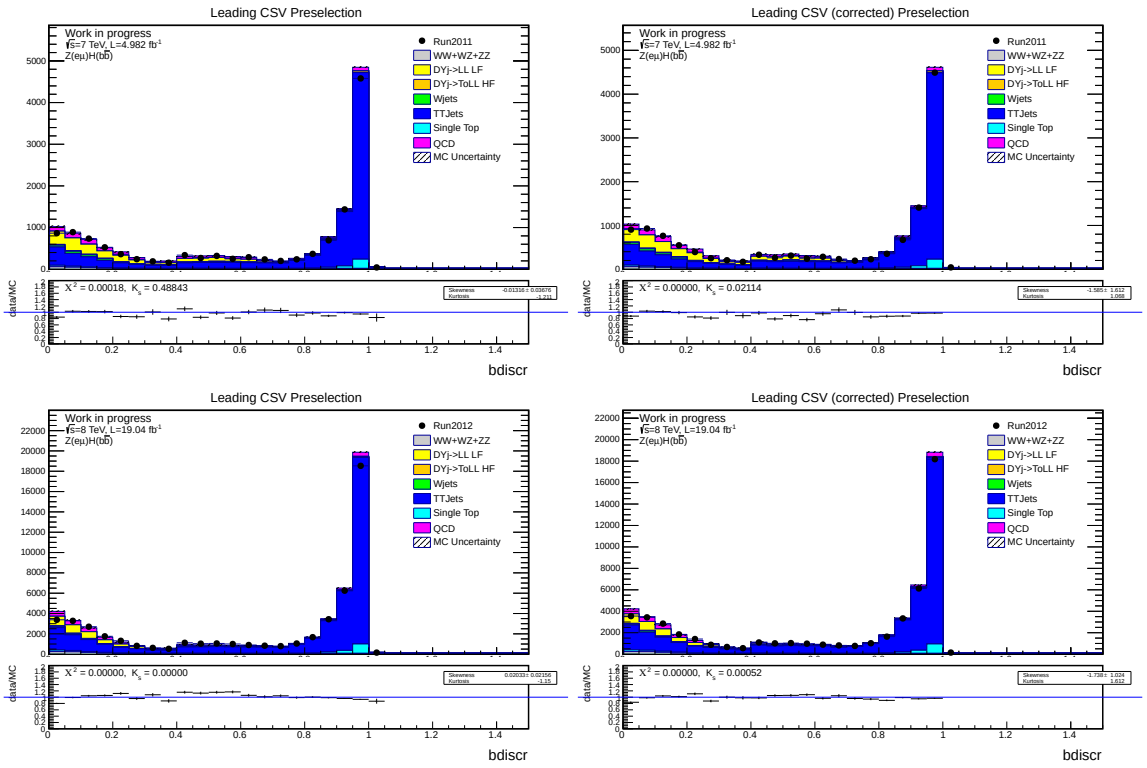

Figure 4.8: CSV discriminant distribution for the Higgs candidate jet with the highest CSV value before (left) and after (right) CSV value is corrected for 2011 (top) and 2012 (bottom) data.

in Ref. 70. The vector $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is calculated as the negative of the vectorial sum of transverse momenta of all particle-flow objects identified in the event, and the magnitude of this vector is referred to as E_{miss}^{T}. This value divided by the scalar sum of E_{T} of all particle-flow objects is referred to as the "E_{miss}^{T} significance."

The $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is corrected using type-I missing transverse energy corrections. This propagates the jet energy corrections to $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ by using jets from the anti-k_T algorithm [65] with an opening parameter of 0.5. The expected resolution of E_{miss}^{T} is estimated on an event-by-event basis by the particle-flow $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ significance algorithm [71]. While no minimum E_{miss}^{T} threshold is required, the distance in ϕ between $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ and the reconstructed mass of the Z candidate are important variables for this analysis. The vector, $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is an input to the SVfit method of reconstructing the ditau mass.

4.6 Vector Boson Reconstruction

Further background rejection is achieved by exploiting the resonant dijet mass, the event topology and the mass peak coming from the Z boson. Reconstruction of the Z boson begins with identification of the electron, muon, and $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ as described in the previous section. Reconstructing the $Z \to \tau\tau$ mass is critical for this analysis. The tau lifetime is 2.906×10^{13} seconds [52], almost instantaneous. Methods of trying to tag taus's displaced vertex, similar to b-tagging, have been unsuccessful. An additional problem in tau reconstruction is that the neutrinos produced in tau decays are not detected by the CMS detector and therefore it is unknown how much energy and momentum each neutrino carries away.

A simple reconstruction of the Z bosons using only the electron and muon, referred to as the visible mass, produces a result for every candidate event, but peaks near 40 GeV, far from the nominal mass of 91 GeV. The collinear approximation (CA) [72] was also investigated as a way of reconstructing the ditau mass, $M_{\tau\tau}$. This method assumes that the neutrinos produced in tau decays are are the only source of E^T_{miss} in the event and the momentum vectors of the neutrinos produced in tau decays are collinear to the momentum vectors of the visible decay products. The CA uses a system of linear equations to approximate the neutrino momenta and to reconstruct the full Z boson mass. This system of equations becomes degenerate in the case where the tau leptons are back-to-back. Also, the CA method fails to find physical solutions, that is, it yields negative neutrino energies, in the case where $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is reconstructed outside the azimuthal range spanned by the visible decay products. It was found that the CA method gave physical solutions only for 40% of signal events. This method was abandoned because a 40% cut is too tight for a channel with such a low branching fraction.

A likelihood based $M_{\tau\tau}$ reconstruction was found to give optimal performance for this analysis. This SVfit algorithm [73] reconstructs the mass of the ditau system based on computed probability functions for ditau decay modes. Probabilities have been calculated based on a set of parameters observable in the CMS detector. The SVfit algorithm returns the ditau mass that corresponds to the highest probability given a set of observed parameters in the event. Solutions are products of two terms: the compatibility of visible tau decay products with tau decay kinematics and compatibility of $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ with the tau hypothesis. The algorithm assumes that the decay products of the two taus are at least separated by

a $\Delta R > 0.3$. As mentioned in the previous section, a requirement of $\Delta R > 0.3$ is already applied to remove electron fakes from poorly reconstructed muons. Input variables to the SVfit algorithm are the four momenta of the visible tau decay products, the corrected $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$, and a covariance matrix of the corrected E_{miss}^{T} resolution. Unconstrained parameters are the invariant mass of each neutrino pair from electron and muon decays and the decay angle in the rest frame. A full description of this method is given in Ref. 73.

The SVfit algorithm gives a small improvement in mass resolution. The root mean square of the distribution is smaller by about 7% as seen in Figure 4.9. The true power of the SVfit algorithm comes from correctly identifying events with true ditau decays. Z+jets and signal event are reconstructed with $M_{\tau\tau}$ very near the nominal Z mass. $t\bar{t}$ events are identified as inconsistent with the ditau mass hypothesis and do not have SVfit values in the Z mass peak. This gives a larger than 10% improvement to the limit on the Higgs cross section as compared to using the visible mass.

4.7 Higgs Boson Reconstruction

Reconstruction of the Higgs boson seems simple, for events with only two jets above 20 GeV there is no ambiguity on which jets constitute the Higgs candidate. However for events with more than two jets, however, we need to determine the best algorithm for picking the jets that came from the Higgs. Initial state and final state radiation produce multiple jets, making it more difficult to choose the jets that decayed from the Higgs boson. Several methods were investigated: two jets with highest p_T, two jets with highest CSV value, and two jets that created the Higgs candidate with the highest p_T. It was found

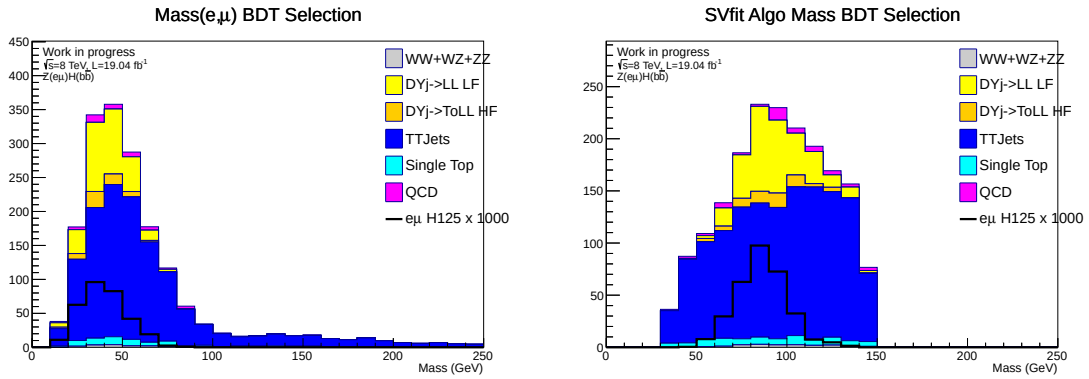

Figure 4.9: Reconstructed Z boson mass for visible mass (left) and SVfit mass (left) for events used in the BDT analysis.

that by choosing the two jets with the highest b-tag value to build the Higgs candidate, the jets coming from the generated Higgs were selected more often than choosing the Higgs candidate with the highest p_T.

It was found that when the b-tag value is below a certain threshold that it is best to choose the jet with the highest p_T. Using this scheme, instead of simply the best two CSV values, improves matching by 8%. Many b-tag thresholds were studied, including the CSV tight, medium and loose working points. Using a CSV value of 0.4 yields the most Higgs candidates with jets coming from a generated Higgs boson. Improved Higgs mass resolution and separation from signal and background was observed with higher correct matches.

In general, requiring a large p_T for the dijet system improves the dijet mass resolution and improves signal significance. A resolution of approximately 15% is achieved for p_T requirement on the Higgs candidate. A resolution of 10% is recorded by the $VH(bb)$ working group when $p_T(\mathrm{jj}) > 100\,\mathrm{GeV}$ is implemented. Even though the resolution is better

at larger p_T for the dijet system, the limit is optimal without a p_T cut for the dijet system because of the increased acceptance. Due to the low branching fraction of this analysis channel, no p_T requirement for the dijet system is applied. For the $e\mu$ analysis a p_T cut for the dijet system may be more appropriate for larger datasets than are currently recorded.

4.8 Event Selection

The final result of this analysis is a 95% confidence level limit on the SM Higgs boson cross section using the 7 TeV and 8 TeV data samples, obtained by fitting the shape of the resulting boosted decision tree (BDT) [53] output to the background only and background plus signal hypotheses. Before the BDT is trained it must be decided which events should be used in the training, referred to as the "BDT selection", in which selection criteria were optimized with respect to the final result. The BDT selection starts from the sample of $e\mu$ events that passed the preselection criteria, then uses cuts on only two variables both to preserve the maximum statistical sample and to avoid bias within the BDT process.

The $e\mu b\bar{b}$ candidates are selected from the sample of events that contain one electron, one muon, $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ and at least two reconstructed jets. The muon and electron with highest p_T but opposite charge are are assumed to come from the Z decay. Minimum p_T thresholds of 20 GeV for the leading lepton and 10 GeV for the second lepton are applied. The jets are required to pass jet identification, $p_T > 20$ GeV, and to lie within $|\eta| < 2.5$. The jets selected to build the Higgs candidate are described in Section 4.7.

4.8.1 Signal and Background Characteristics

Signal events are characterized by a Z boson recoiling against a Higgs boson. For a $Z \to \tau\tau$ candidate reconstructed from an electron and a muon, the p_T from the four neutrinos is not measured resulting in an ambiguity in the direction and p_T of the $Z \to \tau\tau$ candidate. The azimuthal angle between the Z boson and the Higgs boson is sharply peaked at 180 degrees as the virtual Z boson which produces the Z and Higgs bosons is expected to be produced close to rest, therefore the Z boson and Higgs boson would be produced back to back to conserve momentum. It is not possible to measure the pseudorapidity of the missing momentum in hadron collisions, but it is possible to infer the azimuth of $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$. The the azimuthal angle between the Z candidate and the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ are sharply peaked at 0 degrees. In $e\mu$ events we expect most of the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ to come from the four neutrinos from the tau lepton ($e$and μ) decays. We, therefore, expect the Z boson and neutrinos to share a boost direction. A well reconstructed Higgs boson mass and Z boson mass are essential to the ZH signal because we do not expect two high mass objects from any other sources other than our signal and diboson production.

The dominant backgrounds of this analysis come from three sources listed below:

- ***V*+jets:** production of W and Z in association with one or more jets has a topology very similar to signal events, but has a more sharply falling $M(\mathrm{jj})$ distribution. This background has decay characteristic of color radiation and spin that differ significantly from the signal. W+jets events, in which one lepton comes from the W and the other from a semileptonic decay of a b or a fake, do not have event kinematics compatible with the ditau hypothesis. Tagging the Higgs daughter with the highest CSV value

removes a significant part of this background. The azimuthal angle between the Z candidate and the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is uniformly distributed for the V+jets sample because many times the Z is reconstructed from leptons other than its decay products. Requiring the Z candidate and $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ to be collinear removes half of these background events.

- **Top quarks:** production of $t\bar{t}$ pairs, as well as single top quarks in the tW, t-channel, and s-channel processes, is a particularly challenging background for this analysis. For $t\bar{t}$, events include two real W decays, at least two b-jets, and a large intrinsic mass and momentum scales close to the Higgs mass scale. Top quarks tend to be produced back to back producing back-to-back leptons. This background is reduced by requiring the azimuthal angle between the Z candidate and the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ is collinear and by a mass window cut on the ditau mass because this background is often identified as incompatible with the ditau mass hypothesis, however, the large production cross section means that even after cutting out 99% of $t\bar{t}$ events it still remains the dominant background. Single Top events could be eliminated by applying a b-tag on the second Higgs jet, but they make up only 4% of the total background so no such requirement is applied in order to conserve signal.

- **Diboson (WW, WZ, ZZ):** production of vector-boson pairs is another important background as it also produces a resonant dijet mass within a few standard deviations of the signal. Events where one Z decays leptonically and the second Z decays to jets have the same event topology as the signal. The WW mode has the largest diboson production cross section, but these events do not have leptons coming from the same vector boson. Requiring the Z candidate reconstructed from the electron and muon

and $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ to be collinear removes almost 90% of WW events. WZ and ZZ modes can decay into $\mathrm{Z} \to \mathrm{b\bar{b}}$ resulting in two real b-jets. $M(\mathrm{jj})$ resolution is critical for distinguishing these background events from signal.

In addition, QCD multijet events were investigated as a possible source of background. It was found that the combination of stringent requirements on five objects and a few event shape cuts was sufficient to remove the contribution from simulated QCD events. However, a missing contribution between data and simulation was discovered as seen in Figure 4.10 left. It was found that this missing contribution could be modeled using same sign events. These are events from data that have the same lepton identification and cuts applied except the $e\mu$ candidate events have the same sign leptons. No scale factors or event weights are applied to these events as they come directly from data. While it was considered that the difference between simulation and data could be compensated for by scaling the simulation, it was found that better modeling of the data was achieved by using the same sign background. The same sign background is labeled as QCD in the plots in this thesis as it is believed that these events come from the tails of the QCD multijet background. Due to the uncertainty of the origin of these events, we cut out the majority of the QCD background modeled from data by cutting on the CSV of the Higgs candidate jet with the highest CSV value. As seen in Figure 4.11 (left) most of the QCD events are found at low CVS values which is also a signal-poor region. In the 8 TeV analysis the contribution of QCD was reduced to 2.4% from 3.6% of the total background after the BDT selection was applied. In Figure 4.11 (right), one can see that there is very little QCD background in the Higgs mass distribution after a CSV cut of 0.244 was applied. The cut value is the loose

working point from the BTV POG.

4.8.2 Discriminating Variables

Many variables were considered for this analysis. The following variables were found to be most powerful in separating signal from background:

- $\boldsymbol{M(jj)}$**:** dijet invariant mass. $M(\mathrm{jj})$ peaks at M_H for signal with a shoulder at low $M(\mathrm{jj})$ for events where one of the Higgs jets is outside detector acceptance. Diboson and $t\bar{t}$ pairs peak below M_H.

- $\boldsymbol{M(e\mu)}$**:** visible invariant mass of electron and muon from the Z candidate.

- $\boldsymbol{M_{\tau\tau}}$: ditau mass from the SVfit algorithm.

- **CSV:** output of the CSV discriminant. We consider the CSV of the Higgs candidate daughter with the highest CSV value (CSV0) separately from the jet with the lower value (CSV1).

- $\boldsymbol{\Delta\phi(V,H)}$**:** azimuthal opening angle between the momenta of the Higgs candidate and the Z candidate reconstructed from the electron and muon.

- $\boldsymbol{\Delta\phi(Z,\vec{E}_{\mathrm{T}}^{\mathrm{miss}})}$: azimuthal opening angle between the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ and the Z candidate reconstructed from the electron and muon.

- $\boldsymbol{\Delta\phi(H,\vec{E}_{\mathrm{T}}^{\mathrm{miss}})}$: azimuthal opening angle between the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ and momenta of the Higgs candidate.

- $\boldsymbol{\Delta\phi(\mathrm{Second},\vec{E}_{\mathrm{T}}^{\mathrm{miss}})}$**:** azimuthal opening angle between the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ and the lepton with the second highest p_T coming from the Z candidate.

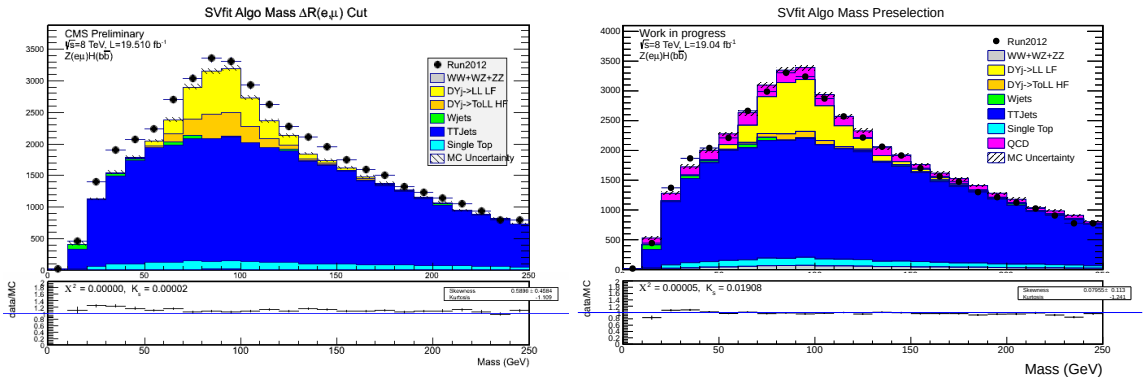

Figure 4.10: The distribution of the reconstruction Z boson mass assuming the ditau hypothesis before (left) and after (right) the same sign background, shown in pink, is added.

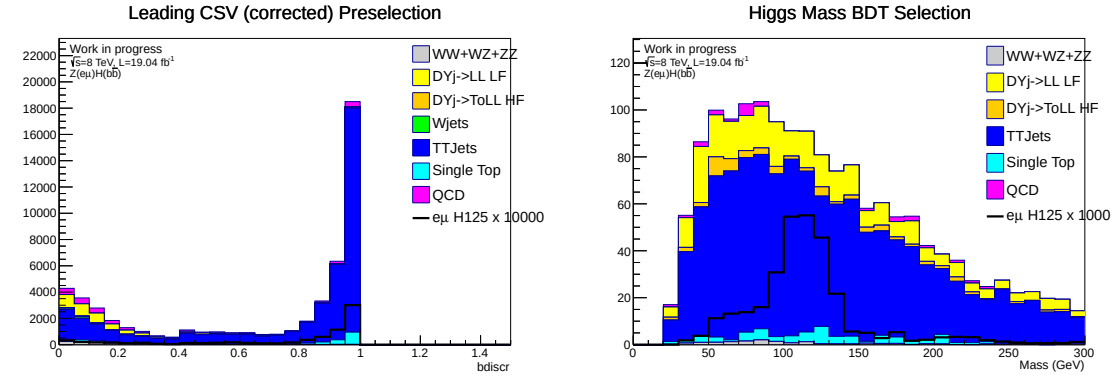

Figure 4.11: The QCD background is concentrated at low CSV value for the Higgs candidate jet with the leading CSV (right). After a CSV cut of 0.244 there is very little QCD background left in the signal region (left).

- **ΔR(jj):** $\Delta \mathrm{R} = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ between the two jets that make up the Higgs candidate.

- **EtaStandDev**: root mean squared of the ηs of the electron, muon, and two jets from the Higgs candidate.

- $\boldsymbol{P_\zeta^{\mathrm{miss}}}$: Projection of the E_{miss}^T onto the bisector of the electron and muon from the Z boson candidate.

- $\boldsymbol{P_\zeta^{\mathrm{vis}}}$: Projection of the visible energy, the combination of the electron and muon momenta in the transverse direction, onto the bisector of the electron and muon.

- $\boldsymbol{M_T}$: transverse mass with respect to the muon or electron is labeled as $M_T\mu$ and M_Te respectively. The transverse mass with respect to lepton is defined as:

$$M_T(l) = \sqrt{2 \cdot p_{\mathrm{T}}(\ell) \cdot E_{miss}^T \left(1 - \cos(\Delta\phi_{\ell,\vec{E}_{\mathrm{T}}^{\mathrm{miss}}})\right)} \quad ; \quad \ell = e, \mu \tag{4.2}$$

Figures 4.12, 4.13, and 4.14 show the signal and background distributions for these discriminating variables.

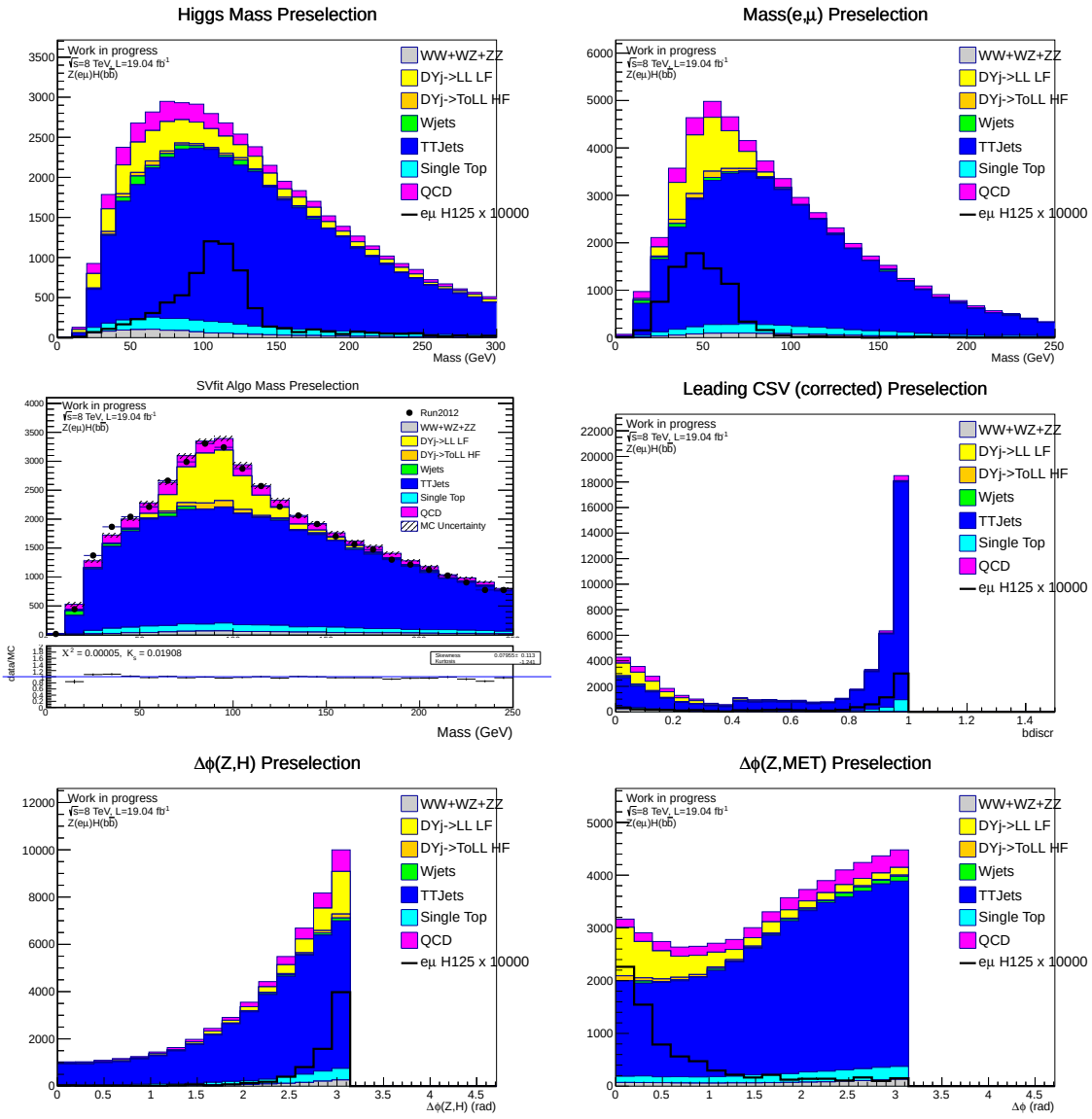

Figure 4.12: Distributions of simulated signal and background events for $M(\mathrm{jj})$ (top left), $M(\mathrm{e}\mu)$(top right), $M_{\tau\tau}$(middle left), CSV0 (middle right), $\Delta\phi(V,H)$ (bottom left) and $\Delta\phi(Z,MET)$ (bottom right) after the preselection has been applied.

4.8.3 Optimization of BDT Selection

The BDT selection retains more events than a traditional cut and count analysis. Two uncorrelated variables, $P_{\zeta}^{\mathrm{miss}}$ and $M_{\tau\tau}$, were used in the optimization to reduce the

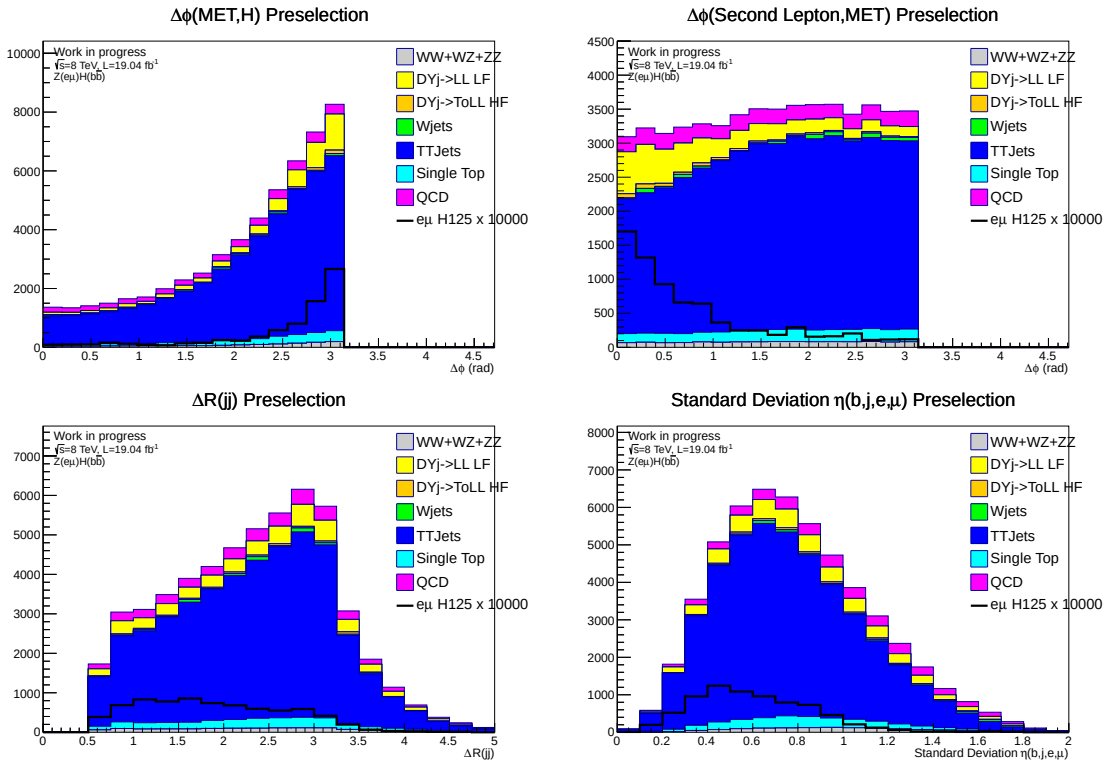

Figure 4.13: Distributions of simulated signal and background events for $\Delta\phi(H, MET)$ (top left), $\Delta\phi(Second, MET)$ (top right), Δ R(jj) (bottom left) and EtaStandDev (bottom right) after the preselection has been applied.

dominant background $t\bar{t}$. Their distributions are shown in Figure 4.15. The BDT selection was optimized by minimizing the asymptotic limit using all event weights and systematics. The full range of cut in two dimensions, show in Figure 4.16 for these two variables was calculated using the Summer12 Monte Carlo. The BDT selection is shown in table 4.20. Note that the P_ζ^{miss} cut value is above the range of the full matrix show in Figure 4.16. One can see that the region of Figure 4.16 with the lowest limits is for high P_ζ^{miss} cut value. Therefore the exercise was extended to higher values of P_ζ^{miss}. The optimal P_ζ^{miss} cut value is show in Table 4.20. Events passing this selection are then passed to the BDT for training.

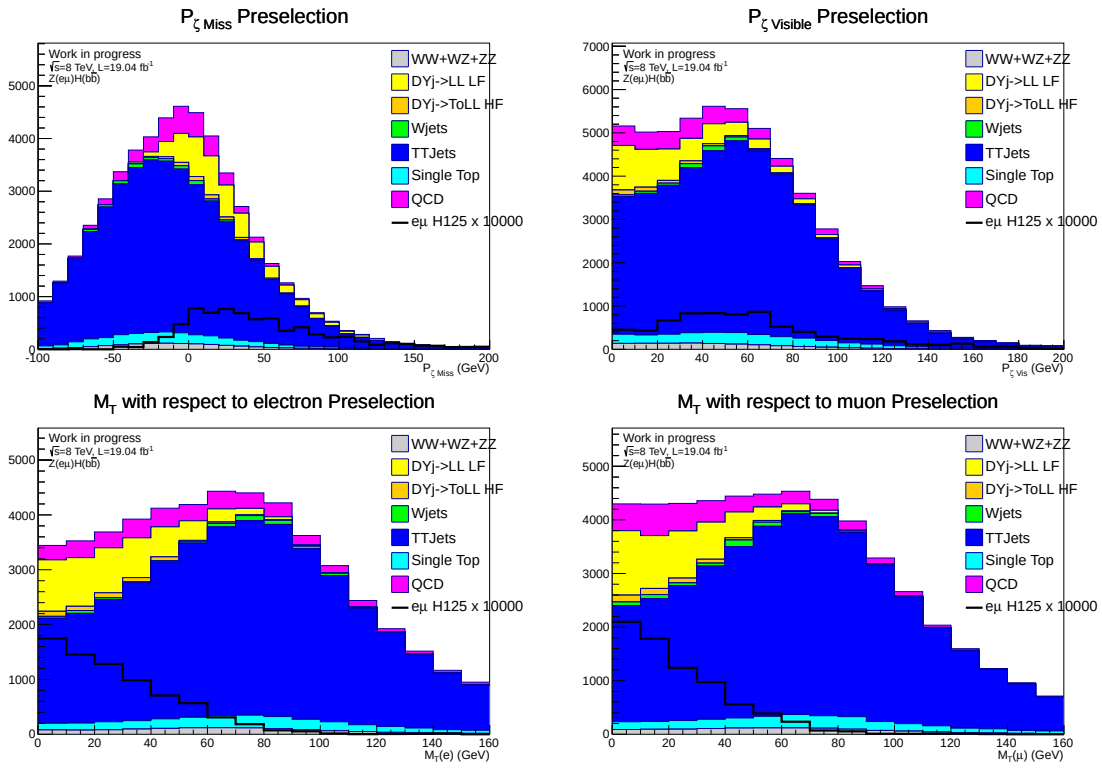

Figure 4.14: Distributions of simulated signal and background events for P_ζ^{miss}(top left), P_ζ^{vis}(top right), $M_T(e)$(bottom left) and $M_T(\mu)$(bottom right) after the preselection has been applied.

4.9 BDT Analysis

We use a boosted decision tree (BDT) as the multivariate classifier for this analysis [74] because it was found that this classifier was most performant at separating signal form background. A BDT is a collection of individual decision tress which form a multivariate classifier by using a weighted combination of a group of individual tress. A decision tree is a sequence of binary splits of data. The data is split into subsets called leaves until a leaf consists of pure signal or background events or until the minimum allowed leaf size (number of events) is reached. For this analysis the minimum number of event in a leaf was

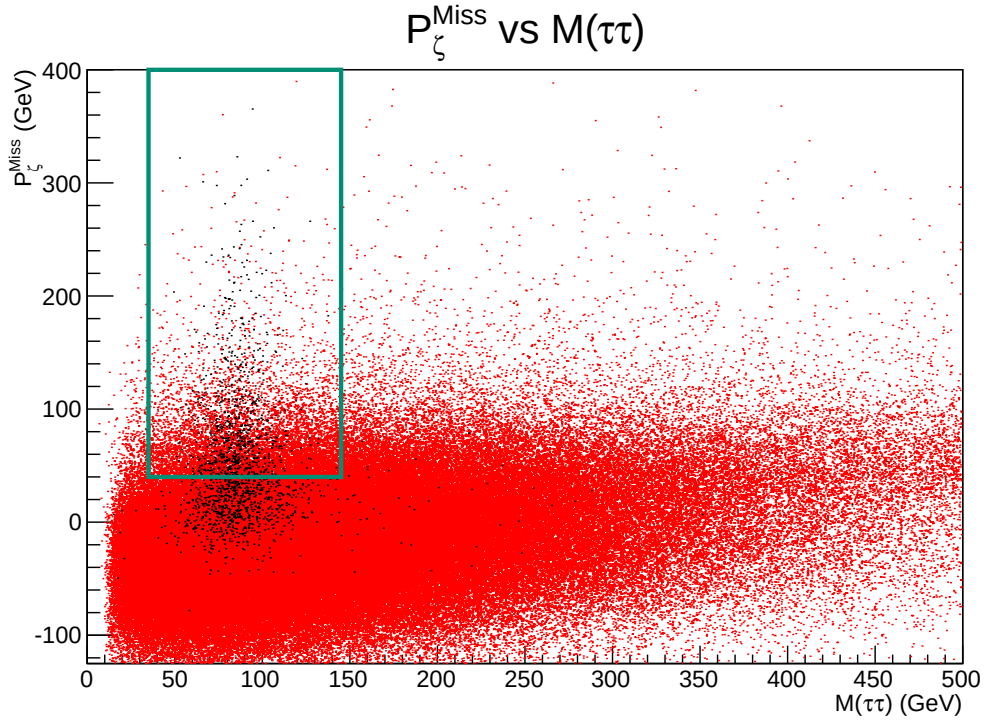

Figure 4.15: P_ζ^{miss} vs $M_{\tau\tau}$ is shown with the signal, $ZH \rightarrow Z(\ell\ell)H(b\bar{b})$, in red and the dominant background $t\bar{t}$ in black. The Higgs mass shown is 125 GeV.

Table 4.20: BDT Selection variables, cut values, and signal efficiency. The efficiency of each criterion is relative to the ones above, starting from the preselection for the 2012 Monte Carlo.

Variable	Cut	Signal Efficiency
$\Delta R(e,\mu)$	> 0.3	99%
$M_{\tau\tau}$	$35 < M(\tau\tau) < 145$	96%
CSV0	> 0.244	88%
p_ζ^{Miss}	> 40	49%
Total Eff.		42%

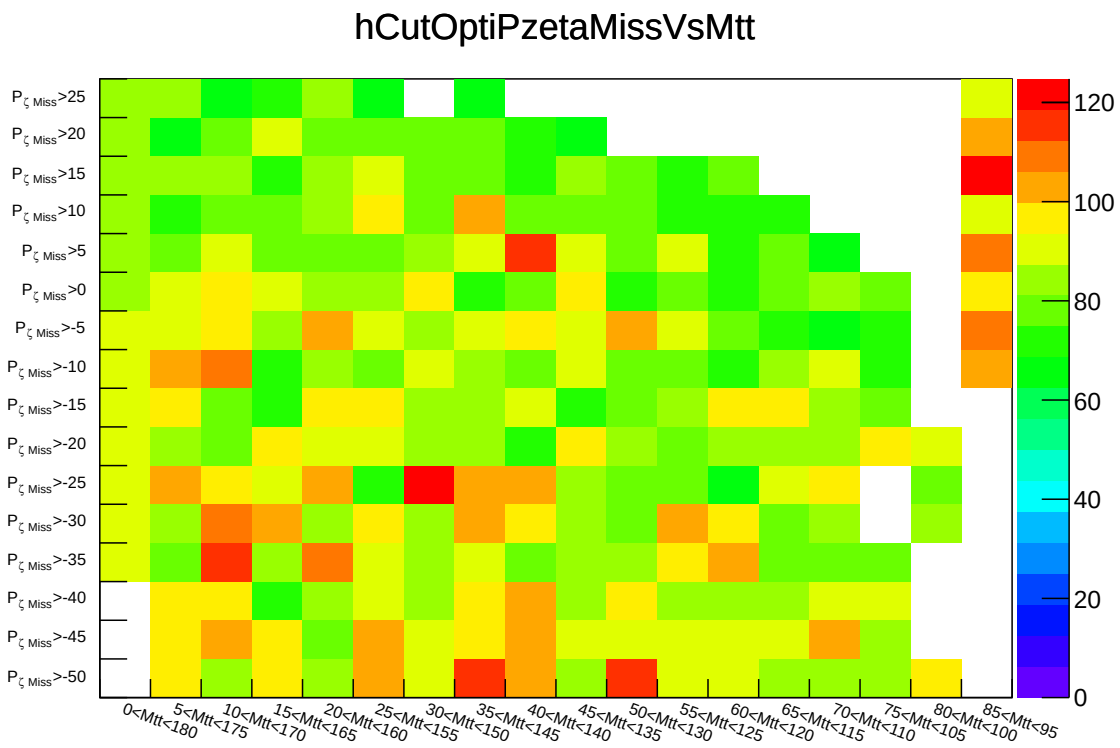

Figure 4.16: Limit values for several cuts values of P_ζ^{miss} on the y-axis and widow cuts on $M_{\tau\tau}$. Cooler colors indicate lower limits.

specified as 100 and the maximum depth of the tree, or the maximum number of splits, was 3. At the end of each tree computation, events that were misclassified are given a larger weight, or boost, in the next iteration of the training [53, 75].

This method is implemented in the TMVA framework [76]. Ten variables are used to extract more information from the events along with an optimized selection. The number of variables used to train the tree were reduced to 10 because it was found that while more variables improved the final limit result, the training became non-predictive. The BDT is trained using $e\mu$ candidate events only as these are the only events with $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ and two leptons. In order to gain statistics for training it was considered to use all $ZHbb$ events where the Z decays into two leptons. However it was found that the BDT was sensitive to the change in topology from $Z \to \mu\mu$ to $Z \to \tau\tau$. Figure 4.17 shows the separation and shape difference between both the signal (blue) and background (red) for the ten variables used by the BDT. Correlations among variables can lead to an unstable training that does not converge. Correlations among variables for signal and background are shown in Figure 4.18. The variables used to train the BDT are listed, for clarity, in Table 4.21 along with their importance, or the number of time the BDT uses the variable to split a node.

The ten variables used for the final BDT, selected from a larger set of variables considered. The most effective variables were taken, excluding those which were highly correlated. The number of variables was determined by minimizing the asymptotic limit and checking for overtraining by the BDT. The overtraining check is shown in Figure 4.19. Half of the simulated events were used to train the BDT, while the other half was used to check good modeling of the BDT output. The BDT is clearly overtrained if the majority of

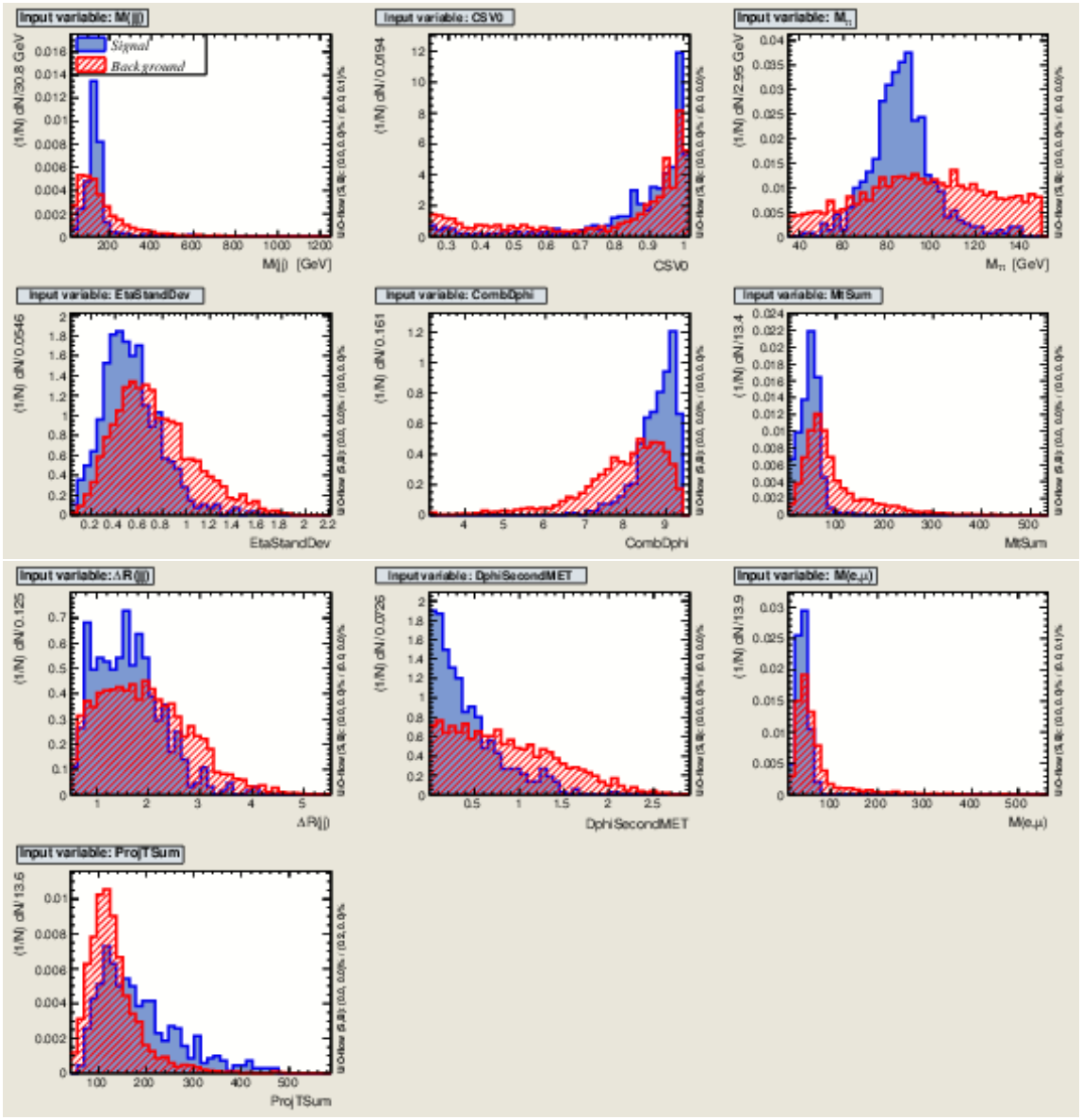

Figure 4.17: Distributions of signal of Higgs mass of 125 GeV (blue) and background (red) for the variables trained in the BDT.

Table 4.21: List of variable used to train the BDT along with variable importance given by TMVA.

Rank	Variable	Variable Importance
1	MtSum	2.06e-01
2	$M_{\tau\tau}$	1.67e-01
3	Δ R(jj)	1.38e-01
4	CombDphi	1.26e-01
5	$M(\mathrm{e}\mu)$	1.01e-01
6	$\Delta\phi(Second, \vec{E}_{\mathrm{T}}^{\mathrm{miss}})$	6.85e-02
7	M(jj)	6.33e-02
8	CSV0	5.08e-02
9	ProjTSum	4.96e-02
10	EtaStandDev	3.07e-02

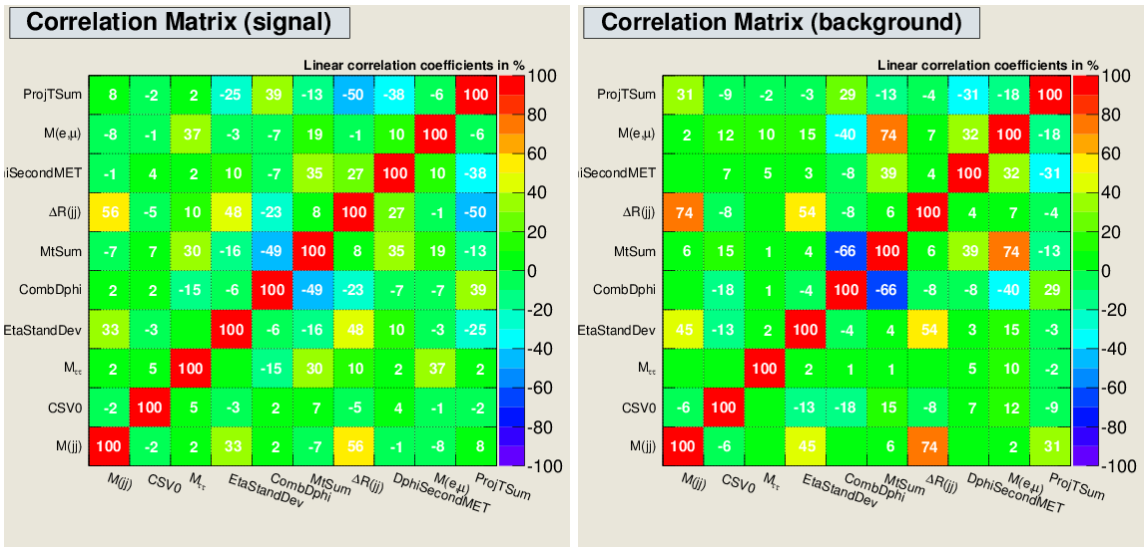

Figure 4.18: Correlations between all BDT training variables for signal of Higgs mass of 125 GeV (left) and background (right).

the signal testing sample is right of the training sample. Sharp differences in shape between the training and testing samples also indicate overtraining. The matching for the training and testing samples via the Kolmogorov-Smirnov test is poor, 0.079. The limited statistics of the signal MC sample, only on the order of 700 raw MC events, makes it difficult to determine if the poor matching is due to overtraining or statistical fluctuations.

The lowest limit was found using fifteen variables. However this divided the signal sample into too many bins. It was found that by using combination of variables the same information could be passed to the BDT without over dividing the signal sample. These combined variables are labled as MtSum, CombDphi, and ProjTSum in the figures. The transverse mass sum (MtSum) is the scalar sum of $M_T(e)$ and $M_T(\mu)$. Signal events are expected to have a lower transverse mass for both the electron and the muon than top events. The combination of $\Delta\phi$ (CombDphi) is the combination of the $\Delta\phi$ of the Higgs candidate, Z boson reconstructed from the electron and muon and the $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$. Signal events are expected to have a $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ along the Z boson direction. Also the Higgs candidate is expect to be recoiling against the Z boson and $\vec{E}_{\mathrm{T}}^{\mathrm{miss}}$ directions.The combination of $\Delta\phi$ is therefore defined as:

$$\mathrm{CombDphi} = \left(\pi - \Delta\phi(Z, \vec{E}_{\mathrm{T}}^{\mathrm{miss}})\right) + \Delta\phi(H, \vec{E}_{\mathrm{T}}^{\mathrm{miss}}) + \Delta\phi(V, H) \tag{4.3}$$

The combination of $\Delta\phi$ should peak at 3π for signal events. Thus the information from all three angles is used by the BDT in this one variable. The sum of P_ζ^{miss} and P_ζ^{vis} is labled as ProjTSum. This name stems from P_ζ being the projection onto the bisector between the electron and muon momentum vectors. After the cut on P_ζ^{miss} in the BDT Selection, both P_ζ^{miss} and P_ζ^{vis} have signal concentrated at lower values.

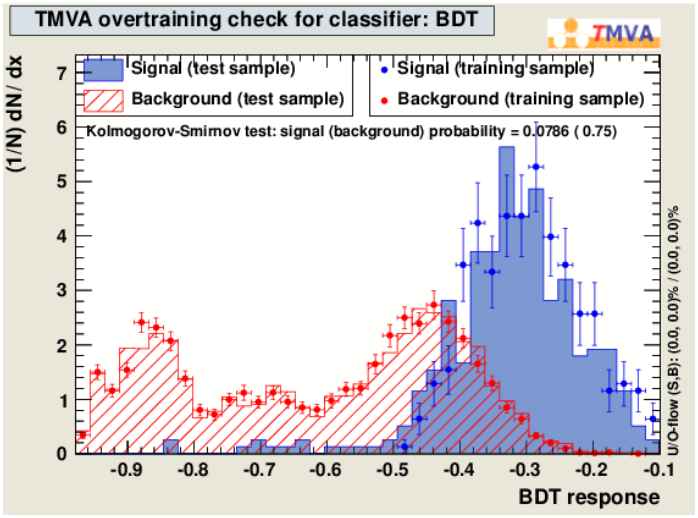

Figure 4.19: Overtraining check performed by TMVA. Signal (blue) and background (blue) distributions are show for both training and testing samples. Reasonable separation between signal and background from the BDT output is observed.

4.10 Background Control Regions

Two control regions (CR), $t\bar{t}$ and Z+light, are defined in order to adjust the MC estimates for the most important background processes. We choose control regions that are enriched in a specific processes to demonstrate that we are modeling those processes well. If data/MC matching is found to be different than unity, then we need to make adjustments to the MC predictions. Ideal control regions are orthogonal to both the signal region and other control regions, but are as close as possible to the signal region in event topology and kinematic scale. The CRs are defined in Table 4.22. The selection are kept as close as possible to the BDT selection while keeping enough statistics in the control region for a good scale factor measurement. Regions with a high b-tag value will be enriched with $t\bar{t}$. Inversely, regions with low b-tag value will be enriched with Z+light, or light flavor jets. Some additional requirements have been applied to ensure higher control region purity. The

selection is the same for 7 TeV and 8 TeV data. Control plots 4.20–4.21 are shown for 7 TeV and 8 TeV data and MC.

Scale factors are found for each control region.The $t\bar{t}$ scale factor is determined first because it has the highest CR purity. This result was then used in the determination of the scale factors for the other control regions. These scale factors correct for cross section discrepancies and residual differences in physics object selection. Tables 4.23 and 4.24 report the predicted yields for each background in each control region and the overall scale factors are applied to background process estimates in the signal region. Scale factor uncertainties are taken as systematic uncertainties and used to calculate the final limit.

Control plots of the BDT output are created to verify the BDT is properly modeling the data. Good matching is shown in Figure 4.22 for $t\bar{t}$ and Z+light control regions.

4.11 Systematics

The result quoted in this analysis is 95% confidence level upper limits on the production of the Higgs boson with respect to the standard model expectation (ϕ_{SM}) for the combined 7 and 8 TeV data sets. Systematic errors affect the final results in terms of signal and background yields and their affect on the shapes of the final BDT distribution. Several sources of systematic uncertainties have been considered and are listed in this section.

Systematic uncertainties are taken into account in two ways: as standard variations of log-normal distribution or as shape uncertainties. The log normal variables are listed here:

- **Luminosity:** an uncertainty of 2.2% is assumed for 2011 luminosity and 5.0% for 2012. This uncertainty is treated as uncorrelated between the two datasets.

Table 4.22: Selection criteria applied for each control region along with the signal BDT selection are listed. Entries marked with a "-" indicated that the variable is not used for that control region. The purity of the control region with respect to other background samples and the overall estimated contribution to the signal region of this background to the overall background in the signal region are also reported.

Variable	$t\bar{t}$	Z+light	signal
$M_{\tau\tau}$	$[35-145]$	$[35-145]$	$[35-145]$
CSV0	> 0.244	< 0.244	> 0.244
p_{ζ}^{miss}	< 40	> 40	> 40
SumMt	-	< 100	-
Purity 2011	82.3%	67.5%	
Contribution to Bkg 2011	62.5%	25.6%	
Purity 2012	89.2%	62.6%	
Contribution to Bkg 2012	70.9%	17.9%	

Table 4.23: 7 TeV data predicted yields for each background obtained from *Fall*11 Monte Carlo samples and scaled to data luminosity of $5.0\,\text{fb}^{-1}$. Lepton trigger efficiency, lepton identification, pile-up, and b-tag reweighting have been applied to all MC numbers. All uncertainties are statistical only. The last row list the scale factor values applied to the Fall11 MC.

Process	$t\bar{t}$	Z+light
TTJets	2474 ± 6.69	163 ± 1.70
DYj→LL LF	182 ± 9.65	1153 ± 22.9
DYj→LL HF	46.3 ± 4.73	51.3 ± 4.82
QCD	127 ± 11.3	236 ± 15.4
W+jets	34.2 ± 9.14	57.1 ± 12.6
Single Top	109 ± 2.59	11.5 ± 0.846
$WW + WZ + ZZ$	13.1 ± 0.854	36.9 ± 1.37
Total MC	2990 ± 19.5	1710 ± 30.8
Total Data	3010 ± 54.8	1500 ± 38.7
Data/MC	1.01 ± 0.02	0.879 ± 0.028

Table 4.24: 8 TeV data predicted yields for each background obtained from *Summer*12 Monte Carlo samples and scaled to data luminosity of 19.0 fb^{-1}. Lepton trigger efficiency, lepton identification, pile-up, and b-tag reweighting have been applied to all MC numbers. All uncertainties are statistical only. The last row list the scale factor values applied to the Summer12 MC.

Process	$t\bar{t}$	Z+light
TTJets	10000 ± 47.0	720 ± 12.6
DYj→LL LF	233 ± 19.6	2790 ± 57.9
DYj→LL HF	85.6 ± 11.2	170 ± 15.1
QCD	310 ± 17.6	635 ± 25.2
W+jets	41.8 ± 21.3	49.7 ± 18.0
Single Top	465 ± 14.6	51.7 ± 4.77
$WW + WZ + ZZ$	43.3 ± 2.07	140 ± 3.61
Total MC	10300 ± 59.5	4380 ± 68.8
Total Data	10500 ± 103	4000 ± 63.2
Data/MC	1.02 ± 0.01	0.91 ± 0.02

- **Lepton Identification Efficiency:** an overall uncertainty of 2% is assumed for electrons and muons separately. This generously covers the statistical uncertainties measured by the tag-and-probe method and accounts for systematic uncertainties of the method. The total bin by bin uncertainty is 1.6% per muon and 1.5% per electron.

- **Cross Section:** we use next-to-next-to leading order cross section calculations with a total uncertainty of 4%. These uncertainties include parton distribution function and scale variations.

- **Background Estimate:** a mix of data-driven methods, simulation and theory uncertainties contribute to the total uncertainty of the background estimates. An uncertainty of 30% is assumed for single top (approximately the uncertainty on the measured cross section) and diboson (assumed to have the same uncertainty as the signal). W+jets uncertainty is taken as 17%, which is the maximum uncertainty for

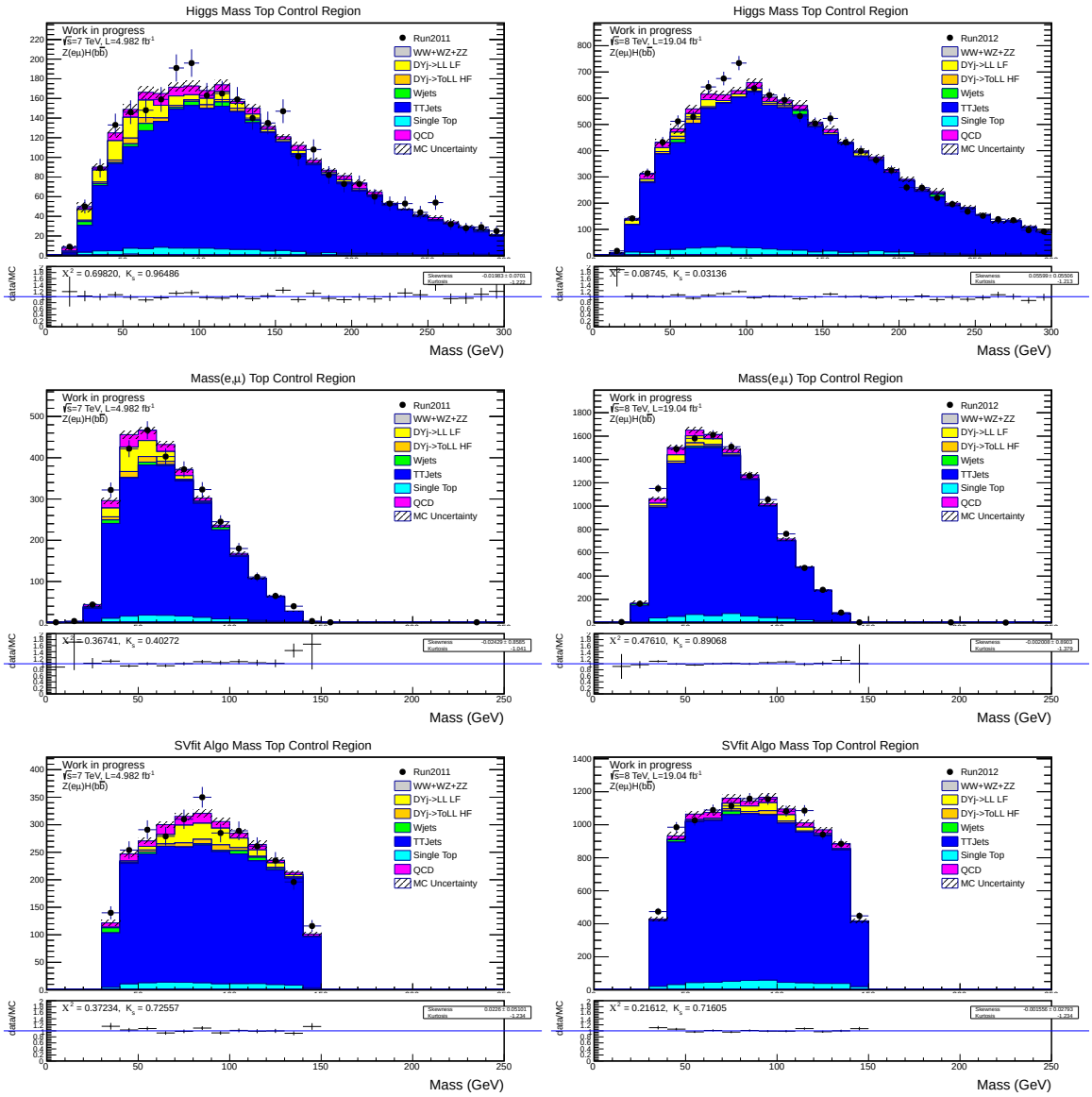

Figure 4.20: Distributions of $t\bar{t}$ control region for 7 TeV (left) and 8 TeV (right) data and MC samples. M(jj) (top), M(eμ) (middle), and $M_{\tau\tau}$ (bottom) are shown.

$WHbb$ analysis [77]. W+jets is taken as a good control region for the $e\mu$ analysis as it is close in kinematics and topology, yet statistically independent. The other backgrounds are measured directly in data. The corresponding uncertainties are from their control regions.

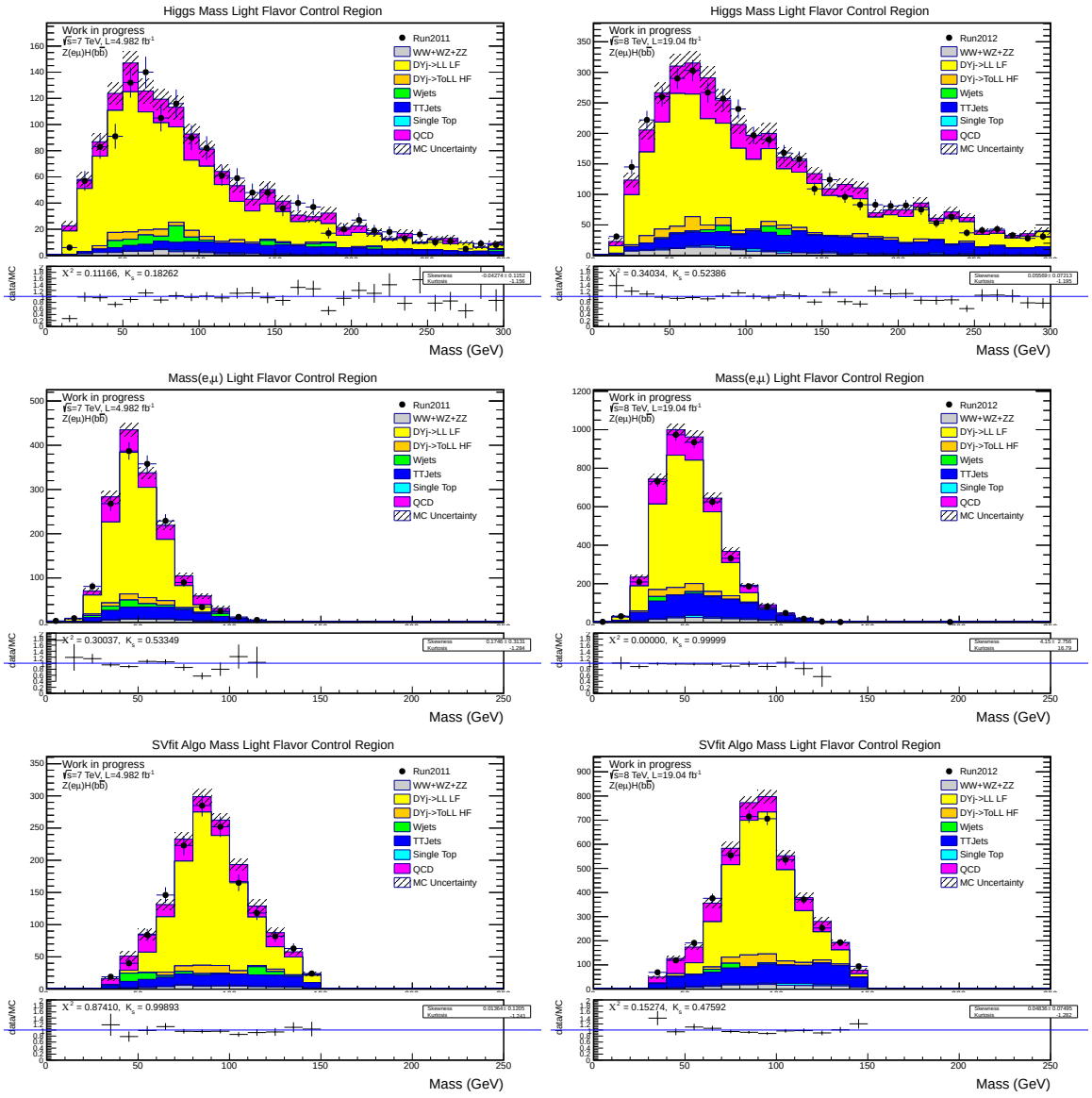

Figure 4.21: Distributions of Z+light control region for 7 TeV (left) and 8 TeV (right) data and MC samples. M(jj) (top), M(eμ) (middle), and $M_{\tau\tau}$ (bottom) are shown.

- **QCD corrections:** For the QCD correction, the relevant comparison is NNLO vs. NLO, where an uncertainty of 10% for ZH is estimated. The $VHbb$ group have checked that the jet veto efficiency in signal MC samples matches the NNLO calculation within this uncertainty.

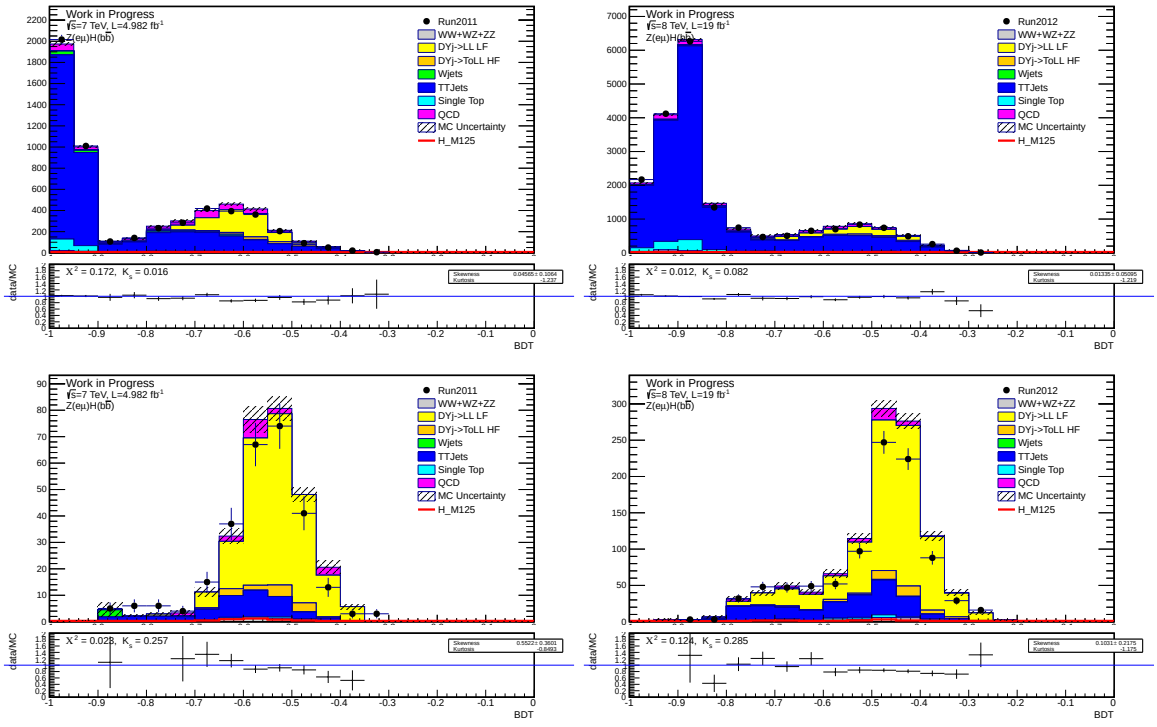

Figure 4.22: Distributions of BDT output in $t\bar{t}$(top) and Z+light (bottom) control regions for 7 TeV (left) and 8 TeV (right) data and MC samples.

- **Unclustered E^T_{miss}:** A 3% systematic uncertainty on the calibration of unclustered E^T_{miss} is applied.

Shape uncertainties are taken into account by propagating these uncertainties to the final BDT distribution. This procedure gives two recomputed BDT distributions per background and signal. The uncertainties define an envelope within which the BDT is allowed to vary. Uncertainties taken into account in this manner are listed below:

- **Jet Energy Scale:** the jet energy scale for each jet is varied by one standard deviation. The jet uncertainty is computed for each jet and is dependent on η and p_T. An uncertainty of 2% for signal and 3% for background is found. While the p_T of the jets

are not directly used by the BDT, variations in the Jet Energy Scale affect the Higgs mass. The daughters from the Higgs candidate are scaled up(down) by 3% and the Higgs mass is recomputed. The BDT is then recalculated using this new Higgs mass.

- **Jet Energy Resolution:** the resolution for each jet is smeared using the jetMET prescription [67]. The resolution is varied by 6% for central jets and 10% for forward jets. Variations in the Jet Energy Resolution affect the Higgs mass.

- ***b*-jet Tagging:** uncertainties were provided by the BTV POG(b-tagging experts) are 6% per b-jet, 12% per c-jet and 15% per fake (light quarks and gluons). The CSV distribution is varied within these uncertainties and the new CSV distributions are used as input to recalculate the BDT distribution.

- **Monte Carlo Statistics:** in addition to including statistical uncertainties in control region scale factors, the BDT distribution is varied by bin-by-bin by one sigma to account for statistical uncertainties.

4.12 Contribution from Other Higgs Channels

When Higgs production mechanisms other than associate production of a Higgs boson with a Z boson were studied, it was found that Higgs events from other Higgs processes and decay modes were anticipated to pass the BDT selection. Table 4.25 and Table 4.26 summarize the cross sections and branching fractions assumed in the other Higgs channels for several mass points. The cross sections are computed at Next-to-Next-to-Leading-Order, as described in Ref. 3. The MC samples are listed in Table 4.27 for 7 TeV

and Table 4.28 for 8 TeV. The number of events from bleed through Higgs events is shown in Tables 4.30 and 4.29. The number of expected events for 7 TeV and 8 TeV are included for several mass points. These numbers may be useful for future Higgs studies at higher center of mass energy and pileup scenarios. While this is not an exhaustive list of Higgs production and decay modes, it does include the dominant processes for the $\tau\tau b\bar{b}$ state. These events are more numerous than the number of expected $ZH \to \tau\tau b\bar{b}$ events, but the event counts are low compared to the overall expected background.

Table 4.25: $\sqrt{s} = 7$ TeV cross sections for Higgs masses between 110 and 150 GeV for the five leading production mechanisms: gluon-gluon fusion (gluglu), vector boson fusion (VBF), production in association with a W boson (WH), production in association with a Z boson (ZH), and production in association with a $t\bar{t}$ pair ($t\bar{t}H$).

Higgs Mass	gluglu	VBF	WH	ZH	$t\bar{t}H$
110	19.8	1.41	0.875	0.472	0.126
115	18.1	1.34	0.755	0.411	0.111
120	16.7	1.28	0.656	0.360	0.098
125	15.3	1.22	0.573	0.316	0.086
130	14.2	1.17	0.501	0.278	0.077
135	13.1	1.12	0.439	0.245	0.068
140	12.2	1.07	0.386	0.217	0.061
145	11.3	1.02	0.341	0.193	0.054
150	10.6	0.98	0.300	0.171	0.049

In the case of gluon fusion (GluGlu) production, there is not expected to be any jets associated with the production of the Higgs. For the $H \to WW$ samples, events are forced to decay to leptons (e,μ, τ). Jets come from other sources in the event like initial state radiation or final state radiation. Higgs bosons produced via Vector Boson Fusion (VBF) have a smaller production cross section than gluon fusion, but jets are expected in the event. However, these jets are expected to be forward so should be largely removed by the requirement $|\eta_j| < 2.5$. Higgs bosons produced in association with a top pair like direct

Table 4.26: $\sqrt{s} = 7\,\text{TeV}$ cross sections for Higgs for masses between 110 and 150 GeV for the five leading production mechanisms: gluon-gluon fusion (gluglu), vector boson fusion (VBF), production in association with a W boson (WH), production in association with a Z boson (ZH), and production in association with a $t\bar{t}$ pair ($t\bar{t}H$).

Higgs Mass	gluglu	VBF	WH	ZH	$t\bar{t}H$
110	25.0	1.81	1.06	0.587	0.189
115	23.0	1.73	0.917	0.512	0.166
120	21.1	1.65	0.797	0.449	0.147
125	19.5	1.58	0.697	0.394	0.130
130	18.1	1.51	0.610	0.347	0.116
135	16.8	1.45	0.535	0.307	0.103
140	15.6	1.39	0.471	0.273	0.092
145	14.6	1.33	0.416	0.242	0.082
150	13.6	1.28	0.368	0.216	0.074

production of $t\bar{t}$ have the same physics objects in the event as $\tau\tau b\bar{b}$ events. In the case where the Higgs boson decays to an electron or muon through a W or τ and a top quark also produces an electron or muon. The $e\mu$ candidate efficiency is higher in this channel because if one of the leptons from the Higgs is not reconstructed or is outside the fiducial acceptance of the detector, a lepton from a top quark decay could still be reconstructed. This means that the event will still produce an $e\mu$ candidate event.

This analysis was designed to be most sensitive to $ZH \to \tau\tau b\bar{b}$ at a Higgs mass of 125 GeV. This can be seen in two ways. First, the $ZH \to \tau\tau b\bar{b}$ events have to highest cut efficiency for the BDT Selection. Second, The $ZH \to \tau\tau b\bar{b}$ events are mostly found in the bins with the highest BDT values. This shows good control in the BDT training. The BDT values for various Higgs channels and the $ZH \to \tau\tau b\bar{b}$ signal samples are shown for a Higgs mass of 125 GeV in Figure 4.23 for 8 TeV and 7 TeV simulation. The plots have been divided by production mode as to not over crowd the plots. Note that there are high scale factors applied to each Higgs contribution so the distribution is visible. A study was

performed measuring the effect of these other Higgs channels on the final limit. It was found that due to their distribution in BDT value $ZH \to \tau\tau b\bar{b}$ still mainly drive the limit value. Even though bleed through Higgs events are more numerous, $ZH \to \tau\tau b\bar{b}$ events occupy the highest BDT value bins.

In Figure 4.24 the ZH production modes is isolated to show the Higgs channels that are closest in topology to $ZH \to \tau\tau b\bar{b}$. The wide ditau mass window of $35 < M_{\tau\tau} < 145$ allows many $H \to \tau\tau$ events to pass the BDT selection in the case where $ZH \to Z(b\bar{b})H(\tau\tau)$ when the Higgs mass is low. Note that these events tend to high BDT values as their topologies and physics objects are similar to $ZH \to \tau\tau b\bar{b}$.

The cross sections of the dominant Higgs production modes for 7 TeV and 8 TeV are show in Figure 4.25. While one would not characterize the curves for the various production cross sections as parallel, it is true that contributions from all of these production modes scale with ZH production. The Higgs events from Higgs channels other than $ZH \to \tau\tau b\bar{b}$ are included as signal in the final result. These events contribute to the sensitivity of this analysis to the SM Higgs in the $e\mu b\bar{b}$ final state. Higgs masses from 110 to 140 GeV are considered. A BDT is trained using $ZH \to \tau\tau b\bar{b} \to e\mu b\bar{b}$ only as the signal. How the kinematics of each signal sample effects the background and contributions from other Higgs production and decay modes is shown in Figure 4.26. While shapes of the distributions do change, the dominance of the original signal $ZH \to \tau\tau b\bar{b} \to e\mu b\bar{b}$ at high values of the BDT output remains across all Higgs masses.

Table 4.27: List of Higgs Fall11 Monte Carlo samples used in this analysis to estimate the contribution of other Higgs channels to the signal region for 7 TeV data. Square brackets indicate MC samples were produced in this range for a Higgs mass in 5 GeV increments.

Mode	Dataset
gluglu $\to$ H $\to \tau\tau$	GluGluToHToTauTau_M-[110-150]_7TeV-powheg-pythia6/Fall11-PU_S6_START42_V14B
gluglu $\to$ H $\to$ WW	GluGluToHToWWTo2LAndTau2Nu_M-[110,115,126,135]_7TeV-powheg-pythia6/Fall11-PU_S6_START42_V14B
gluglu $\to$ H $\to$ WW	GluGluToHToWWToLNuTauNu_M-[120,130,140,150]_7TeV-powheg-pythia6/Fall11-PU_S6_START42_V14B
gluglu $\to$ H $\to$ WW	GluGluToHToWWTo2L2Nu_M-[120,130,140,150]_7TeV-powheg-pythia6/Fall11-PU_S6_START42_V14B
gluglu $\to$ H $\to$ WW	GluGluToHToWWTo2Tau2Nu_M-[120,130,140,150]_7TeV-powheg-pythia6/Fall11-PU_S6_START42_V14B
VBF $\to$ H $\to \tau\tau$	VBF_HToTauTau_M-[110-150]_7TeV-powheg-pythia6-tauola/Fall11-PU_S6_START42_V14B
VBF $\to$ H $\to$ WW	VBF_HToWWTo2LAndTau2Nu_M-[110,115,126,135]_7TeV-powheg-pythia6/Fall11-PU_S6_START42_V14B
VBF $\to$ H $\to$ WW	VBF_HToWWTo2L2Nu_M-[120,130,140,150]_7TeV-powheg-pythia6/Fall11-PU_S6_START42_V14B
VBF $\to$ H $\to$ WW	VBF_HToWWToLNuTauNu_M-[120,130,140,150]_7TeV-powheg-pythia6/Fall11-PU_S6_START42_V14B
VBF $\to$ H $\to$ WW	VBF_HToWWTo2Tau2Nu_M-[120,130,140,150]_7TeV-powheg-pythia6/Fall11-PU_S6_START42_V14B
$t\bar{t}H \to H$	TTH_Inclusive_M-[110-140]_7TeV_pythia6/Fall11-PU_S6_START42_V14B
$VH\_t\bar{t}H \to H \to \tau\tau$	WH_ZH_TTH_HToTauTau_M-[110-150]_7TeV-pythia6-tauola/Fall11-PU_S6_START42_V14B
WH $\to$ H $\to$ b$\bar{\text{b}}$	WH_WToLNu_HToBB_M-[110-135]_7TeV-powheg-herwigpp/Summer11-PU_S4_START42_V11
ZH $\to$ H $\to \tau\tau$	ZH_ZToBB_HToTauTau_M-[110-150]_7TeV-pythia6-tauola/Fall11-PU_S6_START42_V14B

Table 4.28: List of Higgs ($M_{\mathrm{H}} = 125\,\mathrm{GeV}$) Summer12 Monte Carlo samples used in this analysis to estimate the contribution of other Higgs channels to the signal region for 8 TeV data. Samples were ran over 9 Higgs mass points from 110 to 150 GeV in 5 GeV increments with the exception of the inclusive $t\bar{t}H$ sample which was only available up to 140 GeV and WH $H \to b\bar{b}$ which was only available up to 135 GeV.

Mode	Dataset
gluglu $\to$ H $\to \tau\tau$	GluGluToHToTauTau_M-125_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A
gluglu $\to$ H $\to$ WW	GluGluToHToWWTo2LAndTau2Nu_M-125_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A
VBF $\to$ H $\to \tau\tau$	VBF_HToTauTau_M-125_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A
VBF $\to$ H $\to$ WW	VBF_HToWWTo2LAndTau2Nu_M-125_8TeV-powheg-pythia6/Summer12_DR53X-PU_S10_START53_V7A
$t\bar{t}H \to H$	TTH_Inclusive_M-125_8TeV_pythia6/Summer12_DR53X-PU_S10_START53_V7A
$VH\_t\bar{t}H \to H \to \tau\tau$	WH_ZH_TTH_HToTauTau_M-125_8TeV-pythia6-tauola/Summer12_DR53X-PU_S10_START53_V7A
$VH\_t\bar{t}H \to H \to WW$	WH_ZH_TTH_HToWW_M-125_8TeV-pythia6/Summer12_DR53X-PU_S10_START53_V7A
WH $\to$ H $\to$ b$\bar{\text{b}}$	WH_WToLNu_HToBB_M-125_8TeV-powheg-herwigpp/Summer12_DR53X-PU_S10_START53_V7A
ZH $\to$ H $\to \tau\tau$	ZH_ZToBB_HToTauTau_M-125_8TeV-pythia6-tauola/Summer12_DR53X-PU_S10_START53_V7C

Table 4.29: Number of expected Higgs events at $\sqrt{s} = 7\,\text{TeV}$ for $5.0\,\text{fb}^{-1}$ of integrated luminosity for masses between 110 and 150 GeV for the leading production mechanisms and their dominant decay modes to $e\mu b\bar{b}$ after the BDT Selection cuts have been applied. The expected number of $ZH \to \tau\tau b\bar{b}$ events is included for comparison. VH is the combination of WH and ZH channels. A dash(-) indicates no sample was available for the mass point in that particular channel.

Higgs Mass	ZH $\to$ H H(b$\bar{\text{b}}$)Z($\ell\ell$)	gluglu $\to$ H H $\to \tau\tau$	gluglu $\to$ H H $\to$ WW $\to 2\ell 2\nu$	VBF $\to$ H H $\to \tau\tau$	VBF $\to$ H inclusive	H $\to \tau\tau$	VH $\to$ H H $\to$ WW	WH $\to$ H H(b$\bar{\text{b}}$)W($\ell\nu$)
110	0.159	1.852	0.244	0.397	0.056	0.590	0.116	0
115	0.139	1.478	0.412	0.366	0.084	0.540	0.091	0
120	0.124	1.618	0.692	0.316	0.158	0.482	0.103	0
125	0.098	1.330	1.083	0.322	0.224	0.476	0.072	0.0034
130	0.077	1.011	1.347	0.227	0.285	0.444	0.058	0.0051
135	0.061	0.812	1.542	0.165	0.385	0.433	0.042	0.0008
140	-	0.594	2.050	0.135	0.460	0.450	0.029	-
145	-	0.346	-	0.079	-	-	0.021	-
150	-	0.176	2.637	0.050	0.604	-	0.011	-

Table 4.30: Number of expected Higgs events at $\sqrt{s} = 8\,\text{TeV}$ for $19.0\,\text{fb}^{-1}$ of integrated luminosity for masses between 110 and 150 GeV for the leading production mechanisms and their dominant decay modes to $e\mu b\bar{b}$ after the BDT Selection cuts have been applied. The expected number of $ZH \to \tau\tau b\bar{b}$ events is included for comparison. VH is the combination of WH and ZH channels.

Higgs Mass	ZH $\to$ H H(b$\bar{\text{b}}$)Z($\ell\ell$)	gluglu $\to$ H H $\to \tau\tau$	gluglu $\to$ H H $\to$ WW $\to 2\ell 2\nu$	VBF $\to$ H H $\to \tau\tau$	VBF $\to$ H H $\to$ WW $\to 2\ell 2\nu$	$t\bar{t}H$ inclusive	VH $\to$ H H $\to \tau\tau$	VH $\to$ H H $\to$ WW	WH $\to$ H H(b$\bar{\text{b}}$)W($\ell\nu$)
110	0.500	4.074	0.569	1.055	0.190	2.802	0.533	0.008	0.041
115	0.458	3.365	1.072	1.091	0.347	2.505	0.369	0.023	0.033
120	0.377	3.272	1.474	0.996	0.566	2.292	0.341	0.035	0.044
125	0.317	2.767	1.909	0.865	0.818	2.405	0.285	0.022	0.035
130	0.267	2.580	3.935	0.705	1.182	2.239	0.218	0.061	0.028
135	0.213	1.618	4.357	0.603	1.603	2.283	0.162	0.081	0.019
140	0.148	1.184	4.484	0.349	1.765	2.211	0.124	0.094	0.009
145	0.107	0.715	6.817	0.246	2.153	-	0.064	0.113	0.009
150	0.063	0.412	6.838	0.128	2.072	-	0.038	0.063	0.009

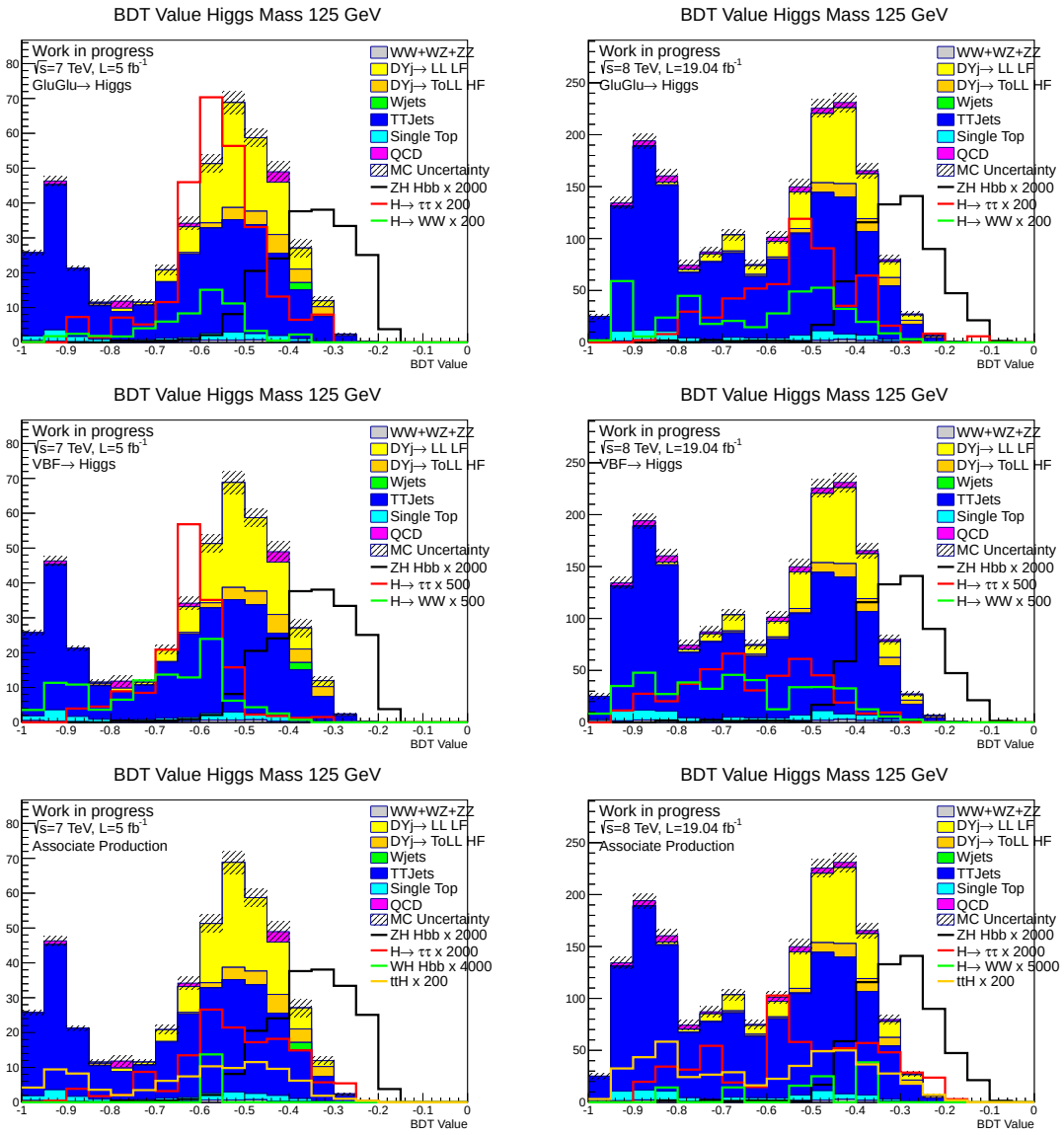

Figure 4.23: BDT distributions for 7 TeV (left) and 8 TeV (right) MC background with $ZH \to \tau\tau b\bar{b}$ (black) overlaid along with gluon-gluon fusion (top), vector boson fusion (center), and associate production (bottom). Note the difference in scale factors for the different Higgs channels.

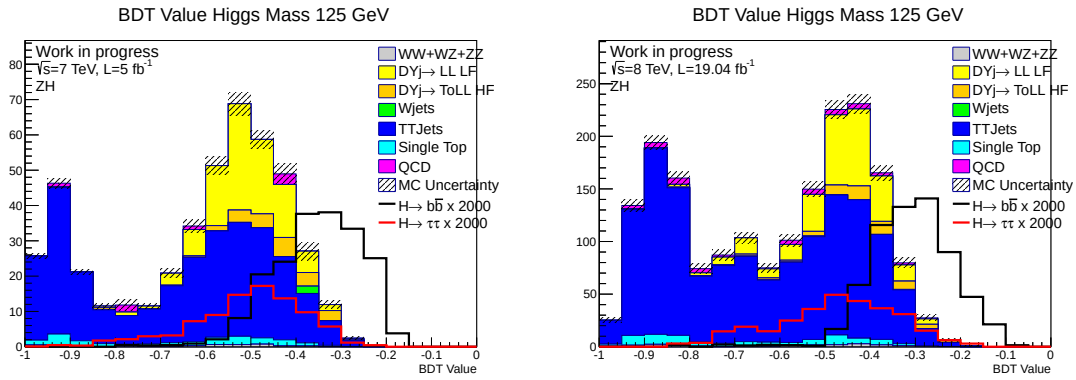

Figure 4.24: BDT distributions for 7 TeV (left) and 8 TeV (right) MC background with $ZH \to H \to b\bar{b}$ (black) overlaid along with $ZH \to H \to \tau\tau$ (red).

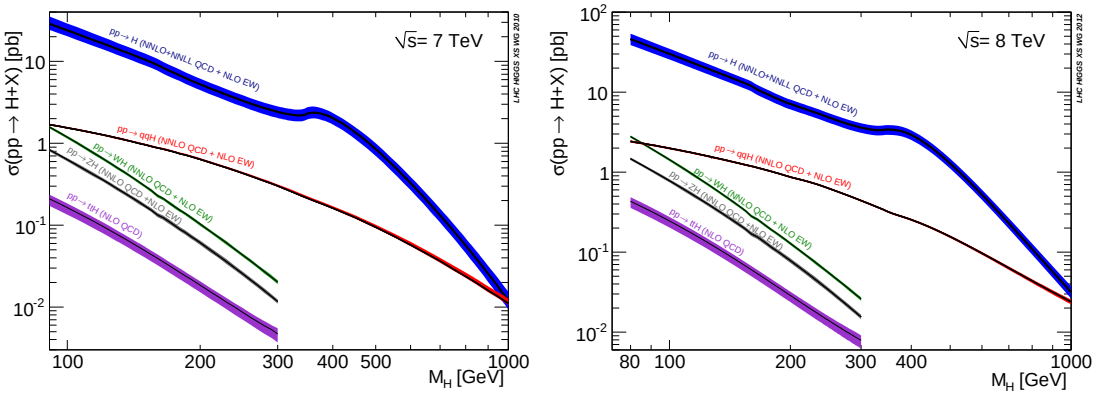

Figure 4.25: Production cross sections at $\sqrt{s} = 7\,\text{TeV}$ (left) and $\sqrt{s} = 8\,\text{TeV}$ (right) of the SM Higgs boson.

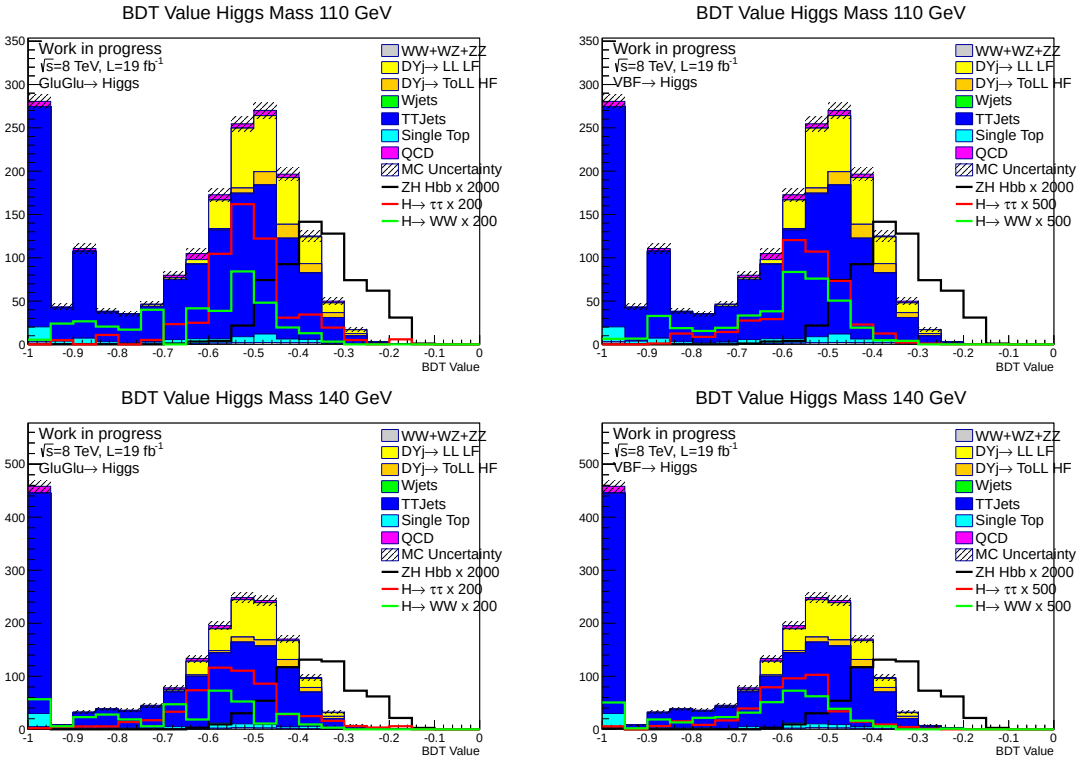

Figure 4.26: BDT distribution for 8 TeV background with $ZH \to H \to b\bar{b}$ (black) overlaid along with $H \to \tau\tau$ (red) and $H \to$ WW(green) for a Higgs boson produced via gluon-gluon fusion (left) and vector boson fusion (right) for a Higgs mass of 110 GeV (top) and 140 GeV (bottom).

4.13 Results

4.13.1 Signal and Background Estimates

This section summarizes the final signal and background estimates for the BDT analysis. Table 4.31 and 4.32 show the MC yields for signal and background BDT selection optimized for a Higgs mass of 125 GeV. Figure 4.27 shows the BDT output. The bins with the highest signal to background ratio do not show data to keep analyst bias from entering this analysis while optimization is on going.

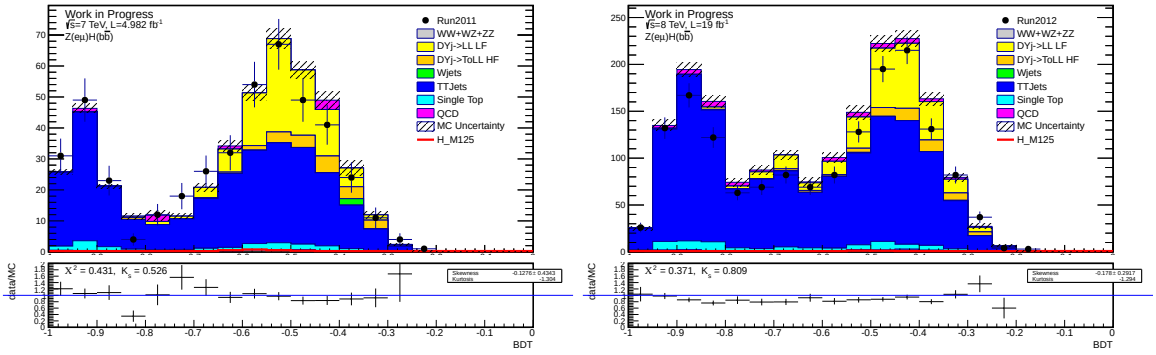

Figure 4.27: Distributions of the BDT value for 7 TeV (left) and 8 TeV (right) for data and MC samples.

Table 4.31: BDT Selection estimates for signal (m_H 125 GeV) and background for 5.0 fb^{-1} from 7 TeV MC.

Cut Variable	Signal	$t\bar{t}$	$Z+$ light	$Z+b\bar{b}$	Single Top	Diboson	W+jets	QCD
$\Delta R(e,\mu) > 0.3$	0.234	9455	2350	200	599	305	409	1208
$35 < M_{\tau\tau} < 145$	0.229	4642	2139	185	292	163	245	749
CSV0$>$ 0.244	0.207	4041	975	134	245	70.9	128	410
$P_\zeta^{\text{miss}} > 40$	0.098	291	119	22.6	19.5	4.15	2.08	7

Table 4.32: BDT Selection estimates
for signal (m_H 125 GeV) and background for 19.0 fb^{-1} from 8 TeV MC.

Cut Variable	Signal	$t\bar{t}$	$Z+$ light	$Z + b\bar{b}$	Single Top	Diboson	W+jets	QCD
$\Delta R(e,\mu) > 0.3$	0.745	41074	4744	497	2608	1245	606	3493
$35 < M_{\tau\tau} < 145$	0.723	19765	4424	454	1206	590	244	1894
CSV0 > 0.244	0.640	16972	1505	277	1017	227	97.1	941
$P_{\zeta}^{\text{miss}} > 40$	0.317	1319	310	60.4	73.3	17.7	2.29	46

4.13.2 Upper Limit Calculation

Limits are determined at the 95% CL following the modified frequentist CLs approach using the roofit base limit calculation tool *combine* as described in Ref. 78. Limits are calculated according to the formulas in Ref. 79, known as the so called Asymptotic likelihood method. These limits are calculated using the maximum likelihood fit to the background only and background plus Higgs boson signal BDT distributions.

To quantify the absence of signal exclusions limits are set on the Higgs boson hypothesis. We define a test statistic q_μ as, used to determine how signal-like or background-like the data is. q_μ is defined as:

$$q_\mu = -2 \ln \frac{\mathcal{L}(\text{obs}|\mu \cdot s(\hat{\theta_\mu}) + b(\hat{\theta_\mu}))}{\mathcal{L}(\text{obs}|\mu \cdot s(\hat{\theta}) + b(\hat{\theta}))}, \quad 0 \leq \hat{\mu} < \mu \tag{4.4}$$

where parameter μ is the signal strength modifier, "obs" stands for the observed data, b is the number and distribution of background events and θ represents the full suite of nuisance parameters. The expected SM Higgs cross section is modified by μ in the case where signal cross section deviates from SM predictions. $\mu \cdot s + b$ is the signal plus background hypothesis. Here "obs" is either the actual experimental *data* or *pseudo-data* used to report the expected limit. The value $\hat{\theta_\mu}$ maximizes the likelihood in the numerator

for a given μ. The likelihood:

$$\mathcal{L}(\text{obs}|\mu \cdot s(\theta) + b(\theta)) = \mathcal{P}(\text{obs}|\mu \cdot s(\theta) + b(\theta) \cdot p(\tilde{\theta}|\theta)) \tag{4.5}$$

Where $\mathcal{P}(\text{obs}|\mu \cdot s(\theta) + b(\theta))$ is a product of probabilities over all bins of discriminant variable distributions and $p(\tilde{\theta}|\theta)$ is the probability density function for all nuisance parameter measurements. Each independent source of systematic uncertainty is assigned a nuisance parameter θ_i that is included in these probability density functions.

We adopt the modified frequentist construction CL_s for calculating the exclusion limit. Two probabilities are associated with the data the probability to obtain a value for the test statistic q_μ larger than the observed value for a) signal + background and b) background only hypothesis. The confidence level is reported at the ratio:

$$CL_s = \frac{CL_{s+b}}{CL_b} = \frac{P(q_\mu \geq q_\mu^{obs}|\mu \cdot s + b)}{P(q_\mu \geq q_\mu^{obs}|b)} \tag{4.6}$$

If $CL_s \geq \alpha$ for $\mu = 1$, we determine that the SM Higgs boson is excluded at the $1 - \alpha$ confidence level. The expected limit, at 95% confidence level, is shown in Figure 4.28 in σ_{obs}/σ_{SM} for 7 TeV and 8 TeV and their combination. The 7 TeV has approximately 1σ more data than expected, while the 8 TeV shows approximately 2σ less data events than predicted by the simulation. The broad deviations from the expected limit are consistent with the poor resolution from b-jets and tau leptons expected in this channel. The dominant systematic uncertainty from comes from statistical uncertainties. The combination of 2011 and 2011 MC increases the statistics significantly reducing the uncertainty on the expected limit.

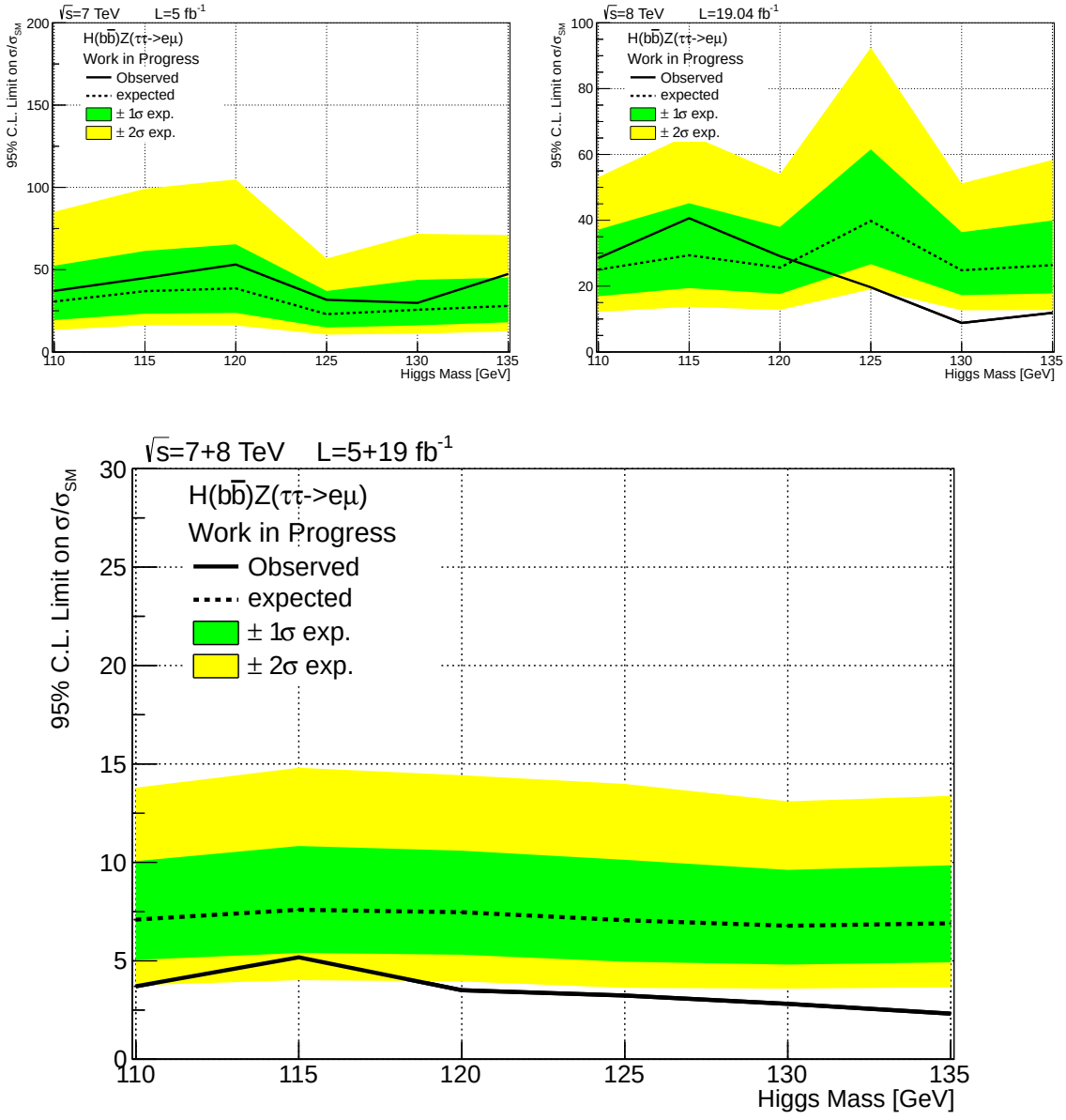

Figure 4.28: Observed and expected 95% confidence level upper limits on the ratio to the SM Higgs cross section for 7 TeV (left) and 8 TeV (right) data and MC samples and their combination (bottom). The black dashed line shows the expected limit from simulation with 1σ uncertainty bands in green and 2σ uncertainty bands in yellow. The solid black line is observed limit from data.

Chapter 5

Conclusion

The search for a Higgs boson when the Higgs is produced in association with a Z boson where the Higgs decays into b quarks and the Z decays into tau leptons has been presented. The case where both taus decay leptonically, one into an electron the other into a muon, has only been considered. This is the first in depth look into the $e\mu b\bar{b}$ final state at the LHC. The large overlap between other Higgs production and decay modes and the originally targeted $ZH \rightarrow \tau\tau b\bar{b}$ channel indicates that careful cross checks must be performed as the branching ratios and cross section of the newly discovered Higgs boson are measured in the new physics run at the LHC. Future coupling measurements at higher Pileup will have to disentangle contributions from many Higgs production and decay modes.

Expected 95% CL upper limits were have been presents. Many of the studies and much of analysis strategy can be applied to other tau decay modes. A fully developed picture of the $\tau\tau b\bar{b}$ final state using all tau decay modes and all 7 and 8 TeV data may be instructional for shaping CMS analysis strategy for the physics run after Long Shutdown

1. The additional luminosity at a high center of mass energy provided by the LHC in the second run will open up many opportunities for probing the characteristics of the Higgs boson at 125 GeV and the search for additional particles in the Higgs sector as one Higgs boson is the simplest case for electroweak symmetry breaking.

The sensitivity to the SM Higgs boson with currently available statistics was presented in the last section of Chapter 4. Current projections of the LHC program estimate a total integrated luminosity of $300\,\mathrm{fb}^{-1}$ [80] with an even higher integrated luminosity of $3000\,\mathrm{fb}^{-1}$ is the High Luminosity LHC project is approved. A factor of over 100 in integrated luminosity would open the opportunity for measurement of Higgs physics in the $\tau\tau b\bar{b}$ final state.

Bibliography

[1] P. W. Higgs, Phys.Lett. **12**, 132 (1964).

[2] CMS Collaboration, Physics Letters B **716**, 30 (2012).

[3] LHC Higgs Cross Section Working Group, CERN-2011-002 (CERN, Geneva, 2011).

[4] F. Halzen and A. D. Martin, *Quarks and Leptons: an Introductory Course in Modern Particle Physics*, Wiley, New York, USA, 1984.

[5] S. Weinberg, Phys.Rev.Lett. **19**, 1264 (1967).

[6] S. Fukuda et al., Phys.Lett. **B539**, 179 (2002).

[7] M. Gell-Mann, Phys.Lett. **8**, 214 (1964).

[8] T.-D. Lee and D. Pines, Phys. Rev. **92**, 883 (1953).

[9] R. Feynman, Phys.Rev. **80**, 440 (1950).

[10] S. Glashow, Nucl.Phys. **22**, 579 (1961).

[11] A. Salam and J. C. Ward, Phys.Lett. **13**, 168 (1964).

[12] S. Weinberg, Phys.Rev.Lett. **31**, 494 (1973).

[13] M. Peskin and D. Schroeder, *An introduction to quantum field theory*, The Advanced Book Program, Addison-Wesley Publishing Company, 1995.

[14] E. Noether, *Invariante Variationenprobleme*, M. A. Tavel's English translation of Noether's Theorems, 1918.

[15] G. Arnison et al., Physics Letters B **122**, 103 (1983).

[16] M. Banner et al., Physics Letters B **122**, 476 (1983).

[17] F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321 (1964).

[18] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, Phys. Rev. Lett. **13**, 585 (1964).

[19] J. Goldstone, A. Salam, and S. Weinberg, Phys. Rev. **127**, 965 (1962).

[20] J. Ellis, J. Espinosa, G. Giudice, A. Hoecker, and A. Riotto, Phys.Lett. **B679**, 369 (2009).

[21] R. Barate et al., Phys.Lett. **B565**, 61 (2003).

[22] Tevatron New Physics Higgs Working Group, CDF Collaboration, D0 Collaboration, (2012).

[23] CMS Collaboration, CMS Physics Analysis Summary **CMS-PAS-HIG-13-005** (2013).

[24] CMS Collaboration, Phys. Rev. Lett. **110**, 081803 (2013).

[25] O. Brüning et al., editors, *LHC Design Report*, CERN-2004-003, Jun. 2004.

[26] L. Evans and P. Bryant, JINST , S08001 (2008).

[27] M. Conte and W. W. MacKay, *An Introduction to the Physics of Particle Accelerators*, World Scientific, 2008.

[28] CERN, CERN's accelerator complex, 2008.

[29] CMS Collaboration, JINST **3**, S08004 (2008).

[30] David Barney, *CMS Slice*, Number CMS-doc-5582, 2012.

[31] M. Fouz, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **573**, 260 (2007), Proceedings of the 7th International Conference on Position-Sensitive Detectors.

[32] A. Dominguez, CMS Technical Design Report for the Pixel Detector Upgrade, Technical Report CERN-LHCC-2012-016. CMS-TDR-11, CERN, Geneva, 2012.

[33] CMS Collaboration, Technical Design Report CMS (2000).

[34] Lucas Taylor, Silicon Strips, 2011.

[35] M. Raymond et al., JINST **7**, C01033 (2012).

[36] Description and performance of the CMS track and primary vertex reconstruction, CMS Paper 2011/001, 2011.

[37] CMS Collaboration, *The CMS electromagnetic calorimeter project: Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 1997.

[38] P. Aspell et al., Results from the 1999 beam test of a preshower prototype, Technical Report CMS-NOTE-2000-001, CERN, Geneva, 2000.

[39] K. W. Bell et al., IEEE Trans. Nucl. Sci. **51**, 2284 (2004).

[40] CMS Collaboration, Journal of Instrumentation **5**, T03012 (2010).

[41] T. Christiansen, CMS Conference Report , 6 (2006).

[42] Precise Mapping of the Magnetic Field in the CMS Barrel Yoke using Cosmic Rays, Technical report, 2010.

[43] CMS Collaboration, *The CMS muon project: Technical Design Report*, Technical Design Report CMS, CERN, Geneva, 1997.

[44] CMS Collaboration, JINST **5**, T03018 (2010).

[45] G. L. Bayatyan et al., Cms tridas project: Technical design report, volume 1: The trigger systems, Technical report.

[46] S. Cittolin, A. Rcz, and P. Sphicas, *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*, Technical Design Report CMS, CERN, Geneva, 2002.

[47] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, Phys. Rev. Lett. **100**, 242001 (2008).

[48] CMS Collaboration, Search for the standard model higgs boson decaying to bottom quarks, Technical report, 2011.

[49] ATLAS Sensitivity to the Standard Model Higgs in the HW and HZ Channels at High Transverse Momenta, Technical Report ATL-PHYS-PUB-2009-088, CERN, Geneva, 2009.

[50] R. Wilken, K. Bloom, R. Clare, A. Dominguez, CMS Analysis Note **AN-12-286** (2012).

[51] CMS Collaboration, CMS Note **2011/430** (2011).

[52] P. D. Group, Phys. Rev. D **86**, 010001 (2012).

[53] H.-J. Yang, B. P. Roe, and J. Zhu, Nucl.Instrum.Meth. **A555**, 370 (2005).

[54] Z. Was, Nucl. Phys. Proc. Suppl. **98**, 96 (2001).

[55] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, Journal of High Energy Physics **2011**, 1 (2011).

[56] T. Sjostrand, L. Lonnblad, S. Mrenna, and P. Z. Skands, (2003).

[57] S. Agostinelli et al., Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506**, 250 (2003).

[58] CMS Collaboration, CMS Physics Analysis Summary **CMS-PAS-PFT-09-001** (2009).

[59] Standard Model Cross Sections, Standard Model Cross Sections for CMS at 7 TeV, 2012.

[60] W. Erdmann and PSI, CMS Internal Note **CMS-IN-11-014** (2011).

[61] A. S. Perloff, Pileup measurement and mitigation techniques in cms, Technical Report CMS-CR-2012-191. CERN-CMS-CR-2012-191, CERN, Geneva, 2012.

[62] CMS Collaboration, Electron reconstruction and identification at $\sqrt{s} = 7$ TeV, CMS Physics Analysis Summary CMS-PAS-EGM-10-004, 2010.

[63] CMS Collaboration, Tag and probe methodology for analyses using electrons and photons, CMS Note CMS-AN-12-116, 2012.

[64] CMS Collaboration, Electron performance with 19.6 fb^{-1} of data collected at $\sqrt{s} = 8$ tev with the cms detector., Technical report, 2013.

[65] M. Cacciari, G. P. Salam, and G. Soyez, JHEP **04**, 063 (2008).

[66] CMS Collaboration, CMS Physics Analysis Summary (2009).

[67] CMS Collaboration, CMS Physics Analysis Summary **CMS-PAS-JME-10-014** (2010).

[68] CMS Collaboration, CMS Physics Analysis Summary **CMS-PAS-BTV-09-001** (2009).

[69] CMS Collaboration, Identification of b-quark jets with the cms experiment, 2010.

[70] CMS Collaboration, Jet performance in pp collisions at $\sqrt{s}$=7 TeV, CMS Physics Analysis Summary CMS-PAS-JME-10-003, 2010.

[71] J. Alexander, L. Gibbons, and A. Khukhunaishvili, Towards a robust met significance with particle flow, CMS Note 2010/400, 2010.

[72] D. Acosta et al., Phys.Rev. **D72**, 072004 (2005).

[73] Tau Working Group, Likelihood based mass reconstruction in events containing tau leptons, 2011.

[74] L. Breiman, Machine Learning **45**, 5 (2001).

[75] Y. Freund and R. E. Schapire, Journal of Computer and System Sciences **55**, 119 (1997).

[76] A. Hocker et al., PoS **ACAT**, 040 (2007).

[77] CMS Collaboration, CMS Physics Analysis Note **AN-12-181** (2011).

[78] CMS Collaboration, Combined results of searches for a Higgs boson in the context of the standard model and beyond-standard models, CMS Physics Analysis Summary CMS-PAS-HIG-12-008, 2012.

[79] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, Eur.Phys.J. **C71**, 1554 (2011).

[80] CMS Collaboration, CMS Note **2016/006** (2012).

[81] CMS Collaboration, JINST **0803**, S08004 (2008).

[82] L. Evans and P. Bryant, Journal of Instrumentation **3**, S08001 (2008).

[83] CMS Collaboration, *CMS Physics Technical Design Report Volume I: Detector Performance and Software*, Technical Design Report CMS, CERN, Geneva, 2006.

[84] M. Aldaya et al., Discovery potential and search strategy for the standard model Higgs boson in the $H \to ZZ^\star \to 4\mu$ decay channel using a mass-independent analysis, CMS Note 2006/106, 2006.

[85] A. Brandt et al., Nucl. Phys. B **514**, 3 (1998).

[86] W. Buchmüller and D. Wyler, Phys. Lett. B **177**, 377 (1986).

[87] CMS Collaboration, J. Phys. G **34**, 995 (2007).

[88] ALEPH, CDF, D0, DELPHI, L3, OPAL, SLD Collaborations, the LEP Electroweak Working Group, the Tevatron Electroweak Working Group, and the SLD Electroweak and Heavy Flavour Groups, Precision electroweak measurements and constraints on the Standard Model, 2010.

[89] I. Bertram et al., A recipe for the construction of confidence limits, Technical Report TM-2104, Fermilab, 2000.

[90] L. Moneta et al., The RooStats Project, in *13^{th} International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT2010)*, SISSA, 2010, PoS(ACAT2010)057.

[91] CMS Collaboration, Search for pair production of first-generation scalar leptoquarks in pp collisions at $\sqrt{s} = 7$ TeV, Submitted to *Phys. Rev. Lett.*, 2010.

[92] CMS Collaboration, Performance of cms muon reconstruction in pp collision events at $\sqrt{s} = 7$TeV, Submitted to *J. Inst.*, 2012.

[93] ATLAS Collaboration, Search for the Higgs boson in the $H \to$WW(*)$\to \ell^+\nu\ell^-\bar{\nu}$ decay channel in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector, Submitted to *Phys. Rev. Lett.*, 2011.

[94] S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.), Handbook of LHC Higgs cross sections: 2. differential distributions, CERN Report CERN-2012-002, 2012.

[95] CMS Higgs to Tau Working Group, CMS Analysis Note **AN-11-390** (2011).

Appendix A

Installation of Low Voltage System of the CSCs

By the summer of 2007, all 468 CSC chambers were installed on the iron return yoke disks on the plus and minus side. However, services like low voltage (LV) and cooling manifolds still needed to be installed. During the Magnet Test Cosmic Challenge (MTCC) [41] of the previous summer, full services were installed for a wedge of the inner most disk on the plus side ME+1. Many lessons were learned, so that ME+1 has a bit of a different configuration than the rest of the CSC system. For example, the peripheral crates on X5 face a different direction on ME+1 than the rest of the detector. Improved safety and ease of working was experienced by turning these crates around. A diagram of the LV system and the five working levels of CMS are shown in Figure A.1.

In the summer of 2007, a team of rotating personnel, including myself, were utilized to install cooling manifolds and low voltage services to the CSC chambers. Much work had

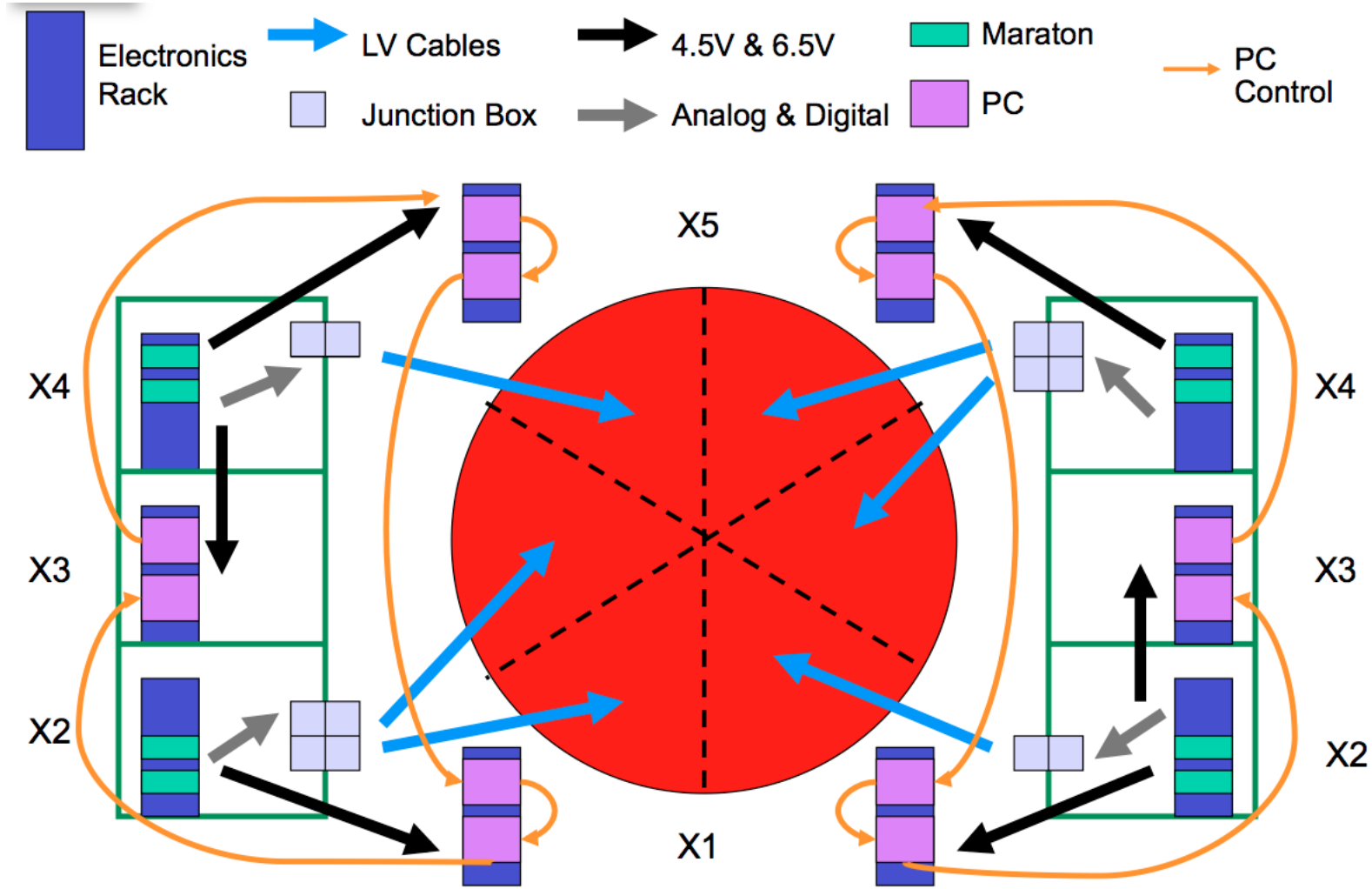

Figure A.1: A diagram of the LV system and the five levels of CMS. The legend at the top lists major components. Arrows are cables

already been done before my arrival. The maratons and AWG4/AWG2 cables [43] , blue LV cables pictured in Figure A.2 (right), had already been installed from the CSC chambers to the junction boxes. These cables needed to be cut to size, laid orderly in cable trays and connected to the junction boxes. A junction box with this work completed is shown in Figure A.2 (right). Cable trays were installed to ensure cables are kept in an orderly configuration as they run from the chamber to the junction box. LV is supplied to the junction box via 4.5 and 6.5 V $\sim$ 50 lb copper cables that run from maratons to the junction boxes. The diagram in Figure A.2 cuts off these cables under the above the junction boxes. In reality, bending these cables to tuck neatly between the two rows of junction boxes was a difficult task.

The other end of these large heavy cables were connected to maratons. Given the large magnetic fluxes expected in the cavern these cables needed to be held securely where they were installed at the maraton. At the junction box, the structure of the junction box itself hold the cables in place in the magnetic field. At the maraton, however, a support structure is not part of the maraton. Metal rails were installed on the electronics rack to hold the 4.5 V and 6.5 V cables in place. Metal brackets were placed around the cables. The metal brackets slid into groves on the rails. Tension between the bracket and rail was increased by tightening a bolt on the top of the metal bracket. This system was able to hold these cables in place with much more force than standard cable ties. A maraton after installation is pictured in Figure A.3. Also note that the cables are to one side to allow access to the PCMB and CANbus telecommunications line.

Control of the LV system is done via a CANbus system via PCMB on the maraton.

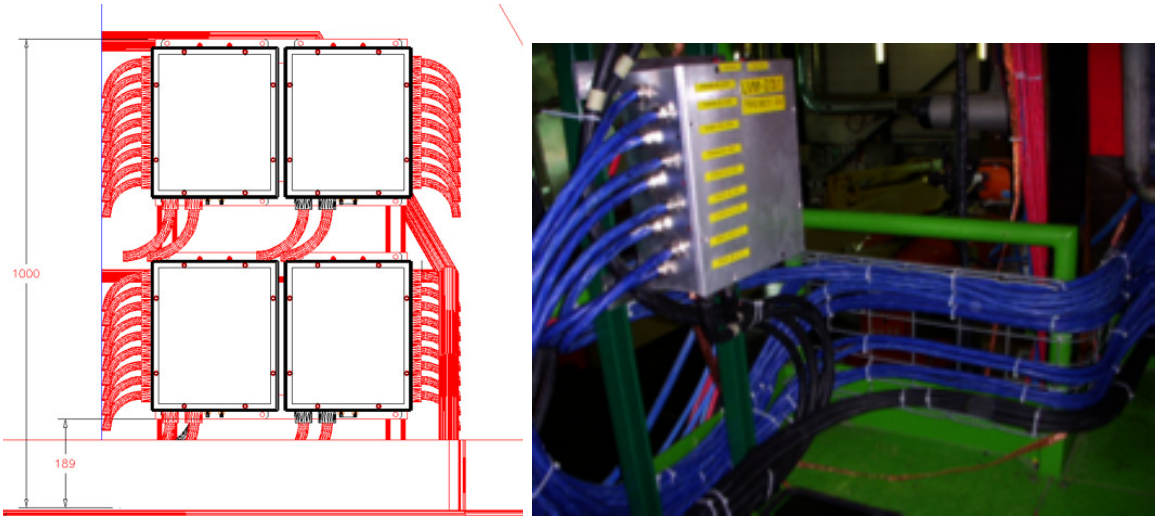

Figure A.2: A diagram of four junction boxes and cables (right) and a picture of a junction box (left) showing all cabling connections. Cable trays, wire baskets attached to the green railing, hold the cables off of the working platform. Cable ties are used to secure cables to the cable tray.

Figure A.3: The back of a maraton with 4.5 V and 6.5 V cables connected is shown. Metal rails attached to the electronics rack (blue) can bee seen along with metal brackets that securely hold the black cables in place.

These are industry standard telephone cables that were installed in a loop on each side of each CSC disk. These cables were much more fragile and had to be installed with care. Kinks in these cables could damage them and interrupt communications. These cables were installed in corners of cable trays and in nooks between larger cables to protect them. As these were industry standard cables, they were purchased in 10 and 30 m lengths. Extra cable had to be properly bundled and placed near the electronics racks. Care was taken so the extra length would not be in the way during future CSC installation efforts.

CSC detector information is transferred out of the experimental cavern into the counting room via fiber optic links. These high speed links are connected to the CSC peripheral crate boards via 1m and 2m patch panel fibers. Each chamber has a DAQ Motherboard (DMB) which produces the optical information. The patch panel fibers must be installed in a coherent way so signals received in the counting room can be associated with specific chambers. These fibers were labeled before being installed in the cavern. Given the delicate nature of these fibers, cable ties could not be used to secure them out of the way from future CSC interventions. These fibers were bundled instead using lengths of velcro. The patch panels, orange patch panel fibers and black velcro holding bundles of fibers are pictured in Figure A.4. Completion of the installation of these fibers in the experimental cavern was carried out by myself during spring 2008.

To avoid potential problems with customs during off-peak hours, the CSC peripheral crate test facility was moved from the ISR building on the main CERN site to building 904 in Prvessin France. This avoided the need to take peripheral crate boards over international boarders in the event an emergency replacement was called for in the middle of

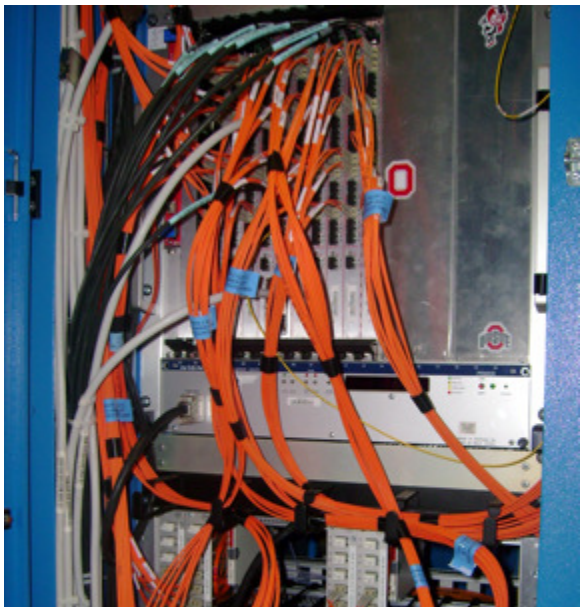

Figure A.4: Orange 1 m fiber optic jumper cables are pictured above. These are connected to patch panels in the lower part of the electronics rack and held in neat bundles by black velcro.

the night. This test facility did not have gas nor HV services, so it was not capable of taking cosmic events, but it did have all necessary equipment to load the latest version of firmware and test peripheral crate electronics boards before they were installed at point 5. Each chamber has a corresponding DMB and Trigger Motherboard (TMB) in a peripheral crate. In addition to DMBs and TMBs peripheral crates has a Clock Control Board (CCB), Muon Port Card (MPC) and VERSA Module Eurocard (VME) crate controller as shown in Figure A.5. The 9 chambers in each peripheral crate make up a trigger sector. A brief description of the function of these boards is below, but more information can be found in Ref. 43.

The TMB combines cathode and anode information. If cathode and anode information is found to be consistent with a muon hit, the information is forwarded to the

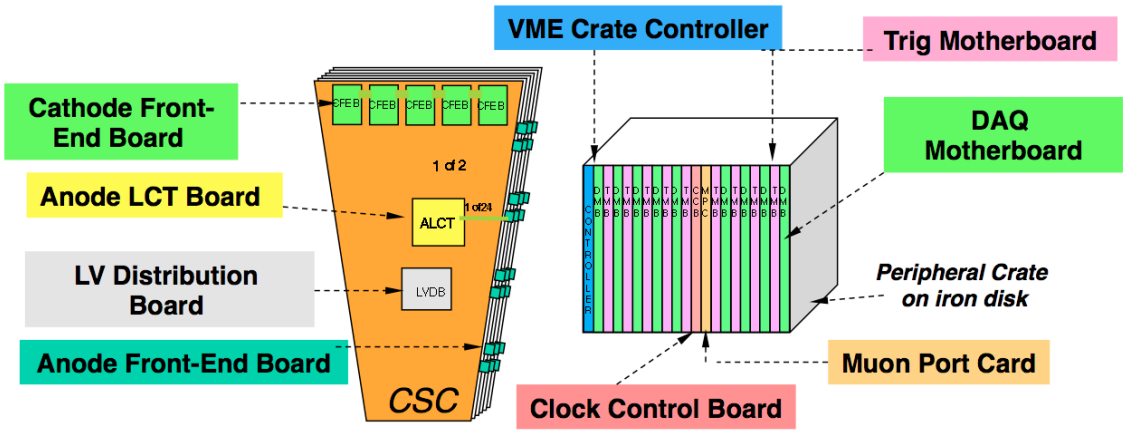

Figure A.5: A diagram of the peripheral crate and on chamber electronics boards are shown.

MPC as a Local Charged Track (LCT). The CCB is responsible to providing a hard-reset to resynchronize all boards of the peripheral crate. It also distributes a 80 MHz clock signal to the MPC. The MPC sorts up to 18 LCTs per event to find the best 3 LCTs to forward to the counting room. Recall in the trigger section of chapter 3, not all detector information is used by the L1 trigger. The sorting performed by the MPC helps to reduce the amount of data processed by the L1. On chamber cathode information is only readout by the DMB if a L1-accept is received from the central DAQ of CMS. The DMB also acts as interface to Slow control. The interfaced functions include downloading of the FPGA, downloading calibration information, downloading commands for turning off bad channels, monitoring of low voltage levels and temperature. The VME reads data from the CSCs and performs one step in the trigger chain, to decide if a given event contains high-pT muons and should be recorded for offline analysis. Analog charge measurements and digital track segments are passed from electronics mounted on the CSCs into the VME boards through the peripheral crate backplane, merging data from 9 CSC chambers (1 trigger sector) in each peripheral

crate. If the resulting pattern matches a template, the data continue to a sector processor in the control room.

Appendix B

Operation of the CMS Detector

Working from the test facility in building 904, the commissioning phase of the CSC started. This involved writing detailed procedures for loading the firmware onto peripheral crate electronics boards when the peripheral board software was still in the development phase. Each board needed to be tested before being installed in the experimental hall. Boards that failed tests the first time were often found to be functional after a power cycle or being reseated in the peripheral crate. Slot 19 of the peripheral crate backplane was found to have problems more often than other slots.

After the peripheral crates were fully populated at point 5, along with gas and cooling services, commissioning in the underground control room commenced. Much of the system had yet to be automated. Pressure of the water cooling circuit had to be checked visually from gauges in the experimental hall. Valves in the cooling circuit were checked every shift to ensure that values had not been inadvertently closed as installation work was still on going in the cavern. During the 2011, run firmware could be written in parallel to all

peripheral crates, this feature was not available at the time. Each cathode strip plane has one extra strip of constant width running next to the set of readout strips. This strip can be pulsed from CSC operation software. This feature was key to developing CSC firmware and commissioning the CSCs. The EMU Collaboration was running two shifts per day at this time. The first shift was used to implement changes in firmware and systematically check every single electronics board installed in the CSC readout. The first shift tried to include as many chambers as possible in the readout system. Fixes were sometimes to write the flash or institute a hard reset on the crate. Other problems were fixed by re-seating a cable or swapping cables to isolate problems. The HV was usually only powered on for the second shift. The second shift kept the system in a more stable state. The goal of the second shift was to verify changes made during the day and take several larger cosmic runs used for alignment and timing.

Commissioning started in 2008 in the underground control room. This was more convenient because along with visual inspections of turbines, water cooling and gas services, underground control room shifts often involved going to the experimental cavern to replace a peripheral crate board or visually confirm response of a board via LED lights on the electronic board. The system turn on procedure was prescriptive and done by a series of manual button clicks that sent soap messages. Each crate was turned on individually, but care needed to be taken so no two chambers were powered up in the same maraton at the same time. As the system grew more stable, operations were moved upstairs to the CMS control room at point 5. These shifts would start by going underground to start high voltage (HV) servers and perform a visual inspection of the system but then operations

were conducted for the most part above ground. In preparation for data taking, the CSCs provided muon triggers from cosmic rays for monthly global CMS runs. With the exception of three day global runs, the HV was turned off at the end of the evening shift because no one was there overnight to ensure the safety of the detector.

An extended global run, referred to as CRUZET (Cosmic RUn at ZEro Tesla), had 24-7 shifts. While CSC operations were run from above ground, walk throughs were performed at the beginning of every shift to verify the pressure of the water cooling loop on each endcap and check for noisy turbine fans. Two CSC shifters were used: one for the Detector Control System (DCS), one for Data Quality Monitoring (DQM).

The timing for CSC triggers was designed with collisions in mind. After the LHC incident September 19, 2008, an extended cosmic run was foreseen. The timing for the upper half of the CSC detector was changed so cosmic muons traveling from the top of the CMS detector down would produce triggers in the muon system. Nearly all of CMS was operational by Fall 2008 during the Cosmic Run At Four Tesla (CRAFT). Approximately 270 million cosmic triggers were recorded while the magnet was operating at 3.8 T. About a fifth of these triggers were triggered by the CSCs [44]. These events were used to measure the efficiency and resolution of the chambers and study the electronic noise of the readout system.

When beam returned in November 2009, it was not known how clean nor how well controlled the beam would be. As a precaution, the CSC system was incrementally turned on. First the outer rings of the CSC were ramped up to full HV. This put the CSC is a state referred to as "outer on". The DCS information is displayed in Figure B.1. If the

rate from the beam was then verified to be reasonable, the inner ring of the CSCs would be ramped up.

Another HV state was create for beam splashes. Beam splashes are when the beam is incident on collimators 150m down the LHC beam pipe from the CMS detector. The LHC beam was accelerated through section by section before circulating all 27 km of the LHC. Beam splash events were used to ensure that the CSC was timed in properly as a spray of particles was expected to hit the CSC at the same time. A beam splash is pictured in Figure B.2. The high influx of particles prompted CSC experts to lower HV setting. A table of beam conditions and corresponding HV settings for difference chambers is listed in Table B.1.

Preparation for first collisions led to many more HV transitions than originally anticipated for the CSC system. During design three states: OFF, ON, and STANDBY were envisioned. Due to unknown beam conditions of the new LHC accelerator, two more states were created: Outer ON and Inner ON. practice runs for these HV transitions were made leading to many HV transitions in a day. It was found that for the ME1/1 chambers the CAEN HV main frame would need to be reset. This caused the ME1/1 chambers to

Table B.1: HV settings for various beam conditions. ME1/1 refers to the inner most chambers of the CSC system. Inner refers to the inner ring of chambers: ME1/2, ME2/1, ME3/1, and ME4/1. Outer refers to the outer ring of chambers: ME1/3, ME2/2, and ME3/2.

Beam Condition	ME1/1 (V)	Inner (V)	Outer (V)
Beam Splash	2300	2700	3000
Beam Injection	2600	3000	3000
Stable Beams	2600	3000	3600
Outer ON $<$ 100 Hz	2600	3600	3600
Inner ON $<$ 100 Hz	2900	3600	2600

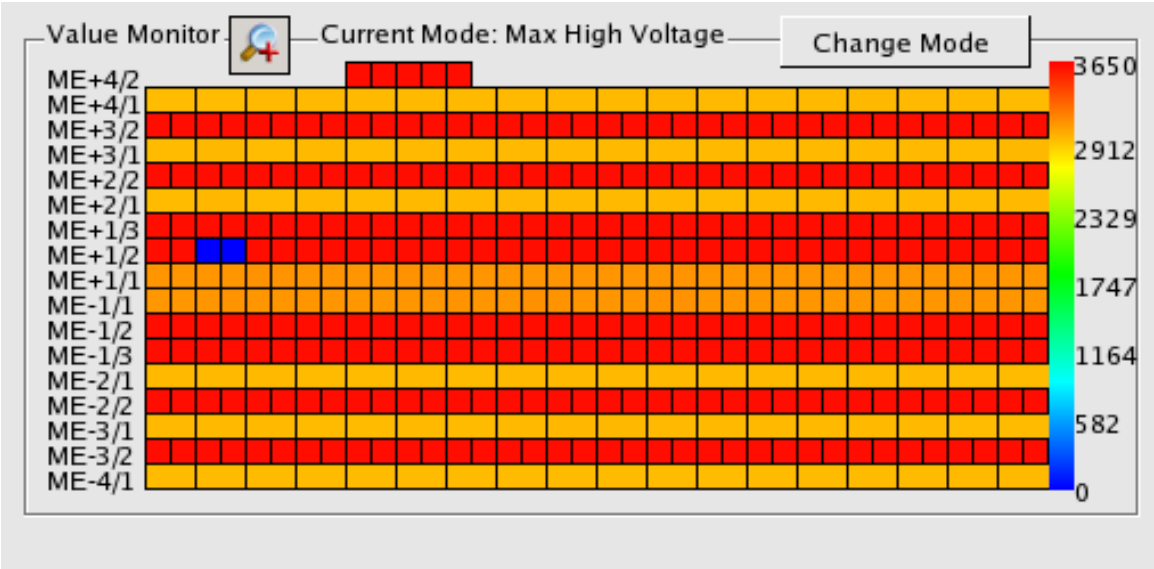

Figure B.1: The max HV of each chamber is shown. The outer chambers have a higher HV than the inner chambers.

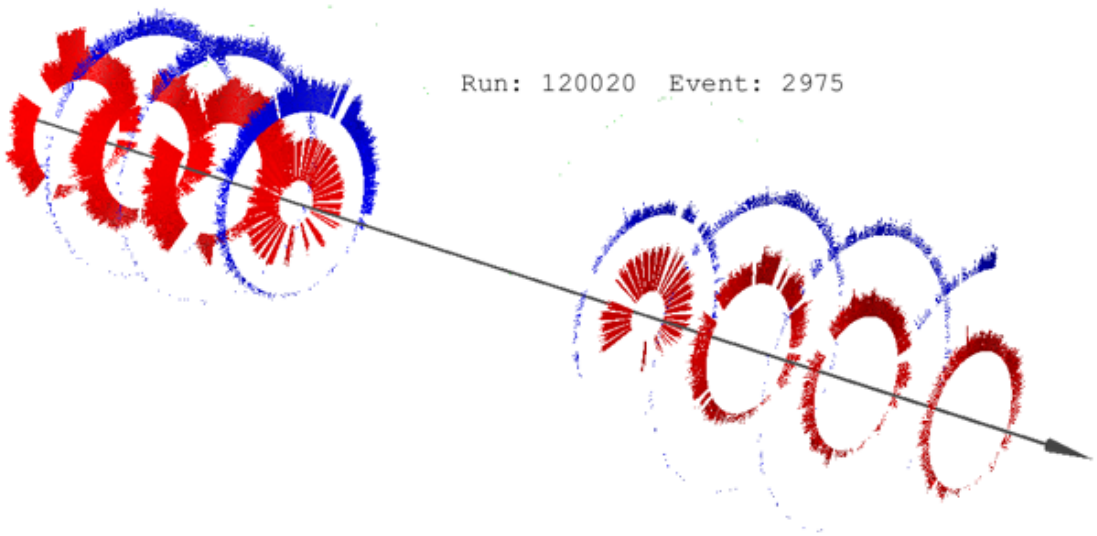

Figure B.2: A beam splash recorded by the CMS detector in November 2009.

go into an error state because DCS would try to set them to a different HV state but the command would not go through. It was found that by increasing the delays in PVSS the number of transitions successfully completed by the CAEN would increase. A more permanent fix was found by updating the CAEN firmware which required an increase in memory.

Before steering the beams in to collision, the LHC circulated single beams in both directions. Halo muons recording during single circulating beams were used for CSC timing and alignment in preparation for collisions. The downstream endcap triggers for halo muons. The upstream endcap's readout is delayed by two bunch crossings.

On November 23, 2009, CMS recorded first collisions at $\sqrt{s} = 900\,\text{GeV}$. Normally, like events used in the analysis described in this thesis, events are from "Golden" json files, or events identified as good by the Physics Data/MC Validation group. However in November 2009, CSC shifters recorded the run number and approximate lumi section that the HV was ramped so that data could be processed as soon as possible to validate the CSC timing and alignment. These collisions were with only one bunch in each beam. The LHC moved quickly to four bunches per beam.

On March 30, 2010, there was a media event at CERN for the first collisions at $\sqrt{s} = 7\,\text{TeV}$. This was the first collisions in a new energy range. For a live webcast, shared with millions around the globe, there were several cameras set up around the CERN laboratory, including one in the CMS control room. The start of an inaugural physics run for a new accelerator has become an more and more rare event in high energy particle physics. Figure B.3 includes a collision event with a muon in the CSC chambers and a few

photos to document this event in this thesis.

The CSC system has about 220K cathode strip channels and 180K anode wire channels read out in groups. In a system with as many electronic channels as are in the CSC system, it is to be expected that hardware problems will lead to a few dead channels. Each CSC wire plane connects to an independent HV power supply channel. Within a plane, the HV is split between up to 5 segments which can be disconnected during LHC runs should a HV or wire noise persist. Many such problems were seen in the first few years of collisions at the LHC. Many of these HV channels would give good data for some time, usually on the order of days. Then, the HV channel would again exhibit a voltage outside of the accepted running parameters. A theory developed by CSC experts that some of these HV channels developed a charge build up that would then dissipate after the HV channel was turned off for a few days. This effect was small $< 1\%$ of all HV channels.

The CSC system continued to be improved and become more automated and stable. As the LHC was providing more and more data the CSC detector moved from two shifters to one. Then the CSCs were given to Global DCS so that no CSC shifter was required to be in the CMS control room starting May 2010. The EMU collaboration took advantage of LHC technical stops, periods where there are no collisions and often no beams, to upgrade online software and replace peripheral crate electronics boards in the experimental cavern. At this time ever effort was made to check and correct all CSC readout problems. There were only 6 bunches with 8×10^{10} protons per bunch in the LHC, but soon the LHC started injecting bunch trains where small problems could cause big losses of data.

Figure B.3: Top shows members of the CMS collaboration in the CMS control room at point 5 (Cessy, France) after successfully recording the first collisions at $\sqrt{s} = 7\,\text{TeV}$ March 30, 2010. Also shown is the author of this thesis giving a live interview (left) and a collision event display including hits in CSC chambers (right).

Much of the commissioning phase will have to be done again after Long Shutdown 1. Many detectors, including the CSC subdetector, are making improvements. The start-up was relatively smooth for the CMS detector. Hopefully this culture of careful validation remains and the start of the new run will see high operational efficiencies as well.