Observation of standard model Higgs boson decays to tau leptons and a search for dark matter with the CMS detector at the LHC

by

Laura Margaret Dodd

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Physics)

at the

University of Wisconsin – Madison

2018

Defended on February 15th, 2018

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

Wesley H. Smith · Bjorn Wiik Professor of Physics

Sridhara R. Dasu · Professor of Physics

Matthew F. Herndon · Professor of Physics

Baha Balantekin · Eugene P. Wigner Professor of Physics

Marshall F. Onellion · Professor of Physics and Materials Science

It takes no more time to see the good side of life than to see the bad.

Jimmy Buffett

Abstract

A search for the standard model (SM) Higgs boson which decays to a pair of τ leptons is presented in this thesis using 35.9 fb^{-1} of 13 TeV data collected in 2016 by the CMS detector at CERN. The final states considered require that (1) one τ lepton decays leptonically while the other decays hadronically ($\mu\tau_{\mathrm{h}}$ or eτ_{h}), (2) both taus decay hadronically ($\tau_{\mathrm{h}}\tau_{\mathrm{h}}$) or (3) one tau lepton decays to an electron while the other decays to a muon (eμ). This thesis specifically focuses on the details of the semi-hadronic decay channels ($\mu\tau_{\mathrm{h}}$ and eτ_{h}), and presents the results of all channels combined. The signal strength, μ, relative to the expectation for the SM Higgs boson is measured to be $\mu = 1.09^{+0.27}_{-0.24}$. An excess is observed (expected) corresponding to 4.9 (4.7) σ using 2016 data. The observed (expected) significance reaches 5.9 (5.9) σ when combined with 7 and 8 TeV data collected at CMS.

Subsequently following the SM search, a search for dark matter (DM) is presented and interpreted in two simplified Higgs-portal models: a baryonic dark matter model and a two Higgs-doublet model. In this thesis, the Higgs boson is required to decay to an opposite-sign tau pair. The tau pair final states considered are eτ_{h}, $\mu\tau_{\mathrm{h}}$, and $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$. The expected 95% confidence level upper limits are produced and compared with the observed limit. No significant deviations beyond the SM prediction are observed.

Acknowledgements

First, I must acknowledge the CMS Collaboration for designing and operating the CMS detector. Without the CMS Collaboration, this thesis would not exist. Additionally I express my gratitude to all those in the University of Wisconsin-Madison physics department and CMS group who have supported me on this journey.

I would like to thank and acknowledge my advisor Wesley for always offering straightforward and sage advice, and for guiding us all. I owe a great deal to Sridhara, a co-advisor, for giving suggestions on how to move forward in difficult situations. And of course, where would any of us be without the HELP@HEP email team; thank you Carl, Dan, Ajit, and Chad for making sure the Tier 2 runs smoothly, replying to my emails at various hours, and giving priority to my jobs when deadlines were looming.

Thank you Bhawna, Cécile, Isobel, and Marià for always being around for coffee breaks and urgent situations. More importantly, thank you for being my good friends. In the most stressful of times, the "UW" chat provided laughs and the latest hubbub via Skype at all hours of the day! To Devin, Tyler, Nate, Nick, Kenneth, Aaron, Tom, James and Usama may your futures be bright.

Most importantly, all this would be impossible without the support from my family and loved ones.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

In this thesis, I discuss the observation of Higgs boson decays to tau pairs $(\tau^+\tau^-)$ and a search for dark matter (DM) produced in association with a Higgs boson which subsequently decays to tau pairs. Both analyses are performed on the 13 TeV LHC proton-proton 35.9 fb^{-1} data set collected in 2016 at the CMS experiment. The search and resulting observation presented in this thesis of $H \to \tau\tau$ is the first single experiment 5 σ observation of the Higgs boson fermionic decay. The DM search presented is also the first search for DM conducted in this final state.

First, I give a general introduction of the standard model (SM) of particle physics and motivate searches for beyond the SM physics (BSM). I then discuss the theory and associated phenomenology of the SM Higgs boson searches and Higgs-portal models. Then a discussion of the CERN LHC and the CMS detector follows. After this information, the reconstruction of physics from detector signals and production of simulation is explained. In Chapter 7, I discuss generalities associated with tau pair final states. The observation of $H \to \tau\tau$ is discussed in Chapter 8. In this analysis, I specialized in observable final states in the detector

where one tau lepton decays hadronically (τ_h) and one tau lepton decays leptonically (ℓ) to either an electron (e) or muon (μ). I participated as a member of a team that examined the four channels: $\mathrm{e}\tau_\mathrm{h}$, $\mu\tau_\mathrm{h}$, $\mathrm{e}\mu$, and $\tau_\mathrm{h}\tau_\mathrm{h}$. More information on the $\tau_\mathrm{h}\tau_\mathrm{h}$ and $\mathrm{e}\mu$ channels can be found in the public results [1, 2]. A search for dark matter produced in association with a $H \to \tau\tau$, as hypothesized in Higgs-portal models, is discussed in Chapter 9. Lastly in Chapter 10, I give an overview of the thesis, significance of results, and outlook for the future.

1.2   Standard model of particle physics

The "standard model" (SM) of particle physics [3, 4] is the current model explaining the interaction between all known fundamental particles and the electromagnetic, weak, and strong forces. The known forces are shown in Table 1.1.

Table 1.1: Forces

Interaction	Mediator	Range (m)	Strength
Strong	gluon g (8 states)	10^{-15}	1
Electromagnetic	photon γ	inf	10^{-3}
Weak	$W^{\pm}/Z$	10^{-18}	10^{-14}
Gravitational	unknown	inf	10^{-43}

All the particles currently included in the SM are shown in Fig. 1.1. There are two main kinds of particles in the SM: fermions and bosons. Fermions are spin-1/2 particles and include leptons and quarks, while bosons are integer-spin and force-carrying particles (mediators). There are three generations of fermions; each generation includes a quark doublet, charged lepton, and neutral lepton (neutrino).

The first generation includes electrons (e), electron neutrinos (ν_e), and up-type (u) and down-type (d) quarks. The second generation includes muons (μ), muon neutrinos (ν_μ), and charm-type (c) and strange-type (s) quarks. The third generation includes taus (τ), tau neutrinos (ν_τ), and top-type (t) and bottom-type (b) quarks. The τ lepton, named after the Greek work for *third*, was discovered and subsequently published in 1975 [5–7]. Within each quark doublet, the up-type quark has charge $+\frac{2}{3}e$, where e is the elementary charge (1.602×10^{-19} Coulombs), and the down-type quarks have charge $-\frac{1}{3}e$. Additionally, all SM particles have an anti-particle which is identical except the quantum state undergoes a CPT transformation. Practically, this CPT transformation results in opposite electric charge [8, 9].

The t and b quark were predicted in 1973 [10]. The b quark was found shortly thereafter in 1977 at FermiLab [11] and the t quark was found in 1995 at FermiLab in $p - \bar{p}$ collisions from the Tevatron [12, 13]. Generally the masses of the fundamental particles increase in each generation, meaning the t quark and τ lepton are more massive than the u quark and electron. The mass hierarchy of the neutrinos is not known. The neutrinos were originally thought to be massless, but now it is known they have a very small mass due to the presence of oscillations. Oscillations indicate that the flavor eigenstates of the neutrinos are not the mass eigenstates which leads to mixing [14–16]. The τ lepton and the t quark are unique from other quarks and leptons in their decays. The τ is unlike the other charged leptons in that it can decay into hadrons (composite quark states). The τ mass is $m_\tau = 1776.86 \pm 0.12$ MeV with a lifetime of 2.9×10^{-13} seconds. The t quark is the heaviest elementary particle in the SM and the lifetime is on the order of 10^{-25} seconds causing it to decay weakly, and never forming a bound state within a hadron [8, 17]

All electrically charged particles interact exchanging photons (γ), which are spin-1 neutral gauge bosons. Quantum electrodynamics (QED), based on U(1) gauge symmetry, which results in the conservation of electric charge, is the first of the quantum field theories (QFT) [18] to be incorporated into the SM. QED formed the basis for the development of the SM in the 20th century [9, 19].

Quarks carry an additional conserved quantum number called color (usually called red, green and blue). A gauge theory based on the $SU(3)_C$ color symmetry contains colored massless gluons which are the force-carrying particles of the strong force, characterized by quantum chromodynamics (QCD). Only color-neutral quark composite particles, "hadrons", are observed in nature; these hadrons are held together by the strong force [9]. Mesons are $q\bar{q}$ bound states, while baryons are composed of three quarks ($qq'q''$) [17]. The proton (neutron) is a baryon with a uud (ddu) state. Recently, the LHCb Collaboration published a potential pentaquark resonance in Ref. [20]. Baryon number, of which each quark contributes 1/3, is

conserved as a global symmetry. There are no known interactions that will violate baryon number [21]. The strong force increases as the particles are separated further, unlike in QED. The result is the range of the strong force is confined to about the size of a proton, 10^{-15} meters; this is called confinement [17]. Deep inelastic scattering experiments have also demonstrated "asymptotic freedom" as the quarks get closer together they may be considered independent from one another [22–25].

The left-handed portion of the quark pairs (u, d), (c, s), (t, b) and lepton pairs (ν_e, e), (ν_μ, μ) and (ν_τ, τ) exhibit an SU(2) symmetry, which results in last remaining gauge bosons in the standard model, the $W^\pm/Z^0$ bosons. The Z^0 boson is neutral and associated with weak neutral currents. The $W^\pm$ boson is associated with charged flavor-changing weak current. The quark mass eigenstates are not the same as the weak quark eignestates. The intergenerational quark mixing in weak decays is described by the unitary Cabibbo-Kobayashi-Maskawa (CKM) matrix [10,26]. In the SM, the electromagnetic and weak forces are unified into a single electroweak force, mediated by the $W^\pm/Z^0/\gamma$ bosons, represented by $SU(2)_L \times U(1)_Y$ symmetry [27]. The U(1) generator weak hypercharge, Y_W, is defined such that

$$\mathcal{Q} = \mathcal{T}_3 + \frac{1}{2}Y_W \tag{1.1}$$

where $\mathcal{Q}$ is the charge and $\mathcal{T}_3$ is the third component of the weak isospin. The right-handed particles have $\mathcal{T}_3 = 0$ and do not interact weakly.

Unbroken $SU(2)_L$ symmetry results in massless gauge bosons, whereas the real W and Z bosons are expected to be massive due to the short-range of the weak interactions. A method of electroweak symmetry breaking (EWSB), the Higgs mechanism, was hypothesized, with the result of providing mass to the $W^\pm$ and Z^0 bosons while being gauge invariant, and predicting the existence of a fundamental scalar boson, the Higgs boson. Gerard t'Hooft proved that the electroweak theory with the Higgs was renormalizable in 1971 [28]. The

massive scalar (spin-0) boson, the Higgs boson (H), was discovered in 2012 at the LHC. A review of the SM history focused on the development of electroweak theory can be found in Ref. [29]. The EWSB and spontaneous symmetry breaking (SSB) resulting in massive $W^{\pm}/Z^0/H$ are discussed in Chapter 2.1.

The Higgs mechanism can also provide mass to the charged leptons and quarks through the Yukawa coupling of the Higgs boson, taking advantage of the existance of their right-handed singlets. The Yukawa mechanism does not explain the mass of the neutrinos; it only works if left- and right-handed chiralities exist for the particle. Currently the mechanism for neutrino mass generation is unknown. Some models, such as the seesaw mechanism and heavy sterile neutrinos, exist to explain the neutrino masses [8]. An experimental measurement of the Higgs boson decays to fermions can directly probe the Yukawa coupling. The τ lepton final state is of particular interest because of the experimental accessibility compared to other fermionic decays; it is the heaviest observed lepton.

Extensive resources are available discussing the SM. A select few are in Refs. [9,17–19,30].

This thesis will specifically discuss H decays to τ pairs, and provides the first single-experiment measurement confirming the existence of the SM Higgs boson coupling to leptons.

Standard Model of Elementary Particles
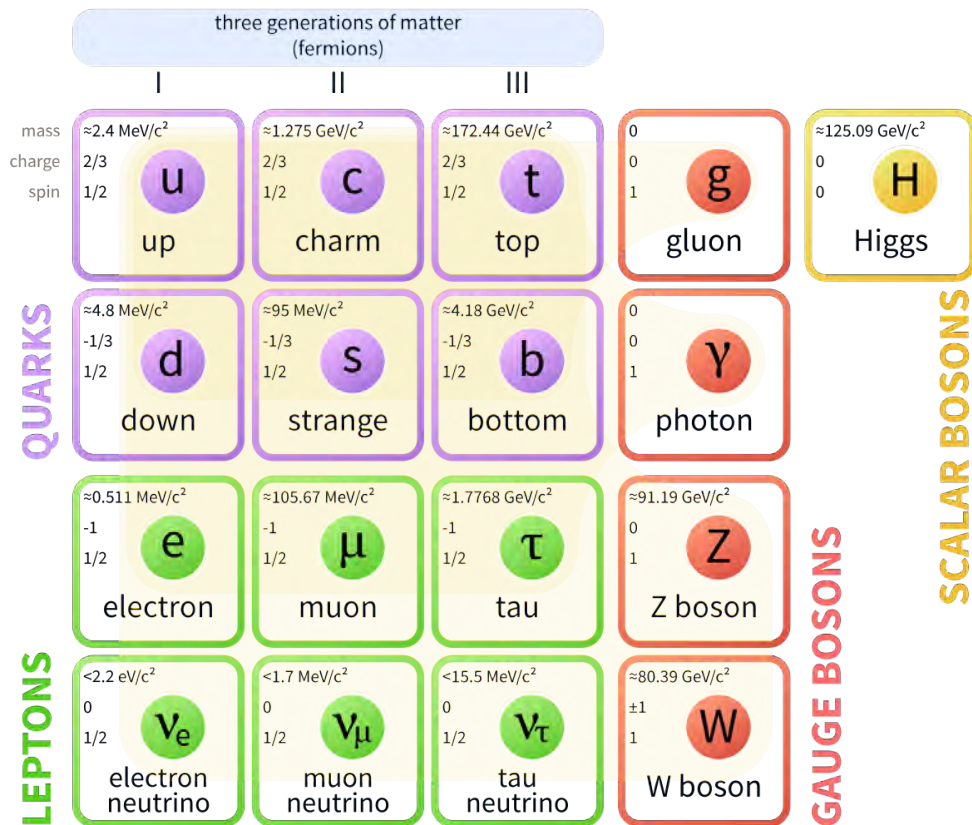
Figure 1.1: Particles in SM, including the newly discovered Higgs boson. Each square includes the mass, charge, and spin of the particle.

1.3   Beyond the standard model

Despite the success of the SM, it is widely thought to be incomplete [31, 32]. Many models exist to explain physics beyond the standard model (BSM).[1] One of the most interesting issues not addressed by the SM is the existence of dark matter (DM).

The first indications of DM existence came in 1933, when Fritz Zwicky measured Coma cluster galaxy velocities by measuring the Doppler shift of their spectra [35]. The measured velocities were two orders of magnitude higher than expected, given the mass-estimate from counting the number of stars. Since velocities are dependent on the radius from the center and the mass of the galaxy, this mismatch between star count and velocity indicated a discrepancy in the mass of the galaxy. The measured invisible mass did not interact electromagnetically and was eventually coined dark matter (DM) [30, 36]. Velocity rotational curves are measured to be roughly constant as a function of the radius; at what galaxy radius the dark matter reduces is not yet known as the dark matter halo extends beyond the visible galaxy curve [8]. Figure 1.2, reproduced from Ref. [37], shows the galactic rotation curve of NGC 6503. The existence of DM halos and DM disks limits the baryonic contribution to dark matter because baryons dissipate energy on a long time scale and would collapse into a disk.

DM and dark energy dominate the universe [8, 38]. Current astrophysical measurements indicate 68% of the universe is composed of dark energy, 27% is DM, and the remaining 5% is considered SM matter. Dark energy is more prevalent than DM, but is not the subject of this thesis. Further reading can be found in Ref. [39].

Rotational curves and gravitational lensing measurements support that DM lives mostly in a spherical halo around galaxies. Rotational curves also currently estimate the DM mass density of the universe to be $\Omega_{DM} \approx 0.2$, where the ratio, Ω, is defined such that

---

[1]I performed 2 BSM $H \to \tau\tau$ searches at 8 TeV [33] and 13 TeV [34] that are not discussed in this thesis. Both were focused on Minimal Supersymmetric Model (MSSM) Higgs searches.
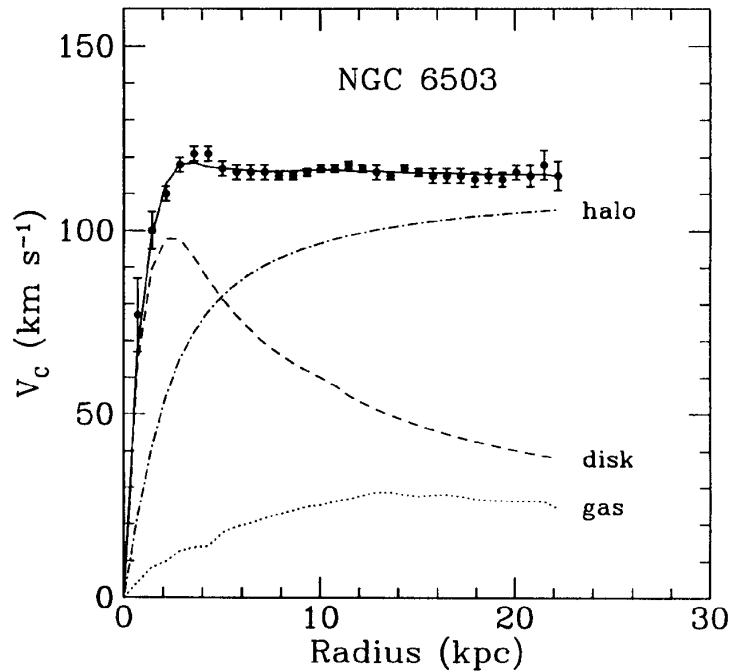
Figure 1.2: NGC 6503 galactic rotation curves showing the various contributions of dark matter needed to agree with observation.

$\Omega_{\text{all matter}} = 1$ corresponds to a flat universe. Simultaneously, this estimate places a lower bound of $\Omega_{\text{DM}} \geq 0.1$ [8]. Using the virial theorem, the average velocities of the dark matter in the halos is estimated to be $< v >= 200$ km/s. The local dark matter density near earth is estimated to be $\rho_{local}^{DM} \approx 0.3 \frac{\text{GeV}}{\text{cm}^3}$ [40].

Halo formation implies a lower limit on the mass range of the DM particle candidate. A scalar DM particle mass must be $m_{\text{spin-0}} \gtrsim 10^{-22}$ eV [40, 41]. The lower bound on the DM mass increases for fermionic particles, $m_{\text{spin-1/2}} \gtrsim 0.7$ keV [40, 42]. Upper bounds on the DM mass come from thermal physics arguments. As the early universe expanded, the rate of DM annihilation slowed as the universe cooled, until "freeze-out" when the DM density became constant. The annihilation rate is a function of the number of DM (χ) particles, n_χ, and the velocity-averaged cross section. Requiring the annihilation rate to be equivalent to the universe expansion (Hubble rate, $\mathcal{H}$) demonstrates that "hot" DM is relativistic at

freeze-out, while "cold" DM is not relativistic. Warm DM occupies the spectrum between relativistic and non-relativistic. Early galactic structures were dependent on DM, and if DM were relativistic these early structures would not exist. Therefore DM is expected to be cold, meaning non relativistic. Additionally galactic DM disks would condense into rotating bars if DM were hot. Other astrophysical measurements of colliding galaxies indicate dark matter only weakly couples to itself [43, 44].

Particle physicists aim to study the underlying fundamental nature of DM and how it can be included in the SM. DM doesn't interact electromagnetically and is stable. None of the current existing SM particles are viable DM candidates. One current theory is that DM is a yet-undiscovered weakly interacting massive particle (WIMP). A massive particle such as this may interact with the Higgs boson. The so-called "WIMP-miracle" with a 100 GeV DM, at the weak scale, produces the observed DM abundance and satisfies its other properties. The WIMP miracle assumes DM annihilation is a $2 \rightarrow 2$ process. Many collider searches focus on these WIMPS at the GeV scale. The DM models in this thesis are included by the LHC DM benchmark paper [45], where DM is assumed to be a fermionic WIMP, and does not interact with the detector.

There are three main ways to search for DM: (1) indirect detection which searches for DM annihilation products from astrophysical sources, (2) direct detection, where dense heavy shielded detectors wait for a DM particle to cause a nuclear recoil from a collision, and lastly (3) collider production, where DM is produced in collisions of standard model particles in colliders. These three methods are summarized in Fig. 1.3.

Indirect detection requires measuring DM collisions and the resulting annihilation from millions of light years away. No indirect detection of DM has been made via telescopes and other technologies. One inconclusive recent search indicates there is a deficit of anti-electrons compared to electrons compared to anti-protons and protons found with the AMS experiment [46]. However the reported signal could be from pulsars as well as DM.
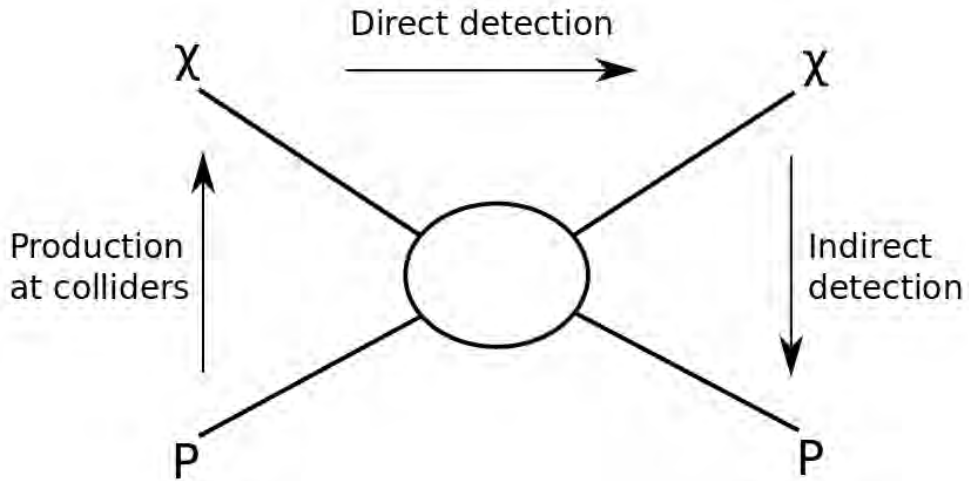
Figure 1.3: Diagram showing the three methods of possible DM, χ, detection. The P can be a proton or another SM particle.

Direct detection requires a DM particle to hit a nucleus in a small experiment on Earth, which has a very low cross-section. The excluded cross sections for a dark matter in the GeV mass range currently extend down to about 10^{-44} cm^2. Despite extensive shielding against backgrounds, and long run-times of experiments, no DM has been found via direct direction. Substantial phase space remains for direct detection experiments to explore.

Lastly it is possible that DM could be produced in SM particle collisions within collider experiments at the LHC, typically via a heavy mediating particle. In this thesis, I will focus on a small subset of DM searches at the CMS experiment. Specifically, I concentrate on a Higgs-portal model search, where a potential link between the dark matter sector and the newly discovered Higgs boson is hypothesized.

Most theoretical models that predict production of DM particles from proton-proton collision must be consistent with the DM density observed in the universe. The thermal relic density places some constraint on the model parameter space to be considered.

1.4 Significance of results

One of the hallmarks of the SM Higgs boson prediction is the Yukawa coupling, theorized to give fermions mass. By measuring the Higgs boson cross section and branching fractions, we can see where possible future deviations of the SM may be. Measuring the fermionic decays of the Higgs boson provides further confirmation of the Yukawa coupling of the Higgs boson and mass generation of charged leptons and quarks. Further studies of the fermionic coupling of the Higgs boson will continue in Run-II of the LHC.

This thesis also presents a search for a 125 GeV $H \to \tau\tau$ in association with missing transverse momentum, interpreted in the context of higgs-portal DM models, using events recorded by the CMS experiment in 13 TeV proton-proton collisions at the LHC collected in 2016. Three final states with the highest branching ratios are considered: $e\tau_{\mathrm{h}}$, $e\tau_{\mathrm{h}}$, and $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$. 95% CL limits are set on a Z' two Higgs-doublet model and a baryonic Z' model with $H + \mathrm{DM}$ final states.

Chapter 2

Standard model 125 GeV Higgs boson

2.1 Theory

Glashow, Weinberg, and Salam worked on unifying the electric force and the weak force, now called the "electroweak" force. Above 100 GeV the electromagnetic and weak forces merge. The resulting theory exhibits an $SU(2)_L \times U(1)_Y$ symmetry.

The UA1 experiment at CERN, which began in 1979, published the discovery of the W/Z bosons in 1983, ultimately culminating in a Nobel Prize [47, 48]. Currently, the masses are measured to be $M_{W^\pm} = 80.385 \pm 0.015$ GeV [8] and the $M_{Z^0} = 91.1876 \pm 0.0021$ GeV [8]. Including the masses by directly adding a mass term in the SM Lagrangian breaks gauge invariance, therefore they must be added via a different mechanism [49].

In 1964, Higgs, Brout, and Englert published papers positing the existence of a scalar boson, and provided a mechanism for electroweak symmetry breaking (EWSB) in order to give the $W^\pm/Z^0$ masses of 80 GeV and 90 GeV respectively [50–53]. EWSB could also solve the problem of massless quarks and massless charged leptons in the SM theory. Guralnik, Hagan and Kibble among others published papers on EWSB in the 1960's [54, 55].

In the above papers, to explain mass generation, a complex scalar, spin-0, doublet with

four degrees of freedom is introduced to the SM. Three degrees of freedom are given to $W^\pm/Z^0$ (V) vector bosons to make them massive, and the last degree of freedom predicts the existence of a new scalar boson, now termed "Higgs boson". It is the only scalar boson predicted in the SM. The Higgs doublet, ϕ, in the form,

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{2.1}$$

is put into the following Lagrangian,

$$\mathcal{L} = (D_\mu \phi)^\dagger (D^\mu \phi) + V(\phi) \tag{2.2}$$

where D_μ is a covariant derivative and the potential $V(\phi)$ is

$$V(\phi) = \mu^2 \phi^\dagger \phi + \frac{\lambda}{4}(\phi^\dagger \phi)^2. \tag{2.3}$$

The values of μ and λ are free parameters. If $\lambda > 0$ and $\mu^2 > 0$, the potential is symmetric with a vacuum expectation value (VEV) of 0. If $\mu^2 < 0$ and $\lambda > 0$, this potential has a degenerate set of non-zero minimum values which can be seen in Fig. 2.1 and a non-zero vacuum expectation value, v. The minima are located in a ring of radius

$$v^2 = -\frac{\mu^2}{\lambda} \equiv 246 \quad \text{GeV}. \tag{2.4}$$

Any fluctuation around the v spontaneously breaks the rotational symmetry. We can choose to break the symmetry by defining the minimum of ϕ to be

$$\phi_{min} \equiv \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{2.5}$$

The solution is now

$$\phi_1 = \phi_2 = \phi_4 = 0, \quad \phi_3^2 = -\frac{\mu^2}{\lambda} \equiv v^2 \tag{2.6}$$

which indicates that three of the four components are null. A small fluctuation around this minima can be expanded and written

$$\phi(x) = \phi_0 + h(x) \tag{2.7}$$

where $h(x)$ describes a Higgs boson. Expanding to second order around the minima produces $V = V_0 + \lambda v^2 h^2$, and indicates a Higgs boson mass of

$$M_H^2 = 2\lambda v^2 \tag{2.8}$$

which needs to be experimentally measured.
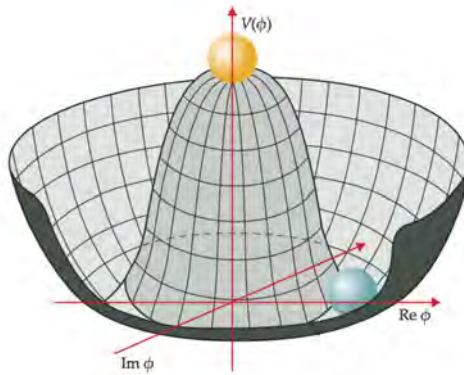
Figure 2.1: $V(\phi)$ potential in Eq. 2.3 showing a non-zero vacuum expectation value.

The potential contains the kinetic term $(D_\mu \phi)^\dagger (D^\mu \phi)$. Using the following field definitions,

$$W_\mu^{1,2(\pm)} = \sqrt{1/2}(A_\mu^1 \mp iA_\mu^2) \tag{2.9}$$

$$Z_\mu = -B_\mu \sin\theta_W + W_\mu^3 \cos\theta_W) \tag{2.10}$$

$$A_\mu = B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W \tag{2.11}$$

and the weak mixing angle[1], θ_W, the matrix coming from the Lagrangian is diagonalized. The W_μ, Z_μ, A_μ are considered as the W boson, Z boson and photon respectively. The weak

---

[1] also called Weinberg angle, is responsible for the diagonalization of the matrix.

mixing angle is defined as $\tan\theta_W = g'/g$ where g is the coupling from U(1) and g' is the coupling from SU(2).

When these fields are included in ϕ, with the VEV above, a mass term emerges from the $(D_\mu\phi)^\dagger(D^\mu\phi)$ term,

$$\mathcal{L}_{\text{mass}} = -M_W^2 W^{+\mu}W_\mu^- - \frac{1}{2}M_Z^2 Z^\mu Z_{mu} \tag{2.12}$$

which identifies the masses as

$$M_W = \frac{1}{2}gv, \quad M_Z = \frac{g^2 + g'^2}{2} \tag{2.13}$$

for the W and Z boson [18].

The Higgs field is hypothesized to give masses to the fermions, solving another conundrum of the SM, by adding a Yukawa interaction with left- and right-handed fermionic fields. A Yukawa interaction is the way Dirac (e.g. fermion) fields can interact with scalar fields. For a given leptonic field, ψ_f, and a scalar field, ϕ, the interaction is $g_f\bar{\psi}_f\psi_f\phi$. The fermions masses can be wrtten in the form

$$m_f = \frac{g_f v}{\sqrt{2}}. \tag{2.14}$$

The Yukawa couplings required for the measured masses of the fermions is shown in Table 2.1. The SM does not explain the values in Table 2.1; they are values from the experimentally measured mass.

Table 2.1: Higgs boson Yukawa couplings for charged leptons and quarks

		1st gen.		2nd gen.		3rd gen.
up-type quark	g_u	2×10^{-5}	g_c	9×10^{-3}	g_t	~ 1
down-type quark	g_d	4×10^{-5}	g_s	8×10^{-4}	g_b	3×10^{-2}
charged lepton	g_e	3×10^{-6}	g_μ	6×10^{-4}	g_τ	1×10^{-2}

2.2 Previous results

The four LEP collaborations searched for Higgs boson between 1989 and 1995 in the ZH channel [56]. They set a 95% CL lower bound on the Higgs boson mass around $m_H = 114.4$ GeV with 2461 pb^{-1} of data collected. The LEP, an $e^+ - e^-$ collider, had the highest likelihood of finding a Higgs boson in the ZH production mode. The electroweak measurements at the time provided a best guess for the Higgs boson mass of $m_{\mathrm{H}} = 81^{+52}_{-33}$ GeV, with a 95% upper exclusion of 193 GeV [56].

Before the LHC startup and after LEP, the Tevatron, a proton-antiproton collider located at Fermilab, also performed searches for the Higgs boson. They ruled out a large portion of phase space around $m_H > 160$, primarily looking for associated and gluon fusion production. Vector boson fusion also contributed but it was subdominant [57, 58].

In 2012, the LHC experiments, CMS and ATLAS, found a scalar boson in several decay channels: $H \to \gamma\gamma$, $H \to ZZ \to 4\ell$, $H \to \mathrm{WW}$. The combination of these channels resulted in a best-fit mass of $125.09 \pm 0.21(stat) \pm 0.11(syst)$ GeV [59]. The signal strength of the observation given only the SM was 1.09 ± 0.11 [60]. The Tevatron collaborations released results around the same time with the full detector Tevatron data sets. CDF and D0 had a broad excess of data between 115 GeV and 140 GeV, with a global significance of 3.1 sigma for the 125 GeV Higgs boson. Additionally, they excluded the phase space between 149 GeV and 182 GeV. The Tevatron results are compatible with the 125 GeV Higgs boson found at the LHC [57, 58, 61].

2.3 Production and phenomenology

Within the standard model (SM), there are four primary Higgs boson production mechanisms possible for proton-proton colliders. The Feynman diagrams for these processes are shown

in Fig. 2.2.

(a) gluon fusion ≈ 50 pb

(b) vector boson fusion ≈ 3.75 pb

(c) V associated production ≈ 2.25 pb total

(d) $t\bar{t}$ associated production ≈ 0.509 pb
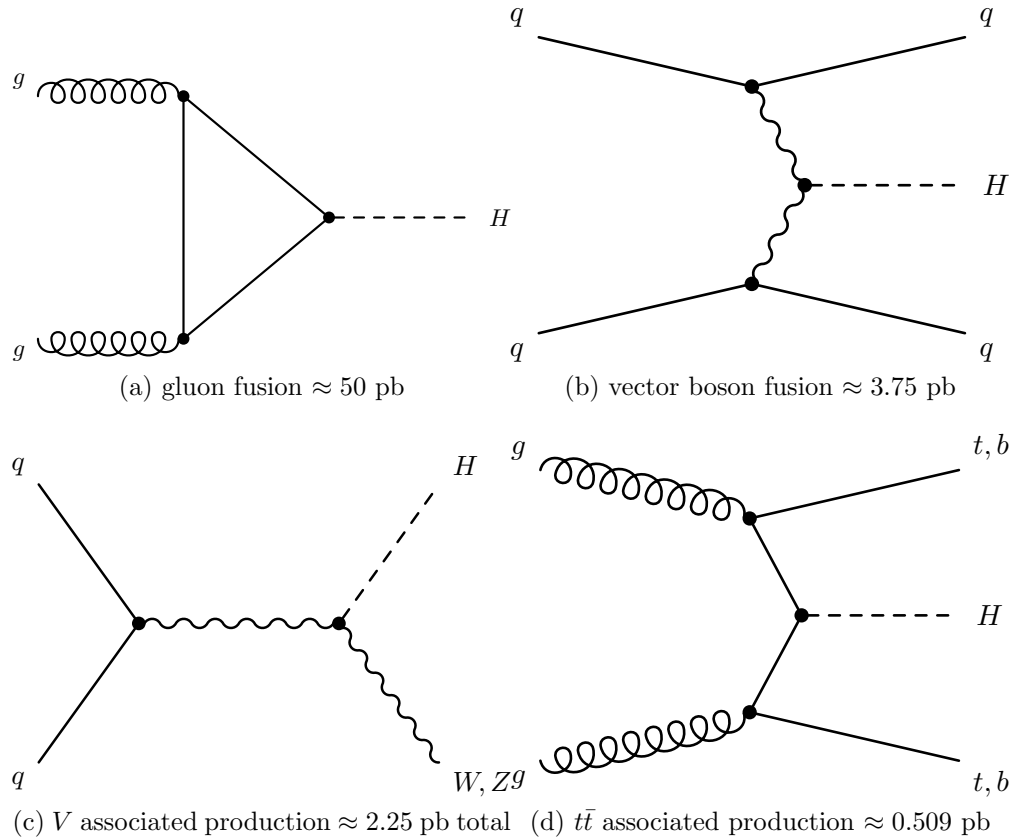
Figure 2.2: Tree-level Higgs boson production Feynman diagrams at the LHC and the predicted cross sections at $\sqrt{s} = 13$ TeV.

Each production mechanism has a trademark phenomenology associated with it. For non-associated production processes, the outgoing partons in the event can be used for potential identification of the underlying process (categorization). In high-energy particle collisions, unbound quarks and gluons will go through the hadronization process. Unbound quarks and gluons are not observed outside of high-energy particle collisions due to confinement; only color-neutral matter states are observed. Any remaining partons and quarks will create a color-neutral quark/gluon particle shower and deposit energy in the detector as various mesons, baryons, and their respective decay products. This hadronization shower is called a

jet. The production of these showers is governed by the strong force, the subject of quantum chromodynamics (QCD) [17].

Gluon fusion, sometimes denoted as ggH in this thesis, may be associated with no jets, or be boosted off of one jet from the initial state. The largest expected production mechanism at the LHC is ggH. The predicted 13 TeV p-p ggH cross section of 50 pb is a factor of ten higher then the next largest production mechanism, vector boson fusion (VBF) with a cross section around 4 pb.

VBF is usually associated with two forward jets with high di-jet invariant mass, originating from the outgoing quarks from the interaction. The associated production modes ($t\bar{t} + H$ and $V + H$) both have a Higgs boson decay in the final state along with either $t\bar{t}$ or V and are most easily categorized by extra leptons in leptonic final states.

The Higgs boson is predicted to couple directly to massive particle pairs. Given a high enough Higgs boson mass, the largest branching fraction would be to $t\bar{t}$. However $H \to t\bar{t}$ is an off-shell decay for the top quarks, therefore this final state is suppressed due to the top mass. Therefore, the largest branching ratio is $H \to b\bar{b}$.[2] The primary searchable experimental Higgs boson decay modes within the SM are $H \to b\bar{b}$, $H \to \tau\tau$, $H \to \gamma\gamma$, $H \to ZZ$, and $H \to$ WW. The branching fractions are dependent on the Higgs boson mass, and on the mass of the decay product. The scalar boson found in 2012 indicated a mass of $125.09 \pm 0.21(stat) \pm 0.11(syst)$ GeV [59]. The branching fractions predicted in the SM for measured mass are shown in Table 2.2. The tree-level Feynman diagrams of Higgs boson decays are shown in Fig. 2.3 and the amplitudes associated with these processes are mostly dictated by the mass of the outgoing particles.

The Higgs boson can decay to γ's, but not at tree-level. For a di-photon decay, the Higgs boson must decay via either a vector boson loop or a fermion, predominately top quark,

---

[2]$b\bar{b}$ frequently denoted bb. In Higgs boson decays it is understood that particles are produced in OS pairs if charged.

Table 2.2: Branching fractions for 125 GeV SM Higgs boson

Higgs boson decay channel	Branching Fraction [%]
$H \to bb$	57.5 ± 1.9
$H \to WW$	21.6 ± 0.9
$H \to gg$	8.56 ± 0.86
$H \to \tau\tau$	6.30 ± 0.36
$H \to cc$	2.90 ± 0.35
$H \to ZZ$	2.67 ± 0.11
$H \to \gamma\gamma$	0.228 ± 0.011
$H \to Z\gamma$	0.155 ± 0.014
$H \to \mu\mu$	0.022 ± 0.001

loop as shown in Fig. 2.4. To *directly* measure the fermionic (Yukawa) coupling of the Higgs boson, the W and Z boson decay channels cannot be used. However, the relative strength of the Higgs coupling to the top quark can be tested with a precise measurement of the ggH production cross section.

(a) vector boson decays   (b) fermionic decays (Yukawa coupling)
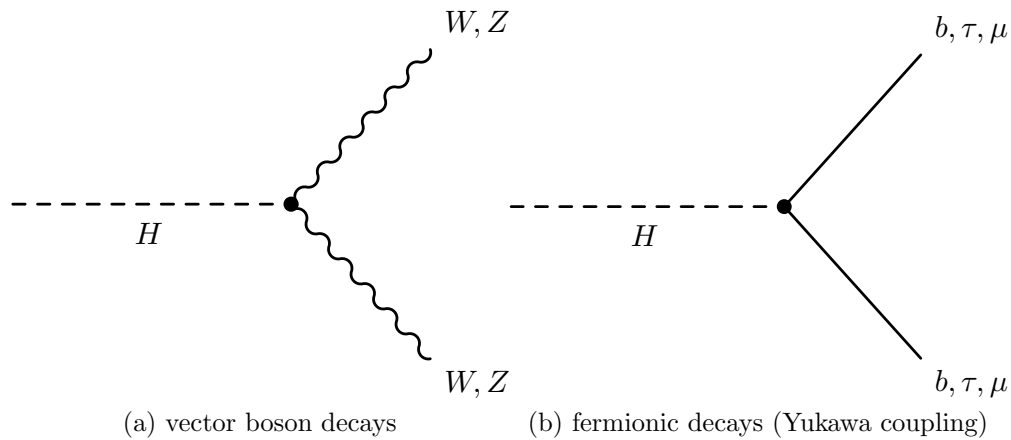
Figure 2.3: Higgs boson decay Feynman diagrams illustrating the direct couplings of the Higgs boson to fermions and bosons

(a) fermion loop   (b) vector boson loop
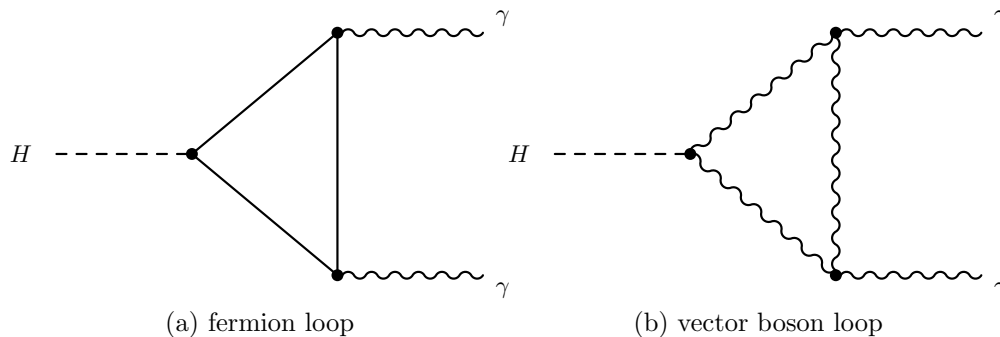
Figure 2.4: Higgs boson decay to $\gamma\gamma$ Feynman diagrams

2.4   Previous LHC results

In order for the scalar boson found to be the Higgs boson there, are certain properties that must be confirmed. First, given the mass of the scalar boson it must match the predicted branching fractions for each final state. Assuming a mass of 125.09 GeV, the ratio of observed to predicted signal strengths are calculated for 6 different decay modes ZZ,WW, $\gamma\gamma$, bb, $\tau\tau$, and $\mu\mu$; the results are shown in Fig. 2.5.

Both CMS and ATLAS observed data compatible with the CP-even spin-0 SM Higgs boson. The spin parity of the Higgs boson is measured in the H $\rightarrow$ WW, H $\rightarrow ZZ$ and H $\rightarrow \gamma\gamma$ decay modes. ATLAS used an effective field theory approach [62], while CMS used the anomalous coupling approach [63] and both confirmed the CP-0 nature of the scalar boson. The CMS spin measurements from 7 and 8 TeV exclude the spin-1 hypothesis at greater than 99% confidence level(CL), due to the presence of the di-photon decay. The spin-2 hypothesis is rejected at greater than 95% confidence level(CL) in the $ZZ \rightarrow 4\ell$ final state [63, 64]. ATLAS rejected several BSM spin models including spin-0 and spin-2 models at more than 99.9% CL [62].

CMS results constrain the Higgs boson lifetime to be $\tau_H < 1.9 \times 10^{-13}s$ at the 95% CL, which corresponds to a lower bound on the width of $\Gamma_H > 3.5 \times 10^{-9}$ MeV [65]. The width
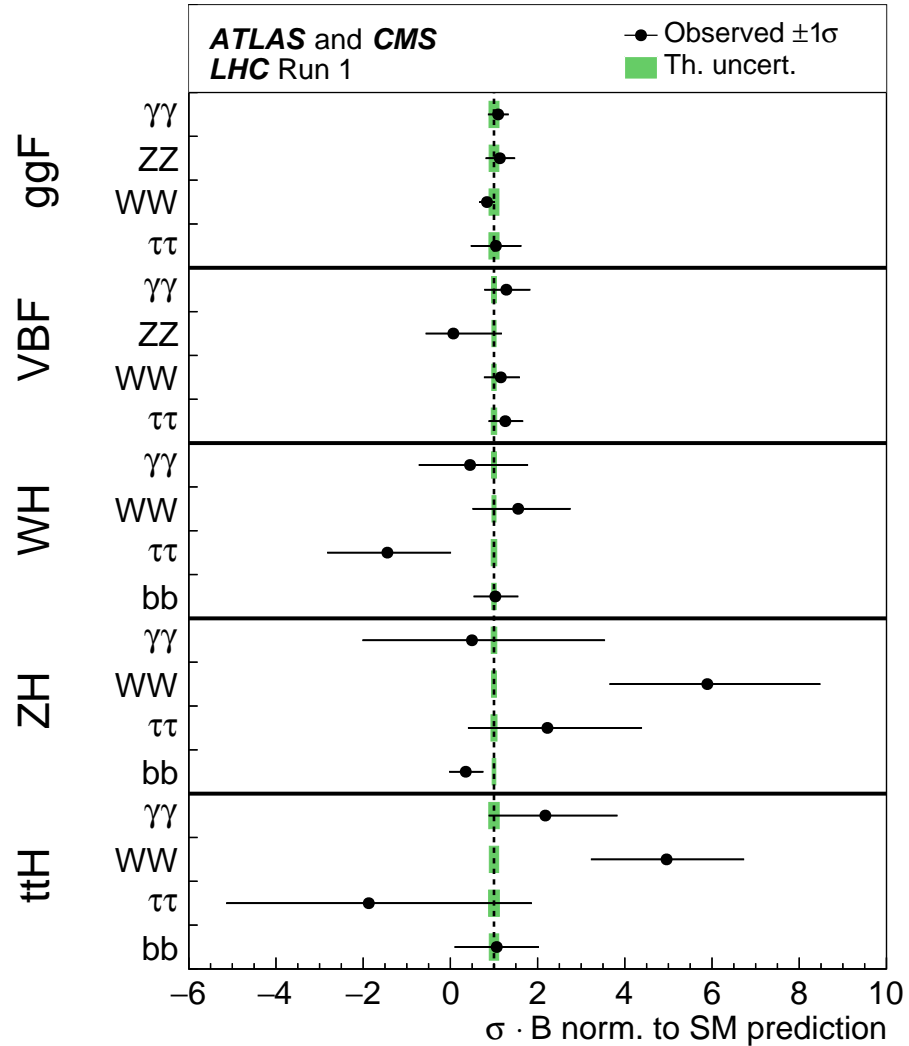
Figure 2.5: Results of CMS+ATLAS Higgs combination comparing the SM prediction to the measured values, split by production mode.

is consistent with the predicted SM Higgs boson width.

One of the last requirements for the 125 GeV Higgs boson to be SM Higgs is the Yukawa coupling. Neither ATLAS nor CMS alone had enough significance in the fermionic decay channels, bb and $\tau\tau$, to claim observation of the Yukawa coupling. As shown in Fig. 2.6, the measured $H \to bb$ decay mode was less than SM prediction. There are three experimentally
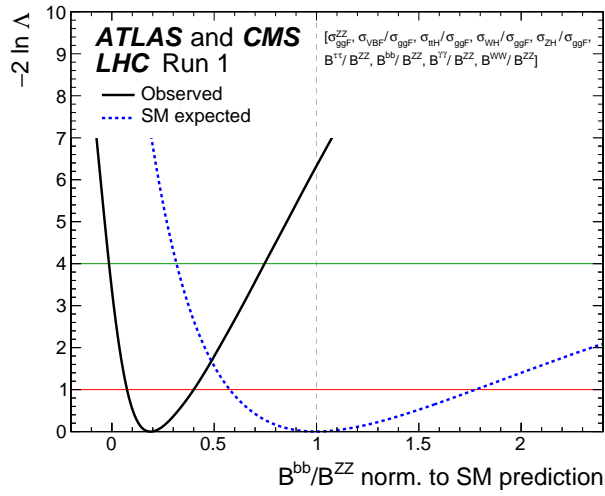
Figure 2.6: Result of comparing the observed branching fraction of $H \to bb$ to $H \to ZZ$.

accessible fermionic Higgs decay modes: bb, $\tau\tau$, and $\mu\mu$. The $\mu\mu$ channel is extraordinarily difficult due to the large falling Drell-Yan background still present around 125 GeV. The bb channel has the highest branching ratio, however it is difficult to trigger the data due to the topology of those events compared with background events. The $\tau\tau$ channel has large backgrounds and neutrinos in the final state. Excluding the $H \to \mu\mu$ final state, ATLAS measured a SM signal strength of $\mu_{H\to bb} = 0.62$ and $\mu_{H\to\tau\tau} = 1.41$, while CMS measured $\mu_{H\to bb} = 0.81$ and $\mu_{H\to\tau\tau} = 0.88$. Despite tension in the $H \to \tau\tau$ channel, the combination of the ATLAS and CMS results led to an observed (expected) 5.5 (5.0) standard deviations for the $\tau\tau$ final state. Whereas the bb final state has a combined observed (expected) significance of 2.6 (3.7) standard deviations [60]. LHC Run-II objectives for CMS include improving these measurements to check possible deviations in the SM.

In Run-II of the LHC, the CMS and ATLAS Collaborations use the Higgs boson as tool for discovery of new physics.[3] If the Higgs boson decays to BSM physics, it should be seen. By combining all the SM Higgs boson searches an upper limit can be placed on the Higgs

---

[3]Included in the 2014 P5 report is "Use the Higgs boson as a new tool for discovery." The report can be found http://www.usparticlephysics.org
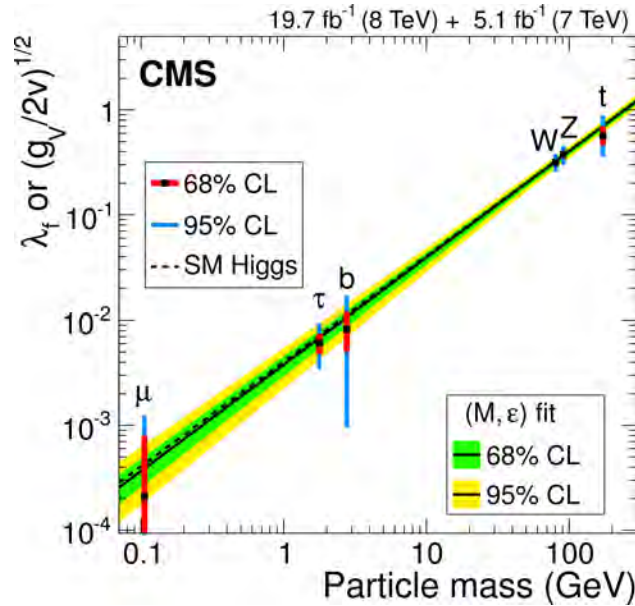
Figure 2.7: Summary of the fits for deviations in the coupling for the generic five-parameter model not effective loop couplings, expressed as a function of the particle mass. Image retrieved from public CMS site.

branching fraction to BSM physics. The Run-I CMS-ATLAS Higgs combination produced an upper limit on the BSM branching fraction of 34% as shown in Fig. 2.8, including invisible decays. In order to decrease the upper limit on the BSM Higgs decays, the measurements of each decay mode should be improved [60].

Overall the reduction of the error bars on each category in Fig. 2.5 must to be reduced to see how well the SM prediction agrees with observation. Better measurements of the $H \to \tau\tau$ decay, which has a relatively[4] high branching fraction of 6.3%, can help place more stringent limits on the BSM branching fraction shown in Fig. 2.8. The error bar on Fig. 2.7 for the τ can be reduced to further check agreement of data with the λ_f. In this thesis, I present the analysis of the $H \to \tau\tau$ decay in the data collected in 2016 by the CMS experiment.

---

[4]Compared with the higher-sensitivity channels ZZ and $\gamma\gamma$
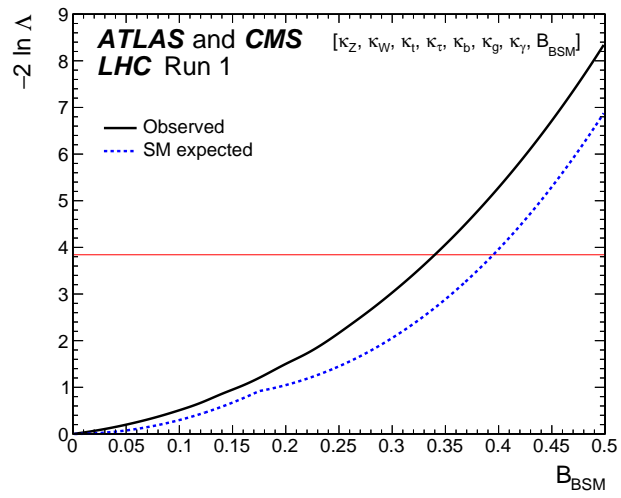
Figure 2.8: Upper limit on BSM branching fraction of Higgs boson including invisible decays of the Higgs.

Chapter 3

Higgs-portal models and mono-Higgs production

Since the Higgs couples to massive particles, and dark matter (DM) is massive, it could be possible for the SM Higgs boson to couple to DM. In the case where a Higgs boson decays "invisibly" to a pair of DM particles it is classified as an invisible Higgs search [66]. The decay should be on-shell for maximum efficiency, that is about < 62 GeV, half the Higgs boson mass.

Other models exist where a Higgs boson is produced in association with a pair of DM particles. This final state is called a "mono-Higgs" [45, 67, 68]. In this final state there the resulting decay of one Higgs boson in addition to an energy imbalance in the detector.

Higgs-portal models are models where the Higgs can interact with the dark sector. In this thesis, I will discuss a search for two Higgs-Portal benchmark models found in Ref. [45] and discussed in the following sections: the two Higgs doublet model (Z'-2HDM) and the baryonic Z' model. It should be noted some "portal" is needed for collider DM detection, for proton collisions to produce DM. The Higgs-portal models discussed must provide the necessary thermal relic abundance of DM observed in the universe.

In many of these theories a new mediating particle besides the DM particle is hypothesized in order to provide a link between the incoming quarks at the LHC and the DM pair.

3.1 Z' two Higgs-doublet model

The two Higgs doublet model (2HDM) is a popular model in several theories [69], for example the minimal supersymmetric standard model (MSSM) [70]. A second Higgs doublet is added to the SM [45, 67]. The two Higgs doublets consist of Φ_u and Φ_d, where the first couples to up-type quarks and the second couples to leptons and down-type quarks (Type-2). The general Yukawa potential for 2HDM models can be written

$$V_{yukawa} = -\sum_{i=1,2}(\mathcal{Q}\tilde{\Phi}_i y_i^u \bar{u} + \mathcal{Q}\Phi_i y_i^d \bar{d} + L\Phi_i y_i^e \bar{e} + h.c.)$$
(3.1)

and the different values assigned, y_i, change the behavior of the couplings. The Type-2 Z'-2HDM model used in this thesis sets $y_1^u = y_2^d = y_2^e = 0$. Now rather than one VEV, there is v_u and v_d. Using this we can write the two Higgs doublets as

$$\Phi_d = \frac{1}{\sqrt{2}}\begin{pmatrix} -\sin\beta H^+ \\ v_d - \sin\alpha h + \cos\alpha H - i\sin\beta A^0 \end{pmatrix}$$
(3.2)

$$\Phi_u = \frac{1}{\sqrt{2}}\begin{pmatrix} \cos\beta H^+ \\ v_u + \cos\alpha h + \sin\alpha H + i\cos\beta A^0 \end{pmatrix}$$
(3.3)

where $H^\pm$ is a charged scalar, A^0 is a neutral pseudoscalar[1], and h/H are neutral CP-even scalars (one being the 125 GeV Higgs). The definition of $\tan\beta$ is $\tan\beta \equiv \frac{v_u}{v_d}$ and α is the mixing angle that diagonalizes the two neutral CP-even Higgs. The parameters are tuned such that h has SM-like couplings.

---

[1]CP-odd

There is one more scalar singlet ϕ added for spontaneous symmetry breaking (SSB), discussed in Section 2.1, resulting in a heavy Z' order of several hundred GeV and above. The masses of the Z and Z' are given by

$$(M_Z^0)^2 = g^2 \frac{v_d^2 + v_u^2}{4\cos^2\theta_w} \tag{3.4}$$

$$(M_{Z'})^2 = g_z^2(z_d^2 v_d^2 + z_u^2 v_u^2 + z_\Phi^2 v_\Phi^2) \tag{3.5}$$

where z_u and z_d are the respective Z' charges for Φ_u and Φ_d. The cross section scales with $(g_z)^2$. The new particles in this model are Z', A^0, $H^\pm$, H^0 and a DM particle, χ. A Feynman diagram for DM production in this model is shown in Fig. 3.1.
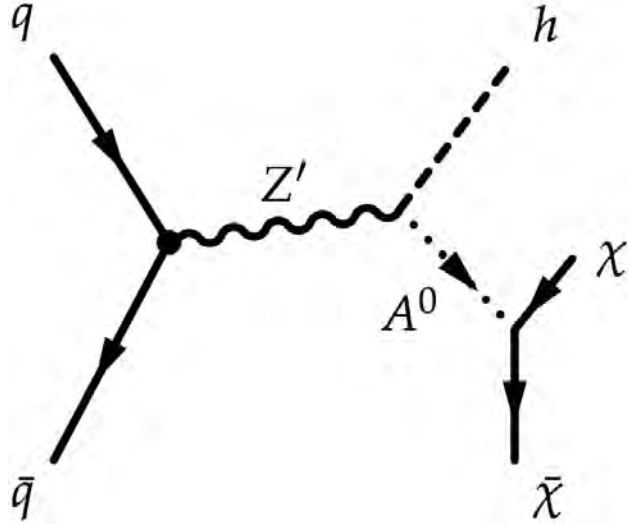
Figure 3.1: Two Higgs-doublet model Feynman diagram

In this model a heavy Z', in this thesis between 600 GeV and 2500 GeV, is produced and decays to a 125 GeV Higgs boson, and the pseudoscalar A^0. The A^0 then decays to a pair of dark matter particles, χ. Coupling the χ to the A^0 is motivated by direct detection constraints and other di-lepton channels [45]. Heavy Z' particle with di-lepton decays have been extensively searched for and ruled out in this mass range [71]. Therefore, only the right-handed SM fermions are charged with the Z' charge in this model, thus avoiding these

limits. The DM in this model could potentially but rarely decay to another dark matter particle plus a qq pair. Constraints from LUX [72] allow the scattering of DM off quarks to be mediated by the H or h.

There are 5 main parameters in the model: M_{A^0}, m_χ, $M_{Z'}$, $\tan\beta$, and g_z. The parameters $\tan\beta$ and g_z do not affect the kinematical distributions, and only affect production rates. If m_χ is less than half of m_A^0, then this parameter may also stay fixed, without large kinematical differences, as shown in Fig. 3.2, reproduced from Ref. [45]. In this thesis, g_z is fixed at $g_z = 0.8$, $\tan\beta = 1$, and $m_\chi = 100$ GeV [45, 67]. The branching fraction of $A^0 \to \chi\chi$ is assumed to be 100%.

The predicted cross sections for each mass point can be seen in Table 3.2. With a $H \to \tau\tau$ branching fraction of about 6% and assuming roughly 15% acceptance of events and 36 fb^{-1}, we could expect to see roughly 300 events for a cross section of 1 pb.

3.2 Baryonic Z' model

The other mono-Higgs model included in this thesis is the "baryonic" model. Three new particles are introduced into the SM: A (1) DM particle, χ, with no SM associated charge except for baryon number, but not considered a qqq state, in the mass range between 1 GeV and 1 TeV and (2) a Z' in the mass range 10 GeV to 10 TeV, which does not decay leptonically, and (3) a higgs (h_B) from broken U(1) baryonic gauge symmetry. The Feynman diagram for the production can be seen in Fig. 3.3.

The baryonic Z' model Higgs boson kinematical distributions are distinct from the Z'-2HDM discussed in the previous section. The Z'-2HDM has resonant production; the mass of the Higgs+$\chi\chi$ pair forms a peak structure at the Z' mass. The baryonic Z' model, with a radiating Higgs, is non-resonant; the mass of the Higgs+$\chi\chi$ does not form a peak at the Z' mass.
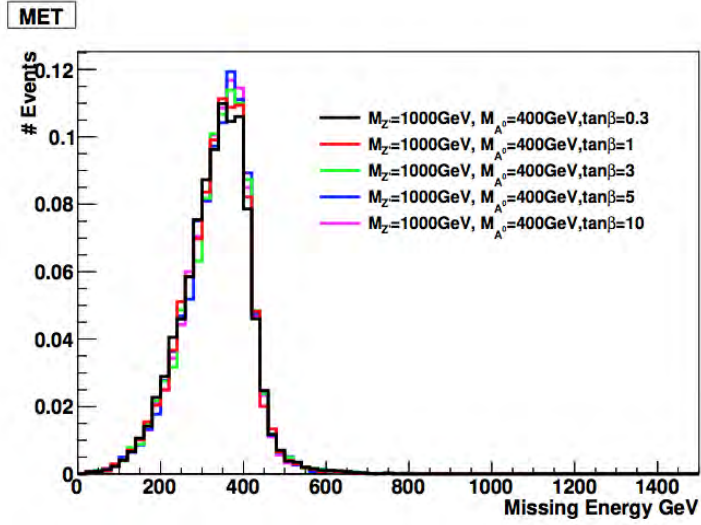
(a) $\tan\beta$ $E_{\mathrm{T}}^{\mathrm{miss}}$ variation
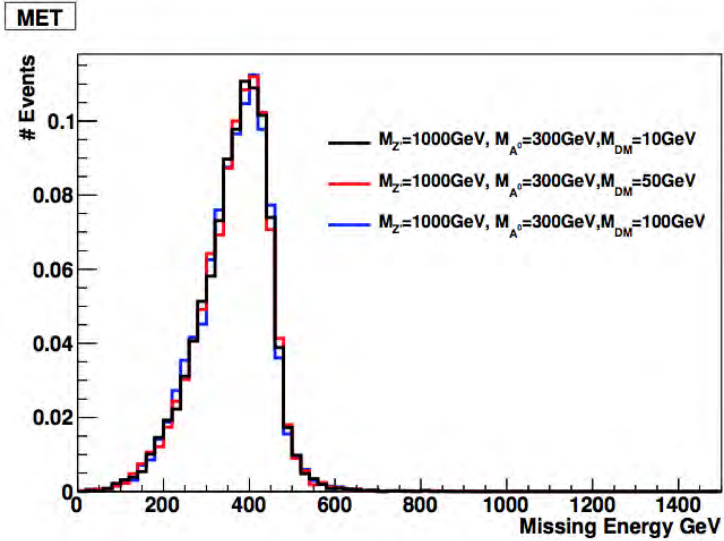
(b) m_χ $E_{\mathrm{T}}^{\mathrm{miss}}$ variation

Figure 3.2: Distributions of generator level MET reproduced from dark matter benchmark paper showing little kinematical differences in the MET spectrum when the $\tan\beta$ parameter is varied and when the m_χ parameter is varied.
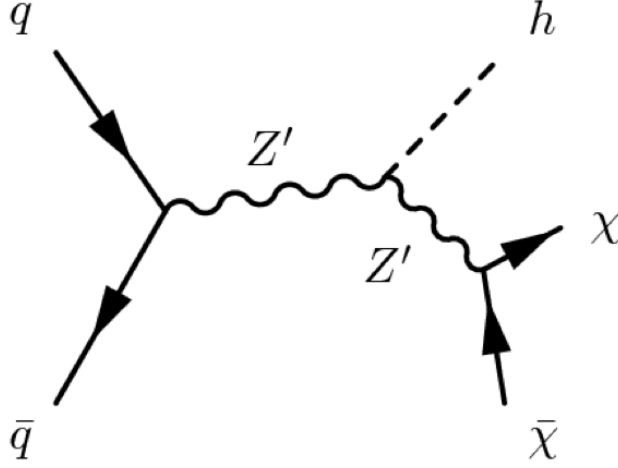
Figure 3.3: New gauge boson, Z', decays to dark matter pairs and radiates a Higgs boson.

The Lagrangian for the Z' describing the interaction with quarks and χ is

$$\mathcal{L} = g_q \bar{q} \gamma^\mu q Z'_\mu + g_\chi \bar{\chi} \gamma^\mu \chi Z'_\mu \qquad (3.6)$$

where the g_q is the quark coupling, which is fixed at one third of the gauge baryonic coupling g_B, and the DM coupling is $g_\chi = B g_B$ for simplicity.

The coupling of the DM to the Z' scales as the baryonic coupling (g_B) and baryon number. The Z' does not couple to any leptons, which avoids the experimental constraints placed on heavy Z' [73, 74]. The effective Lagrangian can be written

$$\mathcal{L}_{eff} = -\frac{g_q g_\chi}{m_{Z'}^2} \bar{q} \gamma^\mu q \bar{\chi} \gamma_\mu \chi \left(1 + \frac{g_{hZ'Z'}}{m_{Z'}^2} h \right). \qquad (3.7)$$

It should be noted that the mono-jet signature, that is a single jet with detector energy imbalance, for dark matter can constrain the first part of the above equation. The first term describes a vector mediator decaying to $\chi\chi$. For the mono-jet search, one of the initial-state quarks radiates a gluon.

The coupling $g_{hZ'Z'}$ can be expressed $g_{hZ'Z'} = \frac{2m_{Z'}\sin\theta}{v_B}$, where θ is the mixing angle between the baryonic Higgs (h_B) and the SM Higgs(h) and v_B is the VEV of h_B. This model has 6 parameters: $M_{Z'}$, m_χ, θ, $g_{hZ'Z'}$, g_q, and g_χ.

The kinematical distributions are not affected by changes in θ, $g_{hZ'Z'}$, g_q, and g_χ, shown in Fig. 3.4 reproduced from Ref. [45]. A scan in the $M_{Z'}$ and m_χ parameters, which do change the kinematic distributions, can be explored. For maximization of the cross section we choose $g_{hZ'Z'}/m_{Z'} = 1$(which will fix the value of θ), $g_q = 1/3$, and $g_\chi = 1$ [45].

The predicted cross sections for each mass point can be seen in Table 3.1. Recall, we could roughly expect to see 300 total events for a process with a cross section of 1 pb for 36 fb^{-1} of data.
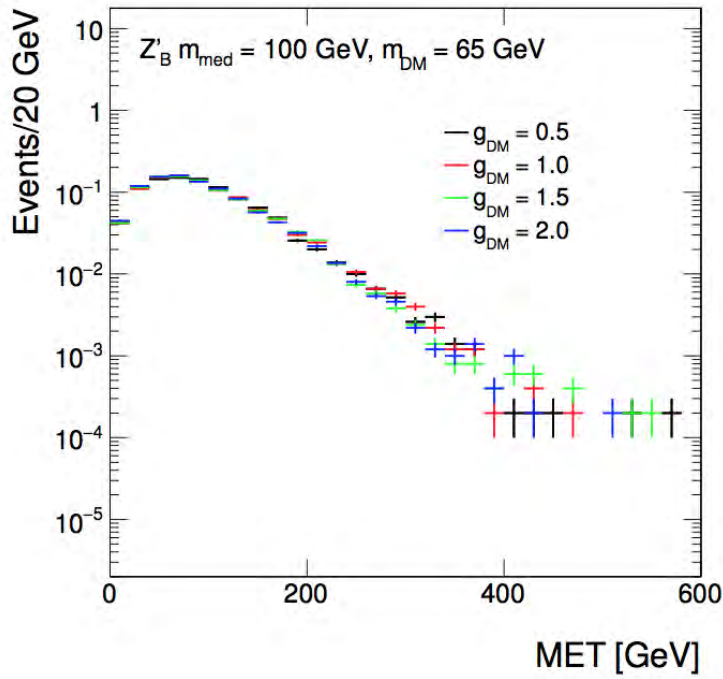
(a) g_{DM} variation
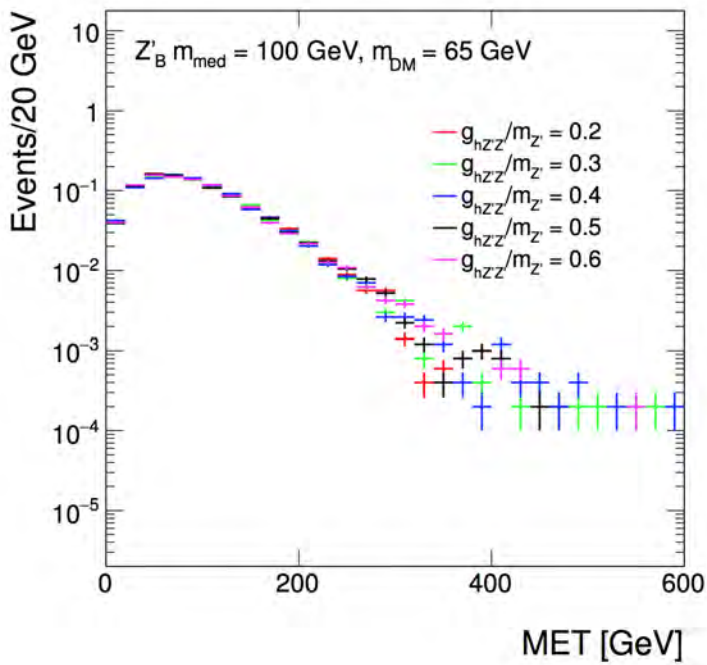
(b) $g_{hZ'Z'}/m_{Z'}$ variation

Figure 3.4: Distributions of generator level MET showing little kinematical differences in the MET spectrum when the DM coupling to the mediator is varied, and when the 125 GeV Higgs coupling to the mediator is varied.

Table 3.1: Select predicted baryonic model cross sections $M_{Mediator}$ and M_χ.

$M_{Mediator}$ GeV	M_χ GeV	Cross Section (pb)
10	1	2.594752001
10	10	0.011928549
10	100	1.78E-05
10	500	4.51E-09
10	1000	2.77E-11
50	1	3.256070229
50	10	3.238556778
50	100	0.000487014
50	500	1.13E-07
50	1000	6.89E-10
100	1	3.180245146
100	10	3.173548501
100	100	0.002702657
100	500	4.58E-07
100	1000	2.77E-09
500	1	1.093670849
500	10	1.094727145
500	50	1.075864949
500	100	0.956036166
500	500	2.06E-05
500	1000	8.08E-08
1000	1	0.201769754
1000	10	0.202000886
1000	50	0.201049075
1000	100	0.199719772
995	500	0.011121583
1000	1000	5.52E-07
1500	1	0.048723009
1500	10	0.048770099
1500	50	0.048894283
1500	100	0.048541216
1500	500	0.035793462
1500	1000	4.08E-06
2000	1	0.013934823
2000	10	0.013954034
2000	50	0.013995043
2000	100	0.013972162
2000	500	0.012671852
1995	1000	0.00072969
10000	1	1.19E-08
10000	10	1.19E-08
10000	100	1.18E-08
10000	1000	6.60E-09

Table 3.2: Select predicted cross sections in pb for the Z'-2HDM model considered for various $M_{Z'}$ and M_A.

$M_{Z'}$ GeV	M_A GeV	M_χ GeV	Cross Section (pb)
600	300	100	0.46742
600	400	100	0.052464
600	500	100	0.0033008
600	600	100	0.00083025
800	300	100	0.28642
800	400	100	0.075285
800	500	100	0.029397
800	600	100	0.00782689
1000	300	100	0.14834
1000	400	100	0.04844
1000	500	100	0.027545
1000	600	100	0.01511
1200	300	100	0.078043
1200	400	100	0.028053
1200	500	100	0.018397
1200	600	100	0.012562
1400	300	100	0.04268
1400	400	100	0.016158
1400	500	100	0.011416
1400	600	100	0.0086372
1700	300	100	0.018507
1700	400	100	0.0073166
1700	500	100	0.0054759
1700	600	100	0.004477
2000	300	100	0.0085964
2000	400	100	0.0034733
2000	500	100	0.0026851
2000	600	100	0.0022923
2500	300	100	0.0026755
2500	400	100	0.001103
2500	500	100	0.00087733
2500	600	100	0.00077499

Chapter 4

Experiment

4.1   The Large Hadron Collider

The Large Hadron Collider (LHC) [75], shown in Fig. 4.1 is a 26.7 km circular particle accelerator and collider installed underground near Geneva, Switzerland, under the French/Swiss border. The LHC has a design center of mass (c.o.m.) energy of 14 TeV, 7 TeV per proton beam, and cost 3.756 billion to build, with an additional 576 million CHF for detector access areas and CERN's computing share. The LHC accelerates protons in counter-rotating beams, and collides the two proton beams at 4 points around the LHC ring. The 2.7-meter-diameter tunnels were originally built for the Large Electron Positron (LEP) collider between 1984 and 1989. LEP was decommissioned in 2000 at which point the installation of the LHC was underway. The LHC is located deep underground to offset the large cost of acquiring land and building new tunnels. The bedrock supporting the tunnel ensures long term stability. The rock above the LHC limits cosmic backgrounds within detectors, while increasing radiation safety by limiting the amount of potential radiation exposure to persons on the surface.

The LHC operated at a c.o.m. energy of 7 TeV in 2011, 8 TeV in 2012 and, after a long

shutdown , operated at 13 TeV in 2015 and 2016. The LHC collides protons at four points for four experiments to collect data.

- The Compact Muon Solenoid (CMS), discussed in Section 4.2, is located at Point 5. It is a multi-purpose detector.

- A Toroidal LHC ApparatuS (ATLAS) located underground at Point 1 is another general-purpose detector, performing similar searches and measurements as CMS. The experiments are important cross-checks for results [76]. ATLAS is a much larger detector than CMS, with calorimeters located outside the magnets

- Large Hadron Collider beauty (LHCb) is an asymmetric non-hermetic detector which studies matter-antimatter asymmetry through b-quark physics [77] and it is located at Point 1.

- A Large Ion Collider Experiment (ALICE) is designed specifically to collect heavy ion collisions at the LHC in order to study the quark gluon plasma [78]. It is located at Point 2.

CMS and ATLAS were designed independently and have completely independent data sets. Reliable cross checks are important for measurements and physics searches. They were built on differing fundamental structures, ATLAS has many gas chambers to measure tracks of particles that pass through, and calorimeters outside the magnets. The stronger CMS magnet allows for better momentum measurement, with the caveat that the calorimeters must be located within the solenoid structure. This thesis only examines data collected in 2016 from protons that collide at Point 5, in the CMS experiment.

The accelerator chain, shown in Fig. 4.1, [75] is composed of many different smaller accelerators in addition to the LHC.
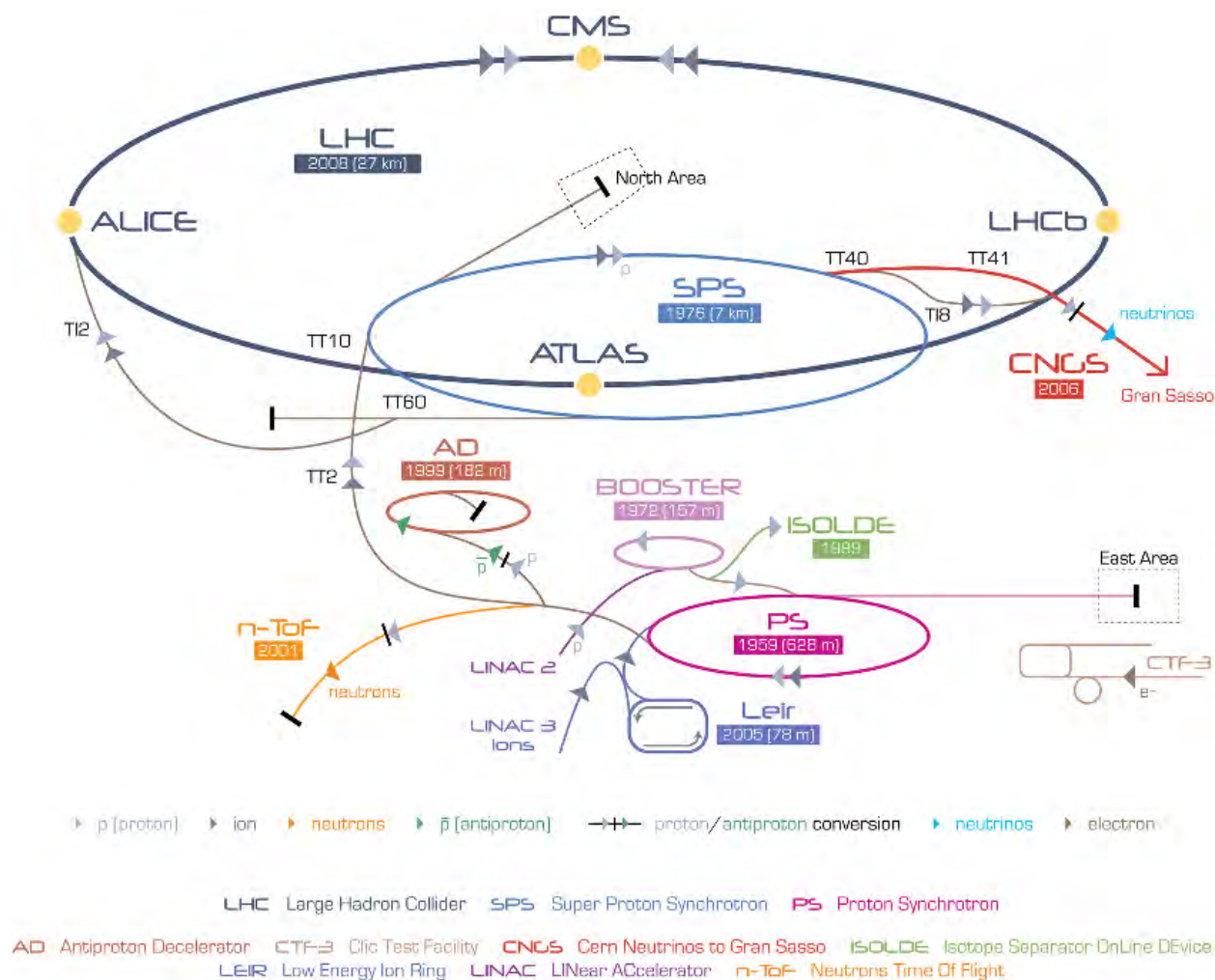
Figure 4.1: LHC accelerator chain showing the 4 main interaction points/experiments: AT-LAS,CMS, LHCb, ALICE.

CMS and ATLAS are located on opposite sides of the LHC ring to allow for these high-luminosity general-purpose experiments to receive adequate luminosity and to increase beam stability.

The protons that are put into the LHC originate from a small red bottle of hydrogen gas near a linear accelerator at the Meyrin site, LINAC2, where the protons are accelerated to 50 MeV. The protons are separated from the electrons by applying an electric field, and then are accelerated via radio-frequency (RF) cavities. At this point, the protons are separated

into bunches. The beam is then injected into the Proton Synchrotron Booster, where the protons are accelerated to 1.4 GeV. Next, the Proton Synchrotron accelerates them to 25 GeV in order to inject them into the Super Proton Synchrotron, where they gain enough energy (450 GeV) in order to be injected and appropriately captured by the LHC's RF cavities.

CROSS SECTION OF LHC DIPOLE
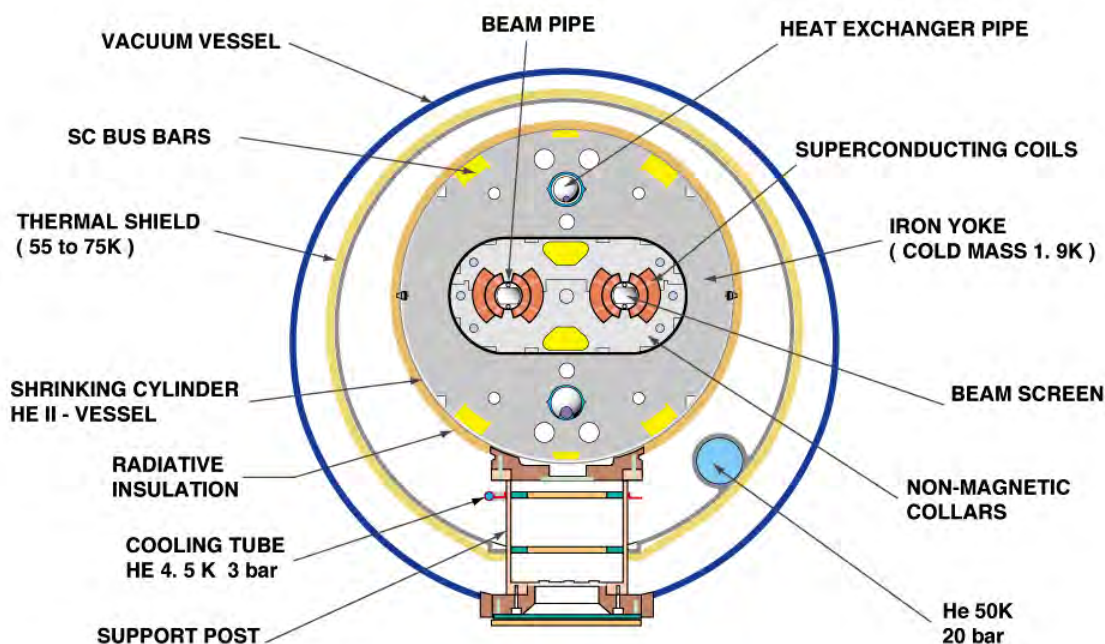
Figure 4.2: LHC dipole cross section

The LHC uses 1232 15-meter-long super-conducting dipole magnets that generate 8.33 Tesla magnetic fields, which bend the proton bunches around the LHC ring. Superconducting magnets are required to obtain the needed magnetic field of 8.33 T to bend the beams. They must be cooled to a temperature of 1.9K to maintain their superconducting state. An image

showing the cross section of the dipole can be seen in Fig. 4.2. The cooling system uses liquid helium. Each proton beam has final energy of 6.5 TeV, corresponding to the protons traveling at $\sim 0.999999990c$. The LHC is made of 8 different sectors (octants), and within each octant is a straight portions and curved portions. Within the LHC dipole, there are two beam pipes, each one bending the protons in opposite directions. The beam pipes are surrounded by superconducting niobium-titanium (NbTi) magnets. Different kinds are magnets are used in the LHC for different purposes, such as bending, and focusing proton beams. Besides the dipoles, there are about 400 quadrupoles, 688 sextupoles and 168 octupoles which aid in squeezing the beam. About 2,500 sextupoles compensate for extraneous fields from the dipoles, 1232 octupole and decapole magnets help to correct for other beam effects in each dipole magnet. The decapole and sextupoles are shown in Fig. 4.3.

While the magnets bend and focus the beams, the 400 MHz superconducting RF cavities inside the LHC accelerate two beams. Each bunch of protons fills one RF cavity.

When the LHC has captured all intended proton bunches, it is called one "fill." About thirty minutes, four cycles of the PS synchrotron and twelve cycles of the SPS are required to inject one fill. The LHC then takes about 30 minutes to ramp up (RAMP) the protons' energy from 450 GeV to 6.5 TeV. The next beam mode after RAMP is FLAT TOP, which indicates the acceleration is complete. The bunches are then squeezed in preparation for collisions in SQUEEZE, to increase the density of each bunch. The final step before stable collisions are declared is ADJUST. During ADJUST, each beam is expertly moved closer and closer to the other beam until they are colliding. After the beams are optimally colliding STABLE BEAMS are declared. The bunches collide every 25 ns.

Once the LHC is filled with 2808 bunches of protons, the beams are left to circulate in the LHC at maximum energy for many hours in order for each experiment to collect data. After many hours, the number of protons in each bunch has decreased substantially. The beam cannot be immediately shut off. Eventually the beam is emptied, or "dumped," at
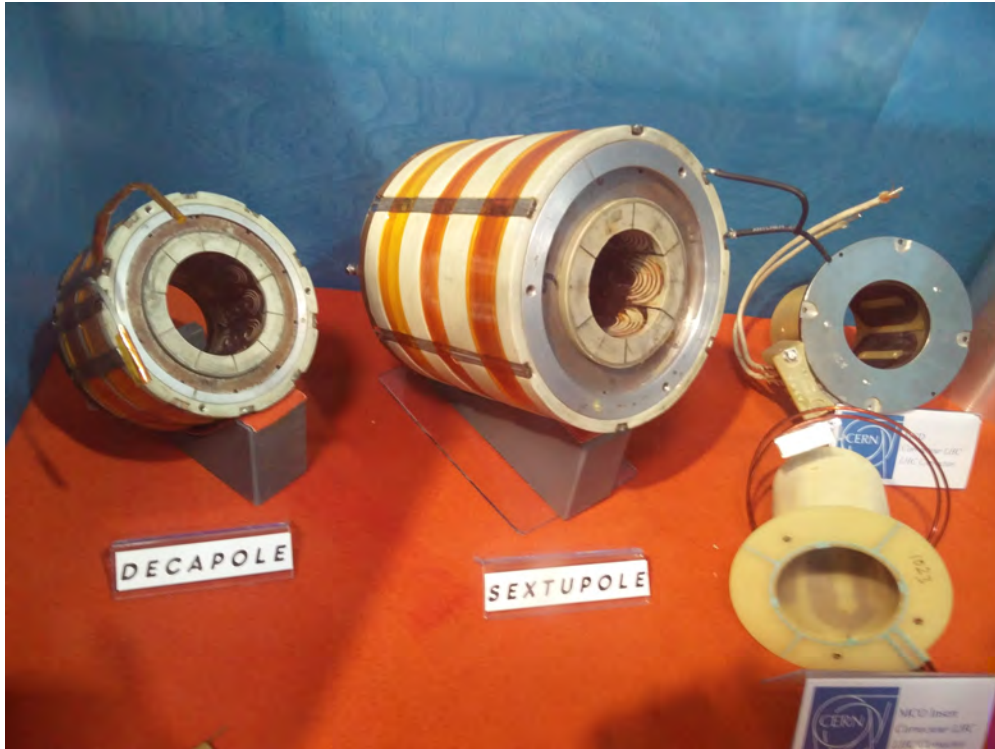
Figure 4.3: Self-taken picture of LHC decapole and sextupole.

Point 6 in octant 6, the adjacent octant to octant 5 where the CMS experiment is located. In order to abort the beam, 3 μs are needed for the LHC dump kicker, so there is a small gap in the proton bunch train, termed abort gap, that allows for the protons to exit the LHC. The beam collides into an absorber in a spiral pattern, a beam sweep, in order to avoid boiling the absorber. The absorber is 7 m long water-cooled graphite cylinder surrounded by steel and concrete. Once the protons are kicked out of the LHC, they travel down a 700 meter-long tunnel and pass through various magnets, which push the beam in a spiral. The 700 meter tunnel-length before the absorber allows for the bunch cross section to increase by a factor of eight. The wider beam helps to avoid excessive damage to the absorber. Figure 4.4 [79] shows the absorber and the spiral beam sweep pattern.

Another critical piece of LHC protection is the magnet quenching system. For the desired magnetic field of the dipoles, the temperature may not exceed the critical temperature, where

(a) Beam Sweep   (b) Absorber
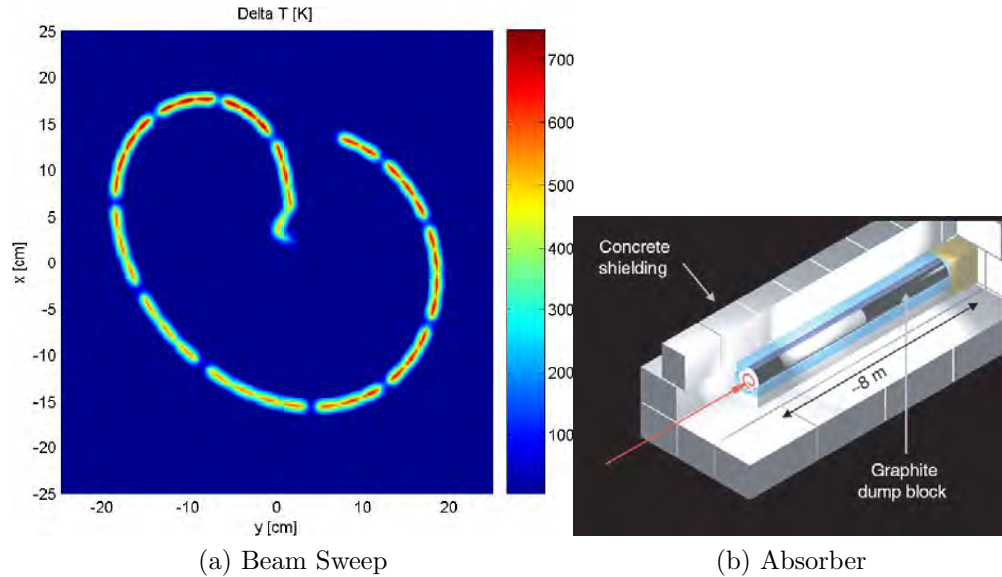
Figure 4.4: Beam sweep example (4.4a) and absorber layout (4.4b)

the magnet becomes normal and will quench. During quench, the temperature of the magnet windings will greatly increase due to resistive losses. A quench could start by a increase in temperature due to cooling system failure, small energy deposit from beam of a few mJ, small movement (micrometer) of the superconducting wire. Advanced quench protective systems are in place [79]. The beam pipe is in direct contact with the 1.9K helium. For effective traversal of the proton beams, the beam pipe is kept under high vacuum between $10^{-7} - 10^{-9}$ Pa.

The beams are then tuned carefully to ensure collisions at several places around the LHC ring. Luminosity, measured in units of *barns*, is the ratio of the rate to a cross section and easily provides the rate of a process given a cross section. It is a useful unit to measure the amount of data being collected at a moment in time (instantaneous luminosity). Integrated luminosity is typically used to describe how much data has been collected over a certain time-range. One barn is equivalent to 10^{-24}cm^2. The instantaneous luminosity delivered to each experiment is based on several beam parameters at the collision point. The luminosity

can be written

$$\mathcal{L} \simeq \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} \mathcal{F}. \tag{4.1}$$

The numerator includes N_b, the number of protons in each bunch, n_b, the number of colliding bunches, f_{rev}, the revolution frequency of bunches at the LHC, γ_r, a relativistic factor. The denominator is essentially how likely the bunches are to collide, and it includes the emittance, ϵ_n, and β^*, approximately the transverse width of the bunch. There is one more factor $\mathcal{F}$ and it accounts for the relativistic crossing angle of the two proton bunches. The various LHC operating parameters can be found in Table 4.1. To maximize luminosity one can decrease the collision angle, maximize the number of particles per bunch, and squeeze each bunch. However there is a tradeoff to be made, because when the number of interactions per bunch increases it becomes harder to reconstruct some events. Multiple interactions per bunch crossing is called *pile-up*, and the reconstruction of pile-up is discussed in Section 5.5. The integrated luminosity delivered to CMS by the LHC is shown in Fig. 4.5.

Table 4.1: LHC design beam and operation conditions between 2010 and 2016.

Year	2010	2011	2012	2015	2016	Design
Center of Mass Energy (TeV)	7	7	8	13	13	14
Energy per Beam (TeV)	3.5	3.5	4	6.5	6.5	7
Proton bunch spacing (ns)	150	50	50	50/25	25	25
N_b ($\times 10^{11}$)	1.2	1.5	1.7	1.15	1.25	1.15
n_b	348	1331	1368	2232	2208	2808
β^*	3.5	1.0	0.6	0.8	0.4	0.55
ϵ_n	2.2	2.3	2.5	3.5	3.0	3.75
Peak Instantaneous $\mathcal{L}$ 10^{34}	0.02	0.35	0.77	0.52	1.53 (above design)	1
Total Integrated $\mathcal{L}$ (fb^{-1})	0.04	6.1	23.3	4.2	40.8	-
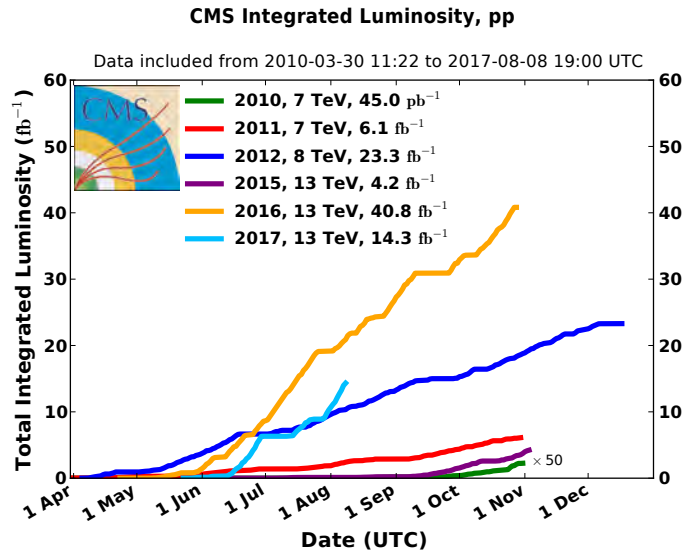
Figure 4.5: Luminosity delivered to CMS by the LHC split by year. Image collected CMS public page.

4.2 The Compact Muon Solenoid detector

The CMS experiment [80] is located at Point 5 of the LHC, in Cessy, France. The CMS experiment captures data during proton-proton and proton-nucleon Runs of the LHC. It is a hermetic general-purpose apparatus with goals of good di-jet and missing energy resolution, good muon identification and resolution, and good charged particle momentum resolution and identification. These goals were chosen in order to have good W and Z boson identification and energy resolution, and search for a Higgs boson with moderate mass, as well as any other new physics that may appear at the LHC energy scale. CMS is 28.7 m long and 15 m in diameter, while weighing 14,000 tons. The entirety of CMS is shown in Fig. 4.6.

One of the main components of a particle detector is the magnet, which is necessary for charge identification. The superconducting niobium-tin magnet in the CMS experiment is the most powerful magnet in the world. The choice of magnet motivated the compact nature of CMS. All the calorimeters are located within the magnet, except for the Outer Hadronic calorimeter (HO) which is designed to indicate when any energy has escaped through the

magnet. CMS is much smaller than ATLAS, which has a diameter and length of 25 m and 46 m respectively. Muons will easily traverse the CMS solenoid as intended, so an extra muon detection system is placed outside the solenoid to allow for extra precision of high p_T muons within CMS. The combination of many different technologies layered together that interact with the particles produced in the collisions work together to maximize the identification and resolutions of the particles and objets of interest.

Geometrically there are two shapes of CMS. The cylindrically-shaped center part is called the barrel. The two circular flat end pieces are called endcaps.
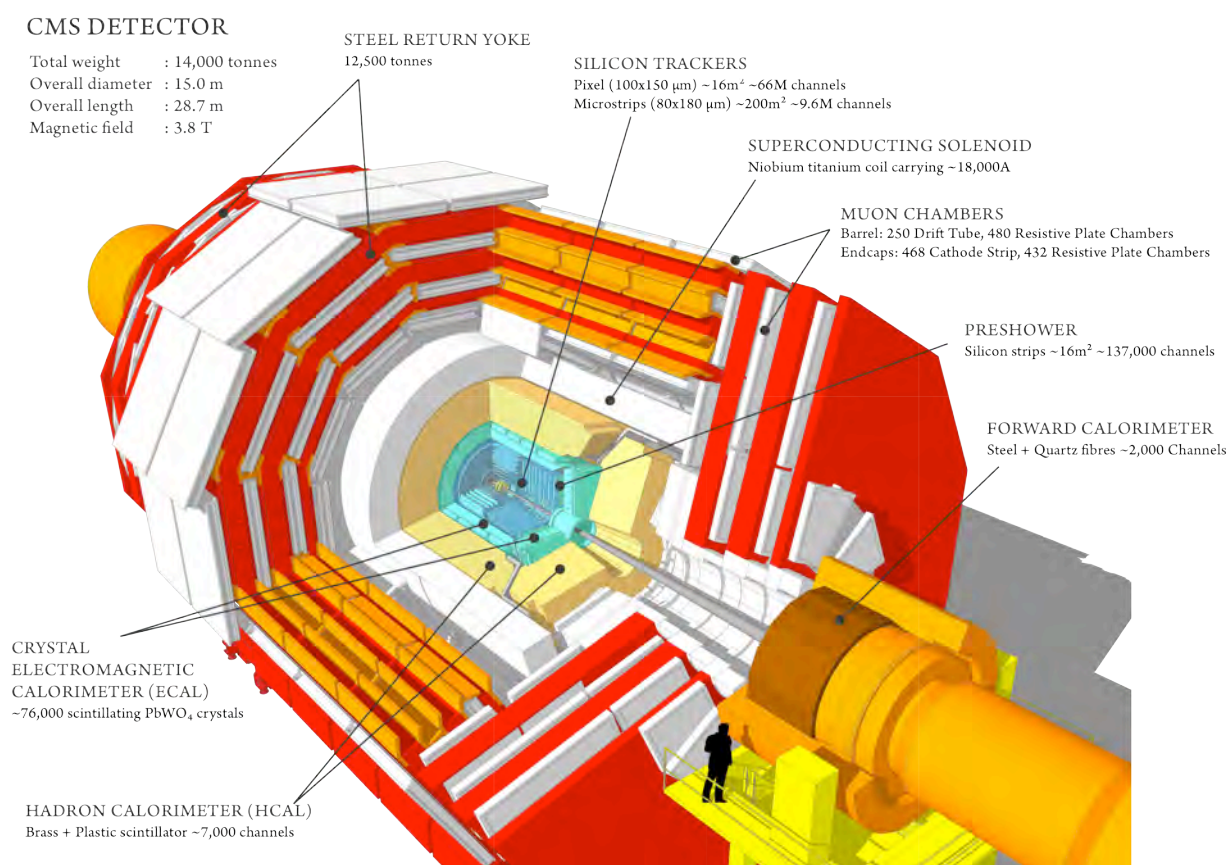
Figure 4.6: CMS layout

4.3 Coordinate system

The coordinates of the CMS detector are defined such that x-axis points to the center of the LHC ring southwards, the y-axis points vertically up from the detector, and the z-axis points westwards along the beam pipe.

The azimuthal angle ϕ, and pseudorapidity, η, are the most commonly used coordinates in CMS.

The azimuthal angle, ϕ, is defined as the angle in the transverse plane to the beam from the x-axis. The polar angle, θ, is measured directly by the detector as the angle from the particle in the rz-plane to the z-axis. A particle going along the beam pipe, would have $\theta = 0$ or $\theta = \pi$. Pseudorapidity, η, defined in Eq. 4.2, is used instead of the polar angle, θ, since η is lorentz invariant. Particle production is constant over η, as shown in Fig. 4.7, from Ref. [81].

$$\eta = -\ln\tan\left(\frac{\theta}{2}\right) \tag{4.2}$$

ΔR is a commonly used variable, constructed from these coordinates. If the CMS detector were to be unrolled ΔR is the area in the $\eta - \phi$ plane, i.e. $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

4.4 Superconducting solenoid

The niobium-tin wire superconducting solenoid is the defining feature of CMS. Inside the solenoid and 3.8T magnetic field is the tracking system, electromagnetic calorimeters, and hadronic calorimeters discussed in Section 4.5, 4.6, and 4.7 respectively. Outside the solenoid there are muon detectors embedded in steel flux return yoke, where the magnetic field is about 2T. The magnet is 5.9 m in inner diameter and 12.9 m long. The large magnetic field allows for precise momentum measurement of charged particles. It is composed of 2168
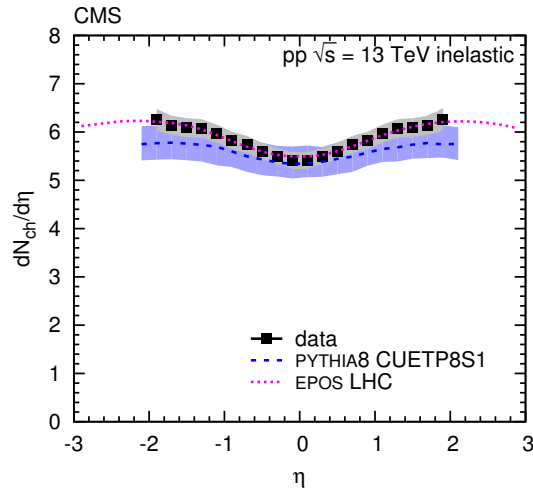
Figure 4.7: Observed and predicted charged track production as a function of η. The distribution observed is roughly flat as expected. This was the first 13 TeV result published by CMS.

turns with a 19.5 kA current. The CMS solenoid has a 2.7 GJ energy storage capacity at full field. While the magnet was designed to reach 4T, it is operated at 3.8T to prolong the lifetime of the magnet. A map of the magnetic field in CMS is shown in Fig. 4.8 [82].

The magnet is cooled to 4.7K with liquid helium, with a total cold mass of 220 tons. If it is quickly turned off, it will warm to a temperature of about 55K. A controlled dump will take about 5 hours with no temperature rise. The solenoid was designed to have a limited number of cycles; once the magnetic field is on during running it is designed to stay on. In 2015 operation, the CMS cooling system was stressed due to liquid helium filters clogging due to debris [83, 84].

For a highly energetic particle detected in CMS, the transverse momentum can be approximated

$$p_T \approx \frac{0.3 L^2 B}{8s}, \tag{4.3}$$

where the magnetic field, B, was measured precisely before operation using flux-loops and hall probes described in Ref. [85] and published in Ref. [82], and the sagitta (s) and distance
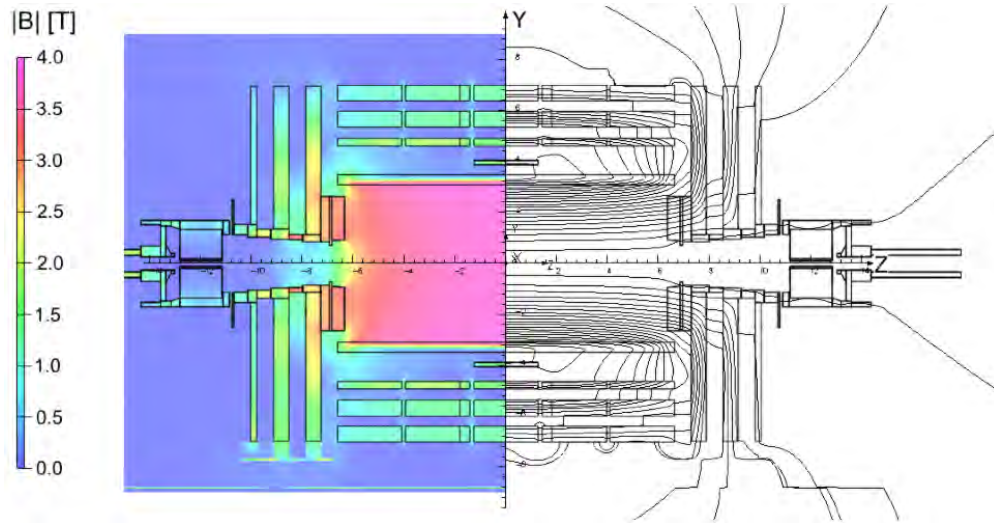
Figure 4.8: Map of the magnetic field inside the CMS detector. There is magnetic flux increment of 6 Wb for each field line.

from interaction point (L) are related to the radius of the track and calculated during event reconstruction, discussed in Chapter 6.

4.5 Tracking system

The CMS tracking system is critical for particle momentum measurement and for searches of any long lived particles with a displaced vertex. The tracking system cross sectional layout is shown in Fig. 4.9. It is designed to record tracks of charged particles. The curvature of the tracks is used for the precision momentum measurement. The tracks are also used to reconstruct secondary vertices of long lived particles, such as b-mesons or other new physics. The total system is 5.8 meters long and 2.5 meters in diameter with several detectors having η coverage up to 2.5. Recall that CMS is similar to an onion; Different detectors are placed in layers around the center of the detector, where the interaction point will be. The tracking system is the first major layer sounding the interaction point.

The innermost layer of the CMS tracker is the pixel detector which has 65 million pixels.

Cylindrical layers are placed near 4 cm, $\sim$ 7 cm, and $\sim$ 10 cm, with disks at the end around 34 cm and 46 cm, which corresponds to about $|\eta| \leq |1.5|$ for hermeticity. The high-granularity of the pixels allows for separation of neighboring proton-proton collisions within the same bunch crossing. Additionally charged tracks are assigned to a specific vertex, given the high-spatial resolution. The charged-hadron-subtraction pile-up estimation technique, discussed in Section 6.6, utilizes this information. Each pixel is 100μm by 150μm. The material is designed to be as minimal as possible to reduce multiple scattering and nuclear interactions. When a charged particle passes through a pixel, ionization creates electron-hole pairs. A voltage difference applied across the chip causes charges to build up on electrodes, and a "hit" is recorded. The pixels are mounted on cooling tubes.

Outside the pixel detector are the tracker strips, which consist of ten layers of silicon strips in the barrel region, and four layers in the endcaps. There are four inner barrel layers, two inner endcaps, six outer concentric layers, and two outer endcaps. The outer layers are not double-sided to reduce cost and material. One "module" usually has two silicon sensors (with several hundred strips) connected to one set of readout electronics. In total, the strips are composed of 15,148 modules, with ten million detector strips and 76,000 chips [86].

The tracker is cooled to 253K (-20C) to prevent lasting radiation damage to the silicon. If the silicon is operated at temperatures similar to above the currents originating from radiation damage are much reduced.

4.6 Electromagnetic calorimeters

The layout of the CMS electromagnetic calorimeter (ECAL) is shown in Fig. 4.10. It extends until $|\eta| = 3$. The goals of the ECAL are: good energy resolution and containment of electrons and photons from Higgs and W/Z bosons. Fine granularity in addition to having small volume are obtained. The small volume is critical to fit the calorimeters within the
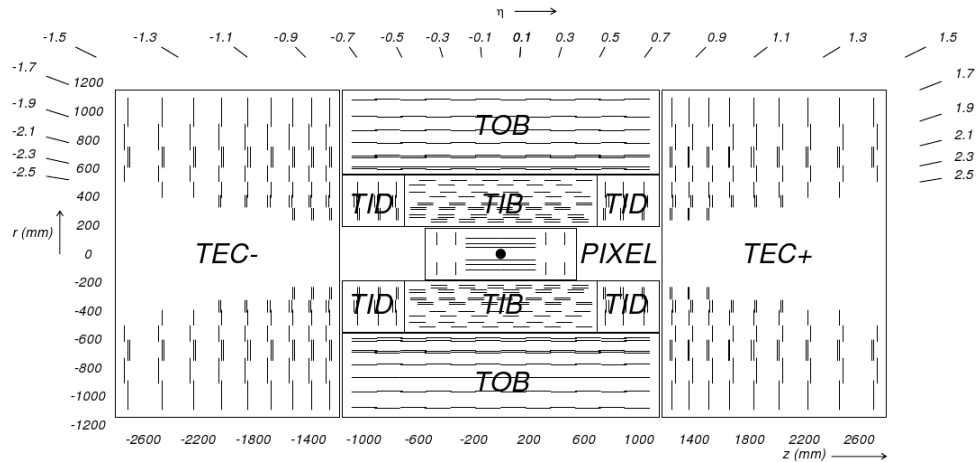
Figure 4.9: CMS tracker layout

magnet, and fine granularity is necessary for particle identification and optimal reconstruction, discussed in Chapter 6. For example, we want to be able to match the electromagnetic energy deposit from an electron with the electron's track.

Lead tungstate (PbWO$_4$) crystals were chosen for the ECAL, due to the high density and good light transmitting properties. The density of lead tungstate is 8.28 g cm^{-3}. When charged particles enter ECAL, the material quickly scintillates 420 nm (blue-green) light. About 80% of the light from the crystal is emitted within one bunch crossing of the LHC. Each crystal is almost 26 radiation lengths long, so the showers are mostly contained within one crystal. One radiation length is 0.89 cm, and the overall length of each crystal is 23 cm. The Molière radius is 2.2 cm therefore 90% of the energy deposit from an electron stays within one crystal, if the electron is centered within the ECAL crystal face. It took 10 years to manufacture all 75,848 crystals with production split between China and Russia. Avalanche photodiodes (APDs) are attached at the end of each crystal to collect scintillation light in the barrel, in the endcaps vacuum photodiodes (VPTs) are used. The barrel extends until $|\eta| < 1.479$, while the endcaps cover the range $1.479 < |\eta| < 3.0$ in pseudorapidity. However there is a slight gap in coverage. In front of the endcaps is a pre-shower detector

made of lead and silicon. The preshower detector is finer in granularity than the ECAL crystals and helps to discriminate between neutral pion decays and photons.

The energy resolution for photons from SM Higgs decays is between 1.1% to 2.6% in the barrel and between 2.2% to 5% in the forward endcaps [87]. The electrons from Z boson decays are even more precise with an energy resolution of 0.4% (0.8%) in the barrel (endcaps) [87].

The energy resolution of ECAL can be written

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right) + \frac{N^2}{E} + C^2 \tag{4.4}$$

where the stochastic term $S = 2.8\%$, the noise term $N = 0.12\,\text{GeV}$, and the intrinsic constant term $C = 0.003$. The stochastic term includes error originating from statistical fluctuations in the showers. The noise term includes noise from the analog to digital conversion and other electronic sources. The intrinsic term accounts for energy leakage and any energy mis-calibration.

ECAL is radiation resistant, however the scintillation response varies during the run as the ECAL crystals become more irradiated. ECAL has a laser system that is constantly monitoring the response of the crystals. The response is measured and corrections are correspondingly updated and applied every week during running. After a long shutdown of the beam the ECAL response gradually improves [88].

4.7 Hadronic calorimeters

Sandwiched between ECAL and the solenoid is the hadronic calorimeter (HCAL). The HCAL is brass/scintillator sampling calorimeter, with interspaced layers of plastic scintillators and brass absorbers. HCAL decided to use heavy brass absorbers, $8.53\,\text{g cm}^{-3}$ to ensure that hadronic showers (e.g. jets) at CMS deposited as much of their energy as possible before the
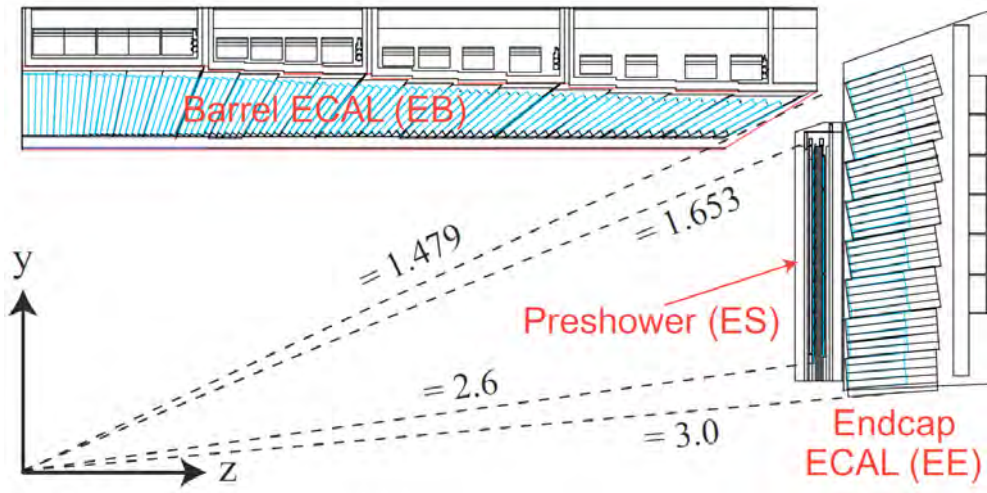
Figure 4.10: CMS ECAL layout

magnet. The HCAL is designed to be as hermetic as possible for accurate missing transverse momentum calculation and forward jets. The HCAL extends until $|\eta| < 5$. If the HCAL were to be placed outside the solenoid, the energy resolution would suffer. Figure 4.11 shows HCAL layout.
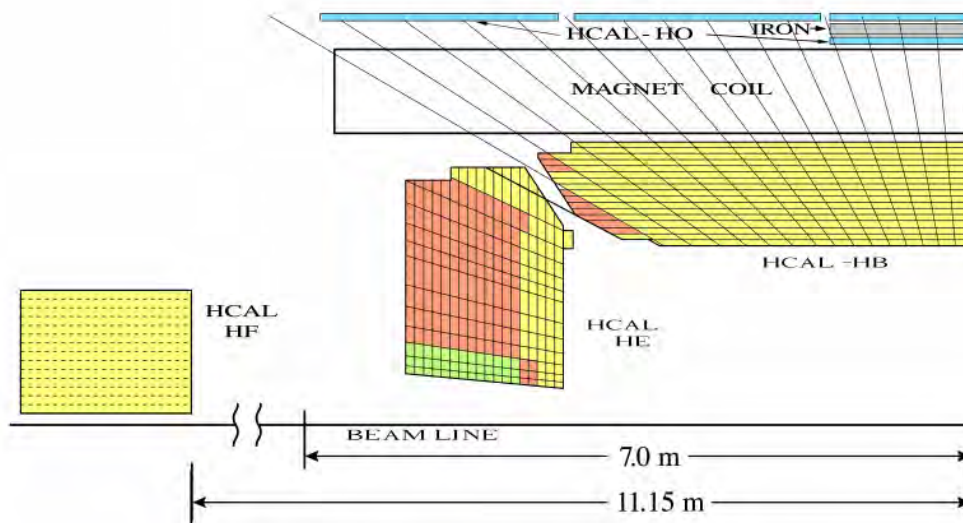
Figure 4.11: CMS HCAL layout

The plastic scintillators have a 7% sampling fraction. Brass was chosen because it is

non-ferromagnetic and has a radiation length of 1.49 cm and an interaction length of 16.42 cm. The interaction length is the mean path of a hadron before interacting with the material and the more important quality for a hadron calorimeter. The HCAL needs to be sufficiently big enough to reduce punch-through, instances where hadron shower remnants escape the calorimeters and leave deposits in the muon system. The barrel section of the HCAL (HB) extends until $|\eta| < 1.3$. It is divided into $\eta \times \phi$ towers with the size 0.087×0.087. The HB is nine meters long, six meters in outer diameter, and one meter thick. It is split into two half barrels, each composed of 18 wedges in ϕ and is about five interaction lengths deep. Wavelength shifting fibers are used for readout. The hadron calorimeter endcap (HE) $(1.305 < |\eta| < 3.0)$ is also composed of brass and scintillator. It is 1.8 m thick, about ten interaction lengths deep. The endcaps experience a very large particle flux. The forward hadron calorimeter (HF) completes the system going out until $|\eta| < 5.2$ and utilizes Cherenkov light radiation-hard quartz fibers in order to survive the high particle flux environment. HF is about 165 cm deep, which corresponds to about 10 interaction lengths.

Charged hadrons have a resolution of

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{115\%}{\sqrt{E}}\right)^2 + (5.5\%)^2 \tag{4.5}$$

for HB/HE measured in the barrel [89] and

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{280\%}{\sqrt{E}}\right)^2 + (11\%)^2 \tag{4.6}$$

in the HF where E is in GeV [90, 91].

Outside of the magnet there is one more HCAL sub-detector, the outer barrel (HO). This helps identify any punch through of particle showers that escape the HB and pass through the solenoid. The total interaction lengths a particle must get though to reach the muon system is 12 interaction lengths in the barrel.

4.8 Muon systems

Outside of the solenoid is the muon system composed of a magnetized steel yoke interspersed with muon-detecting chambers of three types: resistive plate chambers (RPCs), drift tubes (DTs), and cathode strip chambers (CSCs). All three sub-detectors are gaseous detectors that measure muon tracks. The muon system is placed outside the magnet, and is immersed in a magnetic field of 1.9 T due to the saturation of the steel return yoke. Excluding the tracker measurements, the muon system has a 10% resolution on the muon momentum when $|\eta_\mu| < 2.4$ and $p_T(\mu) < 200$. Including the tracker improves the resolution to 2% in the barrel and to 6% in the endcaps [92]. The layout of the muon subsystems is shown in Fig. 4.12.
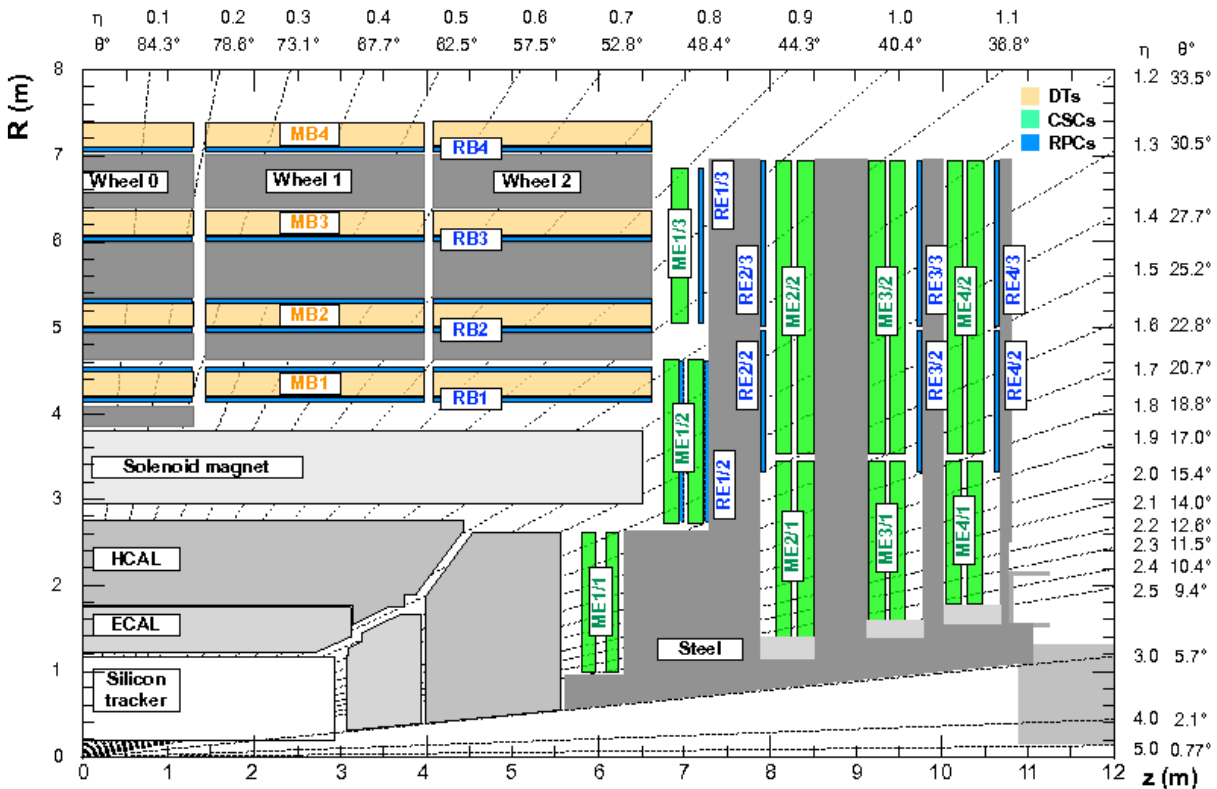
Figure 4.12: CMS muon chambers layout

4.8.1 Drift tubes

Drift Tubes (DTs) are only located in the central barrel region with $|\eta| < 1.2$. There are 12 segments in ϕ and four stations extending radially outward, separated by the layers of the yoke steel. There are five wheels in z. In all there are 250 drift chambers. The chambers are staggered in ϕ, to reduce muons escaping detection though a ϕ-gap. When a charged particle traverses the gas inside the 4 cm wide tube, the gas atoms ionize and the charges are collected by the wire passing through the drift cell, as shown in Fig. 4.13 reproduced from [93]. The drift tubes record the location where charged particles pass through the chambers. It has a 260 μm spatial resolution. Drift chambers are typically inexpensive to construct, however the DTs are slightly slower than the RPCs and CSCs. They do not perform as well in high particle flux and non-uniform magnetic fields [93]. The maximum drift time per cell is almost 400 ns [94].
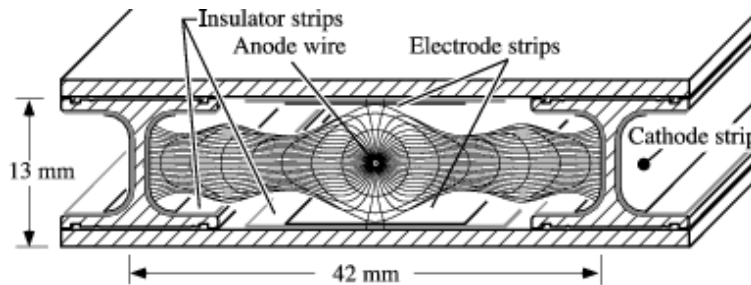
Figure 4.13: Drift cell schematic

4.8.2 Cathode strip chambers

In the endcaps there is a very high particle flux, and uneven magnetic fields. More expensive robust technology is used. The CSCs exist in the pseudorapidity region from 0.9 to 2.4. They have a 40-150 μm spatial resolution and provide muon hits in two dimensions. CSCs have a very fast anode response time, which is useful for triggering [95, 96].

The CSCs are trapezoidal shaped gaseous detectors, shown in Fig. 4.14 [95]. The wires are perpendicular to the strips in all chambers, except the innermost ones on ME1/1, where the wires are tilted 29°.
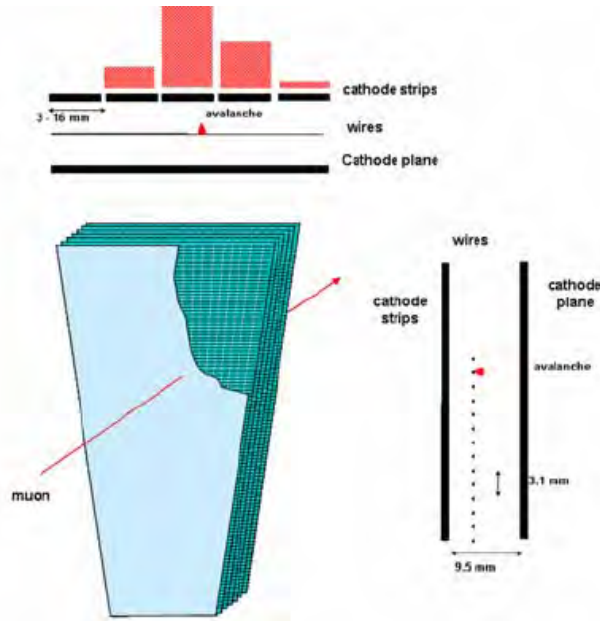
Figure 4.14: CSC schematic showing front and side views, and an image showing the avalanche distribution used to precisely locate muon hits.

In Run I of the LHC, there were 468 CSC chambers in CMS. By 2014, that number increased to include ME4/2 Chambers. ME1/1 was upgraded during LS1, to have better muon coverage in endcap.

4.8.3   Resistive plate chambers

The double-gap parallel plate capacitors are composed of two charged plates surround a layer of gas on each side, with the setup shown in Fig. 4.15. Resistive plate chambers (RPCs) are constructed using six layers of these detectors in the barrel and three layers in the endcaps. The gap is made air-tight to reduce cost and improve functionality. The outer surface is coated with conductive graphite paint. The readout strips are located between the two

gas chambers. When a charged particle passes through the gas, the gas ionizes causing an "avalanche" of electrons. The readout strip measures the collection with a time resolution of 1 ns. The spatial resolution 0.8-1.2 cm is larger than the other two systems. The RPCs extend out to $|\eta| < 2.1$ as shown in Fig. 4.12. The RPC barrel and endcaps geometries differ. In the endcaps, the RPCs are shaped similarly to the trapezoidal CSCs, while in the barrel the strips run parallel to the beam and more closely match the DT configuration.
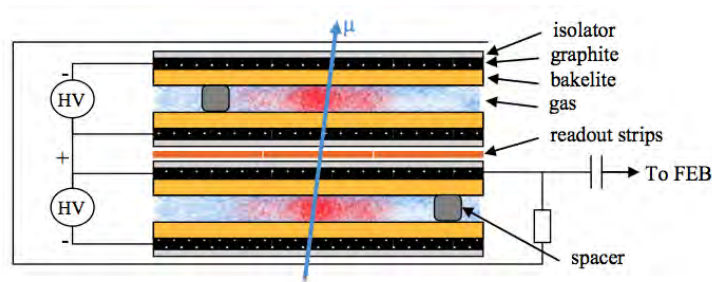
Figure 4.15: CMS RPC layout

4.9 Trigger system

The LHC operates at 40 MHz and storing that much data is impossible with current technology. Each event is about a megabyte. Additionally, the vast majority of LHC collisions are inelastic collisions, and consequently are not of interest. The trigger system does the first pass of finding the interesting collisions. Only the data containing the most interesting possible physics can be stored. With the 25 ns bunch spacing, the trigger system also must operate at 40 MHz, as is is timed in to the LHC clock. The CMS Trigger system is split into two levels: hardware-based Level-1 trigger (L1) and the software-based High-Level Trigger (HLT). The firing of the acceptance will *trigger* the data recorded during that bunch-crossing to be saved offline. Only the data collected via the trigger will be saved for further analysis.

4.9.1   Level-1 trigger

Due to requirements from the CMS tracking system readout, L1 only has 4 microseconds in latency to decide (Level-1 Accept shortened to L1A) initially if a recorded bunch-crossing should be kept for further processing. The output rate of L1 is 100 kHz. The L1 does not have time to analyze nor access the full detector data, such as tracker information and full ECAL granularity.

Before the L1A, the L1 is split in two parallel paths, a calorimetric path and a muon path. The muon path gathers data from the barrel and endcaps for the muon system and makes rudimentary tracks from them. The muon detector track finder systems consist of an endcap region($1.25 < |\eta| < 2.4$), a barrel region($|\eta| < 0.85$), and an overlap region ($0.85 < |\eta| < 1.25$). The resulting tracks are the muon candidates.

The L1 calorimetric path has access to ECAL and HCAL trigger towers, which are mostly uniform $\eta - \phi$ size of 0.087×0.087. The ϕ size of towers is constant through the detector, but the η size gets larger beyond $|\eta| > 2.1$. One ECAL trigger tower is a 5x5 sum of ECAL crystals, with the corresponding HCAL tower located directly behind the ECAL one. The algorithms on the calorimetric path are based around basic physics objects; there are trigger algorithms for e, γ, τ_h, jets and energy sums. Algorithms are developed specifically to operate quickly, while maximizing efficiency of the process of interest and minimizing the acceptance of backgrounds.

L1 Calorimeter Trigger is split further into two layers. The Layer-1 processes and calibrates the incoming trigger towers from ECAL and HCAL to prepare the data-stream for Layer-2 [97,98] Layer-1 is composed of 18 CTP7 [98] processor cards, with 864 total optical links to pass the resulting data to Stage-2 Layer-2. The Layer-2 uses a time-multiplexed card setup and has more of the algorithms to find the objects [99].

Once calorimetric objects are constructed they are sent on the global L1 trigger, along

with the muon objects reconstructed from the muon trigger track-finders, where the L1A is made based on a predetermined menu of triggers, which is prepared ahead of collisions.

4.9.2 High-level trigger

The HLT runs on a commercial computing farm and it is a highly configurable system, since it is software-based. Compared to the L1, it has a long time about 200 ms to reconstruct events and access to the full detector readout. The output rate of the HLT to archive media is about ~ 1000 Hz.

With about 200 ms per event to decide if an event should be accepted, first events are filtered through a hierarchy of simplified algorithms with access to all detector information. Once an event passes through this, the HLT does a full-detector reconstruction so that the decisions are based on quantities which are more similar to offline quantities. This length of time allows for secondary vertex tagging algorithms as well as other complex CPU-intensive algorithms. Most objects are reconstructed according the "particle flow" algorithm discussed in Section 6.2, however τ_h are notably reconstructed with a jet-based isolation. [100]

Similar to L1, there is an HLT trigger menu, used to decide if an event *passes* or *fails* at HLT. The menu consists of ORs various trigger paths. If a full trigger path passes than the event is kept. Each path includes various filters, such as a E_T^{miss} cut or N_{jet} cut. Once an event passes a full path, the event is saved to disk.

4.10 Luminosity measurement

An important measurement at CMS is the luminosity measurement. Five detectors are used at CMS to measure luminosity, which is associated with a rate of an observable. The silicon pixel detector, the barrel DT, the HF, and the Fast Beam Conditions Monitor (BCM1f) and the Pixel Luminosity Telescope (PLT) are all used in the luminosity measurement.

Luminosity is measured offline and online. The online measurement takes place continuously during data-taking, by the PLT and BCM1f as part of the Beam Radiation Instrumentation and Luminosity (BRIL) project. These two detectors run independently from the CMS detector and are synchronized to the LHC clock. Beside luminosity measurement, the BRIL project at CMS aims to protect the existing instrumentation at CMS from radiation damage as well as monitor beam conditions, by providing a safety beam-dump.

Both of these BRIL detectors can provide online luminosity measurements. The PLT is a small silicon pixel detector positioned like endcaps around the beam. The number of fired pixels is used to estimate luminosity. BCM1f is composed of 24 single crystal diamond sensors arranged in a ring around the beam pipe 1.8 m away from the interaction point. They are connected with the LHC clock. The HF can also monitor online luminosity.

The pixel detector and the DT are also good tools for luminosity measurement; they have good performance throughout the run and low occupancy. However both of these systems require the trigger to have fired to collect data, thus they are only useful for an offline measurement.

Offline luminosity is measured by pixel cluster counting (PCC) which should have good linearity until 150 pileup environment [101]. The mean number of pixel clusters in each bunch crossing, $\langle N_{cluster} \rangle$, is used in conjunction with the minimum bias cross section, σ_0, and LHC frequency, f, to estimate luminosity as

$$\mathcal{L} = \frac{\langle N_{cluster} \rangle f}{\langle N_{cluster/interaction} \rangle \sigma_0}. \tag{4.7}$$

Van der Meer (VdM) scans are performed at the beginning of yearly runs, to calculate the size of the beams to get the σ_0 value. VdM scans pass the beams through each other to measure the size of the beams, to determine the cross sections of the colliding beams [102]. The total uncertainty on the luminosity measurement is 2.5% [101].

Chapter 5

Monte Carlo background and signal simulation

Monte carlo (MC) simulation of proton-proton collisions is used for many purposes in high energy physics. The primary purpose, for this thesis is to simulate backgrounds and signal processes, to provide an estimate of the backgrounds, and a model with which to test the data. Some other uses within CMS include planning future upgrades to design detector components, planning trigger menus, and designing reconstruction techniques. The basics steps to model proton-proton collisions at the CMS experiment are briefly summarized in this chapter. The hard scatter and underlying event are modeled with event generators and parton distribution functions (PDFs), and then that event simulation is passed through software that models the passage of particles through matter and the detector response to those particles. The final samples (set of simulated processes) are in the same format as stored data, so they can be reconstructed in the same manner as the data.

There are various MC generators which can simulate signal and background events.

Three main event generators can fully model a hadron-hadron collision event: PYTHIA8 [103], HERWIG++ [104], and SHERPA [105].

Of the above generators, this thesis only utilizes PYTHIA8, which has a large internal library of available processes. The standard $2 \rightarrow 2$ processes from PYTHIA6 and newly available $2 \rightarrow 3$ in PYTHIA8 QCD processes allow for the simulation of the underlying event, parton showering and hadronization. QCD, Electroweak, Higgs, and top LO processes among some other BSM models are also included. However, while PYTHIA8 can indeed be used as a full event generator for some processes, it doesn't have enough hard processes available for the samples in this thesis and it calculates a lower perturbative order. In this thesis, it is used for the showering and hadronization. Apart from PYTHIA8, other generators used in this thesis to calculate the hard scatter process are: POWHEG [106, 107] and MADGRAPH5_AMC@NLO [108,109]. The POWHEG (Positive Weight Hardest Emission Generator) and MADGRAPH5_AMC@NLO generators are interfaced with PYTHIA8 [103] for parton shower, hadronization, as well as τ-lepton decays.

POWHEG 2.0 is used for samples of $t\bar{t}$ [110], and SM 125 GeV scalar boson production [111,112,112–115]. The SM Higgs processes produced in association with a vector boson, WH/ZH processes, are generated with the MiNLO HVJ extension for POWHEG 2.0 [116]. POWHEG 1.0 is used for the single-top-quark [117]. The leading-order tree-level (LO) in perturbative QCD samples with Z+jets, W+jets processes [118], and a 125 GeV scalar boson produced in association with dark matter are generated via MADGRAPH5_AMC@NLO 2.2.2 [109]. All diboson production is computed to at least next-to-leading-order (NLO) in perturbative QCD. MADGRAPH5_AMC@NLO 2.2.2 is used for some multi-boson samples, such as the tri-boson samples (ZZZ,WZZ,WWZ,WWW), when the corresponding POWHEG sample is not available. The various production cross sections can be calculated using each sample's respective generator. Usually the measured cross section within CMS is used, or the cross sections are computed using higher order calculations.

5.1 Hard process generators

To produce a simulated event, first the hard scatter is calculated. The cross section is factorizable and can be split into the hard process, discussed in this section, and the normalization, discussed in Sec. 5.2.

MADGRAPH5_AMC@NLO is the most common generator used in this thesis, it can be run at LO or at NLO. It operates under the principle that the construction and structure of the hard part of the cross section is independent of the actual physical process under consideration. For input, MADGRAPH5_AMC@NLO needs a theoretical model, supplied by Feynman diagrams at LO, and a process (initial state and final state particles). Matrix elements are generated from all relevant Feynman diagrams for the process [119]. At LO, FEYNRULES [120–122] supplies the Feynman diagrams and it is fully automated. At NLO, Feynman diagrams for one order higher in QCD are provided by FEYNRULES while other NLO contributions, such as UV counter-terms, are added separately using a dedicated computation [109].

The AMC@NLO output includes code[1] to evaluate the computed helicity amplitudes and to produce pictorial representations of the Feynman diagrams for the user.

For more realistic predictions in hadron-hadron collisions, one should consider a parton radiating off the initial or final state partons. The AMC@NLO[2] solution involves merging fixed order in terms of the number of additional final state partons calculations with a showered, discussed in Section 5.3, monte carlo.

However for the NLO with parton shower (NLOWPS) setting, there is a potential double counting problem, when matching the 'PS' term to the hard process [109]. Some simulated events are assigned negative weights to avoid this double counting at NLOWPS in MADGRAPH5_AMC@NLO. Specifically the PS term is represented as an analytical form and it

---

[1]Python, C++, Fortran
[2]integrated into MADGRAPH5_AMC@NLO

is subtracted to remove extra NLO overlap. But when the PS analytical term is larger than the hard process, the event can receive a negative weight [109]. These negative weights are included as mentioned in Section 7.5.8.

The other main generator used in this thesis, POWHEG, utilizes a different solution to avoid negative weights, and, as described in the name (Positive Weight Hardest Emission Generator), only generates positive weights [106,107]. Thus POWHEG avoids the loss of statistics, which can be as high as 30%, associated with negative event weights from AMC@NLO. POWHEG is usually the preferred NLO solution when high-statistical precision of certain samples is critical. It first computes the matrix element along with the hardest emission in the event and frequently is able to produce a better event description of the p_T spectrum and hard jets [106,107]. Unlike MADGRAPH5_AMC@NLO, POWHEG is not an automatic generator, so limited processes are available for use. Each process needs to be coded separately and when a process is added to the POWHEG library, it is typically validated and released in a publication. As new recommendations are released, CMS MC production is updated. The SM Higgs processes have been the topic of many publications [111,112,112–115]. This is in contrast to MADGRAPH5_AMC@NLO, where the code to produce the hard process is part of the output and so can be considered roughly automatic.

5.2   Parton distribution functions

At high-energy lepton-lepton colliders, such as LEP, there is no substructure to the electrons. The cross sections can be calculated directly from Feynman rules.

Protons are composite particles, as discussed in Sec. 1.2, and additional information on the proton composition is needed to compute the normalization part of the factorized cross section. This additional information is the "parton distribution function," shortened to PDF. Accurate PDFs must be included for the normalization term to be correct. Without

a realistic model of the initial state, generated final states are not correct. The calculations of these cross sections require perturbative calculations, originating from the strong interactions, and non-perturbative calculations, which use PDFs.

Specifically for LHC production, the proton PDF NNPDF3.0 [123] is used for matrix element generators in this thesis. NNPDF3.0 replaced the older NNPDF2.1 [124]. This has been found to be reliable at the CMS Experiment. [125] The NNPDF3.1 [126] has been released and are reproduced in Fig. 5.1. CMS production of simulated events has switched to these updated recommendations in 2017.
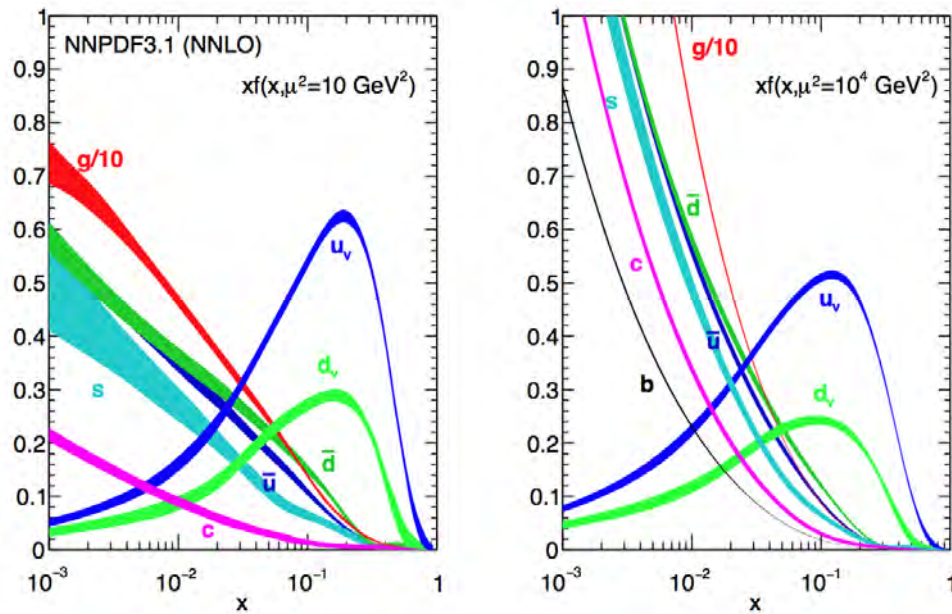
Figure 5.1: The NNPDF3.1 parton distribution dunctions for demonstration of general PDF structure.

5.3 Parton showering

After the hard process is calculated, the simulated partons from the process are showered. Outgoing quarks and gluons(partons) are not "free" and may radiate more, mostly soft

partons. Each produced parton has a probability of radiating another parton within a range of momenta. It is a p_T iterative process and multiple partons can be emitted through a parton shower. In this thesis, the parton shower is simulated perturbatively with PYTHIA8 8.212.

Additionally the underlying event (UE), defined as everything not originating from the hard scatter, is responsible for a large portion of particle multiplicity and energy deposits in LHC collisions [127], The underlying event can come from colliding-remnants of the color-connections between the incoming partons in the protons, or from other multi-parton interactions (MPI), where one proton collision causes several scatters [128, 129]. This is also simulated and the partons from the UE are showered as well. The many PYTHIA8 8.212 parameters affecting the description of the UE and other showering is set to the CUETP8M1 tune [125], where it stands for "CMS Underlying Event Tune Pythia 8 Monash set 1". This is the recommended tune [125], and includes two extra energy-dependent MPI parameters and other updated parameters in the Monash tune [130]. The tune can affect the shape and behavior of the jets, which can therefore in turn affect the jet identification, rate of misidentified taus, and tau efficiency. The effect of different tunes on acceptance can be factored into measurements with tunes of uncertainties on the signal processes.

5.4 Hadronization

There are two models frequently used to model jet hadronization and both are shown in Fig. 5.2. The Lund string model (a) as in Fig. 5.2 is used in PYTHIA8, and in this thesis [17].

Of special notice in this thesis is the treatment of τ decays. In Run-II, τ decays are now simulated in PYTHIA8. Initially, in Run-I TAUOLA [131] was used to decay taus, as PYTHIA6 did not include proper treatment of the τ spin in the decay [131]. Now spin correlations are included, based off the work done in TAUOLA and HERWIG++ [131–133]. All τ decay
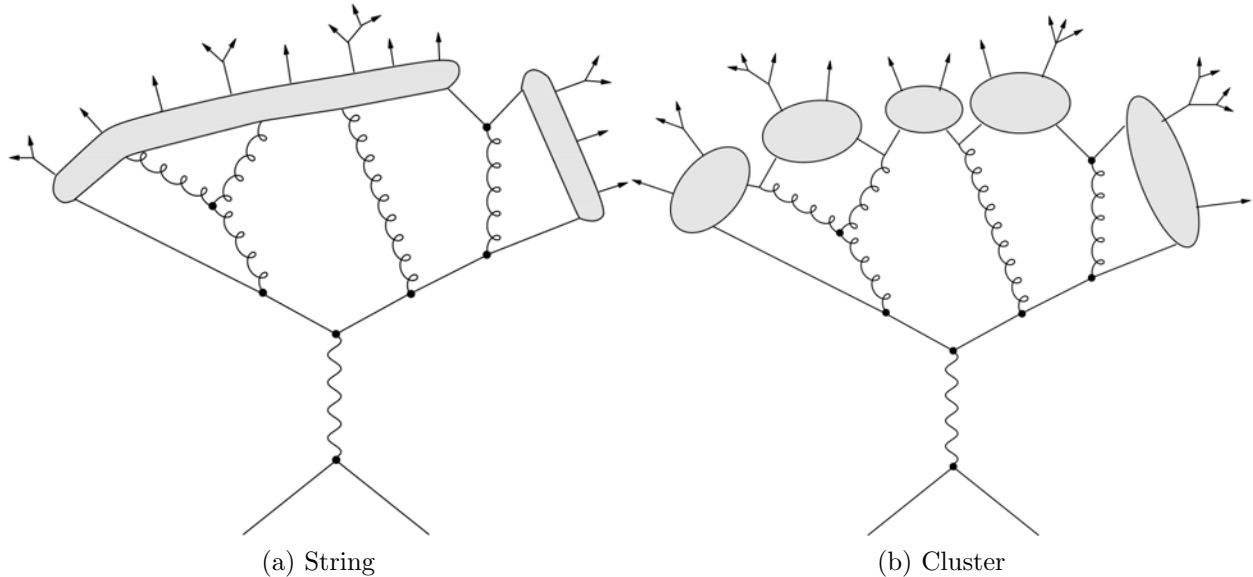
(a) String   (b) Cluster

Figure 5.2: Cartoon demonstrating the differences (a) Lund string model used by PYTHIA8 and the (b) cluster model which is used by SHERPA. Hadronization is represented by the gray bubbles.

modes with branching ratios $> 0.04\%$ are included in simulation [132].

5.5   Pile-up

The number of interactions per bunch crossing at the LHC, called "pile-up," is included the simulation. The number of interactions in a sample is typically a gaussian distribution at a specific luminosity, and intended to be as close as possible to predicted LHC settings. Typically as the luminosity rate increases, the pile-up increases. The same factors that increase luminosity, such as protons per bunch, also increase pile-up. On average the pile-up for the data set collected in 2016 is 27 interactions per bunch crossing. Most proton-proton collisions will not feature a hard process, and instead are usually composed of softer QCD processes. These are called Minimum Bias MB events. To replicate this, many MB events are produced in PYTHIA8. Processes included in the MB events include multi-parton

interactions. The MB simulation is produced and then embedded as background originating from other vertices in the MC. Many of these events need to be generated to avoid repeatedly embedding the same MB event in a sample.

The pile-up events and primary vertices are all contained in the luminous region, the region in the detector where the events are distributed. In this thesis, the primary vertices are required to be within 24 cm of $z = 0$. All the vertices occur within the pixel detector for optimal tracking.

5.6 Detector simulation

Once all the particles are accordingly showered and decayed, the resulting particles are passed through the GEometry ANd Tracking, (GEANT4) [134, 135] toolkit, which simulates the passage of particles through matter and the detector response to those particles. The HEPMC output format is used by the above processes and is interfaced to GEANT4. GEANT4 includes tools for defining geometry of materials, tracking, and detector response. The software is in use by both CMS and ATLAS in the final step of the simulation process. A high-resolution computer model of the CMS detector is input into GEANT4 with the corresponding properties of all material used within the detector, as well as the magnetic and electric fields. All materials are included in the material model, even the electronics wires, cooling systems and any other absorbers. GEANT4 is able to simulate γ,e's,μ's, and charged and neutral hadrons between energies ranging from a few eV to PeV [135]. GEANT4 operates iteratively by moving a particle a certain distance and estimating if an interaction will occur. While it is practically impossible to fully eliminate all differences between the simulated detector conditions and actual conditions during data-taking, GEANT4 does an excellent job predicting behavior of particles within CMS. The level of agreement between data and simulation can be seen in many public plots showing agreement between recorded

events and simulation, for example shown in Fig. 5.3 [136, 137]. GEANT4 is continually validated in the form of routine physics object performance comparisons between collected data and GEANT4 simulation.

The GEANT4 output for CMS simulation includes simulated hits in the detector. These simulated electronics hits in the detector are saved in the same format as collected data. Therefore, when the simulation and data are passed to the reconstruction step, discussed next in Chapter 6, they are identically treated, further minimizing any differences between data and simulation.
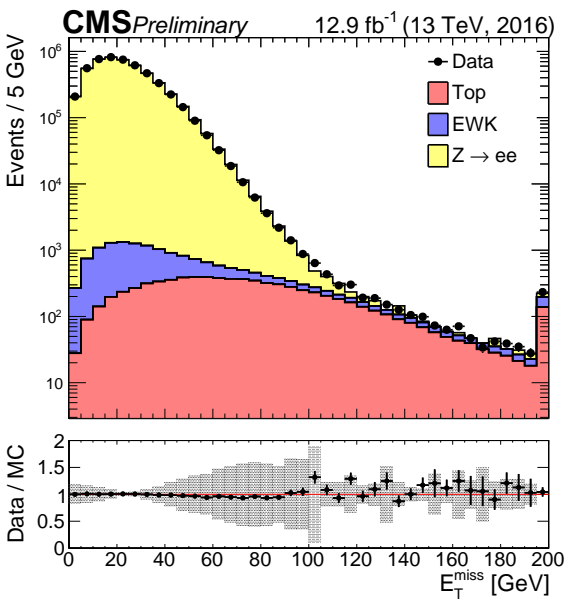
(a) 13 TeV $E_\mathrm{T}^\mathrm{miss}$ distribution
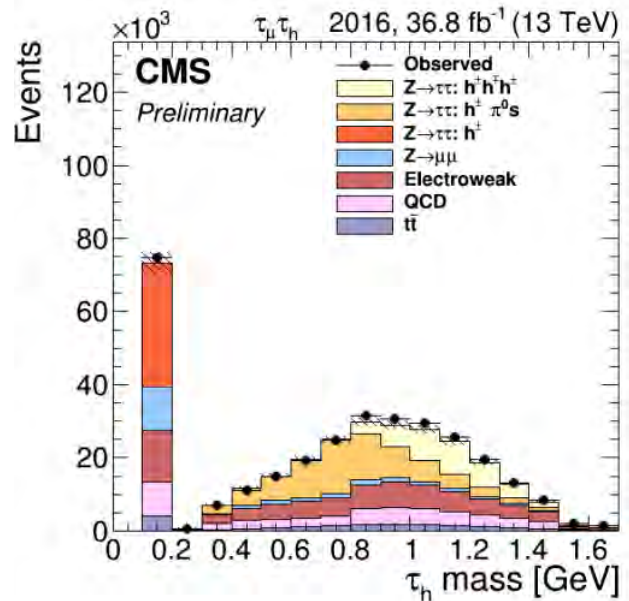
(b) 13 TeV τ mass

Figure 5.3: Reconstructed distributions of (a) missing transverse momenta, $E_\mathrm{T}^\mathrm{miss}$, showing excellent agreement between data and simulation and (b) τ mass from CMS showing data-simulation agreement with a very complicated reconstructed object. This demonstrates the level of agreement obtained through the simulation (and reconstruction) chain for complicated and sensitive variables.

Chapter 6

Event reconstruction and Particle Flow

In this chapter I discuss the overall reconstruction of physics objects from detector signals. CMS physics-object reconstruction utilizes all of the detector information simultaneously to determine the type of particle, reconstruct particles, and provide the four vector of reconstructed objects. For example, rather than treating a hadronization shower from a quark as many particles, it is treated as one object, a "jet," and given a single four-vector. The Particle Flow (PF) algorithm [138] is used by CMS to identify, and reconstruct physics objects from the electronic signals recorded by the detector. The idea behind PF was first implemented by ALEPH [139]. PF reconstructs muons, electrons and isolated photons, and hadrons[1] and non-isolated photons. From these components, jets, taus, and missing energy are further reconstructed using other algorithms. Before PF can reconstruct these objects, two basic objects tracks and clusters are formed, discussed in the next section.

---

[1] the various types of hadrons are not distinguished by PF

6.1 Input to particle flow

Out of the detector we have several electronic signals. "Hits" are recorded in the tracking layers and energy deposits are recorded in calorimeters.

Out of hits, we make tracks based on a Kalman-Filter (KF) [140–142] approach[2]. Initially a seed track, composed of a few hits that could be from a charged particle passing through the tracking layers, is passed to the track finder. Next from that seed track, other physically compatible hits are added in succession to the seed track. The last step in the track finding is the parametric regression that is applied to the track to determine the curvature and the vertex of the track. This last step provides the physically relevant momentum of the charged particle. Additionally there are further requirements on each track: reasonable χ^2, two hits in the pixel detector, at least eight total tracking hits, and a curvature corresponding to a momentum greater than 0.9 GeV. The additional muon tracking layers boost the reconstruction track efficiency of muons to around 99% [94, 143] and significantly improve the muon momentum resolution for muons above 200 GeV. All the tracks are identified as coming from a specific vertex. The vertices are reconstructed [144] after the tracks are identified. The vertex which is identified as the vertex of interest is called the "primary vertex," discussed in Section 6.9, is chosen after all the objects have been reconstructed.

At this stage of tracking about 80% of tracks are reconstructed [145]. Iterative track finding [144], also called combinatorial track finding, is deployed. Iterative track finding uses the same method as above but further creates tracks out of the remaining hits in 10 tracking steps. A track with eight hits indicates it has traversed the beam pipe, pixel detector, the inner tracker, and part of the outer tracker before undergoing a significant nuclear interaction, therefore indicating a high-quality track. Tracks with eight hits do not receive additional quality criteria cuts on the track fit, χ^2, and on track compatibility with

---

[2]similar to a global Least-Square

primary vertex. After each successive round the hits are removed from the tracker, and remain available for the next iteration of track finding to be used to reconstruct a track. This continues until no more tracks are found. The last iterations of track finding are used to find "displaced tracks", that is tracks which start slightly displaced from the vertex. This is done in seven iterative steps. As the tracks become higher energy they become straighter, and more likely to be mis-reconstructed, as shown in Fig. 6.1 reproduced from [138]. Any high-p_T track missed in this charged tracking step, may be measured by deposits in other sub-detectors, but potentially mis-identified by type with poorer energy resolution. Nuclear interactions between the charged particles and the silicon tracker may cause a "kink" to appear in the tracks, this is also considered in the iterative tracking program [138, 146].

(a) Efficiency   (b) Mis-reconstruction
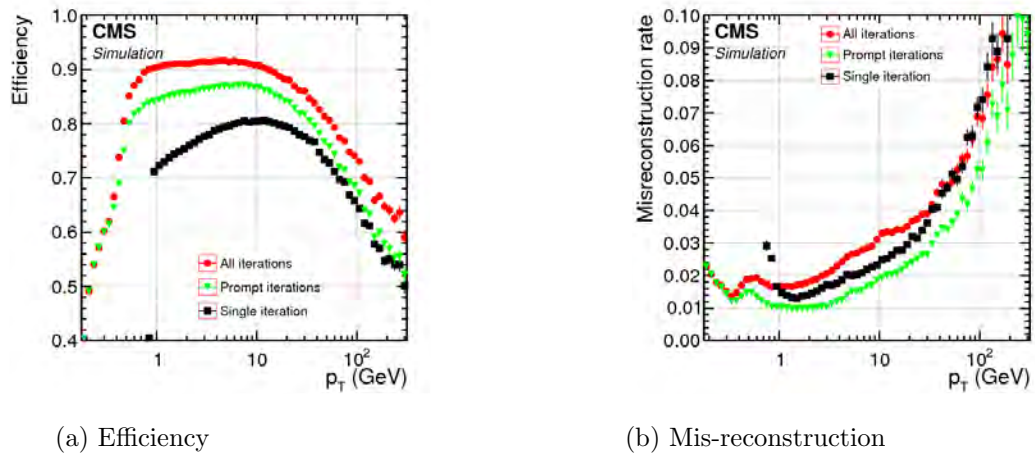
Figure 6.1: Efficiency and mis-reconstruction rate of tracks in CMS.

Muons rarely interact in the calorimeters of CMS. They are entirely built from tracks. These muon track candidates are input to PF. Standalone, tracker and global muons are the three candidates for PF construction. The hits from only the outer muon detectors form standalone muon candidates, while those from the inner tracker with one hit in the muon system are defined as tracker muons. Global muons are those where the muon track is found in both the tracker and the muon system. In this thesis, only global muons are

considered [1,2]. Global muons are usually above 10 GeV because they must pass through the steel return yoke. A very small percentage of charged hadrons with significant momenta will be mis-identified as global muon tracks, due to "punch-through," where the hadrons escape through the calorimeters and leave tracks in the muon system [143]. Another large source of problematic background muons are produced from in-flight hadron-decays within jets.

Calorimeter clusters are passed to the PF reconstruction. Calorimeter clusters are made in ECAL and HCAL sub-detectors separately. A local maxima (cluster seed) in a sub-detector is identified between four (HCAL) to eight (ECAL) cells and a topological cluster is built around the seed within an envelope. An iterative algorithm is used to refine these clusters [87,138]. These clusters are further calibrated separately by sub-detectors to optimize response. ECAL is calibrated using photons from π^0 decays [87]. HCAL is calibrated using 50 GeV charged pions which did not interact in ECAL. However charged hadrons interact differently in the ECAL, so charged hadrons receive a separate calibration when interacting with both ECAL and HCAL. They are derived from K_L^0 simulation, and checked against charged hadrons collected in data for closure.

Information given to the PF algorithm for electrons include "superclusters," which are made up of significant energy clusters in ECAL of electrons and nearby bremsstrahlung (brem) photons. These superclusters provide identification and energy measurement for isolated electrons, but for non-isolated electrons a special tracking algorithm is employed to ensure that ECAL energy is properly associated with the electron and not to other objects. The tracking information also helps correct for energy not included in the supercluster. The tracks found from iterative tracking with $p_T > 2$ GeV are considered as seeds for electron tracking to ensure no electrons are missed as input to PF reconstruction. Electrons are more likely to radiate than charged hadrons. The track seed's cluster energy must be close to the track momentum. These selected tracks are fit again with a Gaussian-Sum Filter

(GSF) [147]. The electron will radiate more in the tracker. The GSF is more robust to changes in track p_T [147]. Additional discrimination between electrons and charged hadrons is provided from a small BDT. The input to the electron BDT includes the ratio of the χ^2 of the GSF track fit to the χ^2 of the KF track fit, the amount of energy lost along the track, number of hits, and the extrapolated difference between the associated ECAL inner surface and the associated ECAL cluster. All the track information and already calibrated calorimeter clusters are passed to the PF reconstruction described next.

6.2 Particle Flow

The guiding principle of particle-flow (PF) reconstruction [138] is the combination of information from different parts of the detector to produce the best particle measurement and identification. The PF algorithm attempts to reconstruct physics objects, rather than have standalone tracks, deposits, and hits interpreted into objects separately. All particles start out at the interaction point and pass through the tracker; the direction of the curvature of the tracks identifies the charge of the particles, while the bend determines the momenta. Photons and electrons both leave deposits in the ECAL, but they are distinguishable by the presence of tracks. Muons curve in the inner tracking system, similar to electrons, but muons easily traverse the calorimeters and leave tracks in the outer muon detectors allowing for separation of muons from both charged hadrons and electrons. Figure 6.2 summarizes how different particles interact in the different technologies used in each layer of CMS.

A critical tenet of PF is fine-enough granularity to separate particle deposits. The first step of particle flow is the "linking algorithm," where certain hits and tracks are grouped such that information from all sub-detectors is available for one primitive "object." However PF can only link together track and clusters information if the resolution and granularity of the detector systems is sufficient enough to resolve individual particles and unambiguously

associate the information from two different detector systems. Further details of each reconstruction object can be found below. Once an element is used in PF, it is usually no longer considered for further processing of PF objects.

6.3   Photons

Photons can be identified by a strip of deposits in the ECAL, with little or no deposit in the HCAL. The photons may pair produce and in turn be recorded as a strip by the detector. Photons are also characterized by a lack of associated tracks in the tracker. In this thesis, photons are only used within the tau reconstruction where a τ may include several pions in the decay. This will be discussed more in depth in Section 6.8.

6.4   Muons

Muons appear as curved tracks in the tracker and in the muon system. In this thesis the global muon definition, where the muon track exists in both the tracker and the muon system, is required. Furthermore the muon is required to be well isolated. The PF algorithm uses track and cluster information to better calculate the amount of energy deposited around the muon. To reject muons coming from heavy hadron decays or mis-identified muons from punch-through, the energy from clusters and tracks within a cone of $\Delta R < 0.4$ around the muon is required not to exceed 10% of the muon p_T. For non-isolated muons, identification criteria beyond isolation are employed [138, 143]. There is a p_T assignment available from several different muon fits. For muons with $p_T < 200$ GeV, the chosen p_T assignment comes from the tracker-only fit since the muon is sufficiently curved. For higher p_T muons, the momentum is taken from four different track fits with the smallest χ^2. The global-track fit is normally chosen for high-p_T muons [138, 143].
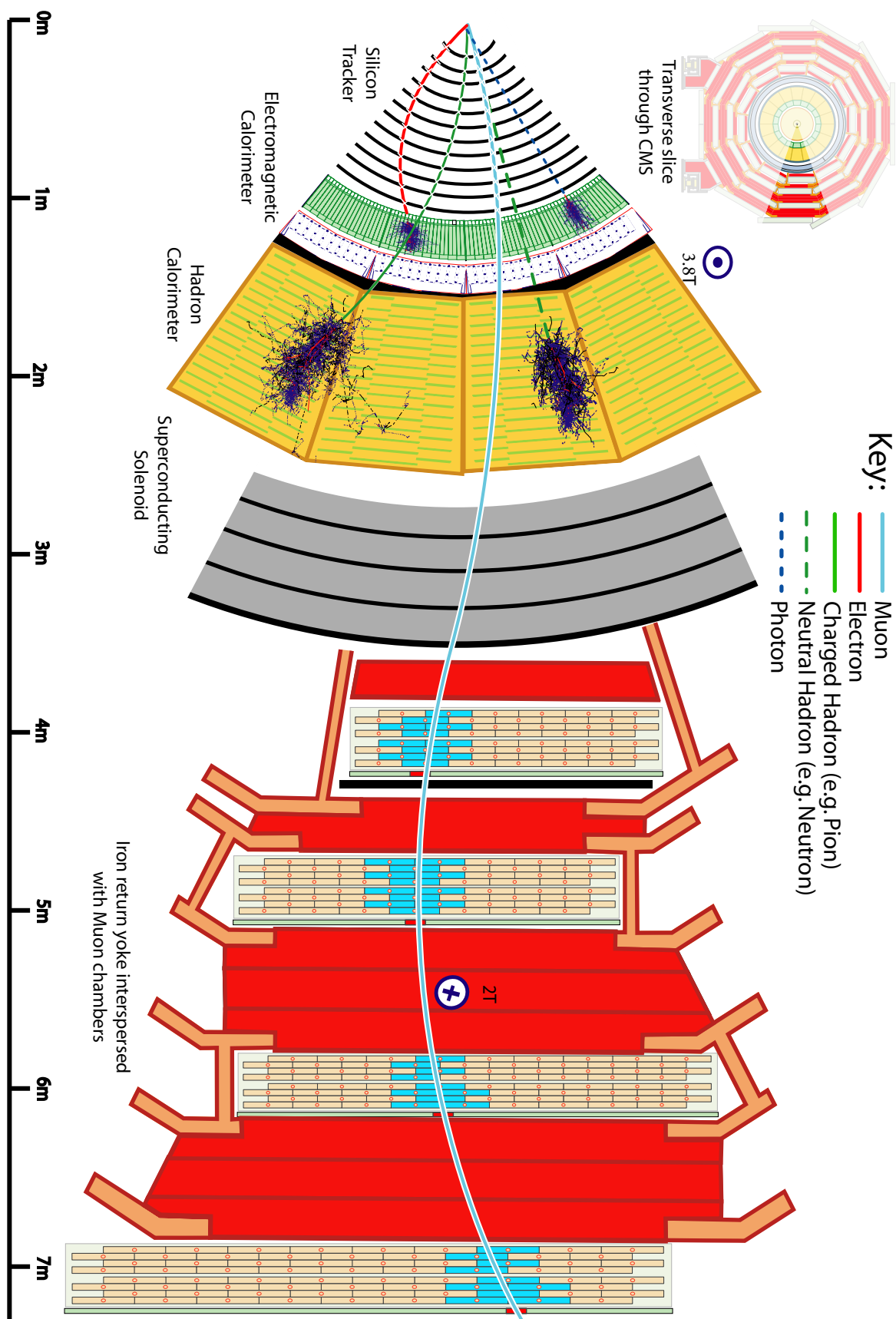
Figure 6.2: Passage of particles through a cross sectional slice of the CMS Detector.

In this thesis, there are several additional cuts applied to muons to be considered "good". For example, the muon track must originate from within 0.2 cm along the z-direction and within 0.045 cm in the transverse direction from the primary vertex [1, 2].

There are several working points used in this thesis: loose, medium, and tight. A loose muon is a basic PF muon, and is required to exist in the tracker. A medium muon is defined as first being identified as a loose muon, with valid fraction of inner tracker hits above 80%. Muons must have a segment compatibility score of at least 0.451 ensuring the muon is reasonably compatible with the tracker muon **OR** pass the following cuts:

- Segment compatibility score of at least 0.303,

- Normalized χ^2 of track less than three,

- Kink-finder χ^2 less than 20 to reduce muons from in-flight decays,

- Compatibility χ^2 between the standalone (muon system-only) muon and the tracker muon s than 12.

Tight muons are PF muons are loose muons that pass all the following cuts:

- Global muon, requires a global track output using both the tracker track and the standalone track to make a global track,

- Global track χ^2 fit less than 10 suppresses hadronic punch through,

- Global track requires at lest one muon chamber hit in the global fit,

- At least two muon stations include muon segments, which brings id further in harmony with the muon trigger logic,

- Loose impact parameter cuts (which we further tightened as mentioned above) of $d_Z < 0.5$ cm and $d_{xy} < 0.2$ cm,

- At least one hit in the pixel detector is required,

- At least five hits in the tracking layers.

The relative isolation, shown in Eq. 6.1 is used to reduce the number of muons from in-flight hadron decays.

$$I = (\sum p_T{}^{\text{charged}} + \max[0, \sum p_T{}^{\text{neutral}} + \sum p_T{}^\gamma - 0.5 \times \sum p_T{}^{\text{PU}}])/p_T{}^l \qquad (6.1)$$

Here, $\sum p_T{}^{\text{charged}}$, $\sum p_T{}^{\text{neutral}}$, and $\sum p_T{}^\gamma$ are, the scalar sums of transverse momenta from charged hadrons associated with the primary vertex, neutral hadrons, and photons, respectively. $\sum p_T{}^{\text{PU}}$ is the sum of transverse momenta of charged hadrons not associated with the primary vertex. We expect the number of charged particles to be roughly double the neutral particles for PU. This 0.5 factor helps in predicting the amount of neutral pile-up in the isolation sum. All sums are taken inside a solid angle cone surrounding the electron, with a ΔR cone of 0.4. The primary goal of requiring isolated leptons is to reduce leptons not originating from the hard process.

6.5   Electrons

Electrons are distinguished by having tracks, and ECAL deposits with minimal or no associated HCAL deposits. The seed for electron identification is a GSF track, that may be linked to a ECAL cluster. Any HCAL clusters linked with the ECAL clusters must contain no more than 10% of the ECAL cluster energy [138, 148]. Once the electron is identified the included PF candidates (tracks and clusters) are removed from use in further reconstruction.

Additionally, for "good" electron identification an MVA technique is applied [148, 149]. There are two different MVA working points available, one with 80% efficiency with higher purity and one with 90% efficiency with lower purity. The working points for this are shown

in Table 6.1. The isolation used is 15% relative isolation, in Eq. 6.1 around the electron in a cone $\Delta R < 0.3$ [1, 2, 138].

Table 6.1: MVA working points for electrons applicable during the 2016 data-taking period

	80% Efficiency	90% Efficiency
barrel ($|\eta| < 0.8$) p$_T$ above 10 GeV	> 0.941	> 0.837
barrel ($|\eta| > 0.8$) p$_T$ above 10 GeV	> 0.899	> 0.715
endcap p$_T$ above 10 GeV	> 0.758	> 0.357

Cut-based electron identification is used in this thesis for rejection of events. The used working point, veto, has an efficiency of about 95% of the reconstructed electrons. Variables included for discrimination are an energy weighted shower shape variable $\sigma_{i\eta i\eta}$, the η and ϕ size of the deposit, the H/E ratio, and a measure of the bremsstrahlung of the electron. Notably in this thesis the cut-based veto WP allows up to two missing inner pixel hits for the GSF track associated with the electron, instead of one missing hit usually in place in other working points.

In this thesis, additional vertex cuts are applied to ensure the electron is near the primary vertex: $d_Z < 0.2$ cm and $d_0 < 0.045$ cm.

6.6   Jets

Jets in this thesis are constructed from charged and neutral hadron groups in PF using the anti-k_t (AK) algorithm [150], with a radius of $\Delta R = 0.4$ (AK4). This algorithm tends to produce more circular-jets in the $\eta - \phi$ plane, than other available algorithms and is well suited to reconstruct jets from quark and gluon decays. A larger radius AK jet, such as AK8, can be used for boosted object identification with jet-substructure techniques. The Cambridge-Aachen (CA) jet clustering, which tends to produce larger and more irregular jets, is available for use as well, but not used in this thesis. CA jets are better-suited

to finding low-p_T jets with substructure [151]. Both AK and CA are implemented in the FastJet 3.0.1 [152] software used at CMS. The performance of the jet reconstruction and calibration for the 2016 collected data set is discussed in Ref. [151]. Over 98% of jets are reconstructed *and* identified as good within CMS, with $p_T > 30$ GeV [151]. After the PF jets are reconstructed, the jet four-vectors are calibrated specifically for jets [138,151]. The estimated contribution from pile-up is first subtracted from the jet. The charged hadrons clustered into the jet which are identified as coming from a vertex *other* than the primary vertex, are removed from the jet. This method, called Charged Hadron Subtraction (CHS), is used on jets in this thesis [153]. Other high efficiency selections (loose) are applied on variables such as the number of particles in the jet to reduce jets from pileup and detector noise [151]. There is an MVA technique applied to identify b-jets, with various available working points [154]. It takes into consideration any secondary vertices if the b or c-hadron decayed after traveling a small distance, of order 50 μm [154]. The significance of the impact parameter of tracks, defined as the impact parameter divided by the uncertainty, is shown in Fig. 6.3 reproduced from [154].

6.7 Missing transverse momenta

In this thesis, Type-I E_T^{miss} [137, 138] is used. The negative sum of all of the transverse momenta of recalibrated jets, and PF objects (charged hadrons, electrons, and muons) is the E_T^{miss}. Type-1 E_T^{miss} indicates that the recalibrated jets discussed above are used in the calculation. The performance of E_T^{miss} is discussed in Ref. [137]. The excellent data/mc agreement of E_T^{miss} as shown in Fig. 5.3 demonstrates the excellent reconstruction of jets and energy resolution of tracks in CMS.
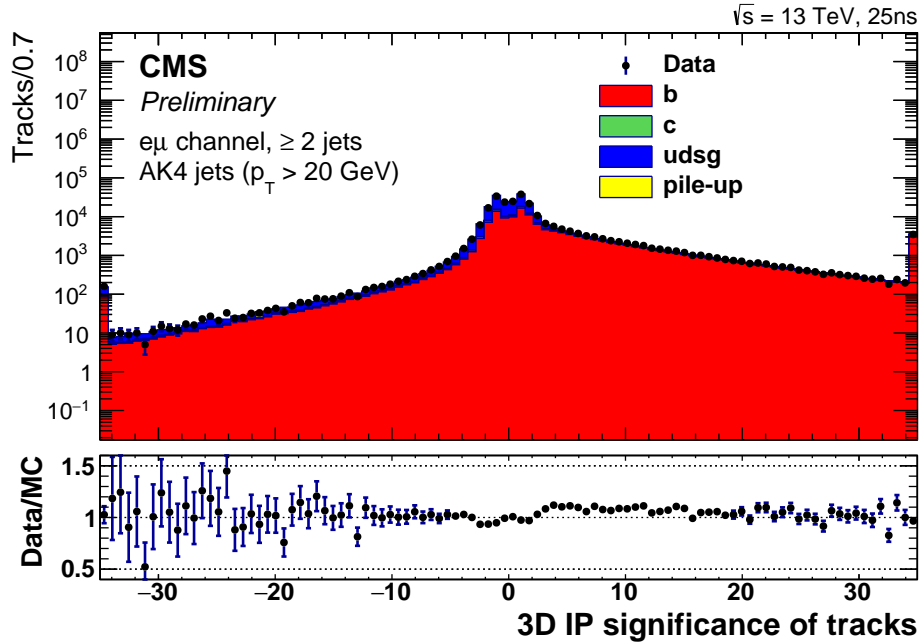
Figure 6.3: 3D impact parameter significance of tracks showing the long positive distribution of heavy flavor (b) tracks.

6.8 Taus

Taus are the only lepton which decays to hadrons, given that the tau mass is 1.78 GeV. When the τ decays via a virtual W to the other leptons (e/μ), the decay is reconstructed according to the previous sections. The electron from a τ_e decay is reconstructed as an electron. However sixty percent of of τ decays are hadronic (τ_{h}). For optimal efficiency it is essential these decays be reconstructed.

The tau reconstruction algorithm used at CMS is the Hadron-Plus-Strips (HPS). Hadron-plus-strips (HPS) [155,156] is the algorithm used at CMS to reconstruct hadronic tau decays, denoted τ_{h}. For an HPS tau, one or three charged hadrons are reconstructed near a narrow ECAL deposit which is shaped as a ϕ-strip. The charged hadrons are produced directly in the τ_{h} decay, while the ϕ-strip comes from the $\pi^0 \rightarrow \gamma\gamma$ decay. Any electrons from photon conversions will bend in the magnetic field. The result is a "strip"-shaped deposit

in ECAL. HPS requires the $p_T(strip) > 2.5$ GeV [156]. There are currently three main modes of reconstructed taus used: one-prong $(h^\pm)$, one-prong one-π^0 $(h^\pm\pi^0)$, and three-prong $(h^\pm h^\pm h^\pm)$. The charge of the τ_h is required to be compatible with a tau: $|q_\tau| = 1$.

First the tau reconstruction candidates are seeded by PF AK4 jets, discussed in Section 6.6 and the highest PF charged hadron from within the jet is chosen as the lead track of the tau candidate. Nearby tracks are added to the tau candidate. If a nearby ϕ-strip is found, it is added to the τ_h.

In Run-I, the HPS ϕ-strips were fixed in size: 0.05×0.2 in the $\eta - \phi$ plane. However the size of the ECAL deposit from π^0 decays is variable. Lower-p_T electrons, from a lower p_T π^0, will bend more in the magnetic field and create a larger strip. Likewise, a higher p_T π^0 may create a more collimated strip shorter in ϕ. With a fixed-size strip, lower momentum τ_h-decay products may contribute to their own isolation cone, and therefore could be self-rejected when isolation requirements are applied. τ_h can also contribute to the isolation calculation, if a charged pion/hadron in the decay undergoes a nuclear interaction with the tracker, and produces secondary particles which are not counted as part of the tau object.

In Run-II of the LHC, at CMS, the HPS reconstruction was updated in order to mitigate the low-p_T τ_h isolation contribution. In this thesis, dynamic strip reconstruction [156] is used. The strip position is computed using a p_T weighted average of the electron and photon contributions. Electrons and photons are added to the strip within an envelope depending on the p_T of the electron or photon being added. The size of the 95% envelope used to determine the size of the nearby strip is demonstrated in Fig. 6.4 [156, 157]. The fitted envelope functions used are

$$f_\eta(p_T) = 0.20 \cdot p_T^{-0.66}$$
$$g_\phi(p_T) = 0.35 \cdot p_T^{-0.71}$$

(6.2)

The mass of the reconstructed τ_h is shown in Fig. 5.3 where the various mass spectra of

(a) η-size   (b) ϕ-size
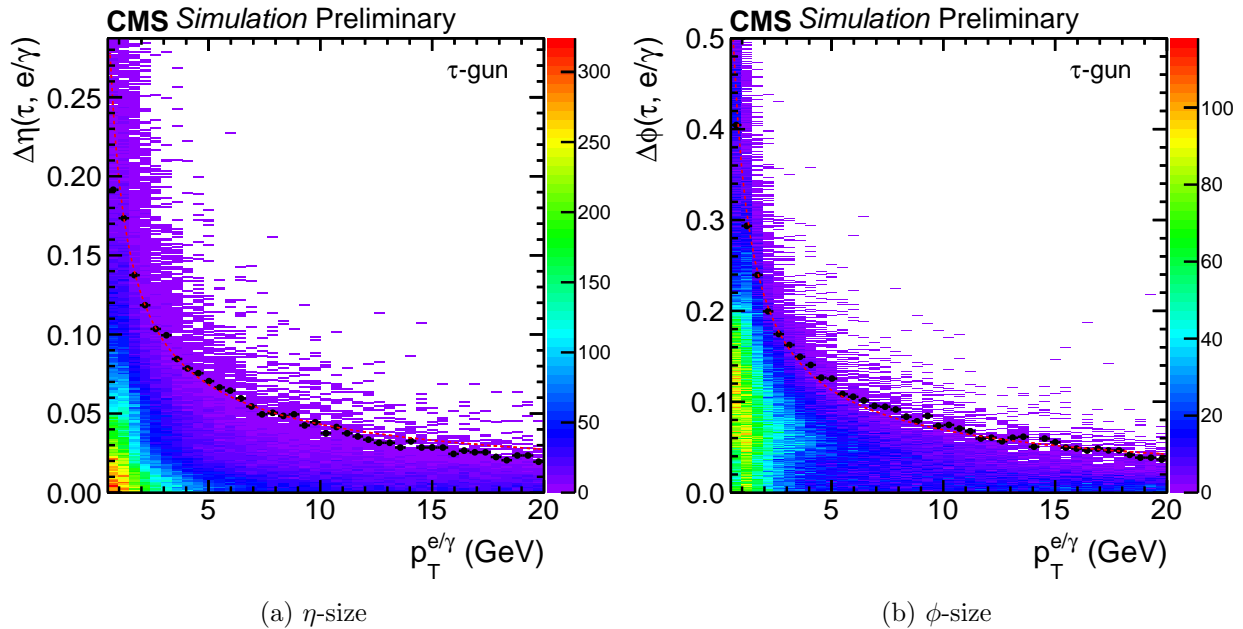
Figure 6.4: Dynamic strip sizing 95% envelope shown in η and ϕ.

the charged hadrons is evident. The spike at 139 MeV is the mass of the charged pion. The other two peaks represent the 770 MeV ρ resonance and the 1260 MeV a_1 resonance.

The largest source of mis-identified τ_h come from quark/gluon jets. An Multivariate Analysis (MVA) discriminator is developed to reduce the jet-τ_h-fakes. The MVA considers the following information as input: charged isolation sum information, neutral isolation sum information, reconstructed τ_h decay mode, transverse impact parameter, and secondary impact parameter and other lifetime information. The trained MVA discriminator primarily rejects jet-τ_h-fakes. There are various working points provided, with different purities and efficiencies, shown in Fig. 6.5.

In addition to reducing jet-fake-τ_h, ℓ-tau-fakes must also be reduced. Anti-lepton discriminators are developed for rejecting electrons and muons that fake τ_h. One way to reject a muon-τ_h-fake is to check if a PF muon exists within a certain distance of the tau ($dR < 0.4$), if so the tau should be rejected as a likely muon-τ_h-fake. The loose working point rejects

(a) τ_h Efficiency

(b) τ_h Mis-identification
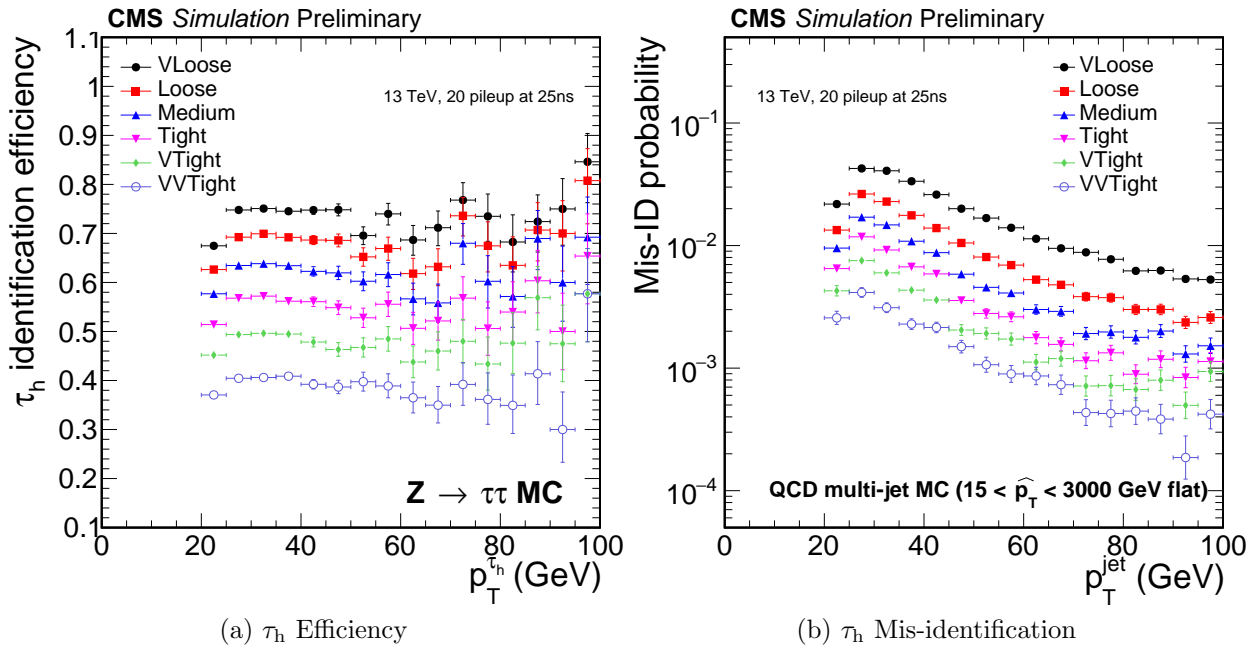
Figure 6.5: MVA tau identification efficiencies and mis-identification probabilities and mis-identification probabilities fr available working points.

a tau candidate if there are at least two hits in muon stations within the isolation cone, or if the energy recorded in the calorimeters is less than twenty percent of the momentum of the leading track. The tight working point will reject any τ_h if any muon-station hit is recorded in the isolation cone. Both anti-muon τ_h discriminants are over 99% efficient for $Z \rightarrow \tau\tau$, while reducing the mis-identification rate to 1.77×10^{-3} for the loose working point and 7.74×10^{-4} for the tight working point. To reject electron-τ_h-fakes, an MVA is deployed with variables such as the number of photons in the isolation cone, relative HCAL recorded-energies, and mass of the τ_h. The performance of the anti-electron MVA is shown in Fig. 6.6.

(a) τ_h efficiency

(b) electron-τ_h-fake rate
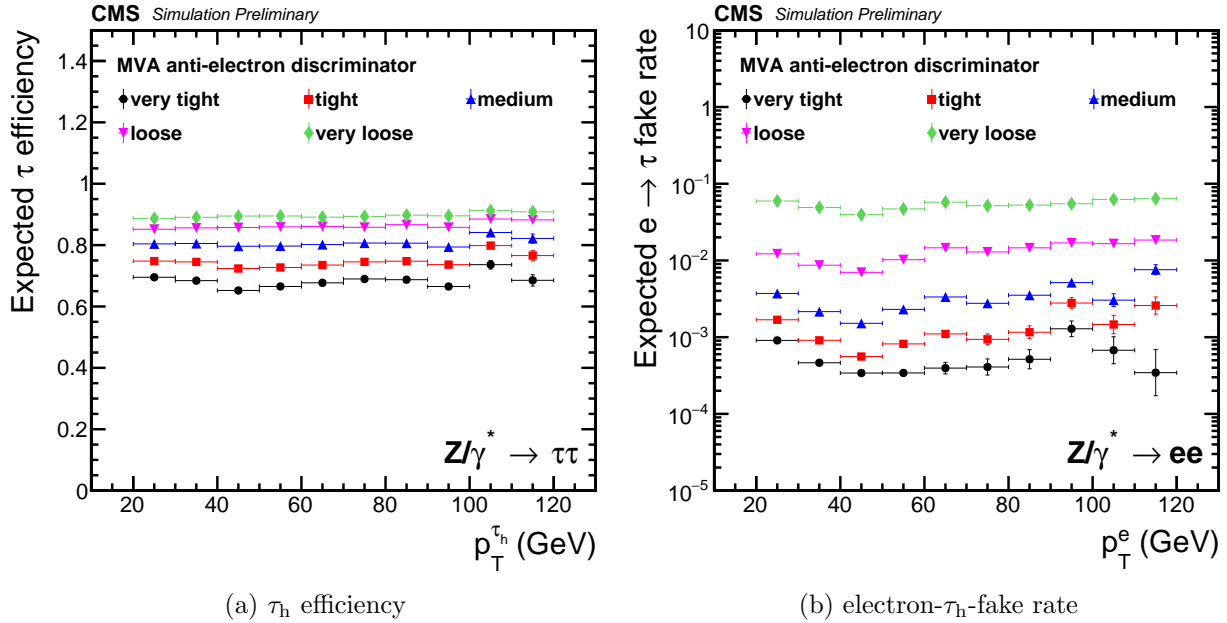
Figure 6.6: MVA anti-electron tau identification efficiencies and mis-identification probabilities for available working points.

6.9   Primary vertex

The vertex where the physics of interest is coming from is called the primary vertex. It is identified as the vertex with the largest value of $\sum(p_T^2)$, where the sum is taken over the transverse momentum (p_T) of clustered objects and E_T^{miss}. The E_T^{miss} is only included if it is greater than $2 \times \sqrt{\sum E_T}$. There may be additional requirements on the quality of the primary vertices, such as the number of degrees of freedom for the vertex should be greater than four and the $|z| < 24$ cm. Choosing a quality primary vertex is especially important when selecting events with large E_T^{miss}.

Chapter 7

Tau pair selection and event weights

Both analysis performed for this thesis involve tau pair searches. In this chapter I discuss tau pair selection, applicable event weights, and general systematics related to the selection of tau pairs.

7.1 Channels and final states

Tau particles can decay leptonically to muons (μ) or electrons (e) in addition to two neutrinos(ν), or hadronically (τ_{h}) with one additional neutrino. Figure 7.1 shows several possible decays of the τ lepton.

The various possible final states for a τ-pair decay are: ee$+4\nu$, $\mu\mu+4\nu$, e$\mu+4\nu$, e$\tau_{\mathrm{h}}+3\nu$, $\mu\tau_{\mathrm{h}}+3\nu$, and $\tau_{\mathrm{h}}\tau_{\mathrm{h}}+2\nu$. Within each analysis, the final states are separated by "channel." For example, the e$\mu+4\nu$ final state corresponds to the eμ channel. The channels considered for a specific analysis depend on the physical process under consideration. For example, the $\tau\tau \rightarrow$ ee (ee) channel is insensitive to 125 GeV Higgs decays so it is not included in Chapter 8. The $H \rightarrow \tau\tau \rightarrow$ ee $+ 4\nu$ mass peak aligns very closely with the Z boson peak, albeit with a lower resolution compared to $H \rightarrow \tau\tau \rightarrow \mu\mu + 4\nu$. The $\mu\mu$ final state is also
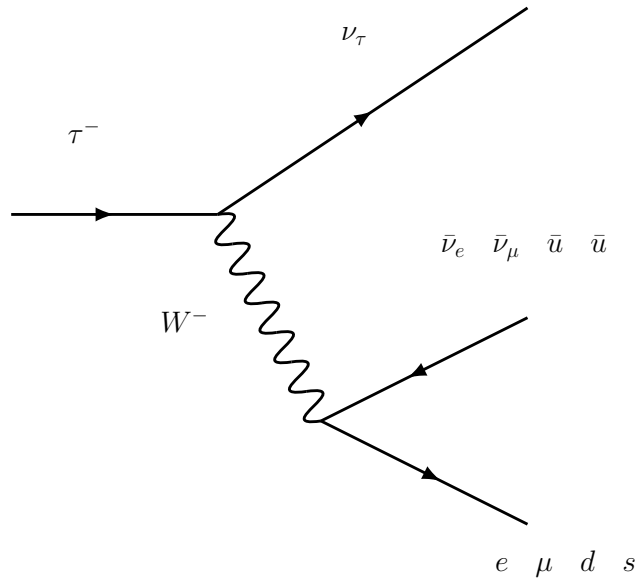
Figure 7.1: τ lepton decay Feynman diagram

excluded due to low sensitivity. The eμ channel is more promising, relative to ee and $\mu\mu$, for a 125 Higgs search, since no resonant behavior from the Z boson is found in this final state. The ee and $\mu\mu$ final sates are excluded from Chapter 8 and 9. The eμ channel is excluded from Chapter 9 as well, since it has a low branching ratio compared with the other three channels considered.

Overall the τ_{h} will retain a larger percentage of the momentum from the original tau particle, while in the leptonic decays the neutrinos have considerably larger share of momenta. Table 7.1 shows the channels considered in Chapter 8 and Chapter 9.

7.2   Event selection

An event is considered if a tau pair is found, with the objects selected passing requirements discussed in Chapter 6 such as

- well-identified isolated ℓ (muon or electron) or hadronic tau, τ_{h}

Table 7.1: Included τ-pair channels

Channel	Chapter 8	Chapter 9	Fraction
Channel	Standard Model $H \to \tau\tau$	$H \to \tau\tau + E_{\mathrm{T}}^{\mathrm{miss}}$	of $\tau\tau$
$\tau_{\mathrm{h}}\tau_{\mathrm{h}}$	Y	Y	42%
$\mu\tau_{\mathrm{h}}$	Y	Y	23%
$e\tau_{\mathrm{h}}$	Y	Y	23%
$e\mu$	Y	N	6%
ee	N	N	3%
$\mu\mu$	N	N	3%

- well-identified isolated hadronic tau, τ_{h}

- no other electrons or muons found in event

Further selections are chosen depending on analysis, and online trigger selection criteria.

7.2.1 Available triggers

Kinematic selections for tau pairs is largely determined by the online trigger threshold used.

The $\ell\tau_{\mathrm{h}}$ channels may trigger off both the ℓ and the τ_{h}, or only the ℓ. Without any trigger constraints, the minimum $p_{\mathrm{T}}(\tau_{\mathrm{h}})$ requirement is $p_{\mathrm{T}}(\tau_{\mathrm{h}}) > 20$ GeV. In the $\ell\tau_{\mathrm{h}}$ channels, the minimum $p_{\mathrm{T}}(\ell)$ will always be determined by the online trigger threshold. Table 7.2 shows the available triggers for all final states.

The Level-1 available triggers in the 2016 data-taking run were Level-1 muon $p_{\mathrm{T}}(18)$, electron $p_{\mathrm{T}}(21)$, and τ_{h} $p_{\mathrm{T}}(32)$. However in order to conserve rate at output, we do not save every event with those requirements. The number of events passing these selections is capped at 100kHz, from here it is passed to the HLT where there are additional cuts. The cuts at L1 affect the cuts at HLT, and therefore affect the offline analysis cuts. For example the di-τ_{h} trigger has an online Level-1 cut at 32 GeV and an online HLT p_{T} cut at 35 GeV. Therefore the analysis selection will have a minimum p_{T} requirement for the $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ channel of $p_{\mathrm{T}}(40)$ GeV in order to be significantly above the turn-on region. In the turn-on

region, the trigger efficiency is increasing towards the plateau region, where trigger efficiency is constant.

Channel	HLT Trigger	Runs available
$\mu\tau_h$	Isolated HLT Muon $p_T > 24$ GeV	B,C,D,E,F,G,H
$\mu\tau_h$	Isolated HLT Tracker Muon $p_T > 24$ GeV	B,C,D,E,F,G,H
$\mu\tau_h$	Isolated HLT Muon $p_T > 22$ GeV	B,C,D,E,F
$\mu\tau_h$	Isolated HLT Tracker Muon $p_T > 22$ GeV	B,C,D,E,F
$\mu\tau_h$	η-restricted Isolated HLT Muon $p_T > 22$ GeV	C,D,E,F,G,H
$\mu\tau_h$	η-restricted Isolated HLT Tracker Muon $p_T > 22$ GeV	C,D,E,F,G,H
$\mu\tau_h$	η-restricted Isolated HLT Muon $p_T > 19$ GeV and Loose Isolated $\tau_h > 20$	B,C,D,E,F,G,H
	Only one L1 seed required	
$\mu\tau_h$	η-restricted Isolated HLT Muon $p_T > 19$ GeV and Loose Isolated $\tau_h > 20$	B,C,D,E,F,G,H
	Muon L1 Seed and Tau L1 Seed Required	
$e\tau_h$	Isolated HLT Electron $p_T > 25$ GeV and $|\eta| < 2.1$	B,C,D,E,F,G,H
$\tau_h\tau_h$	Double Isolated HLT Tau $p_T > 35$ GeV and $|\eta| < 2.1$	B,C,D,E,F,G,H
	Type of isolated changed halfway through year, efficiencies are lumi-averaged.	

Table 7.2: List of relevant HLT triggers for each final state and the available runs.

7.2.2 Third lepton veto

A common cut used in τ-pair analyses is an extra-lepton veto. An event is rejected if an extra, not already considered in the tau pair, "loose" muon or electron is found in the event. Vetoing events with extra muons and electrons ensures that there is no overlap between the different channels and also reduces Z+jets contribution when the events contain misidentified jet-τ_h-fakes.

For example, if in a $\mu\tau$ event in addition to the τ_μ candidate there is another opposite-sign μ, then the event is vetoed as it is likely that the reconstructed hadronic tau was faked by a jet. The vetoing electron "loose" criteria is defined: $p_{T_e} > 10$ GeV, $|\eta_e| < 2.5$, electron MVA-based 90% efficiency working point and passes conversion veto. The electron track has ≤ 1 missing inner track hit, and the isolation (iso$_e < 0.3$). The vetoing muon "loose"

criteria is as follows: $p_{T\mu} > 10\text{GeV}$, $|\eta_\mu| < 2.4$, and μ is loosely isolated ($\text{iso}_\mu < 0.3$) while passing the medium ID, discussed in Chapter 6.

7.3 Tau pair backgrounds

Any process that can produce two isolated leptons, or one isolated lepton with additional jets, or multijet processes are considered as backgrounds, since jets can fake τ_h at a relatively high rate. The cross sections of considered background processes are listed in Table 7.3. The following sections discuss the major contributing background processes.

7.3.1 Backgrounds with τ

The main resonant and irreducible background is the Drell-Yan process. The Feynman diagram for the Drell-Yan process is shown in Fig. 7.2. The E_T^miss from this process originates from the τ-decay. The transverse mass of one of the tau candidates with the E_T^miss, written as $M_{T,(\tau,E_\text{T}^\text{miss})}$ and defined as

$$M_T = \sqrt{2p_\text{T}^\tau E_\text{T}^\text{miss}\left(1 - cos(\phi)\right)} \tag{7.1}$$

where p_T^τ and E_T^miss are transverse momentum of one τ lepton candidate and missing transverse energy. This observable, $M_{T,(\tau,E_\text{T}^\text{miss})}$, is close to 0 GeV and usually less than 50 GeV. The cross section used for the normalization of this process is about 5746 pb.

The $t\bar{t}$ process also contributes tau-pairs, but the tau pair mass is non-resonant. Figure 7.3 shows the leading production Feynman diagrams for the $t\bar{t}$ process. The transverse mass of one of the tau candidates with the E_T^miss ($M_{T,\tau,E_\text{T}^\text{miss}}$) is usually larger than 50 GeV, unlike the Drell-Yan process. Top quark decays include b-quarks which are reconstructed as jets, as discussed in Section 6.6. The cross section used to normalize the simulation for this process is about 831 pb.

(a) Z boson decays to 2 leptons

(b) Z boson decays to 2 leptons plus jet
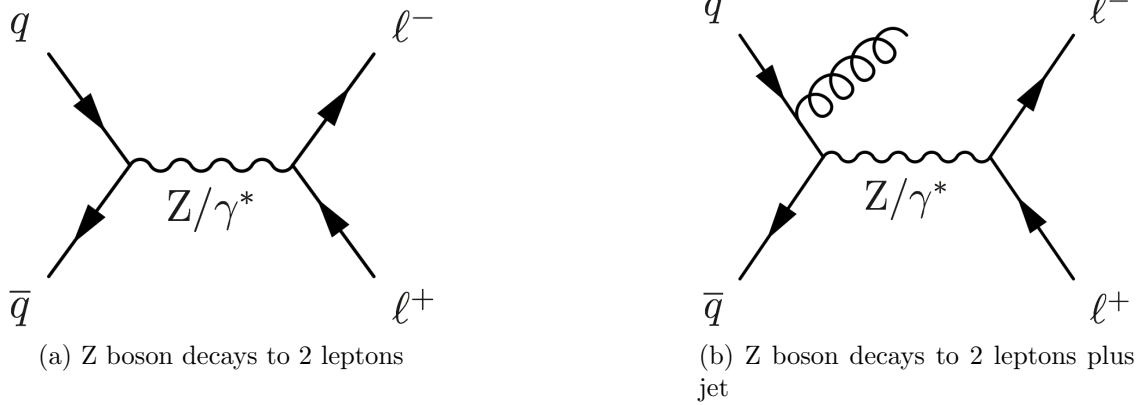
Figure 7.2: Drell-Yan process Feynman diagram in 7.2a, and Drell-Yan process with additional gluon in 7.2b.

(a) leading gg contribution

(b) sub-leading gg contribution
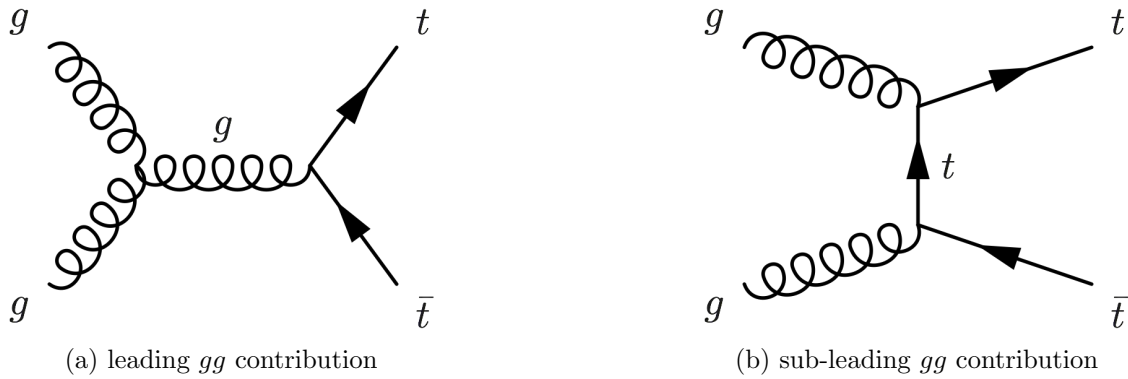
Figure 7.3: gg-fusion $t\bar{t}$ production Feynman diagrams

The last major contribution of real tau pairs comes from various diboson and tri-boson processes. Tri-boson processes have very small cross sections; these processes can contribute when the V decays hadronically. The cross sections for the various multi-boson processes are shown in Table 7.3.

7.3.2 Backgrounds with fake-τ_h

The fake-τ_h process with the largest cross section of about a mb, is QCD/multijet. However the tau MVA identification greatly reduces the contribution from this process as shown in

Background Process	Cross section (pb)
Z+jets Inclusive Jet Production	5746.64
$t\bar{t}$	831.76
QCD (multijet)	Data-Driven
W+jets Inclusive Jet Production	60959.8
$WZ \to 1\ell 3\nu$	3.05
$WZ \to 1\ell 1\nu 2Q$	10.71
$WZ \to 2\ell 2Q$	5.595
Single $\bar{t}$ + W	35.6
Single t + W	35.6
Single t	80.95
Single $\bar{t}$	136.02
$WW \to 1\ell 1\nu 2Q$	1.212
$ZZ \to 2\ell 2Q$	3.22
$WW \to 2\ell 2\nu$	12.178
$ZZ \to 2\ell 2\nu$	0.564 pb
EWK $WZ \to 3\ell\nu$	4.708
$WZ \to 1\ell 1\nu 2Q$	10.71
$WZ \to 1\ell 3\nu$	3.05
$WZ \to 2\ell 2Q$	5.595
$ZZ \to 4\ell$	1.212
EWK $W^{-} + 2Jets$	20.25
EWK $W^{+} + 2Jets$	25.62
EWK $Z \to \ell\ell + 2Jets$	3.987
EWK $Z \to \nu\nu + 2Jets$	10.01
$gg\ H \to WW \to 2\ell 2\nu$	1.001
VBF $H \to WW \to 2\ell 2\nu$	0.0858

Table 7.3: NLO cross sections for considered backgrounds. In this table, ℓ represents all three generations of charged leptons (e,μ,τ.)

Fig. 6.5. In the $\ell\tau_{\text{h}}$ channel, the transverse mass of the ℓ and the $E_{\text{T}}^{\text{miss}}$, $M_{T,(\ell,E_{\text{T}}^{\text{miss}})}$ will peak near 0 GeV with a long tail ending around 90 GeV.

The second largest background with expected mis-reconstructed (fake) τ_{h} is the W+jets process. The W+jets process cross section is about 60959 pb, which is larger than Z+jets by a factor of ten. Feynman diagrams for this process are shown in Fig. 7.4. In the $\ell\tau_{\text{h}}$ channel, the transverse mass of the ℓ and the $E_{\text{T}}^{\text{miss}}$, $M_{T,(\ell,E_{\text{T}}^{\text{miss}})}$ will peak near the W boson mass, 80 GeV.

(a) leading contribution   (b) sub-leading contribution
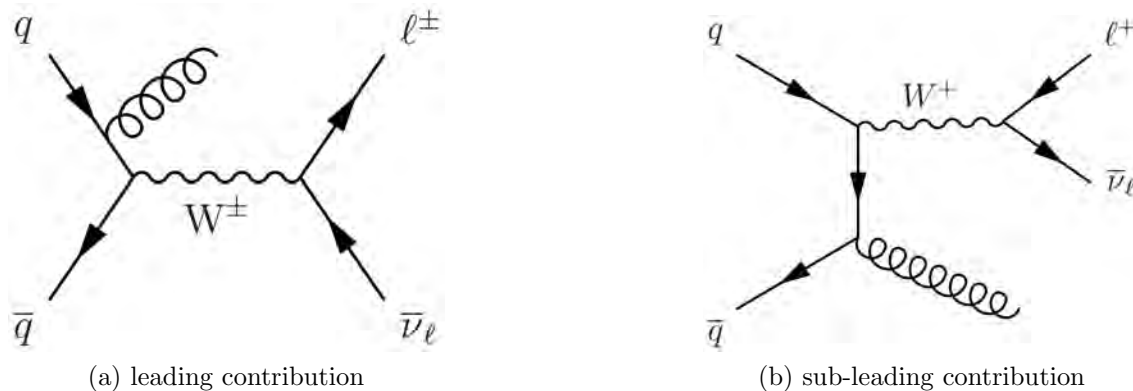
Figure 7.4: W+1 jet production Feynman diagrams

7.4 Tau pair background methods

7.4.1 Multijet/QCD background estimation

The QCD/multijet background can be estimated using the ABCD method indicated in Table 7.4. Our signal region is region A, with *isolated* and *opposite-sign* tau pair. We get the opposite-sign isolated contribution (A) from the same-sign isolated (C) region, which is corrected by the ratio of B/D. We assume the equality $A \times D = B \times C$ within an uncertainty, which facilitates describing the A-contribution as $A = C \times \frac{B}{D}$. The $\frac{B}{D}$ ratio is dependent on basic identification selections, and in a more selective final state, this ratio of $\frac{B}{D}$ is recalculated.

	Isolated τ-cand.	Anti-Isolated τ-cand
Opposite-sign	A	B
Same-sign	C	D

Table 7.4: ABCD Method: Region A is our signal region.

To summarize, the isolated opposite-sign QCD contribution is obtained from the isolated same-sign region multiplied by an $SS \rightarrow OS$ factor ($\frac{B}{D}$) measured in the anti-isolated τ-candidate region.

7.5   Event weights

Various weights are applied on an event-by-event basis to simulation to improve the data-simulation agreement. Every simulated event is given a weight. Weights are taken into account when calculating expected event contributions by giving some simulated events more impact, that is a higher weight, than other events. The derivation and application of these weights is described in the following sections. The tau selection used in the following section for the plots is generally low pt ($p_T^{\tau_h} > 20$ GeV and $p_T^{\mu} > 23$ GeV), and low transverse mass($M_{T,(\tau_h, E_T^{miss})} < 50$ GeV, and third lepton vetos applied in order to select events compatible with a Z boson or a Higgs boson.

7.5.1   Pileup reweighting

Pileup (PU) reweighting is applied to the simulation in order to have similar PU distributions between data and simulation. An estimated "true" pile–up, the number of pp interactions per bunch crossing, distribution is produced. The minimum-bias cross section, 69.2 mb as measured by CMS, is used to re-weight the number of interactions per bunch crossing. This measured cross section is dependent on beam energy and other beam conditions.

The number of reconstructed primary vertices is correlated with the number of pp interactions. The data/simulation distribution for the number of primary vertices in the $\mu\tau$ channel before and after this reweighting are shown in Fig. 7.5. The agreement is improved after PU reweighting is applied.

7.5.2   Tau identification efficiency scale factors

In order to correct for differences observed in hadronic tau identification between data and simulation, scale factors are calculated and applied. The measurements are derived using a tag-and-probe method in $Z \to \mu\tau_h$ events. The tag-and-probe method targets Z boson

(a) Before PU reweighting (b) After PU reweighting
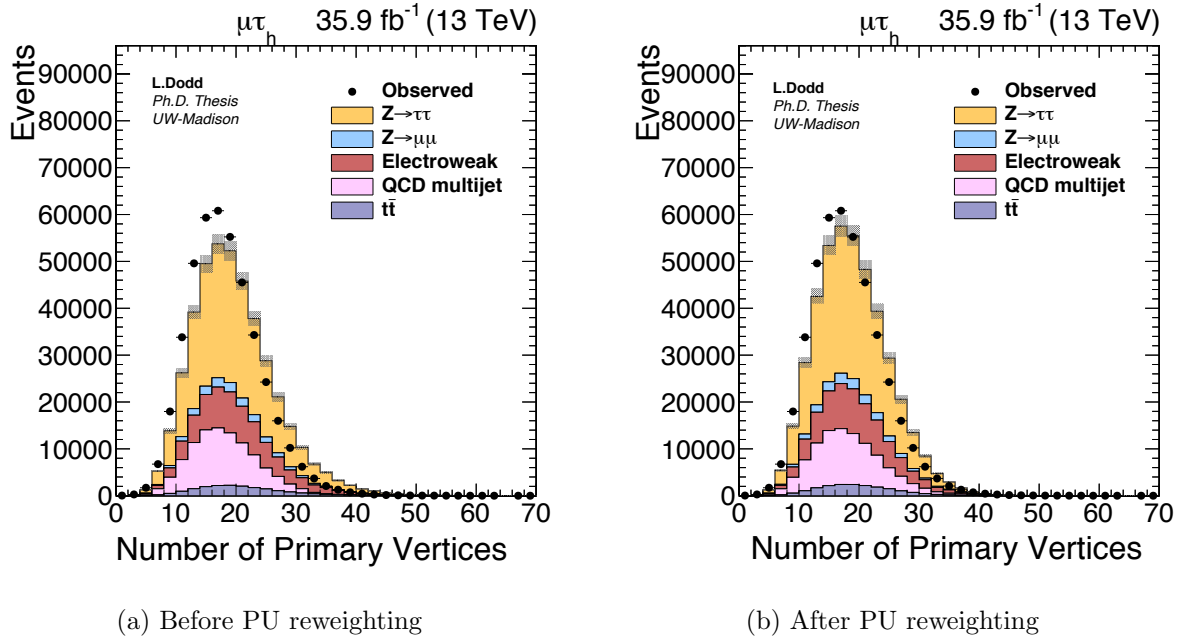
Figure 7.5: Primary vertex distribution before and after the pile–up reweighting weight application as described in the text for the $\mu\tau_h$ final state, using the measured minimum-bias cross section 69.2 mb.

decays. The method operates in the principle that if two isolated opposite-sign particles are found with an invariant mass compatible with the Z boson mass, it is highly likely both particles came from the Z boson.

First, an isolated well-reconstructed lepton is "tagged", if there is an opposite-sign particle compatible with a Z decay the identification efficiencies of the second reconstructed particle can be "probed." For τ pairs, the $\mu\tau_h$ final state is frequently chosen. The μ is required to be the isolated, reconstructed "tag" lepton, and the loosely identified τ_h is used as the "probe." The loose τ_h identification requires that the tau reconstruction succeeds and the tau has at least one track with p_T greater than 5 GeV. Events are divided into two categories: the loosely defined tau candidate (the probe) either passes or does not pass (fails) the working point of the discriminator under study. This forms the two categories: "pass" and "fail." The $Z \to \mu\tau_h$ contribution to each region is adjusted with a simultane-

ous maximum likelihood fit that considers the tau identification efficiency as the parameter of interest. The pass and fail category contributions are anti-correlated. Any change in the identification parameter in the passing region negatively affects the contribution in the failing region and vice versa.

A $Z \to \mu\mu$ control region is added to the $Z \to \mu\tau_h$ tag-and-probe maximum likelihood fit to further control the expected event count from the Drell-Yan process, the μ identification, and the luminosity measurement [156–158]. The fit is done on two variables, the invariant visible mass of the tau pair (m_{vis}) and the number of tracks within a cone of the reconstructed tau candidate. The results for the hadronic tau reconstruction found via the tag-and-probe method in the $\mu\tau_h$ channel is extrapolated to generator-matched τ_h.

Figure 7.6 shows the pass and fail categories, described above, after the simultaneous maximum likelihood fit from this tag-and-probe measurement of the MVA tight τ identification efficiency [156–158].

The tight MVA τ_h discriminator working point, which has an efficiency around 50%, has a scale factor of 0.95 ± 0.05 applied to real tau backgrounds to account for differences in efficiencies between simulation and data. The scale factors calculated for each working point are shown in Table 7.5.

Table 7.5: General hadronic tau MVA identification corrections

MVA Working Point	Scale Factor Correction
VLoose	0.99 ± 0.05
Loose	0.99 ± 0.05
Medium	0.97 ± 0.05
Tight	0.95 ± 0.05
VTight	0.93 ± 0.05

If there are two hadronic taus in the final state, for example in $Z \to \tau_h\tau_h$ events, the weights for each τ_h are multiplied together. There is currently no recommendation to apply τ_h efficiency as a function of p_T or η of the τ_h candidate. Figure 7.7 shows the agreement in
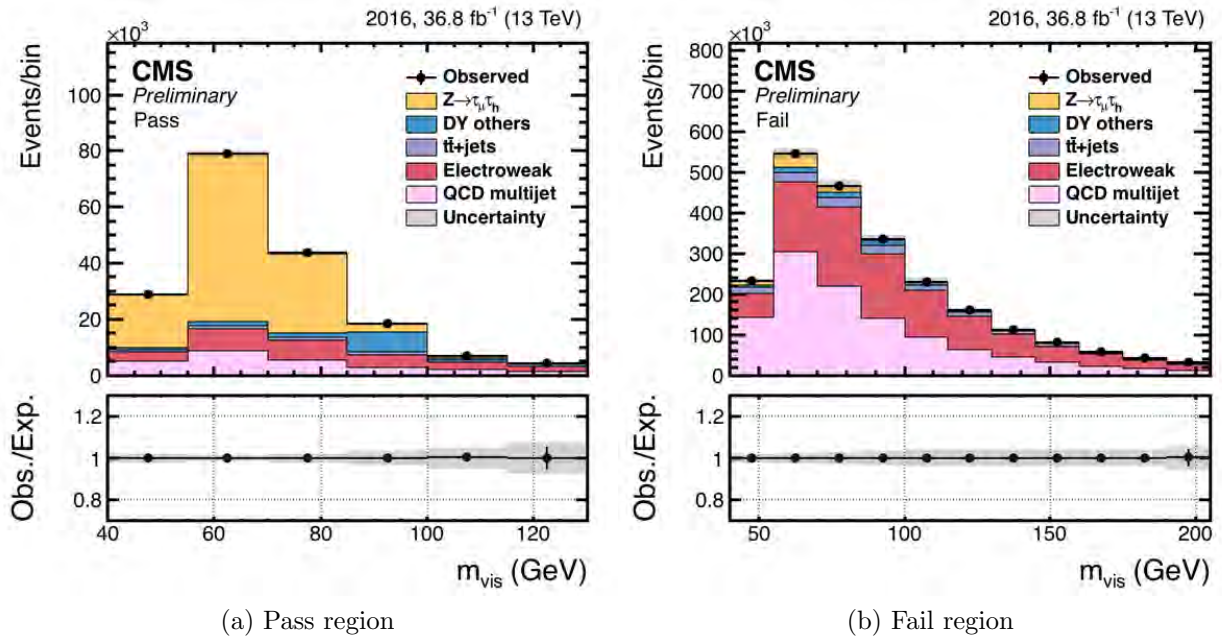
Figure 7.6: Tag-and-probe measurement of the tau identification efficiency scale factor in $Z \to \mu\tau_\mathrm{h}$ events, for the MVA tight isolation working point, using the visible mass between the muon and the tau candidate as an observable. The distributions shown are the visible mass of the μ and τ_h postfit in the "pass" region, (7.6a), and the "fail" region, (7.6b). All τ_h decay modes are included; p_T is required to be greater than 20 GeV and $|\eta| < 2.3$.

the p_T and η distributions.

7.5.3 Anti-lepton discriminator tau-ID scale factors

The anti-lepton discriminator scale factors are applied to the simulated backgrounds where a reconstructed tau is matched to a generator-level prompt(not from a τ decay) lepton. The scale factor for the anti-muon discriminator is applied to background simulation where a μ fakes a τ_h, despite the additional application of the anti-electron discriminator. Similarly the scale factor for the anti-electron discriminator is applied in the case where an e fakes a τ_h. The scale factors are found with a tag-and-probe method in Z+jets events.

In the "fail" region, i.e. tau candidates fail the anti-lepton discriminator under study, the major contribution comes from $Z \to \ell\ell$ events, which form a narrow peak around the

(a) Inclusive η distribution and low $\mu + E_\mathrm{T}^\mathrm{miss}$ trans-verse mass $(M_{T,(\mu,E_\mathrm{T}^\mathrm{miss})} < 50$ GeV$)$

(b) Inclusive p_T distribution and low $\mu + E_\mathrm{T}^\mathrm{miss}$ trans-verse mass $(M_{T,(\mu,E_\mathrm{T}^\mathrm{miss})} < 50$ GeV$)$
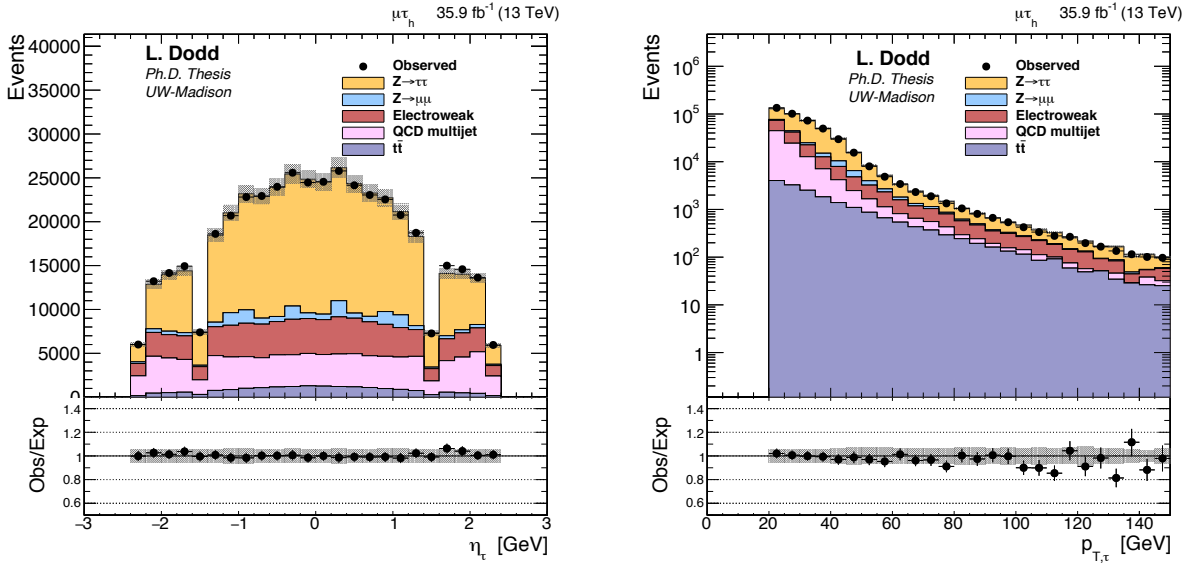
Figure 7.7: η and p_T distribution of the τ_h for taus selected for the analysis discussed in Chapter 8, with the tau p_T cut reduced to 20 GeV. The inclusive distributions are flat in η^{τ_h} and $p_\mathrm{T}{}^{\tau_\mathrm{h}}$

Z boson mass. The scale factors used in the analysis depend on the pseudorapidity of the tau candidate, and are indicated in Table 7.6. The uncertainties typically range between 5% and 30%, and are related to the more statistically limited populations of the "pass" region.

7.5.4 Lepton identification, isolation and trigger scale factors

Lepton identification and isolation efficiencies are measured in data and simulation using a tag-and-probe technique as described in Section 7.5.2. The efficiency ratio $\epsilon_{data}/\epsilon_{sim.}$ is applied as a weight on an event-by-event basis to correct the simulation. Efficiencies are measured in $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ events, in bins of lepton p_T and $|\eta|$. The kinematic selections on the "tag" and "probe" mirror the analysis selections. These measured values are used for the e and μ objects in the $e\tau_\mathrm{h}/\mu\tau_\mathrm{h}$ channels.

Table 7.6: Efficiency scale factors for the discriminators used to reject prompt leptons in the tau identification process.

Tau discriminator	η range	Scale factor
anti-Electron MVA discriminator VLoose WP	$\|\eta\| < 1.460$	1.213 ± 0.07
anti-Electron MVA discriminator Loose WP	$\|\eta\| < 1.460$	1.320 ± 0.04
anti-Electron MVA discriminator Medium WP	$\|\eta\| < 1.460$	1.323 ± 0.07
anti-Electron MVA discriminator Tight WP	$\|\eta\| < 1.460$	1.402 ± 0.11
anti-Electron MVA discriminator VTight WP	$\|\eta\| < 1.460$	1.207 ± 0.18
anti-Electron MVA discriminator VLoose WP	$\|\eta\| > 1.558$	1.375 ± 0.05
anti-Electron MVA discriminator Loose WP	$\|\eta\| > 1.558$	1.380 ± 0.04
anti-Electron MVA discriminator Medium WP	$\|\eta\| > 1.558$	1.527 ± 0.13
anti-Electron MVA discriminator Tight WP	$\|\eta\| > 1.558$	1.900 ± 0.30
anti-Electron MVA discriminator VTight WP	$\|\eta\| > 1.558$	1.968 ± 0.46
anti-Muon discriminator Loose WP	$\|\eta\| < 0.4$	1.010 ± 0.02
anti-Muon discriminator Tight WP	$\|\eta\| < 0.4$	1.263 ± 0.07
anti-Muon discriminator Loose WP	$0.4 < \|\eta\| < 0.8$	1.007 ± 0.03
anti-Muon discriminator Tight WP	$0.4 < \|\eta\| < 0.8$	1.364 ± 0.28
anti-Muon discriminator Loose WP	$0.8 < \|\eta\| < 1.2$	0.870 ± 0.03
anti-Muon discriminator Tight WP	$0.8 < \|\eta\| < 1.2$	0.854 ± 0.04
anti-Muon discriminator Loose WP	$1.2 < \|\eta\| < 1.7$	1.154 ± 0.13
anti-Muon discriminator Tight WP	$1.2 < \|\eta\| < 1.7$	1.712 ± 0.5
anti-Muon discriminator Loose WP	$1.7 < \|\eta\| < 2.3$	2.281 ± 0.26
anti-Muon discriminator Tight WP	$1.7 < \|\eta\| < 2.3$	2.324 ± 0.5

Efficiencies are found whether the "probe" lepton passes or fails, since the probe lepton is from a Z boson; it is defined as

$$eff = \frac{\text{num. passing probes}}{\text{total num. of probes}}. \tag{7.2}$$

The passing probes are found from fitting di-lepton invariant mass in the window around the Z boson mass, $75 < m_{ll} < 105$ GeV using an decreasing exponential function to model the background and two asymmetric gaussians to model the signal. The scale factor applied to simulated events is given by the ratio $\epsilon_{data}/\epsilon_{sim.}$ for the (p_T, η) bin of the selected lepton.

Trigger efficiencies for the single muon were measured in data using a tag-and-probe

technique, where both the tagged *and* probed lepton pass the selection required at analysis level. The $\tau_h \tau_h$ trigger scale factors provided centrally by a working group are lumi-weighted to account for different online triggers during 2016 data-taking; the Run BCDEFG and Run H data efficiency curves are luminosity weighted to provide an average turn on curve for all 2016 data. The $\tau_h \tau_h$ triggers are calculated and applied on per leg basis and are dependent on $p_T{}^{\tau_h}$, and τ_h decay mode.

7.5.5 Reweighting of LO MadGraph5_aMC@NLO Drell-Yan samples

To constrain the contribution of the $Z \to \tau\tau, \ell\ell$ events, a dedicated control sample of $Z \to \mu\mu$ events is collected in data with a single muon trigger, and compared to simulation. It is built by requiring two well identified and isolated opposite-sign muons with p_T greater than 25 GeV, with an invariant mass between 70 and 110 GeV. Events with additional isolated muons or electrons are rejected. The purity of this sample exceeds 99% and is used to correct simulations in order to better reproduce the Z+jets observed kinematical distributions. The derived weights are then corrected to prevent introducing a general contribution variation of the Drell-Yan background; these only have a shape effect. The $\mu\mu$ distributions before and after reweighting are shown in Fig. 7.9 and Fig. 7.8, whereas the weights are illustrated in Table 7.7.

An additional reweighting based on $p_T(\ell\ell/\tau\tau)$, and m_{jj}, can be applied to reproduce specific kinematic distributions of observed events. Since leading-order simulation is used, some kinematics distributions vary from data. The $\mu\mu$ control region is used to correct the distributions. The corrections obtained are applied to the generator-level quantities in $\tau\tau$ decays. The corrections are shown in Table 7.8.

Table 7.7: LO MADGRAPH5_AMC@NLO reweighting scale factors

m_Z range (GeV) Z p_T range (GeV)	0-60	60-120	120-160	160-200	200-240	more than 240
			Correction per Z p_T and m_Z			
> 300	1.56	1.03	1.06	1.09	0.85	0.92
220-300	0.78	1.13	1.35	1.28	1.45	1.47
180-220	1.47	1.18	1.17	1.43	1.21	1.41
140-180	1.21	1.17	1.28	1.42	1.30	1.44
120-140	1.29	1.17	1.27	1.21	1.17	1.28
100-120	1.14	1.17	1.22	1.35	1.41	1.61
80-100	1.09	1.16	1.21	1.09	1.42	1.48
60-80	1.09	1.12	1.16	1.21	0.97	1.26
40-60	1.12	1.05	1.13	1.15	1.21	1.47
20-40	1.04	0.99	1.07	1.17	1.22	1.24
0-20	1.00	0.99	1.02	1.06	1.10	1.11

Table 7.8: Additional corrections applied to generator level quantities in Drell-Yan $\tau\tau$ decays.

tau pair $p_{T_{vis}}$ (GeV)	Correction
0-100	0.971
100-150	0.975
150-200	0.960
200-250	0.964
250-300	0.934
> 300	0.942
leading di-jet mass m_{jj} (GeV)	Correction
300-700	1.070
700-1100	1.090
1100-1500	1.055
> 1500	1.015

(a) $\mu\mu$ mass, before reweighting

(b) $\mu\mu$ mass, after reweighting
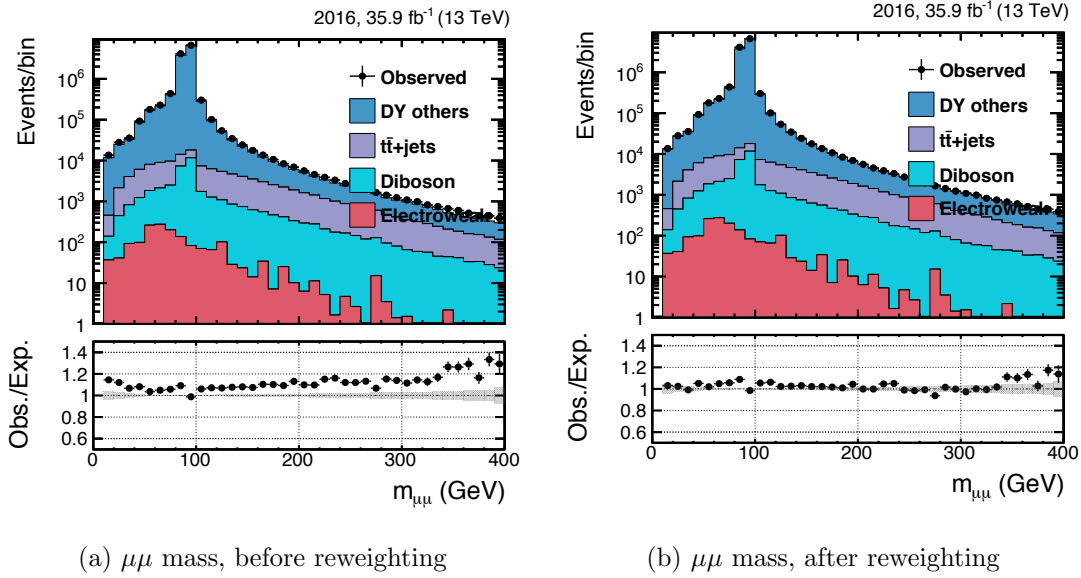
Figure 7.8: Di-muon mass distributions in $Z \to \mu\mu$ data before and after the reweighting discussed in this section.

(a) $\mathrm{p_T}(\mu\mu)$, before reweighting

(b) $\mathrm{p_T}(\mu\mu)$, after reweighting
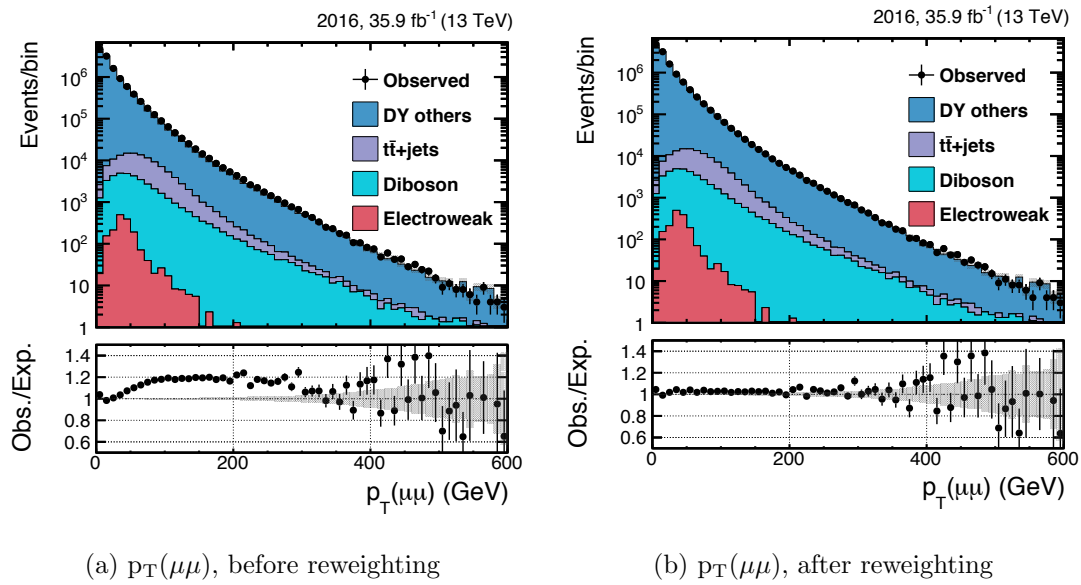
Figure 7.9: Di-muon $\mathrm{p_T}$ distributions in $Z \to \mu\mu$ data before and after the reweighting discussed in this section.

7.5.6 Top p_T reweighting

There is evidence that the top p_T distribution in $t\bar{t}$ events is softer in data than in simulation [159–162]. Corrections are applied as a function of the t quark p_T shown in Eqs. 7.3 and 7.4 to account for this mis-modeling.

$$w = \sqrt{SF(t) + SF(\bar{t})} \tag{7.3}$$

where

$$SF(p_T) = \exp(0.0615 - 0.0005 \times p_T). \tag{7.4}$$

These weights are only applied to $t\bar{t}$ and not to single top samples. The uncertainty on the correction is equal to the value of the correction.

7.5.7 Recoil corrections

Recoil corrections are applied to correct for any differences of $\vec{E}_T^{\mathrm{miss}}$ in the simulated samples of the Drell-Yan, W+Jets and Higgs production [1, 137, 163]. Electrons and Muons are reconstructed with better resolutions, 1-4% and 1-6% [144] respectively, than jets which have between 5-15% resolution. The hadronic resolution of the jets will contribute the most to the E_T^{miss} resolution when the boson p_T is low. Corrections are only applied to bosons that recoil against hadronic activity to ensure agreement between data and simulation.

The corrections are performed on the variable U defined as,

$$\vec{U} = \vec{E}_T^{\mathrm{mis}} - \vec{p}_{T,\nu}. \tag{7.5}$$

The variable, U, is the vectorial difference of the E_T^{miss} and the total neutrino transverse momentum, $p_{T tot,\nu}$, originating from the decay of the Z, W or Higgs boson. The corrections are dependent on the number of jets in the event and the boson p_T.

The parallel and perpendicular U components are fit. These fits are performed in bins of boson p_T (0-10, 10-20, 20-30, 30-50, and > 50 GeV), and bins of number of jets (N_{jets}

= 0, 1 or at least 2) [1, 2, 163]. The jet distribution falls off quickly, so above 2 jets we can group them together in the calculation for statistics. The corrections are propagated to $E_{\mathrm{T}}^{\mathrm{miss}}$ using Eq. 7.5 and applied to the simulated samples.

(a) Without recoil corrections

(b) With recoil corrections
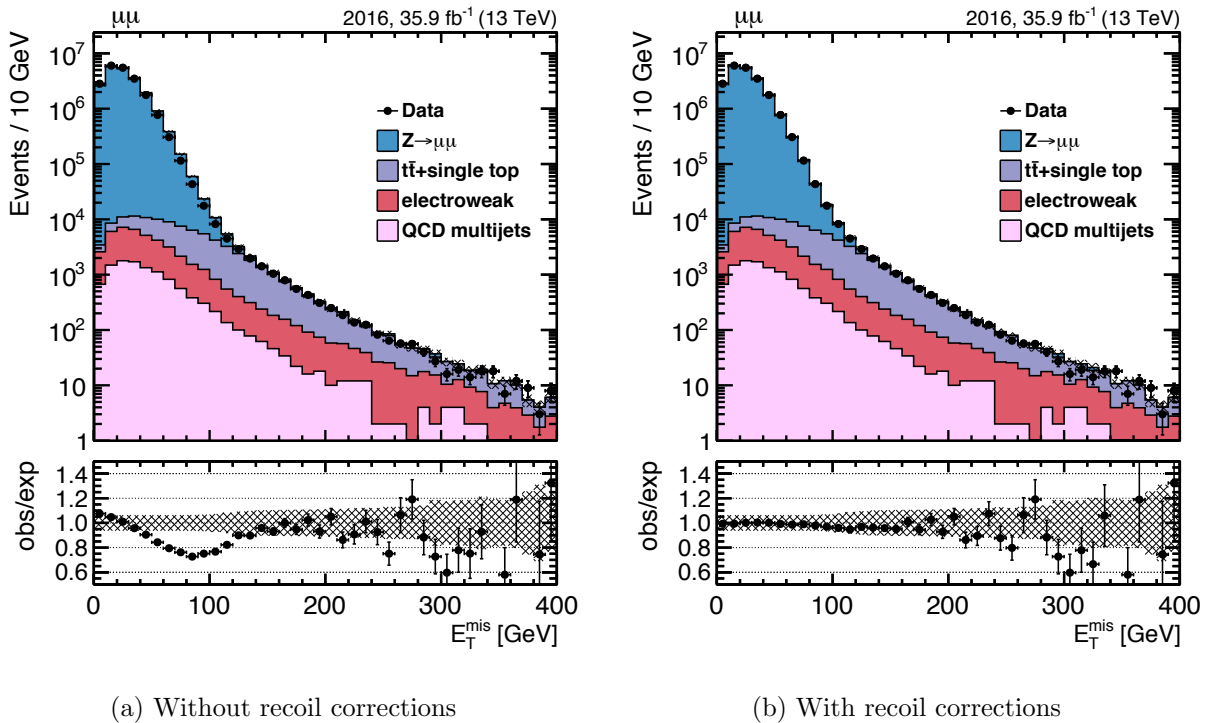
Figure 7.10: Effect of applying recoil corrections to the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution in the $Z \to \mu\mu$ selection. The recoil corrections improve the agreement between data and simulation.

The effect on $Z \to \mu\tau_{\mathrm{h}}$ is shown in Fig. 7.11. The recoil correction improved modeling of $E_{\mathrm{T}}^{\mathrm{miss}}$ in events with recoiling bosons. Above a boson p_{T} of 50 GeV, recoil corrections are not necessary. If no noticeable disagreement is seen in the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution, these corrections are not needed.
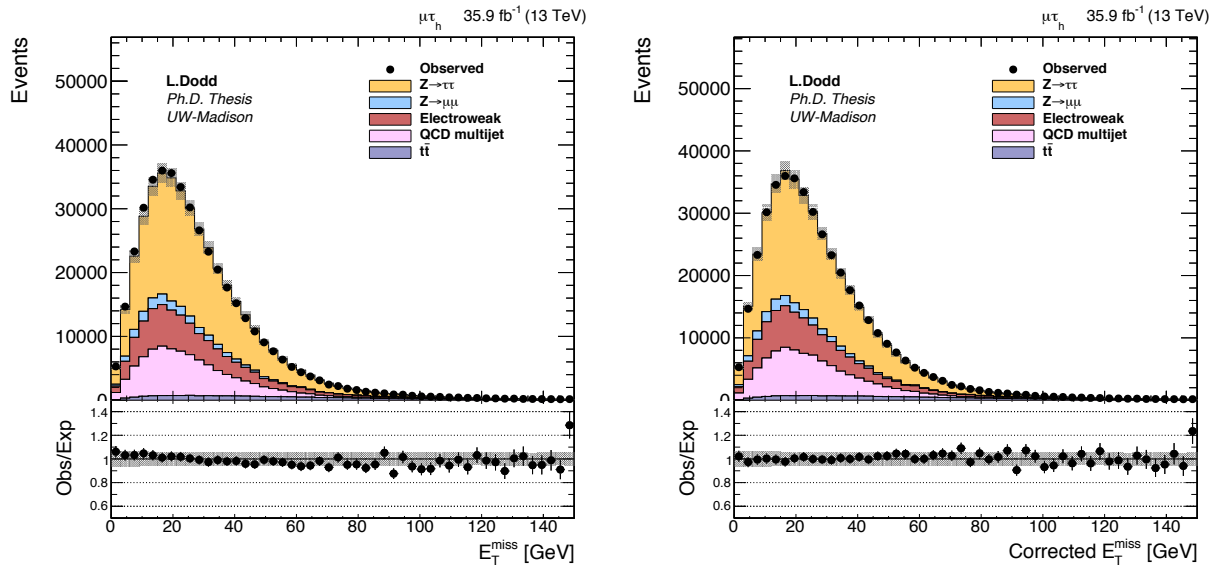
(a) Without recoil corrections   (b) With recoil corrections

Figure 7.11: Effect of applying recoil corrections to the E_T^{miss} distribution in the $Z \to \mu\tau$ selection. The recoil corrections improve the agreement between data and simulation.

7.5.8   Generator event weights and luminosity

Generator weights are applied on an event-by-event basis. Samples produced with the AMC@NLO generator contain both positive and negative event weights as discussed in Section 5.1. Event weights for simulation are scaled to the expected contributions for each sample. Negative weights are taken into account when present. Negative weights reduce the effective luminosity of the simulation.

7.6   Tau pair systematics

7.6.1   Uncertainties related to object reconstruction and identification

Hadronic tau reconstruction is usually the main source of experimental uncertainties in a di-τ final state. The τ_h identification efficiencies for genuine τ_h leptons sum up in quadrature to an overall rate uncertainty of 5%, including reconstruction efficiencies, and tracking efficiencies. This number is partially uncorrelated among channels because the τ_h candidates are required to pass different channel-dependent working points of the anti-lepton discriminators. Additionally, the triggers applied to each channel further motivates this de-correlation. The trigger efficiency uncertainty per τ_h leg amounts to an additional 5% uncertainty, which leads to a total of 10% uncertainty for processes estimated from simulations in the $\tau_\mathrm{h}\tau_\mathrm{h}$ final state.

An uncertainty of 1.2% for the visible energy scale of genuine τ_h leptons affects both the distributions and the contributions of signal and background processes. It is uncorrelated among the 1 prong, 1 prong + π^0, and 3 prong decay modes. The four-vector of the tau is scaled up(down) by 1.2% and the resulting scaled-up(-down) τ_h is propagated through the analysis chain resulting in a shifted $m_{\tau\tau}$ and $E_\mathrm{T}^\mathrm{miss}$. The 1.2% magnitude of the tau energy scale uncertainty was measured in $Z \rightarrow \mu\tau_\mathrm{h}$ events with a tag-and-probe measurement and subsequently inflated to take into account tracking uncertainties. This uncertainty is applied uniformly across all τ_h matched in simulation to a hadronically-decaying tau. Even in the most boosted categories between Chapter 8 and Chapter 9, the reconstructed τ_h typically have moderate p_T ($p_T <$ 200 GeV) and are reconstructed in the barrel of the detector, and always within tracker acceptance. The τ_h tracks cleanly bend in the magnetic field, so we treat the visible energy scale of genuine τ_h leptons as fully correlated for all τ_h leptons

considered in the analysis for a given decay mode.

For events where μ or e are misidentified as τ_h, i.e. $Z \to$ ee events in the eτ_h final state and $Z \to \mu\mu$ events in the $\mu\tau_\mathrm{h}$ final state, the τ_h identification leads to prefit rate uncertainties of 12(25%) for e(μ)-faking-τ_h.

In the decay channels with muons or electrons, the uncertainties in the muon and electron identification, isolation, and trigger efficiencies lead to rate uncertainties of 2% for muons, and 2% for electrons. The electron energy scale in eτ_h uncertainty is negligible compared to the τ_h energy scale uncertainty. In all channels, the effect of the uncertainty in the muon energy scale is negligible, we do not expect many muons with $p_\mathrm{T}(\mu) > 200$ GeV.

The $E_\mathrm{T}^\mathrm{miss}$ scale uncertainties [164], which are computed event-by-event, affect the event distributions through the event selection, and are propagated through the di-τ mass reconstruction. The unclustered $E_\mathrm{T}^\mathrm{miss}$ scale uncertainties come from four independent sources related to the CMS sub-detectors: tracker, ECAL, HCAL, and HCAL forward. The clustered $E_\mathrm{T}^\mathrm{miss}$ scale uncertainties are related to uncertainties in the jet energy scale measurement, which lead to uncertainties in the $E_\mathrm{T}^\mathrm{miss}$ calculation. The uncertainties on the jet energy scale depend on the jet p_T and jet η [151, 165, 166].

The $Z \to \tau\tau$ background normalization is corrected as a function of $p_T(\ell\ell)$, $m_{\ell\ell}$, and m_{jj}, derived from the agreement between data and background prediction in the $Z \to \mu\mu$-enriched region, which was the used to obtain the correction in Section 7.5.5. In simulated Z+jets samples, a shape uncertainty of 10% of the Z p_T reweighting correction, as described in Section 7.5.5, is considered.

The uncertainties on the W + jets contribution, in the case of the eτ_h and $\mu\tau_\mathrm{h}$ final states, are typically derived through the inclusion of dedicated control regions in the fit. They account for the statistical limitation of observed data, the effective luminosity of the W+jets simulation sample, and the systematic uncertainties of other processes in the control regions. The W+jets contribution in the $\tau_\mathrm{h}\tau_\mathrm{h}$ channel is much reduced form the $\ell\tau_\mathrm{h}$ channels,

due in part to lepton vetos and higher required $p_T^{\tau_h}$ thresholds.

In the $e\tau_h$ and $\mu\tau_h$ final states, uncertainties from the fit of the control regions with leptons passing relaxed isolation conditions are considered together with a 20% uncertainty that accounts for the extrapolation from the relaxed isolation region to the isolated signal region.

The combined systematic uncertainty in the background contribution arising from diboson and single-top-quark production processes is estimated to be 5% based on recent CMS measurements.

7.6.2   Other uncertainties

The uncertainty on the measured luminosity is measured to 2.5% precision, as in Ref. [101], is considered for simulated processes. This uncertainty is removed for data-driven background contributions. The luminosity measurement is described in Section 4.10.

Chapter 8

Observation of standard model Higgs boson decay to τ leptons

In this chapter I discuss the event selection, background estimation, signal extraction, further systematics and events weights beyond what was discussed in Chapter 7, and results for the $H \to \tau^+\tau^-$ analysis on 35.9 fb^{-1} collected in 2016 with the CMS experiment. The 2016 results are combined with Run-I $H \to \tau\tau$ results in Ref. [163]. More information on the $\tau_h\tau_h$ and eμ channels is found in Refs. [1, 2].

The overall structure consists of four channels: $\mu\tau_h$, eτ_h, $\tau_h\tau_h$, and eμ. Each channel is further split into three categories (0-jet, boosted, and vbf[1]) where the last two categories, boosted and vbf, are primarily targeting ggH and VBF production respectively. Each category in each channel is analyzed two-dimensionally, where one variable is a di-τ mass variable and the second variable is dependent on channel and category. Control regions are added into a maximum likelihood fit to control QCD, W + jet, and $t\bar{t}$ backgrounds.

---

[1]Lowercase *vbf* denotes the category targeting *VBF* production.

8.1 Event selection

8.1.1 Trigger requirements

The physics object of interest, e.g. muon, is required to be the object that fired the trigger, instead of an in-flight decay, or misidentified other object. The application of this cut allows for accurate application of trigger scale factors. All possible triggers are summarized in Table 7.2.

For the $\mu\tau_\mathrm{h}$ channel events are recorded using the single isolated muon trigger with an online L1 threshold of 18 GeV and an HLT threshold of 22 GeV and the $\mu - \tau_\mathrm{h}$ cross trigger, with the HLT $\mathrm{p_T}(\mu) > 19$ GeV and the HLT $\mathrm{p_T}(\tau_\mathrm{h}) > 20$ GeV.

Events with a muon with $\mathrm{p_T}$ between 20 and 23 GeV are required to fire the cross trigger. Events with a muon with $\mathrm{p_T}$ greater than 23 GeV are required to fire the single muon trigger with the online threshold of 22 GeV. The corresponding trigger scale factors, calculated on a $Z \to \mu\mu$ tag-and-probe study, are applied as a function of muon $\mathrm{p_T}$. Including events with a muon with $\mathrm{p_T}$ between 20 and 23 GeV increases the signal acceptance by about 25%, and the analysis sensitivity increases by about 10%.

For the $e\tau_\mathrm{h}$-channel, this thesis considers events collected using the single electron HLT trigger, which requires an HLT isolated electron with $|\eta < 2.1|$ and $\mathrm{p_T}$ greater than 25 GeV. The offline electron $\mathrm{p_T}$ requirement is greater than 26 GeV to avoid the turn-on region of the trigger. The scale factors, calculated in a $Z \to ee$ tag-and-probe study, account for the trigger differences between data and simulation. They are applied as a function of the $\mathrm{p_T}$ and also correct for brief mismatch between Level-1 electron $\mathrm{p_T}$ threshold and HLT $\mathrm{p_T}$ threshold during 2016 data, which results in a broader turn-on curve.

8.1.2 Kinematic selections

An event is selected if at least one $\mu\tau_{\rm h}(e\tau_{\rm h})$ pair is found. In case more than one $\tau_{\rm h}$ is found in the event, the more-isolated $\tau_{\rm h}$ is used for this analysis. Using isolation-sorting instead of $p_{\rm T}$-sorting increases signal efficiency by less than 1%. The $\mu(e)$ and $\tau_{\rm h}$ must have opposite-sign (OS) electric charges and must be separated by $\Delta R > 0.5$.

The kinematic selections are shown in Table 8.1 for all final states. The trigger requirement is defined by a combination of trigger objects with $p_{\rm T}$ over a given threshold, indicated inside parentheses. The offline selections for each final state are listed in the lepton selection column. The muon identification, electron identification, and $\tau_{\rm h}$ identification criteria are previously defined and discussed in Sections 6.4, 6.5, and 6.8, respectively. Most kinematic selections are driven by online trigger thresholds. The $p_{\rm T}(\tau_{\rm h})$ cut for the $\ell\tau_{\rm h}$-channel was raised to 30 GeV to reduce further the QCD background.

Moreover, events are vetoed that contain additional leptons as described in Section 7.2.2.

Table 8.1: Kinematic selection requirements for the four $\tau\tau$ decay channels.

Final state	Trigger requirement	Lepton selection		
		p_T (GeV)	η	Isolation
$\mu\tau_{\rm h}$	$\mu(22)$	$p_{\rm T}^{\mu} > 23$	$|\eta^{\mu}| < 2.1$	$I_{rel}^{\mu} < 0.15$
	–	$p_{\rm T}^{\tau_{\rm h}} > 30$	$|\eta^{\tau_{\rm h}}| < 2.3$	MVA $\tau_{\rm h}$ ID
	$\mu(19)$	$20 < p_{\rm T}^{\mu} < 23$	$|\eta^{\mu}| < 2.1$	$I_{rel}^{\mu} < 0.15$
	$\tau_{\rm h}(21)$	$p_{\rm T}^{\tau_{\rm h}} > 30$	$|\eta^{\tau_{\rm h}}| < 2.3$	MVA $\tau_{\rm h}$ ID
$e\tau_{\rm h}$	$e(25)$	$p_{\rm T}^{e} > 26$	$|\eta^{e}| < 2.1$	$I_{rel}^{e} < 0.1$
	–	$p_{\rm T}^{\tau_{\rm h}} > 30$	$|\eta^{\tau_{\rm h}}| < 2.3$	MVA $\tau_{\rm h}$ ID
$\tau_{\rm h}\tau_{\rm h}$	$\tau_{\rm h}(35)$ (leading)	$p_{\rm T}^{\tau_{\rm h}} > 50$	$|\eta^{\tau_{\rm h}}| < 2.1$	MVA $\tau_{\rm h}$ ID
	$\tau_{\rm h}(35)$ (sub-leading)	$p_{\rm T}^{\tau_{\rm h}} > 40$	$|\eta^{\tau_{\rm h}}| < 2.1$	MVA $\tau_{\rm h}$ ID
$e\mu$	$e(12)$ (sub-leading)	$p_{\rm T}^{e} > 13$	$|\eta^{e}| < 2.5$	$I_{rel}^{e} < 0.15$
	$\mu(23)$ (leading)	$p_{\rm T}^{\mu} > 24$	$|\eta^{\mu}| < 2.4$	$I_{rel}^{\mu} < 0.2$
	$e(23)$ (leading)	$p_{\rm T}^{e} > 24$	$|\eta^{e}| < 2.5$	$I_{rel}^{e} < 0.15$
	$\mu(8)$ (sub-leading)	$p_{\rm T}^{\mu} > 15$	$|\eta^{\mu}| < 2.4$	$I_{rel}^{\mu} < 0.2$

8.1.3 $\mu\tau_{\mathrm{h}}$ event selection

The muon is required to pass the medium muon identification discussed in Section 6.4. The medium muon working point has around 99% efficiency, and a 0.65% rate of muons from in-flight meson decays. This working point had two run-dependent versions applied for 2016 data-taking where the fraction of valid tracker hits is > 0.49 for Run B-F, and > 0.8 for Run G-H. The scale factors are lumi-averaged to account for the small differences in the working point versions.

The τ_{h} candidate is required to be within tracker acceptance, $\eta_{\tau_{\mathrm{h}}} < 2.3$, pass the tau reconstruction algorithm, and be identified as a 1- or 3-prong tau. The τ_{h} is also required to pass the tight working point of the MVA discriminant discussed in Section 6.8 with an efficiency around 50% for $H \to \tau\tau$ simulation, and a $jet \to \tau_{\mathrm{h}}$ fake rate of 10^{-3}.

The selected τ_{h} must match to the primary vertex, satisfying the requirement $d_{Z\tau} < 0.2$ cm. To reduce contribution from background processes where a prompt lepton is misidentified as a hadronic tau the selected τ_{h} must satisfy strict anti-muon identification and looser anti-electron identification. The anti-muon working point is chosen specifically to reduce the large $Z \to \mu\mu$ background, where the μ fakes τ_{h}.

8.1.4 eτ_{h} event selection

The electron is identified with the non-triggering MVA ID algorithm in Section 6.5, using a working point that corresponds to an efficiency of 80%. The longitudinal and transverse impact parameters of the electron track with respect to the primary vertex are less than 0.2 cm and less than 0.045 cm, respectively. The electron track has at most 1 missing inner hit and passes the conversion veto. The offline $p_{T,e} > 26\mathrm{GeV}$, which is 1 GeV higher than the trigger threshold, and $|\eta_e| < 2.1$, corresponding to the trigger acceptance. The offline p_T selection must be tighter than the trigger and the relative isolation, shown in Eq. 6.1,

measured in a cone with $\Delta R = 0.3$, should be smaller than 0.1.

The reconstructed hadronic tau candidate identification and selection is the same as the $\mu\tau_{\mathrm{h}}$ channel, except for the anti-lepton discriminators; the tight working point of the anti-electron discriminator is applied to suppress $Z \to ee$. The loose working point of the anti-muon discriminator is applied to reject other multi-leptonic backgrounds.

8.1.5 Transverse mass

The transverse mass, M_T, reconstructed from a muon and missing transverse energy is defined in Eq. 8.1:

$$M_T = \sqrt{2p_{\mathrm{T}}^{\ell}E_{\mathrm{T}}^{\mathrm{miss}}\left(1 - cos(\phi)\right)} \tag{8.1}$$

where p_{T}^{ℓ} and $E_{\mathrm{T}}^{\mathrm{miss}}$ are transverse momentum of the ℓ (muon or electron) and missing transverse energy, and ϕ is an angle between these two four-vectors in the transverse plane. The distribution of M_T in signal and background processes is shown in Fig. 8.1. The peak near the W boson mass is visible.

The transverse mass in signal processes, gluon fusion and vector boson-fusion is expected to be small, while for some background processes, such as W+jets and $t\bar{t}$, it is expected to be large. Thus, selected events must satisfy a requirement $M_T < 50$ GeV, which is obtained by an optimization scan using $S/\sqrt{S+B}$, shown in Fig. 8.2 for the $\mu\tau_{\mathrm{h}}$ channel.
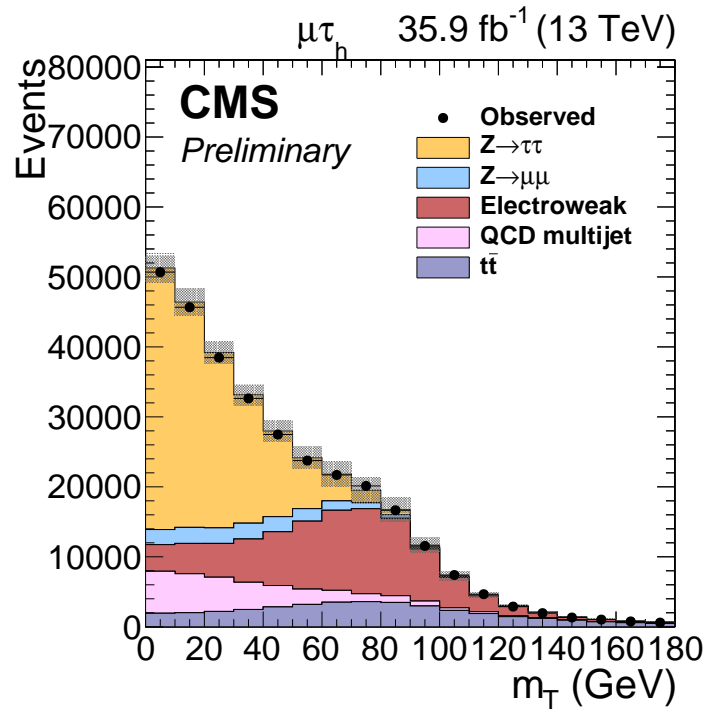
Figure 8.1: Prefit observed and expected distributions of the transverse mass between the muon and E_T^{miss} , in the $\mu\tau_h$ final state.
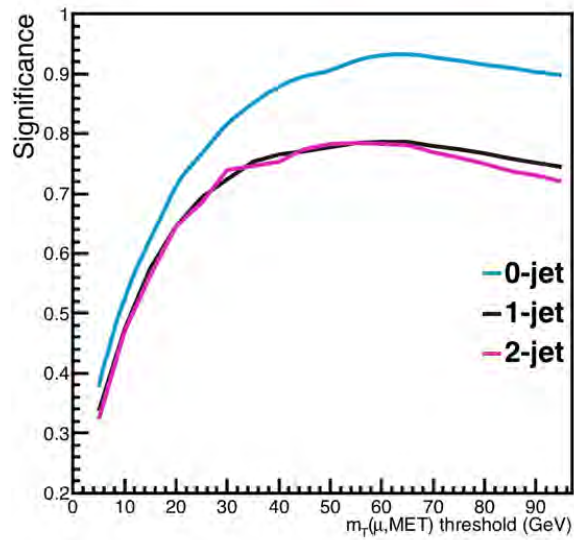
Figure 8.2: Signal significance as a function of the M_T cut applied in the $\mu\tau_h$ selection, split into rough 0-jet(gluon fusion), 1-jet (gluon fusion), and 2-jet (VBF) categories.

The event cut flow showing the observed number of events after each stage in the selection process is shown in Table 8.2.

Process	$\mu\tau_{\mathrm{h}}$ Events	$e\tau_{\mathrm{h}}$ Events
Dataset/Trigger	786751488	942127424
Extra Lepton Veto and OS tau pair	9731414	6141090
Tight lepton identification	845627	343863
$M_T < 50$ GeV	492668	171301
Total expected VBF contribution	109.9	39.6
Total expected ggH contribution	1140.7	356.0

Table 8.2: Number of observed events in selection process and expected contribution of VBF and ggH after these selection cuts are applied.

8.2   Signal extraction

In Run-I, the analysis was split into various categories targeting various production phase spaces. However, the thesis choses a method that better scales to higher integrated luminosities.

Each channel is split into categories depending on production mode. There is the 0-jet category, the boosted category, and the vbf category. The expected Higgs boson production mechanism for the 0-jet category is 98% gluon fusion events and it is dominated by Z+jets. Despite the overwhelming irreducible background, the 0-jet category adds to the final result by providing a high statistics region for the maximum likelihood fit in order to constrain some uncertainties, such as the tau energy scale.

The vbf category targets specifically Higgs boson produced via VBF, by requiring at least two jets with a large di-jet mass. Only 57% of the Higgs boson yield expectation in this category is from VBF. The remaining 43% is from gluon fusion events with extra jets. The boosted category is expected to contain mostly Higgs bosons boosted from gluon fusion and signal from vector boson fusion where a jet is missed in reconstruction. About 80% of

signal events in this category are from gluon fusion production, while the remaining events are split between vector boson production, about 12%, and VH associated production with hadronic decays 8%. The definition of the three categories is given in Table 8.3, with the contributions in each category given in Table 8.3.

8.2.1 Two-dimensional extraction

A two-dimensional (2D) signal extraction is performed. Depending on the production mechanism of the Higgs, there are other variables besides the mass variable that can separate signal and background. Vector boson fusion is highly dependent on the di-jet mass of the two leading jets in the event, thus m_{jj} mass is chosen as the second extraction variable to target the outgoing partons from the vector boson fusion process. The Higgs boson in gluon fusion production is frequently boosted compared with background, therefore the $\tau\tau$ p_T is chosen as the second extraction variable to target gluon fusion events.

Figure 8.4 demonstrates several 2D distributions for targeted signal and leading background contributions. To fully profit from the differences in the second dimension, these 2D distributions are unrolled in slices of the second variable, with the choices visible in Fig. 8.3.

	0-jet	VBF	Boosted
		Selection	
$\tau_h\tau_h$	No jet	$\geq$2 jets, $p_T^{\tau\tau} > 100\,\mathrm{GeV}$, $\Delta\eta_{jj} > 2.5$	Others
$\mu\tau_h$	No jet	$\geq$2 jets, $m_{jj} > 300\,\mathrm{GeV}$, $p_T^{\tau\tau} > 50\,\mathrm{GeV}$, $p_T^{\tau_h} > 40\,\mathrm{GeV}$	Others
$e\tau_h$	No jet	$\geq$2 jets, $m_{jj} > 300\,\mathrm{GeV}$, $p_T^{\tau\tau} > 50\,\mathrm{GeV}$	Others
$e\mu$	No jet	2 jets, $m_{jj} > 300\,\mathrm{GeV}$	Others
		Observables	
$\tau_h\tau_h$	$m_{\tau\tau}$	$m_{jj}, m_{\tau\tau}$	$p_T^{\tau\tau}, m_{\tau\tau}$
$\mu\tau_h$	τ_h decay mode, m_{vis}	$m_{jj}, m_{\tau\tau}$	$p_T^{\tau\tau}, m_{\tau\tau}$
$e\tau_h$	τ_h decay mode, m_{vis}	$m_{jj}, m_{\tau\tau}$	$p_T^{\tau\tau}, m_{\tau\tau}$
$e\mu$	p_T^{μ}, m_{vis}	$m_{jj}, m_{\tau\tau}$	$p_T^{\tau\tau}, m_{\tau\tau}$

Figure 8.3: Category definitions and second dimension used in signal extraction.

The numbers of expected events in each category is shown in Table 8.3.
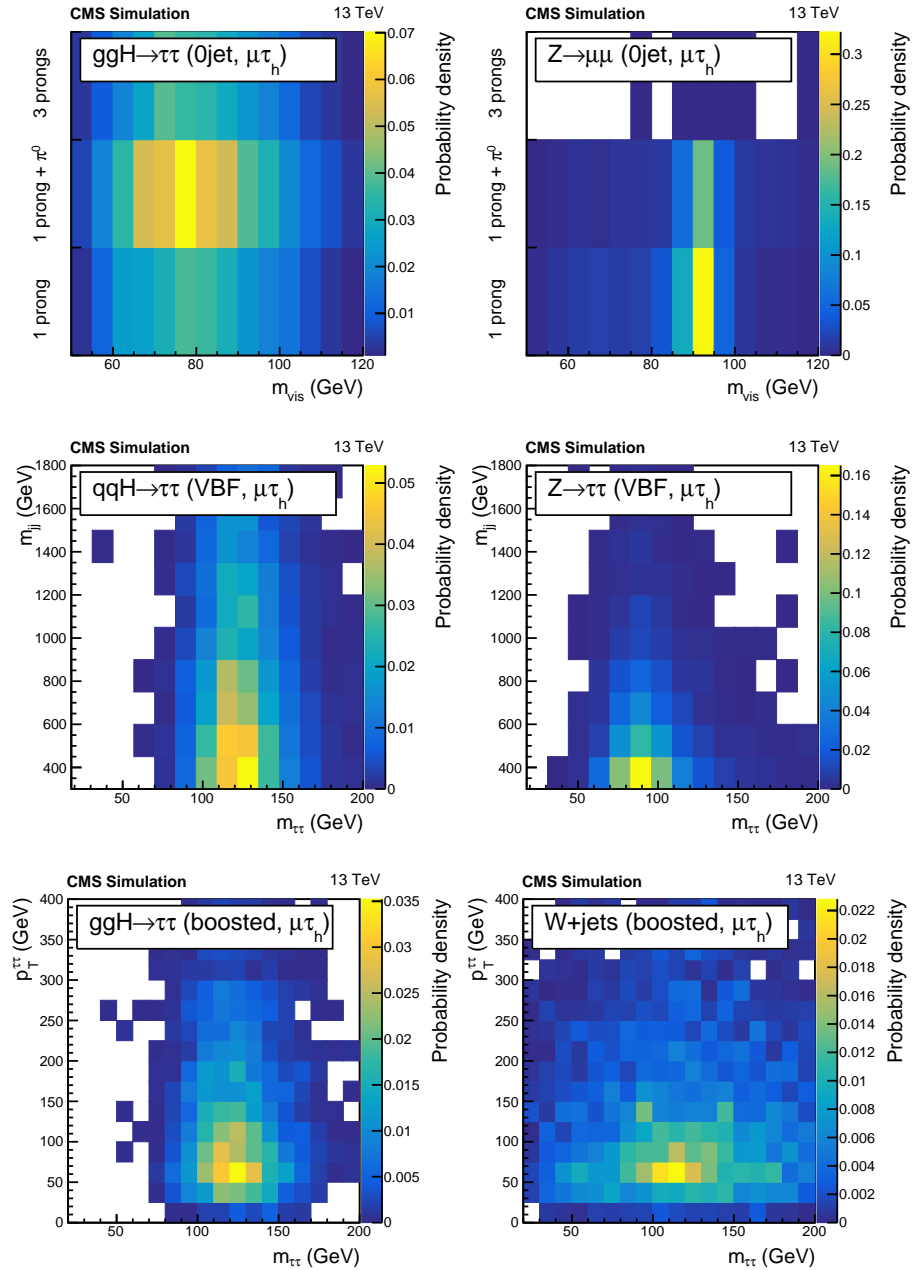
Figure 8.4: Two-dimensional distributions of various signal and background processes demonstrating clear differences between signal and background when compared two-dimensionally.

Process	$\mu\tau_{\mathrm{h}}$ Events	$e\tau_{\mathrm{h}}$ Events
Observed $M_T < 50$ GeV	492668	171301
Observed 0-jet	128571	41160
Observed boosted	60127	21250
Observed vbf	2927	2088
Total expected VBF events	109.9	39.6
exp. 0-jet	5.74	1.77
exp. boosted	58.3	18.5
exp. vbf	29.5	16.9
Total expected ggH events	1140.7	356.0
exp. 0-jet	569.1	187.2
exp. boosted	359.8	130.6
exp. vbf	31.2	18.1

Table 8.3: Number of observed events for $\mu\tau_{\mathrm{h}}$ and eτ_{h} in each category, with expected number of ggH and VBF events.

8.3 Background estimation

Simulated samples used in this analysis are reconstructed and stored in a reduced format. These samples all belong to the samples produced specifically to model 2016 running conditions. All samples are generated for p–p collisions at a center-of-mass energy of 13 TeV.

Various processes listed in Table 8.4 are included as backgrounds. Any background with one or more isolated lepton and/or jets is considered as background. QCD/multijet is typically not isolated, but due to the large cross section it is a large process for the $\ell\tau_{\mathrm{h}}$ channel concentrated at low M_T. For example $WZ \to 3\ell 1\nu$ is considered as a background despite employing an additional lepton veto. If one lepton is out of acceptance, the other two leptons can be considered as the tau-candidate pair.

The signal simulation samples included in the analysis and the considered cross sections are in Table 8.5. For each Higgs boson mass point there is a different branching fraction used as listed in Table 8.6.

Background Simulations	Cross section (pb)
QCD (multijet)	Data-Driven
Drell-Yan Inclusive Jet Production	4954.0 pb (LO)
Drell-Yan 1-jet	1012.5 pb (LO)
Drell-Yan 2-jet	332.8 pb (LO)
Drell-Yan 3-jet	101.8 pb (LO)
Drell-Yan 4-jet	54.8 pb (LO)
Drell Yan $M > 150$	6.657 pb (LO)
$t\bar{t}$	831.76 pb
W+jets Inclusive Jet Production	50380 pb (LO)
W+jets 1-jet	9644.5 pb (LO)
W+jets 2-jet	3144.5 pb (LO)
W+jets 3-jet	954.8 pb (LO)
W+jets 4-jet	485.6 pb (LO)
$WZ \to 1\ell 3\nu$	3.05
$WZ \to 1\ell 1\nu 2Q$	10.71
$WZ \to 2\ell 2Q$	5.595
Single $\bar{t}$ + W	35.6
Single t + W	35.6
Single t	80.95
Single $\bar{t}$	136.02
$WW \to 1\ell 1\nu 2Q$	1.212
$ZZ \to 2\ell 2Q$	3.22
$VV \to 2\ell 2\nu$	11.95
EWK $WZ \to 3\ell\nu$	4.708
$WZ \to 1\ell 1\nu 2Q$	10.71
$WZ \to 1\ell 3\nu$	3.05
$WZ \to 2\ell 2Q$	5.595
$ZZ \to 4\ell$	1.212
EWK $W^- + 2Jets$	20.25
EWK $W^+ + 2Jets$	25.62
EWK $Z \to \ell\ell + 2Jets$	3.987
EWK $Z \to \nu\nu + 2Jets$	10.01
$gg\ H \to WW \to 2\ell 2\nu$	1.001
VBF $H \to WW \to 2\ell 2\nu$	0.0858

Table 8.4: Background simulation samples included. A LO-NLO factor of 1.16 is considered for the Z+jets samples, and 1.21 for the W+jets samples listed in Table 8.4.

Signal sample	m_H (GeV)	cross section (pb)
gluon fusion	110	57.9
gluon fusion	120	52.22
gluon fusion	125	48.58
gluon fusion	130	45.31
gluon fusion	140	36.0
vector boson fusion	110	4.434
vector boson fusion	120	3.935
vector boson fusion	125	3.782
vector boson fusion	130	3.637
vector boson fusion	140	3.492
ZH	110	1.309
ZH	120	0.994
ZH	125	0.884
ZH	130	0.790
ZH	140	0.6514
$W^+(^-)$	110	1.335 (0.8587)
$W^+(^-)$	120	0.9558 (0.6092)
$W^+(^-)$	125	0.840 (0.5328)
$W^+(^-)$	130	0.7414 (0.4676)
$W^+(^-)$	140	0.6308 (0.3940)

Table 8.5: Signal samples included in the thesis and their respective cross sections include NNLO+NNLL QCD and NLO EW corrections.

m_H (GeV)	$\tau\tau$ Branching fraction (%)
110	0.0791
120	0.0698
125	0.0627
130	0.0541
140	0.0360

Table 8.6: $H \rightarrow \tau\tau$ branching fractions depending on Higgs boson mass

8.3.1 Drell-Yan

For all Z+Jets backgrounds a correction related to the generator-level mass and p_T is applied as described in Section 7.5.5.

The simulation Z+jet modeling is checked in $Z \to \mu\mu$ events. The $Z \to \mu\mu$ enriched region is obtained by selecting events with two opposite-sign muons passing the medium muon ID with relative isolation less than 0.15, $p_T > 26$ GeV, and $|\eta| < 2.1$. The events are required to fire the single isolated muon trigger with an online threshold of 24 GeV. The purity is increased by requiring the visible mass of the muon-pair to be between 70 and 110 GeV. Events with additional electrons or muons are vetoed. The events are divided into the same 2D categories, as described in Section 8.2.1. The data/simulation distribution agreement is shown in Fig. 8.5 for the 0-jet, vbf, and boosted categories.

The Higgs boson p_T, which is one of the fitting variables in the boosted category, is equivalent to the di-muon p_T in $\mu\mu$ events. It is well modeled in simulation because the simulation has already been reweighted to match the Z boson mass and p_T distributions in data, as discussed in Section 7.5.5. Some data/MC differences are observed as a function of the invariant mass of the leading jets in the vbf category. The weights based on p_T and di-jet mass are remeasured after selecting muons with higher p_T thresholds in order to be closer to the kinematic selection in the signal regions, where the taus before decaying should have large p_T to give, for example, a muon with $p_T > 20$ GeV and a τ_h with $p_T > 30$ GeV. In the case of the vbf category, the weights become flat when selecting muons with a higher p_T threshold, as shown in Fig. 8.6 for a threshold of 50 GeV on both muons. Therefore half of the measured weight with a threshold of 25 GeV are applied to the nominal shapes, whereas the weights obtained with 25 GeV as a threshold, and no applied scaling, are considered as two alternative shapes for systematic uncertainties.

The Z+Jets backgrounds are separated into three main components based on whether

(a) 0-jet

(b) boosted

(c) vbf
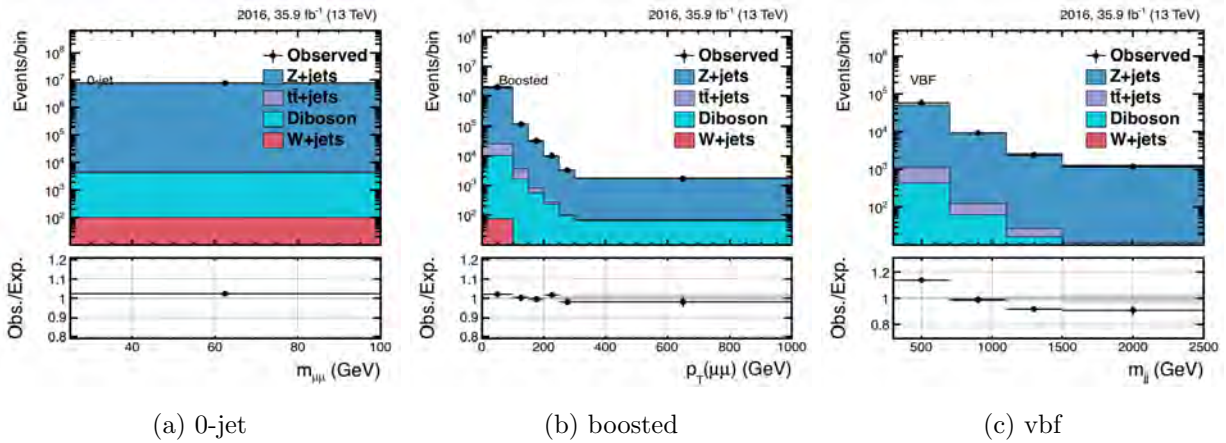
Figure 8.5: Data/simulation comparison in a $Z \to \mu\mu$ control region. The data/simulation ratio are applied as scale factors for the $Z \to \tau\tau$ background in the signal region.

(a) Muon $p_T > 25$ GeV

(b) Muon $p_T > 50$ GeV

(c) Muon $p_T > 70$ GeV
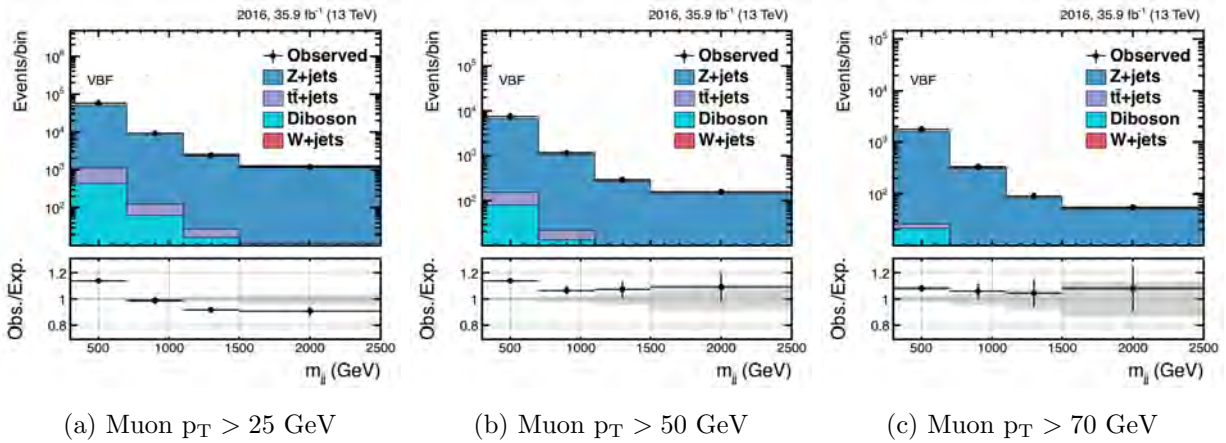
Figure 8.6: Data/simulation comparison in a $Z \to \mu\mu$ control region with vbf category criteria, for varying muon p_T thresholds.

or not the τ_h candidate is matched at generator-level to a true generated τ_h. The Z+jets simulated distribution is split depending on if the tau is classified as a matched to a generator-level hadronically-decaying tau, a prompt lepton, or none (i.e. jet-fake-τ_h).

8.3.2   W+jets background

Selected tau pair events with the requirement $M_T > 80$ GeV are used to check the normalization of the W+jets background. This high-M_T sideband is dominated by the W+jets process, but there are contributions from other processes as well. In particular with the increase of QCD cross section in Run II at the LHC, we find a larger fraction of QCD in the high-M_T region than in the Run I analysis [163]. New background methods are developed to account for this infiltration of QCD in the W+jets region.

The W+jets distribution for the low-M_T signal ($M_T < 50$ GeV) region is extracted from simulation using the relaxed selection[2] is used described in Section 8.3.4. The relaxed selection does not bias the limit more than about 3%, while reducing statistical fluctuations.

The expected W+jets contribution in the signal region is initially extracted from simulation. Then the expected W+jets contribution is scaled such that the W+jets opposite-sign $M_T > 80$ GeV control region exhibits good agreement.

Assuming the other backgrounds are controlled by the fit, the dominant process in this phase space is W+jets providing a relatively pure, ranging from 50% to 72% depending on category, region to estimate the W+jets. The postfit control regions are shown in Fig. 8.29. The W+jets is scaled on the input to the maximum likelihood fit, but it is allowed to fluctuate in the global maximum likelihood fit in case a more optimal contribution is found. The W_{SF}, defined as

$$W_{SF} = \frac{W_{\text{observed,OS,W}}}{W_{\text{Sim. yield,OS, W}}} \tag{8.2}$$

will be used to derive the QCD contribution in the $M_T < 50$ GeV region as well.

The W+jets estimation is repeated for each category in $\mu\tau_\text{h}$ and $e\tau_\text{h}$ separately. In practice the vbf category at high m_T has a low W+jets purity, and no difference is expected

---

[2]The relative isolation of the muon (electron) is required to be between 0.15 and 0.30 (0.10 and 0.30) and the τ_h MVA identification working point used is relaxed from medium to tight.

between the W+jets data/simulation scaling in the boosted and vbf regions because they both target events with several jets. Therefore, two control regions (0-jet and boosted) with $M_T > 80$ GeV are included as control regions in the final fit to constrain dynamically the normalization of the W+jets process. The W+jets in the vbf category signal region is linked to the boosted control region and normalization of the W+jets process.

8.3.3 QCD/multijet

QCD multijet contribution is estimated using the same-sign(SS) low-transverse mass ($M_T < 50$) sideband, denoted $_{SS,Low}$. The shape of the distribution is relaxed according to Section 8.3.4 and does not bias the limit more than about 3% as described in the section. This method is outlined in Section 7.4.1.

In the SS $M_T < 50$ GeV region, the contribution from expected simulated background processes is subtracted from data. The subtracted W+jets simulation is scaled as in Eq. 8.2. The remaining yield is scaled by a factor describing ratio of opposite-sign to same-sign events in a QCD dominated region containing anti-isolated muons (electrons). This factor is measured independently for each category and these results are summarized in Table 8.7. The factor is measured in an anti-isolated lepton category $I_{rel} < 0.3(0.25)$ and $I_{rel} > 0.1(0.15)$ with OS tau pair, where the QCD contribution is taken from the anti-isolated SS tau pair region. The QCD contribution in this region is fit using a maximum likelihood fit treating QCD as the signal strength. The resulting signal strength is the $OS \rightarrow SS$ factor used in the isolated regions. Any non closure with the maximum likelihood fit of QCD in this region, with a generic scaling treatment, is treated as additional uncertainty. The effect of varying the W+jets expected contribution on the QCD expected contribution was used as an extra systematic uncertainty. Additionally, the statistical-only uncertainty of the scale factors is included in the error estimate of the $SS \rightarrow OS$.

The postfit distributions in the regions with anti-isolated muons, from which the scale factors are extracted, are shown in Fig. 8.27. The vbf category QCD contribution is controlled by the boosted category due to low final state statistics in the anti-isolated region. The distributions used for extraction are shown in Fig. 8.7.

In summary the QCD expectation is found as follows:

$$QCD_{Yield,OS,Low} = \text{QCD OS/SS factor} \times$$

$$(\text{DATA}_{\text{SS,Low}} - \text{VV,Z+jets,TT,Other}_{\text{SS,Low}} - \text{W}_{SF} \times (\text{W}_{\text{SS,Low}}))$$

(8.3)

The QCD distribution is similarly taken from Eq. 8.3, although the simulation distributions are obtained with the relaxed selection for the boosted and vbf categories.

Category	OS/SS QCD ratio	Error	Fraction of W+jets in high M_T same-sign	Fraction of QCD in high M_T opposite-sign
0-jet $\mu\tau_{\text{h}}$	1.07	15 %	0.72	0.18
boosted $\mu\tau_{\text{h}}$	1.06	15 %	0.50	0.16
vbf $\mu\tau_{\text{h}}$	1.00	30 %	0.50	0.28

Table 8.7: QCD OS/SS ($SS \rightarrow OS$) ratios measured in the $\mu\tau_{\text{h}}$ final state.

Table 8.7 shows the relative fractions of of W+jets and QCD in the W+jets region, and the $SS \rightarrow OS$ results for the $\mu\tau_{\text{h}}$ channel. The eτ_{h} results are shown in Table 8.8.

Control regions for the QCD obtained in the opposite-sign region with anti-isolated taus are added to the fit to adjust the QCD normalization dynamically.

Category	OS/SS QCD ratio	Error
0-jet	1.00	15 %
boosted	1.28	15 %
vbf	1.0	30 %

Table 8.8: QCD OS/SS ($SS \rightarrow OS$) ratios measured in the eτ_{h} final state.

(a) $\mu\tau_h$ vbf anti-isolated
(b) $e\tau_h$ vbf anti-isolated
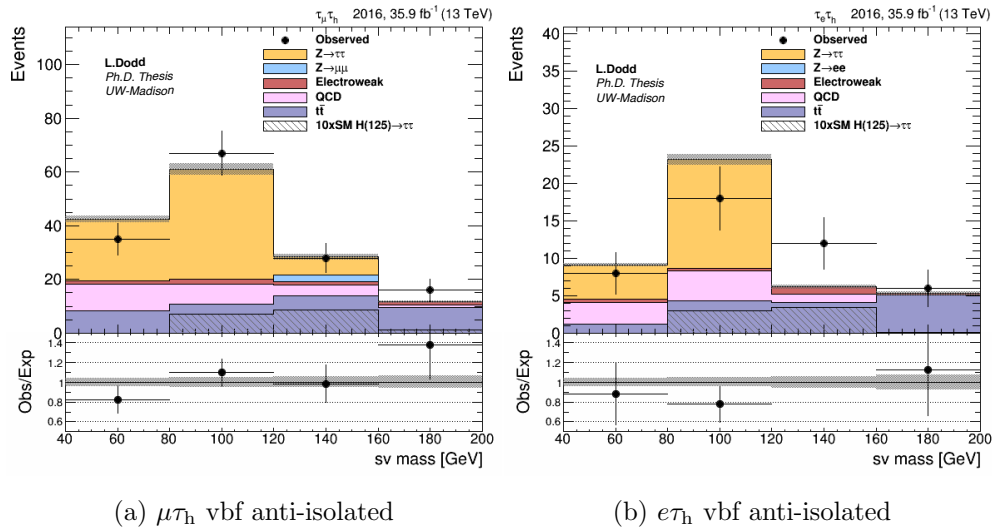
Figure 8.7: Prefit anti-isolated vbf regions used to compute the QCD OS/SS scale factors for the vbf category. The uncertainty on the measured factors is constrained by the boosted category in the fit, with little difference on the final result.

8.3.4 Relaxed selection

To improve statistical accuracy of some background processes a relaxed selection is used. The relative isolation of the muon (electron) is required to be between 0.15 and 0.30 (0.10 and 0.30) and the τ_h MVA identification working point used is relaxed from medium to tight. All other selections remain the same. The selection is determined so that it does not bias the shape of the distribution. Kolmogorov-Smirnov tests were performed between the nominal and various loosened isolation working points to determine the most relaxed shape that is still compatible with the original nominal shape. The criteria used to determine compatibility is Kolmogorov-Smirnov (KS) test values between the two histograms must be ≥ 0.10. A KS test result $\ll 1$ indicates incompatibility. The relaxed selection that uses looser isolation criteria for muons (electrons in the $e\tau_h$ final state) is chosen. The relative delta-beta corrected muon isolation is loosened to < 0.3 (from < 0.15) and the tau ID is loosened to the Medium MVA working point. The extra lepton vetoes are the same as in the

signal region, as well as the other τ_h or muon (electron) identification criteria. The relaxed selection is most important in the boosted and vbf categories.

The KS test indicates the relaxed selection is not incompatible with the nominal selection. However, the KS test does not show whether any other potential biases could be expected by using a relaxed selection. The QCD and W+jets shapes are relaxed and expected limits are computed with the (non-relaxed) Asimov data set for the vbf and boosted categories. The relaxed shapes do not affect the expected limits. Only the W+jets and QCD processes use a relaxed shape.

8.3.5   Other backgrounds

Most other backgrounds, listed in the Table 8.4, are separated into two main components: those backgrounds with fake-τ_h and those backgrounds with generator-level matched hadronic taus. Separation allows for easier application of additional corrections and systematic uncertainties to τ_h-fakes. The requirement to be matched to a generator-level hadronic tau necessitates summing the four-vectors of all visible tau decay products excluding muons and electrons. The sum of the decay products must be larger than 15 GeV. The normalization of the $t\bar{t}$ background is estimated from a $t\bar{t}$-enriched control region. The diboson and single-t samples are grouped together.

8.3.6   Mass extraction techniques

The `SVfit` algorithm [167] is used to reconstruct the $\tau\tau$ mass that is used as final observable in the boosted and vbf categories. The `SVfit` algorithm described in [167] returns the most likely value of the $m_{\tau\tau}$. The inputs are the four vectors of each tau candidate, the type of decay for each tau, the missing transverse energy in the event, and the uncertainty on the missing transverse energy (provided as a covariance matrix). The `SVfit` algorithm assumes

that all the missing energy in the event comes from the τ decay. This assumption results in incorrect mass reconstruction for events with E_T^miss from other sources than the neutrinos in the τ decay, for example $ZH \to \tau_\text{h}\tau_\text{h}\nu\nu$ will have a mass reconstruction very far from 125 GeV. The SVfit reconstructed mass compared to the visible mass of the $\tau\tau$ pair (m_vis) is shown in Fig. 8.8. SVfit is a computationally intensive algorithm and every event is run by the algorithm 10,000 times and the most likely $\text{m}_{\tau\tau}$ of these 10,000 is returned.
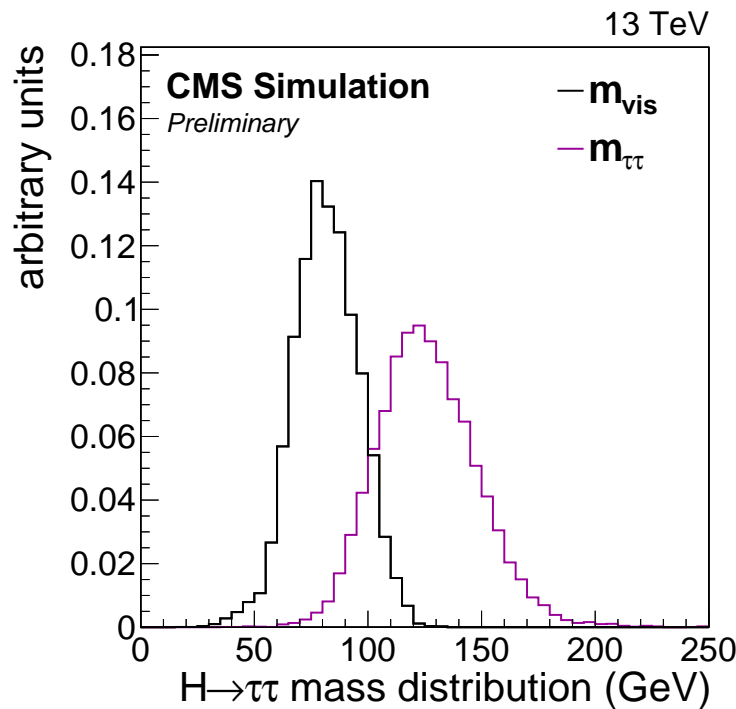
Figure 8.8: Normalized distributions of the visible mass, m_vis, and of the SVfit mass, $\text{m}_{\tau\tau}$, for a signal sample with a SM Higgs boson of mass $m_H = 125$ GeV decaying to a pair of τ leptons in the $\mu\tau_\text{h}$ final state. The $\text{m}_{\tau\tau}$ value is centered over 125 GeV, while m_{vis} is close to 90 GeV. The simulated events pass the required event selections and the $M_T < 50$ GeV criteria is applied.

For 0-jet categories, the visible mass of the tau-pair candidates is used for signal extraction. Separation between signal and the $Z \to \ell\ell$ background is much better with the m_{vis} than with the SVfit. The leptons from Z+jets that fake taus have a reconstructed mass at

90 GeV, while $Z \to \tau\tau$ peaks closer to 60 GeV. Since the $H \to \tau\tau$ broadly peaks near 90 GeV, it is beneficial to use visible mass as our mass reconstruction variable. Comparison of inclusive visible mass and `SVfit` distributions in a 0-jet category can be seen in Fig. 8.9, to motivate that decision.

(a) 0-jet $\mu\tau_{\mathrm{h}}$ m_{vis} (b) 0-jet $\mu\tau_{\mathrm{h}}$ m_{svfit}
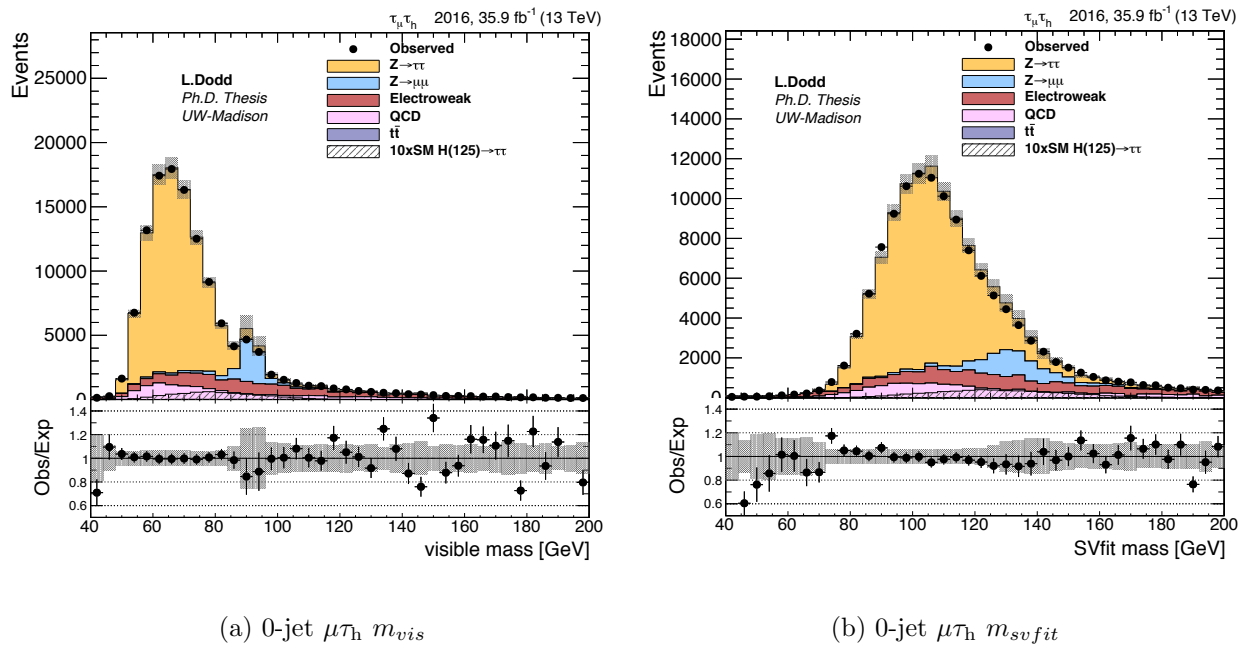
Figure 8.9: Prefit comparison of `SVfit` mass, $m_{\tau\tau}$, and visible mass, m_{vis}, in events with 0 jets. In the above plots the $p_{T,\tau} > 20$ GeV is lowered to 20 GeV and additional corrections on the $\ell \to \tau_{\mathrm{h}}$ are not applied. The difference in the $\tau\tau$ mass resolution between `SVfit` and m_{vis} is visible for the $\ell \to \tau_{\mathrm{h}}$ fakes ($Z \to \mu\mu$).

8.3.7   0-jet signal extraction

The 0-jet category is helpful for constraining systematics and targeting gluon fusion production. The distribution is unrolled in three bins of reconstructed decay τ decay mode: 1-prong, 1 prong + π^0(s), 3-prong. The main advantage of separating the reconstructed

τ decay modes is that the $Z \to \mu\mu$ background with a muon faking a τ_h is absent in the 3-prong decay mode, whereas it is a large contribution in the 1 prong decay mode.

The anti-lepton discriminators were measured by targeting $Z \to \tau\tau$ pairs and the results were decay mode inclusive. However we find a variation in the energy scale and the SF depending on the decay mode of the reconstructed τ_h, seen in Figs. 8.10 and 8.11. This energy scale and normalization is corrected for the $\ell \to \tau_h$ according to Table 8.9.

(a) 0-jet $\mu\tau_h$ m_{vis} 1-prong τ_h (b) 0-jet $\mu\tau_h$ m_{vis} 1-prong $+ \pi^0$(s) τ_h
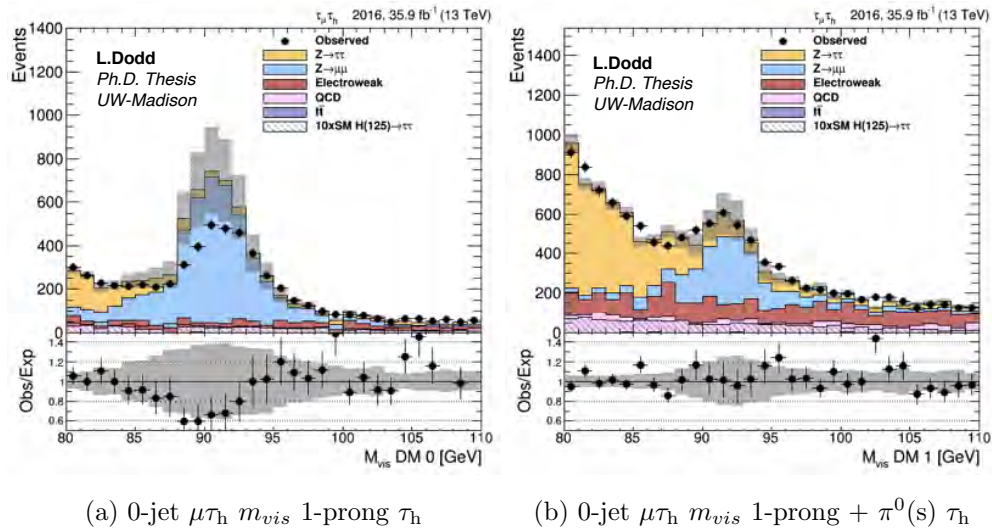
Figure 8.10: Prefit and uncorrected comparison of visible mass, m_{vis}, in the 0-jet category per τ_h decay mode in 1-prong and 1-prong $+ \pi^0$(s) for demonstration of misidentified ℓ-fake-τ_h energy scale. Agreement is improved post-correction.

	1-Prong Decay Mode	1-Prong $+ \pi^0$(s)	3-Prong
$\mu\tau_h$ τ_h ES Correction	$-0.2\% \pm 1.5\%$	$1.5\% \pm 1.5\%$	- $\pm 1.5\%$
$\mu\tau_h$ $\mu \to \tau_h$ fake rate correction	0.74 ($\pm 25\%$)	1.0 ($\pm 25\%$)	-
$e\tau_h$ τ_h ES Correction	$9\% \pm 3\%$	$3\% \pm 0.5\%$	-
$e\tau_h$ $e \to \tau_h$ fake rate correction	0.98 ($\pm 12\%$)	1.49 ($\pm 12\%$)	-

Table 8.9: 0-jet $\mu \to \tau_h$ and e $\to \tau_h$ fake rate corrections.

(a) 0-jet $e\tau_\mathrm{h}$ m_{vis} 1-prong τ_h (b) 0-jet $e\tau_\mathrm{h}$ m_{vis} 1-prong + π^0(s) τ_h
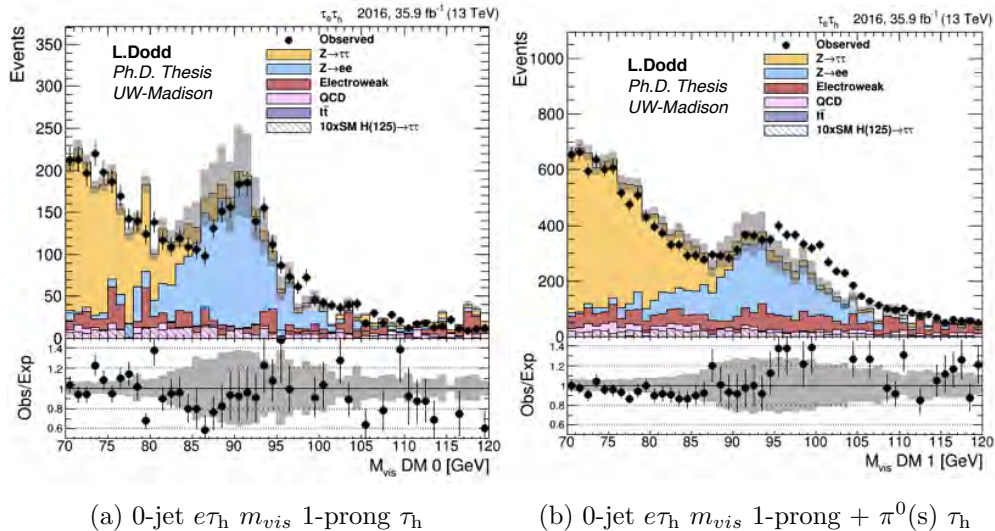
Figure 8.11: Prefit comparison of visible mass in the 0-jet category per τ_h decay mode in 1-prong and 1-prong + π^0(s) for demonstration of misidentified ℓ-fake-τ_h energy scale. Agreement is improved post-correction.

8.4 Other final states

Two other final states were included in the analysis, eμ and $\tau_\mathrm{h}\tau_\mathrm{h}$. These categories are presented briefly in the following sections, as they are input into the final maximum likelihood fit together with eτ_h and $\mu\tau_\mathrm{h}$ channels.

8.4.1 eμ

An e $-\mu$ cross trigger is used online with relatively low thresholds. The eμ channel is dominated by $t\bar{t}$ background. A cut on the D_ζ observable, shown in Eq. 8.4 was first used by CDF and provides a measure on how back-to-back the $E_\mathrm{T}^\mathrm{miss}$ and $\tau_e\tau_\mu$ candidates are; it reduces $t\bar{t}$ by 60%. The observable reconstruction is shown in Fig. 8.12 [168].

$$D_\zeta = P_\zeta - 1.85 P_\zeta^{vis}$$

$$\text{with } P_\zeta = (\vec{P}_{T,1}^{vis} + \vec{P}_{T,2}^{vis} + \vec{P}_T^{\text{mis}})\frac{\vec{\zeta}}{|\vec{\zeta}|} \tag{8.4}$$

$$\text{and } P_\zeta^{vis} = (\vec{P}_{T,1}^{vis} + \vec{P}_{T,2}^{vis})\frac{\vec{\zeta}}{|\vec{\zeta}|}$$
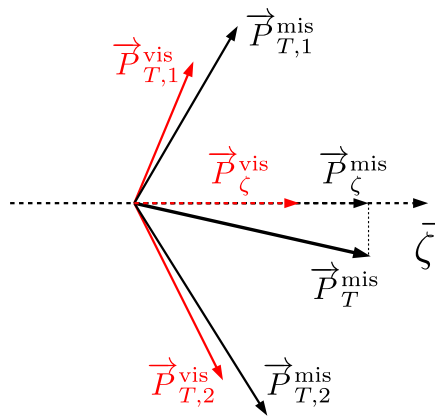
Figure 8.12: Reconstruction of P_ζ

The final unrolled two-dimensional distributions are shown in Figs. 8.15, 8.17, and 8.16.

8.4.2 $\tau_h \tau_h$

The $\tau_h \tau_h$ is improved over the Run-1 result, due to the improved Level-1 trigger for online τ and an MVA based tau identification algorithm that allows for the reduced fake rate and higher efficiency. Kinematic cuts online are 40 GeV for the $\tau_h \tau_h$ channel, with offline kinematic cuts for the leading and sub-leading tau $p_T > 50, 40$ GeV. The $\tau_h \tau_h$ channel has a large background from Z+jets and QCD, with QCD being the largest background for the channel inclusively. The QCD is estimated from an anti-isolated tau region.

The final unrolled two-dimensional distributions are shown in Figs. 8.24, 8.26, and 8.25.

8.5 Fit model

Results are extracted by a simultaneous fit of the three categories (0 jet, VBF, boosted) per final state($e\tau_{\mathrm{h}}$,$\mu\tau_{\mathrm{h}}$,$\tau_{\mathrm{h}}\tau_{\mathrm{h}}$,and $e\mu$). The systematic uncertainties included as nuisance parameters in the fit are detailed in Section 7.6.1. In addition to these signal regions, several control regions are included in the fit.

First, a region to control the $t\bar{t}$ background is introduced in the fit. It is defined with the same baseline selection as used for the $e\mu$ final state, except that the cut on P_ζ is inverted in the boosted category: $P_\zeta > -10$ (boosted). To reduce the small Drell-Yan contamination in this region, the visible invariant mass between the muon and the electron is required to be greater than 90 GeV. The diboson contamination is reduced by requiring at least one jet in the final state. The purity achieved is about 85%. No requirement is made on the number of b-tagged jets [154]. Only one bin is considered for the control region distribution, as shown in Fig. 8.13. No dedicated uncertainty is assigned to the extrapolation from this single control region to the various categories in the different final states. The already considered top p_T reweighting uncertainty covers uncertainties due to missing orders, which could affect the low-to-high P_ζ extrapolation scale factor. The data/MC agreement in the $t\bar{t}$ control region is constant despite changing kinematical cuts to mirror each category. The uncertainty from the fit in the control region is automatically propagated to the $t\bar{t}$ background in the signal region.

Second, control regions for the W+jets background in the $e\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ final states, are also added to the fit model. One control region per category and per final state is included. The control regions are defined in the same way as the signal regions, except events with medium b-tagged jets are vetoed to reduce $t\bar{t}$ in W+jets control region and the m_T cut is modified: $m_T > 80$ GeV. Again, only one bin per distribution is considered. The three categories can be seen in Figs. 8.29 and 8.27. W+jets has already been scaled prefit so that
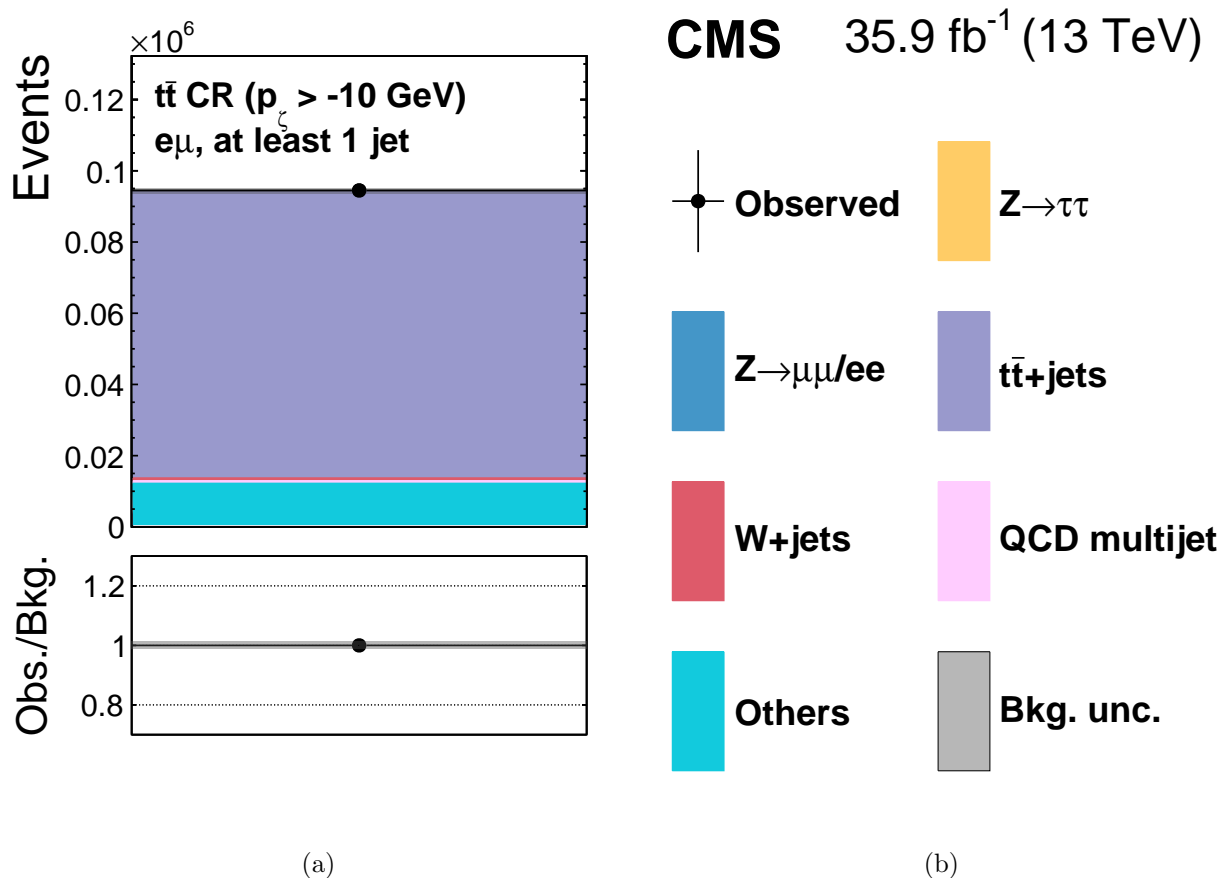
Figure 8.13: $t\bar{t}$ control region is included in simultaneous fit for all channels and categories.

background predictions match data, but it is free to float in the fit.

Third, control regions for the QCD background in the $e\tau_\mathrm{h}$ and $\mu\tau_\mathrm{h}$ final states are also added to the fit model. One control region per category and per final state is included. The control regions are defined in the same way as the signal regions, except that the muon/electron isolation requirements are inverted. The relative isolation of the muon (electron) is required to be between 0.15 and 0.30 (0.10 and 0.30). Four bins per distribution are included to benefit from the different mass distributions of the QCD, W+jets, and Drell-Yan processes. One bin can have more Drell-Yan, while another can have more W. The QCD control regions are shown in Fig. 8.27. The $\ell\tau_\mathrm{h}$ control regions are the only control regions

that have several bins included, this allows for better control of normalization, since each bin has a different processes contributing to it; other background methods are flatter and nothing significant is to be gained by dividing the distributions further.

It can be noted that no control region for the Drell-Yan background is included, but previously described scale factors and uncertainties from the $Z \to \mu\mu$ region have been taken into account to create the background shapes in the signal regions.

In the fit the systematic uncertainties are represented by nuisance parameters that are varied according to their probability density function. A log-normal probability density function is assumed for the nuisance parameters affecting the event yields of the various background contributions, whereas systematic uncertainties that affect the distributions are represented by nuisance parameters whose variation results in a continuous perturbation of the spectrum [169] and which are assumed to have a Gaussian probability density function. Overall, the statistical uncertainty of the collected data is the dominant source of uncertainty for all combined results.

8.6   Systematic uncertainties

General uncertainties applicable to tau pair final states are discussed in Section 7.6. In addition to those uncertainties, additional uncertainties applicable to this analysis are described in this section.

8.6.1   Uncertainties related to object reconstruction and identification

Only in the 0-jet category of the $e\tau_h$ and $\mu\tau_h$ channels, the relative contribution of τ_h in a given reconstructed decay mode is allowed to fluctuate by 3%, chosen based upon observed

decay mode agreement. For example, this migration uncertainty will increase or decrease the expected number of selected 3-prongs taus relative to 1-prong taus. Subtle effects may appear in hadronic tau efficiency depending on decay mode. Adding a decay mode migration uncertainty in other channels and categories results in negligible difference and is therefore not applied outside of $\ell\tau_h$ 0-jet categories.

Using m_{vis} and the reconstructed τ_h decay mode as the two dimensions in the 0-jet category of the $\mu\tau_h$ and eτ_h channels helps in reducing the uncertainty of the e(μ)-faking-τ_h from 12%(25%) after the signal extraction fit: the uncertainty in the rate of muons or electrons misidentified as τ_h becomes on the order of 5%. For events where quark or gluon jets are misidentified as τ_h, an uncertainty of 20% per 100 GeV of τ_h p_T accounts for potential mis-modeling of the jet$\to\tau_h$ misidentification rate as a function of the τ_h p_T. The E_T^{miss} scale uncertainties are among the most impactful nuisances in the analysis, together with the visible τ_h energy scale uncertainties.

The maximum likelihood fit constrains the visible τ_h energy scale uncertainty to about 0.3%, from the original value of 1.2% measured in a $\mu\tau_h$ tag-and-probe. This reduction of the uncertainty is explained by the addition of two other final states with τ_h candidates (eτ_h and $\tau_h\tau_h$), by the higher number of events in the simulations, and by the finer categorization that leads to regions with a larger $Z \to \tau\tau$ purity, than used by tag-and-probe τ_h-energy scale measurement.

Jet energy scale uncertainties depend on the jet p_T and jet η [151, 165, 166]. The 27 independent sources of jet energy scale uncertainty are considered as uncorrelated, and are propagated to the number of reconstructed jets and m_{jj}. These uncertainties therefore affect categorization and the unrolling in the vbf category. These jet uncertainties are treated as shape uncertainties in the fit, however some jet energy scale uncertainties do not have smoothly varying up and down templates due to limited statistical precision in some phase spaces. The 68% up and down jet energy scale templates are passed through a cleaning

algorithm. The following criteria are used to decide if, for a jet energy scale template for a specific background in a given category and channel, we keep the template, demote it from a shape to a log normal uncertainty, or remove the uncertainty altogether:

- If the up and down histograms are identical, the uncertainty is dropped.

- If less then 10% of the bins in a down/up template are shifted from nominal, the uncertainty is made log-normal.

- If there is at least one bin in a down/up template that changes by $> 10\%$ from nominal and the total change in acceptance is $< 2\%$ then the uncertainty is made log-normal.

- If there is one bin in a down/up template that changes by $> 10\%$ from nominal and the total change in acceptance is $> 2\%$: (1) if one bin has a $> 500\%$ variation, then made log-normal (2) otherwise, the shape is kept.

- No down/up template has a bin shifted by more than 10%. Shape is kept.

Additionally, if the total change in acceptance for up/down is less than 0.5%, the uncertainty is removed.

The Stewart-Tackmann [170] uncertainties are applied, increasing the uncertainty on best fit value by a few percent. These uncertainties are related to jet-related variables in Higgs production, so they are applied only to signal processes [170].

8.6.2 Background estimation uncertainties

Included in the simultaneous fit is a $t\bar{t}$-enriched region, obtained in the $e\mu$ channel by inverting the P_ζ cut. The $t\bar{t}$-enriched region is shown in Fig. 8.13 with the inverted cut, $P_\zeta > -10$ GeV. This $t\bar{t}$ region controls the normalization of $t\bar{t}$ in all channels and categories. The resulting normalization uncertainty is about 5%.

The $Z \to \tau\tau$ background normalization is corrected as a function of $p_T(\ell\ell)$, $m_{\ell\ell}$, and m_{jj}, derived from the agreement between data and background prediction in the $Z \to \mu\mu$-enriched region, which was the used to obtain the correction in Section 7.5.5. The extrapolation uncertainties, related to kinematic differences, to go from $Z \to \mu\mu$ to $Z \to \tau\tau$ in different categories, range between 3 and 10% depending on the category. This is found by varying the μ p_T cuts in the $Z \to \mu\mu$. In addition, shape uncertainties related to the uncertainties in the applied corrections are considered. These uncertainties can reach 20% for some ranges of m_{jj} in the vbf category.

The uncertainties on the W + jets event contributions, in the case of the $e\tau_h$ and $\mu\tau_h$ final states, are derived through the inclusion of dedicated control regions in the fit. They account for the statistical limitation of observed data, the effective luminosity of the W+jets simulation sample, and the systematic uncertainties of other processes in the control regions. An uncertainty in the extrapolation from the high-M_T control regions to the low-M_T signal regions is included and ranges from 5 to 10%. It is obtained by comparing the agreement of the M_T distributions of simulated and observed $Z \to \mu\mu$ events where one of the muons is replaced by E_T^{miss} to replicate W+jets events, after multiplying the mass of the reconstructed parent boson in the rest frame by the ratio of the W boson mass to the Z boson mass. In the $e\mu$ and $\tau_h\tau_h$ channels, where the W+jets background is entirely estimated from simulation, the uncertainty on the contribution of this small background is equal to 20% which takes into account jet-τ_h-fake fluctuations between data and simulation.

In the $e\tau_h$ and $\mu\tau_h$ final states, uncertainties from the fit of the control regions with leptons passing relaxed isolation conditions are considered together with a 20% uncertainty that accounts for the extrapolation from the relaxed isolation region to the isolated signal region. In the $\tau_h\tau_h$ final state, an uncertainty that ranges from 3 to 15% accounts for small non-closures in dedicated QCD control regions. It adds up to the uncertainty extracted from fitting the control region with τ_h passing relaxed isolation criteria described previously.

The combined systematic uncertainty on the expected diboson and single-top background contributions is 5%.

8.6.3   Signal prediction uncertainties

The theoretical uncertainties arising from changes in rate and acceptance for the signal processes come from several sources: PDF variations, renormalization and factorization scales, and underlying event and parton shower (UEPS) modeling. The PDF uncertainty for the ggH, VBF, WH, and ZH production modes amounts to 3.2, 2.1, 1.9, and 1.6%, respectively [171]. The renormalization uncertainty for the ggH, VBF, WH, and ZH production modes amounts to 3.9, 0.4, 0.7, and 3.8%, respectively [171].

The uncertainty on the $H \rightarrow \tau\tau$ branching fraction is composed of 3 different independent sources and adds in quadrature to 2.1% [171]. Changes in signal acceptance when comparing default POWHEG 2.0 [106, 107] to HERWIG++ [104], which go up to 7% in the boosted category, account for UEPS uncertainties. The α_s uncertainty is less than 1%.

The Higgs boson p_T distribution in the POWHEG 2.0 simulations is tuned to match the NNLO+NNLL prediction from HRES2.1 [115, 172].

The QCD scale uncertainty is less than 1% for VBF production, however changing the QCD scale can change the m_{jj} and $p_T(H)$ distributions for ggH production. QCD scale uncertainties are added as shape uncertainties for each appropriate category. That is the QCD scale shape uncertainty, a function of m_{jj}, is applied in the boosted category and applied as a function of $p_T(H)$ for the boosted category. The QCD scale uncertainty increases linearly with $p_T(H)$ and m_{jj}.

8.6.4   Other uncertainties

Bin-by-bin uncertainties are considered. These uncertainties are related to the finite number of simulated events, or to the limited number of events in data control regions. If simulations and control regions had infinite statistics, then these uncertainties would not exist. They are uncorrelated across different samples and across bins of a single distribution. The combined effect of these bin uncertainties has a large impact on the precision of the analysis, the largest impact is on the less statistically-populated vbf category. The levels are chosen such that the expected limit sensitivity is unchanged, but limit computation time is reduced.

The systematic uncertainties considered in the analysis are summarized in Fig. 8.14.

8.7   Results

To search for an excess of $\tau\tau$ events from the SM Higgs boson over backgrounds events, a global maximum likelihood fit is made simultaneously of the following distributions:

- 0-jet, boosted, vbf categories for $e\tau_{\mathrm{h}}$ shown respectively in Figs. 8.18, 8.20, and 8.19.

- 0-jet, boosted, vbf categories for $\mu\tau_{\mathrm{h}}$ shown respectively in Figs. 8.21, 8.23, and 8.22.

- 0-jet, boosted, vbf categories for $\tau_{\mathrm{h}}\tau_{\mathrm{h}}$ shown respectively in Figs. 8.24, 8.26, and 8.25.

- 0-jet, boosted, vbf categories for $e\mu$ shown respectively in Figs. 8.15, 8.17, and 8.16.

- 12 Control Regions composed of (1) six regions for QCD multijet shown in Figs. 8.27 and 8.28, (2) W+jets for $\ell\tau_{\mathrm{h}}$ for 0-jet and boosted categories shown in Fig. 8.29, where the boosted constraints are passed to the vbf categories, and (3) one $t\bar{t}$ control region shown in Fig. 8.13 to control normalization across all channels and categories.

To more clearly show the observed signal, the events are plotted based on the decimal logarithm of the signal to signal-plus-background ratio, of each bin of the distributions used

Source of uncertainty	Prefit	Postfit (%)
τ_h energy scale	1.2% in energy scale	0.2–0.3
e energy scale	1–2.5% in energy scale	0.2–0.5
e misidentified as τ_h energy scale	3% in energy scale	0.6–0.8
μ misidentified as τ_h energy scale	1.5% in energy scale	0.3–1.0
Jet energy scale	Dependent upon p_T and η	—
$\vec{p}_T^{\,miss}$ energy scale	Dependent upon p_T and η	—
τ_h ID & isolation	5% per τ_h	3.5
τ_h trigger	5% per τ_h	3
τ_h reconstruction per decay mode	3% migration between decay modes	2
e ID & isolation & trigger	2%	—
μ ID & isolation & trigger	2%	—
e misidentified as τ_h rate	12%	5
μ misidentified as τ_h rate	25%	3–8
Jet misidentified as τ_h rate	20% per 100 GeV τ_h p_T	15
$Z \rightarrow \tau\tau/\ell\ell$ estimation	Normalization: 7–15%	3–15
	Uncertainty in $m_{\ell\ell/\tau\tau}$, $p_T(\ell\ell/\tau\tau)$, and m_{jj} corrections	—
W + jets estimation	Normalization (eμ, $\tau_h\tau_h$): 4–20%	—
	Unc. from CR (eτ_h, $\mu\tau_h$): $\simeq$5–15	—
	Extrap. from high-m_T CR (eτ_h, $\mu\tau_h$): 5–10%	—
QCD multijet estimation	Normalization (eμ): 10–20%	5–20%
	Unc. from CR (eτ_h, $\tau_h\tau_h$, $\mu\tau_h$): $\simeq$5–15%	—
	Extrap. from anti-iso. CR (eτ_h, $\mu\tau_h$): 20%	7–10
	Extrap. from anti-iso. CR ($\tau_h\tau_h$): 3–15%	3–10
Diboson normalization	5%	—
Single top quark normalization	5%	—
$t\bar{t}$ estimation	Normalization from CR: $\simeq$5%	—
	Uncertainty on top quark p_T reweighting	—
Integrated luminosity	2.5%	—
b-tagged jet rejection (eμ)	3.5–5.0%	—
Limited number of events	Statistical uncertainty in individual bins	—
Signal theoretical uncertainty	Up to 20%	—

Figure 8.14: Summary of systematics, including postfit constraints
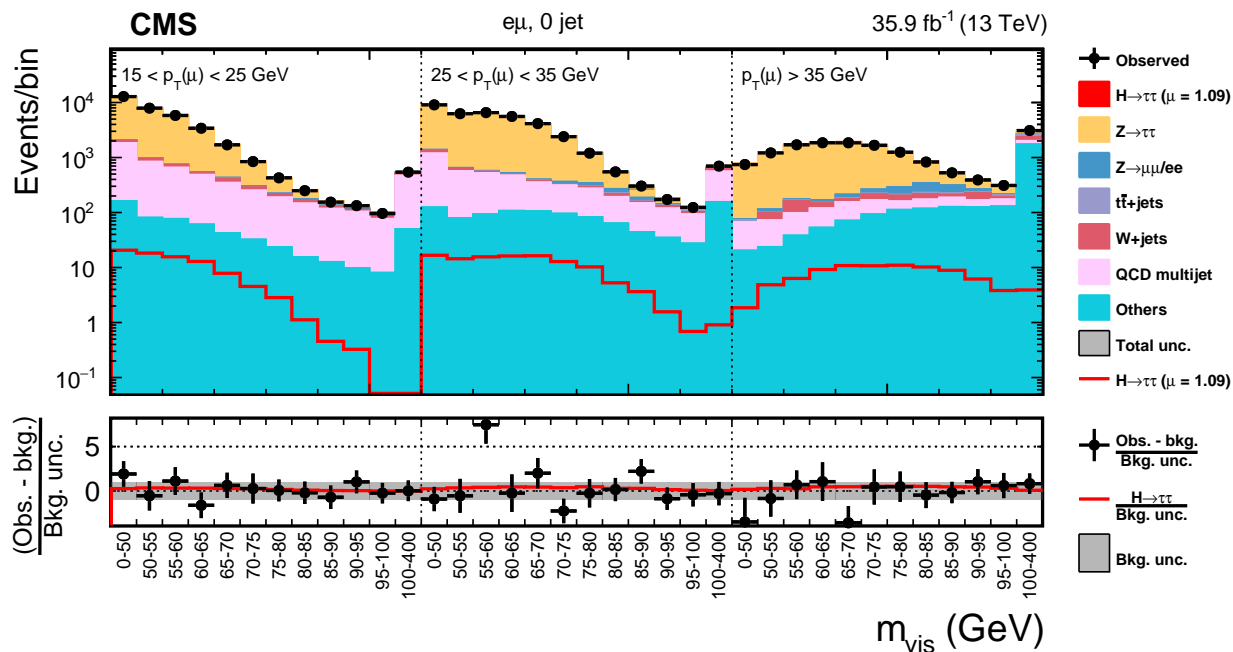
Figure 8.15: Observed and predicted 2D distributions in the 0-jet category of the eµ final state, with 86345 observed events.
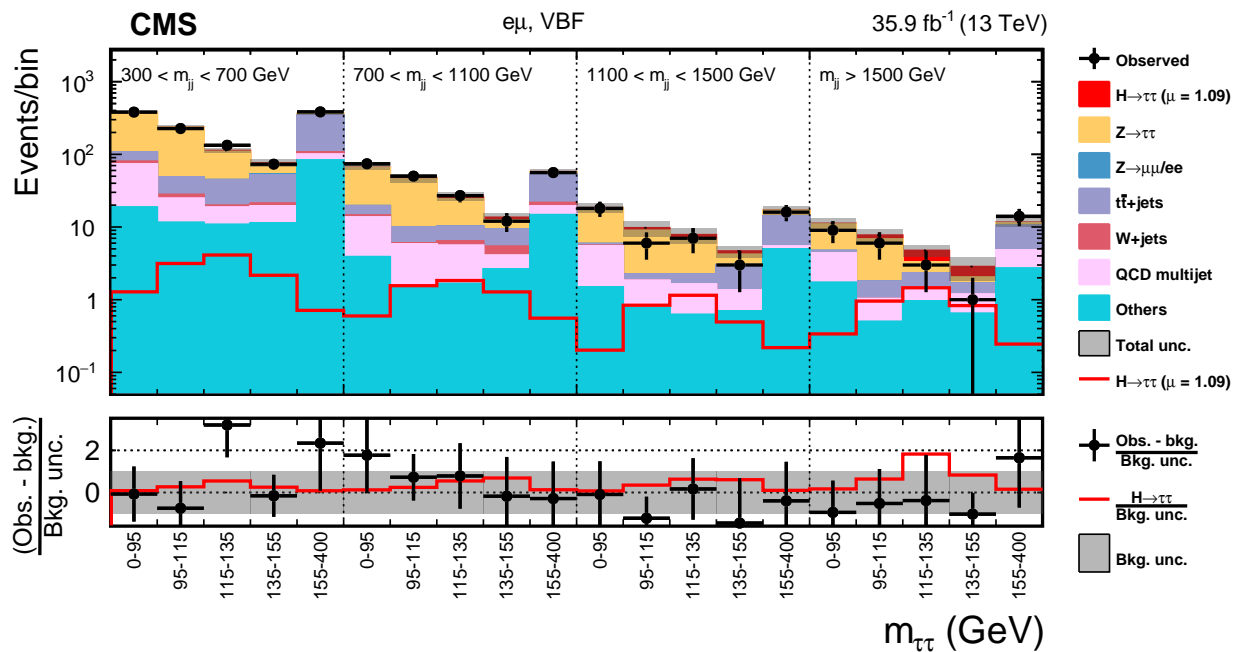
Figure 8.16: Observed and predicted 2D distributions in the vbf category of the eµ final state, with 1497 observed events.
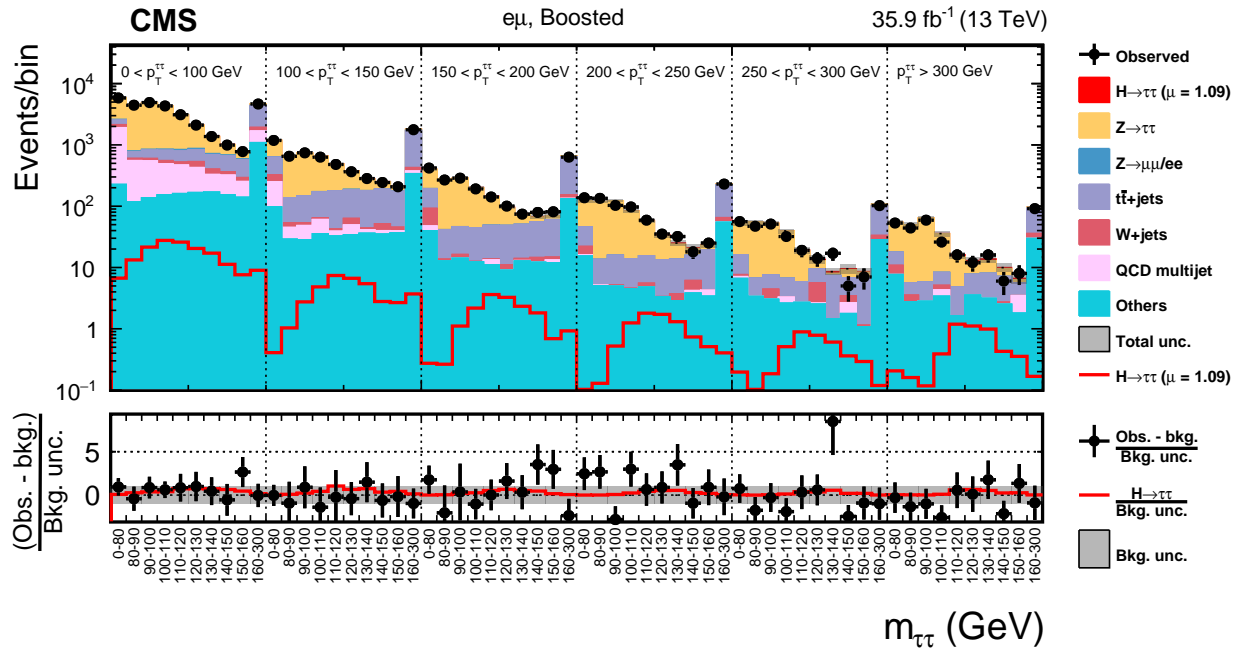
Figure 8.17: Observed and predicted 2D distributions in the boosted category of the eμ final state, with 42777 observed events.
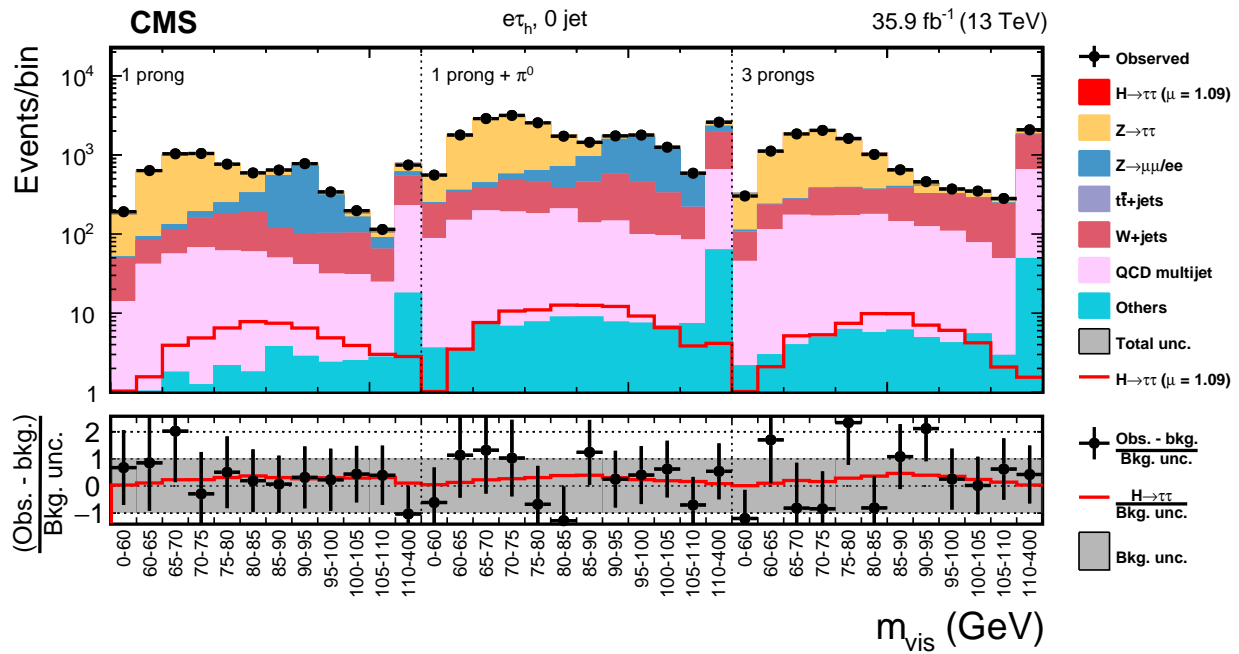
Figure 8.18: Observed and predicted 2D distributions in the 0-jet category of the $e\tau_h$ final state, with 41160 observed events.
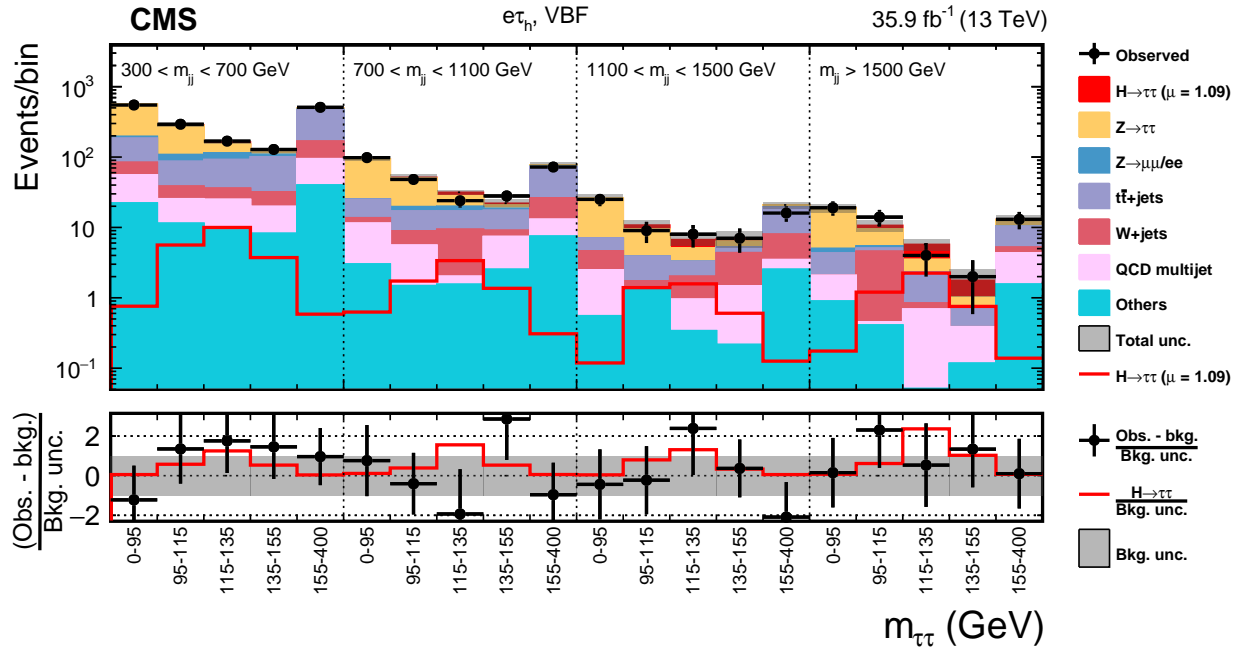
Figure 8.19: Observed and predicted 2D distributions in the vbf category of the eτ_h final state, with 2088 observed events.
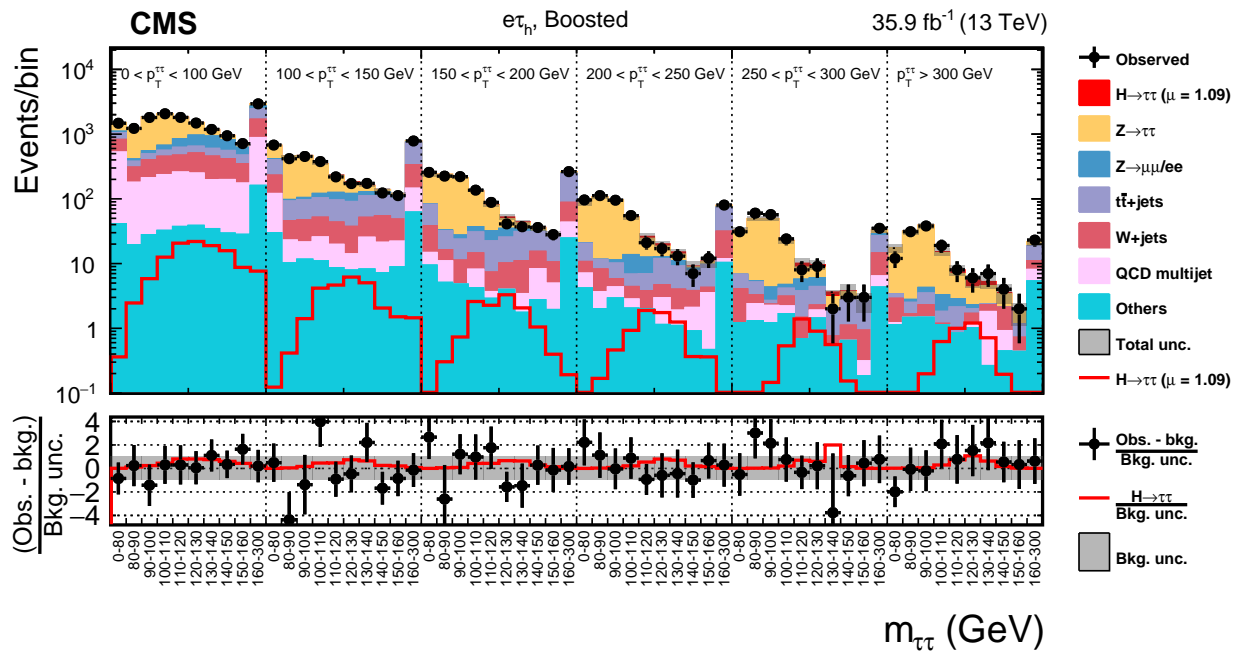
Figure 8.20: Observed and predicted 2D distributions in the boosted category of the eτ_h final state, with 21250 observed events

.
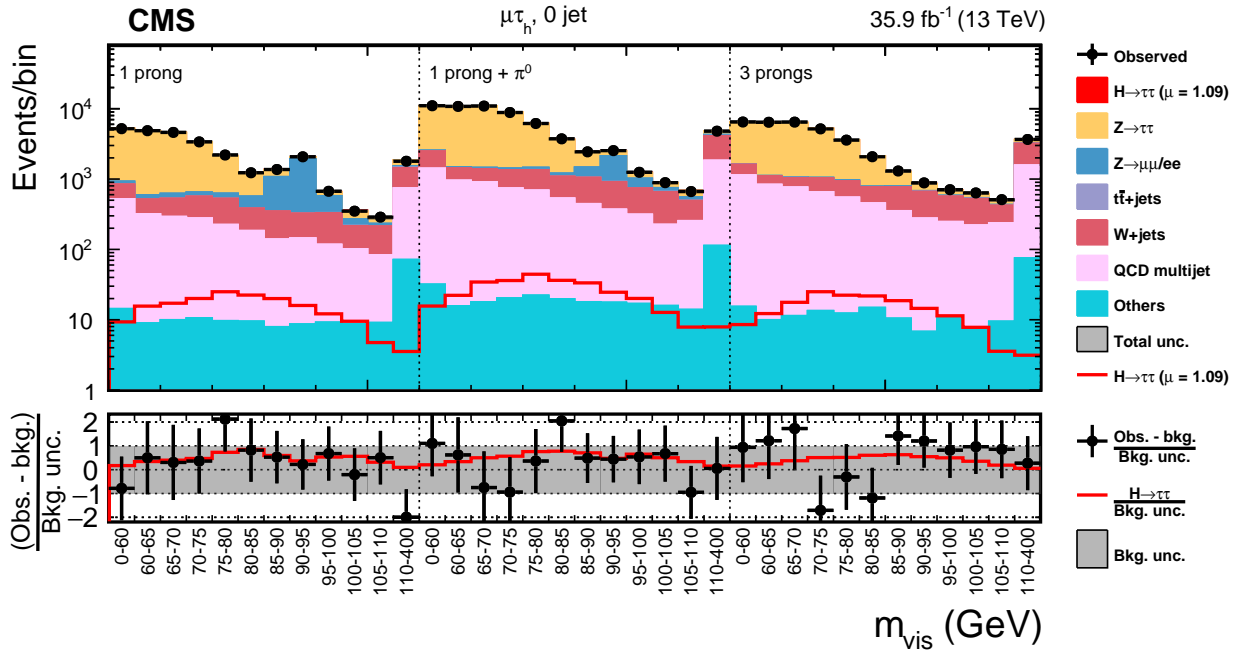
Figure 8.21: Observed and predicted 2D distributions in the 0-jet category of the $\mu\tau_\mathrm{h}$ final state, with 128571 observed events.
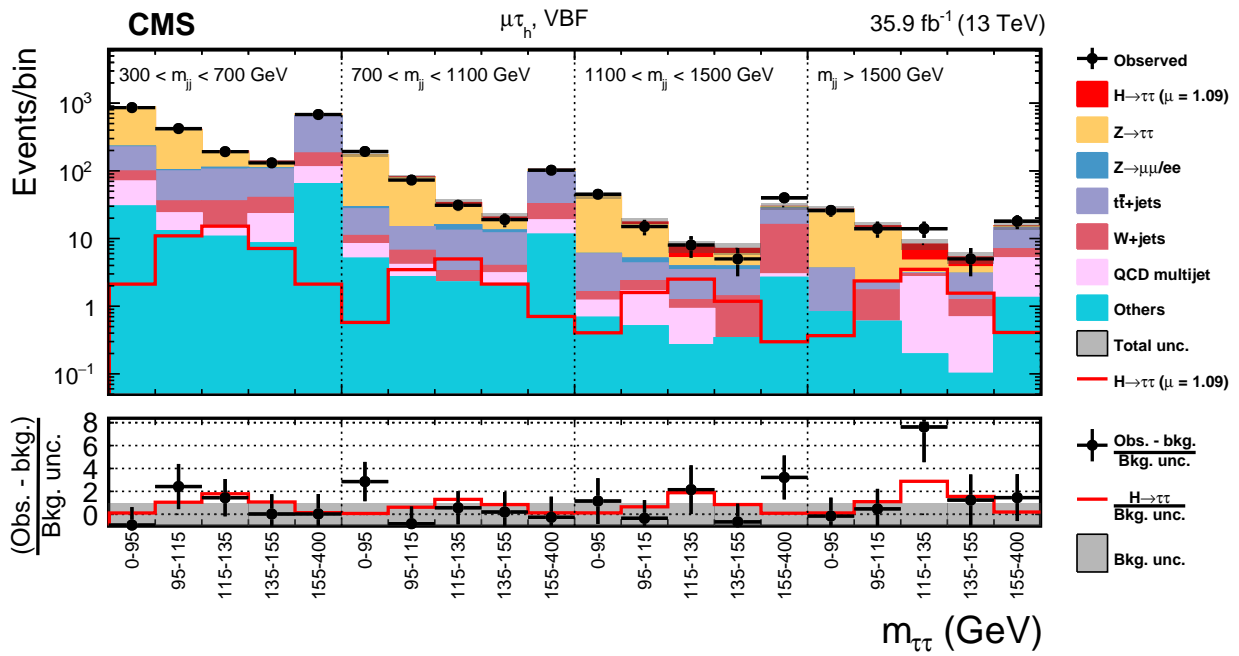
Figure 8.22: Observed and predicted 2D distributions in the vbf category of the $\mu\tau_\mathrm{h}$ final state, with 2927 observed events.
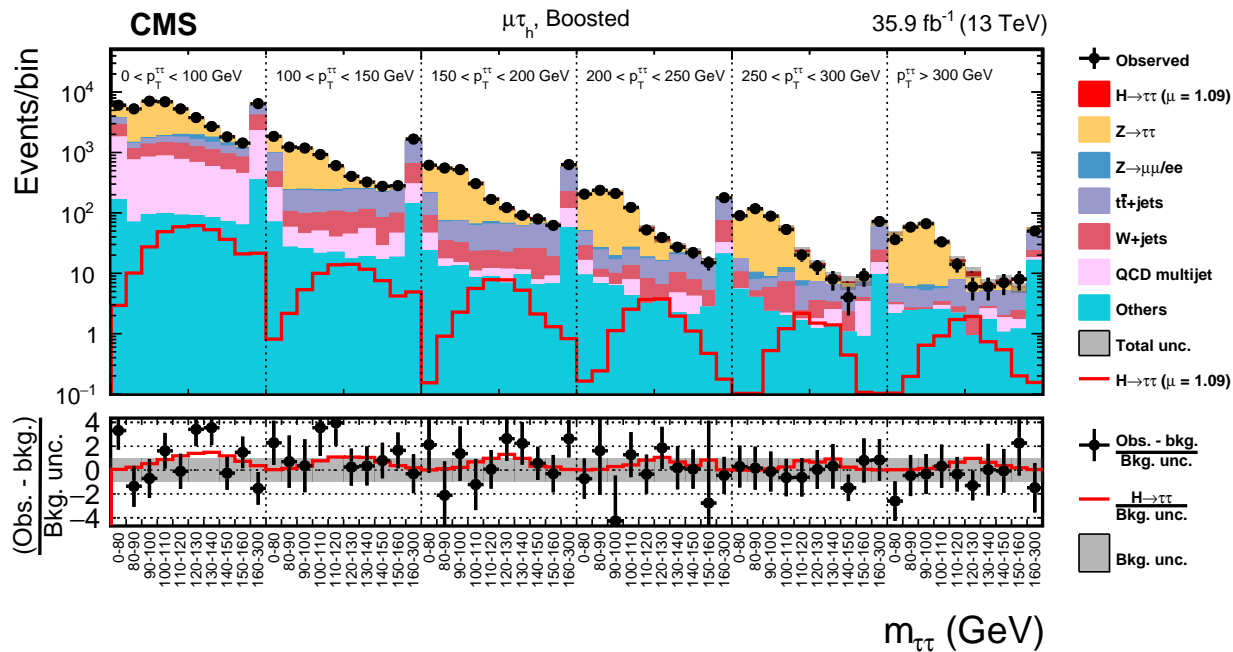
Figure 8.23: Observed and predicted 2D distributions in the boosted category of the $\mu\tau_\mathrm{h}$ final state, with 60127 observed events.
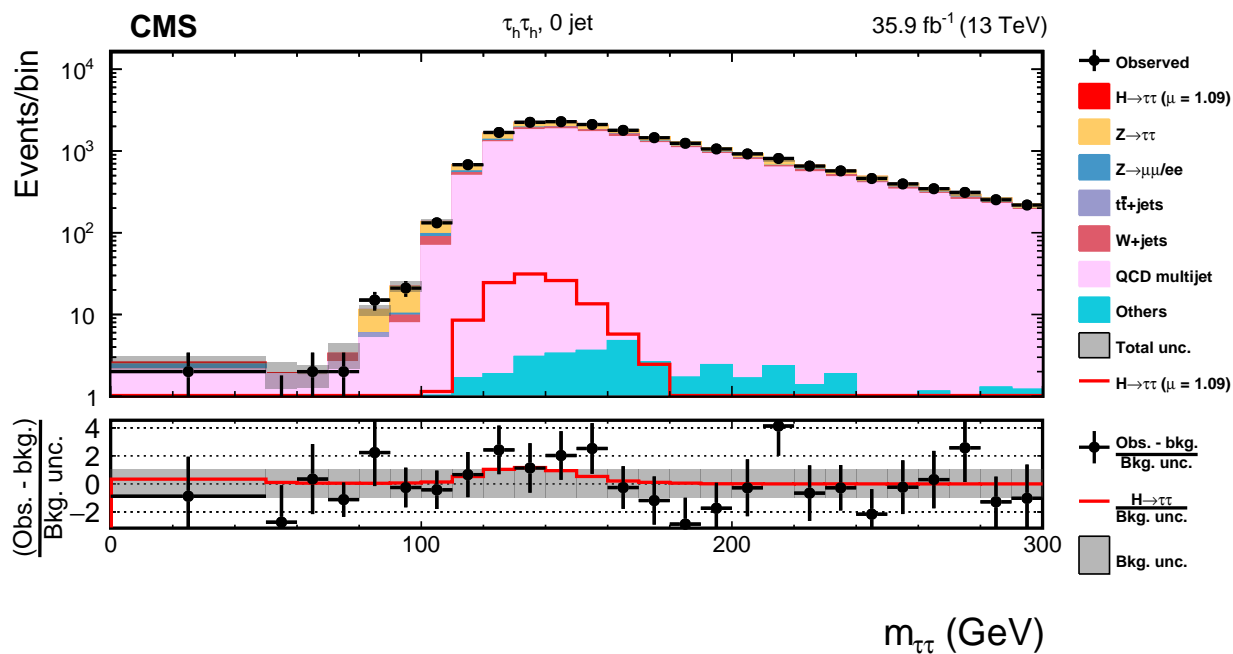
Figure 8.24: Observed and predicted 2D distributions in the 0-jet category of the $\tau_\mathrm{h}\tau_\mathrm{h}$ final state, with 18505 observed events.
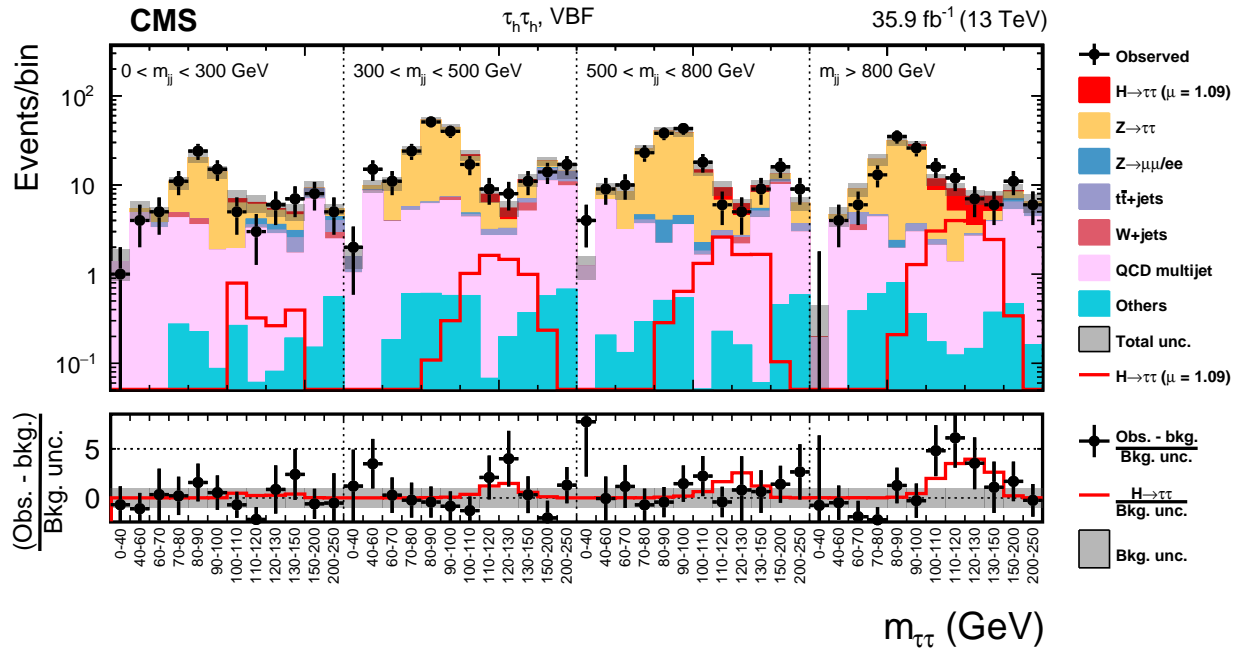
Figure 8.25: Observed and predicted 2D distributions in the vbf category of the $\tau_h\tau_h$ final state, with 645 observed events.
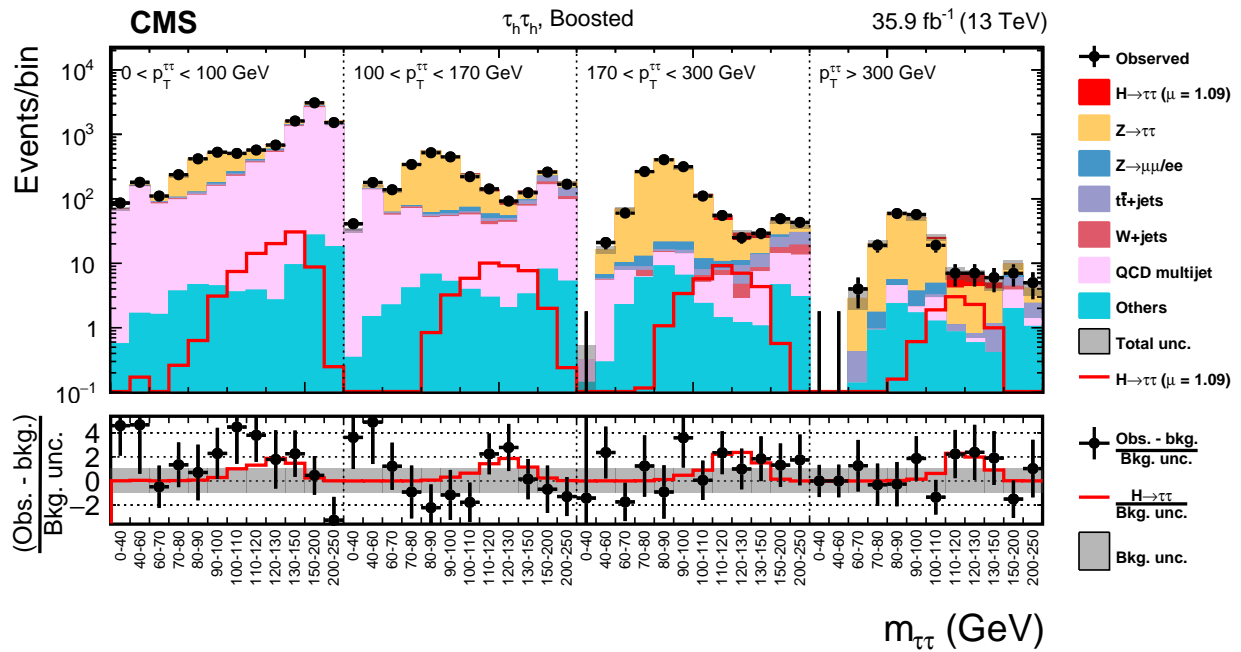
Figure 8.26: Observed and predicted 2D distributions in the boosted category of the $\tau_h\tau_h$ final state, with 13732 observed events.
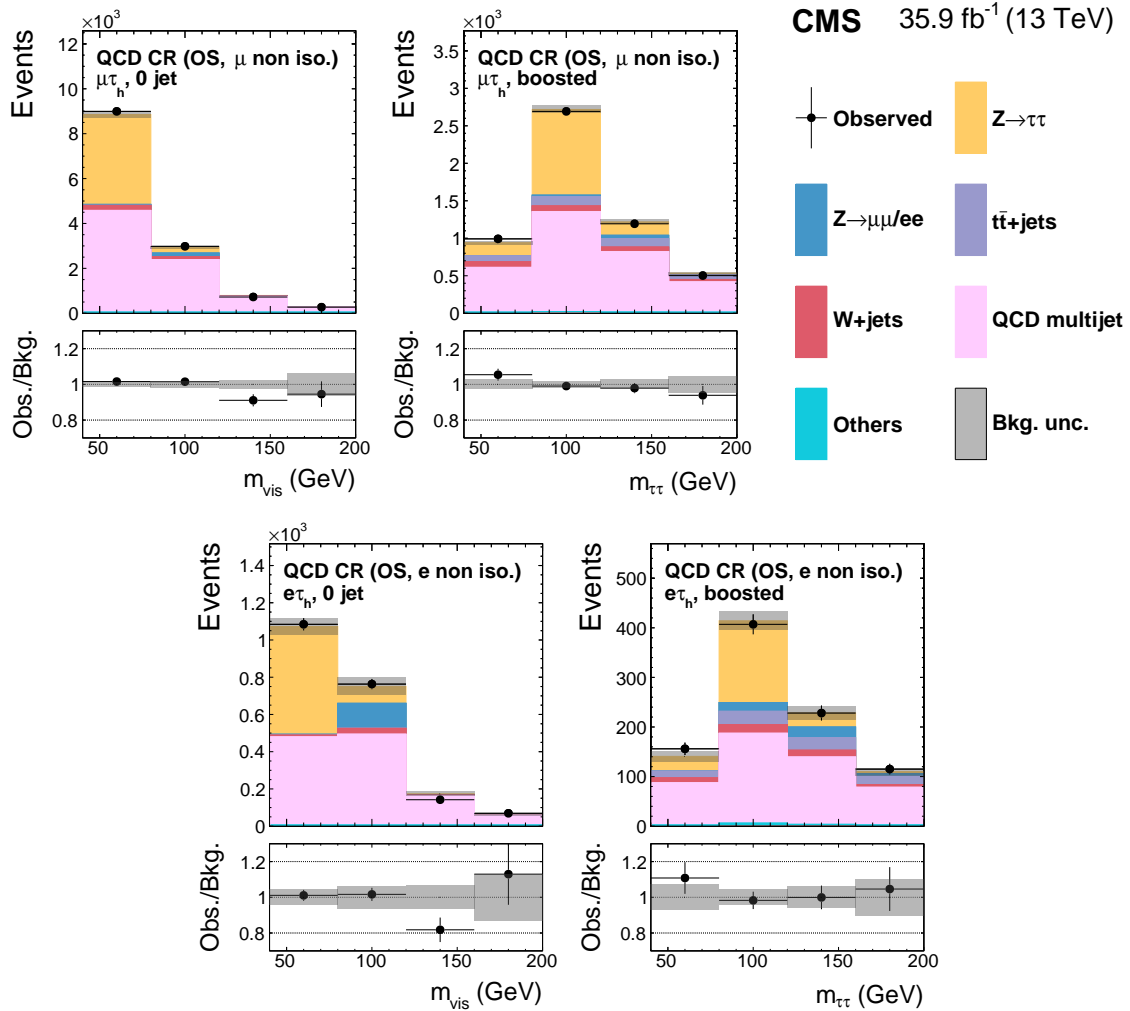
Figure 8.27: All postfit predicted contributions and observed events for $\ell\tau_{\mathrm{h}}$ QCD control regions. This is the same phase space where the OS/SS ratios are measured, so the uncertainty of the ratios is included in the final fit.
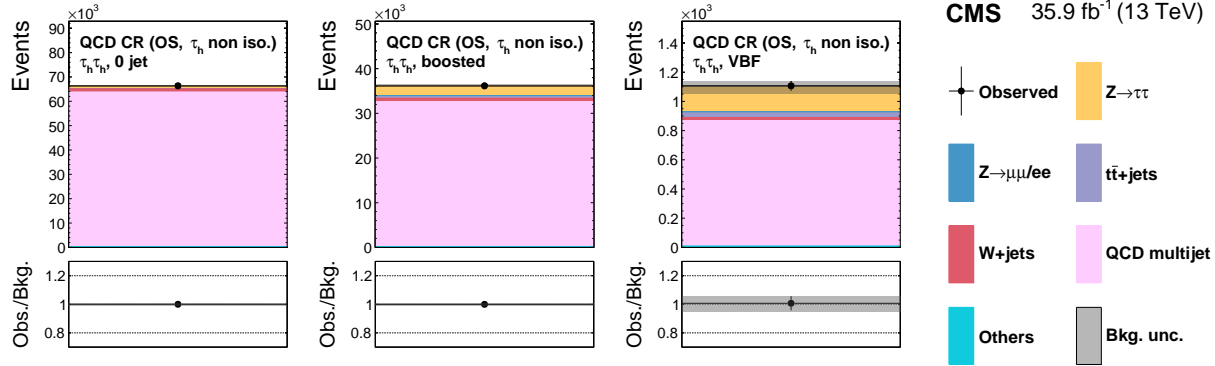
Figure 8.28: All postfit predicted contributions and observed events for $\tau_h \tau_h$ QCD control regions.
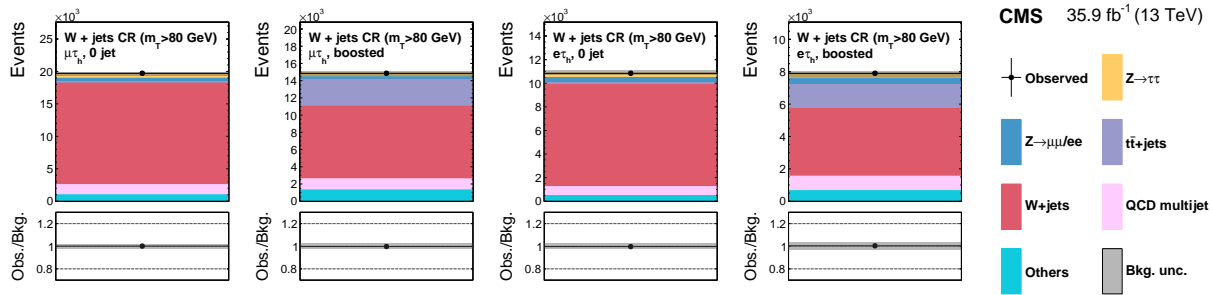
Figure 8.29: All postfit predicted contributions and observed events for $\ell \tau_h$ W+jets control regions.

for signal extraction. Therefore the bins contributing the most to our result are found closest to 0, seen in Fig. 8.31. The excess of observed events over the expected SM background expectations is visible above -1. The expected contributions and observed events with $\log(S/(S+B)) > -0.9$ are printed in Table 8.30. $\tau_h \tau_h$ contributes the most to the highly significant region, with $\mu \tau_h$ being next, followed by $e\tau_h$ and $e\mu$. These events are the most significant in the final result and separately plotted in Fig. 8.31.

By calculating the local p-values with a profile-likelihood ratio test statistics we can quantify the excess of observed events [173, 174]. As shown in Fig. 8.32, the observed significance for a SM Higgs boson with $m_H = 125$ GeV is 4.9 standard deviations, for an

Process	eμ	eτ_h	$\mu\tau_\mathrm{h}$	$\tau_\mathrm{h}\tau_\mathrm{h}$
Z $\to \tau\tau$	5.8 ± 2.2	21.2 ± 3.3	34.6 ± 4.9	89.1 ± 6.9
Z $\to$ ee/$\mu\mu$	0.0 ± 0.0	2.9 ± 0.2	3.7 ± 0.2	5.0 ± 0.2
t$\bar{\mathrm{t}}$+jets	1.9 ± 0.1	10.4 ± 0.3	22.2 ± 1.8	13.9 ± 0.5
W + jets	0.8 ± 0.02	4.0 ± 0.3	6.6 ± 1.3	7.6 ± 0.8
QCD multijet	2.1 ± 0.3	3.3 ± 2.5	5.0 ± 1.3	35.5 ± 2.1
Other backgrounds	1.4 ± 0.1	5.2 ± 0.2	6.1 ± 0.2	7.3 ± 0.2
ggH, H $\to \tau\tau$	0.6 ± 0.1	5.0 ± 0.6	6.0 ± 0.6	27.4 ± 2.1
VBF H $\to \tau\tau$	2.8 ± 0.3	5.1 ± 0.5	12.55 ± 1.0	17.5 ± 1.0
VH, H $\to \tau\tau$	0.0 ± 0.0	0.3 ± 0.0	0.2 ± 0.0	1.3 ± 0.1
Total backgrounds	12.1 ± 2.2	46.5 ± 4.1	77.7 ± 5.5	156.2 ± 7.3
Total signal	3.4 ± 0.4	10.9 ± 0.8	19.2 ± 1.4	48.3 ± 2.6
Observed	11	54	91	207

Figure 8.30: Background and signal expectations, together with the number of observed events, in the signal region bins that have $\log(S/(S+B)) > -0.9$, where S and B are, respectively, the number of expected signal and background events in those bins. The background uncertainty accounts for all sources of background uncertainties, systematic as well as statistical, after the global fit. The contribution from "other backgrounds" includes events from diboson and single-top-quark production, as well as Higgs boson decays to a pair of W bosons.

expected significance of 4.7 standard deviations. The observed p-value is the largest at $m_H = 125$ GeV. Scanning the negative-log likelihood as a function of m_H, and fitting the obtained distribution with a parabola, gives the best-fit mass $\hat{m}_H$ and its 68% CL interval: $\hat{m}_H = 127 \pm 6$ GeV in Fig. 8.33.

Relative contributions of channels and categories to the significance is shown in Fig. 8.34. The addition of the 0-jet category improves the final result around 20% mostly due to systematic constraints, the $S/(S+B)$ weighted distribution is also shown in Fig. 8.34.

The corresponding best-fit value for the signal strength μ is $\hat{\mu} = 1.09^{+0.27}_{-0.26}$ at $m_H = 125$ GeV. The uncertainty on the best-fit signal strength can be decomposed into four
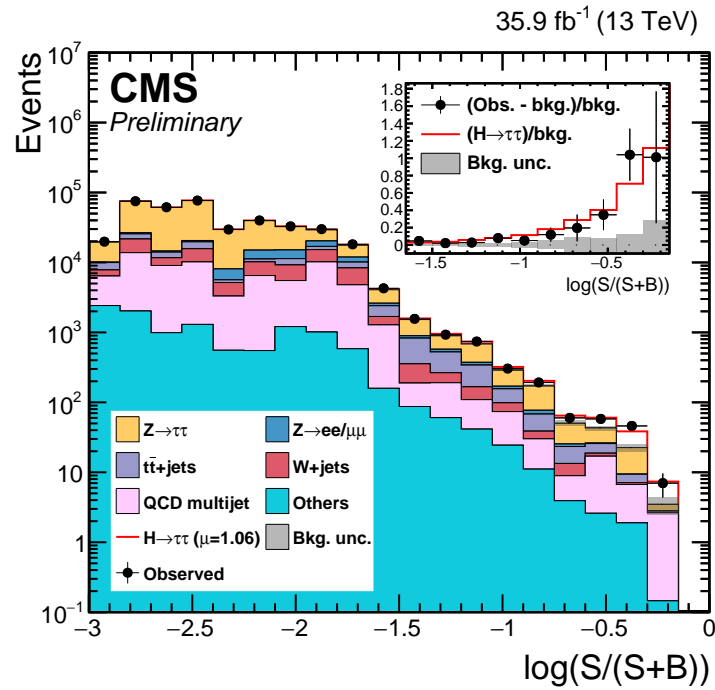
(a) Separated by process
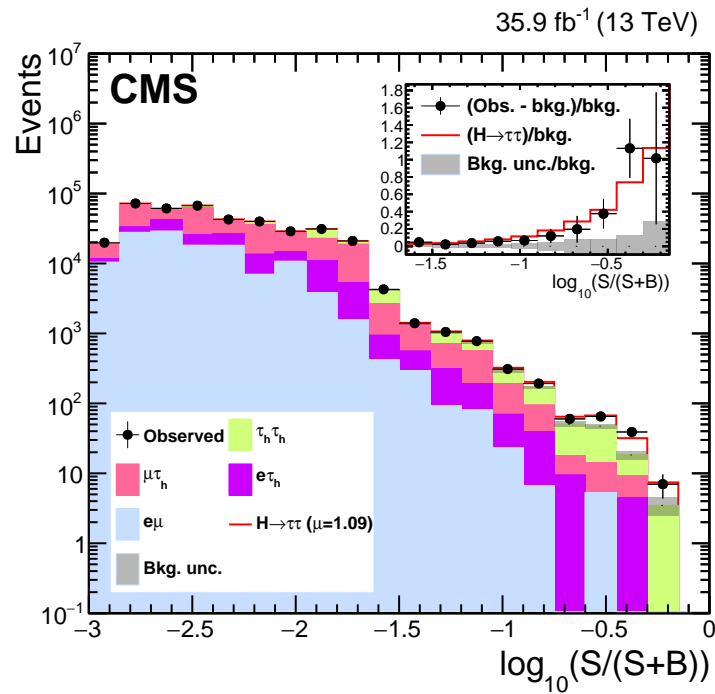
(b) Separated by category

Figure 8.31: Logarithm of $S/(S + B)$ for each bin of the mass distributions used to extract the results, in all signal regions.
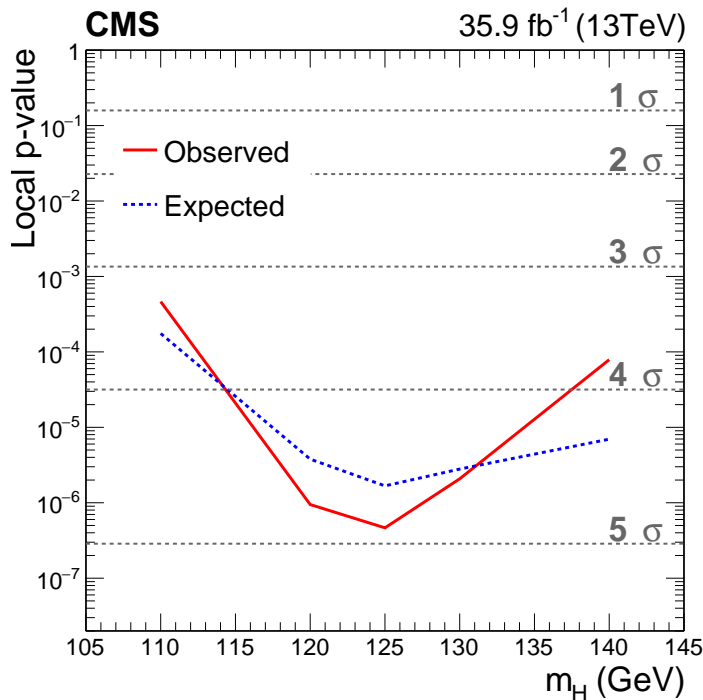
Figure 8.32: Local p-value and significance as a function of the SM Higgs boson mass hypothesis. The observation (black) is compared to the expectation (blue) for a Higgs boson with a mass $m_H = 125$ GeV. The background includes Higgs boson decays to a pair of W bosons, with $m_H = 125$ GeV.

components as shown in Fig. 8.35: theory uncertainties, bin-by-bin statistical uncertainties on the backgrounds, other systematic uncertainties, and statistical uncertainty. In that format, the best-fit signal strength uncertainties are (1) $^{+0.12}_{-0.12}$ (bin-by-bin), related to limited statistical precision of the samples, (2) $^{+0.13}_{-0.12}$ (systematic) account for systematic uncertainties mostly including background estimation uncertainties, and (3) $^{+0.15}_{-0.15}$ (statistical) uncertainties on the limited amount of observed events. The remainder about 12% is mostly due to theoretical uncertainties on Higgs boson production. If we had unlimited simulation and unlimited data, we would have no bin-by-bin and statistical error as well as other reduced systematic uncertainties.

The individual best-fit signal strengths per channel and per category are given in Fig. 8.36;
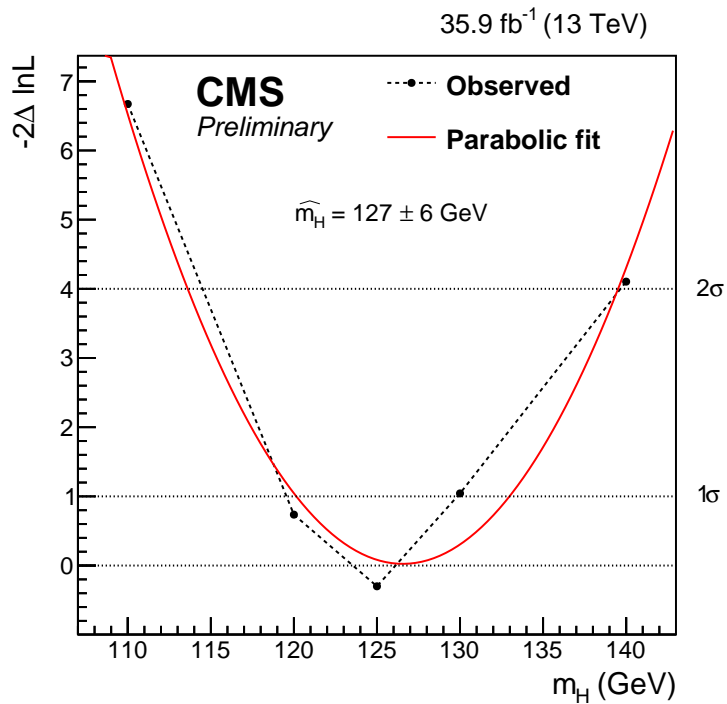
Figure 8.33: Scan of the negative log-likelihood difference, $-2\Delta \ln L$, as a function of m_H. A parabola is fitted to the observed, and the 68% CL interval on the best fit mass is $\hat{m}_H = 127 \pm 6$ GeV.

they demonstrate the consistency by channel and category of the observation with the SM Higgs boson hypothesis. The observed signal strength of the VBF process is $\mu = 1.07 + 0.45 - 0.43$ which is the best precision measurement CMS has done on the VBF production cross section as of the time of publication. [2]. This signal strength and uncertainty on the cross section is used for the 13 TeV qqH cross section for the CMS cross section summary plot shown in Fig. 8.37. Other final states lack statistics for VBF measurements or cannot easily categorize events with extra jets.

A likelihood scan is performed for $m_H = 125$ GeV in the (κ_V, κ_f) parameter space, where κ_V and κ_f quantify, respectively, the ratio between the measured and the SM value for the couplings of the Higgs boson to vector bosons and fermions. For this scan, Higgs boson decays to pairs of W bosons are considered as part of the signal. All nuisance parameters
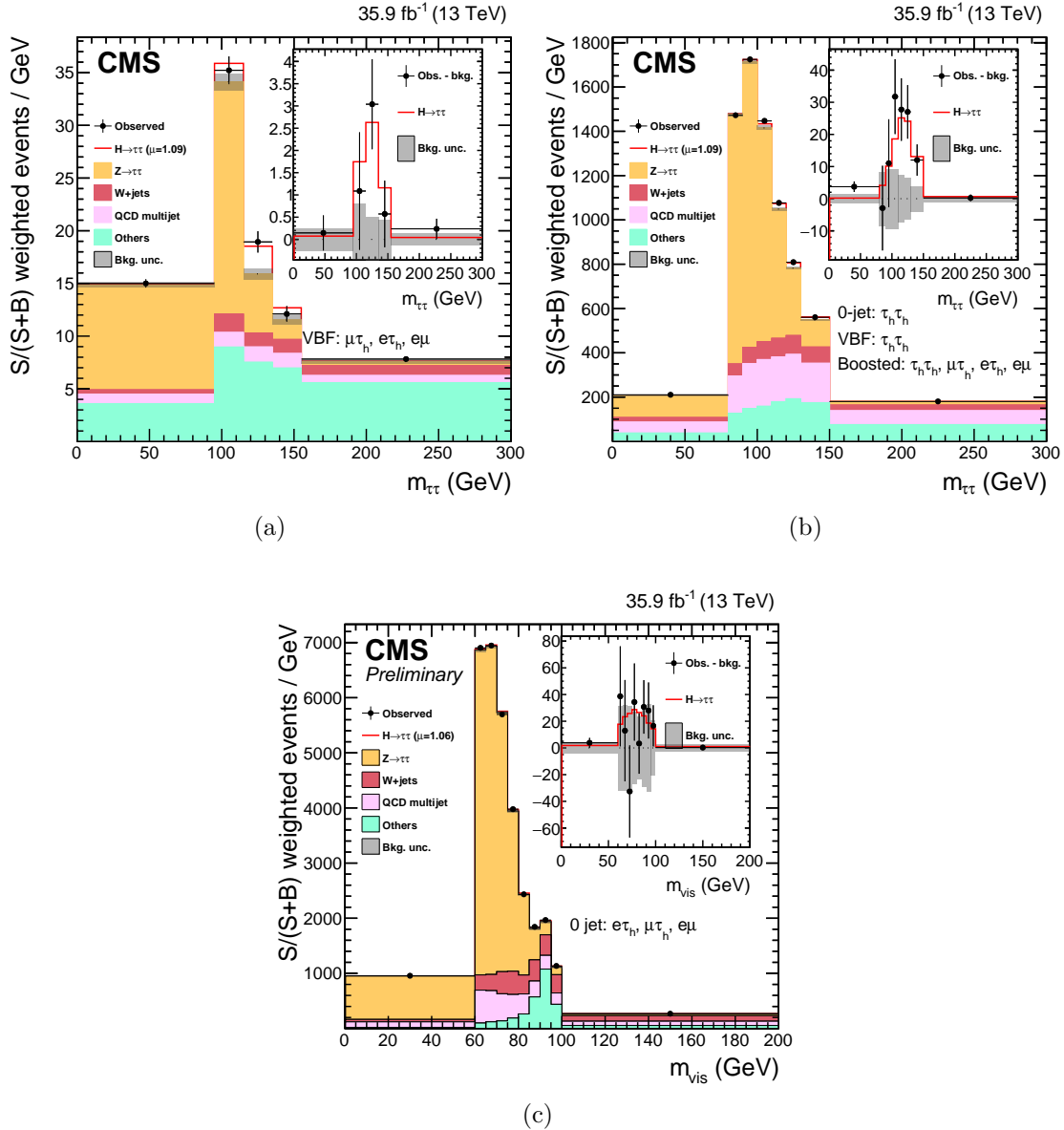
Figure 8.34: Combined observed and predicted $S/(S + B)$ weighted $m_{\tau\tau}$ distributions, for select channels and categories and the m_{vis} distributions for the $\ell\tau_h$ channel. Each slice of the unrolled distribution has a weight; the slices are then folded back together with the more significant slices getting more weight. The others contribution includes $H \to WW$, diboson processes, $t\bar{t}$ and $Z \to \ell\ell$. The peak behavior seen in Fig. 8.34c is due to the $Z \to \ell\ell$ contamination in the 0-jet category in the visible mass spectrum, causing a peak around 90 GeV.
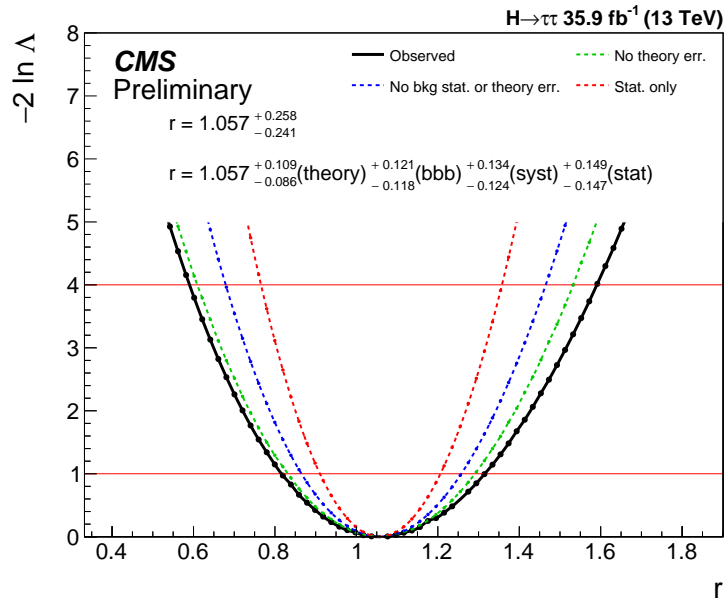
Figure 8.35: Profile likelihood ratio as a function of the signal strength parameter, excluding Stewart-Tackmann uncertainties. The effect of limited statistical precision of simulation is shown in this plot as the red-dashed line. The limited number of simulated events has a larger effect on our final result than all other uncertainties. The theory uncertainties and best fit $\hat{\mu}$ value were updated for publication.

are profiled for each point of the scan. As shown in Fig. 8.38, the observed likelihood contour is consistent with the SM expectation of κ_V and κ_f equal to unity. The κ_V and κ_f scan touches the y-axis with the addition of the Stewart-Tackman uncertainties [170]. If the true value of the κ_V parameter were 0, then the result we measure would still be within 2 sigma. This is due to the contamination of gluon fusion signal events in the vbf category.

The analysis described here was performed in combination with the CMS 8 TeV Run-I analysis, where only the signal strength between the two run periods was kept correlated to provide a more conservative combination. The combination with the Run-1 result produces 5.9 sigma significance expected and observed, with a $\hat{\mu} = 0.98 \pm 18$ with a $m_\text{H} = 125.09$. This result is the most significant direct measurement of the fermionic coupling of the Higgs boson.
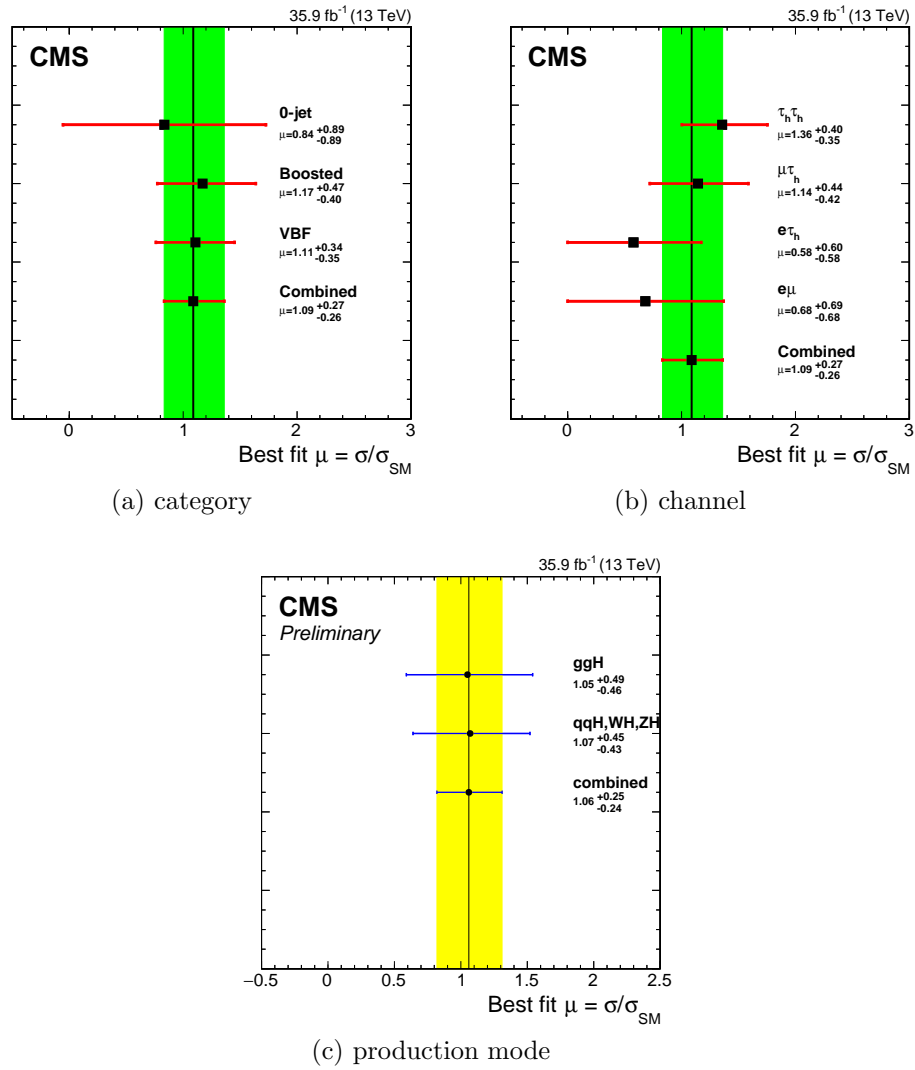
Figure 8.36: Best-fit signal strength for $m_H = 125$ GeV split by (a) category, (b) channel, and (c) production mode. (a) and (b) are included in [1], while (c) was included in preliminary result [2].
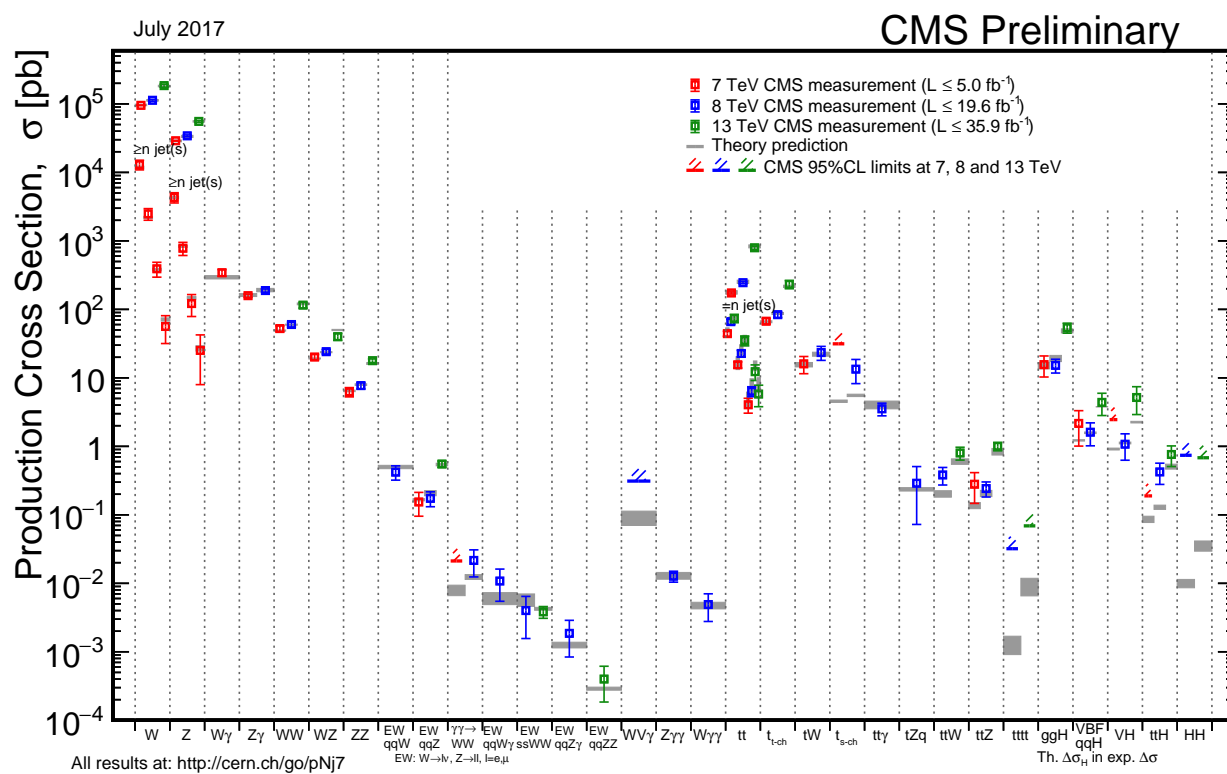
Figure 8.37: CMS cross section summary plot. This result is used for the 13 TeV *qqH* point.
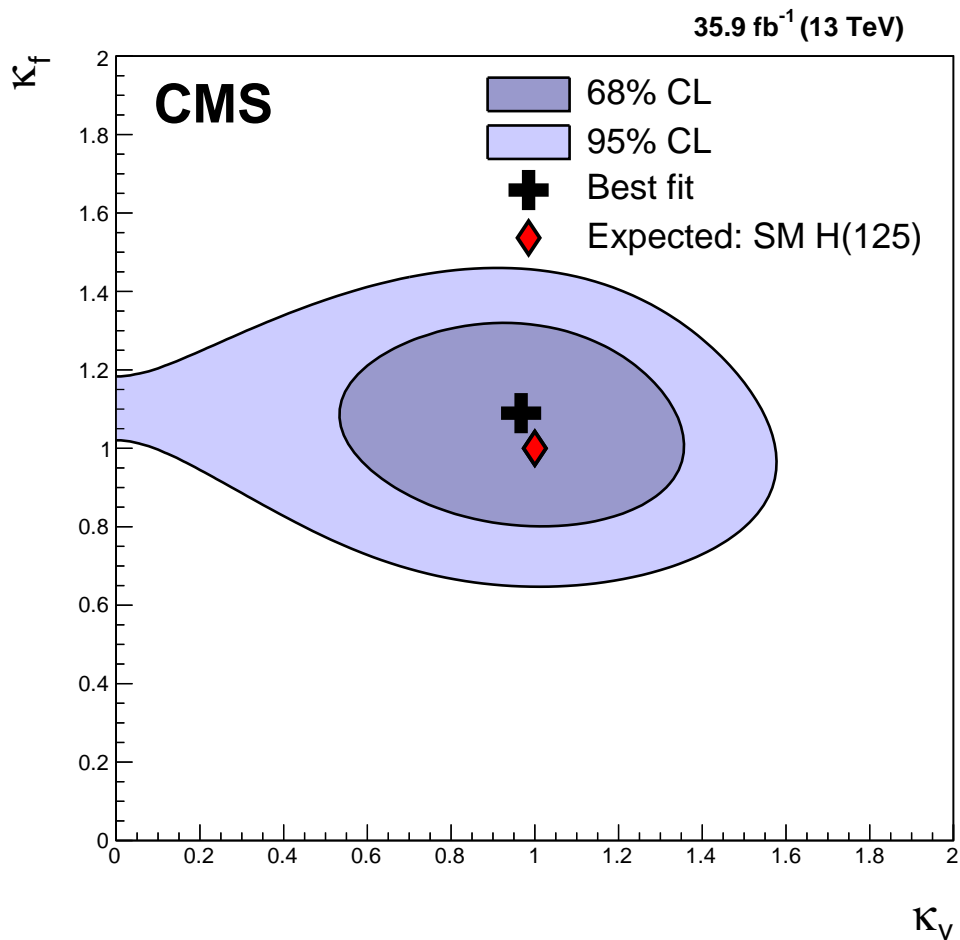
Figure 8.38: Negative log-likelihood difference for κ_V vs κ_f parameter space where $m_{\mathrm{H}} = 125.09$. The H $\rightarrow$ WW process is treated as a signal process unlike elsewhere, since it is an expected decay mode of the SM Higgs boson.

Chapter 9

Search for dark matter produced in association with $H \to \tau\tau$

In this chapter I present a search for SM Higgs boson produced in association with dark matter (DM) pairs. The expected and observed limits are presented on the cross section for different mediator and DM masses for the simplified models discussed in Chapter 3. Many details of searching for a pair of τ particles is discussed in Chapter 7. In this chapter I discuss the $\tau_h\tau_h$ final state where both τ leptons decay hadronically, and the $\ell\tau_h$ final states where one tau decays leptonically to an electron or muon and the other decays hadronically.

9.1 Search for missing transverse energy in association with $H \to \tau\tau$

We consider the three τ lepton pair final states with the highest branching fractions, which are labeled as $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$. The final state $e\mu$ is not included, due to the low branching fraction for that final state. The ee and $\mu\mu$ final states are not included; they have low final-state branching fractions.

9.2 Event selection

In the $e\tau_h$ and $\mu\tau_h$ channels, one of the τ leptons decays leptonically to an electron or a muon (ℓ) and the other τ lepton decays hadronically (τ_h). The two final states together are denoted $\ell\tau_h$. In the third channel, $\tau_h\tau_h$, both τ leptons decay hadronically. Signals from all CMS subsystems are combined by a particle-flow algorithm [138], discussed in Section 6.2, to reconstruct final objects. The individual leptons, hadrons, and photons in each event are used to construct higher level objects: jets, hadronically decaying taus, and missing transverse momentum (E_T^{miss}).

For this analysis, jets (Section 6.6) are reconstructed from hadrons using the anti-k_T algorithm [150, 151]. Taus are identified via the HPS algorithm discussed in Section 6.8. Events with jets from b-quark decays are excluded in order to reduce the background from $t\bar{t}$ events. The combined secondary vertex (CSV) algorithm is used to identify the displaced vertices characteristic of b-quark jets, as described in Section 6.6. A medium level of rejection, corresponding to 70% efficiency and 1% mistag rate, is required [154]. Extra lepton vetos are applied, as described in Section 7.2.2.

Triggers based on the presence of an electron (muon) are used to select events in the $e\tau_h$ ($\mu\tau_h$) channel. In the $\tau_h\tau_h$ channel, the trigger requires the presence of two isolated τ_h

objects. The single electron (muon) triggers with an HLT threshold of 24(25) GeV is used. The double hadronic trigger has an online object threshold of 35 GeV for each tau. The triggers are listed in Table 7.2.

Offline kinematic selections are shown in Table 9.1.

Table 9.1: Selection criteria for the four $\tau\tau$ decay channels. The online HLT threshold requirement for the trigger is given in the first column by the number inside the parentheses.

Final state	Trigger requirement	Lepton selection		
		p_T (GeV)	η	Isolation
$\mu\tau_{\mathrm{h}}$	$\mu(24)$	$p_T{}^{\mu} > 26$ $p_T{}^{\tau_{\mathrm{h}}} > 20$	$|\eta^{\mu}| < 2.4$ $|\eta^{\tau_{\mathrm{h}}}| < 2.3$	$I^{\mu}_{\mathrm{rel}} < 0.15$ Tight MVA τ_{h} ID Tight muon rejection Loose electron rejection
$e\tau_{\mathrm{h}}$	$e(25)$	$p_T{}^{e} > 26$ $p_T{}^{\tau_{\mathrm{h}}} > 20$	$|\eta^{e}| < 2.1$ $|\eta^{\tau_{\mathrm{h}}}| < 2.3$	$I^{e} < 0.1$ Tight MVA τ_{h} ID Loose muon rejection Tight electron rejection
$\tau_{\mathrm{h}}\tau_{\mathrm{h}}$	$\tau_{\mathrm{h}1,2}(35)$	$p_T{}^{\tau_{\mathrm{h}1}} > 55$ $p_T{}^{\tau_{\mathrm{h}2}} > 40$	$|\eta^{\tau_{\mathrm{h}1}}| < 2.1$ $|\eta^{\tau_{\mathrm{h}2}}| < 2.1$	Loose MVA τ_{h} ID Loose muon rejection Loose electron rejection

The electrons and muons in the $e\tau_{\mathrm{h}}$ and $\mu\tau_{\mathrm{h}}$ channels are required to have p_{T} greater than 26 GeV, exceeding the trigger online thresholds for the single electron and muon triggers. Electrons (muons) with $|\eta| < 2.1$ (2.4) are used. An $e\tau_{\mathrm{h}}$ ($\mu\tau_{\mathrm{h}}$) event is required to have an electron (muon) passing the an MVA-trained [148] (cut-based) identification criteria and an isolation requirement of $I_{\mathrm{rel}} < 0.10$ (0.15) , where I_{rel} is defined in Eq. (6.1) with an isolation ΔR solid angle cone of 0.3 (0.4) surrounding the electron (muon). Specifically the electron requirements match those exactly in Section 8.1.4 detailed in Section 6.5; the muon uses the tight cut-based discriminator discussed in Section 6.8.

Hadronic taus in all channels satisfy at least the loose working point of a MVA-based tau isolation measure [155–157], discussed in Section 6.8, plus additional tighter discriminators

for the $\ell\tau_\mathrm{h}$ channels for prompt lepton rejection. All τ_h taus are required to be identified as decaying via one of the three modes recognized with HPS tau reconstruction in Section 6.8. For the $\ell\tau_\mathrm{h}$ channels, the hadronic taus must have $\mathrm{p_T}$ greater than 20 GeV, and $|\eta| < 2.3$. In the $\tau_\mathrm{h}\tau_\mathrm{h}$ channel the minimum tau $\mathrm{p_T}$ requirement is raised to 40 GeV, exceeding the double hadronic tau trigger online threshold of 35 GeV each. The $\tau_\mathrm{h}\tau_\mathrm{h}$ pair is produced with the minimum 40 GeV $\mathrm{p_T}$ criteria. Only after the $\tau_\mathrm{h}\tau_\mathrm{h}$ pair candidate is identified, the leading τ_h cut is further raised to $\mathrm{p_T}(\tau_\mathrm{h})_1 > 55$ GeV.

Table 9.2: Number of observed events in selection process.

Process	$\mu\tau_\mathrm{h}$ Events	$e\tau_\mathrm{h}$ Events	$\tau_\mathrm{h}\tau_\mathrm{h}$ Events
Dataset/Trigger	564814720	942127424	379366752
Extra lepton veto and OS tau pair	6600148	5751308	4246205
B-tagged jet veto	5901391	5474527	4037097
Full identification criteria	535026	258505	175422

Table 9.2 shows the observed number of events after each stage of the selection. The number of events, starting with passing the trigger, collected is shown in the first line of Table 9.2, where the minimum $\tau_\mathrm{h}\tau_\mathrm{h}$ pair kinematics are $\mathrm{p_{T}}_1^{\tau_h} > 40$ and $\mathrm{p_{T}}_2^{\tau_h} > 40$. The second line of the table requires the minimum identification of loose tau pairs: electrons and muons have relative isolation less than 0.3, taus are reconstructed with a decay mode and pass the loose working point of the tau MVA discriminator [155–157], and the tau candidates are opposite sign. The third line of the table is the step when events are rejected due to the presence of a CSV medium working point b-tagged jet [154]. The last line in the table is the application of: the anti-lepton discriminators, discussed in Section 6.8, the tight working point for the hadronic tau identification for $\ell\tau_\mathrm{h}$, the relative isolation requirements less than 0.1(0.15) for e (μ) and the τ_h1 in the $\tau_\mathrm{h}\tau_\mathrm{h}$ channel is raised to 55 GeV from 40 GeV. The last line in Table9.2 may be referred to as the "inclusive" selection.

9.3 Background estimation

The background contributions from W + jets and the QCD processes are estimated using data-driven approaches. The procedure to estimate these processes relies on control regions, which are included in the maximum likelihood fit to extract the results, detailed in Sections 9.4 and 9.5. Simulation is used to estimate the backgrounds from $t\bar{t}$, Z+jets, 125 GeV Higgs boson processes, single top and multi-boson processes. The largest diboson contributions are $ZZ \to 2L2\nu$ and WW $\to 2L2\nu$ for all channels and the $WZ \to 3L1\nu$ process is present for the $\tau_h\tau_h$ channel. The $t\bar{t}$ process is checked in a b-tagged jet enriched region. These processes are all produced at NLO QCD precision, with additional weights applied, as discussed in Table 9.3, for higher-order corrections for diboson, and to reproduce observable top p_T distribution in data. Agreement for these large backgrounds is checked, and found to be appropriately match observation within the applied uncertainties.

The weights and other simulation corrections discussed in Section 7 are used. Table 9.3 summarizes the relevant weights used. Recoil corrections discussed in Section 7.5.7 are not applied, as those corrections were produced for parton-binned W+jets simulation. In this analysis, an H_T binned W+jets simulation is used with other generator p_T event weights. A τ-pair pt cut applied in Section 9.4 to further reduce dependence on recoil corrections.

The distribution of the W + jets background is estimated from simulation by requiring the same selection as the signal region, but the MVA isolation of the τ candidates is relaxed from the Tight MVA WP to the Loose MVA WP, as detailed in Section 8.3.4, to increase the statistical precision of the distribution. To constrain the normalization of the W + jets background, a control region (CR) enriched in W + jets events is constructed by inverting the isolation criteria on the τ_h candidates while keeping a loose MVA isolation requirement. The CR is included in the maximum likelihood fit and the resulting normalization found for the W + jets contribution in the control region similarly changes the normalization of the

Table 9.3: Applicable event weights and scale factors

Weight	Comment
cross section	ensures expected contributions match recorded 35870 pb^{-1}
AMCATNLO negative weights	results in reduction of effective luminosity of simulation described in Section 7.5.8
Drell-Yan	reproduces observed $Z - p_T/m_Z$ distributions described in Section 7.5.5
W+jets	NLO EWK correction applied as function of W boson generator p_T [175–177]
ZZ	NNLO QCD corrections and NLO EWK correction for $2\ell2\nu$ processes applied as a function of generator leading boson p_T [178]
WW	NLO EWK corrections for $2\ell2\nu$ processes [179] applied as function of generator E_T^{miss}
$t\bar{t}$ reweighting	observed $t\bar{t}$ p_T spectrum is softer than predicted described in Section 7.5.6
τ_h ID and anti-ℓ-τ_h ID	corrects for differences in hadronic τ reconstruction described in Section 7.5.2 and in Section 7.5.3
Trigger	Event weight is applied as a function of reconstructed object p_T discussed in Section 7.5.4, relevant for $\ell\tau_h$ channel. The $\tau_h\tau_h$ channel uses the same trigger event weight as in Ref [1], and is additionally applied as a function of reconstructed p_T *and* reconstructed τ_h decay mode.
Pile-up weights	discussed in Section 7.5.1 and applied to improve agreement of pile-up distribution between data and simulation
CSV b-tagged jet [154] weight	Event weight applied to correct differences between data and simulation [154] for medium working point discussed in Section 6.6.

W + jets contribution in the signal region.

To estimate the QCD background, a sideband in data is obtained by requiring the τ candidates to have the same sign. No significant amount of signal and of background with opposite-sign τ_h is expected in this region because the τ_h charge misidentification is at the percent level, and the charge misidentification for electrons and muons is even smaller. All simulated backgrounds are subtracted from observed events in the sideband, and the remaining distribution is classified as QCD background.

The contribution of QCD events with opposite-sign τ candidates in the signal region is obtained by multiplying the QCD background obtained in the same-sign region by a scaling factor. The scaling factor, close to unity and shown in Table 9.4, is measured from events with anti-isolated τ candidates and consist mostly of low $E_T^{miss} < 50$ GeV, which do not overlap with events selected in the signal region. For the $\ell\tau_h$ channels, the factors are 1.11+/-15% and 1.22+/-20%, for $\mu\tau_h$ and eτ_h respectively. The basis for the extrapolation to high E_T^{miss} ($E_T^{miss} \gtrsim 100$ GeV) and boosted pair region are in Table 9.4. The regions used for measurement of these factors are shown in Fig. 9.1. The scaling factor for the $\tau_h\tau_h$ channel is taken as unity, with an inflated uncertainty from 2% used in the public result for Chapter 8 to 20%, for extrapolation. After the signal extraction cuts for the $\tau_h\tau_h$ channel, the QCD contribution is nearly negligible. To increase the statistical precision of the QCD distribution, the isolation is relaxed for the $\ell\tau_h$ channels (see Section 8.3.4), while conserving obtained normalization as detailed above. The same-sign control region is added to the maximum likelihood fit to constrain the normalization of the QCD background in the signal region, for all channels. Any mis-modeling of W+jets would show up as an additional non-closure in the QCD control region, so all the regions are included in the simultaneous fit.

[e$\tau_{\rm h}$]

(a) $\mu\tau_{\rm h}$
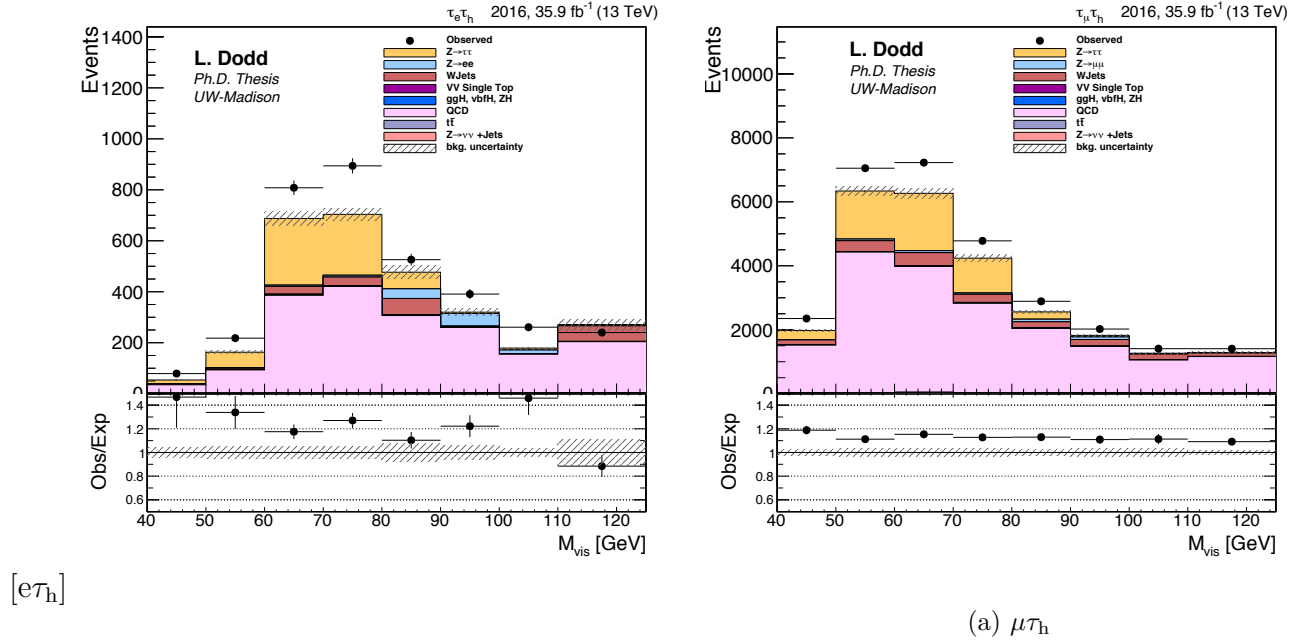
Figure 9.1: OS and anti-isolated lepton phase space using dedicated anti-isolated trigger and identification scale factors. The QCD contribution is taken from the same-sign region. The non-closure in the region allows calculation of the OS/SS factor for the $\mu\tau_{\rm h}$ and e$\tau_{\rm h}$ final states.

Table 9.4: e$\tau_{\rm h}$ and $\mu\tau_{\rm h}$ SS→OS factors. The "stat-only" method only includes statistical uncertainty and is is the scale factor calculated when assuming all disagreements come from QCD normalization. For example, if one observes 20 events, with 10 expected QCD events and 5 expected events from other backgrounds, the "stat-only" method would return approximately 1.5 ± 0.60.

e$\tau_{\rm h}$	"stat-only" stat uncert	Max. likelihood fit	incl.	total
inclusive	1.26 +/- 0.16	1.22 -0.11/+0.11	1.22+/-10%	1.22+/- 20%
$E_{\rm T}^{\rm miss}$ >20 GeV	1.31 +/- 0.20	1.34 -0.13/+0.13		-
$E_{\rm T}^{\rm miss}$ >40 GeV	1.32 +/- 0.67	1.49 -0.18/+0.19		add 5%non-closure
p$_{\rm T}(H)$ >30 GeV	1.38 +/- 0.40	1.46 -0.16/+0.17		add 5%non-closure
$\mu\tau_{\rm h}$	"stat-only" stat uncert	Max. likelihood fit	incl.	total
inclusive	1.18+/- 0.09	1.11 -0.07/+0.08	1.11+/- 10%	1.11+/-15%
$E_{\rm T}^{\rm miss}$ >20 GeV	1.16 +/- 0.11	1.11 -0.08/+0.08		-
$E_{\rm T}^{\rm miss}$ >40 GeV	1.14 +/- 0.20	1.23 -0.09/+0.09		add 5% met uncert
p$_{\rm T}(H)$ >30 GeV	1.16 +/- 0.34	1.15 -0.10/+0.11		-

9.4 Signal extraction

The event sample is split into three exclusive channels $\mu\tau_\mathrm{h}$, $e\tau_\mathrm{h}$, and $\tau_\mathrm{h}\tau_\mathrm{h}$. In each category the p_T of the di-τ system, visible $m_{\tau\tau}$, and $E_\mathrm{T}^\mathrm{miss}$ are chosen to extract the signal. The final selection criteria are shown in Table 9.5, and were chosen as the maximal values of $S/\sqrt{S+B}$. The optimization is performed in steps of 5 GeV first on $E_\mathrm{T}^\mathrm{miss}$, then on $p_{\mathrm{T}\tau\tau}$. Additionally, visible $m_{\tau\tau}$ is required to be less than 125 GeV to ensure that the SM Higgs boson is the di-τ resonance we seek and $\Delta R_{\tau\tau} < 2.0$. The $\tau\tau$ decays of the SM Higgs boson always have a reconstructed invariant visible mass < 125 GeV since energy is lost to neutrinos. Table 9.6 shows the observed number of events in each category after each signal selection.

Table 9.5: Signal extraction additional criteria

variables	criteria
$E_\mathrm{T}^\mathrm{miss}$	> 105GeV
$p_\mathrm{T}(H)$	> 65GeV
$m_{\tau\tau}$	< 125GeV
$\Delta R_{\tau\tau}$	< 2.0

The limit is extracted from the total transverse mass ($M_{T,tot}$) distribution where $M_{T,tot}$ is defined in Eq. (9.1). Other limit extraction techniques and observables were tested, and this variable performed the best compared with other extraction techniques. For instance, a 2D extraction shown in Fig. 9.2 had an expected upper limit comparable but 10% worse for the most sensitive mass point. The same unrolling procedure done in Chapter 8 was attempted for this analysis. The simpler $M_{T,tot}$ extraction is chosen for this thesis.

$$M_{T,tot} = \sqrt{(E_T^{\tau_1} + E_T^{\tau_2} + E_T^{miss})^2 - (p_x^{\tau_1} + p_x^{\tau_2} + p_x^{miss})^2 - (p_y^{\tau_1} + p_y^{\tau_2} + p_y^{miss})^2} \qquad (9.1)$$

The distributions used to extract the limits are shown in Fig. 9.5.
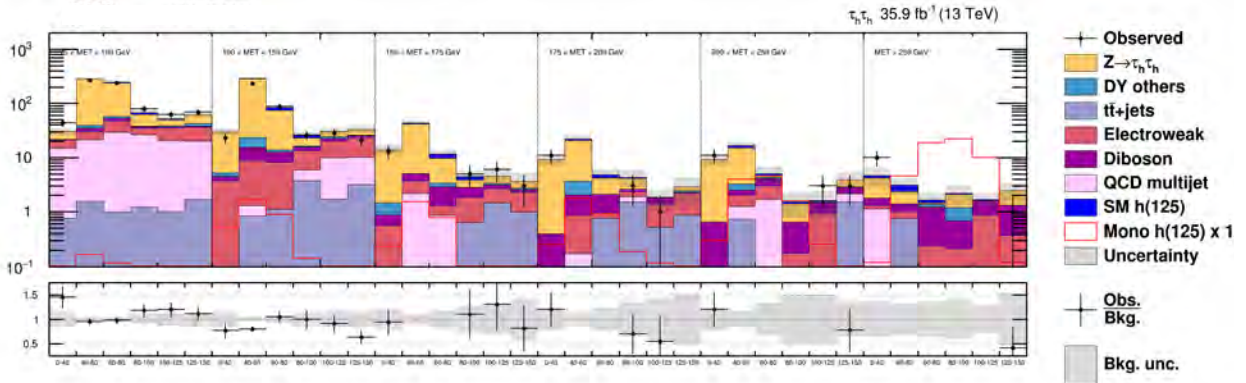
(a) $\tau_h\tau_h$
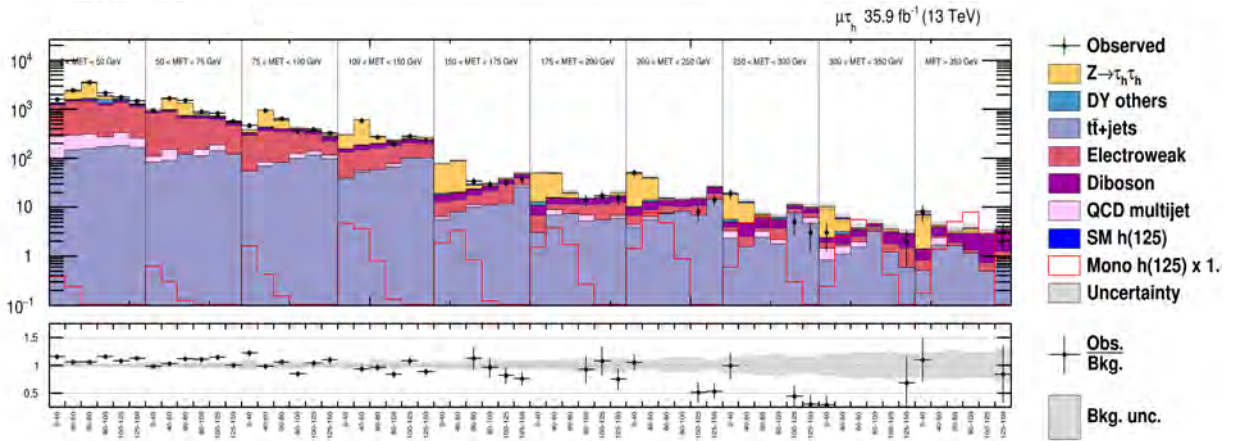
(b) $\mu\tau_h$

Figure 9.2: The visible mass spectrum is unrolled in slices of E_T^{miss} . All uncertainties are considered as functions of E_T^{miss} . The most sensitive category, $\tau_h\tau_h$, and the $\mu\tau_h$ category in the 2D unrolling procedure are still blinded in potentially sensitive areas. The red line demonstrates the 1 pb Z'-2HDM for the $M_{Z'} = 1200$ GeV and $M_{A^0} = 300$ mass points.

Process	$\mu\tau_h$ Events	eτ_h Events	$\tau_h\tau_h$ Events
Full identification criteria	535026	258505	175422
E_T^{miss} > 105 GeV	10127	4837	1244
$p_T(H) > 65$ GeV	4783	2559	958
$\Delta R_{\tau\tau} < 2.0$ GeV	2880	1577	655
$m_{vis} < 125$ GeV	2798	1533	629

Table 9.6: Number of observed events in selection process.

9.5 Systematic uncertainties

The examination of the CMS data for the presence or absence of a mono-Higgs signal contribution is affected by several sources of systematic uncertainty. These uncertainties are expected to affect the normalization of signal and/or background processes and, in several instances, also act to change the distribution of the predicted observable.

The uncertainties from Section 7.6.1 are considered.

The dominant uncertainty in this analysis is the effect of limited Monte Carlo statistical precision in the sensitive region. The effect of the limited simulation affects the final result by up to 40% for some mass points. An uncertainty on the statistical precision of each process in each bin of the distribution is considered. The statistical uncertainty is considered a nuisance in the simultaneous fit if the relative uncertainty on a process in a bin is greater than 5% in the signal region and 10% in the control region. None of the statistical uncertainties resulting for these requirements are merged with any others. There are no bin-by-bin uncertainties considered for the QCD control region, since the same contribution (scaled by a factor) is included in the signal region. We avoid double counting this statistical uncertainty.

The next leading uncertainty is the EWK correction on the WW $\rightarrow 2\ell 2\nu$ prediction [179]. The entire correction is treated as the uncertainty on the contribution. Additional theoretical uncertainties on the modeling of the underlying event and parton shower, parton distribution functions and factorization and normalization scale variations have been included for signal processes [180]. There is not significant variation seen for the pdf and the factorization and normalization scale uncertainties as a function of $E_{\mathrm{T}}^{\mathrm{miss}}$. We use an uncertainty of 4% and 2%, for pdf and α_s respectively.

A 4% uncertainty is assigned to $t\bar{t}$, and a 2% uncertainty is assigned to diboson and single-top to account for changes in overall normalization arising from uncertainties on b-jet tagging performance and mistag rates. A 5% uncertainty is assigned to Z+jets, SM Higgs processes

for b-jet mistag uncertainty, while all other processes including signal receive a 2% mistag uncertainty. The 15%(20%) systematic uncertainty is applied to the QCD contribution for $\mu\tau_h(e\tau_h)$ to account for differing background contributions in same-sign region. All the background systematics in the same-sign region are propagated to the total QCD uncertainty, which approaches 40%.

The W+jets normalization receives an extrapolation uncertainty of 20% in the $\ell\tau_h$ tau channels, while the $\tau_h\tau_h$ channel receives a normalization uncertainty of 10%. The top quark p_T spectrum is reweighted since there is still evidence at CMS that the spectrum is softer in data than in simulation [159, 181]. A shape uncertainty is used for the uncertainty on the double hadronic trigger. Recall, at the HLT τ_h are reconstructed in larger jet cones, and fluctuation of energy in this cone due to pile-up or other HLT cluster mis-calibration will affect the p_T of the reconstructed τ_h and be reflected in the turn on curve. If an HLT τ_h transverse momenta is shifted 3%, there is a 12% difference in acceptance of the tau. At 60 GeV, a 3% energy shift of an HLT tau τ_h leads to a 2 % difference. This variation is included as a systematic uncertainty, with a 1.2% energy scale considered.

To allow for differing jet-τ_h-fake fake rates between data and simulation, a shape uncertainty, capped at 200 GeV is added as function of τp_T to backgrounds with generator-level matched τ_h-fakes. Above 200 GeV, the fake rate is roughly flat. In this analysis, there are not many taus beyond 200 GeV. The fake rate is measured in an anti-isolated region and fit with a second-order polynomial resulting in roughly $1.25 - 0.005 \times p_T(\tau_h) - 0.00001 \times p_T(\tau_h)^2$.

In simulated events, an asymmetric high-p_T tau efficiency uncertainty is applied only to generator-matched hadronic taus. The measurements for the τ_h identification scale factors, discussed in Section 7.5.2, consist of τ_h originating from Z bosons, so the high-p_T statistical precision on the identification efficiency is low. Additionally around 200 GeV the τ_h tracks are fairly straight, accompanied by larger electromagnetic and hadronic calorimeter deposits. The high-p_T τ_h efficiency is measured using off-shell W and Z bosons and extrapolating

to higher efficiencies using a fit. The fit of the measurements produced an asymmetric uncertainty on the efficiency where the one sigma corresponds to +5%/1 TeV and -35%/1 TeV.

Table 9.7: Summary of uncertainties considered in the simultaneous fit

Uncertainty	Affected processes	Change in acceptance or shape		
		$e\tau_h$	$\mu\tau_h$	$\tau_h\tau_h$
Tau ID correlated	Sim. processes	4.5%	4.5%	-
Tau ID uncorrelated	Sim. processes	2%	2%	9%
Electron ID & trigger	Sim. processes	2%	-	-
Muon ID & trigger	Sim. processes	-	2%	-
e misidentified as τ_h	$Z \to$ ee	12%	-	-
μ misidentified as τ_h	$Z \to \mu\mu$	-	25%	-
Jet misidentified as τ_h	$Z +$ jets	Shape		
τ_h energy scale (per decay mode)	Sim. processes	1.2% on energy scale		
Jet energy scale on met	Sim. processes	Shape		
E_T^{miss} energy scale	Sim. processes	Shape		
Luminosity 13TeV	Sim. processes	2.5%		
Norm. W + jets	W + jets	up to 21%		
Norm. $t\bar{t}$	$t\bar{t}$	6%		
Norm. diboson	Diboson	5%		
Norm. single top	Single top	5%		
Norm. QCD	QCD	up to 20%		
Tau trigger	Sim. processes	-	-	Shape
Z+jets LO-NLO reweighting	Z+jets	Shape		
Top p_T reweighting	$t\bar{t}$	Shape		
Higher-order (NLO) EWK/QCD	W + jets and $VV \to 2\ell2\nu$	Shape		
$\tau\tau$ branching ratio	Higgs processes	1.7%		
Renormalization scale	Signal	4%		
Theory: PDF	Signal	2%		
Limited number of events (bin-by-bin)	All processes	Shape only		

9.6 Results

The systematics described in the previous section are treated as nuisance parameters in the global maximum likelihood fit.

The nuisance parameters are varied according to their probability density function, as described in Section 8.5 for Chapter 8. Five control regions, shown in Figs. 9.3 and 9.4 respectively, are used to control the normalizations. Nuisance parameters and normalizations are correlated across regions. The global maximum likelihood fit has the ability to constrain some of the input shape uncertainties and normalization uncertainties, which can produce a more discriminating limit. The resulting postfit distributions from a background-only global maximum likelihood fit are in Fig 9.5.
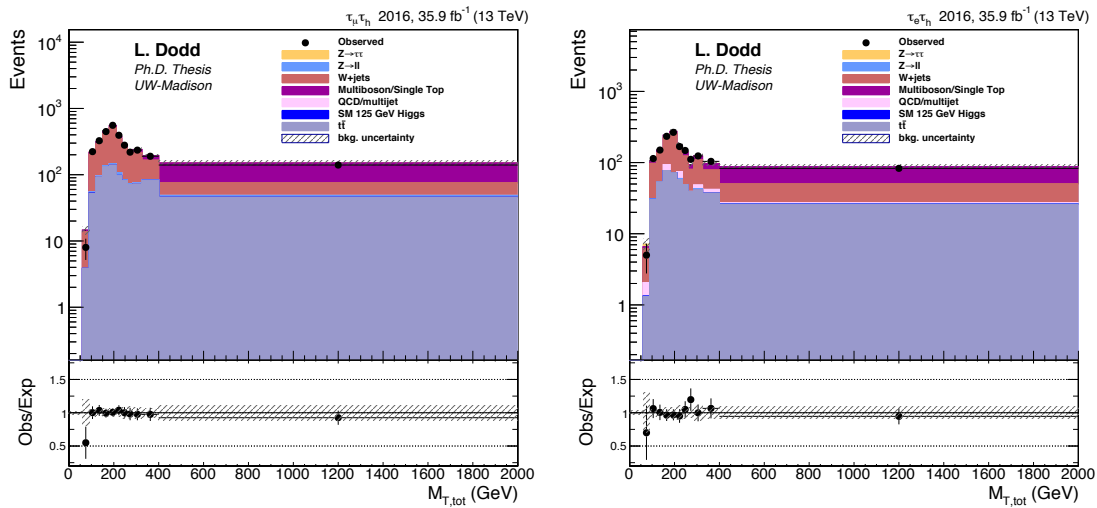
(a) $\mu\tau_h$ W+jets Control Region (Anti-Isolated Opposite-Sign)

(b) eτ_h W+jets Control Region (Anti-Isolated Opposite-Sign)

Figure 9.3: W+jets control regions for systematics and normalization propagation
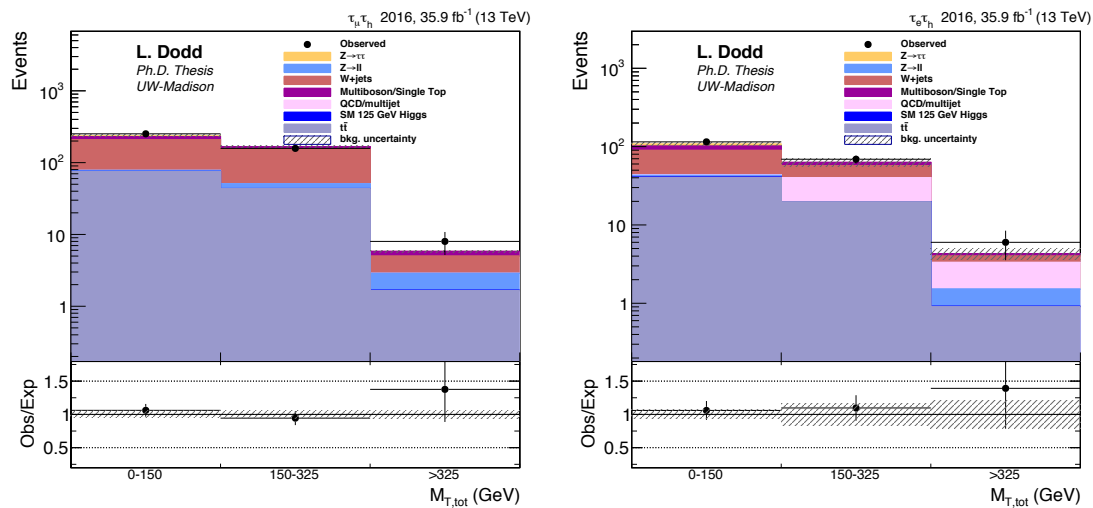
(a) $\mu\tau_h$ QCD Control Region (Isolated Same-Sign)

(b) eτ_h QCD Control Region (Isolated Same-Sign)
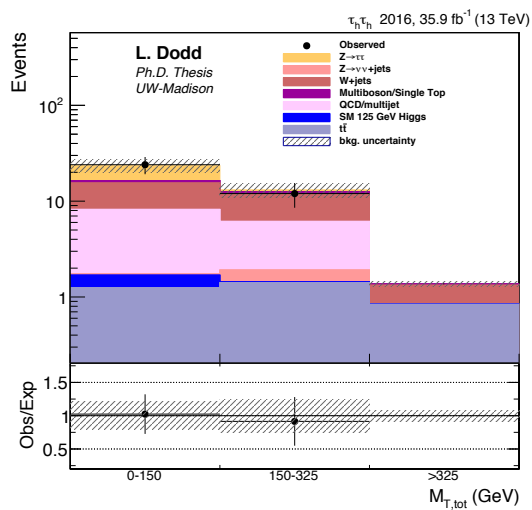
(c) $\tau_h\tau_h$ QCD Control Region (Same-Sign)

Figure 9.4: Multijet control regions for systematics and normalization propagation

The fit to extract the limit is performed on the M_T distributions of the three τ channels ($e\tau_h$, $\mu\tau_h$, and $\tau_h\tau_h$) together with the W+jets and QCD control regions. The simultaneous fit takes the systematics previously discussed into consideration. The extracted postfit contributions and observed values in the sensitive region, $M_{T,tot} > 260$ GeV, of the analysis are shown in Table 9.8. The observed number of events is in good agreement with the predicted SM backgrounds.

Table 9.8: Estimated background contributions and observed events for $M_{T,tot} > 260$ GeV in the signal region for 35.9 fb^{-1} of 2016 data. The total expected contributions include the statistical and systematic error. The $Z \to \ell\ell$ contribution is roughly split 50% ℓ-faking-τ_h and 50% jet-faking-τ_h.

Process	$\mu\tau_h$	$e\tau_h$	$\tau_h\tau_h$
W + jets QCD	32.54 ± 6.18	13.11 ± 2.18	3.79 ± 2.59
$t\bar{t}$	24.83 ± 2.04	13.75 ± 1.60	4.24 ± 1.30
125 GeV H	0.72 ± 0.06	0.48 ± 0.08	1.21 ± 0.08
Multi-boson	21.53 ± 1.46	12.34 ± 0.99	7.30 ± 0.63
$Z \to \tau\tau$	0.14 ± 0.53	0.00 ± 0.01	3.57 ± 1.24
$Z \to \ell\ell$	2.00 ± 1.33	0.84 ± 1.87	-
$Z \to \nu\nu$	-	-	0.37 ± 0.25
Total expected	81.77 ± 6.31	40.50 ± 3.26	20.48 ± 2.97
Observed	81.00 ± 9.00	38.00 ± 6.16	26.00 ± 5.10
Expected Zprime1200A300 events	5.75 ± 0.27	3.52 ± 0.16	4.78 ± 0.33

The results for the Z'-2HDM interpretation are shown in Fig. 9.6, where the expected 95% CL upper limits are computed using a profile-likelihood ratio and the modified frequentist method [182, 183]. The results are interpreted for the baryonic Z' model shown in Fig. 9.7. In the two interpretations above, the observed upper limits are presented in blue, while the expected is in black with the $\pm$ one sigma bands are shown in red.

No significant deviation from the SM background prediction is observed. Several mass points can be excluded from the considered models. We are able to exclude the Z'-2HDM model for the $M_A^0 = 300$ GeV between the Z' masses of $M_{Z'} = [800, 1000]$ GeV. We are able to exclude the baryonic Z' model for the mass point $M_{Z'} = 10$ GeV and $M_\chi = 1$.
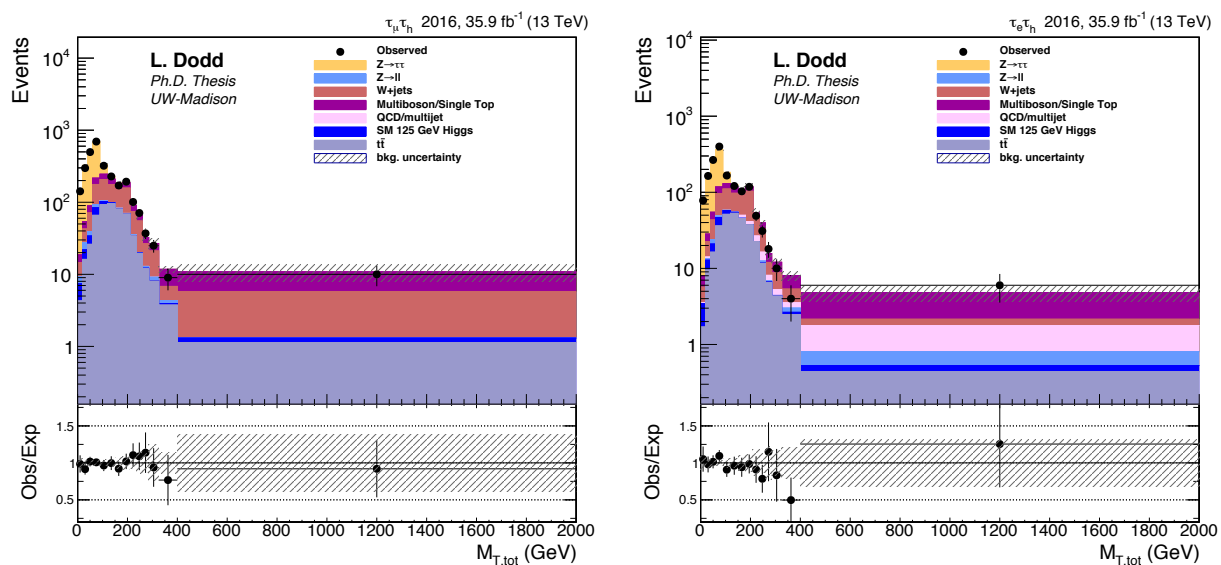
(a) $\mu\tau_h$

(b) $e\tau_h$
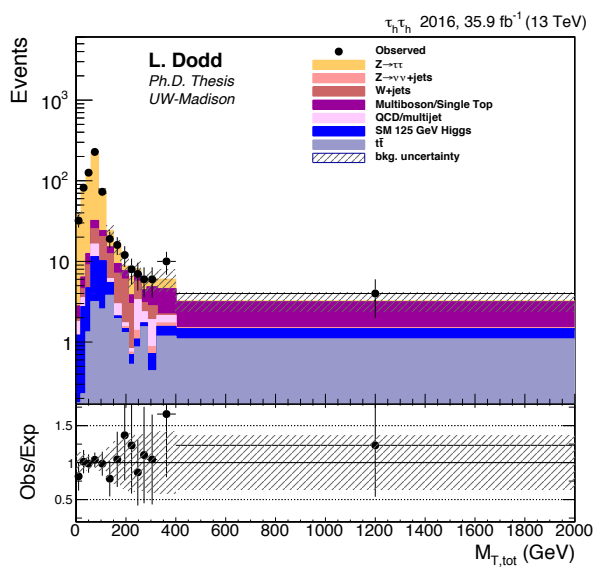
(c) $\tau_h\tau_h$

Figure 9.5: Postfit distributions, without including any signal in the fit.
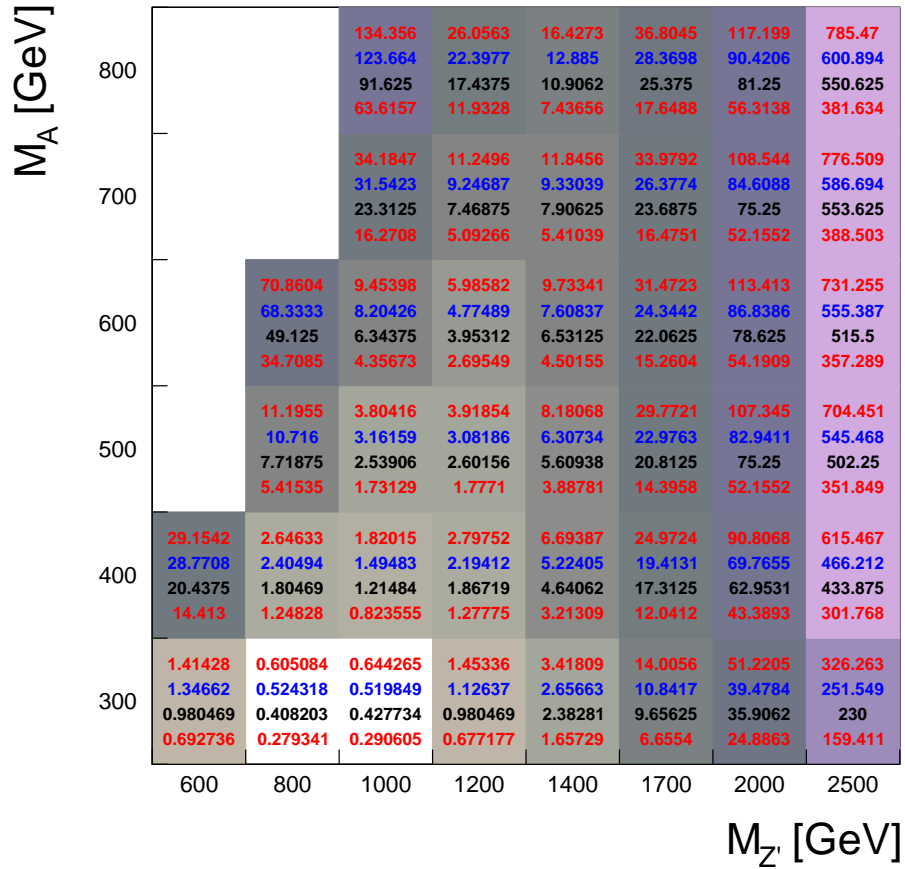
(a) Z'-2HDM model

Figure 9.6: 95% C.L. upper limit on σ/σ_{SM} for the Z'-2HDM. Z' mass scanned for varying A^0 values. The observed limit is shown in blue and the expected limit in black. The $+1$ and -1 sigma bands are printed on the plot. The colors visually represent the expected limit on σ/σ_{SM} presented via log scale.
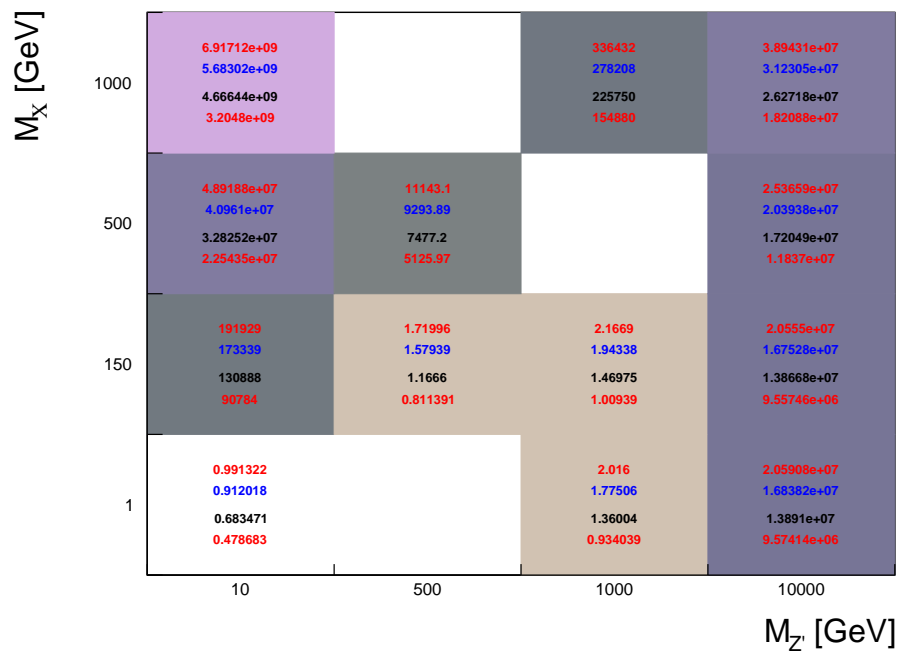
(a) Select mass points for baryonic-Z' model

Figure 9.7: 95% C.L. upper limit on σ/σ_{SM} for the baryonic-Z' model. Z' mass scanned for varying M_{DM} values. The observed limit is shown in blue and the expected limit in black. The +1 and -1 sigma bands are printed on the plot. The colors visually represent the expected limit on σ/σ_{SM} presented via log scale.

The results presented in above Figs. 9.6 and 9.7 are produced from a limited grid of model mass points. Generally the kinematics of the models smoothly vary from one mass point to the next, and exploiting this the range of masses excluded is smoothed and expanded.

Weights are developed as a function of generator-level variables to extrapolate between nearby mass points. Closure tests are performed, ensuring the weights accurately extrapolate from one mass point to another; closure tests entail weighting one mass point to another already simulated mass point and checking agreement of generator level variables. These closure tests agree within 5%, and given the statistical precision of the results, can be considered negligible. This interpolation method for prediction between mass point produces a much smoother grid, shown in Fig. 9.8. Further work includes producing more baryonic simulation in order to produce a similar 2D exclusion plot for the baryonic Z' model.

With more baryonic sample simulation availability, the analysis in this thesis will combine with the mono-Higgs($\gamma\gamma$) final state to produce a more discriminating final limits than either analysis can produce individually. In the ,ono-$\tau\tau/\gamma\gamma$ combination, the luminosity, MET and JET energy scales, and various theoretical uncertainties are be correlated. A result of this rough combination with the baryonic Z' model is shown in Fig. 9.9. The observed event list in this thesis was compared with that used in the mono-Higgs($\gamma\gamma$) analysis and no events overlap.

The combination of the $\gamma\gamma$ and $\tau\tau$ final states produces a slightly stronger limit, more than doubling the exclusion phase space of $\tau\tau$ before the combination.
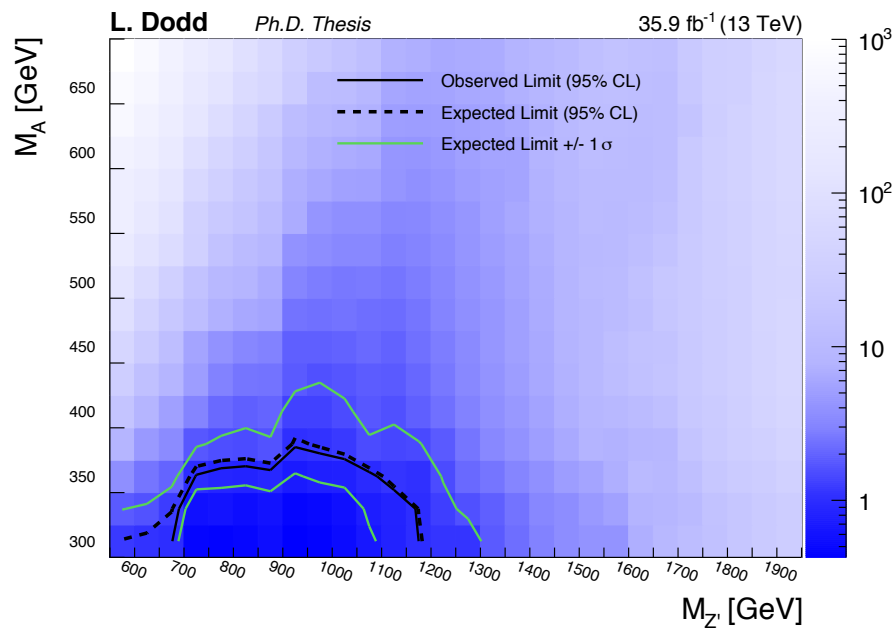
Figure 9.8: Z'-2HDM interpolated 2D exclusion plot for the Z'-2HDM model. Z' mass scanned for varying M_{A^0} values. The color scale on this plot represents the expected 95% CLs upper limits on the production. The portion of excluded phase space is below the observed solid-black line.
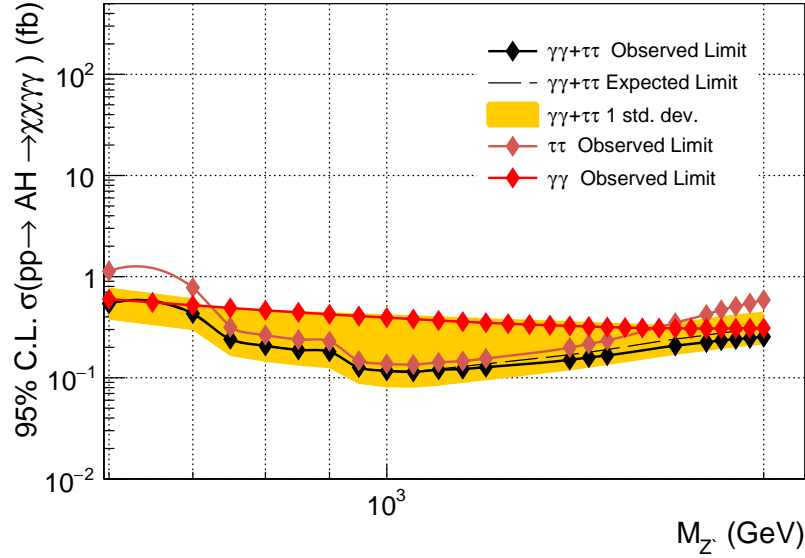
(a) Z'-2HDM, $M_A^0 = 300$ GeV
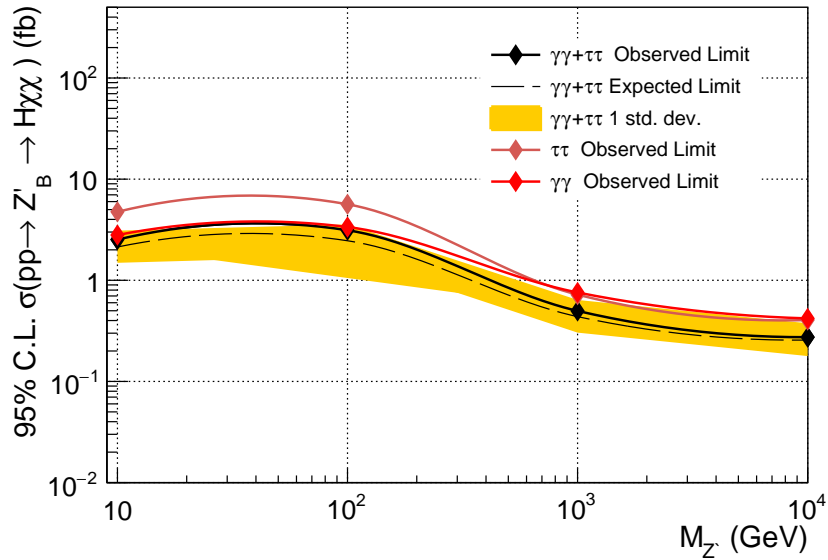
(b) Baryonic Z', $M_\chi = 1$ GeV

Figure 9.9: 95% C.L. upper limit on σ for the Z'-2HDM model in (a) and the baryonic Z' in (b). The combination of $\tau\tau$ and $\gamma\gamma$ roughly doubles the exclusion of $\tau\tau$ alone for the baryonic Z' model.

Chapter 10

Summary and Outlook

10.1 Standard model Higgs decays to τ leptons

10.1.1 Summary

This thesis presented the measurement of the $H \rightarrow \tau\tau$ decay within the Compact Muon Solenoid (CMS) experiment using events recorded in 13 TeV proton-proton collisions in 2016 at the Large Hadron Collider (LHC). The data set corresponds to an integrated luminosity of 35.9 fb^{-1}. The τ leptons are allowed to decay to hadrons or to leptons, and four final states ($e\mu$, $e\tau$, $\mu\tau$,and $\tau\tau$) are considered. An excess of data over background is found at the level of 4.9 standard deviations for a SM Higgs boson with mass of $m_H = 125$ GeV, compared with an expected 4.7 standard deviations. A maximum likelihood fit is performed to extract the best fit result for the signal strength of $\mu = 1.09^{+0.27}_{-0.24}$ times the SM expectation. In combination with the 7 and 8 TeV CMS result, the expected and observed result are both 5.9 standard deviations.

10.1.2   Outlook

Further steps include reducing the uncertainty on the Higgs boson vector boson fusion and gluon fusion cross sections. Additionally the tau pair final state provides a useful handle on measuring the polarization and searching for anomalous couplings of the Higgs boson; any CP violation in Higgs decays could be BSM physics. For further precision, the WH and ZH final states can be included, given that the observed event list from this thesis does not overlap with any analysis developed for WH and ZH searches. With an expected full Run-II data set of 150 fb^{-1}, one could expect potential SM deviations occurring most likely on the order of one to two sigma, as no significant deviations have yet been seen on 36 fb^{-1}. The overall cross section of VBF production and the $H \rightarrow \tau\tau$ coupling may measured at roughly the 10% level with 150 fb^{-1}. Other H final states may measure the gluon fusion production cross section tat better than the 10% level.

10.2   Dark matter in association with Higgs decays to τ leptons

10.2.1   Summary

Additionally this thesis presented a search for dark matter in association with a $H \rightarrow \tau\tau$. Limits were placed on the baryonic Z' and Z'-2HDM models, with no significant deviations from expected standard model production found. Z' mass points between $M_{Z'} = [800, 1000]$ GeV and for the $M_A^0 = 300$ GeV were excluded in the Z'-2HDM models, while only the one mass point for the baryonic Z' model, given a rough grid, is excluded: $M_{Z'} = 10$ GeV and $M_\chi = 1$. A combination of $H \rightarrow \gamma\gamma + DM$ and this thesis approximately double the range of excluded mass points with interpolation.

10.2.2 Outlook

Additional short-term steps for the mono-Higgs dark matter search include scanning the full mass point grid in combination with $\gamma\gamma$ for both baryonic Z' and Z'-2HDM models using interpolated mass points for a smooth grid. Searches for these models should be repeated on the full Run-II LHC data set. The strategy used in this thesis could be optimized further in order to (1) reduce the E_T^miss cut needed offline to reduce backgrounds and (2) increase acceptance by considering boosted tau pair topologies. While $H \to \tau\tau + E_\text{T}^\text{miss}$ is not the most discriminating of the mono-Higgs final states given SM Higgs boson decays, it may be more interesting for other models that include additional scalar bosons that preferentially couple to the third generation and have associated dark matter in the final state. The analysis of this final state should be continued in the LHC program.

With the expected full Run-II data set of 150 fb^{-1}, if no new physics exists and all mono-Higgs final states are combined, the Z'-2HDM model will most likely be fully excluded given a 100% branching ratio of $A \to \chi\chi$. The full combination for the baryonic-Z' model may be mostly excluded except for mass points with cross sections smaller than about 10^{-5} pb. If new physics does exist, the expected significance from mono-H($\tau\tau$) would probably not reach 5 sigma.

Bibliography

[1] Albert M Sirunyan et al. Observation of the Higgs boson decay to a pair of tau leptons. 2017.

[2] CMS Collaboration. Observation of the SM scalar boson decaying to a pair of τ leptons with the CMS experiment at the LHC. 2017.

[3] S. Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264, 1967.

[4] Abdus Salam. Weak and electromagnetic interactions. In Nils Svartholm, editor, *Elementary particle physics: relativistic groups and analyticity*, page 367, Stockholm, 1968. Almqvist & Wiksell. Proceedings of the eighth Nobel symposium.

[5] Martin L. Perl et al. Evidence for Anomalous Lepton Production in e+ - e- Annihilation. *Phys. Rev. Lett.*, 35:1489–1492, 1975.

[6] Martin L. Perl et al. Properties of Anomalous e mu Events Produced in e+ e- Annihilation. *Phys. Lett.*, 63B:466, 1976.

[7] Martin L. Perl et al. Properties of the Proposed tau Charged Lepton. *Phys. Lett.*, 70B:487–490, 1977.

[8] C. Patrignani et al. Review of Particle Physics. *Chin. Phys.*, C40(10):100001, 2016.

[9] Francis Halzen and Alan Martin. *Quarks & Leptons: An introductory course in modern particle physics*. John Wiley & Sons, New York, USA, 1984.

[10] Makoto Kobayashi and Toshihide Maskawa. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys.*, 49:652–657, 1973.

[11] S. W. Herb et al. Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions. *Phys. Rev. Lett.*, 39:252–255, 1977.

[12] F. Abe et al. Observation of top quark production in $\bar{p}p$ collisions. *Phys. Rev. Lett.*, 74:2626–2631, 1995.

[13] S. Abachi et al. Observation of the top quark. *Phys. Rev. Lett.*, 74:2632–2637, 1995.

[14] Royal Swedish Acad. Sci. Neutrino oscillations. *The Universe*, 3(4):38–50, 2015.

[15] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, 1998.

[16] Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by ^8b solar neutrinos at the sudbury neutrino observatory. *Phys. Rev. Lett.*, 87:071301, Jul 2001.

[17] R. Keith Ellis, W. James Stirling, and B. R. Webber. QCD and collider physics. *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.*, 8:1–435, 1996.

[18] Mark Srednicki. *Quantum Field Theory*. Cambridge Univ. Press, Cambridge, 2007.

[19] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.

[20] Roel Aaij et al. Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \rightarrow J/\psi K^- p$ Decays. *Phys. Rev. Lett.*, 115:072001, 2015.

[21] K. S. Babu et al. Working Group Report: Baryon Number Violation. In *Proceedings, 2013 Community Summer Study on the Future of U.S. Particle Physics: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, July 29-August 6, 2013*, 2013.

[22] David J. Gross and Frank Wilczek. Ultraviolet Behavior of Nonabelian Gauge Theories. *Phys. Rev. Lett.*, 30:1343–1346, 1973.

[23] H. David Politzer. Reliable Perturbative Results for Strong Interactions? *Phys. Rev. Lett.*, 30:1346–1349, 1973.

[24] D. J. Gross and Frank Wilczek. Asymptotically Free Gauge Theories. 1. *Phys. Rev.*, D8:3633–3652, 1973.

[25] H. David Politzer. Asymptotic Freedom: An Approach to Strong Interactions. *Phys. Rept.*, 14:129–180, 1974.

[26] Nicola Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, 10:531–533, 1963. [,648(1963)].

[27] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961.

[28] Gerard 't Hooft. Renormalizable Lagrangians for Massive Yang-Mills Fields. *Nucl. Phys.*, B35:167–188, 1971.

[29] T. W. B. Kibble. History of electroweak symmetry breaking. *J. Phys. Conf. Ser.*, 626(1):012001, 2015.

[30] David J Griffiths. *Introduction to elementary particles; 2nd rev. version.* Physics textbook. Wiley, New York, NY, 2008.

[31] R. Rosenfeld. Physics Beyond the Standard Model. In *Proceedings, 8th CERNLatin-American School of High-Energy Physics (CLASHEP2015): Ibarra, Ecuador, March 05-17, 2015*, pages 159–164, 2016.

[32] Hitoshi Murayama. Physics Beyond the Standard Model and Dark Matter. In *Les Houches Summer School - Session 86: Particle Physics and Cosmology: The Fabric of Spacetime Les Houches, France, July 31-August 25, 2006*, 2007.

[33] Vardan Khachatryan et al. Searches for a heavy scalar boson H decaying to a pair of 125 GeV Higgs bosons hh or for a heavy pseudoscalar boson A decaying to Zh, in the final states with $h \to \tau\tau$. *Phys. Lett.*, B755:217–244, 2016.

[34] CMS Collaboration. Search for a neutral MSSM Higgs boson decaying into $\tau\tau$ with 12.9 fb^{-1} of data at $\sqrt{s} = 13$ TeV. 2016.

[35] F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110–127, 1933.

[36] V. C. Rubin, N. Thonnard, and W. K. Ford, Jr. Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/. *Astrophys. J.*, 238:471, 1980.

[37] Katherine Freese. Review of Observational Evidence for Dark Matter in the Universe and in upcoming searches for Dark Stars. *EAS Publ. Ser.*, 36:113–126, 2009.

[38] Stefano Profumo. *An Introduction to Particle Dark Matter*. World Scientific, 2017.

[39] P. J. E. Peebles and Bharat Ratra. The Cosmological constant and dark energy. *Rev. Mod. Phys.*, 75:559–606, 2003.

[40] Mariangela Lisanti. Lectures on Dark Matter Physics. In *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: New Frontiers in Fields and Strings (TASI 2015): Boulder, CO, USA, June 1-26, 2015*, pages 399–446, 2017.

[41] Wayne Hu, Rennan Barkana, and Andrei Gruzinov. Cold and fuzzy dark matter. *Phys. Rev. Lett.*, 85:1158–1161, 2000.

[42] Shunsaku Horiuchi, Philip J. Humphrey, Jose Onorbe, Kevork N. Abazajian, Manoj Kaplinghat, and Shea Garrison-Kimmel. Sterile neutrino dark matter bounds from galaxies of the Local Group. *Phys. Rev.*, D89(2):025017, 2014.

[43] Maxim Markevitch, A. H. Gonzalez, D. Clowe, A. Vikhlinin, L. David, W. Forman, C. Jones, S. Murray, and W. Tucker. Direct constraints on the dark matter self-interaction cross-section from the merging galaxy cluster 1E0657-56. *Astrophys. J.*, 606:819–824, 2004.

[44] Gianfranco Bertone, Dan Hooper, and Joseph Silk. Particle dark matter: Evidence, candidates and constraints. *Phys. Rept.*, 405:279–390, 2005.

[45] Daniel Abercrombie et al. Dark Matter Benchmark Models for Early LHC Run-2 Searches: Report of the ATLAS/CMS Dark Matter Forum. 2015.

[46] M. Aguilar et al. First Result from the Alpha Magnetic Spectrometer on the International Space Station: Precision Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5350 GeV. *Phys. Rev. Lett.*, 110:141102, 2013.

[47] G. Arnison et al. Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at s**(1/2) = 540-GeV. *Phys. Lett.*, 122B:103–116, 1983. [,611(1983)].

[48] G. Arnison et al. Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider. *Phys. Lett.*, 126B:398–410, 1983.

[49] J. D. Jackson and L. B. Okun. Historical roots of gauge invariance. *Rev. Mod. Phys.*, 73:663–680, Sep 2001.

[50] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321, 1964.

[51] P. W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508, 1964.

[52] P. W. Higgs. Broken symmetries, massless particles and gauge fields. *Phys. Lett.*, 12:132, 1964.

[53] P. W. Higgs. Spontaneous Symmetry Breakdown without Massless Bosons. *Phys. Rev.*, 145:1156, 1966.

[54] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.*, 13:585, 1964.

[55] T. W. B. Kibble. Symmetry Breaking in Non-Abelian Gauge Theories. *Phys. Rev.*, 155:1554, 1967.

[56] R. Barate et al. Search for the standard model Higgs boson at LEP. *Phys. Lett.*, B565:61–75, 2003.

[57] CDF Collaboration. Search for a low-mass standard model higgs boson in the $\tau\tau$ decay channel in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Rev. Lett.*, 108:181804, May 2012.

[58] T. Aaltonen et al. Higgs Boson Studies at the Tevatron. *Phys. Rev.*, D88(5):052014, 2013.

[59] Georges Aad et al. Combined Measurement of the Higgs Boson Mass in *pp* Collisions at $\sqrt{s}$ = 7 and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett.*, 114:191803, 2015.

[60] Georges Aad et al. Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV. *JHEP*, 08:045, 2016.

[61] T. Aaltonen et al. Evidence for a particle produced in association with weak bosons and decaying to a bottom-antibottom quark pair in Higgs boson searches at the Tevatron. *Phys. Rev. Lett.*, 109:071804, 2012.

[62] Georges Aad et al. Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector. *Eur. Phys. J.*, C75(10):476, 2015. [Erratum: Eur. Phys. J.C76,no.3,152(2016)].

[63] Vardan Khachatryan et al. Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV. *Eur. Phys. J.*, C75(5):212, 2015.

[64] Serguei Chatrchyan et al. Measurement of the properties of a Higgs boson in the four-lepton final state. *Phys. Rev.*, D89(9):092007, 2014.

[65] Vardan Khachatryan et al. Limits on the Higgs boson lifetime and width from its decay to four charged leptons. *Phys. Rev.*, D92(7):072010, 2015.

[66] Vardan Khachatryan et al. Searches for invisible decays of the Higgs boson in pp collisions at sqrt(s) = 7, 8, and 13 TeV. *JHEP*, 02:135, 2017.

[67] Asher Berlin, Tongyan Lin, and Lian-Tao Wang. Mono-Higgs Detection of Dark Matter at the LHC. *JHEP*, 06:078, 2014.

[68] Linda Carpenter, Anthony DiFranzo, Michael Mulhearn, Chase Shimmin, Sean Tulin, and Daniel Whiteson. Mono-Higgs-boson: A new collider probe of dark matter. *Phys. Rev.*, D89(7):075017, 2014.

[69] Nathaniel Craig, Jamison Galloway, and Scott Thomas. Searching for Signs of the Second Higgs Doublet. 2013.

[70] Csaba Csaki. The Minimal supersymmetric standard model (MSSM). *Mod. Phys. Lett.*, A11:599, 1996.

[71] Vardan Khachatryan et al. Search for heavy resonances decaying to tau lepton pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 02:048, 2017.

[72] D. S. Akerib et al. First results from the LUX dark matter experiment at the Sanford Underground Research Facility. *Phys. Rev. Lett.*, 112:091303, 2014.

[73] Vardan Khachatryan et al. Search for narrow resonances in dilepton mass spectra in proton-proton collisions at $\sqrt{s} = 13$ TeV and combination with 8 TeV data. *Phys. Lett.*, B768:57–80, 2017.

[74] Search for a high-mass resonance decaying into a dilepton final state in 13 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-EXO-16-031, CERN, Geneva, 2016.

[75] Lyndon Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001, 2008.

[76] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[77] A. Augusto Alves, Jr. et al. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.

[78] K. Aamodt et al. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008.

[79] R. Schmidt et al. Protection of the CERN Large Hadron Collider. *New J. Phys.*, 8:290, 2006.

[80] S. Chatrchyan et al. The cms experiment at the cern lhc. *JINST*, 3:S08004, 2008.

[81] Vardan Khachatryan et al. Pseudorapidity distribution of charged hadrons in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett.*, B751:143–163, 2015.

[82] S Chatrchyan et al. Precise Mapping of the Magnetic Field in the CMS Barrel Yoke using Cosmic Rays. *JINST*, 5:T03021, 2010.

[83] T M Taylor. The Magnets for the LHC Experiments; 1999 ed. (LHC-Project-Report-339. CERN-LHC-Project-Report-339):6 p, Dec 1999.

[84] D Delikaris, J P Dauvergne, Giorgio Passardi, J C Lottin, J P Lottin, and C Lyraud. The cryogenic system for the superconducting solenoid magnet of the CMS experiment. Technical Report LHC-Project-Report-165. CERN-LHC-Project-Report-165, CERN, Geneva, Jan 1998.

[85] V. I. Klyukhin, D. Campi, B. Cure, A. Gaddi, H. Gerwig, J. P. Grillet, A. Herve, R. Loveless, and R. P. Smith. Developing the Technique of Measurements of Magnetic Field in the CMS Steel Yoke Elements With Flux-Loops and Hall Probes. *IEEE Transactions on Nuclear Science*, 51:2187–2192, October 2004.

[86] CMS Collaboration. Commissioning and performance of the cms silicon strip tracker with cosmic ray muons. *Journal of Instrumentation*, 5(03):T03008, 2010.

[87] Serguei Chatrchyan et al. Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV. *JINST*, 8:P09009, 2013. [JINST8,9009(2013)].

[88] P Adzic. Energy resolution of the barrel of the cms electromagnetic calorimeter. *Journal of Instrumentation*, 2(04):P04004, 2007.

[89] Victor Daniel Elvira. Measurement of the Pion Energy Response and Resolution in the CMS HCAL Test Beam 2002 Experiment. Technical Report CMS-NOTE-2004-020, CERN, Geneva, Sep 2004.

[90] Serguei Chatrchyan et al. Identification and Filtering of Uncharacteristic Noise in the CMS Hadron Calorimeter. *JINST*, 5:T03014, 2010.

[91] CMS-HCAL Collaboration. Design, Performance and Calibration of the CMS Forward Calorimeter Wedges. Technical Report CMS-NOTE-2006-044, CERN, Geneva, Feb 2006.

[92] Giovanni Abbiendi. The CMS muon system in Run2: preparation, status and first results. *PoS*, EPS-HEP2015:237, 2015.

[93] S Chatrchyan et al. Performance of the CMS Drift Tube Chambers with Cosmic Rays. *JINST*, 5:T03015, 2010.

[94] Serguei Chatrchyan et al. The performance of the CMS muon detector in proton-proton collisions at sqrt(s) = 7 TeV at the LHC. *JINST*, 8:P11002, 2013.

[95] J Hauser, D Acosta, E Boyd, B Bylsma, R Cousins, A A Drozdetsky, S Durkin, J Gilmore, J Gu, S Haapanen, A Korytov, S Lee, T Ling, A Madorsky, M Matveev, M Mey, B Mohr, J Mumford, P Padley, G Pawloski, J Roberts, B Scurlock, H Stck, V Yu Valuev, G Veramendi, J Werner, and Y Zheng. Experience with Trigger Electronics for the Cathode Strip Chamber System of CMS. 2004.

[96] S Chatrchyan et al. Performance of the CMS Cathode Strip Chambers with Cosmic Rays. *JINST*, 5:T03018, 2010.

[97] B. Kreis and others. Run 2 Upgrades to the CMS Level-1 Calorimeter Trigger. *JINST*, 11(01):C01051, 2016.

[98] A. Svetek, M. Blake, M. Cepeda Hermida, S. Dasu, L. Dodd, R. Fobes, B. Gomber, T. Gorski, Z. Guo, P. Klabbers, A. Levine, I. Ojalvo, T. Ruggles, N. Smith, W.H. Smith, J. Tikalsky, M. Vicente, and N. Woods. The calorimeter trigger processor card: the next generation of high speed algorithmic data processing at CMS. *JINST*, 11(02):C02011, 2016.

[99] A Tapper and Darin Acosta. CMS Technical Design Report for the Level-1 Trigger Upgrade. Technical Report CERN-LHCC-2013-011. CMS-TDR-12, Jun 2013.

[100] Daniele Trocino. The CMS High Level Trigger. *J. Phys. Conf. Ser.*, 513:012036, 2014.

[101] CMS Collaboration. CMS Luminosity Measurements for the 2016 Data Taking Period. Technical Report CMS-PAS-LUM-17-001, CERN, Geneva, 2017.

[102] S van der Meer. Calibration of the effective beam height in the ISR. Technical Report CERN-ISR-PO-68-31. ISR-PO-68-31, CERN, Geneva, 1968.

[103] Torbjrn Sjstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.

[104] M. Bahr et al. Herwig++ Physics and Manual. *Eur. Phys. J.*, C58:639–707, 2008.

[105] T. Gleisberg, Stefan. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter. Event generation with SHERPA 1.1. *JHEP*, 02:007, 2009.

[106] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.

[107] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.

[108] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5 : Going Beyond. *JHEP*, 06:128, 2011.

[109] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[110] John M. Campbell, R. Keith Ellis, Paolo Nason, and Emanuele Re. Top-Pair Production and Decay at NLO Matched with Parton Showers. *JHEP*, 04:114, 2015.

[111] Simone Alioli, Keith Hamilton, Paolo Nason, Carlo Oleari, and Emanuele Re. Jet pair production in POWHEG. *JHEP*, 04:081, 2011.

[112] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. NLO Higgs boson production via gluon fusion matched with shower in POWHEG. *JHEP*, 04:002, 2009.

[113] Oliver Brein, Abdelhak Djouadi, and Robert Harlander. NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders. *Phys. Lett. B*, 579:149, 2004.

[114] V. Ravindran, J. Smith, and W. L. van Neerven. NNLO corrections to the total cross section for Higgs boson production in hadron-hadron collisions. *Nucl. Phys. B*, 665:325, 2003.

[115] D. de Florian, G. Ferrera, M. Grazzini, and D. Tommasini. Higgs boson production at the LHC: transverse momentum resummation effects in the H $\to \gamma\gamma$, H $\to$ WW $\to \ell\nu\ell\nu$ and H $\to ZZ \to 4\ell$ decay modes. *JHEP*, 06:132, 2012.

[116] Gionata Luisoni, Paolo Nason, Carlo Oleari, and Francesco Tramontano. $HW^\pm/HZ +$ 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO. *JHEP*, 10:083, 2013.

[117] Rikkert Frederix, Emanuele Re, and Paolo Torrielli. Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO. *JHEP*, 09:130, 2012.

[118] Johan Alwall et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J.*, C53:473–500, 2008.

[119] Lance J. Dixon. A brief introduction to modern amplitude methods. In *Proceedings, 2012 European School of High-Energy Physics (ESHEP 2012): La Pommeraye, Anjou, France, June 06-19, 2012*, pages 31–67, 2014.

[120] Neil D. Christensen and Claude Duhr. FeynRules - Feynman rules made easy. *Comput. Phys. Commun.*, 180:1614–1641, 2009.

[121] Neil D. Christensen, Priscila de Aquino, Celine Degrande, Claude Duhr, Benjamin Fuks, Michel Herquet, Fabio Maltoni, and Steffen Schumann. A Comprehensive approach to new physics simulations. *Eur. Phys. J.*, C71:1541, 2011.

[122] Adam Alloul, Neil D. Christensen, Cline Degrande, Claude Duhr, and Benjamin Fuks. FeynRules 2.0 - A complete toolbox for tree-level phenomenology. *Comput. Phys. Commun.*, 185:2250–2300, 2014.

[123] Richard D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015.

[124] Richard D. Ball, Valerio Bertone, Francesco Cerutti, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Jose I. Latorre, Juan Rojo, and Maria Ubiali. Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO. *Nucl. Phys.*, B855:153–221, 2012.

[125] Vardan Khachatryan et al. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J.*, C76(3):155, 2016.

[126] Richard D. Ball et al. Parton distributions from high-precision collider data. 2017.

[127] Darin Acosta, Filippo Ambroglini, Paolo Bartalini, Albert de Roeck, Livio Fano, Richard D Field, and Khristian Kotov. The Underlying Event at the LHC. Technical Report CMS-NOTE-2006-067, CERN, Geneva, Jun 2006.

[128] Rick Field. Min-Bias and the Underlying Event at the LHC. In *Proceedings, 31st International Conference on Physics in collisions (PIC 2011): Vancouver, Canada, August 28-September 1, 2011*, 2012.

[129] T. Sjostrand and Peter Z. Skands. Multiple interactions and the structure of beam remnants. *JHEP*, 03:053, 2004.

[130] Peter Skands, Stefano Carrazza, and Juan Rojo. Tuning PYTHIA 8.1: the Monash 2013 Tune. *Eur. Phys. J.*, C74(8):3024, 2014.

[131] P. Golonka, B. Kersevan, T. Pierzchala, E. Richter-Was, Z. Was, and M. Worek. The Tauola photos F environment for the TAUOLA and PHOTOS packages: Release. 2. *Comput. Phys. Commun.*, 174:818–835, 2006.

[132] Philip Ilten. Tau Decays in Pythia 8. *Nucl. Phys. Proc. Suppl.*, 253-255:77–80, 2014.

[133] David Grellscheid and Peter Richardson. Simulation of Tau Decays in the Herwig++ Event Generator. 2007.

[134] René Brun, F. Bruyant, Federico Carminati, Simone Giani, M. Maire, A. McPherson, G. Patrick, and L. Urban. GEANT4: Detector Description and Simulation Tool. 1994.

[135] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth.*, A506:250–303, 2003.

[136] Tau-Id performance with full 2016 dataset using $Z \rightarrow \tau_\mu \tau_h$ events. Technical report, Jan 2017.

[137] Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector. Technical Report CMS-PAS-JME-16-004, CERN, Geneva, 2016.

[138] Albert M Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. 2017.

[139] D. Buskulic et al. Performance of the ALEPH detector at LEP. *Nucl. Instrum. Meth.*, A360:481–506, 1995.

[140] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

[141] R. Fruhwirth. Application of kalman filtering to track and vertex fitting. *Nucl. Instrum. Meth. A*, 262:444, 1987.

[142] W. Adam, B. Mangano, T. Speer, and T. Todorov. Track reconstruction in the CMS tracker. 2005.

[143] Serguei Chatrchyan et al. Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV. *JINST*, 7:P10002, 2012.

[144] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014.

[145] Paolo Azzurri. Track Reconstruction Performance in CMS. *Nucl. Phys. Proc. Suppl.*, 197:275–278, 2009.

[146] Vardan Khachatryan et al. CMS Tracking Performance Results from early LHC Operation. *Eur. Phys. J.*, C70:1165–1192, 2010.

[147] W. Adam, R. Fruhwirth, A. Strandlie, and T. Todorov. Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC. *eConf*, C0303241:TULT009, 2003. [J. Phys.G31,N9(2005)].

[148] Vardan Khachatryan et al. Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV. *JINST*, 10(06):P06005, 2015.

[149] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, I. Puljak, C. Rovelli, R. Salerno, and Y. Sirois. Electron reconstruction in CMS. *Eur. Phys. J.*, C49:1099–1116, 2007.

[150] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[151] CMS Collaboration. Jet algorithms performance in 13 TeV data. 2017.

[152] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur. Phys. J.*, C72:1896, 2012.

[153] Pileup Removal Algorithms. Technical Report CMS-PAS-JME-14-001, CERN, Geneva, 2014.

[154] Identification of b quark jets at the CMS Experiment in the LHC Run 2. Technical Report CMS-PAS-BTV-15-001, CERN, Geneva, 2016.

[155] Vardan Khachatryan et al. Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS. *JINST*, 11(01):P01019, 2016.

[156] Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2. Technical Report CMS-PAS-TAU-16-002, CERN, Geneva, 2016.

[157] CMS Collaboration. Tau energy scale and $\mu \to \tau$ misidentification rate estimated with early 2016 data using Z events. Jul 2016.

[158] CMS Collaboration. Performance of tau identification with 2016 data at $\sqrt{s} = 13$ TeV. Mar 2017.

[159] Measurement of the differential cross section for $t\bar{t}$ production in the dilepton final state at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-TOP-16-011, CERN, Geneva, 2016.

[160] Vardan Khachatryan et al. Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV. *Phys. Rev.*, D95(9):092001, 2017.

[161] Vardan Khachatryan et al. Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV. *Eur. Phys. J.*, C75(11):542, 2015.

[162] Vardan Khachatryan et al. Measurement of the $t\bar{t}$ production cross section in the all-jets final state in pp collisions at $\sqrt{s} = 8$ TeV. *Eur. Phys. J.*, C76(3):128, 2016.

[163] Serguei Chatrchyan et al. Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons. *JHEP*, 05:104, 2014.

[164] CMS Collaboration. Performance of missing transverse momentum reconstruction algorithms in proton-proton collisions at $\sqrt{s} = 8$ TeV with the CMS detector. CMS Physics Analysis Summary CMS-PAS-JME-12-002, 2012.

[165] Vardan Khachatryan et al. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *JINST*, 12(02):P02014, 2017.

[166] Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS. *JINST*, 6(arXiv:1107.4277. CMS-JME-10-011. CERN-PH-EP-2011-102):P11002. 67 p, Jul 2011.

[167] Lorenzo Bianchini, John Conway, Evan Klose Friis, and Christian Veelken. Reconstruction of the higgs mass in h $\tau\tau$ by dynamical likelihood techniques. *Journal of Physics: Conference Series*, 513(2):022035, 2014.

[168] D. Jang. *Search for MSSM Higgs decaying to tau pairs in $p\bar{p}$ collision at $\sqrt{s} = 1.96$ TeV at CDF*. PhD thesis, Rutgers University, 2006. FERMILAB-THESIS-2006-11.

[169] J.S. Conway. Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra. (arXiv:1103.0354):115–120. 6 p, Mar 2011. Comments: Presented at PHYSTAT 2011, CERN, Geneva, Switzerland, January 2011, to be published in a CERN Yellow Report.

[170] Iain W. Stewart and Frank J. Tackmann. Theory Uncertainties for Higgs and Other Searches Using Jet Bins. *Phys. Rev.*, D85:034011, 2012.

[171] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2016.

[172] Massimiliano Grazzini and Hayk Sargsyan. Heavy-quark mass effects in Higgs boson production at the LHC. *JHEP*, 09:129, 2013.

[173] ATLAS and CMS Collaborations, LHC Higgs Combination Group. Procedure for the LHC Higgs boson search combination in Summer 2011. Technical Report ATL-PHYS-PUB 2011-11, CMS NOTE 2011/005, CERN, 2011.

[174] Serguei Chatrchyan et al. Combined results of searches for the standard model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Lett. B*, 710:26, 2012.

[175] Albert M Sirunyan et al. Search for dark matter produced with an energetic jet or a hadronically decaying W or Z boson at $\sqrt{s} = 13$ TeV. *JHEP*, 07:014, 2017.

[176] Stefan Kallweit, Jonas M. Lindert, Philipp Maierhfer, Stefano Pozzorini, and Marek Schnherr. NLO electroweak automation and precise predictions for W+multijet production at the LHC. *JHEP*, 04:012, 2015.

[177] Stefan Kallweit, Jonas M. Lindert, Philipp Maierhofer, Stefano Pozzorini, and Marek Schnherr. NLO QCD+EW predictions for V + jets including off-shell vector-boson decays and multijet merging. *JHEP*, 04:021, 2016.

[178] Albert M Sirunyan et al. Search for new physics in events with a leptonically decaying Z boson and a large transverse momentum imbalance in proton-proton collisions at $\sqrt{s} = 13$ TeV. 2017.

[179] S. Kallweit, J. M. Lindert, S. Pozzorini, and M. Schnherr. NLO QCD+EW predictions for $2\ell2\nu$ diboson signatures at the LHC. *JHEP*, 11:120, 2017.

[180] Jon Butterworth et al. PDF4LHC recommendations for LHC Run II. *J. Phys.*, G43:023001, 2016.

[181] Measurement of the inclusive and differential tt production cross sections in lepton + jets final states at 13 TeV. Technical Report CMS-PAS-TOP-16-008, CERN, Geneva, 2016.

[182] Thomas Junk. Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth.*, A434:435–443, 1999.

[183] Alexander L. Read. Presentation of search results: The CL(s) technique. *J. Phys.*, G28:2693–2704, 2002. [,11(2002)].