A Study of the Standard Model Higgs Boson Decaying to a Pair of Tau Leptons with the CMS Detector at the LHC

by

Tyler Henry Ruggles

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 10 May 2018

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

Sridhara Dasu · Professor of Physics
Wesley H. Smith · Bjorn Wiik Professor of Physics
Matthew F. Herndon · Professor of Physics
Aki Hashimoto · Professor of Physics
Marshall F. Onellion · Professor of Physics and Materials Science

CERN-THESIS-2018-064
10/05/2018

© Copyright Tyler Henry Ruggles 2018

All Rights Reserved

Abstract

This thesis presents a 5.5 standard deviation observation of the Higgs boson decaying to fermions using the data collected at the LHC at 13 TeV center-of-mass energy. The studied dataset corresponds to an integrated luminosity of $35.9\,fb^{-1}$. The best fit signal strength for the $H \rightarrow \tau\tau$ process is measured to be $\mu = 1.24^{+0.29}_{-0.27}$, consistent with standard model predictions. Unique event categories are used targeting the leading Higgs boson production processes, gluon fusion, vector boson fusion, and associated production. This provides signal regions sensitive to Higgs boson couplings to both fermions and vector bosons. These two Higgs boson couplings are measured and are consistent with standard model predictions within one standard deviation. This 5.5 standard deviation observation of the $H \rightarrow \tau\tau$ process and the consistency of the Higgs boson couplings with the standard model provide confirmation of the Higgs boson Yukawa couplings to fermions. This is evidence that the Higgs field provides mass for the τ lepton in addition to the vector bosons.

Acknowledgements

I would like to thank the many people who made this thesis and my time in graduate school possible and enjoyable. Above all, I want to thank my wife, Caroline, who has been supportive of me through these five crazy, stressful, and amazing years of graduate school and life in Wisconsin and Switzerland. Thank you for all of your love and for everything you have done for me, for us, and for our child we are expecting in two months. You are wonderful.

My parents encouraged and enabled my curiosity and passion for science from a young age. Through their steadfast love and support these past 31 years, I have been able to pursue many interesting paths, both academic and in life. Thank you for everything.

Thank you to Peter and Betsy Hunt, my wife's parents, who have listened to me mention the Higgs boson and tau leptons a few too many times. Thank you to my brother Erik and my brother-in-law Will for entertaining these same conversations at family gatherings.

I want to thank my advisor Professor Sridhara Dasu for inviting me into the Wisconsin CMS group and for encouraging me despite my lack of previous training in computer programming. Thank you to Professors Wesley Smith and Matthew Herndon as well who have assembled a great team of graduate students, postdocs, research scientists, and engineers. It has been an interesting, fun, and great learning experience working as part of the Wisconsin CMS team.

Thank you to my favorite postdocs, Isobel Ojalvo, Cécile Caillol, and Maria Cepeda. You three have been invaluable resources, great support, and good friends these past years.

Thank you Alexander Savin for pushing me and always being ready to discuss

physics questions at all hours of the day or night even on weekends. Thank you Aleš Svetek for your good humor when being woken up in the middle of the night to check the optical power of the many links in our trigger system.

I want to thank all of the Wisconsin CMS graduate students of the past few years and current ones for great friendship, physics discussions, and life discussions. Specifically, thank you Laura Dodd for being great to work with and for making the office environment more fun. I know I was not always receptive to your bursts of chatter on Monday mornings. I appreciated them nonetheless. Thank you to Nick Smith and Kenneth Long who took French 101 and many physics courses with me and moved across the world to CERN with me. It has been a crazy trip. Also, thank you to Usama Hussain, Tom Perry, Aaron Levine, Devin Taylor, Nate Woods, and James Buchanan. It has been fun.

Contents

Abstract i

Acknowledgements ii

List of Figures **ix**

List of Tables **xxii**

1 Introduction **1**

1.1 The Standard Model of Particle Physics 3

1.2 The Standard Model: Experimental Context 7

2 Higgs Phenomenology **9**

2.1 Standard Model Symmetries 9

2.2 Electroweak Symmetry Breaking 11

2.3 Higgs Yukawa Couplings 12

2.4 Higgs Production 14

2.4.1 Gluon Fusion 15

2.4.2 Vector Boson Fusion 16

2.4.3 Associated Production 17

2.5 Higgs Boson Decays 17

2.6 Higgs Boson: Experimental Results 20

3 The CMS Experiment and the LHC 23

3.1 The LHC 23

3.1.1 LHC Proton Acceleration 25

3.1.2 LHC Magnets 26

3.2 The CMS Experiment 28

3.2.1 Geometry 30

3.2.2 Superconducting Magnet 31

3.2.3 Inner Tracking System 31

3.2.4 Electromagnetic Calorimeter 34

3.2.5 Hadronic Calorimeter 35

3.2.6 Muon System 38

3.2.7 Trigger and Data Acquisition 42

4 Simulation 48

4.1 Hard-Scattering Process 50

4.2 Parton Distribution Functions 51

4.3 Underlying Event 52

4.4 Parton Showers 52

4.5 Fragmentation and Hadronization 53

4.6 Pileup 55

4.7 Detector Simulation 55

5 Object Reconstruction and Selection 57

5.1 Particle Flow Input 57

5.1.1 Particle Flow Tracks 58

5.1.2 Particle Flow Energy Clusters 63

5.1.3 Particle Flow Candidates 65

5.2 Event Level Quantities 70
    5.2.1 Primary Vertex Reconstruction 71
    5.2.2 Missing Transverse Energy 71
5.3 Composite Object Identification and Selection 72
    5.3.1 Jets 72
    5.3.2 b-jet Identification 72
    5.3.3 Taus 73

6 Higgs → $\tau\tau$: Gluon Fusion and Vector Boson Fusion 77

6.1 Overview 78
    6.1.1 Event Selection 79
    6.1.2 Triggers 79
    6.1.3 Baseline Object Selection 80
    6.1.4 Categorization 84
6.2 Data Set 89
6.3 Monte Carlo Samples 89
6.4 Mass Reconstruction 90
6.5 Background Estimation 92
6.6 Monte Carlo Corrections 98
    6.6.1 Pileup Reweighting 99
    6.6.2 Tau Identification Efficiency 99
    6.6.3 Tau Energy Correction 100
    6.6.4 Lepton Identification and Isolation Efficiencies 100
    6.6.5 Trigger Efficiencies 100
    6.6.6 Drell–Yan Reweighting 102
    6.6.7 Hadronic Recoil Corrections 103

6.6.8 Generator Event Weights 104

6.6.9 Luminosity Weighting 104

6.7 Systematic Uncertainties 104

6.7.1 Integrated Luminosity Uncertainties 105

6.7.2 Object Reconstruction, Identification and Trigger Uncertainties 105

6.7.3 Tau Fake Rate 106

6.7.4 Energy Scales 106

6.7.5 b-Tagging Uncertainties 109

6.7.6 Background Estimation Uncertainties 109

6.7.7 Theoretical Uncertainties for Higgs Boson 113

6.7.8 Other Uncertainties 114

6.7.9 Systematic Model Summary 115

6.8 Results 117

6.8.1 Two-Dimensional Distributions 117

6.8.2 Results Extraction 121

6.8.3 Analysis Sensitivity Details 121

6.8.4 Mass Distributions 123

6.8.5 Signal Significance And Signal Strength 124

6.8.6 Impact of Uncertainties 125

7 Higgs $\to \tau\tau$: WH and ZH Associated Production 129

7.1 Overview 130

7.1.1 Triggers 131

7.1.2 Event Selection 132

7.1.3 Baseline Object Selection 135

7.2 Data Set 136

7.3 Monte Carlo Samples 136

7.4 Mass Reconstruction 138

7.5 Background Estimation 138

7.6 Monte Carlo Corrections 146

7.7 Systematic Uncertainties 146

7.7.1 Simulated Background Estimation Uncertainties 147

7.7.2 Reducible Background Estimation Uncertainties 147

7.7.3 Theoretical Uncertainties for Higgs Boson 149

7.8 Results 150

7.8.1 Signal Region Details 150

7.8.2 Analysis Sensitivity Details 151

8 Combined $H \to \tau\tau$ Results 158

8.1 Signal Strength and Significance 159

8.2 Higgs Boson Couplings 160

9 Conclusions 163

9.1 Future Studies 164

Bibliography 166

List of Figures

1.1 The fundamental particles of the SM and some of their properties including their: mass, electric charge, and spin. The units for mass are reported as electron volts divided by the speed of light (c) squared and use scientific notation prefixes. M for million, G for billion. 3

1.2 Diagram showing the bosons arranged into a central column with the fermions in the upper corners. The blue lines linking particles and groups of particles together indicate that those fermions can be influenced by force associated to that mediator boson. The Higgs boson plays a central role in the SM, coupling to all massive particles. 5

2.1 The potential $V(\phi)$ from Equation 2.3 showing a non-stable state at the origin and a stable state in the circular trough. 13

2.2 Feynman diagrams representing the leading Higgs boson production processes. Progressing in order of largest to smallest in production cross sections: (top left) ggH, (top right) VBF, (bottom left) WH and ZH processes, and (bottom right) $t\bar{t}H$. 15

2.3 The calculated Higgs boson production cross sections and their uncertainties as a function of the Higgs boson mass, are shown. The ggH process is denoted as pp → H in the figure. The CMS and ATLAS experiments have determined $m_H = 125.09\,\text{GeV}$ [1]. 16

2.4 The different theorized Higgs boson decay process are shown as a as a function of the Higgs boson mass. The CMS and ATLAS experiments have determined $m_H = 125.09\,\text{GeV}$ [1]. 18

2.5 The best fit values for the signal strength of the listed Higgs boson production processes and decay processes. A value of 1 indicates perfect agreement with the SM. The error bars indicate the 1σ intervals. The green shaded bands indicate the theoretical uncertainties in the predictions. . 22

3.1 Diagram of the LHC accelerator complex showing the five accelerators used to accelerate protons to their design collision energy of 7 TeV. For the LHC Run II, the highest energy achieved for protons is 6.5 TeV. . . 25

3.2 Cross section of an LHC superconducting dipole magnet showing the two beam pipes in the middle surrounded by the the superconducting coils. 27

3.3 A cutaway diagram of the CMS detector showing the many subdetectors within. 29

3.4 Schematic cross section of the CMS tracker. Each line represents a detector module. Double lines indicate back-to-back modules which deliver stereo hits. 32

3.5 Total thickness t of the inner tracker material expressed in units of interaction lengths λ_l (left) and radiation lengths X_0 (right), as a function of the pseudorapidity η. 33

3.6 Longitudinal view of the CMS detector depicting the ECAL subdetector. The crystals are inclined towards the interaction region. 35

3.7 Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters. The effective thickness of the HB detector increases with increasing $|\eta|$. 37

3.8 Layout of one quadrant of CMS. The four DT stations in the barrel (MB1-MB4, green), the four CSC stations in the endcap (ME1-ME4, blue), and the RPC stations (red) are shown. 38

3.9 Schematic of a drift tube cell showing drift lines leading to/from the anode wire and isochrones three of which are seen as the concentric lines surrounding the anode wire. The voltages applied to the electrodes are +3,600V for wires, +1,800V for strips, and -1,200 V for cathodes. 39

3.10 (left) Schematic of a CSC module composed of 7 trapezoidal panels with 6 gas gaps between the panels. The cut away of the top panel reveals the anode wires used to detect ionized molecules from a passing muon. The anode wires run horizontally and the cathode strips run vertically in this orientation. (right) Two diagram showing the inner dimensions of a CSC gas gap. The two views show how the orthogonal configuration of the anode wires and cathode strip can be used to localize the positions of a transversing muon. 40

3.11 Diagram of a double-chamber RPC. The gas chamber is 2 mm wide and is surrounded by two bakelite layers 2 mm thick each. 41

3.12 (left) System flow chart of the L1 Trigger, showing the complete trigger system, and input detector subsystems. (right) The L1 calorimeter trigger showing the number of cards and links between each system. 44

3.13 Monte-carlo VBF $H \to \tau\tau$ based efficiency distribution showing the cone-based HLT algorithm in red and the HPS-based HLT algorithm in blue. For nearly identical efficiency performance it is expected that the HLT rate can be reduced by up to 20%. 47

4.1 (left) Lund String Model flux tubes connecting a quark-antiquark pair. (right) A space-time representation of the hadronization of a quark-antiquark pair. The final representation depicts a seven mesons final state. 54

5.1 A schematic of a slice of an x-y cross section of the CMS detector showing different physics-objects such as electrons, photons, charged and neutral hadrons, and muons propagating outwards from the collision region within the detector. The schematic shows how tracks are linked to energy deposits and in which subdetectors different particles deposit most of their energy on average. 59

5.2 Efficiency (left) and misreconstruction rate (right) of the global combinatorial track finder (black squares) which is the first pass through the tracking algorithm. The prompt iterations of the tracking method (green triangles) show the results after all iterations based on seeds with at least one hit in the pixel detector are completed. The final results after all iterations (red circles) includes iterations with displaced seeds. Efficiency and misreconstruction rate are plotted as a function of the track p_{T}, for charged hadrons in multijet events without pileup interactions. Only tracks with $|\eta| < 2.5$ are considered. The efficiency is displayed for tracks originating from within 3.5 cm of the beam axis and ±30 cm of the nominal center of CMS along the beam axis. 61

5.3 (left) Calibration coefficients obtained from single $\mathrm{K_L^0}$s in the barrel as a function of their true energy E. The blue triangles show the calibrations for hadrons depositing energy only in the HCAL. The red circles (green squares) show the ECAL (HCAL) calibration for hadrons depositing energy in both the ECAL and HCAL. (right) Relative energy response (dashed curves) for the raw (blue) and calibrated (red) energy, and energy resolution (solid curves). Both for single $\mathrm{K_L^0}$s in the barrel as a function of their true energy E. 65

5.4 The reconstructed invariant mass of the τ_h candidate. A spike is seen at 140 MeV for the 1-prong τ_h decay mode where the mass is assigned equal to the mass of a $\pi^{\pm}$. The 1-prong+π^0 decay mode is seen to peak around 770 MeV while the 3-prong decay mode centers around 1260 MeV. . . . 74

5.5 Distance in η (left) and ϕ (right) between τ_h and e/γ, that are due to hadronic tau decay products, as a function of e/γ p_T. The size of the window is larger in the ϕ direction due to bending in the magnetic field. The dotted line shows the 95% quantile while the red line shows the fit to the 95% quantile. The red line is used to define the widths of the dynamic strip. 75

6.1 (Left) Diagram showing the construction of the p_ζ and p_ζ^{vis} projections. (Right) An example $p_\zeta - 0.85\,p_\zeta^{\mathrm{vis}}$ distribution is shown for a similar but not overlapping selection in the $e\mu$ channel. This distribution is from an analysis focusing on specifically studying the $Z \to \tau\tau$ process [2]. In the $H \to \tau\tau$ analysis, the Higgs boson $p_\zeta - 0.85\,p_\zeta^{\mathrm{vis}}$ spectrum aligns very closely with the $Z \to \tau\tau$ distribution shown here. 84

6.2 Two dimensional distributions for Higgs boson events passing event selection for each category are shown (left). The specific Higgs boson process shown is the one specifically targeted for each category. Distributions for select dominant background processes are shown (right). The rows correspond to the three categories: 0-jet (top), VBF (center), and Boosted (bottom). All distributions are from the $\mu\tau_h$ decay channel. 88

6.3 Mean number of interactions per bunch crossing for the 2016 proton-proton run at 13 TeV. The cross section is taken to be 80 mb. 91

6.4 the high-m_T control regions enriched in the $W + \mathrm{jets}$ background used in the maximum likelihood fit, together with the signal regions, to extract the results. The normalization of the predicted background distributions corresponds to the result of the global fit. These regions, defined with $m_\mathrm{T} > 80\,\mathrm{GeV}$, control the yields of the $W + \mathrm{jets}$ background in the $\mu\tau_h$ and $e\tau_h$ channels. 93

6.5 Schematic depicting the range of τ_h MVA working points, defined in Section 5.3.3, ranging from very loose (`VLoose`) to very tight (`VTight`). The signal region is depicted in orange where both τ_h meet the `Tight` or `VTight` criteria. The QCD multijet relaxed isolation estimation region is depicted in blue and does not overlap with the signal region. 96

6.6 Control regions enriched in the QCD multijet background used in the maximum likelihood fit, together with the signal regions, to extract the results. The normalization of the predicted background distributions corresponds to the result of the global fit. These regions, a formed by selecting events with opposite-sign τ_h candidates passing relaxed isolation requirements with at least one of them failing `Tight Tau MVA` isolation. These regions control the yields of the QCD multijet background in the $\tau_h\tau_h$ channel. . . 97

6.7 Control region enriched in $t\bar{t}$ background, used in the maximum likelihood fit, together with the signal regions, to extract the results. The normalization of the predicted background distributions corresponds to the result of the global fit. This region, defined by inverting the p_ζ requirement and rejecting events with no jet in the $e\mu$ final state, is used to estimate the yields of the $t\bar{t}$ background in all channels. 98

6.8 Comparison of the trigger efficiencies measured for a single leg of the double-τ_h with a muon+τ_h monitoring trigger. The efficiencies shown here are used in the $\tau_h\tau_h$ channel. The specific distribution shown is for genuine τ leptons and includes all used decay modes. 101

6.9 (Left) For reference, previously shown signal region and QCD multijet relaxed isolation estimation region. (Right) Schematic of the further loosened MVA isolation (defined in Section 5.3.3) selections used for the QCD multijet validation. The `Tighter Validation Region` (red) in the right plot is treated just like the `Signal Isolation Region` on the left plot. The `Looser Validation Region` (pink) in the right plot is treated just like the `Looser Isolation Region` on the left plot. 112

6.10 Observed and predicted 2D distributions in the VBF category of the $\tau_h\tau_h$ channel. The two variables are the mass, $m_{\tau\tau}$, and the di-jet mass, m_{jj}. The background uncertainty band accounts for all sources of background uncertainty. The signal is shown both as a stacked filled histogram and as an open overlaid non-stacked histogram. 118

6.11 Observed and predicted 2D distributions in the VBF category of the $\mu\tau_h$ channel. The description of the histograms is the same as in Figure 6.10. 118

6.12 Observed and predicted 2D distributions in the Boosted category of the $\tau_h\tau_h$ channel. The description of the histograms is the same as in Figure 6.10.119

6.13 Observed and predicted 2D distributions in the Boosted category of the $\mu\tau_h$ channel. The description of the histograms is the same as in Figure 6.10. 119

6.14 Observed and predicted distributions in the 0-jet category of the $\tau_h\tau_h$ channel. The description of the histograms is the same as in Figure 6.10. 120

6.15 Observed and predicted 2D distributions in the 0-jet category of the $\mu\tau_h$ channel. The description of the histograms is the same as in Figure 6.10. 120

6.16 Distribution of the decimal logarithm of the ratio between the best fit signal and the sum of the best fit signal and best fit background expectations in each bin of the mass distributions used to extract the results. All signal regions and channels are included. Background contributions are broken down by channel. The inset shows the corresponding difference between the observed data and best fit background distributions divided by the best fit background expectation. The best fit signal expectation is also divided by the background expectation in the inset. 122

6.17 Combined observed and predicted $m_{\tau\tau}$ distributions. The left pane includes the VBF category of the $\mu\tau_h$, $e\tau_h$ and $e\mu$ channels, and the right pane includes all other channels that make use of $m_{\tau\tau}$ instead of m_{vis} for the signal strength fit. The normalization of the predicted background distributions corresponds to the result of the global fit, while the signal is normalized to its best fit signal strength. The mass distributions for a constant range of the second dimension of the signal distributions are weighted according to $S/(S+B)$, where S and B are computed as the signal or background contribution in the mass distributions. The "Others" background contribution includes events from diboson, $t\bar{t}$, and single top quark production, as well as Higgs boson decay to a pair of W bosons and Z bosons decaying to a pair of light leptons. The background uncertainty band accounts for all sources of background uncertainty, systematic as well as statistical, after the global fit. The inset shows the corresponding difference between the observed data and expected background distributions, together with the signal expectation. The signal yield is not affected by the reweighting. 125

6.18 Local p-value and significance as a function of the SM Higgs boson mass hypothesis. The observation (red, solid) is compared to the expectation (blue, dashed) for a Higgs boson with a mass $m_H = 125.09\,\text{GeV}$. The background includes Higgs boson decays to pairs of W bosons, with $m_H = 125.09\,\text{GeV}$. 126

6.19 Best fit signal strength μ per category (left) and channel (right), for $m_H = 125.09\,\text{GeV}$. The constraints from the global fit are used to extract each of the individual best fit signal strengths. The combined best fit signal strength is $\mu = 1.09^{+0.27}_{-0.26}$. 126

6.20 1D likelihood scans which were used to extract the uncertainty associated with each of the four uncertainty components: theoretical uncertainties, bin-by-bin statistical uncertainties on the backgrounds, other systematic uncertainties, and the statistical uncertainty of the data gathered. The difference between the nominal scan in black and the scans represented as dotted lines show the effect of freezing out certain uncertainty components from the analysis. Freezing out an uncertainty component shows how the analysis results would change if there was zero uncertainty associated with that component. For example, the green dotted curve shows how the analysis results would change if theoretical Higgs boson uncertainties were reduced to zero. 128

7.1 Pre-fit p_{T} distributions showing statistical uncertainties only for the four leptons in the $\ell\ell e\tau_h$ final states. (top left) and (top right) leading and subleading ℓ from Z, (bottom left) and (bottom right) e and τ_h from Higgs boson candidate. The WH and ZH signals are summed as $VHiggs$ and multiplied by a factor of 10 times their SM expected yield. 142

7.2 Pre-fit p_{T} distributions showing statistical uncertainties only for the four leptons in the $\ell\ell\mu\tau_h$ final states. (top left) and (top right) leading and subleading ℓ from Z, (bottom left) and (bottom right) μ and τ_h from Higgs boson candidate. The WH and ZH signals are summed as $VHiggs$ and multiplied by a factor of 10 times their SM expected yield. 143

7.3 Pre-fit p_{T} distributions showing statistical uncertainties only for the four leptons in the $\ell\ell\tau_h\tau_h$ final states. (top left) and (top right) leading and subleading ℓ from Z, (bottom left) and (bottom right) leading and subleading τ_h from Higgs boson candidate. The WH and ZH signals are summed as $VHiggs$ and multiplied by a factor of 10 times their SM expected yield. 144

7.4 Pre-fit p_{T} distributions showing statistical uncertainties only for the four leptons in the $\ell\ell e\mu$ final states. (top left) and (top right) leading and subleading ℓ from Z, (bottom left) and (bottom right) e and μ from Higgs boson candidate. The WH and ZH signals are summed as $VHiggs$ and multiplied by a factor of 10 times their SM expected yield. 145

7.5 The postfit $m_{\tau\tau}$ distributions used to extract the signal shown for the (top left) $eee\tau_h$, (top right) $\mu\mu e\tau_h$, (bottom left) $ee\mu\tau_h$, and (bottom right) $\mu\mu\mu\tau_h$ final states. The distributions show full uncertainties. The WH and ZH, $H \rightarrow \tau\tau$ signal processes are summed together and shown as VH, $H \rightarrow \tau\tau$ with a best-fit $\mu = 2.5$. VH, $H \rightarrow \tau\tau$ is shown both as a stacked filled histogram and an open overlaid histogram. In these distributions the ZH, $H \rightarrow \tau\tau$ process contributes more than 99% of the total of VH, $H \rightarrow \tau\tau$. 151

7.6 The postfit $m_{\tau\tau}$ distributions used to extract the signal shown for the (top left) $ee\tau_h\tau_h$, (top right) $\mu\mu\tau_h\tau_h$, (bottom left) $eee\mu$, and (bottom right) $\mu\mu e\mu$ final states. The distributions show full uncertainties. The WH and ZH, $H \rightarrow \tau\tau$ signal processes are summed together and shown as VH, $H \rightarrow \tau\tau$ with a best-fit $\mu = 2.5$. VH, $H \rightarrow \tau\tau$ is shown both as a stacked filled histogram and an open overlaid histogram. In these distributions the ZH, $H \rightarrow \tau\tau$ process contributes more than 99% of the total of VH, $H \rightarrow \tau\tau$. 152

7.7 The postfit $m_{\tau\tau}$ distributions used to extract the signal shown for all 8 ZH final states combined. The distribution shows full uncertainties. The left half of the distribution is the Low-$L_{\mathrm{T}}^{\mathrm{Higgs}}$ region while the right half corresponds to the High-$L_{\mathrm{T}}^{\mathrm{Higgs}}$ region. The definitions of the $L_{\mathrm{T}}^{\mathrm{Higgs}}$ regions in this distribution are the same as those used in Figures 7.5 and 7.6 and are final state dependent. The WH and ZH, $H \to \tau\tau$ signal processes are summed together and shown as VH, $H \to \tau\tau$ with a best-fit $\mu = 2.5$. VH, $H \to \tau\tau$ is shown both as a stacked filled histogram and an open overlaid histogram. In this distribution the ZH, $H \to \tau\tau$ process contributes more than 99% of the total of VH, $H \to \tau\tau$. 153

7.8 Postfit mass distributions in the $e\mu\tau_h$ (top left), $\mu\mu\tau_h$ (top right), $e\tau_h\tau_h$ (bottom left), and $\mu\tau_h\tau_h$ (bottom right) final states. The distributions show full uncertainties. The WH and ZH, $H \to \tau\tau$ signal processes are summed together and shown as VH, $H \to \tau\tau$ with a best-fit $\mu = 2.5$. VH, $H \to \tau\tau$ is shown both as a stacked filled histogram and an open overlaid histogram. In these distributions the WH, $H \to \tau\tau$ process contributes between 91%–93% of the total of VH, $H \to \tau\tau$. 155

7.9 Postfit mass distributions of the four WH final states combined together. The distributions show full uncertainties. The WH and ZH, $H \to \tau\tau$ signal processes are summed together and shown as VH, $H \to \tau\tau$ with a best-fit $\mu = 2.5$. VH, $H \to \tau\tau$ is shown both as a stacked filled histogram and an open overlaid histogram. In this distribution the WH, $H \to \tau\tau$ process contributes 92% of the total of VH, $H \to \tau\tau$. 156

7.10 Distribution of the decimal logarithm of the ratio between the expected signal, corresponding to the best fit value $\mu = 2.5$, and the sum of expected signal and expected background in each bin of the mass distributions used to extract the results, in all signal regions. The background contributions are separated based on the final states, WH versus ZH. The inset shows the corresponding difference between the observed data and expected background distributions divided by the background expectation, as well as the signal expectation divided by the background expectation. 157

8.1 Best fit signal strength per Higgs boson production process, for $m_H = 125.09\,\text{GeV}$. The constraints from the combined global fit are used to extract each of the individual best fit signal strengths. The combined best fit signal strength is $\mu = 1.24^{+0.29}_{-0.27}$. 160

8.2 Scan of the negative log-likelihood difference as a function of κ_V and κ_f, for $m_H = 125.09\,\text{GeV}$. All nuisance parameters are profiled for each point. This scan is a combination of the ggH and VBF targeted analysis with the WH and ZH targeted analysis. For reference, the results for just the ggH and VBF targeted analysis are also presented and correspond to the updated ggH modeling. For this scan, all $H \to WW$ and $H \to ZZ$ processes are treated as signal. 162

List of Tables

1.1 The fundamental forces, their mediator particles, and the relative strength of the force. There has been no observed mediator for the gravitational force. 5

3.1 LHC beam characteristics after the LHC has increased the energy of the proton beam from the input 450 GeV to the target collision energy. . . . 28

5.1 The clustering parameters used for ECAL and HCAL energy deposit clustering. The ECAL endcap requires an additional seed E_{T} threshold because the detector noise increases as a function of $|\eta|$. 64

5.2 Decay modes for τ^- leptons including leptonic decays and hadronic decays. The $h^{\pm}$ stand for $\pi^{\pm}$ or $K^{\pm}$. Inverting all of the "-" for "+" will give the decay modes for τ^+ leptons. Nearly 65% of τ leptons decay hadronically to τ_h. 74

6.1 For each channel the HLT p_{T} threshold and $|\eta|$ range is listed along with a description of the HLT path where "iso" stands for "isolated." Changes in the available triggers with respect to the data collection era are noted. 81

6.2 Kinematic, identification and isolation selection requirements for the four di-τ channels. 82

6.3 Category selection and observables used to build the 2D kinematic distributions. The events failing the 0-jet and VBF selection are included in the Boosted category and are denoted by “Others”. 87

6.4 NLO cross sections for considered backgrounds. In this table, ℓ represents all three generations of charged leptons, e,μ,τ. In some cases the production mechanism is listed: quarks (qq) versus gluons (gg). 90

6.5 The category dependent scale factors used to adjust the QCD multijet yield to correspond to the expected yield in the signal region. The large uncertainty on the VBF scale factor is due to the limited amount of QCD multijet events in the VBF same-sign region. 97

6.6 Energy corrections applied to simulated genuine τ_h. The energy corrections are measured and applied depending on the reconstructed decay mode of the τ_h. 100

6.7 QCD multijet uncertainties for the $\tau_h\tau_h$ channel for each category. The “Estimate” / “Predicted” column is the closure found for the QCD multijet estimation method validation. The statistical uncertainty on the closure value is found from adding in quadrature the statistical uncertainty of each of the four regions used in the comparison (SS vs. OS and Looser vs. Tighter Isolation). The QCD method systematic uncertainty is calculated as the deviation from unity for the closure summed with the statistical uncertainty on the closure test. The QCD method statistical uncertainty is derived from the uncertainty in the same-sign regions used to calculate the loose-to-tight scale factor. The final uncertainty values used in the analysis, the QCD total uncertainty is the sum in quadrature of the QCD method systematic uncertainty and QCD method statistical uncertainty. 112

6.8 Sources of systematic uncertainty. If the global fit to the signal and control regions, described in the next section, significantly constrains these uncertainties, the values of the uncertainties after the global fit are indicated in the third column. The acronyms CR and ID stand for control region and identification, respectively. 116

6.9 Best-fit background and signal expectations, together with the number of observed events, for highly sensitive bins in the signal region. High sensitivity bins are defined by $\log_{10}(S/(S+B)) > -0.9$, where S and B are the number of best fit expected signal events for a Higgs boson with a mass $m_H = 125.09\,\text{GeV}$ and of best fit expected background events. The background uncertainty accounts for all sources of background uncertainty, systematic as well as statistical, after the global fit. The contribution from "other backgrounds" includes events from diboson and single top quark production. The contribution from Higgs boson decays to a pair of W bosons is zero in these bins. 123

7.1 Kinematic selection requirements for WH and ZH events. The trigger requirement is defined by a combination of trigger candidates with p_{T} over a given threshold (in GeV), indicated inside parentheses. The pseudorapidity thresholds come from trigger and object reconstruction constraints. The trigger requirements for the ZH events are defined by the Z boson decay products, either $Z \to ee$ or $Z \to \mu\mu$. 132

7.2 Electron, muon, and τ_h selection criteria for each final state in the associated production $H \to \tau\tau$ analysis. In the $e\mu\tau_h$ final state there are two different working points listed for electron and muon rejection. Anti-e VL, anti-μ T applies for events where the electron and τ_h are same charge and anti-e T, anti-μ L applies for events where the muon and τ_h are same charge. 137

7.3 NLO cross sections for considered backgrounds. In this table, ℓ represents all three generations of charged leptons, e,μ,τ. In some cases the production mechanism is listed: quarks (qq) versus gluons (gg). 139

7.4 Sources of systematic uncertainty 149

7.5 Background and signal expectations for the WH final states, together with the number of observed events, for the post-fit signal region distributions. The background uncertainty accounts for all sources of background uncertainty, systematic as well as statistical, after the global fit. The contribution from "Rare" includes events from triboson, $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}H$ production, and other rare processes. 154

7.6 Background and signal expectations for the ZH final states, together with the number of observed events, for the post-fit signal region distributions. The ZH final states are each grouped according to the Higgs boson decay products. $\ell\ell$ covers both $Z \to \mu\mu$ and $Z \to ee$ events. The background uncertainty accounts for all sources of background uncertainty, systematic as well as statistical, after the global fit. The contribution from "Rare" includes events from triboson, $t\bar{t}Z$, $t\bar{t}H$ production, and other rare processes. 154

8.1 Best fit signal strength and significance for three fit scenarios: ggH and VBF, associated production, and the combination with the updated ggH modeling. 159

Chapter 1

Introduction

This thesis presents a 5.5 standard deviation observation of the Higgs boson decaying to a pair of tau leptons ($\tau^+\tau^-$). The Higgs boson couplings to fermions and vector bosons are measured and are consistent with standard model predictions within one standard deviation. The observation of the Higgs boson decaying to $\tau^+\tau^-$ and the consistency of the Higgs boson couplings with the standard model provide confirmation of the Higgs boson Yukawa couplings to fermions. This is evidence that the Higgs field provides mass for the τ lepton in addition to the vector bosons.

This study includes two analyses which target different ways the Higgs boson is produced [3, 4]. The analyses are performed using 13 TeV center-of-mass energy proton-proton collision data from the CERN LHC. The data is collected by the CMS experiment and corresponds to an integrated luminosity of 35.9 fb^{-1}. The results here constitute an important milestone in the effort to better understand the fundamental properties of nature and the Higgs boson, one of the fundamental particles of the standard model (SM) of physics.

The standard model of physics is a mathematical framework for explaining the interactions and behavior of the fundamental particles observed in nature. It has been built up and defined through the 1950s and 60s culminating in the theoretical prediction of the existence of a neutral scalar boson, now called the Higgs boson. The SM incorporates descriptions of three of the four fundamental forces of nature: the strong force, the electromagnetic force, and the weak force.

The Higgs boson eluded observation by experimental particle physicists for 40 years after the establishment of its theoretical prediction. In 2012, the Higgs boson was discovered by the CMS and ATLAS collaborations at CERN [5, 6, 7]. With this discovery, all particles predicted and described in the SM have been observed. Based on research leading up to today, the SM is the best tested theory of nature at the fundamental level. Overall, the SM shows remarkable consistency between theoretical predictions and the resulting experimental observations.

Since the discovery of the Higgs boson, the focus of the experimental particle physics community has transitioned from Higgs boson "discovery" mode to Higgs boson "measurement" mode. High energy particle physics experiments are dedicating a vast portion of their research effort and person power towards efforts to measure the Higgs boson properties as precisely as possible. Many of of these properties are firmly predicted by theory. Affirmation or negation of these predictions, such as how often a Higgs boson decays into a pair of τ leptons, are critical to further testing the merits of the SM. Confirmation of the predicted Higgs boson properties would further support the SM, along with the myriad previous experimental results. Significant discrepancies between the SM theoretical predictions and the observed Higgs boson properties could point to flaws in the SM and would lead to a more full and complete understanding of nature.

1.1 The Standard Model of Particle Physics

The standard model of particle physics [8, 9, 10] is currently the best mathematical framework for predictions and explanations of the behavior of the fundamental particles of nature, excluding gravitational interactions. Fundamental particles can be grouped together by their characteristics and arranged diagrammatically as in Figure 1.1. The particles are split vertically into categories: fermions which have spin $\frac{1}{2}$, the vector bosons which have spin 1, and the Higgs boson with spin 0. In general, the fermions constitute matter while the vector bosons are the mediators of the fundamental forces.

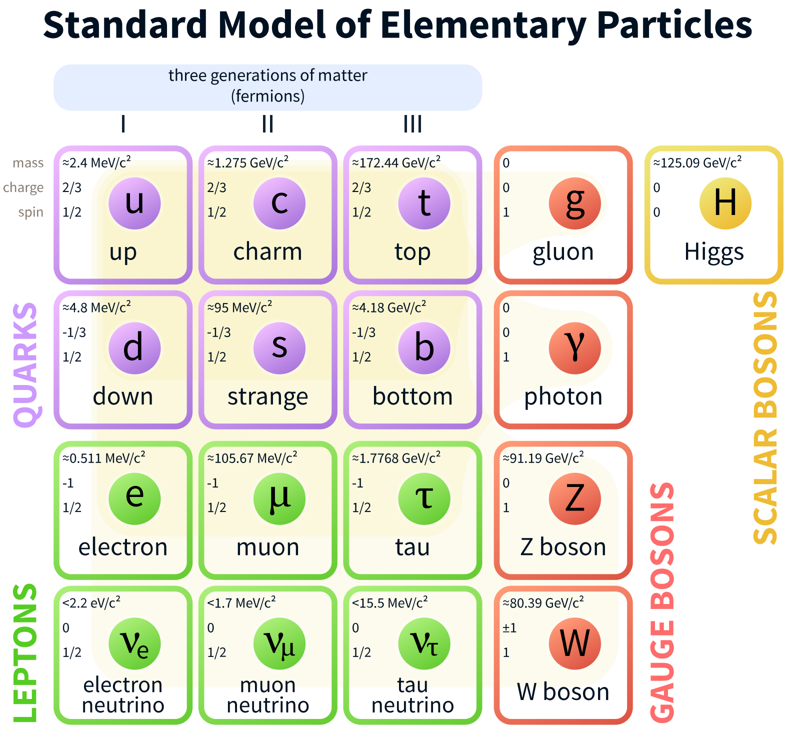

Standard Model of Elementary Particles
three generations of matter (fermions)
I
II
III
mass
charge
spin
≈2.4 MeV/c²
2/3
1/2
u
up
≈1.275 GeV/c²
2/3
1/2
c
charm
≈172.44 GeV/c²
2/3
1/2
t
top
0
0
1
g
gluon
≈125.09 GeV/c²
0
0
H
Higgs
≈4.8 MeV/c²
-1/3
1/2
d
down
≈95 MeV/c²
-1/3
1/2
s
strange
≈4.18 GeV/c²
-1/3
1/2
b
bottom
0
0
1
γ
photon
≈0.511 MeV/c²
-1
1/2
e
electron
≈105.67 MeV/c²
-1
1/2
μ
muon
≈1.7768 GeV/c²
-1
1/2
τ
tau
≈91.19 GeV/c²
0
1
Z
Z boson
<2.2 eV/c²
0
1/2
νe
electron neutrino
<1.7 MeV/c²
0
1/2
νμ
muon neutrino
<15.5 MeV/c²
0
1/2
ντ
tau neutrino
≈80.39 GeV/c²
±1
1
W
W boson
QUARKS
LEPTONS
GAUGE BOSONS
SCALAR BOSONS

Figure 1.1: The fundamental particles of the SM and some of their properties including their: mass, electric charge, and spin. The units for mass are reported as electron volts divided by the speed of light (c) squared and use scientific notation prefixes. M for million, G for billion.

The fermions can be further grouped into either quarks or leptons based on whether they interact strongly or not. Fundamental particles that carry a charge can interact via the force associated with that charge. Thus, strong interactions of the quarks are described in the SM by assigning them what is called a color charge. Quarks have a fractional electric charge of either $-\frac{1}{3}e$ or $\frac{2}{3}e$ where e is the elementary charge (1.602×10^{-19} Coulombs). Where as the leptons are colorless (carry no color charge) and have integer electric charge of $0e$ or $-1e$. In Figure 1.1 the fermions are arranged according to what is called their mass "generation" with more massive particles appearing to the right in the third mass generation column.

The first mass generation column composes the fundamental particles which make up the ordinary matter interact with every day. Up-quarks and down-quarks are the fundamental constituents within protons and neutrons which build the atoms and molecules making up the paper pages of this thesis or the metals and plastics in your computer. Electrons are the remaining fundamental particles we are familiar with and are also part of the basic structure of atoms. The electron neutrino is less familiar because it only interacts with the other particles through the weak force and does not directly contribute to the basic stable atoms.

In addition to the particles shown in Figure 1.1, there exist antiparticles. Each SM particle has an antiparticle, though some particles, such as the photon are their own antiparticle. Antiparticles have the same mass as their "normal" particle pair except they have opposite: electric charge, color charge, and weak charge. The antiparticle partner of the electron is the positron which is sometimes called an antielectron. Antiparticles can be created in many types of interactions in particle physics experiments and are relatively common. One characteristic of antiparticles is that when a particle and its corresponding antiparticle are in close proximity they can annihilate each other resulting in a burst of energy. Antiparticles are denoted in this thesis with a “bar” over the top of a particle symbol. For example a top-quark is t while an antitop-quark is $\bar{t}$.

The fundamental forces, their mediator particles, and the relative strength of that force are listed in Table 1.1. The relatively weak strength of the gravitational force is what allows the SM to still successfully predict the behavior of particles despite not including the gravitational force.

Fundamental Force	Force Mediator	Relative Strength
Strong	gluon (g)	1
Electromagnetic	photon (γ)	10^{-3}
Weak	W and Z bosons	10^{-14}
Gravitational	unknown	10^{-43}

Table 1.1: The fundamental forces, their mediator particles, and the relative strength of the force. There has been no observed mediator for the gravitational force.

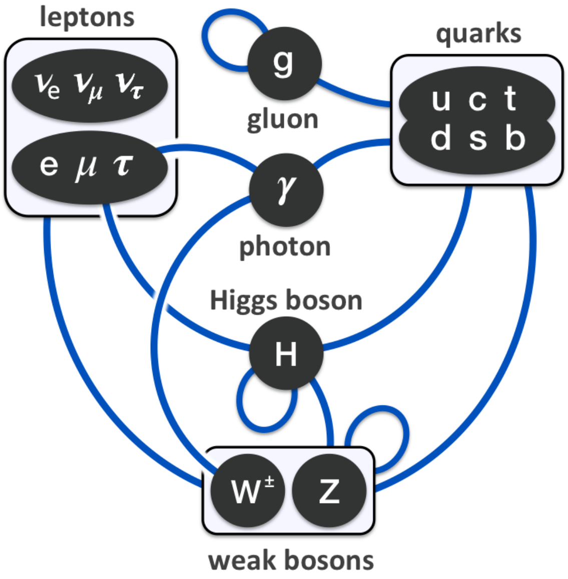


Figure 1.2: Diagram showing the bosons arranged into a central column with the fermions in the upper corners. The blue lines linking particles and groups of particles together indicate that those fermions can be influenced by force associated to that mediator boson. The Higgs boson plays a central role in the SM, coupling to all massive particles.

The strong force has the largest relative strength of the fundamental forces but the reach, or distance over which the force can be felt, is very limited and is confined to the sub-atomic scale, 10^{-15}m. The strong force is experienced between particles with a color charge, exclusively gluons and quarks. Figure 1.2 shows a diagram of the SM particles with lines linking force mediating bosons with the particles which experience that force. For example, a line links the quarks with the gluons representing the strong force.

The electromagnetic force follows after the strong force in order of largest relative strength. The reach of the electromagnetic force decreases with distance as $\frac{1}{r^2}$. Despite its infinite reach, the electromagnetic force is not experienced on the macroscopic scale because stable matter is composed of equal portions positively charged and negatively charged matter. This leads to an overall neutral electrical charge for the universe. The electromagnetic force is experienced by all electrically charged particles: quarks, the charged leptons, and the $W^{\pm}$ bosons. This force is mediated by the photon which has neutral electric charge.

The next force in descending order of relative strength is the weak force. The weak force is experienced by all of the leptons and the quarks and is mediated by the $W^{\pm}$ and Z bosons. It is responsible for familiar phenomena such as the radioactive decay of atoms. Beta decay is one example of radioactive decay where, within an atomic nucleus, a neutron is transformed into a proton and an electron and an electron antineutrino (more on antiparticles following). Fundamentally, what happens is the quark composition within the proton changes, thereby changing the proton to a neutron. This process is mediated by a W^{-} boson which subsequently decays to an electron and the antielectron neutrino.

The final and weakest force is the one we are most familiar with, the gravitational force. Just like the electromagnetic force, the reach of the gravitational force is infinite and decreases with distance as $\frac{1}{r^2}$. Yet, unlike the electromagnetic force, gravity is felt over extremely large distances. This is because gravity is a purely attractive force which acts on all massive particles. The force we are most colloquially familiar with is actually the weakest of the four fundamental forces. Because the gravitational force is so weak, when talking about the effect on a single particle, it can be ignored for all of the particle physics calculations throughout this thesis.

1.2 The Standard Model: Experimental Context

There was a gap of 40 years between the prediction of the Higgs boson and its discovery. Many of the particles making up the SM were not discovered when the Higgs boson was originally being theorized. In fact, the existence of quarks or the discovery of gluons, the mediator of the strong force, were still to happen. The same is true for the W and the Z bosons, the mediators of the weak force. The decades after the 1960s saw discovery after discovery, slowly piecing together and validating the SM.

The internal structure of protons was illuminated by deep inelastic scattering experiments carried out at SLAC which eventually led to the observation of the three least massive quarks: up (u), down (d), and strange (s) in 1969 [11, 12]. In 1974, the J/Ψ particle, a composite particle made from a charm quark (c) and a charm anti-quark ($\bar{c}$) was discovered [13, 14]. The bottom quark (b) was discovered in 1977 via the decays of a new particle, the Upsilon meson [15]. The top quark (t) was the last quark of the three known generations discovered and was not found until 1995 at Fermilab [16, 17]. The gluon which mediates the strong force for all of the quarks was discovered in 1979 at DESY [18].

Beyond the partons, physicists discovered the weak neutral current [19, 20] which led to the discoveries of new bosons, specifically the mediators of the weak force. In 1983 the W and the Z bosons were discovered [21, 22]. These two bosons were the most massive fundamental particles at the time of their discovery with masses of 84.4 GeV and 91.2 GeV, respectively. A very important discovery laying the foundation for the analyses in this thesis is the discovery of the third generation τ lepton, which was discovered in 1975 by Martin Perl [23]. Following 25 years after the discovery of the τ lepton was the discovery of the τ neutrino by the DONUT collaboration [24]. While far from an exhaustive list, these many discoveries give an indication of the strong background of experimental research supporting the SM.

As more pieces of the SM fell into place and particle accelerators became more powerful, searches for the Higgs boson were conducted at multiple experiments in the 1990s [25, 26, 27, 28] and early 2000s [29, 30]. Before discussing these experimental results and those that followed leading to the discovery of the Higgs boson by the CMS and ATLAS collaborations at CERN in 2012 [5, 6, 7], I first provide a more detailed discussion of the Higgs boson itself in Chapter 2.

Chapter 2

Higgs Phenomenology

Electromagnetic and weak interactions are described using very similar relativistic quantum mechanics theory and can be described in a single unified electroweak theory. Despite this, these forces are manifestly non-symmetric in their behavior. The photon traverses the universe at the highest speeds possible while the W and Z bosons decay before they traverse subnuclear distances. The theoretical framework which breaks the symmetry of the electroweak theory (EWSB) is the Brout–Englert–Higgs mechanism which results in masses for the W and Z bosons, confining the weak force to subnuclear distances, while leaving the photon massless and capable of traversing great distances. The EWSB also predicts a massive scalar boson, the Higgs boson, whose properties are entirely determined by the parameters of the SM, except for its mass and self-coupling.

2.1 Standard Model Symmetries

Many people are attracted to physics because of the beauty they see in the patterns of nature. PW Anderson, a physics Nobel laureate, stated, “it is only slightly overstating the case to say that physics is the study of symmetry" [31]. One way of expressing many types of patterns mathematically is with symmetries, properties which remain invariant under certain transformations. Noether's theorem, first proven in 1915, states that symmetries result in conserved quantities [32]. The SM is defined by a group of symmetries representing certain conserved properties. Specifically, the properties of the fundamental particles can be related to representations of Lie algebras [33]. Using Lie algebra, the SM is represented by the symmetries of the unitary product group $\mathrm{SU}(3)_\mathrm{C} \times \mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y}$. These three terms all reflect symmetries and conserved quantities discussed below.

Relating back to Noether's theorem, each of the symmetries above describes a unique conserved quantity. The matter fields are invariant under phase rotations. This rotational symmetry is described by the U(1) group. Hypercharge, denoted as Y, is the conserved quantity for this group.

Pairs of matter fields, such as the electron and the same-flavor neutrino, the electron-neutrino, can transform into each other through weak interactions. This is described through the SU(2) symmetry group. Each left-handed charged lepton and the same-flavor neutrino form a weak isospin doublet which couples to the W and Z bosons, the mediators of the weak force, allowing these transformations. Weak isospin is the conserved quantity of the weak interactions.

Quantum chromodynamics (QCD) provides a description of the strong force and the interactions between the particles with a color charge, the quarks and gluons. In QCD each quark and anti-quark is represented by a color triplet. Strong interactions are described by $\mathrm{SU}(3)_\mathrm{C}$ where the "C" stands for color denoting the conserved quantity of strong interactions, color charge.

2.2 Electroweak Symmetry Breaking

Glashow, Weinberg, and Salam demonstrated a unified description of the electromagnetic force and weak force which merge at energies of order 100 GeV into the electroweak force, described by the electroweak Lagrangian [8, 9, 10]. The electroweak Lagrangian defines a gauge field theory which is invariant under the $\mathrm{SU}(2)_\mathrm{L} \times \mathrm{U}(1)_\mathrm{Y}$ symmetry group. Electric charge (Q), the third component of weak isospin (T_3), and hypercharge (Y) act together to define a conserved quantity in the SM, $Q = T_3 + \frac{Y}{2}$. This is represented by the hypercharge $\mathrm{U}(1)_\mathrm{Y}$ symmetry group.

Prior to the introduction of Electroweak Symmetry Breaking (EWSB), the electroweak Lagrangian describes massless W and Z bosons. However, as previously mentioned, the W and Z bosons discovered in 1983 are two of the most massive particles [21, 22]. The EWSB mechanism rescues what would be a colossal disagreement between theoretical prediction and experimental results. The introduction of EWSB to the theory preserves the structure of the electroweak interactions and succeeds in endowing the W and Z bosons with mass.

Electroweak symmetry breaking is an application of spontaneous symmetry breaking to the electroweak Lagrangian. It is achieved via the Brout–Englert–Higgs mechanism [34, 35, 36, 37, 38, 39], leading, in its minimal version, to the prediction of the existence of one physical neutral scalar particle, commonly known as the Higgs boson. The EWSB mechanism proposes a self-interacting complex doublet scalar field,

$$\phi = \begin{pmatrix} \phi_\alpha \\ \phi_\beta \end{pmatrix} = \sqrt{\frac{1}{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \tag{2.1}$$

that is applied to the electroweak Lagrangian,

$$\mathcal{L} = \left(D_\mu \phi\right)^\dagger \left(D^\mu \phi\right) - V(\phi), \tag{2.2}$$

where D^{μ} is the covariant derivative and the potential, $V\big(\phi\big)$, is expressly defined with two terms,

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

The Lagrangian describes a system of four scalar particles, the ϕ_i, of equation 2.1 each with mass μ. To achieve EWSB, the constants in the potential $V\big(\phi\big)$ are defined such that $\mu^2 < 0$ and $\lambda > 0$. The choice $\mu^2 < 0$ and $\lambda > 0$ lead to a potential $V\left(\phi\right)$ as shown in Figure 2.1. Expanding around a chosen minima of the potential $V\big(\phi\big)$, v, and taking $\phi_1 = \phi_2 = \phi_4 = 0$ and $\phi_3^2 = -\frac{\mu^2}{\lambda} \equiv v^2$, ϕ becomes

$$\phi = \sqrt{\frac{1}{2}}\begin{pmatrix} 0 \\ v \end{pmatrix}, \tag{2.4}$$

where the scalar doublet, ϕ acquires a non-zero vacuum expectation value (VEV), $v \approx 246\,\text{GeV}$. In this process a neutral and two charged massless Goldstone bosons [40] are generated. These Goldstone bosons mix with the fields corresponding to the broken $\text{SU}(2)_\text{L} \times \text{U}(1)_\text{Y}$ symmetries giving masses to the W and Z bosons. The single remaining component of the original complex doublet ϕ becomes a new fundamental scalar particle known as the Higgs boson with a mass

$$m_H^2 = 2\lambda v^2, \tag{2.5}$$

but the value of this is not predicted.

2.3 Higgs Yukawa Couplings

In addition to providing a mechanism for the W and Z bosons to gain mass in the SM, the EWSB mechanism also provides a mechanism for the fermions to acquire mass through their interactions with the Higgs boson via Yukawa couplings. A Yukawa

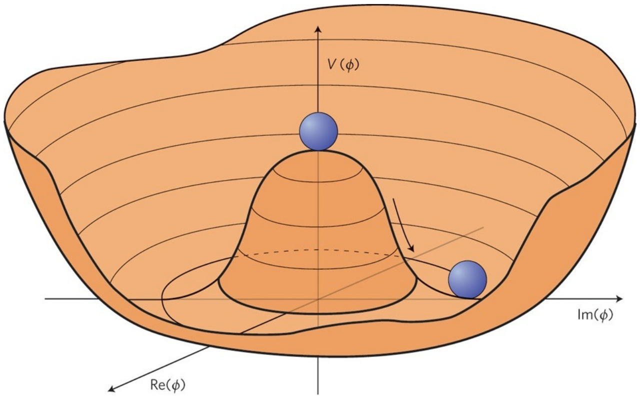


Figure 2.1: The potential $V(\phi)$ from Equation 2.3 showing a non-stable state at the origin and a stable state in the circular trough.

coupling is an interaction between a scalar field and a Dirac field similar to,

$$V \approx h_f \bar{\psi}_f \phi \psi_f, \tag{2.6}$$

where the Dirac fields, ψ, describe fermions and the scalar field, ϕ, is taken to be that of the Higgs boson. The introduction of a Yukawa interaction linking together the fermions and the Higgs boson, results in massive fermions where their mass can be written as,

$$m_f = \frac{h_f v}{\sqrt{2}}, \tag{2.7}$$

which covers the masses for the nine charged fermions: the three charged leptons and six quarks. Neither the EWSB mechanism nor the Yukawa interaction provide insight into the larger variety of fermion masses. Instead, the fermion masses are taken as free parameters of the SM and the values h_f represent the Yukawa coupling parameter for each of the fermions.

The mass of nine charged fermions are known experimentally, most to very high precision. The τ lepton is known to be $1776.82 \pm 0.16\,\mathrm{MeV}$ [41]. Knowing this, the τ lepton to Higgs boson Yukawa coupling can be calculated from Equation 2.7

and compared against experimental results. Exploring the Higgs boson to fermion decay processes is the most promising way to directly probe the Higgs boson Yukawa couplings. The Higgs boson decay processes are discussed below in Section 2.5 where this discussion continues.

2.4 Higgs Production

The Higgs boson is produced through interactions with the particles it couples to in the SM Lagrangian. Understanding the different production mechanisms for the Higgs boson allows experimentalists to search for unique signatures in collisions to better help separate Higgs bosons from the myriad other interactions occurring in collisions at the LHC. At a hadron collider like the LHC, the relevant Higgs boson production mechanisms begin from initial states of quarks and gluons. The Higgs boson only couples to massive particles, eliminating direct gluon to Higgs boson processes as is seen in Figure 1.2. However, the Higgs boson production process at the LHC which occurs most often begins with two gluons in the initial state and is called gluon fusion discussed below.

Feynman diagrams of the leading Higgs boson production processes for proton-proton based colliders are shown in Figure 2.2. The cross sections for the leading Higgs boson production processes for a center-of-mass energy of 13 TeV are shown in Figure 2.3. The values for the leading processes are approximately: gluon fusion (ggH) $\approx 49\,\mathrm{pb}$, vector boson fusion (VBF) $\approx 3.8\,\mathrm{pb}$, W boson associated production (WH) $\approx 1.4\,\mathrm{pb}$, and Z boson associated production (ZH) $\approx 0.88\,\mathrm{pb}$ [42]. The top-quark associated process ($t\bar{t}H$), which is found to be relatively insignificant in the following analyses, has a cross section approximately half the size of the next smallest one, ZH, where $t\bar{t}H \approx 0.51\,\mathrm{pb}$ and accounts for less than 1% of the total Higgs boson cross section.

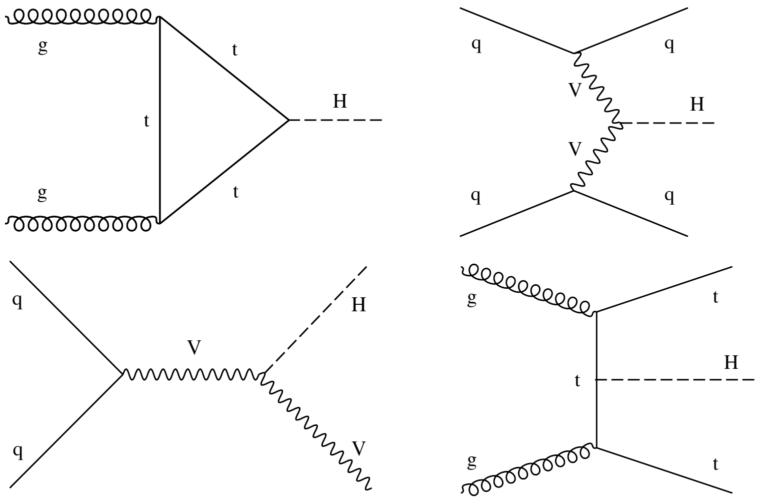


Figure 2.2: Feynman diagrams representing the leading Higgs boson production processes. Progressing in order of largest to smallest in production cross sections: (top left) ggH, (top right) VBF, (bottom left) WH and ZH processes, and (bottom right) $t\bar{t}H$.

2.4.1 Gluon Fusion

The Higgs boson does not couple directly to gluons. Thus, the ggH process is mediated by a virtual top quark loop, Figure 2.2. There are contributions from the other lighter quarks; however, their contributions in the loop are suppressed proportional to m_q^2. Despite the ggH process having the largest cross section, it is the hardest process to isolate from background events. A portion of ggH produced Higgs bosons are produced in conjunction with one or multiple “jets” (“jets” are defined in Section 5.3.1). In these cases, conservation of momentum in the transverse plane will result in the jets and the Higgs boson recoiling off each other with their transverse momenta being equal in magnitude and in opposite directions. This reoil can lead to a boosted topology where the Higgs boson has a large transverse momentum which is a unique event signature. In this thesis the boosted event topology is used as a

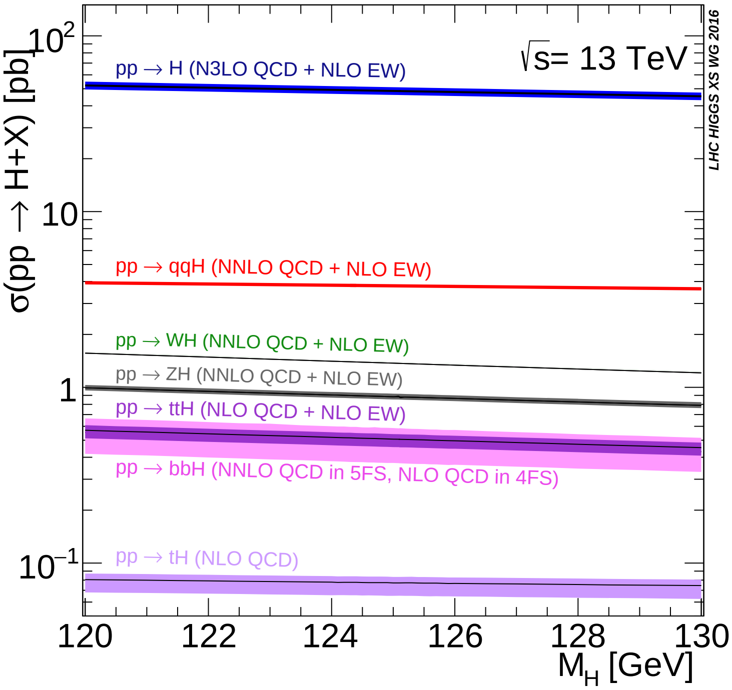


Figure 2.3: The calculated Higgs boson production cross sections and their uncertainties as a function of the Higgs boson mass, are shown. The ggH process is denoted as pp → H in the figure. The CMS and ATLAS experiments have determined $m_H = 125.09\,\text{GeV}$ [1].

handle in the “Boosted” event selection category discussed in Section 6.1.4. However, the majority of ggH events are produced without additional jets. These events are categorized into the “0-Jet” category which is less sensitive because of the lack of a clear handle to separate ggH events with zero jets from the background.

2.4.2 Vector Boson Fusion

The second largest Higgs boson production process is VBF. This process originates by the scattering of two quarks or anti-quarks. The scattering is mediated by the t- or u-channel exchange of a W or Z boson. The Higgs boson is radiated off of the W or Z boson [43], Figure 2.2. The quarks or anti-quarks which are scattered lead to the generation of high-energy jets. This is a highly unique event signature and is used in nearly all Higgs analyses to identify Higgs bosons produced via VBF. In this thesis, events with two high-energy jets consistent with VBF topology are categorized into a "VBF" category for analysis.

2.4.3 Associated Production

The Higgs boson associated production mechanism makes up the third and fourth largest Higgs boson cross sections for WH and ZH respectively. Both processes are depicted in the lower left Feynman diagram in Figure 2.2 where the V represents vector bosons covering both W and Z. Associated production is often called Higgs-strahlung in reference to bremsstrahlung which is the process where a decelerating charged particle emits radiation, often an electron emitting a photon. In the case of Higgs-strahlung, the vector boson created from two initial state quarks or anti-quarks emits a Higgs boson. The associated production event topology is quite unique with the presence of a W or Z boson plus a Higgs boson. In this thesis there is an analysis targeted at each of these process, WH and ZH. We focus on events where the vector bosons decay leptonically: $W^{\pm} \rightarrow \ell^{\pm} + \bar{\nu}$ and $Z \rightarrow \ell^{+}\ell^{-}$.

2.5 Higgs Boson Decays

After a Higgs boson is produced, it will decay rapidly with a predicted lifetime of 1.6×10^{-22}s [44] corresponding to a decay length of 4.8×10^{-12}cm. When created inside of the CMS detector, a Higgs boson will always decay within the LHC beam pipe which has a radius measured in centimeters not picometers. In all Higgs boson analyses, experimentalists search for the signatures of the decay products in energy deposits in their detectors, not the Higgs boson itself. There are multiple possible decay paths each with its own branching ratio; the largest branching ratio processes are shown in Figure 2.4. In this thesis, the Higgs boson decay process to a pair of τ leptons will be denoted $H \to \tau\tau$ where the ± are dropped from the τ superscript for convenience. This same approach is used for $H \to WW$ and $H \to ZZ$. The $H \to \tau\tau$ process has a branching ratio of approximately 6.3% with a relative theoretical uncertainty of ±5.7% [42] for $m_H = 125.09\,\text{GeV}$. This places $H \to \tau\tau$ below $H \to b\bar{b}$, $H \to WW$, and $H \to$ gg in branching fraction. The $H \to$ gg processes is mediated through a quark loop just like the ggH process.

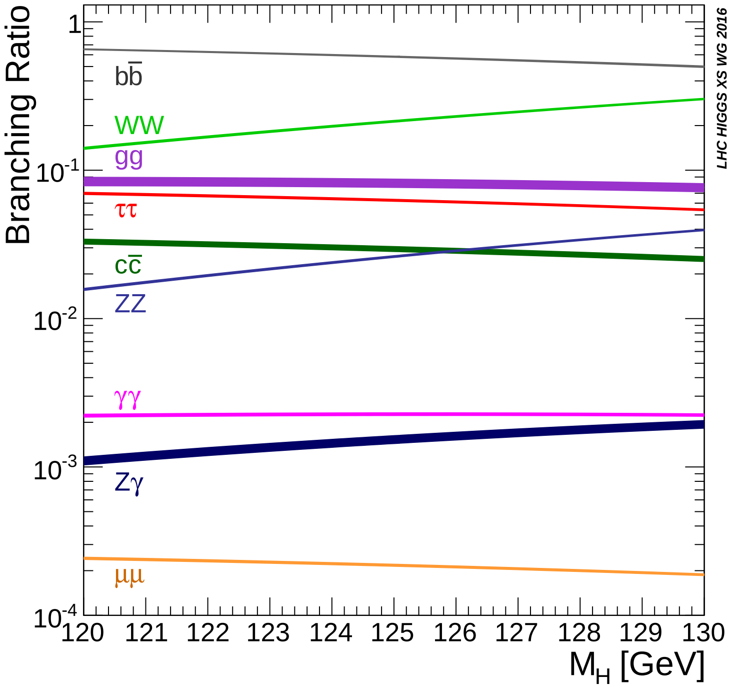


Figure 2.4: The different theorized Higgs boson decay process are shown as a as a function of the Higgs boson mass. The CMS and ATLAS experiments have determined $m_H = 125.09\,\text{GeV}$ [1].

The leading Higgs boson decay processes each have certain advantages and disadvantages for measuring the properties of the Higgs boson. The dominant Higgs boson decay process, $H \to b\bar{b}$, comprises well over half of all Higgs boson decays yet suffers from very large background processes, making it difficult to distinguish Higgs bosons. Additionally, the $m_{\bar{b}b}$ resolution is limited, making the expected Higgs boson signal a broad distribution on top of a large background [41]. The $H \rightarrow WW$ process has a large branching ratio, but this is reduced after requiring $W^{\pm} \rightarrow \ell^{\pm} + \bar{\nu}$ where ℓ is a charged lepton. Additionally, the presence of neutrinos in the final state decreases the Higgs boson mass resolution, $\approx 20\% \, m_H$ [41]. The $H \rightarrow$ gg is an extraordinarily difficult process to attempt to observe at a hadron collider because the final state provides no helpful handles other than m_H to disentangle it from the dominant multijet backgrounds.

The $H \rightarrow \tau\tau$ branching fraction is the largest of the Higgs boson decays directly resulting in leptons. It provides a unique opportunity to probe the Higgs boson Yukawa couplings to the leptons. The $H \rightarrow b\bar{b}$ and $H \rightarrow \tau\tau$ decay processes and the $t\bar{t}H$ production process can all directly probe the Higgs boson Yukawa couplings to fermions. The mass resolution for the Higgs boson reconstructed from the $H \rightarrow \tau\tau$ process suffers like the previously mentioned decays. The neutrinos resulting from the decay of the τ leptons make the mass reconstruction non-trivial. One large benefit when studying $H \rightarrow \tau\tau$ versus $H \rightarrow b\bar{b}$ is that each τ has a 35% probability to decay leptonically to either an electron or a muon (Table 5.2) which provides a clean handle for selecting the event. The other 65% of τ decays result in hadronic τ decays which provide a cleaner handle for event selection compared to the $H \rightarrow b\bar{b}$ bottom-quark decays (Section 5.3.3). This reduces the relative size of the backgrounds with event signatures similar to the Higgs boson in $H \rightarrow \tau\tau$ events compared to those for the $H \rightarrow b\bar{b}$ process and is one of the significant benefits of studying the Higgs boson through the $H \rightarrow \tau\tau$ process.

In this thesis, the various Higgs boson production cross sections and branching ratios for Higgs boson production, and their corresponding uncertainties are taken from References [42, 45, 46] and references therein.

2.6 Higgs Boson: Experimental Results

Searches for the Higgs boson were conducted at multiple experiments such as the searches at the LEP at CERN in the 1990s [25, 26, 27, 28]. In the datasets corresponding to these searches, there were not enough potential Higgs boson events for discovery. Instead, these searches all resulted in placing limits on the possible mass and cross section of the Higgs boson. The Tevatron at Fermilab was active in Higgs boson searches through the early 2000s with multiple analyses targeting the same decay process studied in this thesis, $H \to \tau\tau$. Similar to the LEP results, these analyses placed limits on the possible mass and cross section of the Higgs boson [29, 30]. After the full analysis of the Tevatron dataset, there was evidence for the existence of a Higgs boson [47, 48]. However, just like LEP, there were not enough potential Higgs boson events to qualify as a discovery.

The discovery of the Higgs boson required higher collision energies than those provided by the Tevatron, which reached a maximum center-of-mass collision energy of 1.96 TeV. The LHC at CERN was designed to deliver this increase in collision energy and in 2010 the LHC started delivering proton-proton collisions at up to 7 TeV; an increase to 8 TeV followed in 2012. Using the proton-proton collision data with center-of-mass energy 7 and 8 TeV, a particle compatible with the Higgs boson expectations was observed by the CMS and ATLAS experiments at the CERN LHC in events where the Higgs boson decays to ZZ, $\gamma\gamma$, and WW [5, 6, 7].

Using this same set of data, other analysts pursued the $H \to \tau\tau$ decay process and the CMS Collaboration showed evidence for the $H \to \tau\tau$ process with an observed significance of 3.2 based on an expected significance of 3.7 standard deviations (s.d.)

for a Higgs boson mass of 125 GeV [49]. The ATLAS experiment reported evidence for the $H \rightarrow \tau\tau$ process with an observed (expected) significance of 4.5 (3.4) s.d. for a Higgs boson mass of 125 GeV [50]. The combination of results from both experiments yields an observed (expected) significance of 5.5 (5.0) s.d. [51].

Further analysis from both experiments, described in References [52, 53, 54, 55, 56, 57], established that the measured properties of the new particle, including its spin, charge-parity properties, and coupling strengths to SM particles, are consistent with those expected for the Higgs boson predicted by the SM. The mass of the Higgs boson (m_H) has been determined to be $125.09 \pm 0.21(stat.) \pm 0.11(syst.)$ GeV, from a combination of ATLAS and CMS measurements [1]. An example of these measurements can be seen in Figure 2.5 which shows the best fit value for the signal strength for the Higgs boson production and decay processes. As can be seen, the majority of the $\tau\tau$ decay process measurements shows agreement with the SM predictions for the Higgs boson. The measurements corresponding to $H \rightarrow \tau\tau$ with the Higgs bosons produced in association with a W boson (WH) and the measurement of the Higgs boson produced with two top-quarks ($t\bar{t}H$) show non-significant deviation from the SM prediction. However, the uncertainties are sizable and are represented by the size of the 1σ band. These $t\bar{t}H$ Run-I results have been updated with the same dataset studied in this thesis. The updated $t\bar{t}H$, $H \rightarrow \tau\tau$ results are compatible with the SM prediction within 1σ [58]. Further data collection and analysis will continue to reduce the size of the 1σ uncertainty bands and will show if deviations are statistically significant or if they are temporary fluctuations that are smoothed out with more data collection.

This thesis builds on these previous experimental results at 7 and 8 TeV and measures the properties of the Higgs boson in the $H \rightarrow \tau\tau$ decay process at 13 TeV center-of-mass collision energy. We establish the first 13 TeV observation of the $H \rightarrow$

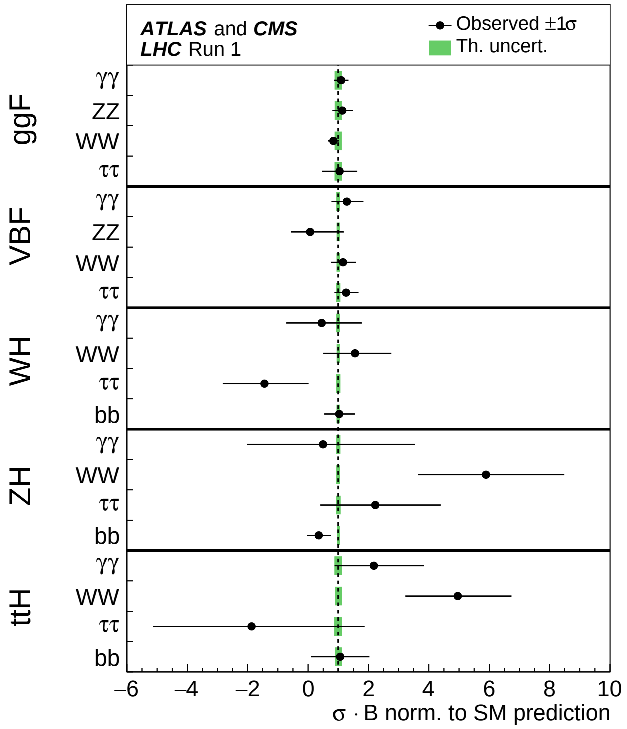


Figure 2.5: The best fit values for the signal strength of the listed Higgs boson production processes and decay processes. A value of 1 indicates perfect agreement with the SM. The error bars indicate the 1σ intervals. The green shaded bands indicate the theoretical uncertainties in the predictions.

$\tau\tau$ process with an observed (expected) significance of 5.5 (4.8) s.d. [4]. The best fit signal strengths are measured similar to what is seen in Figure 2.5. Additionally, the couplings of the Higgs boson to fermions and vector bosons is measured and is discussed in the final results section of this thesis, Chapter 8.

Chapter 3

The CMS Experiment and the LHC

The CERN accelerator complex takes hydrogen gas, strips the electrons off of the protons, accelerates the protons and creates proton-proton collisions in the middle of the Compact Muon Solenoid (CMS) detector. The CMS detector is a large general purpose physics detector designed to precisely measure the energy of particles emanating from the proton-proton collisions to allow physicists to reconstruct the details of the initial collisions. Smooth operation of the LHC and the delivery of the proton-proton collisions and high efficiency operation of the CMS detector are both necessary to the data we gather and analyze.

3.1 The LHC

The LHC is a hadron accelerator and collider built in the existing 26.7 km tunnel, under France and Switzerland, that was originally used for the Large Electron-Positron (LEP) collider. It is a two ring system with counter-rotating beams which are accelerated by radio frequency (RF) cavities located in dedicated sections along the LHC ring. The LHC uses superconducting magnets to bend the path of the protons around the circular path of the ring and also to steer the proton beams into collisions. The original design center-of-mass energy for the LHC was 14 TeV, 7 TeV per beam. The 27 km LHC tunnel is on average 100 m underground. The original LEP tunnel was built underground to help offset the high cost of land acquisition in the Swiss and French countrysides and to reduce the emission of radiation from the experiments into the countryside. Additionally, there are physics benefits from housing the detectors underground coming from a reduced rate of cosmic ray background [59].

The LHC particle collisions take place at four locations around the ring where the two beams cross each other. These are the regions where the LHC can steer the proton beams into collisions. The four crossing regions and their corresponding detectors are:

- The Compact Muon Solenoid (CMS), one of the two general purpose detectors which is discussed in detail in Section 3.2 and is located at Point 5
- A Toroidal LHC ApparatuS (ATLAS), the other of the two general purpose detectors located at Point 1
- Large Hadron Collider beauty (LHCb), a specialized b-physics experiment measuring the parameters of CP violation focused on exploring the matter-antimatter asymmetry in the universe
- A Large Ion Collider Experiment (ALICE), specifically designed to study quark-gluon plasmas from heavy ion collisions

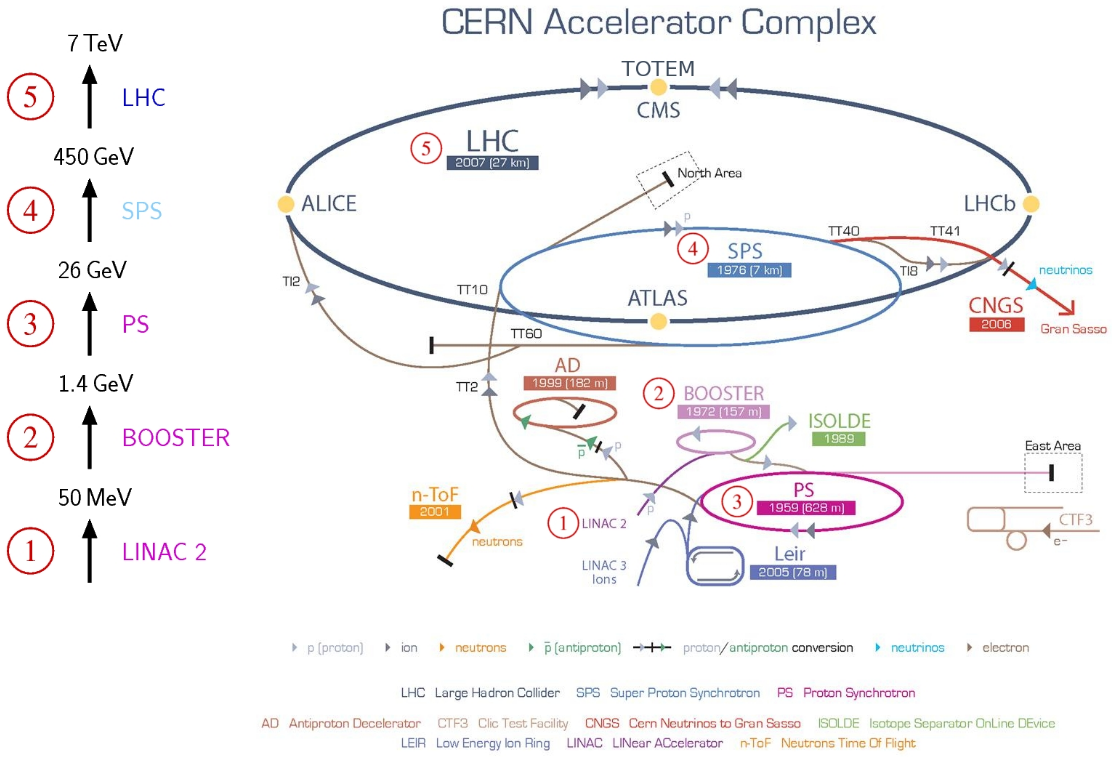

7 TeV
5
LHC
450 GeV
4
SPS
26 GeV
3
PS
1.4 GeV
2
BOOSTER
50 MeV
1
LINAC 2
CERN Accelerator Complex
TOTEM
CMS
5
LHC
2007 (27 km)
North Area
ALICE
LHCb
p
TT40
TT41
4
SPS
1976 (7 km)
TI8
neutrinos
CNGS
2006
Gran Sasso
TI2
TT10
ATLAS
TT60
AD
1999 (182 m)
TT2
2
BOOSTER
1972 (157 m)
ISOLDE
1989
East Area
p̄
p
n-ToF
2001
neutrons
1
LINAC 2
p
3
PS
1959 (628 m)
CTF3
e−
LINAC 3
Ions
Leir
2005 (78 m)
p (proton)
ion
neutrons
p̄ (antiproton)
proton/antiproton conversion
neutrinos
electron
LHC Large Hadron Collider
SPS Super Proton Synchrotron
PS Proton Synchrotron
AD Antiproton Decelerator
CTF3 Clic Test Facility
CNGS Cern Neutrinos to Gran Sasso
ISOLDE Isotope Separator OnLine DEvice
LEIR Low Energy Ion Ring
LINAC LINear ACcelerator
n-ToF Neutrons Time Of Flight

Figure 3.1: Diagram of the LHC accelerator complex showing the five accelerators used to accelerate protons to their design collision energy of 7 TeV. For the LHC Run II, the highest energy achieved for protons is 6.5 TeV.

3.1.1 LHC Proton Acceleration

The LHC accelerator complex has been built up over the years as technology progressed, physics goals advanced, and continued funding allowed for the expansion and repurposing of previous accelerators. The LHC acceleration injector chain begins with the LINAC2 linear accelerator which accelerates protons from rest to an energy of 50 MeV. The protons come from a bottle of hydrogen gas which feeds the source chamber of the LINAC2. A large electric field is applied in the source chamber which breaks the electron-proton bonds separating the electrons from the protons. The positive electric charge of the protons allows them to be accelerated by electric fields. The LINAC2 and the subsequent circular accelerators all use radio frequency (RF) cavities to accelerate the protons. With the electrons removed, a group of protons can begin progressing through the accelerator chain; each grouping of protons accelerated in the LINAC2 eventually contributing to one of the circulating proton bunches used for collisions in the LHC.

The LINAC2 feeds into the circular Proton Synchrotron Booster accelerator boosting the protons to an energy of 1.4 GeV, or 91.6% c, followed by the Proton Synchrotron further accelerating the protons to 26 GeV, 99.9% c. Next comes the Super Proton Synchrotron accelerating the protons to 450 GeV, the input energy for the LHC. From 450 GeV the LHC will accelerate the proton bunches to their maximum energy of 6.5 TeV before collisions [59]. A diagram of the LHC accelerator complex can be seen in Figure 3.1.

3.1.2 LHC Magnets

The path of the proton bunches is controlled by magnetic fields throughout the accelerator chain. In all of the circular accelerators, dipole magnetic fields are applied at right angles to the protons direction of travel which applies a force orthogonal to the magnetic field and proton velocity bending the trajectory into a circular path. The LHC relies on 1232 15-m long superconducting dipole magnets, see Figure 3.2, to maintain the roughly circular path around the LHC. The maximum magnetic field the superconducting dipole magnets can attain is 8.33 T which corresponds to a proton beam energy of 7 TeV. However, for physics operations in Run-II, the field was limited to 6.5 TeV due to the time consuming process of retraining the dipole magnets to 7 TeV [60]. The LHC dipole magnets must be cooled by liquid helium to 1.9 Kelvin to maintain their superconducting characteristics [61]. The superconducting coils are made from NbTi Rutherford cables [62].

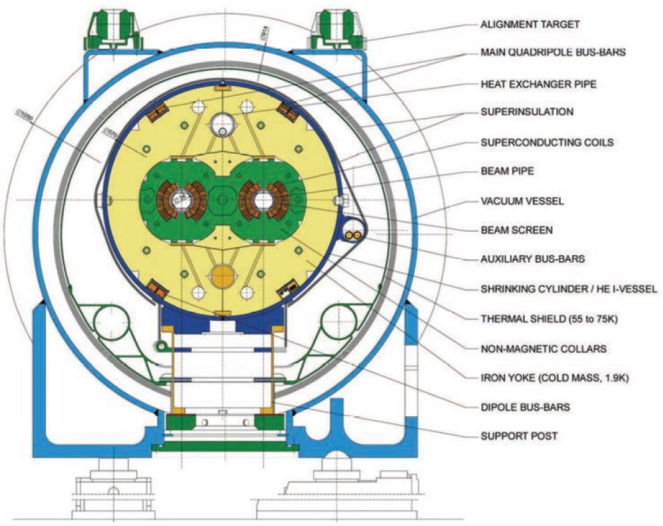

ALIGNMENT TARGET
MAIN QUADRIPOLE BUS-BARS
HEAT EXCHANGER PIPE
SUPERINSULATION
SUPERCONDUCTING COILS
BEAM PIPE
VACUUM VESSEL
BEAM SCREEN
AUXILIARY BUS-BARS
SHRINKING CYLINDER / HE I-VESSEL
THERMAL SHIELD (55 to 75K)
NON-MAGNETIC COLLARS
IRON YOKE (COLD MASS, 1.9K)
DIPOLE BUS-BARS
SUPPORT POST

Figure 3.2: Cross section of an LHC superconducting dipole magnet showing the two beam pipes in the middle surrounded by the the superconducting coils.

The LHC superconducting dipole magnets are complemented by 392 main superconducting quadrupole magnets, 688 sextupole magnets, and 168 octupole-type magnets. These additional magnets are dedicated to proton beam focusing, keeping the beam from spreading out laterally at the collision point. The beam has much more lateral spread throughout the rest of the LHC ring [61, 59].

The accelerator chain and superconducting magnets of the LHC provide a proton beam segmented into discrete bunches which can be focused, defocused, and otherwise adjusted to deliver a target rate of proton-proton collisions [63]. Basic proton beam design parameters are detailed in Table 3.1 [63]. The bunch spacing for the proton bunches in the beam corresponds to a bunch crossing rate, or collision rate, of 40 MHz. The bunch length is measured in ns which is the relevant value for determining the frequency of the LHC RF cavities (400 MHz) used to accelerate the beam; this length roughly corresponds to 1 meter.

	Units	LHC Design	2016 LHC Operations
Center of Mass Energy	[TeV]	14	13
Energy per Beam	[TeV]	7	6.5
Peak Luminosity	[$\mathrm{cm}^{-2}\mathrm{s}^{-1}$]	1×10^{34}	1.53×10^{34}
Number of Bunches	N/A	2808	2208
Bunch Spacing	[ns]	25	25
Protons per Bunch	N/A	1.15×10^{11}	1.25×10^{11}
Bunch Length, total (4σ)	[ns]	2.5	2.5

Table 3.1: LHC beam characteristics after the LHC has increased the energy of the proton beam from the input 450 GeV to the target collision energy.

3.2 The CMS Experiment

The Compact Muon Solenoid (CMS) detector is located at Point 5 along the LHC ring near Cessy, France. The design of the CMS detector was inspired by decades of previous high energy particle physics detectors and has as one of its primary purposes the detection and measurement of the characteristics of the Standard Model Higgs boson. The detector was designed and built targeting high efficiency detection and excellent energy resolution for specific Higgs boson decay modes. The strong, 3.8 T magnetic field, is created by a superconducting magnet with a central bore of 6.3 m. There is enough room inside the magnet to fully contain a track system, the electromagnetic calorimeter, and the hadronic calorimeter. Having the electromagnetic calorimeter within the magnet is the key unique feature of CMS that allows it to have such good electromagnetic energy resolution. Additionally, the high granularity of the electromagnetic calorimeter and superb energy resolution make it perfect for reconstructing high-energy (> few GeV) electrons and photons. The superconducting magnet combined with the muon spectrometer provide precise momentum resolution. Charged hadrons are precisely measured in the tracker with momenta ranging from a few hundred MeV to a TeV. The detector is a large cylinder 15.0 m in diameter and 28.7 m in length. It has a mass of approximately 14,000,000 kg.

The CMS detector is composed of many individual subdetectors. Progressing radially outwards from the collision point, there is first the silicon pixel tracker then the silicon strip systems. Next is the electromagnetic calorimeter followed by the hadronic calorimeter. These systems are all contained within the bore of the CMS superconducting magnet. After the superconducting magnet, there are additional gas ionization detector subsystems for muon detection embedded within the steel flux-return yoke. In the central region of the detector, there is an outer portion of the hadronic calorimeter, and lastly, the muon spectrometer. These systems can all be viewed in Figure 3.3.

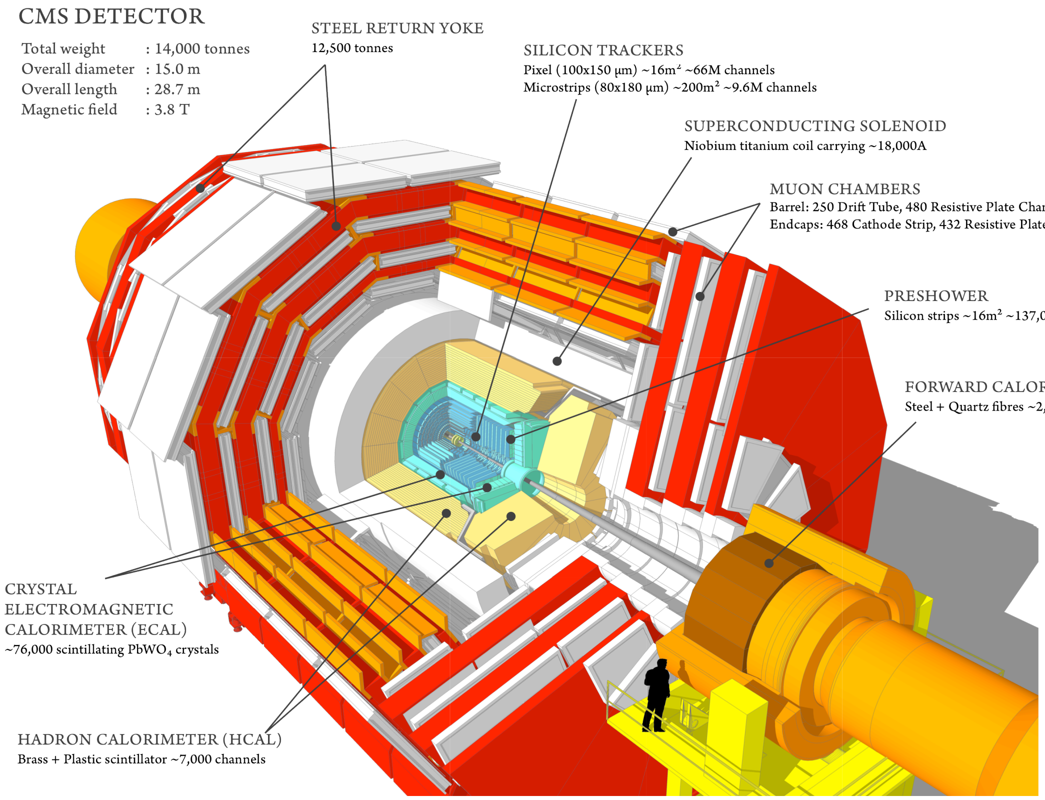

CMS DETECTOR
Total weight : 14,000 tonnes
Overall diameter : 15.0 m
Overall length : 28.7 m
Magnetic field : 3.8 T
STEEL RETURN YOKE
12,500 tonnes
SILICON TRACKERS
Pixel (100x150 μm) ~16m² ~66M channels
Microstrips (80x180 μm) ~200m² ~9.6M channels
SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying ~18,000A
MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 468 Cathode Strip, 432 Resistive Plate Chambers
PRESHOWER
Silicon strips ~16m² ~137,000 channels
FORWARD CALORIMETER
Steel + Quartz fibres ~2,000 Channels
CRYSTAL ELECTROMAGNETIC CALORIMETER (ECAL)
~76,000 scintillating PbWO₄ crystals
HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator ~7,000 channels

Figure 3.3: A cutaway diagram of the CMS detector showing the many subdetectors within.

Information from the 40 MHz proton-proton collisions is aggregated by a central data acquisition system and filtered by a two-tiered trigger system before storage [64]. The first level of the trigger system, the L1 Trigger, is composed of custom hardware processors and uses information from the calorimeters and muon detectors to select interesting events at a rate of around 100 kHz. The second level of the trigger, known as the high-level trigger (HLT), consists of a farm of commercial processors further filtering out less interesting data events resulting in a final event rate of about 1 kHz which is stored for further processing and analysis.

3.2.1 Geometry

To understand the CMS detector, an understanding of the geometry and coordinate system used by the CMS experiment is necessary. The CMS detector is a hermetic, cylindrical particle detector surrounding a central region where the proton-proton collisions delivered by they LHC occur. This central point, inside the LHC beam pipe is designated as the origin for the CMS coordinate system, (0,0,0) in (x, y, z) coordinates. Positive x points towards the center of the LHC ring. Positive y points vertically upwards. The positive z is along the LHC ring in the clockwise direction when viewed from above. These (x, y, z) coordinates are commonly transformed into quasicylindrical coordinates when referring to particles, $(p_{\mathrm{T}}, \eta, \phi)$. For a particle, p_{T} is,

$$p_{\mathrm{T}} \equiv \sqrt{p_x^2 + p_y^2} \tag{3.1}$$

η is defined using θ based on spherical coordinates,

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

ϕ is defined in the $x - y$ plane.

3.2.2 Superconducting Magnet

The CMS superconducting magnet, as is noted in the collaboration name, is a solenoid magnet and is one of the most fundamental pieces of the CMS experiment. The superconducting magnet bends the trajectories of charged particles within the CMS detector. The curvature of the flight path of a charged particle can be used to help calculate the momentum of the particle in accordance with the Lorenetz force.

The CMS superconducting magnetic is constructed from a 4-layer winding of stabilized reinforced NbTi conductor. When in operation, the magnet is kept in a superconducting state by cooling it with liquid helium to a temperature of 4.6 Kelvin. With a nominal current of 19.14 kA, the superconducting solenoid magnet is able to produce a roughly uniform magnetic field of 3.8 T within its central bore [65]. The magnetic field created is approximately 100,000 times stronger than the Earth's magnetic field. The stronger the magnetic field the more highly curved a particle's trajectory will be, leading to better momentum and energy measurements. The central bore of the magnet is only 6.3 m in diameter leaving limited room for the CMS tracker and calorimeter systems.

3.2.3 Inner Tracking System

The inner tracking system is composted of two subdetectors, the pixel tracker and the strip tracker. They are designed to deliver precise and efficient measurement of the trajectories of charged particles propagating outwards from the proton-proton collisions. Beyond reconstructing tracks, the information from the tracking systems can be used to reconstruct the proton collision points and secondary decay vertices.

The tracking systems surround the interaction point and have a length of 5.8 m and diameter of 2.5 m. Both the pixel and strip trackers have cylindrical barrel layers and disk-like endcap layers. A schematic of the tracking systems can be seen in Figure 3.4.

The pixel tracker has three barrel layers at radii between 4.4 cm and 10.2 cm. The incredibly high granularity and low relative occupancy of the pixel tracker is helpful for track seeding and vertex reconstruction discussed in the track reconstruction section 5.1.1. The silicon strip tracker is composed of 10 barrel detection layers extending outwards to a radius of 1.1 m. During the 2016-2017 extended year end technical stop, the pixel tracker was upgraded to have a fourth detection layer. As 2017 data is not used in the analyses presented here, details of this upgrade are omitted. The endcaps of the pixel tracker consist of 2 disks which extend the η range of the detector to $|\eta| < 2.5$. The same η range is covered in the strips tracker with 12 disks on each side of the barrel.

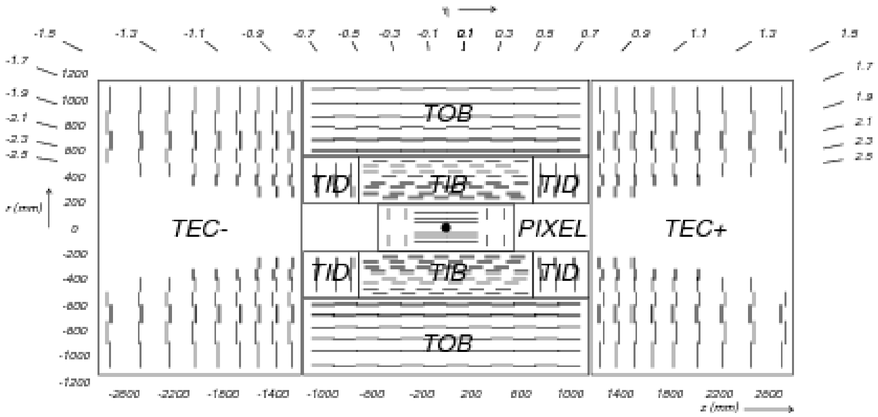

η
-1.5
-1.3
-1.1
-0.9
-0.7
-0.5
-0.3
-0.1
0.1
0.3
0.5
0.7
0.9
1.1
1.3
1.5
-1.7
-1.9
-2.1
-2.3
-2.5
1.7
1.9
2.1
2.3
2.5
1200
1000
800
600
400
200
0
-200
-400
-600
-800
-1000
-1200
r (mm)
TOB
TID
TIB
TID
TEC-
PIXEL
TEC+
TID
TIB
TID
TOB
-2600
-2200
-1800
-1400
-1000
-600
-200
200
600
1000
1400
1800
2200
2600
z (mm)

Figure 3.4: Schematic cross section of the CMS tracker. Each line represents a detector module. Double lines indicate back-to-back modules which deliver stereo hits.

The pixel tracker contains 66 million individual pixels. The strips tracker contains 9.3 million individual strips. The high granularity is necessary because of the extremely high particle flux through the tracker and need to reconstruct tracks which are close to each other. The tracker measures the p_{T} of charged hadrons in the barrel region with a resolution of 1% for $p_{\mathrm{T}} < 20\,\mathrm{GeV}$. Details of track reconstruction are discussed in Section 5.1.1.

A serious concern during the design of the tracker systems was the amount of material used to build the systems. All of the electronics, hardware, cooling systems, and wiring contribute to the tracker material budget. Material located between the interaction region and the calorimeters will reduce the precision of the calorimeter energy measurements as well as potentially degrade the track-based measurement of the energy of a particle. This is because of potential electromagnetic and/or nuclear interactions between particles and the tracker material which will absorb energy and redirect the trajectory of a particle. Figure 3.5 shows the material budget in terms of interaction lengths and radiation lengths as a function of η for the tracker systems.

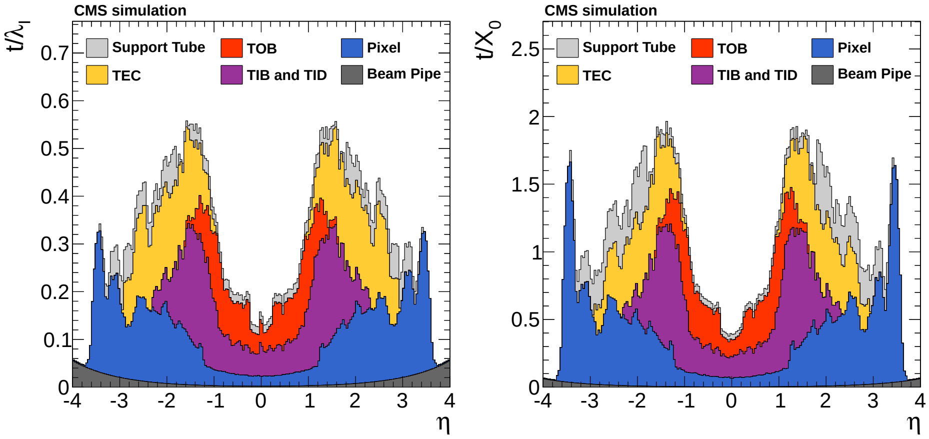


Figure 3.5: Total thickness t of the inner tracker material expressed in units of interaction lengths λ_l (left) and radiation lengths X_0 (right), as a function of the pseudorapidity η.

3.2.4 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) is a homogeneous calorimeter made of lead tungstate ($PbWO_4$) crystals. There are 61,200 individual crystals mounted in the barrel region with 360 crystals completing a ring in ϕ and 170 crystals spanning the barrel η range, $|\eta| < 1.479$. The ECAL endcap detectors cover the η range $1.479 < |\eta| < 3.0$. The ECAL subsystems can be seen in Figure 3.6. The ECAL barrel crystal cross-section is approximately 0.0174×0.0174 in (η, ϕ), or about $22 \times 22\,\mathrm{mm}^2$ on the front face of a crystal [65]. Twenty-two millimeters is also the Molière radius for ($PbWO_4$), meaning an electromagnetic shower could be contained by as few as 4 crystals. The fine grain resolution is extremely helpful for precise energy measurements of electrons, photons, and π^0 all of which contribute to better τ_h reconstruction and more more effective τ_h identification and isolation.

The ECAL barrel crystals are 230 mm in length corresponding to a radiation length of 25.8 X_0 which is sufficient to contain 98% of the energy of electrons and photons up to 1 TeV. Despite the large number of radiation lengths, the crystals have an interaction length, λ_l, of about 1.0. This causes roughly two thirds of the hadrons passing through the ECAL to start showering within the ECAL [65].

As electrons and photons interact with the crystals, energy is deposited at a rate of about 4.5 photoelectrons per MeV. The scintillation decay time of the ($PbWO_4$) crystals is approximately the same order of magnitude as the LHC bunch crossing time; about 80% of the light is emitted in 25 ns [66]. This provides the fast response necessary to separate out-of-time ECAL energy deposit from a given LHC bunch crossing. To read out the energy deposited in each barrel crystal, there is an avalanche photodiode (APD) mounted to the backside-facing surface of the crystal. For the endcap crystals, a more radiation hardened readout is used, vacuum phototriodes (VPTs). typical ECAL electronics noise $\sigma_{\mathrm{noise}}^{\mathrm{ECAL}}$ is measured to be ≈ 40 (150) MeV per crystal in noise the barrel (endcaps). The ECAL barrel energy resolution for electrons in beam tests has been measured and is given by Equation 3.3 [67].

$$\frac{\sigma}{E} = \frac{2.8\%}{\sqrt{E/\,\mathrm{GeV}}} \oplus \frac{12\%}{E/\,\mathrm{GeV}} \oplus 0.3\% \tag{3.3}$$

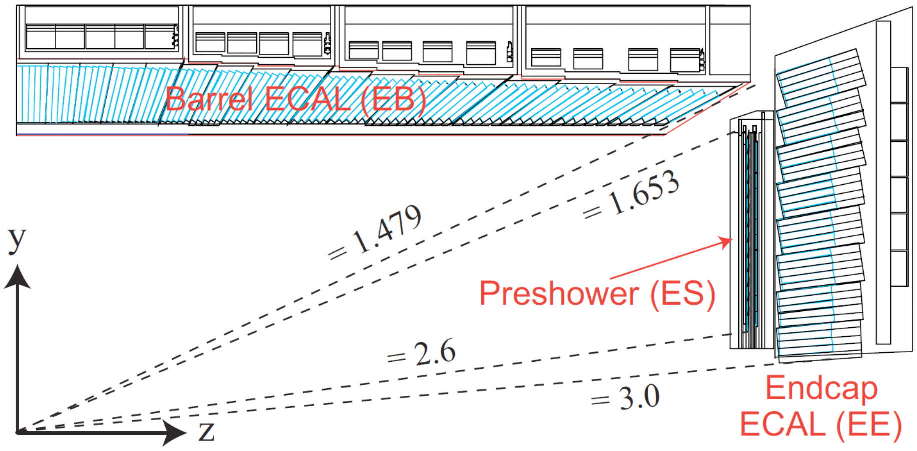


Figure 3.6: Longitudinal view of the CMS detector depicting the ECAL subdetector. The crystals are inclined towards the interaction region.

3.2.5 Hadronic Calorimeter

The hadronic calorimeter (HCAL) provides necessary energy measurements to reconstruct neutral hadrons and total event energy sums, see Section 5.2.2. These particle are the primary measurable constituents in “jets” and hadronically decay τ leptons. Additionally, measurement of the transverse energy imbalance from a collision can indicate the presence of neutrinos or possibly exotic particles [68]. The HCAL is segmented into four calorimeters: the barrel detector, the endcap detector, the forward hadronic detector, and the barrel outer detector. The barrel detector is a sampling calorimeter covering $|\eta| < 1.3$. It is built from flat brass absorber plates interleaved with plastic scintillator segments to measure and readout the deposited energy. The

HCAL barrel has coarser granularity than the ECAL; each segment is 0.087×0.087 in (η, ϕ) which covers the same area as 25 ECAL crystals.

The HCAL barrel occupies the space between the ECAL and the superconducting magnet from a radial distance of 1.77 m to 2.95 m form the beam line. Particles propagating outwards at an angle of $\eta = 0$ will encounter the thinnest portion of the HCAL barrel corresponding to 5.82 λ_l. The HCAL barrel effective thickness increases with polar angle (θ) as $1/sin\theta$, resulting in 10.6 λ_l at $|\eta| = 1.3$. The electromagnetic crystal calorimeter in front of HB adds about 1.1 λ_l of material [65].

To increase the effective thickness in the barrel region, $|\eta| < 1.3$, an outer hadronic calorimeter is placed outside the superconducting magnet. This helps to identify late starting showers and to measure the shower energy deposited after the HCAL barrel detector. The outer HCAL detector utilizes the solenoid coil as an additional absorber equal to $1.4/sin\theta$ interaction lengths [65]. The HCAL endcap system partially overlaps with the barrel detector and ranges from $1.3 < |\eta| < 3.0$. The HCAL detector is depicted in Figure 3.7.

The forward hadronic calorimeter, located 11.2 m from the interaction point, extends the HCAL η coverage over $3 < |\eta| < 5.2$. The high particle and radiation flux, at high $|\eta|$ values severely limits the types of detector systems which can survive years of LHC operating conditions. On average, 760 GeV per proton-proton interaction is deposited into the forward calorimeters, compared to only 100 GeV for the rest of the detector [65]. Because of this, the forward hadronic calorimeter is not the same style sampling calorimeter like the rest of HCAL. Instead it uses Cherenkov-based, radiation-hard, fused-silica quartz fibers embedded in grooves within the steel structure of the absorber material which is a more radiation hardened technology.

The energy resolution of the HCAL barrel has been measured using test beams [69], and is:

$$\frac{\sigma}{E} = \frac{115\%}{\sqrt{E/\,\mathrm{GeV}}} \oplus 5.5\% \tag{3.4}$$

For the forward HCAL system the energy resolution is [70],

$$\frac{\sigma}{E} = \frac{280\%}{\sqrt{E/\,\mathrm{GeV}}} \oplus 11\% \tag{3.5}$$

These two equations show the much poorer energy resolution of HCAL compared to ECAL. There are multiple reasons for this include the fact that the HCAL is a sampling calorimeter. Much of the energy deposits in HCAL occur in absorber layers reducing the amount of photons entering the scintillator material available to the readout electronics. Additionally, typical HCAL electronics noise $\sigma^{\mathrm{HCAL}}_{\mathrm{noise}}$ is measured to be $\approx 200\,\mathrm{MeV}$ per tower.

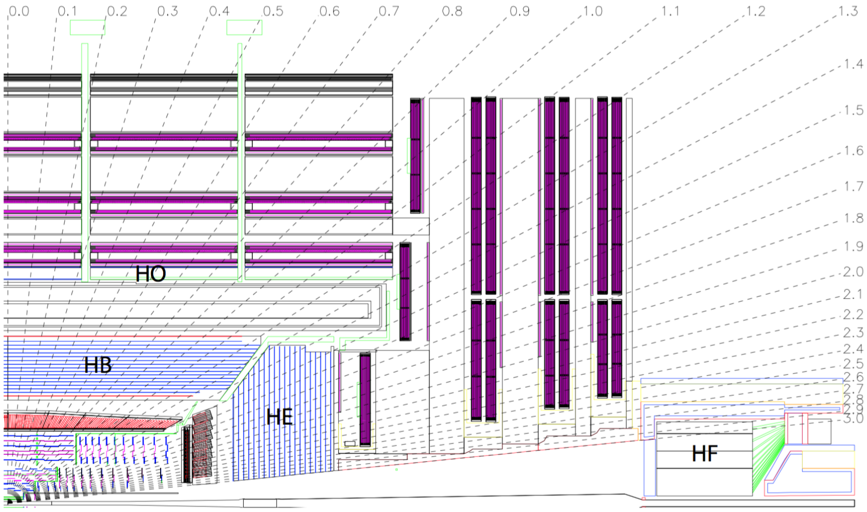


Figure 3.7: Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters. The effective thickness of the HB detector increases with increasing $|\eta|$.

3.2.6 Muon System

The muon spectrometer is important to the physics goals of the CMS experiment. Excellent muon reconstruction, identification, and momentum measurement lead to a high purity of reconstructed muons and makes the system very useful for selecting data events to store. The muon spectrometer is composed of three subsystems, the drift tubes (DT) which cover $|\eta| < 1.2$, the cathode strip chambers (CSC) which cover $0.9 < |\eta| < 2.4$, and the resistive plate chambers (RPC) covering $|\eta| < 1.6$, see Figure 3.8.

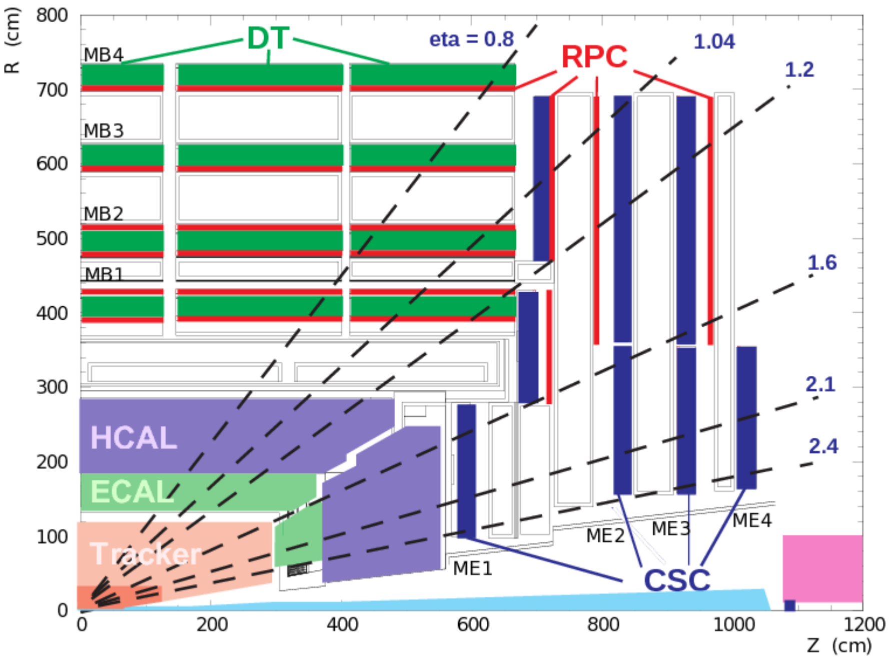

R (cm)
800
700
600
500
400
300
200
100
0
0
200
400
600
800
1000
1200
Z (cm)
DT
MB4
MB3
MB2
MB1
eta = 0.8
RPC
1.04
1.2
1.6
2.1
2.4
HCAL
ECAL
Tracker
ME1
ME2
ME3
ME4
CSC

Figure 3.8: Layout of one quadrant of CMS. The four DT stations in the barrel (MB1-MB4, green), the four CSC stations in the endcap (ME1-ME4, blue), and the RPC stations (red) are shown.

The muon subsystems are each embedded within the magnet's flux-return yoke. Each system detects particles through gas ionization techniques. When passing through ionizing gas detectors, charged particles leave a trail of ionized gas molecules which can be detected and measured.

The drift tubes are used in the barrel region because of the relatively low occupancy rate and the relatively low strength of the local magnetic field (which is well contained in the flux-return yoke through the barrel region). The DTs rely on 172,000 2.4 m long sensitive wires maintained at a high voltage to detect the ionized traces of passing charged particles. The wires are housed in a drift cell with a maximum path length and time of drift of 21 mm and 380 ns using a gaseous mixture of 85% Argon + 15% CO_2 [68]. The drift cell size and drift time are small enough to avoid high occupancy levels during collisions data taking. A detailed schematic of a DT cross-section is in Figure 3.9.

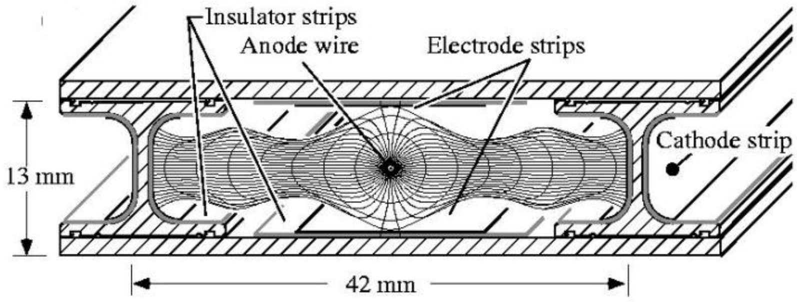


Figure 3.9: Schematic of a drift tube cell showing drift lines leading to/from the anode wire and isochrones three of which are seen as the concentric lines surrounding the anode wire. The voltages applied to the electrodes are +3,600V for wires, +1,800V for strips, and -1,200 V for cathodes.

Progressing outwards from the barrel region towards higher $|\eta|$ values, the background particle flux increases as does the non-uniformity of the magnetic field in the spacing between the flux-return yoke. These considerations lead to a different required muon detection technology. In the two endcap regions of CMS the muon system uses cathode strip chambers arranged as large disks much like the other subsystem endcaps. The CSCs have a fast response time, fine segmentation, and are more radiation resistant than the DTs. The CSC system is depicted in Figure 3.10 showing the relative arrangement of the anode wires to the cathode strips. The muon coordinate along the anode wires (the ϕ coordinate) is obtained by interpolating charges induced on strips [71]. The CSC subsystem has a nominal gas mixture of 40% Argon + 50% CO_2 + 10% CF_4.

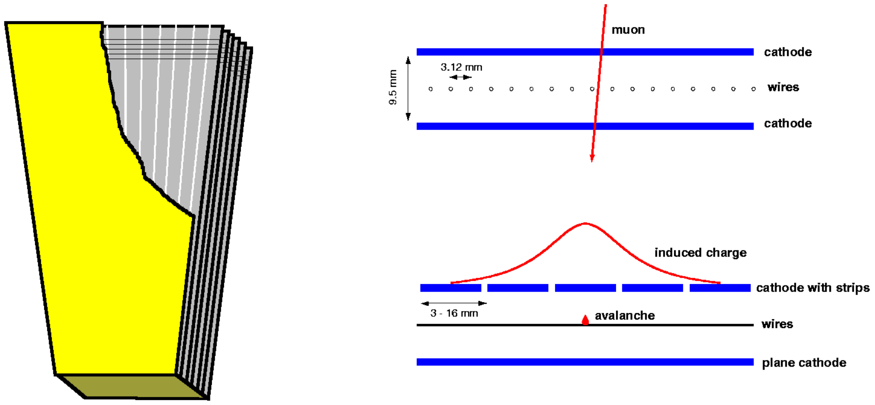


Figure 3.10: (left) Schematic of a CSC module composed of 7 trapezoidal panels with 6 gas gaps between the panels. The cut away of the top panel reveals the anode wires used to detect ionized molecules from a passing muon. The anode wires run horizontally and the cathode strips run vertically in this orientation. (right) Two diagram showing the inner dimensions of a CSC gas gap. The two views show how the orthogonal configuration of the anode wires and cathode strip can be used to localize the positions of a transversing muon.

To increase the timing precision of the muon spectrometer a third subsystem was built from RPCs which cover the barrel and a portion of the endcap region of the detector, $|\eta| < 1.6$. The RPCs are gaseous parallel-plate ionizing gas detectors with two gas chambers per module. The RPC system is used to compliment the excellent spacial resolution of the DTs and CSCs with a high precision timing measurement that can resolve which LHC bunch crossing a certain energy deposit is associated with. The timing resolution is comparable to that of scintillator-based detectors [72]. The RPCs are also used to resolve x-y ambiguity from the DTs and CSCs when applicable.

The RPCs are constructed with two chambers per module that are both operated in avalanche mode. Common pick-up readout strips are located in between the two chambers. With the common readout strips, the total induced signal is the sum total of the signal in each of the chambers. The two-fold signal allows the RPCs to operate at a lower voltage than a single chamber detector and leads to increased efficiency. The RPC subsystem is operated with a nominal gas mixture of $96.2\%\,\mathrm{R134a} + 3.5\%\,i\mathrm{C_4H_{10}} + 0.3\%\,\mathrm{SF_6}$ [68]. An example of the double-chamber construction is in Figure 3.11.

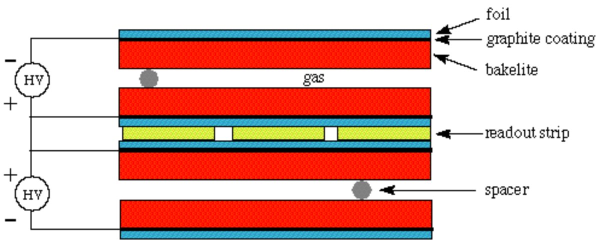


Figure 3.11: Diagram of a double-chamber RPC. The gas chamber is 2 mm wide and is surrounded by two bakelite layers 2 mm thick each.

3.2.7 Trigger and Data Acquisition

The CMS Trigger and Data Acquisition (DAQ) system is designed and built to read-out detector information from each of the CMS subdetectors, make a rapid assessment if the the content associated with a specific LHC bunch crossing is worth storing for future physics analysis, and finally deliver the event content to storage. The 40 MHz bunch crossing rate of the LHC makes this a remarkable challenge. The Trigger and DAQ system must operate rapidly and in close synchronization so that the event content read out by each subdetector is attributed to the proper LHC bunch crossing. In total, the trigger system reduces the throughput rate from the initial 40 MHz to a storage rate of around 1 kHz. In certain circumstances, a technique called "data parking" is used to store lower priority events passing the HLT. Parked events undergo offline reconstruction many months after their initial collections. This can help increase the effective HLT rate towards 2 kHz. This technique was not used directly for any of the data events analyzed in this thesis. Storing more events for physics analysis is always desired, but with each event requiring roughly 1 MB of disk space for storage in its raw format, storing many more events is not currently an option for the CMS experiment based on available storage space [68].

Level-1 Trigger

The CMS Trigger system is broken down into two tiers. This approach relies on the first level, the Level-1 (L1) Trigger, to assess a reduced set of muon and calorimeter data from each LHC bunch crossing and make an initial determination of whether an event appears interesting. The L1 Trigger reduces the input 40 MHz rate to a more manageable 100 kHz. The L1 Trigger is restricted to 4 μs of latency to make the determination of whether the event should be passed the the next tier or whether the event should be deleted [64]. The speed and throughput requirements of the L1

Trigger necessitate that the system is built from dedicated, custom hardware. The latency requirement eliminates the possibility of using inner tracker information, due to its 7 μs readout time, as well as full ECAL crystal-level granularity.

The L1 Trigger utilizes input from the calorimeter subsystems, ECAL and HCAL, and the muon subsystems, DT, CSC, and RPC. A systems flow chart of the L1 Trigger is in Figure 3.12. The L1 Trigger is split into two paths, the calorimeter system and the muon system, which merge at the micro global trigger (μGT) for the final L1 Trigger decision. The muon system is composed of three muon triggers which are geometrically divided by $|\eta|$. The Barrel Muon Track Finder covers $|\eta| < 0.85$, the Overlap Muon Track Finder covers $0.85 < |\eta| < 1.25$, and the Endcap Muon Track Finder covering $1.25 < |\eta| < 2.4$. The resulting tracks from the three track finder systems are the L1 muon candidates.

Information from the calorimeter subdetectors is aggregated by the Calorimeter Layer-1 trigger where initial processing and calibrations are completed. Each of the 18 Layer-1 cards (CTP7s) receive calorimeter information from a single 20 degree ϕ slice of the detector [73]. The HCAL and ECAL input granularity is 0.087×0.087 in (η, ϕ) through the barrel. In the L1 Trigger, ECAL information is represented as a trigger tower, the sum of 5×5 groupings of ECAL crystals [64].

Following the 18 Layer-1 CTP7s is the Calorimeter Layer-2 trigger which further aggregates data from a singular event so that one Layer-2 hardware card (MP7) has calorimeter data for the full detector [73]. This allows computation of full event-based characteristics like the missing transverse energy. Refer to Section 5.2.2 for a description of missing transverse energy. A detailed layout of the L1 calorimeter trigger can be found in Figure 3.12 beginning with the 18 Layer-1 cards and progressing to the 9 main Layer-2 MP7s plus 3 spares. The Layer-2 MP7s execute rapid, firmware-based, algorithms to reconstruct basic physics objet candidates such as electrons and γs, hadronically decaying τ leptons, “jets”, and energy sums [73]. For a description of “jets”, see Section 5.3.1. The calorimeter-based and muon system-based physics objects are sent to the μGT where the final decision is made whether to pass an event along to the HLT or not.

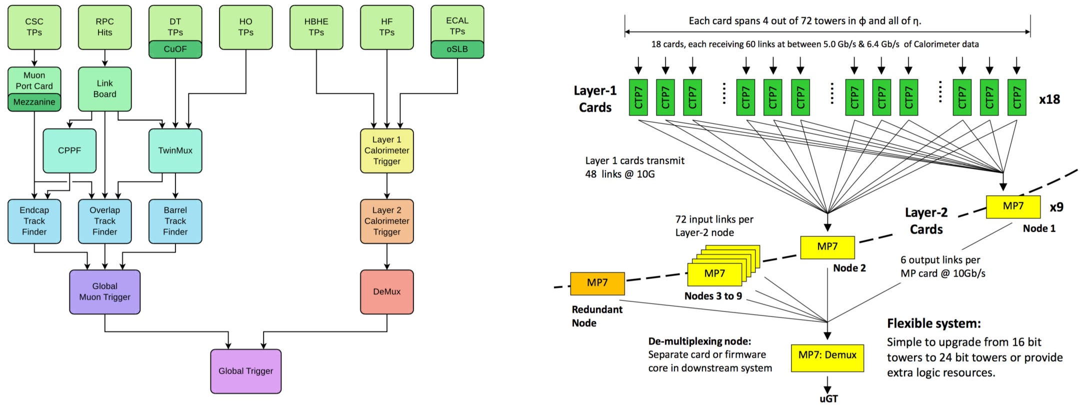

CSC TPs
RPC Hits
DT TPs
CuOF
HO TPs
HBHE TPs
HF TPs
ECAL TPs
oSLB
Muon Port Card
Mezzanine
Link Board
CPPF
TwinMux
Layer 1 Calorimeter Trigger
Endcap Track Finder
Overlap Track Finder
Barrel Track Finder
Layer 2 Calorimeter Trigger
Global Muon Trigger
DeMux
Global Trigger
Each card spans 4 out of 72 towers in ϕ and all of η.
18 cards, each receiving 60 links at between 5.0 Gb/s & 6.4 Gb/s of Calorimeter data
Layer-1 Cards
CTP7
x18
Layer 1 cards transmit 48 links @ 10G
72 input links per Layer-2 node
Layer-2 Cards
MP7
x9
Node 1
Node 2
6 output links per MP card @ 10Gb/s
Nodes 3 to 9
Redundant Node
De-multiplexing node: Separate card or firmware core in downstream system
MP7: Demux
Flexible system: Simple to upgrade from 16 bit towers to 24 bit towers or provide extra logic resources.
uGT

Figure 3.12: (left) System flow chart of the L1 Trigger, showing the complete trigger system, and input detector subsystems. (right) The L1 calorimeter trigger showing the number of cards and links between each system.

The Calorimeter Layer-1 Trigger systems was designed and built by the University of Wisconsin–Madison and is currently operated by a team of engineers, research scientists, and graduate students. I was part of the team of graduate students maintaining the system and would regularly be on-call in case of any trigger related issues. The excellent design and engineering that built the calorimeter trigger Layer-1 CTP7 hardware and firmware has resulted in a very robust system. More often than not, the Layer-1 on-call helps representatives from other subdetector systems diagnose their detector and link related problems.

HLT

Events which pass the L1 Trigger progress to the HLT. The significantly reduced event rate, from 40 MHz to 100 kHz, allows for a higher latency trigger which has access to information from all detector subsystems at full granularity. Specifically, the HLT has access to full ECAL crystal granularity and information from the inner tracker subsystem. At the start of Run 2, the HLT was composed of 15,000 commercial CPU cores running Scientific Linux [74]. The HLT reconstructs events in a similar way to what is used offline for event reconstruction, see Section 5. The exception is that the HLT algorithm is optimized to run with a hierarchical set of algorithms of increasing execution time and complexity. Some sacrifices in precision are made mostly related to the track reconstruction algorithms [64].

The HLT event processing is centered around the concept of an `HLT path`. An `HLT path` is a sequence of algorithmic processing steps containing steps which reconstruct physics objects and their attributes and other steps which make selections filtering out objects which fail certain requirements [64]. Well constructed HLT paths will save the most CPU-intensive steps for downstream in the path after many of the events have been filtered away by failing earlier selections.

A good example of an `HLT path` which progresses from simple steps towards more complex and CPU-intensive steps is that used for hadronically decaying τ (τ_h) lepton reconstruction. A common approach to separate τ_h from quark and gluon induced "jets" is to select well-isolated objects; the HLT relies on this exact approach. The first step of τ_h paths is to reconstruct basic "jet" 5.3.1 objects from calorimeter information only. Events which reconstruct calorimeter based "jets" passing selection criteria move to the next sequence which adds track information from the inner tracker to calculate the isolation of the associated "jet". Only after "jets" are determined to be well-isolated is the full HLT event reconstruction run [64]. This reduces computing time significantly and keeps the τ_h paths within the roughly 200 ms average processing time requirement for the HLT.

Just like the L1 Trigger, the HLT is used to decide if an event should progress to the next stage of data processing. In this case, an event which *passes* an `HLT path` is saved for future data processing. Events which *fail* all `HLT paths` are not stored. The final output rate of the HLT is about 1000 Hz on average over the full length of a run [74].

Tau HLT Upgrades in 2018

In 2017 and 2018, I have worked to help improve the τ_h `HLT paths` used at CMS. For all previous years, the τ_h `HLT paths` have used a cone-based τ_h reconstruction. The cone-based reconstruction uses a signal cone of $\Delta\mathrm{R} = 0.18$ containing the τ lepton decay products, and has an isolation annulus of $0.18 < \Delta\mathrm{R} < 0.45$. Improvements in the algorithm are designed to increase the alignment between the offline τ_h reconstruction algorithm and the HLT algorithm. Offline reconstruction uses the Hadron Plus Strips (HPS) algorithm, see Section 5.3.3. I have transferred the offline HPS algorithm to the HLT, optimized its expected performance and tested it against the previous cone-based reconstruction. For a similar selection efficiency the HPS algorithm is able to achieve a HLT rate reduction of roughly 20% 3.13. The HPS `HLT paths` will be tested during initial 2018 data taking. If online performance follows expectations, all of the primary τ_h `HLT paths` will be updated to using the HPS algorithm.

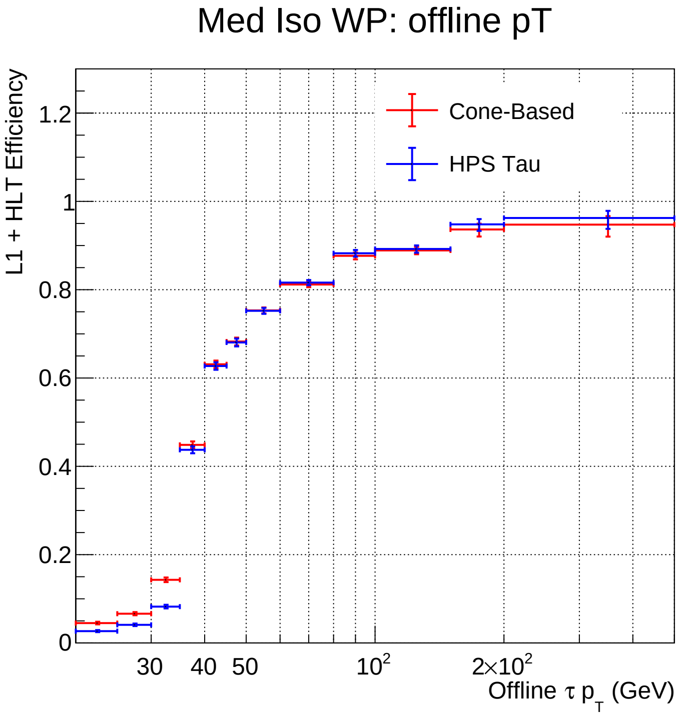


Figure 3.13: Monte-carlo VBF $H \rightarrow \tau\tau$ based efficiency distribution showing the cone-based HLT algorithm in red and the HPS-based HLT algorithm in blue. For nearly identical efficiency performance it is expected that the HLT rate can be reduced by up to 20%.

Chapter 4

Simulation

In this chapter I discuss how high energy particle collisions are simulated. Simulated collisions, based on knowledge of the standard model, provide the foundation of our predictions in high energy particle physics. The wealth of particle physics knowledge accumulated over the last 60 years has provided a testing grounds to compare experimental results with simulated particle collisions. Based on past experiments and extensive work from the simulation community, we are able to simulate and model the the proton-proton collisions which take place in the CMS detector with a high degree of precision. In this thesis, we are specifically interested in comparing two physics scenarios: the standard model without the existence of a 125 GeV Higgs boson versus the standard model with the existence of a 125 GeV Higgs boson. Both of these predictions rely on the standard model processes and Higgs boson processes being well modeled via simulation. In the following sections, I briefly discuss how events are simulated for proton-proton collisions, how the initial products of the collision decay, the steps of parton showering, fragmentation and hadronization, and how the decay products are simulated to interact with a model of the CMS detector.

There are many different Monte Carlo (MC) event generators which can contribute to the simulation of high energy particle collisions and decays. There are three main MC event generators which are commonly used to model the full proton-proton collision: PYTHIA [75], HERWIG [76], and SHERPA [77] including both hard scattering processes and subsequent evolution of the products to final state products.Several other generators, including POWHEG [78, 79] and MADGRAPH5_aMC@NLO [80, 81], are also used to model only hard scattering processes and, in this thesis, interface to PYTHIA8 [75] for evolution to the final state products. PYTHIA8 includes a set of physics models which can simulate the particle evolution from a few-body hard-scattering process through to a complex multiparticle final states.

To produce the physics processes used in this thesis we rely on MC event generators besides PYTHIA8 for the hard-scattering process. Then we switch back to PYTHIA8 to evolve the hard-scatter products through to their final state products. The choice to stitch together multiple MC generators is made when the internal hard-scattering physics library in PYTHIA8 is not as complete as other generators or other generators are able to achieve higher order accuracy in their treatment of the hard-scattering process.

The signal samples with a Higgs boson produced through gluon fusion (ggH), vector boson fusion (VBF), or in association with a W or Z boson (WH or ZH), are all generated at next-to-leading order (NLO) in perturbative quantum chromodynamics [82] with the POWHEG 2.0 [83, 79, 84, 85] generator. Details on these four Higgs boson processes can be found in the Higgs Phenomenology chapter Section 2.4 and Figure 2.2. The MINLO HVJ extension of POWHEG 2.0 is used for the WH and ZH simulated samples [86]. The POWHEG 2.0 generator is used for the $t\bar{t}$, qq $\rightarrow ZZ$, and WZ standard background processes while POWHEG 1.0 is used for simulating single top quark production. Additionally, POWHEG is used for the $WH \rightarrow WWW$, $ZH \rightarrow ZWW$, and $H \rightarrow ZZ$ backgrounds which all include non-signal Higgs bosons production and decay. The MadGraph5_aMC@NLO generator is used for the Z+jets, W+jets, $t\bar{t}Z$, and triboson processes to simulate events at leading order (LO) [87]. The MadGraph5_aMC@NLO generator is used for diboson and $t\bar{t}W$ production simulated at next-to-LO (NLO) [88]. The gg $\rightarrow ZZ$ process is generated at leading order (LO) with MCFM [89].

4.1 Hard-Scattering Process

The hard-scattering process is the main process of interest in a simulated event. It is the portion of the simulation describing the interaction and annihilation of the two partons from the colliding protons. The hard-scattering process creates the signal and background processes of interest in this thesis, such as W, Z, and Higgs bosons. For all simulated signal and background samples, the hard-scattering process is defined by the standard model Lagrangian. The specific processes studied in this thesis can be represented by matrix elements and their resulting Feynman diagrams.

The MadGraph5_aMC@NLO generator can be operated at both LO and NLO accuracies. After supplying MadGraph5_aMC@NLO with a process of interest including the input partons and resulting particles, it generates the Feynman rules describing the process and can translate them into Feynman diagrams [90]. At LO, this process is fully automated using FeynRules [90, 91]. At NLO, the NLO QCD, the process is fully automated via FeynRules; other NLO contributions must be added with dedicated computations [81].

The Powheg event generator is used for production of the Higgs boson signal processes in this thesis. Powheg does not have the convenient automation which comes with MadGraph5_aMC@NLO, instead each physics process must be coded separately. However, the large amount of interest in Higgs boson physics has led to a thoroughly developed and validated process library resulting in NLO accuracy [84, 85, 92, 93, 94].

4.2 Parton Distribution Functions

One of the unique complexities present at a hadron collider, which is avoided with a lepton collider, is accounting for the substructure of the colliding hadrons. We must account for this to create proper simulations of LHC proton-proton collisions. The internal structure of the proton has been probed over the past half century through deep inelastic scattering experiments [12, 11]. Experimental results show the internal structure of protons revealing the existence of quarks and gluons. It is these quarks and gluons which collide and interact in proton-proton collisions. The probability for each quark or gluon to have a certain fraction of the proton's momentum is given by the parton distribution function (PDF). The PDFs depend on the energy scale of the proton. All simulations in this thesis rely on PDFs for an energy scale of 6.5 TeV matching the LHC beam.

The specific PDFs which are used in this thesis are provided by `NNPDF3.0` with the exact PDF set being `NNPDF30_nlo_as_0118` [95, 96]. The `NNPDF3.0` PDFs are derived using a global dataset including, but not limited to, data from HERA, ZEUS, ATLAS, LHCb, and CMS. Functional forms derived from theoretical QCD predictions with electroweak corrections are fit to the available data [95]. The PDFs provide proper normalization based on the probabilistic distribution of the available partons and their energies as inputs to the hard-scattering process.

4.3 Underlying Event

In addition to the particles resulting from the hard-scattering portion of a proton-proton collision, there are additional particles resulting from what is called the underlying event. The underlying event consists of interactions between partons which are not directly associated with hard-scattering event, for example secondary partons within the colliding protons. Underlying event collisions are softer and lower energy in nature which limits the range of the resulting physics processes. The most common underlying event interactions result in two partons interacting resulting in a single parton which then combines with other partons to form outgoing hadrons. This is only an intermediary stage as color confinement eliminates the possibility of a single isolated parton propagating through the CMS detector, this will be revisited shortly.

The underlying event in Monte Carlo event generators characterize the kinematics and composition of soft "jets". There are specific "jet" related physics observables which are sensitive to the characteristics of the underlying event [97, 98]. For a description of "jets" see Section 5.3.1. The initial underlying event tune for PYTHIA8 is the Monash Tune [99]. A CMS specific tuning of PYTHIA8 has been constructed based off of the parameters of the Monash Tune but adding some additional energy-dependent parameters to the fit. The tune is based on data from CDF and 7 TeV CMS data and is called `CUETP8M1`. Considering `CUETP8M1` is derived based largely on data from CMS, there is very good agreement between CMS data and simulations based on the `CUETP8M1` tune [97].

4.4 Parton Showers

While matrix elements are used to calculate the hard-scattering event, the use of matrix elements to fully describe the production and decay of all of the partons is technically not possible with current best Monte Carlo generators [100]. Instead, parton shower programs handle radiation of soft gluons from color charged particles and photons from electrically charged particles [101]. Initial state radiation (ISR) occurs when an incoming charged particle radiates before entering the hard-scattering process [41]. When an outgoing particle radiates, it is called final state radiation (FSR). Partons from ISR, FSR, the underlying event and the hard-scatter process are decayed using the parton shower method.

Parton shower evolution can be viewed as a probabilistic process where each produced parton has a probability of radiating another parton. Partons iteratively emit other partons through three dominant processes: $q \to gq$, $g \to gg$, and $g \to q\bar{q}$. Swap q for $\bar{q}$ to complete the set.

Technically, there is a matching and merging of the hard-scatter matrix element with the parton shower contribution. This does not affect the total process cross section. The Z+jets and W+jets samples, which were simulated with MadGraph5_aMC@NLO at LO, use MLM jet matching and merging [87]. The diboson production samples simulated with MadGraph5_aMC@NLO at NLO use the FxFx jet matching and merging [88]. The Monte Carlo generators used in this thesis are interfaced with Pythia 8.212 [75] to model the parton shower evolution.

4.5 Fragmentation and Hadronization

As mentioned previously, due to color confinement, quarks and gluons cannot exist in isolation and must be transformed in color-neutral final states [102]. To resolve the lone quarks and gluons which resulted from the hard-scatter process and the parton shower, the partons undergo hadronization. Hadronization is the process where hadrons are formed out of quarks and gluons. There are multiple phenomenological models of hadronization, the specific one used in this thesis is the Lund String Model [103]. The Lund String Model of jet hadronization is based the observation that within a meson system the quark-antiquark potential rises linearly with the distance between quarks [104]. In the model, this potential energy is stored in flux lines connecting the quark-antiquark pair, see left pane Figure 4.1. As a quark-antiquark pair with high energy moves apart the energy stored between them in the flux lines increases. If the energy is sufficiently high the flux line can be severed and a second quark-antiquark pair created from the freed energy. This can be visualized in the right pane of Figure 4.1. This process repeats itself creating stable mesons and baryons where the low energy (anti)quarks oscillate about each other stably, represented as the zig-zag lines in Figure 4.1.

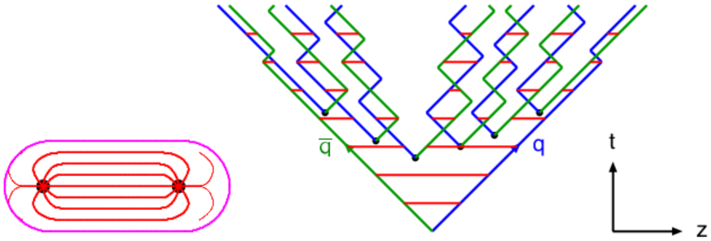


Figure 4.1: (left) Lund String Model flux tubes connecting a quark-antiquark pair. (right) A space-time representation of the hadronization of a quark-antiquark pair. The final representation depicts a seven mesons final state.

Pythia8 also decays vector boson sometimes creating pairs of quarks where it then models the subsequent parton shower, fragmentation, and hadronization. In addition, Pythia8 is also used to model τ lepton decays where the spin correlations are fully included. All τ decay modes with branch ratios greater than 0.04% are included in the simulation [105].

4.6 Pileup

There were 27 proton-proton collisions per bunch crossing on average in the 2016 data collected by CMS (Figure 6.3). For most bunch crossings all of these proton-proton collisions are soft scattering events and, with no hard process present, the event is unlikely to be stored for future physics analysis because it will fail the Level-1 or High Level Trigger selections. When there is a hard-scattering process present in a bunch crossing, the soft scattering collisions will still be present and must be modeled. To emulate this effect in simulated events, in addition to the two protons involved in the hard-scattering interaction, additional soft proton-proton collisions are added. These collisions are referred to as “pileup” and are generated using PYTHIA8. The distribution of the number of soft scattering events added to a simulated sample is intended to align with the distribution observed in the 2016 data. Because alignment is never 100% perfect immediately after the samples are simulated, the simulated events are weighted to adjust their distribution to the data based on the measured instantaneous luminosity for each bunch crossing which varies with time and changes through an LHC fill, see Section 6.3.

4.7 Detector Simulation

At this stage of event simulation, the simulated events model our best approximation of the proton-proton collisions and resulting decays and hadronization taking place in the CMS detector. There is one critical last step to complete the event simulations. The decay products must interact with a simulated model of the CMS detector. The result of this final stage is simulated events that are stored in the same data format as the data gathered by the detector itself, raw energy deposits and tracker hits.

The GEANT4 software toolkit is used to simulate the CMS detector [106]. At its most basic, GEANT4 is a toolkit for simulating the passage of particles through matter. It contains a vast library of functionality allowing the creation of model physics detectors made out of any desired material and in any desired geometrical configuration. To initially validate the GEANT4 modeling of the CMS detector, test beam data and collision data are used. Good agreement is seen between the data and the simulated particles passing through the CMS detector for the energy response and resolution for pions and protons [107].

After the decay products of the simulated events have passed through the GEANT4 simulation, they are stored in the same format as data is originally gathered. From this point forward, data and simulated events are reconstructed with the exact same algorithms.

Chapter 5

Object Reconstruction and Selection

At CMS, particle and object reconstruction draws on input from all detector subsystems simultaneously to build particle tracks and cluster together energy deposits, to link together these tracks and energy deposits to construct basic physics-objects such as electrons and charged hadrons, and to build composite objects such as "jets" and hadronically decaying τ leptons. Event based quantities such as the $\vec{E}_T^{miss}$ are also reconstructed. To achieve all of this, CMS uses its Particle Flow (PF) reconstruction algorithm [108]. The particle flow concept has been used in the past by other experiments such as ALEPH at LEP [109]. CMS is the first experiment to fully utilize the particle flow technique in a hadron collider environment.

5.1 Particle Flow Input

The CMS detector and the PF reconstruction algorithm compliment each other. The CMS detector features a highly-segmented tracker well suited to precise track reconstruction, a fine-grained electromagnetic calorimeter necessary to separate the individual energy deposits from particles within “jets” and for efficient photon and electron identification, a hermetic hadron calorimeter for the measurement and identification of charged and neutral hadrons, a strong magnetic field for the measurement of the momenta of charged particles and to separate the calorimeter energy deposits of charged and neutral hadrons within “jets”, and a muon spectrometer for muon identification and to separate the muon tracks from other tracks in the tracker. A schematic of a slice of the CMS detector and different physics-objects transversing the detector subsystems can be see in Figure 5.1. The different detector systems all contribute necessary pieces of information to the PF reconstruction. From the raw detector signals two classes of PF objects are created, tracks and energy clusters.

5.1.1 Particle Flow Tracks

Energy deposits, often called “hits”, are recorded by the pixel and strip tracker and the muon detectors after a collision . From these, charged particle tracks are reconstructed in subsequent layers mapping the progression of charged particles from the beam axis outwards into the detector volume. Attempting to reconstruct tracks from every possible hit combination quickly because intractable considering the over 70 million tracker pixels and strips which can each record a hit. PF uses a combinatorial track finder based on Kalman Filtering (KF) [110]; the algorithm is broken down into three successive steps.

- Generate seed tracks from a few hits which are compatible with a charged particle trajectory

- Gather other hits along the seed track trajectory when propagated through the rest of the tracker subsystem

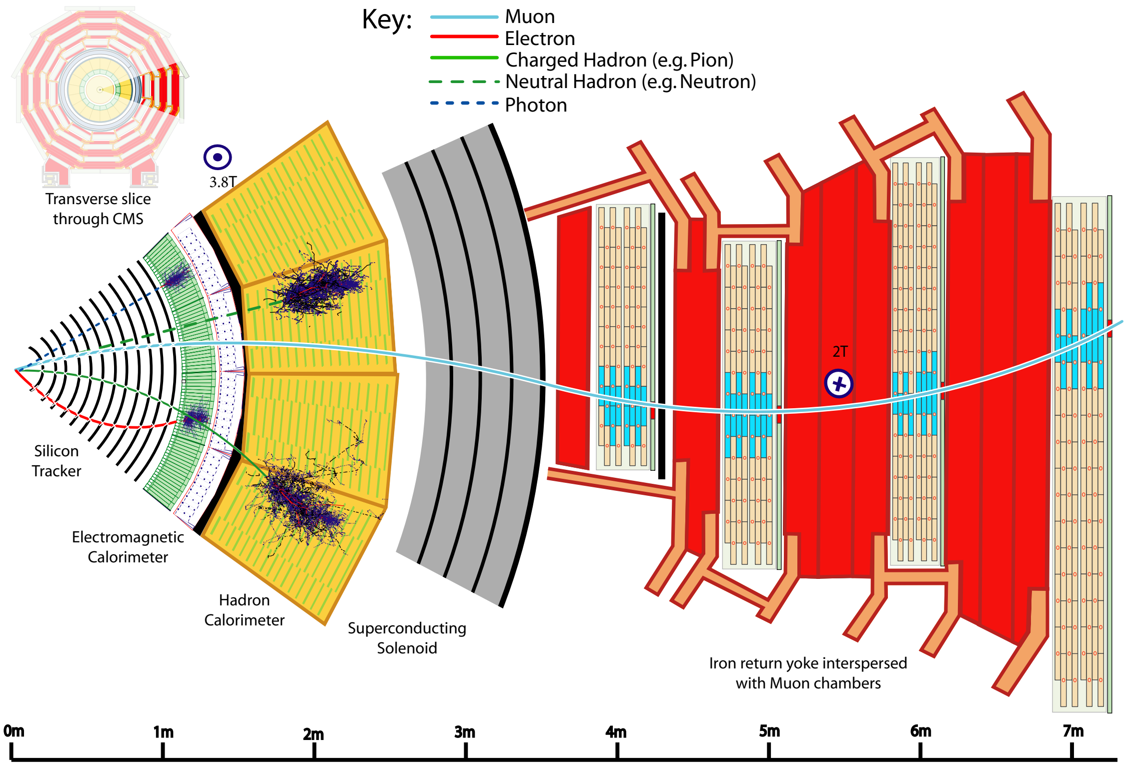

Key:
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon
Transverse slice through CMS
3.8T
2T
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
Iron return yoke interspersed with Muon chambers
0m
1m
2m
3m
4m
5m
6m
7m

Figure 5.1: A schematic of a slice of an x-y cross section of the CMS detector showing different physics-objects such as electrons, photons, charged and neutral hadrons, and muons propagating outwards from the collision region within the detector. The schematic shows how tracks are linked to energy deposits and in which subdetectors different particles deposit most of their energy on average.

- Fit the final track to determine track properties such as the origin, transverse momentum, and direction.

Only tracks meeting certain quality standards are kept for analysis. These tracks must be seeded with two hits in consecutive layers in the pixel detector, and are required to be reconstructed with at least eight tracker hits in total, and with at most one missing hit along the track trajectory. Tracks must also have a curvature corresponding to a momentum greater than 0.9 GeV.

There is a balance between imposing tight quality cuts on reconstructed tracks, which increases the purity of genuine track within the reconstructed track collection, but also decreases the efficiency for reconstructing genuine tracks, and loosening quality cuts to reconstruct genuine tracks with a higher efficiency and lower purity. After an initial pass through what is called the global combinatorial track finder, which has the stringent track quality criteria imposed by the eight-hit requirement mentioned above, the efficiency to reconstruct genuine tracks is roughly 80% for charged pions with $p_{\mathrm{T}} = 10\,\mathrm{GeV}$, and 99% for isolated muons. This corresponds to a misreconstructed track rate of about 2.5% for charged pions with $p_{\mathrm{T}} = 10\,\mathrm{GeV}$ [108], see Figure 5.2. As hits are recorded in a reconstructed track, they are removed from the available, unused hits which can be combined in subsequent passes through the tracking algorithm.

There are ten passes through the tracking algorithm [108] in total. Each pass loosens the track quality criteria, such as χ^2 and number of hits requirements, beyond the stringent initial criteria. This helps to reconstruct difficult to build tracks. Tracks can be difficult to reconstruct for multiple reasons such as detector inefficiencies leading to missing detector hits, or particles originating from elsewhere in the detector besides along the beam axis. These tracks could be from hadrons interacting within the tracker material before reaching the eight-hits threshold, or from the decay of particles with finite lifetimes. Additionally, there can be difficulty disentangling the many tracks within a collimated "jet" where many of the tracks are close to or nearly overlapping with one another. The misreconstruction rate is suppressed in each iterative step despite the loosening quality criteria by the removal of the hits which are previously incorporated into a reconstructed track. This suppresses the random hit-to-seed association in the next iteration and allows moderate efficiency gains for only small misreconstruction losses. The efficiency and misreconstruction rate in Figure 5.2 shows the results for i) the initial pass through the tracking algorithm, ii) the results after all passes which require the track seed to contain hits in the pixel detector, and iii) the final results after considering displaced tracks [111]. Displaced tracks are tracks which do not originate from the collision region. They can be created from a number of processes included Bremsstrahlung and the inelastic nuclear interaction of particles with the tracker material. After all iterations the efficiency is about 90% for a charged pion with $p_{\mathrm{T}} = 10\,\mathrm{GeV}$ for a misreconstruction rate of 3%.

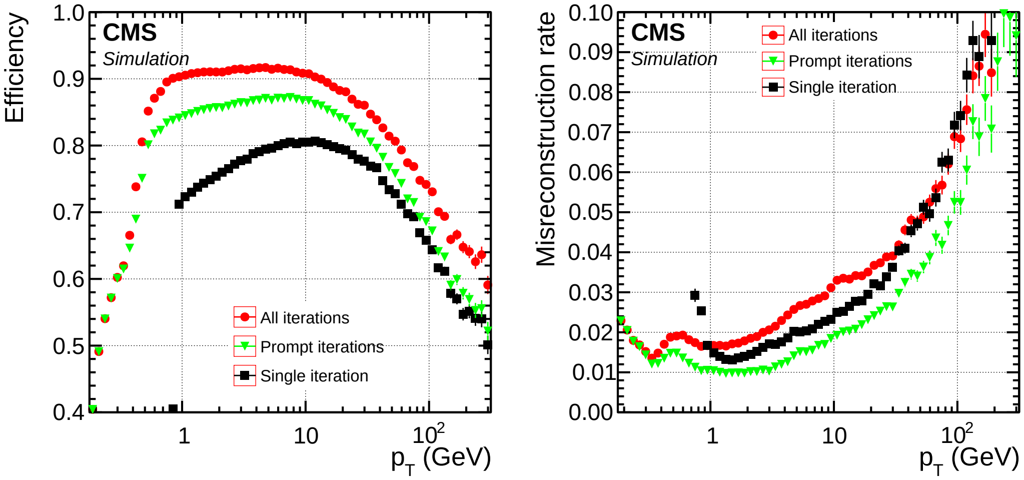


Figure 5.2: Efficiency (left) and misreconstruction rate (right) of the global combinatorial track finder (black squares) which is the first pass through the tracking algorithm. The prompt iterations of the tracking method (green triangles) show the results after all iterations based on seeds with at least one hit in the pixel detector are completed. The final results after all iterations (red circles) includes iterations with displaced seeds. Efficiency and misreconstruction rate are plotted as a function of the track p_{T}, for charged hadrons in multijet events without pileup interactions. Only tracks with $|\eta| < 2.5$ are considered. The efficiency is displayed for tracks originating from within 3.5 cm of the beam axis and ±30 cm of the nominal center of CMS along the beam axis.

Tracks which are likely associated with electrons receive special treatment in PF. Electrons will often emit bremsstrahlung radiation while propagating through the tracker. When energetic photons are radiated from an electron, the pattern recognition in the KF algorithm may have difficulty accommodating the sudden change in electron momentum. This can cause the track to be reconstructed with a smaller number of hits than would be associated with the true electron path. A new collection of tracks is created based on a preselection on the number of hits and the χ^2 for the reconstructed KF-based tracks. The new collection, which is a subset of the KF-based tracks are fit again with a Gaussian-sum filter (GSF) [112]. Instead of modeling the energy loss of particles as a single gaussian probability density function (PDF) like the KF algorithm does, the GSF models the energy loss as a mixture of multiple gaussian PDFs. The GSF fitting algorithm is more adapted to electrons than the KF algorithm. It allows for sudden and substantial energy losses along the trajectory. This additional freedom allows much better fits for the electron-based tracks and provides more precisep_{T} estimates [112].

Muon tracks are built from a combination of the pixel and strip tracker (inner tracker) information and the muon spectrometer information. The hits in the muon spectrometer are quite pure in genuine muon hits. This is because the calorimeters and the solenoid absorb the vast majority of non-muon particles, except neutrinos, before they reach the muon spectrometer. There are three different categories of muon tracks reconstructed using the KF technique:

- Standalone muons are based on hits within each DT, CSC, and RPC detector. Hits in the DT and CSC detectors are clustered to form track segments that are used as seeds for the pattern recognition in the muon spectrometer to gather other hits in the all three muon systems along the track trajectory.
- Global muons are reconstructed from a standalone-muon track which is matched to a track from the inner tracker. If the trajectories of the two tracks propagate onto a common surface they are considered compatible and the hits from the inner tracker and from the standalone-muon track are combined and fit to form a global-muon track.

- Tracker muons are built from an extrapolation of a track from the inner track system to a minimum of one compatible muon segment within the muon spectrometer system. The inner tracker-based track must have a $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$ and a total momentum $p > 2.5\,\mathrm{GeV}$.

About 99% of the muons produced within the geometrical acceptance of the muon system are reconstructed either as a global muon or a tracker muon and very often as both [108]. For muons with $p_{\mathrm{T}} > 200\,\mathrm{GeV}$, the momentum resolution, based on the inner track, is improved by the inclusion of the track extension to the muon system. For muons with $p_{\mathrm{T}} < 200\,\mathrm{GeV}$, the inner tracker already provides a precise measurement of their momentum.

5.1.2 Particle Flow Energy Clusters

Energy clusters make up the second basic PF building block. The PF energy clustering algorithm used to construct the energy clusters serves multiple purposes:

- Detect and measure the energy and direction of stable neutral particles (photons and neutral hadrons)
- Separate neutral particles from charged hadron energy deposits
- Reconstruct and identify electrons and all accompanying bremsstrahlung photons
- Assist the energy measurement of charged hadrons for which the track parameters were not determined accurately; this is primarily the case for low-quality and high-p_{T} tracks

The clustering is performed separately in each subdetector, the ECAL barrel and endcaps and HCAL barrel and endcaps with an aim of a high detection efficiency even for low-energy particles and the ability to separate close energy deposits. Clustering begins with seed hits which have an energy above the seed hit threshold and an energy larger than the energy of the adjacent hits. The fine granularity of the ECAL barrel allows for the clustering to consider all eight adjacent hits, four on the sides and four on the corners. The HCAL barrel has a more coarse granularity. Only the four adjacent sides are considered when selecting a seed candidate. From the starting seed, topological clusters are grown outward by aggregating hits with at least a corner in common with a cell already included in the cluster [108]. Hits must have an energy in excess of two times the subdetector noise level to be considered for clustering. The energy thresholds for seeding a cluster and cluster inclusion threshold are in Table 5.1.

	ECAL		HCAL	
	barrel	endcaps	barrel	endcaps
Cell E threshold (MeV)	80	300	800	800
Seed Number closest cells	8	8	4	4
Seed E threshold (MeV)	230	600	800	1100
Seed E_{T} threshold (MeV)	0	150	0	0

Table 5.1: The clustering parameters used for ECAL and HCAL energy deposit clustering. The ECAL endcap requires an additional seed E_{T} threshold because the detector noise increases as a function of $|\eta|$.

Residual energy calibrations are applied to the ECAL and HCAL energy clusters. The calibrations are designed to account for the effects of the hit energy thresholds which will always result in a smaller amount of energy being incorporated into a cluster than was measured by the detector for a given single object. In the ECAL, the residual energy calibration is determined from simulated single photon events. This generic calibration is applied to all ECAL clusters prior to the hadron cluster calibration mentioned next.

ECAL and HCAL energy clusters are linked together as a potential hadronic decay energy deposit if their positions in (η, ϕ) overlap. For a hadronic decay, the total calorimeter response (ECAL + HCAL) depends on the fraction of the shower energy deposited in the ECAL, and is not linear with energy. The ECAL and HCAL cluster energies are calibrated to get an estimate of the true hadron energy. Simulated single neutral hadrons, specifically K^0_Ls, are used for the hadronic decay response calibration seen in Figure 5.3. The applied calibrations in the left plot lead to excellent agreement in the calorimeter response in the right plot.

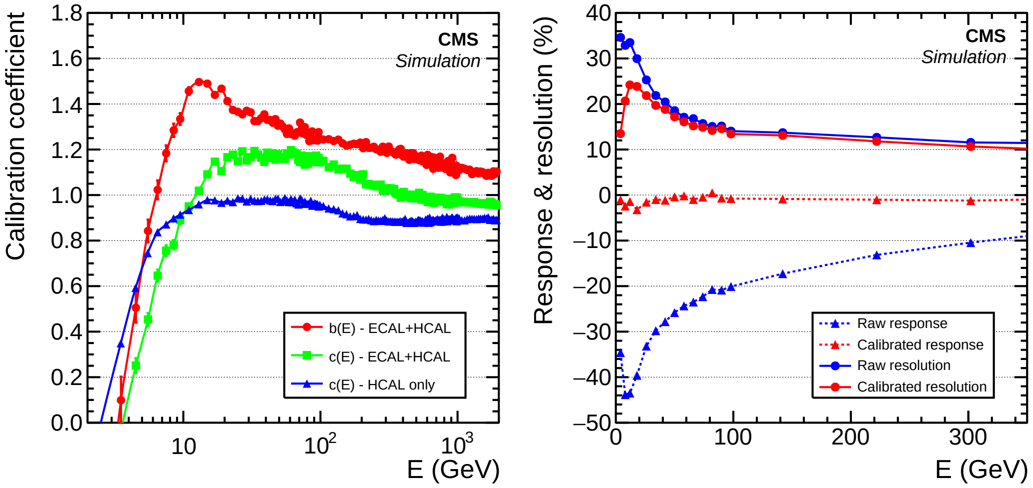


Figure 5.3: (left) Calibration coefficients obtained from single K^0_Ls in the barrel as a function of their true energy E. The blue triangles show the calibrations for hadrons depositing energy only in the HCAL. The red circles (green squares) show the ECAL (HCAL) calibration for hadrons depositing energy in both the ECAL and HCAL. (right) Relative energy response (dashed curves) for the raw (blue) and calibrated (red) energy, and energy resolution (solid curves). Both for single K^0_Ls in the barrel as a function of their true energy E.

5.1.3 Particle Flow Candidates

In general, a given particle (physics-object) propagating through the CMS detector is expected to result in multiple PF elements (tracks and energy clusters) in various CMS subdetectors. The reconstruction of a particle begins with a linking algorithm that connects the PF elements from different subdetectors. For example, tracks are linked to energy clusters if the extrapolated trajectory of the track aligns within the angular acceptance of an energy cluster in (η, ϕ). Energy clusters can be linked between subdetectors as mentioned in the case of hadronic energy deposits above which can span both the ECAL and HCAL. A link is established when the cluster position in the more granular calorimeter (ECAL) is within the cluster envelope in the less granular calorimeter (HCAL). Once PF elements are linked together they are referred to as a PF block. Particle Flow candidates are selected from the PF blocks based on compatibility with different physics-object characteristics and designated quality cuts.

Muons

Isolated global muons are selected from the global muon track collection by looking at the inner tracker tracks and calorimeter energy deposits within a distance $\Delta\mathrm{R} < 0.3$ to the muon trajectory. The sum of the p_{T} of these tracks and of the E_{T} of the energy deposits is required not to exceed 10% of the muon p_{T}. This isolation criterion alone successfully reject hadrons that would otherwise be misidentified as muons. No further selection is applied to these muon candidates. If the muon track $p_{\mathrm{T}} < 200\,\mathrm{GeV}$, then the momentum assigned to the muon PF candidate is that of the inner track. For muon tracks with $p_{\mathrm{T}} > 200\,\mathrm{GeV}$, the momentum assigned is the momentum associated with the smallest χ^2 probability from multiple different track fits: tracker only, tracker and first muon detector plane, global, and global without the muon detector planes featuring a high occupancy [113]. The PF elements and blocks that make up an identified PF muon candidate are masked against further processing to prevent their inclusion in other PF candidates.

In these analyses, there are several additional criteria applied beyond being a basic PF muon candidate. These analyses use muons which pass two different PF muon identification working points: `PF ID Loose` and `PF ID Medium` [114]. The `PF`

`ID Loose` working point is only slightly tighter than the baseline criteria for a PF muon candidate; a `PF ID Loose` muon must be either a global or a tracker PF muon. This selection is highly efficient for prompt muons.

The `PF ID Medium` muon working point requires first that muons pass `PF ID Loose` and then applies additional track-quality and muon-quality requirements on the different muon tracks which are linked to the PF muon candidate. The number of valid inner tracker hits must be greater than 80%. Additionally, either one of the following criteria must be met:

- Option 1 - “Tight Segment Compatibility”
    - Candidate has a segment compatibility [115] score of at least 0.451 which ensures that the track is reasonably compatible with inner tracker-based track
- Option 2 - “Good Global Muon”:
    - PF muon candidate is a global muon
    - The normalized track $\chi^2 < 3$
    - The compatibility χ^2 between the standalone muon track and the inner tracker muon is less than 12
    - The muon track kink-finder, which is designed to remove muons produced from in-flight decays, must have a value less than 20
    - Candidate has a segment compatibility score of at least 0.303, which is looser than the value required in Option 1

The `PF ID Medium` muon working point is very efficient for prompt muon selection but does bring some additional reduction in fake object selection.

To reject non-prompt or misidentified muons and electrons, a relative lepton isolation is defined as:

$$I^{\ell} \equiv \frac{\sum_{charged} p_{\mathrm{T}} + \max\left(0, \sum_{neutral} p_{\mathrm{T}} - \frac{1}{2}\sum_{charged,PU} p_{\mathrm{T}}\right)}{p_{\mathrm{T}}^{\ell}}. \tag{5.1}$$

Some of the quantities mentioned here refer to descriptions in the following sections. In this expression, $\sum_{charged} p_{\mathrm{T}}$ is the scalar sum of the transverse energy of the charged particles originating from the primary vertex (see Section 5.2.1) and located in a cone of size $\Delta\mathrm{R} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.4\,(0.3)$ centered on the muon (electron) direction. The sum, $\sum_{neutral} p_{\mathrm{T}}$, represents a similar quantity for neutral particles. The contribution of photons and neutral hadrons originating from pileup vertices is estimated from the scalar sum of the transverse energy of charged hadrons in the cone originating from pileup vertices, $\sum_{charged,PU} p_{\mathrm{T}}$. This sum is multiplied by a factor of 1/2, which corresponds approximately to the ratio of neutral to charged hadron production in the hadronization process of inelastic pp collisions, as estimated from simulation. The expression p_{T}^{ℓ} stands for the p_{T} of the lepton. Isolation requirements used in the following analyses for electrons and muons range from $I^{\ell} < 0.1$ to $I^{\ell} < 0.25$ depending on the signal efficiency and background rejection needs of the specific final state.

Electrons and Prompt Photons

With the muon related PF elements masked from further processing, electron and prompt photon identification begins. Electron identification is based on information from the inner tracker tracks and the calorimeter energy clusters. Because of bremsstrahlung radiation, electron tracks can be much more kinked than those for other particles as was discussed above in relation to the GSF algorithm track fitting. Additionally, because of radiated bremsstrahlung photons, the resulting energy clusters from an electron bending in the magnetic field can be spread out in the ϕ direction. PF electrons are built from the linking of a GSF track to an ECAL-based energy cluster. To suppress the amount of charged hadrons faking electrons, the sum of the energies measured in HCAL hits behind ($\Delta \mathrm{R} < 0.15$) the electron-linked ECAL energy cluster must not exceed 10% of the ECAL-based energy cluster energy [108]. The energy assignment for an electron candidate is obtained from a combination of the calibrated ECAL energy with the momentum of the GSF track. The electron direction is chosen to be that of the GSF track.

Before being saved to the PF electron collection, electron candidates must satisfy additional identification criteria targeted at reducing electron fakes. Up to fourteen variables are fed into a Boosted Decision Tree (BDT) which determines the passing PF electrons [108]. The BDT input variables include track and energy cluster details such as:

- Amount of energy radiated off the GSF track
- Distance between the GSF track extrapolation to the ECAL entrance and the position of the ECAL cluster
- Ratio between the HCAL and ECAL energies
- The KF and GSF track χ^2 values, and
- The numbers of inner tracker hits

Photon candidates are seeded by ECAL energy clusters with no matching KF or GSF track. They are retained as PF photons if they are isolated from other tracks and calorimeter energy clusters, and if the ECAL energy distribution and the ratio between the HCAL and ECAL energies, H/E, are compatible with those expected from a photon shower. Similar to the PF masking after the muon reconstruction, tracks and energy clusters used to build PF electron and photons are masked from further processing, simplifying the task ahead for charged and neutral hadron identification.

There are additional electron identification requirements used in these analyses which are tighter than the PF electron criteria. The additional identification criteria rely on a multivariate (MVA) discriminant which combines many of the same variables used in the PF electron BDT [116] and adds some additional energy cluster distribution variables such as $\sigma_{i\eta i\eta}$ and $\sigma_{i\phi i\phi}$ cluster shape covariance. The analyses use two different MVA working points, one with 90% signal efficiency and one with 80% efficiency.

Charged and Neutral Hadrons

Once muons, electrons, and isolated photons are identified and removed from the available PF blocks, the remaining particles to be identified are hadrons resulting from jet fragmentation and hadronization. The ECAL and HCAL energy clusters which are not linked to any tracks are turned into PF neutral hadrons while energy deposits successfully linked to a track are turned into PF charged hadrons. Non-isolated photons are indistinguishable from the neutral hadron group. Charged and neutral hadrons form the last sets of fundamental physics-objects which are reconstructed by PF. The next PF steps involve reconstructing composite object such as "jets" and hadronically decaying tau leptons and calculating event quantities such as the primary vertex and $\vec{E}_T^{miss}$.

5.2 Event Level Quantities

There are multiple event level quantities that require input from all PF physics-objects for their calculation.

5.2.1 Primary Vertex Reconstruction

The original location of the pp collisions which gave rise to a given event can be found by tracing the reconstructed physics-object tracks back to the collision region and grouping together tracks that share a common origin; the origins are called vertices. In any given recorded collision there will usually be a singular hard-scatter pp collision and multiple soft-scatter collisions. The hard-scatter vertex is identified as the vertex with the largest quadratic sum of the p_{T} of the associated physics-objects and is called the primary vertex [108]. The other vertices are referred to as the pileup vertices. The calculation of the primary vertex is specifically needed for the identification of composite objects such as "jets" and hadronically decaying τ leptons.

5.2.2 Missing Transverse Energy

The CMS detector can detect and measure the energy of all standard model particles with the exception of neutrinos which leave the detector undetected. The neutrino energy contribution to an event can be estimated using the missing transverse energy, $\vec{E}_T^{miss}$. The $\vec{E}_T^{miss}$ is calculated from the missing transverse momentum vector which is defined to balance the vectorial sum of the transverse momentum of all particles.

$$\vec{p}_{\mathrm{T,PF}}^{\mathrm{miss}} = -\sum_{i=1}^{N_{\mathrm{particles}}} \vec{p}_{\mathrm{T},i} \tag{5.2}$$

All particles reconstructed in the event are used to determine the missing transverse energy, $\vec{E}_T^{miss}$ [117]. The specific $\vec{E}_T^{miss}$ used in these analyses is Type-1 $\vec{E}_T^{miss}$ which is adjusted for the effect of jet energy corrections.

5.3 Composite Object Identification and Selection

Composite objects reconstructed by PF are clustered together into physics-objects which likely resulted from the jet fragmentation or hadronization process. The three groups used in these analyses are discussed below.

5.3.1 Jets

Jets are collections of energy deposits and tracks within a defined conical area radiating outward from the collision region. They are created when a quark or gluon undergoes the hadronization process. Jets are reconstructed with an anti-k_{T} clustering algorithm implemented in the FASTJET library [118, 119, 120]. The anti-k_{T} clustering is based on the grouping together of neutral and charged PF candidates within a distance parameter $\Delta\mathrm{R} = 0.4$. Charged PF candidates not associated with the primary vertex are not considered when building jets. A correction is applied to jet energies to adjust for the contribution to the jet energy from additional *pp* interactions within the same or nearby bunch crossings. The energy of a jet is corrected via calibrations based on simulation and data [121].

5.3.2 b-jet Identification

The combined secondary vertex (CSV) algorithm is used to identify jets that likely resulted from a b quark decay, a “b-jet” [122, 123]. Algorithms designed for b-jet identification exploit the long lifetime of b-hadrons which can be present in jets originating from the hadronization of b-quarks. This long lifetime results in a decay of the b-hadron that is displaced with respect to the primary vertex. The CSV algorithm uses track-based lifetime information together with the reconstructed secondary vertices associated with the jet to provide a likelihood ratio discriminator for b-jet identification. The b jet identification working point chosen in the $H \rightarrow \tau\tau$ analyses gives an efficiency for identifying genuine b jets of about 70%, and for misidentifying light flavor jets as b jets of about 1% [122].

5.3.3 Taus

Hadronically decaying τ leptons are reconstructed in PF with the hadron-plus-strips (HPS) algorithm [124, 125] which is seeded with the collection of anti-k_{T} jets discussed above. The HPS algorithm reconstructs τ_h candidates based on their compatibility with one of the primary $\tau \rightarrow \tau_h$ decay modes listed in Table 5.2. A large number of decay modes include decays of a π^0 which immediately decay as $\pi^0 \rightarrow \gamma\gamma$. The π^0 are reconstructed in strips of (η, ϕ) elongated in the ϕ direction (due to magnetic bending) to catch photon conversions. Since the start of Run-II, the HPS algorithm has moved from using a fixed width strip, $0.05\eta \times 0.20\phi$, to a dynamic width strip where the width depends on the p_{T} of the electron or photon used to seed the strip. More dynamic strip details are provided in the following paragraph. Based on the number of charged hadrons and π^0s, the τ_h candidate is assigned a decay mode. The invariant mass of the τ_h is calculated from all hadrons and π^0 which are assigned as the τ_h decay products. For the τ decay modes involving a meson resonance, there is a mass window requirement that necessitates that the τ_h invariant mass be consistent with the meson resonance. τ_h with only a single charged hadron are assigned a mass equal to the mass of a $\pi^{\pm}$, 140 MeV. Figure 5.4 shows the reconstructed mass for hadronically decaying taus in 2012 data using a selection with high genuine tau purity in the $\mu\tau_h$ final state.

The dynamic strip reconstruction has been optimized to best reconstruct the π^0

Decay Mode	Meson Resonance	$\mathcal{B}$ (%)
$\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$		17.8
$\tau^- \rightarrow \mu^- \bar{\nu}_\mu \nu_\tau$		17.4
$\tau^- \rightarrow h^- \nu_\tau$		11.5
$\tau^- \rightarrow h^- \pi^0 \nu_\tau$	$\rho(770)$	26.0
$\tau^- \rightarrow h^- \pi^0 \pi^0 \nu_\tau$	a1(1260)	10.8
$\tau^- \rightarrow h^- h^+ h^- \nu_\tau$	a1(1260)	9.8
$\tau^- \rightarrow h^- h^+ h^- \pi^0 \nu_\tau$		4.8
Other modes with hadrons		1.8
Total leptonic modes		35.2
Total hadronic modes		64.8

Table 5.2: Decay modes for τ^- leptons including leptonic decays and hadronic decays. The $h^\pm$ stand for $\pi^\pm$ or $K^\pm$. Inverting all of the “-” for “+” will give the decay modes for τ^+ leptons. Nearly 65% of τ leptons decay hadronically to τ_h.

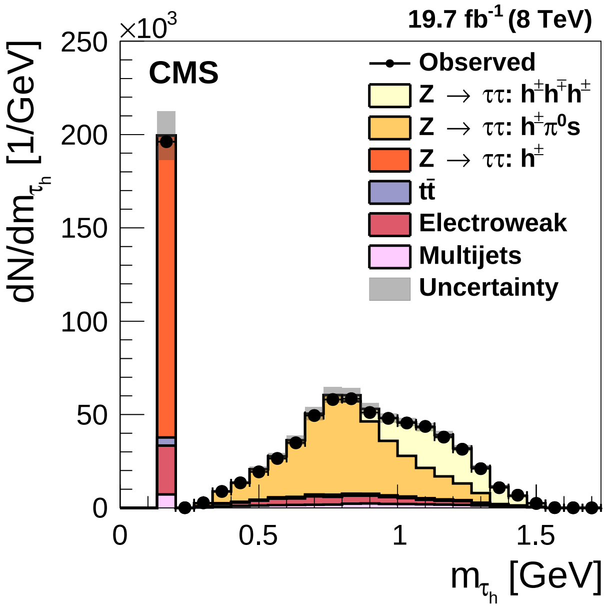


Figure 5.4: The reconstructed invariant mass of the τ_h candidate. A spike is seen at 140 MeV for the 1-prong τ_h decay mode where the mass is assigned equal to the mass of a $\pi^\pm$. The 1-prong+π^0 decay mode is seen to peak around 770 MeV while the 3-prong decay mode centers around 1260 MeV.

decay products. It was found that the fixed width strips used in Run-I occasionally allowed the π^0 decay products to escape the boundaries of the strip and would result in the escaped particle being added to the isolation sums for the τ_h and not contributing to the energy or momentum of the τ_h. Widening the fixed width strips can be used to catch these escaping particles. However, considering that more boosted τ_h will have a more collimated structure, it is also helpful to narrow the strip at higher τ_h p_T to suppress pileup or other non-τ contributions when possible. The strip widths are calibrated to simulations that target retaining 95% of the π^0 decay products, see Figure 5.5. Strips containing one or more electron or photon constituent and passing a cut of $p_\mathrm{T} > 2.5\,\mathrm{GeV}$ for the sum of the electrons and photons are kept as π^0 candidates.

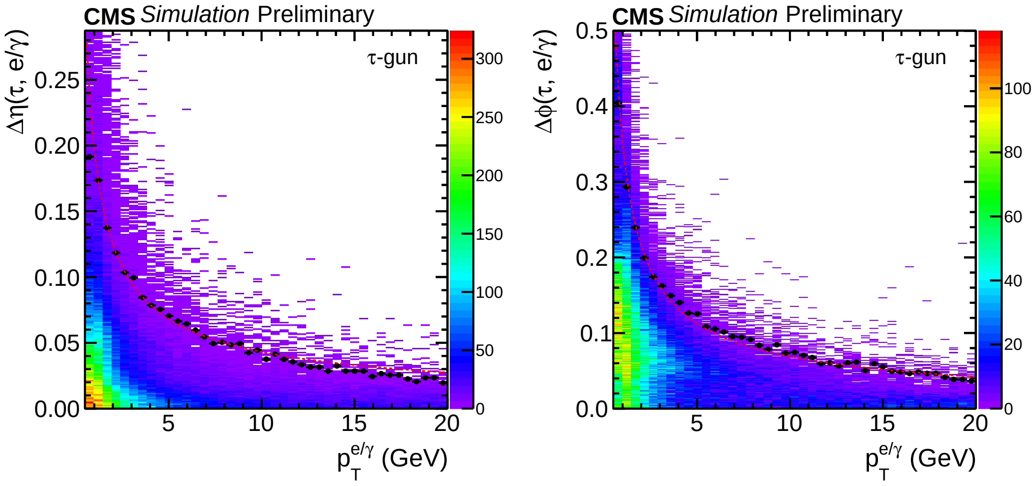


Figure 5.5: Distance in η (left) and ϕ (right) between τ_h and e/γ, that are due to hadronic tau decay products, as a function of e/γ p_T. The size of the window is larger in the ϕ direction due to bending in the magnetic field. The dotted line shows the 95% quantile while the red line shows the fit to the 95% quantile. The red line is used to define the widths of the dynamic strip.

A multivariate (MVA) discriminator [126, 125], including isolation, shape-based variables and lifetime information, is used to reduce the rate for quark and gluon initiated jets to be identified as τ_h candidates. The working point used in the $H \rightarrow$

$\tau\tau$ analysis, `Tight Tau MVA`, has an efficiency of about 60% for genuine τ_h, with about 1.5% misidentification rate for quark- and gluon-initiated jets, for a p_{T} range typical of τ_h originating from a Z boson. A looser working point is used in the Higgs associated production analysis ZH final states, `Medium Tau MVA`, which as an efficiency of 65% for genuine τ_h with a 2% misidentification rate. In the WH final states, the `Very Tight Tau MVA` is used in certain circumstances and has a 55% efficiency for a 1% misidentification rate. Details on the τ_h MVA working points can be found in Reference [125].

Electrons and muons can both be reconstructed as τ_h candidates, usually into the 1-prong or 1-prong+π^0 decay modes. For electrons this can happen easily when an electron emits a bremsstrahlung photon mimicking a π^0 in the decay mode reconstruction. MVA discriminants have been developed which specifically target the suppression of electrons and muons being misidentified as τ_h [124, 125]. A range of anti-e and anti-μ discriminant working points are used in these analyses. The choice of which working point to use is tailored towards suppressing the dominant backgrounds in different final states and is discussed in detail in the analysis Chapters 6 and 7.

Chapter 6

Higgs $\rightarrow \tau\tau$: Gluon Fusion and Vector Boson Fusion

This chapter describes a study of Higgs boson production and subsequent decay to a pair of τ leptons using CMS proton-proton collision data gathered in 2016. This chapter specifically discusses a study targeting the gluon fusion (ggH) and vector boson fusion (VBF) Higgs boson production mechanisms. A later chapter, 7, discusses a study focusing on the associated production Higgs boson mechanism. This gluon fusion and VBF targeted study is the first $H \rightarrow \tau\tau$ analysis performed using center-of-mass energy 13 TeV data from the LHC. Combining these 13 TeV results with 7 TeV and 8 TeV CMS $H \rightarrow \tau\tau$ results we produce the first single experiment observation of the $H \rightarrow \tau\tau$ process, observed at the 5.9 σ confidence level. Additionally, this study provides the strongest constraints on VBF Higgs production to date.

6.1 Overview

This chapter specifically focuses on studying the Higgs boson produced via the gluon fusion or the VBF production mechanisms. A study of the Higgs boson produced in associated production with WH/ZH is presented in Chapter 7. This study utilizes the full 2016 pp dataset collected by CMS corresponding to $35.9\,fb^{-1}$ of integrated luminosity. In the following pages the symbol ℓ refers to electrons and muons and τ_h refers to hadronically decaying τ leptons while τ_e and τ_μ refer to leptonically decay τ to electrons and muons, respectively. We study all possible $\tau\tau$ final state combinations with the exception of two electron and two muon final states because of the low $\tau\tau \to \tau_e\tau_e/\tau_\mu\tau_\mu$ branching fractions and high background from $Z \to ee/\mu\mu$. The $H \to \tau\tau$ final states which are studied are: $\tau_e\tau_h$ denoted here as $e\tau_h$, $\tau_\mu\tau_h$ denoted as $\mu\tau_h$, $\tau_e\tau_\mu$ denoted here as $e\mu$, and lastly, $\tau_h\tau_h$ denoted as $\tau_h\tau_h$. This combination of final states covers about 94% of all possible $\tau\tau$ final states. The different $\tau\tau$ final states will be referred to as different channels in the following pages. We ensure uniqueness between the four studied channels be applying veto criteria to events based on the number of reconstructed loosely identified electrons and muons, see Section 6.1.3. This ensures that no data or simulated event is double counted in two channels.

Selected events are classified into three different categories targeting different characteristics of the gluon fusion and VBF production topologies. The categories are defined according to the number and kinematics of the associated jets in each event along with the reconstructed p_{T}^{Higgs}. A number of different control regions enriched in various background processes are used in the final fit for signal extraction. This allows the fit to simultaneously adjust and constrain the normalization and shape of processes targeted by a control region. The backgrounds which are targeted with dedicated control regions are: W+jets, QCD multijet, and $t\bar{t}$.

6.1.1 Event Selection

There are specific baseline criteria applied to all electrons, muons, τ_h, and jets for every event. Depending on the final state, additional requirements are placed on these objects based on suppressing target backgrounds or to meet trigger requirements and analysis optimization. The baseline criteria ensure that each object is well reconstructed, well identified, and consistent with a objects that are searched for as part of the analysis strategy.

6.1.2 Triggers

Selected events are required to to have passed a high level trigger consistent with their categorized final state channel. For leptons at the HLT, there are multiple isolation and identification working points. Depending on the physics needs of an HLT path, a working point will be selected that leads to the highest efficiency possible while maintaining a rate which fits into the allocated rate budget. Descriptions of the working points can be found in References [64, 125, 116].

The HLT paths used in this analysis, the lepton p_T thresholds, and $|\eta|$ requirements are detailed in table 6.1. For the $\mu\tau_h$ channel, a combination of single muon and muon-τ_h cross triggers are used. The cross trigger allows for a lower p_T threshold on the muon increasing signal acceptance. In contrast, due to the HLT menu available in 2016, the $e\tau_h$ cross triggers do not bring a substantial increase in acceptance and were not used in this analysis. For the $e\tau_h$ channel, events are selected using a trigger requiring a single tightly isolated electron. For the $e\mu$ channel, events are selected with electron-muon cross triggers. For the $\tau_h\tau_h$ channel, events are selected with an

HLT path requiring two loosely isolated τ_h.

Due to changing data gathering conditions and a changing HLT menu throughout 2016 data taking, the 2016 dataset is broken down into different time dependent eras beginning with “B” and sequentially progressing to “H.” During the final eras of 2016, the instantaneous luminosity and the average pileup per event increased compared to the initial eras. Because of this, some HLT paths became disabled or prescaled to keep the overall L1 Trigger and HLT rate within the allocated budget. This study avoids using any prescaled HLT paths.

As the instantaneous luminosity increased the HLT paths for the $\mu\tau_h$, $e\mu$, and $\tau_h\tau_h$ channel changed to maintain the budgeted rate. The single muon triggers introduced a tighter $|\eta|$ restrictions. Towards the end of 2016, the $e\mu$ triggers introduced a Δz requirement matching the leptons to the primary vertex (Section 5.2.1) to reduce rate. The isolation criteria was changed for the $\tau_h\tau_h$ channel triggers to maintain a reasonable rate. The τ_h isolation changed from being based on purely charged energy deposits, which are defined by energy deposits linked to a track, to being based on a combination of charged and neutral energy deposits. Neutral energy deposits are not linked to a track.

In all channels, the electrons, muons, and τ_h in each event must be matched to within $\Delta R < 0.5$ with the associated HLT object which triggered the event.

6.1.3 Baseline Object Selection

All electrons and muons are selected in accordance with Sections 5.1.3 and 5.1.3. In addition, they must meet the minimum requirement that the distance of closest approach to the primary vertex satisfies $|d_z| < 0.2$ cm along the beam direction, and $|d_{xy}| < 0.045$ cm in the transverse plane. Ensuring compatibility with the primary vertex is consistent with the predicted infinitesimal life-time of a Higgs boson. The

Channel	Trigger ($p_{\mathrm{T}}/\|\eta\|$) Req.	High Level Trigger Description	Eras
$\mu\tau_h$	$\mu(22/2.4)$	single iso muon	B-F
	$\mu(22/2.1)$	single iso muon	C-H
	$\mu(19/2.1)\,\&\,\tau_h(20/2.1)$	iso muon & loose iso τ_h	All Eras
$e\tau_h$	$e(25/2.1)$	single tight iso electron	All Eras
$\tau_h\tau_h$	$\tau_h(35/2.1)\,\&\,\tau_h(35/2.1)$	two iso τ_h (charged isolation)	B-G
	$\tau_h(35/2.1)\,\&\,\tau_h(35/2.1)$	two iso τ_h (combined isolation)	H
$e\mu$	$e(12/2.5)\,\&\,\mu(23/2.4)$	loose iso muon & loose iso electron	B-F
	$e(12/2.5)\,\&\,\mu(23/2.4)$	loose iso muon & loose iso electron & DZ req.	G-H
	$e(23/2.5)\,\&\,\mu(8/2.4)$	loose iso muon & loose iso electron	B-F
	$e(23/2.5)\,\&\,\mu(8/2.4)$	loose iso muon & loose iso electron & DZ req.	G-H

Table 6.1: For each channel the HLT p_{T} threshold and $|\eta|$ range is listed along with a description of the HLT path where "iso" stands for "isolated." Changes in the available triggers with respect to the data collection era are noted.

HPS reconstruction of τ_h detailed in section 5.3.3 can involve combining together multiple tracks and π^0s coming from intermediary τ decay products. Due to these intermediary products, the reconstructed d_{xy} for τ_h are often larger than those for electrons and muons. Because of this, the primary vertex matching criteria are relaxed for τ_h and only require $|d_z| < 0.2$ cm.

The offline selection criteria for all electrons, muons, and τ_h are motivated and constrained by the High Level Trigger requirements of their path. Specifically, the offline p_{T} criteria applied are always higher than the HLT p_{T} threshold to ensure a stable measurement and application of trigger efficiencies. An offline p_{T} threshold applied right at the HLT p_{T} threshold makes measurement of the steeply rising efficiency at the HLT threshold absolutely critical. This is very difficult to do perfectly and would lead to very large trigger systematics at low p_{T} in the turn-on region for a fractional gain in acceptance. Additionally, selected objects are required to meet the same $|\eta|$ restrictions as are imposed by the HLT paths.

All selected electrons, muons, and τ_h must be well identified and isolated from overlapping energy deposits and reconstructed objects. This study uses identification criteria provided centrally by the CMS Physics Object Groups (POGs). All identification and isolation working points used have been selected through an optimization process selecting for increased analysis sensitivity. The optimum working points strike a balance between signal efficiency and background rejection. For electrons an MVA-based ID is used in both the $e\tau_h$ and $e\mu$ channels which has been tuned to provide 80% electron selection efficiency, Section 5.1.3. Muons in both the $\mu\tau_h$ and $e\mu$ channels require the Particle Flow Medium ID, Section 5.1.3. The selection of τ_h relies on MVA-based working points. The τ_h MVA-based working points combine both object identification and object isolation together into a single set of working points, Section 5.3.3. Table 6.2 details the p_T, $|\eta|$, identification and isolation criteria for all electrons, muons, and τ_h selected in the study.

To prevent data and simulated events from being used in multiple final states, we remove events with a loosely identified electrons and muons. The loosened identification is consistent in η with Table 6.2 while the p_T threshold is lowered to $p_\mathrm{T} > 10\,\mathrm{GeV}$. For electrons we use: MVA 90% WP, $I^e < 0.3$, while for muons we use: PF ID Medium, $I^\mu < 0.3$.

Channel	p_T (GeV)	η	Identification	Isolation
$\mu\tau_h$	$p_\mathrm{T}^\mu > 20$	$\lvert\eta^\mu\rvert < 2.1$	PF ID Medium	$I^\mu < 0.15$
	$p_\mathrm{T}^{\tau_h} > 30$	$\lvert\eta^{\tau_h}\rvert < 2.3$	MVA τ_h ID Tight	MVA τ_h ID Tight
$\tau_h\tau_h$	Leading $p_\mathrm{T}^{\tau_h} > 50$	$\lvert\eta^{\tau_h}\rvert < 2.1$	MVA τ_h ID Tight	MVA τ_h ID Tight
	Subleading $p_\mathrm{T}^{\tau_h} > 40$	$\lvert\eta^{\tau_h}\rvert < 2.1$	MVA τ_h ID Tight	MVA τ_h ID Tight
$e\tau_h$	$p_\mathrm{T}^e > 26$	$\lvert\eta^e\rvert < 2.1$	MVA 80% WP	$I^e < 0.1$
	$p_\mathrm{T}^{\tau_h} > 30$	$\lvert\eta^{\tau_h}\rvert < 2.3$	MVA τ_h ID Tight	MVA τ_h ID Tight
$e\mu$	$p_\mathrm{T}^e > 13$	$\lvert\eta^e\rvert < 2.5$	MVA 80% WP	$I^e < 0.15$
	$p_\mathrm{T}^\mu > 15$	$\lvert\eta^\mu\rvert < 2.4$	PF ID Medium	$I^\mu < 0.2$

Table 6.2: Kinematic, identification and isolation selection requirements for the four di-τ channels.

Signal extraction categories rely on the details of reconstructed jets, or lack there of, in each event. Jets are reconstructed using the anti-k_T algorithm with distance parameter $\Delta\mathrm{R} = 0.4$ [118]. Charged hadrons that are not consistent with the primary vertex are removed from the anti-k_{T} clustering. Jets are only considered if they pass they pass the loose working point of the PF Jet ID discriminator [127, 128]. Jets must have $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 4.7$. The p_{T} and η requirements are altered for cases where the jet is identified as having likely originated from a b-quark. Jets likely originating from a b-quark are considered if they pass the combined secondary vertex (CSV) Medium working point are then labeled as b-tagged jets, Section 5.3.2. B-tagged jets have a relaxed p_{T} requirement but much tighter η requirement: $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.4$. The tightened η requirement ensures that b-tagged jets are located within the detector volume fully covered by the CMS pixel and strip tracker. Lastly, all jets must be separated from the selected electrons, muons, and τ_h by $\Delta\mathrm{R} > 0.5$.

Depending on the di-τ_h channel, there are specific topological cuts targeted at significantly reducing the contribution of certain background processes in the signal region. The large $W+$jets cross section combined with a non-negligible jet $\rightarrow \tau_h$ fake rate leads to a large $W+$jets contribution in the $\ell\tau_h$ channels. This contribution is significantly reduced at the cost of minimal signal events by cutting on the transverse mass, m_{T}. Where the m_{T} selection is defined as,

$$m_{\mathrm{T}} \equiv \sqrt{2p_{\mathrm{T}}^{\ell}p_{\mathrm{T}}^{\mathrm{miss}}[1-\cos(\Delta\phi)]} < 50\,\mathrm{GeV}, \tag{6.1}$$

where p_{T}^{ℓ} is the transverse momentum of the lepton ℓ, and $\Delta\phi$ is the azimuthal angle between its direction and the $\vec{E}_T^{miss}$.

In the $e\mu$ channel, the large $t\bar{t}$ background is reduced by requiring $p_\zeta - 0.85\,p_\zeta^{\mathrm{vis}} > -35\,\mathrm{GeV}$ in the "0-jet" and "Boosted" categories, or $p_\zeta - 0.85\,p_\zeta^{\mathrm{vis}} > -10\,\mathrm{GeV}$ in the "VBF" category where the categories are defined in the following section. p_ζ is the component of the $\vec{E}_T^{miss}$ projected along the bisector of the transverse momenta of the two leptons and p_ζ^{vis} is the sum of the components of the lepton transverse momenta along the same direction [129], also see Figure 6.1 for visual reference. The p_ζ selection criteria has a high signal efficiency. The $\vec{E}_T^{miss}$ is typically oriented in the same direction as the visible di-τ system in signal events because the $\vec{E}_T^{miss}$ is the results of neutrinos from the signal τ leptons. The orientation of the $\vec{E}_T^{miss}$ with respect to the di-τ system is much less predictable in $t\bar{t}$ events. In addition, events with a b-tagged jet are discarded to further suppress the $t\bar{t}$ background in the $e\mu$ channel.

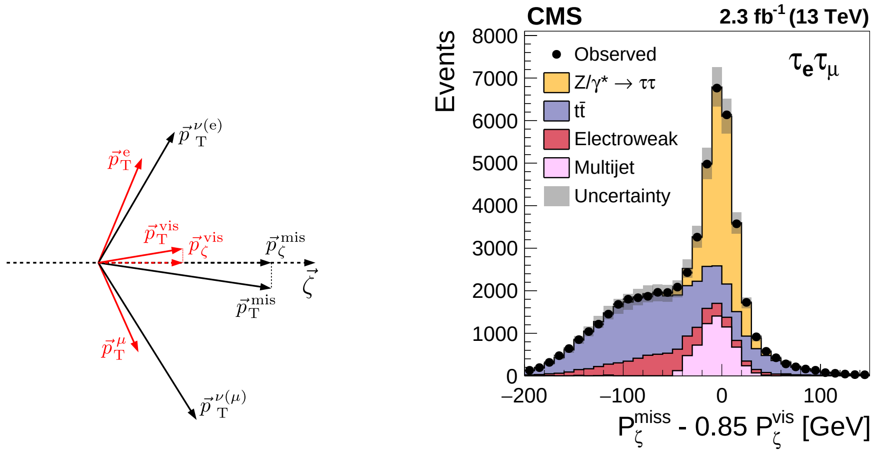


Figure 6.1: (Left) Diagram showing the construction of the p_ζ and p_ζ^{vis} projections. (Right) An example $p_\zeta - 0.85\,p_\zeta^{\mathrm{vis}}$ distribution is shown for a similar but not overlapping selection in the $e\mu$ channel. This distribution is from an analysis focusing on specifically studying the $Z \to \tau\tau$ process [2]. In the $H \to \tau\tau$ analysis, the Higgs boson $p_\zeta - 0.85\,p_\zeta^{\mathrm{vis}}$ spectrum aligns very closely with the $Z \to \tau\tau$ distribution shown here.

6.1.4 Categorization

Selected events are split into three mutually exclusive categories per decay channel. The categories are designed to target different aspects of the gluon fusion and VBF Higgs boson production mechanisms. In each category the two variables that maximize the $H \to \tau\tau$ sensitivity are chosen to build two-dimensional (2D) distributions.

The three categories are defined as:

- 0-jet: This category targets Higgs boson events produced via gluon fusion (Section 2.4). The two variables chosen to extract the results are m_{vis} and the reconstructed τ_h decay mode in the $\mu\tau_h$ and $e\tau_h$ decay channels. In the $e\mu$ channel, m_{vis} and the p_{T} of the muon are used. The $Z \to \ell\ell$ background is large in the 1-prong and 1-prong + π^0(s) τ_h decay modes in the $\mu\tau_h$ and $e\tau_h$ channels. By using only m_{vis} instead of $m_{\tau\tau}$ the $\vec{E}_T^{miss}$ does not smear the mass reconstruction for $Z \to \ell\ell$ which retains a nice sharp peak. The sharp Z peak of $Z \to \ell\ell$ provides an excellent handle to distinguish $Z \to \ell\ell$ from the other background process and helps constrain the associated uncertainties for electrons and muons faking τ_h. As electrons and muons almost exclusively fake 1-prong and 1-prong + π^0(s) τ_h, the reconstructed τ_h decay mode is used as the second of the 2D variables in the $\mu\tau_h$ and $e\tau_h$ channels. Additionally, the lack of $Z \to \ell\ell$ in the 3-prong decay mode results in increased signal significance for that region. Examples of the 2D distributions for the signal and $Z \to \ell\ell$ background in the 0-jet category of the $\mu\tau_h$ decay channel are shown in Fig. 6.1.4 (top). In the $\tau_h\tau_h$ decay channel, only one observable, $m_{\tau\tau}$, is considered because of the low event yields due to the relatively high p_{T} thresholds on the τ_h at trigger level, and because of the sharply falling τ_h p_{T} distribution. Simulations indicate that about 98% of signal events in the 0-jet category correspond to Higgs bosons produced via the gluon fusion production mechanism.

- VBF: This category targets Higgs boson events produced via the VBF process (Section 2.4). The presence of jets from the hard scattering process in VBF production leads the study to heavily utilize jet kinematics and the jet topology in the VBF category. Events are selected with at least two (exactly two) jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ in the $\tau_h\tau_h$, $\mu\tau_h$, and $e\tau_h$ ($e\mu$) channels. In the $\mu\tau_h$, $e\tau_h$, and $e\mu$ channels, the two leading jets are required to have an invariant mass,

m_{jj}, larger than 300 GeV. The variable $p_{\mathrm{T}}^{\tau\tau}$, defined as the magnitude of the vectorial sum of the $\vec{p_{\mathrm{T}}}$ of the visible decay products of the τ leptons and $\vec{E}_T^{miss}$, is required to be greater than 50 (100) GeV in the $\mu\tau_h$ and $e\tau_h$ ($\tau_h\tau_h$) channels to reduce the contribution from $W+$jets backgrounds. This selection criterion also suppresses the background from quantum chromodynamics (QCD) multijet events. In addition, the p_{T} threshold on the τ_h candidate is raised to 40 GeV in the $\mu\tau_h$ channel. The two leading jets in the $\tau_h\tau_h$ channel should be separated in pseudorapidity by $\Delta\eta > 2.5$. The $\Delta\eta > 2.5$ cut in the $\tau_h\tau_h$ channel significantly reduces the contributions from QCD multijet events at the cost of very few signal events because of the large jet $\Delta\eta$ in genuine VBF events. The two observables used in the VBF category are $m_{\tau\tau}$ and m_{jj} for all channels. Example 2D distributions for the signal and $Z \to \tau\tau$ background in the VBF category of the $\mu\tau_h$ decay channel are shown in Fig. 6.1.4 (center). Integrating over the whole m_{jj} phase space, up to 57% of the signal events in the VBF category are produced via the VBF production mode. This proportion increases with m_{jj} which allows disambiguation of the gluon fusion and VBF signals in the highest m_{jj} ranges.

- Boosted: This category contains all selected events that do not enter one of the previous categories, namely events with one jet and events with several jets that fail the specific requirements of the VBF category. The Boosted category contains a mix of gluon fusion events produced in association with one or more jets (78–80% of signal events), VBF events where one of the jets escaped detection or has low m_{jj} (11–13%), as well as Higgs bosons produced in association with a W or a Z boson decaying hadronically (4–8%). Because these gluon fusion events failed the 0-jet category, the Higgs boson will be recoiling off of one or more jets making the $p_{\mathrm{T}}^{\tau\tau}$ a natural choice for the second distribution

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

Table 6.3: Category selection and observables used to build the 2D kinematic distributions. The events failing the 0-jet and VBF selection are included in the Boosted category and are denoted by “Others”.

variable with $m_{\tau\tau}$ as the other of the 2D variables. Most background processes, including $W+$jets and QCD multijet events, typically have low $p_{\mathrm{T}}^{\tau\tau}$. Example 2D distributions for the signal and $W+$jets background in the Boosted category of the $\mu\tau_h$ decay channel are shown in Fig. 6.1.4 (bottom).

In Figure 6.1.4, the background processes are chosen for illustrative purpose for their separation from the signal. The $Z \to \mu\mu$ background in the 0-jet category is concentrated in the regions where the visible mass is close to 90 GeV and is negligible when the τ_h candidate is reconstructed in the 3-prong decay mode. The $Z \to \tau\tau$ background in the VBF category mostly lies at low m_{jj} values whereas the distribution of VBF signal events extends to high m_{jj} values. In the Boosted category, the W+jets background, which behaves similarly to the QCD multijet background, is rather flat with respect to $m_{\tau\tau}$, and is concentrated at low $p_{\mathrm{T}}^{\tau\tau}$ values.

The three categories and the variables used to build the 2D distributions are summarized in Table 6.3.

The results of the analysis are extracted with a global maximum likelihood fit based on the 2D distributions in the various signal regions, and on the control regions,

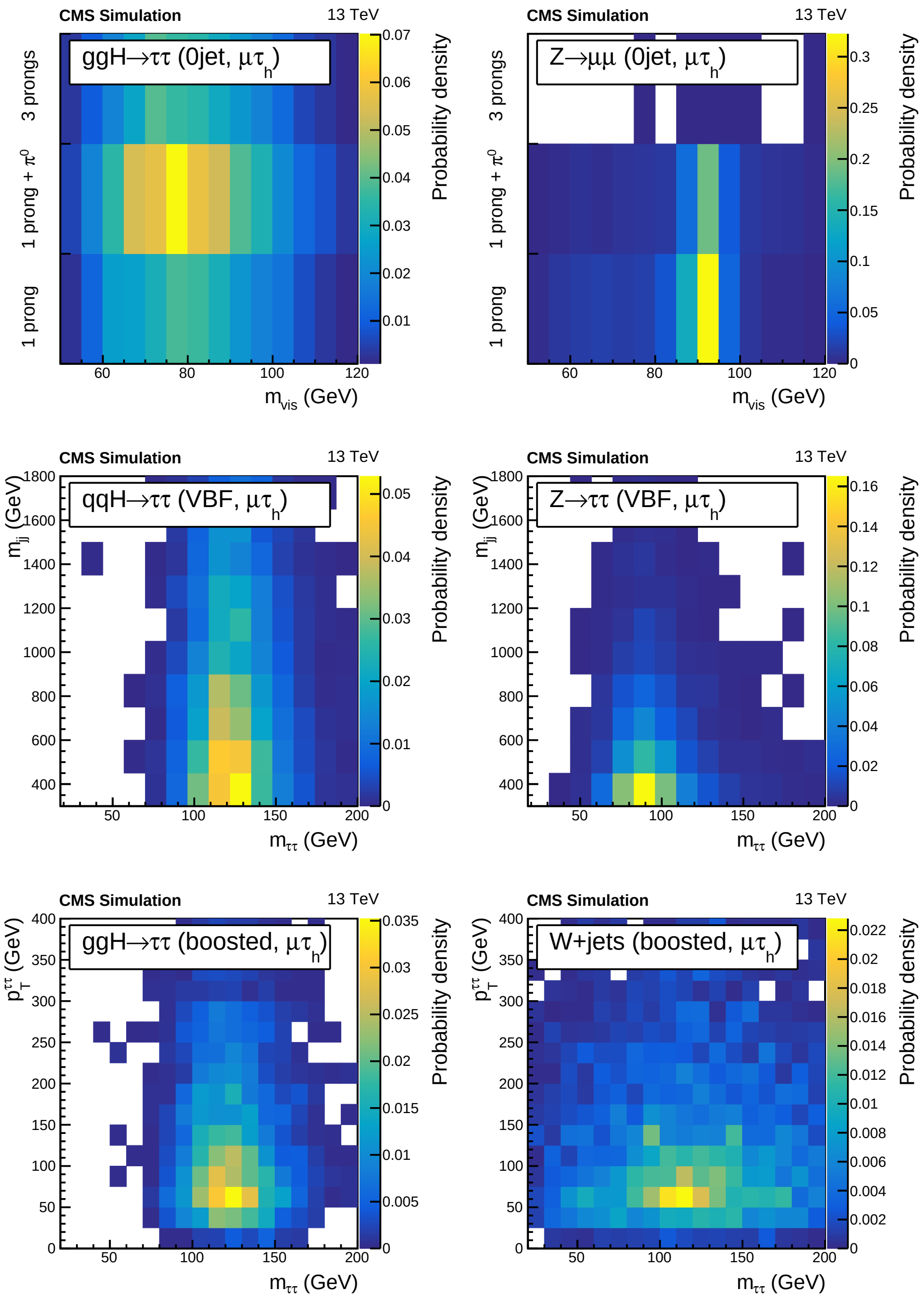


Figure 6.2: Two dimensional distributions for Higgs boson events passing event selection for each category are shown (left). The specific Higgs boson process shown is the one specifically targeted for each category. Distributions for select dominant background processes are shown (right). The rows correspond to the three categories: 0-jet (top), VBF (center), and Boosted (bottom). All distributions are from the $\mu\tau_h$ decay channel.

detailed in Section 6.5, that constrain the normalizations of the main backgrounds.

6.2 Data Set

The ggH and VBF targeted $H \to \tau\tau$ study utilizes the full 2016 pp dataset collected by CMS corresponding to $35.9\, fb^{-1}$ of integrated luminosity. The data were gathered at center-of-mass energy 13 TeV. As data are gathered at CMS, the conditions of the CMS experiment and its subdetectors are recorded. The CMS experiment uses an offline validation process to ensure that only high quality data is used in future analyses. Data collected while the CMS detector is in a faulty state is flagged as such and skipped. The CMS experiment collected $35.9\, fb^{-1}$ of integrated luminosity worth of data usable for physics analysis in 2016.

In this analysis only a subset of the total data marked as good is used. During data gathering, data is filtered into primary datasets (PDs) corresponding to which HLT trigger made the determination to save a given event. The HLT triggers used in this analysis, detailed in table 6.1, correspond to: the Single Electron/Photon PD, the Single Muon PD, the Muon and Electron PD, and the Tau PD.

6.3 Monte Carlo Samples

Signal and background processes are modeled with samples of simulated events. For details on the production for simulated events, see Section 4. The simulated background processes are all scaled to their NLO cross section in Table 6.4. The cross sections of the Higgs boson production processes are listed in Section 2.4.

For each simulated event, a number of additional pileup interactions is simulated and added. The number of pileup interactions added is based on best efforts to match the simulated events to the pileup in data which is estimated from the measured instantaneous luminosity for each bunch crossing. The average number of pileup interactions in the 2016 CMS data is approximately 27 interactions per bunch crossing, Figure 6.3.

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

Table 6.4: NLO cross sections for considered backgrounds. In this table, ℓ represents all three generations of charged leptons, e,μ,τ. In some cases the production mechanism is listed: quarks (qq) versus gluons (gg).

6.4 Mass Reconstruction

The visible mass of the $\tau\tau$ system, m_{vis}, can be used to separate the $H \rightarrow \tau\tau$ signal events from the large contribution of irreducible $Z \rightarrow \tau\tau$ events. However, the neutrinos from the τ lepton decays carry a large fraction of the τ lepton energy. Lep-

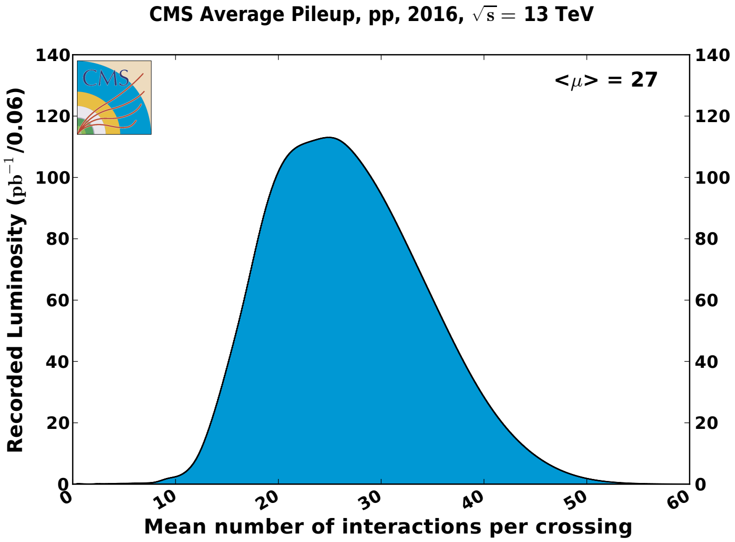


Figure 6.3: Mean number of interactions per bunch crossing for the 2016 proton-proton run at 13 TeV. The cross section is taken to be 80 mb.

tonically decaying τ produce two neutrinos through their decay while hadronically decaying τ produce a single neutrino, see Table 5.2. The presence of the neutrinos reduces the discriminating power of m_{vis}. The SVFIT algorithm [130] combines the $\vec{E}_T^{miss}$ with the four-vectors of both τ candidates to calculate a more accurate estimate of the mass of the parent boson, denoted as $m_{\tau\tau}$. The SVFIT algorithm is based on a likelihood method which, given the input parameters, in discrete $m_{\tau\tau}$ slices, estimates the compatibility of $m_{\tau\tau}$ with the inputs. The $m_{\tau\tau}$ value for an event is taken from the $m_{\tau\tau}$ slice with the highest compatibility. A more detailed description of the algorithm can be found in Reference [130].

The resolution of $m_{\tau\tau}$ is between 15 and 20% depending on the $\tau\tau$ final state. Both mass variables, m_{vis} and $m_{\tau\tau}$ are used in the analysis, as detailed in Section 6.1.4. The m_{vis} variable is preferred over $m_{\tau\tau}$ when the background from $Z \to \ell\ell$ events is large.

6.5 Background Estimation

A number of different background estimation techniques are employed in this analysis. In all cases all Monte Carlo based backgrounds used in the techniques have corrections applied to them. These corrections are discussed in the following section, 6.6, and are derived to increase the agreement between the simulated backgrounds and data.

Drell–Yan Background

The largest irreducible source of background is the Drell–Yan production of $Z/\gamma^* \rightarrow \tau\tau, \ell\ell$. Proper shape and normalization for this leading background is critical to the success of the analysis. In certain Drell–Yan enriched regions of phase space discrepancies between data and simulation can reach as much as 20% for distributions of $m_{\ell\ell}$, $p_{\mathrm{T}}^{\ell\ell}$, and m_{jj} in events with two or more jets. A dedicated, high purity, $Z/\gamma^* \rightarrow \mu\mu$ control region was used to measure reweighting factors for application to all Drell–Yan simulated events and is detailed in the Monte Carlo Corrections section 6.6.6.

$W +$ jets Background

The $W +$ jets process is modelled using simulation. In the $\tau_h\tau_h$ and $e\mu$ channels the $W +$ jets contribution is small compared to other backgrounds. In these two channels both the shape and yield are taken from simulation based on the samples and cross sections in Table 6.4. The background from $W +$jets production contributes significantly to the $\mu\tau_h$ and $e\tau_h$ channels, when the W boson decays leptonically and a jet is misidentified as a τ_h candidate. In the $\ell\tau_h$ channels the shape of the background is from the simulated samples in Table 6.4. The yield is estimated using data in a $W +$jets enriched dedicated side-band region defined based on a transverse mass cut,

$m_{\mathrm{T}} > 80\,\mathrm{GeV}$. In this high-$m_{\mathrm{T}}$ side-band region, the $W+$jets process is scaled so that the sum total of expected background events is equivalent to the sum total of observed data events. The scaling factor is then applied in the low transverse mass signal regions, $m_{\mathrm{T}} < 50\,\mathrm{GeV}$. A scaling factor is measured for the 0-jet and Boosted categories for both $e\tau$ and $\mu\tau$ resulting in four $W+$jets scaling values. The scaling factor measured in the Boosted category is extrapolated to the VBF category. The $W+$jets purity in these side-band regions varies from about 50% in the Boosted category to 85% in the 0-jet category.

The high-m_{T} side-bands, described above, are included as control regions in the final fit. These high-m_{T} W+jets control regions can be seed in Figure 6.4. The control regions have one bin because they are used solely to constrain the normalization of the $W+$jets process.

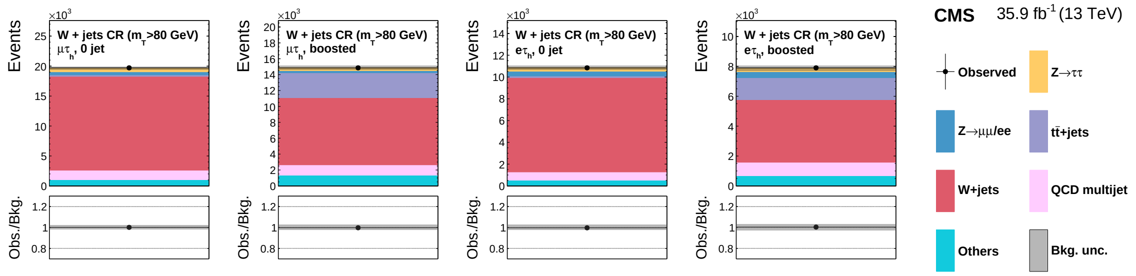


Figure 6.4: the high-m_{T} control regions enriched in the $W+$jets background used in the maximum likelihood fit, together with the signal regions, to extract the results. The normalization of the predicted background distributions corresponds to the result of the global fit. These regions, defined with $m_{\mathrm{T}} > 80\,\mathrm{GeV}$, control the yields of the $W+$jets background in the $\mu\tau_h$ and $e\tau_h$ channels.

QCD Multijet Background

The QCD multijet process constitutes an important source of reducible background in the $\tau_h\tau_h$ and $\ell\tau_h$ channels. The QCD multijet process is subdominant in the $e\mu$ channel. For all four channels, the QCD multijet, is estimated from data. Side-band control regions are constructed to estimate the shape and the yield of the QCD multijet background in these channels.

QCD multijet estimation for the $\ell\tau_h$ channels follows:

1. Raw yield is extracted using a sample where the ℓ and the τ_h candidates have the same-sign charge. Using this sample, the QCD multijet process is estimated from data by subtracting the contribution of the Drell–Yan, $t\bar{t}$, diboson, and $W+$jets processes.

2. The 2D distributions of the QCD multijet background are estimated from a region with same-sign leptons, same as the yield estimate, except the isolation of the ℓ and τ_h candidates is additionally relaxed to reduce the statistical fluctuations in the distributions. The contributions from other background process are subtracted from data to extract the QCD multijet shape template in this region.

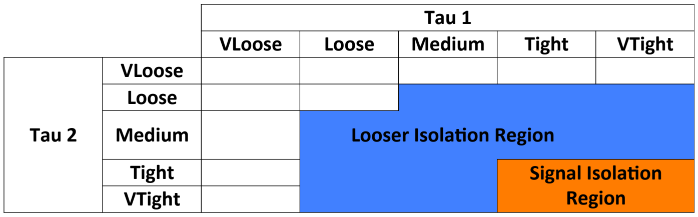
3. The raw yield obtained above is corrected to account for observed differences between the background composition in the same-sign and opposite-sign regions. An extrapolation factor between the same-sign and opposite-sign regions is determined by comparing the yield of the estimated QCD multijet background for events with ℓ candidates passing inverted isolation criteria, in the same-sign and opposite-sign regions. This extrapolation factor is allowed to adjust in the final fit via the inclusion of dedicated QCD multijet control regions for these channels.

The same technique is used in the $e\mu$ channel, except no control region is included in the fit because QCD multijet events contribute little to the total background in this decay channel.

In the $\tau_h\tau_h$ channel, the large QCD multijet background is estimated differently than the other channels. Instead of deriving the QCD multijet shape template from the same-sign region, the template is instead derived from the background-subtracted opposite-sign region. The reason for this is that the shape of the QCD multijet distributions have statistically significant differences in their distribution shape between the same-sign and opposite-sign regions. The difference in these shape templates is evaluated using Kolmogorov–Smirnov (KS) tests [131]. Specifically, this study relied on a two sample, binned KS test which tested the compatibility of the QCD multijet estimated background in the same-sign region with that estimated in the opposite-sign region. Beyond the KS tests showing the incompatibility of same-sign and opposite-sign shape templates, different isolation requirements were used to test the compatibility of QCD multijet shapes across different isolation ranges for i) the same-sign events, and ii) for the opposite-sign events. These tests showed a much higher degree of shape compatibility for events with different isolation requirements, but the same charge configuration. Because of this, a relaxed isolation criteria for both τ_h leptons, shown in Figure 6.5, is used to derive the QCD multijet template for the $\tau_h\tau_h$ channel.

The selection used to estimate the QCD multijet background shape template and raw yield requires that at least one of the τ_h must fail the `Tight Tau MVA` requirement (Section 5.3.3) which is the isolation selection used to define the signal region. This selection ensures that the side-band sample is disjoin from the signal region. In this region, the QCD multijet background shape and raw yield are obtained by subtracting the contribution of the Drell–Yan, $t\bar{t}$, and $W+\text{jets}$ processes from the data.

A scaling factor is required to adjust the raw yield estimated in the relaxed isolation region to the expected QCD multijet yield in the signal extraction region. This extrapolation factor is estimated in the same-sign charge region. Two same-sign QCD



		Tau 1				
		VLoose	Loose	Medium	Tight	VTight
Tau 2	VLoose					
	Loose			Looser Isolation Region		
	Medium		Looser Isolation Region			
	Tight		Looser Isolation Region		Signal Isolation Region	
	VTight		Looser Isolation Region		Signal Isolation Region	

Figure 6.5: Schematic depicting the range of τ_h MVA working points, defined in Section 5.3.3, ranging from very loose (`VLoose`) to very tight (`VTight`). The signal region is depicted in orange where both τ_h meet the `Tight` or `VTight` criteria. The QCD multijet relaxed isolation estimation region is depicted in blue and does not overlap with the signal region.

multijet samples are constructed using 1) the same isolation requirements as the signal region with all selections are identical except for the τ_h charge configuration. And, 2) a second region with the same relaxed isolation requirements as the region used to estimate the opposite-sign QCD multijet shape template and raw yield. From these two samples a “Relaxed-Region-to-Signal-Region” scaling factor is calculated as,

$$\text{Relaxed-Region-to-Signal-Region SF} = \frac{\text{SS Signal Region Yield}}{\text{SS Relaxed Region Yield}} \tag{6.2}$$

where (SS) is same-sign. Multiplying the raw yield estimated above by this scaling factor results in the estimated QCD multijet contribution in the signal region. The “Relaxed-Region-to-Signal-Region” scale factors for the three $\tau_h\tau_h$ categories are listed in table 6.5. The uncertainties associated with these scale factors are propagated through the QCD multijet estimation process and combined with results from closure tests to estimate a final category dependent systematic and statistic uncertainty for the QCD multijet estimation process.

The events selected with opposite-sign τ_h candidates passing relaxed isolation requirements form control regions, shown in Fig. 6.6, and are used in the global fit to extract the results.

Category	"Relaxed-Region-to-Signal-Region" Scale Factor
0-jet	0.25 ± 1.0%
VBF	0.18 ± 9.8%
Boosted	0.23 ± 1.4%

Table 6.5: The category dependent scale factors used to adjust the QCD multijet yield to correspond to the expected yield in the signal region. The large uncertainty on the VBF scale factor is due to the limited amount of QCD multijet events in the VBF same-sign region.

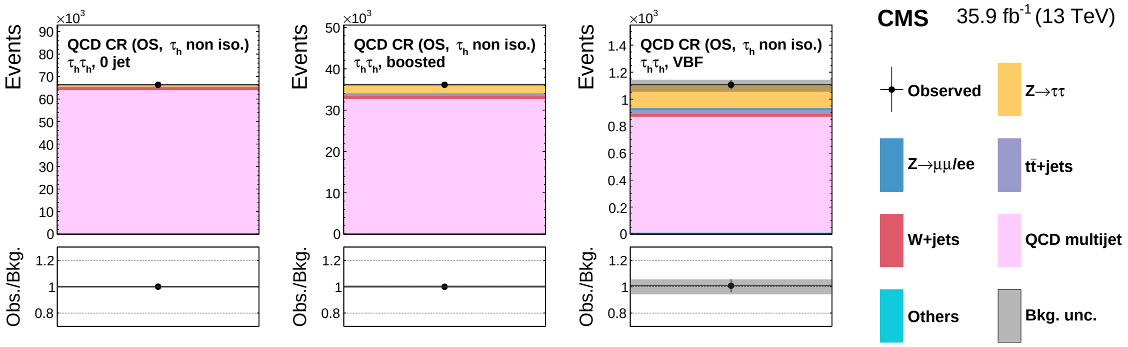


Figure 6.6: Control regions enriched in the QCD multijet background used in the maximum likelihood fit, together with the signal regions, to extract the results. The normalization of the predicted background distributions corresponds to the result of the global fit. These regions, a formed by selecting events with opposite-sign τ_h candidates passing relaxed isolation requirements with at least one of them failing `Tight Tau MVA` isolation. These regions control the yields of the QCD multijet background in the $\tau_h\tau_h$ channel.

$t\bar{t}$ Background

The $t\bar{t}$ production process is one of the main backgrounds in the $e\mu$ channel. In all channels $t\bar{t}$ is predicted from simulation, see Chapter 4. The normalization is adjusted using a $t\bar{t}$-enriched control region orthogonal to the signal region. This control region is included in the global fit. The control region is defined using events in the $e\mu$ channel and is made orthogonal from the signal region by inverting the p_ζ requirement to: $p_\zeta - 0.85\, p_\zeta^{\text{vis}} < -35\,\text{GeV}$. The region is enriched in $t\bar{t}$ by requiring that events should contain at least one jet. The yield of $t\bar{t}$ in all channels and categories is adjusted by this singular $t\bar{t}$ control region. It is shown in Figure 6.7.

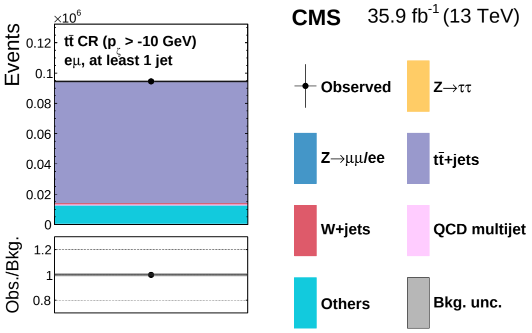


Figure 6.7: Control region enriched in $t\bar{t}$ background, used in the maximum likelihood fit, together with the signal regions, to extract the results. The normalization of the predicted background distributions corresponds to the result of the global fit. This region, defined by inverting the p_ζ requirement and rejecting events with no jet in the $e\mu$ final state, is used to estimate the yields of the $t\bar{t}$ background in all channels.

Other background processes, such as, contributions from diboson and single top quark production, are estimated from simulation using the samples and cross sections listed in Table 6.4.

6.6 Monte Carlo Corrections

Corrections are applied to the simulated Monte Carlo samples to help correct for measured differences between observed data and expectations based on simulation. Many of these corrections are designed to account for differences in reconstruction and identification efficiencies for objects between data and simulation. These corrections are derived in fully orthogonal regions from the analysis signal regions whenever possible. There are however a number of instances where the $\mu\tau_h$ channel is used to derive corrections. In these cases the measurement region used to derive corrections will overlap with a subset of the analysis signal regions. This overlap is discussed in the Systematic Uncertainties section 6.7.

6.6.1 Pileup Reweighting

During 2016 data taking, the LHC delivered *pp* collisions to CMS with an average of approximately 27 interactions per bunch crossing. Monte Carlo samples are generated with additional pileup interactions added to the hard-scattering process for each event. A reweighting technique is used which improves the alignment between the pileup interactions in data and those in simulation. This is critical because there are some reconstructed event characteristics such as $\vec{E}_T^{miss}$ which have different performance in resolution and response depending on the number of pileup interactions in an event.

6.6.2 Tau Identification Efficiency

The reconstruction efficiency for hadronically decaying taus is observed to be different in data and in simulated samples. Correction factors are derived by the Tau POG in the $\mu\tau_h$ channel using a tag-and-probe method to the method defined in Reference [125]. Essentially, the Tau ID Efficiency measurement selects $Z/\gamma^* \to \tau\tau \to \mu\tau_h$ events on the Z mass peak and performs a fit with the $Z \to \mu\tau_h$ process treated as a signal. The degree to which the $Z \to \mu\tau_h$ process is scaled up to down is the resulting Tau ID Efficiency scale factor. The measured scale factor is 0.95 ± 5% for all τ_h passing `Tight Tau MVA`. This correction factor is applied to all simulated τ_h which are matched at the generator level to τ leptons.

6.6.3 Tau Energy Correction

The energy of each simulated τ_h, which matches to a τ at the generator level, is corrected to better represent the observed energy of τ_h in data. The correction is measured and applied as a function of τ_h decay mode with each correction listed in table 6.6.3. The measurement of this correction is performed in the $\mu\tau_h$ final state by the Tau POG [125]. The effect of the shifted τ_h energy is fully propagated through to the $\vec{E}_T^{miss}$ for each event.

τ_h Decay Mode	Energy Correction
1-prong	-1.8% ± 1.2%
1-prong+π^0	+1.0% ± 1.2%
3-prong	+0.4% ± 1.2%

Table 6.6: Energy corrections applied to simulated genuine τ_h. The energy corrections are measured and applied depending on the reconstructed decay mode of the τ_h.

6.6.4 Lepton Identification and Isolation Efficiencies

Similar to the Tau ID Efficiency measured above, the electron and muon reconstruction, identification, and isolation efficiencies are measured in both data and simulation. Correction factors equal to $\epsilon_{data}/\epsilon_{MC}$ are derived as a function of both lepton p_{T} and lepton η. The electron (muon) efficiencies are measured in $Z \rightarrow ee$ ($Z \rightarrow \mu\mu$) events using a tag-and-probe method.

6.6.5 Trigger Efficiencies

Trigger efficiencies are measured in data and in simulation for all of the channels. For the channels which trigger on τ_h, the $\mu\tau_h$ and $\tau_h\tau_h$ channels, the τ_h efficiency is measured using tag-and-probe in the $\mu\tau_h$ final state. The efficiencies for the $\tau_h\tau_h$ channel are measured with dedicated muon+τ_h monitoring triggers which are triggers with the same τ_h requirements as the $\tau_h\tau_h$ trigger while they have a muon instead of a second τ_h to enable the tag-and-probe method. The tag-and-probe is conducted using the Single Muon dataset. The selection for the τ_h trigger efficiency measurements requires one well identified and isolated muon per event which is required to fire a single muon trigger, see Table 6.1 and Reference [125]. All events with at least one isolated τ_h identified by its decay mode are considered. Of those events those passing the trigger under study are counted. The ratio of the count of events passing the trigger to those considered for the study defines the trigger efficiency. Trigger efficiencies are measured as a function of τ_h: p_{T}, decay mode, and generator matching status. Figure 6.8 shows the trigger efficiency versus p_{T} for the double-τ_h trigger.

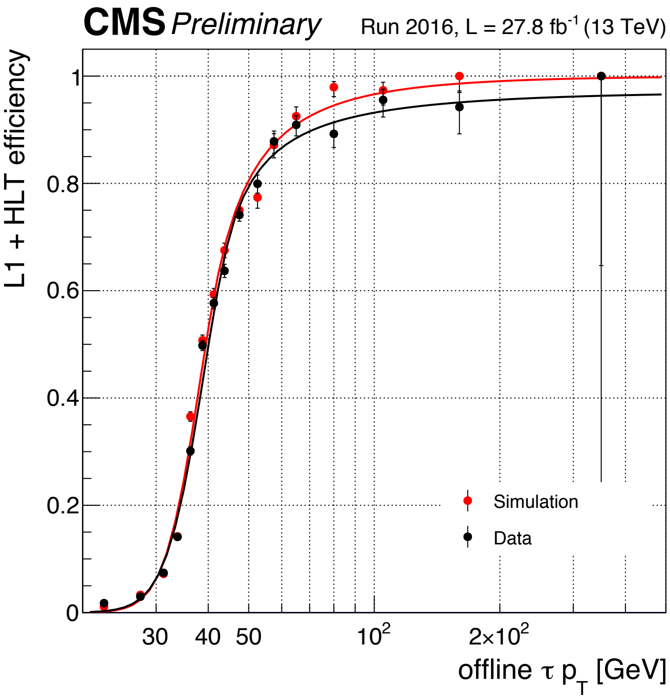


Figure 6.8: Comparison of the trigger efficiencies measured for a single leg of the double-τ_h with a muon+τ_h monitoring trigger. The efficiencies shown here are used in the $\tau_h\tau_h$ channel. The specific distribution shown is for genuine τ leptons and includes all used decay modes.

The trigger efficiencies for electrons and muons are also measured with a tag-and-probe method. Similar to the electron and muon identification and isolation efficiencies, the electron (muon) trigger efficiencies are measured in $Z \to ee$ ($Z \to \mu\mu$) events. The denominator selection for the probe lepton requires the same identification and isolation criteria used in the analysis. The efficiencies are measured as a function of lepton p_T and η.

6.6.6 Drell–Yan Reweighting

The simulated Monte Carlo Drell–Yan sample used in the analysis is simulated at leading order (LO) using the MadGraph5_aMC@NLO generator discussed in section 4. To assess Monte Carlo to data agreement a dedicated $Z/\gamma^* \to \mu\mu$ control region with high Drell–Yan purity was used. The $Z/\gamma^* \to \mu\mu$ control region selected events from the Single Muon PD and requires the presence of two well-identified muons (Table 6.2) with $p_\mathrm{T} > 25\,\mathrm{GeV}$ and an invariant mass between 70 and 110 GeV. With this selection criteria over 99% of the events are attributed to $Z/\gamma^* \to \mu\mu$ decays.

Differences in the distributions of $m_{\ell\ell}$ and $p_\mathrm{T}^{\ell\ell}$ between data and in simulations are observed in this control region. To correct for this, 2D weights based on these variables are derived and applied to simulated $Z/\gamma^* \to \tau\tau, \ell\ell$ events in the signal region of the analysis. These corrections range from 1%–10% with a relative uncertainty of 10%. In addition, corrections depending on m_jj are derived from the $Z/\gamma^* \to \mu\mu$ region and applied to the $Z/\gamma^* \to \tau\tau, \ell\ell$ simulation for events with at least two jets passing the VBF category selection criteria. These corrections range from 2%–6% and have an assigned uncertainty of 100%. The uncertainty covers the range of different corrections which can be measured based on varying the measurement region phase space. After this reweighting, good agreement between data in the $Z/\gamma^* \to \mu\mu$ region and simulation is found for all other variables used in the analysis. These derived corrections are also applied to the simulated samples of electroweak production of

Z bosons. Electroweak produced Z bosons contribute up to 8% of the Z boson production in the VBF category.

6.6.7 Hadronic Recoil Corrections

Recoil corrections are applied to the Drell–Yan and $W+$jets simulated background events as well as the simulated signal events for Higgs bosons produced via the gluon fusion or VBF mechanisms. Recoil corrections correct for the mismodeling of $\vec{E}_T^{miss}$ in the simulated samples. A variable, $\vec{U}$, is defined as the vectorial difference of the measured $\vec{E}_T^{miss}$ and the total transverse momentum of neutrinos originating from the decay of the boson.

$$\vec{U} = \vec{E}_T^{miss} - \vec{p}_{\mathrm{T},\nu} \tag{6.3}$$

For Z bosons decaying leptonically to ℓ, $\vec{U}$ can be expressed as,

$$\vec{U} = \vec{p}_{\mathrm{T,B}} - \vec{H}_{\mathrm{T}} \tag{6.4}$$

where $\vec{H}_{\mathrm{T}}$ is the transverse momentum of hadronic recoil and $\vec{p}_{\mathrm{T,B}}$ is the leptonic recoil and is the transverse momentum of leptonically decaying Z, W, or Higgs bosons.

Like the Drell–Yan Reweighting mentioned above, $\vec{U}$ is measured in $Z \to \mu\mu$ events. It is observed that the $\vec{H}_{\mathrm{T}}$ from simulation in $Z \to \mu\mu$ does not align with the $\vec{H}_{\mathrm{T}}$ from data. A reweighting method is derived to align the simulated distribution with data. The $Z/\gamma^* \to \mu\mu$ final state is used because of the absence of neutrinos in the Z decay making the comparison of the simulation to data and derivation of the reweighting technique far more simple. For the measurement and application of the corrections $\vec{U}$ is decomposed into two components, one which is parallel to $\vec{p}_{\mathrm{T,B}}$ ($\vec{U_1}$) and one which is orthogonal ($\vec{U_2}$).

6.6.8 Generator Event Weights

Generator weights are applied on a per event basis. Samples produced with the MadGraph5_aMC@NLO generator contain both positive and negative event weights.

6.6.9 Luminosity Weighting

The event weights for simulated events are scaled to the expected yields for each sample. A per-event weight (W) is defined based on the number of generated events (N), the best theory predicted cross section for the process (σ), and the integrated luminosity of the dataset being modeled ($\mathcal{L}$), 35.9 fb^{-1}.

$$\mathrm{W} = \mathcal{L}\,\frac{\sigma}{\mathrm{N}} \tag{6.5}$$

For samples with negative generator weights, N is redefined as the sum total of the event weights to properly account for the negative event weights.

6.7 Systematic Uncertainties

The systematic uncertainty model for the $H \rightarrow \tau\tau$ study was critical to the success of the analysis. The goal of the analysis was to reach an observed significance for the $H \rightarrow \tau\tau$ process of 5.0σ. To achieve this the description of the expected background processes and Higgs boson signal needed to be as close to data as possible. The previous Monte Carlo Reweighting section 6.6 detailed this process. Many of the corrections and weights applied to the simulated samples have associated uncertainties which must be represented in the global fit. Other uncertainties come from the measurement of the data itself such as the uncertainty on the integrated luminosity. In addition to these, there are also theoretical uncertainties on aspects such as the production cross sections for the simulated backgrounds and Higgs boson signals. These are all detailed in the sections below.

6.7.1 Integrated Luminosity Uncertainties

The total data taken by the CMS experiment in 2016 that is marked usable for physics is $35.9\,fb^{-1}$. The uncertainty on this $35.9\,fb^{-1}$ integrated luminosity value amounts to 2.5% [132].

6.7.2 Object Reconstruction, Identification and Trigger Uncertainties

The total uncertainty in the τ_h identification efficiency measurement for genuine τ_h leptons is 5% as mentioned in Section 6.6.2. The τ_h are required to pass different identification criteria per channel specifically aimed at reducing the chance of fake τ_h originating from different background processes in each final state. Because of the different anti-e and anti-μ discriminators applied in each channel, the 5% uncertainty is not fully correlated among all the di-τ channels.

The trigger efficiency uncertainty per τ_h candidate results from the uncertainties on the efficiency fit function where the fit is shown in Figure 6.8 for the $\tau_h\tau_h$ channel. The uncertainties result from limited statistics in the $e\tau_h$ and $\mu\tau_h$ channels as a fit function is not used for application of the efficiencies for these channels. The τ_h trigger uncertainties amounts to 5%. This results in a total trigger uncertainty of $5\% + 5\% = 10\%$ for processes estimated from simulation in the $\tau_h\tau_h$ decay channel.

In the decay channels with muons or electrons, the uncertainties for the muon and electron identification, isolation, and trigger efficiencies lead to a total rate uncertainty of 2% for muons and 2% for electrons.

6.7.3 Tau Fake Rate

For events where muons or electrons are misidentified as τ_h candidates, the τ_h identification leads to rate uncertainties of 25% and 12%, respectively, per reconstructed τ_h decay mode [125]. These fake τ_h dominantly come from $Z \rightarrow \mu\mu$ events in the $\mu\tau_h$ decay channel and $Z \rightarrow ee$ events in the $e\tau_h$ decay channel. Using m_{vis} and the reconstructed τ_h decay mode as the observables in the 0-jet category of the $\mu\tau_h$ and $e\tau_h$ channels helps constrain these uncertainties during the signal extraction fit. The post-fit uncertainty in the rate of muons or electrons misidentified as τ_h per decay mode becomes of the order of 5%. This is an 80% constraint for the muon misidentified as τ_h rate and a 40% constraint on the electron misidentified as τ_h rate.

6.7.4 Energy Scales

Visible Tau Energy Scale

An uncertainty of 1.2% in the visible energy scale of genuine τ_h leptons affects both the shape of the signal and background distributions and the expected yields [125]. The 1.2% uncertainty is treated as uncorrelated among the different τ_h decay modes: 1-prong, 1-prong+π^0, and 3-prong. The magnitude of the uncertainty was determined based on the uncertainties on the measured τ_h energy corrections in Section 6.6.3. The measurement was conducted in the $\mu\tau_h$ final state and does have some overlap with the $\mu\tau_h$ channel signal regions. However, the event overlap is found to be less than 50% between the τ_h energy correction and visible τ_h energy scale measurement region and the $\mu\tau_h$ signal region used in this analysis. To prevent over constraining the uncertainties from fitting the data twice, the measured uncertainties which are approximately 0.5% are inflated to the value used in the analysis, 1.2%.

When extracting the results from the maximum likelihood fit discussed in Section 6.8, the fit constrains the visible τ_h energy scale uncertainty to about 0.3%, a 75% constraint, for all decay modes. The constraint mostly comes from highly populated regions with high τ_h purity, for example, the 0-jet and Boosted categories of the $\mu\tau_h$ and $\tau_h\tau_h$ channels. The constraint on the size of the uncertainty is largely explained by the addition of two other decay channels with τ_h candidates ($\tau_h\tau_h$ and $e\tau_h$) in addition to to the $\mu\tau_h$ measurement region which significantly increase the available statistics reducing the uncertainty. Additionally, the higher number of events in the Monte Carlo simulations and the finer categorization which leads to regions with a high $Z \to \tau\tau$ event purity also contribute to the post-fit constraint on the visible tau energy scale. Even in the most Boosted analysis categories, reconstructed τ_h candidates typically have moderate p_T (less than 100 GeV) and are found in the barrel region of the detector. Tracks are well measured in the CMS detector in this p_T range. Therefore, the visible energy scale of genuine τ_h leptons is fully correlated for all τ_h leptons reconstructed in the same decay mode, irrespective of their p_T and η. The uncertainties in the visible energy scale for genuine τ_h leptons together contribute an uncertainty of 5% to the measurement of the signal strength.

Energy Scale for Fake Taus

The energy scale uncertainty for muons or electrons misidentified as τ_h candidates is 1.5 or 3%, respectively [125]. It is uncorrelated between reconstructed τ_h decay modes. The fit constrains these uncertainties to about one third of their initial values. For events where quark- or gluon-initiated jets are misidentified as τ_h candidates, a linear uncertainty that increases by 20% per 100 GeV in τ_h p_T accounts for a potential mismodeling of the jet$\to$ τ_h misidentification rate as a function of the τ_h p_T in simulations. The uncertainty has been determined from a region enriched in $W+$jets events, using events in the $\mu\tau_h$ final state, characterized by a large transverse mass between the $\vec{E}_T^{miss}$ and the muon [124, 125].

Jet Energy Scale

The uncertainties in the jet energy scale depend on the p_{T} and η of the jet [121]. They are factorized into 27 distinct uncorrelated contributions. These factorized uncertainties target jet related mismodeling for different CMS subdetectors, namely ECAL and HCAL, uncertainties in jet energy corrections, timing dependence, high p_{T} jet extrapolations uncertainties, and large $|\eta|$ jet related uncertainties. The uncertainties affect the p_{T} of each jet in an event. As the p_{T} for each jet increases or decreases, the number of jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ is recomputed as is the m_{jj} for each event. The adjusted number of jets and m_{jj} are used to determine the new, shifted, event categorization. For example, events can shift between the 0-jet and Boosted category if a jet has a p_{T} value near 30 GeV that shifts into or out of categorization as a jet by the analysis (jet $p_{\mathrm{T}} > 30\,\mathrm{GeV}$). As m_{jj} is shifted, the shifted m_{jj} value is used to plot the event in the 2D VBF distribution.

MET Energy Scale

The $\vec{E}_T^{miss}$ scale uncertainties [133] are computed event-by-event and affect the normalization of various processes through the event selection in the $\ell\tau_h$ channels which employ the low-m_{T}, $m_{\mathrm{T}} < 50\,\mathrm{GeV}$ criteria. The $\vec{E}_T^{miss}$ scale uncertainties also affect event distributions through the propagation of these uncertainties to the di-τ mass $m_{\tau\tau}$. The $\vec{E}_T^{miss}$ scale uncertainties which have the largest impact on the signal significance in this analysis are from jet related uncertainties and unclustered energy deposits in the detector. The unclustered energy deposits come from four independent sources related to the tracker, ECAL, HCAL, and forward calorimeter subdetectors [134]. The jet related $\vec{E}_T^{miss}$ scale uncertainties arise from uncertainties in the jet energy scale measurement. As the scale of the jet energy is varied, these changes are propagated to the $\vec{E}_T^{miss}$. The combination of both sources of uncertainties in the $\vec{E}_T^{miss}$ scale leads to an uncertainty of about 10% in the measured signal strength.

Electron And Muon Energy Scale

The uncertainty in the electron energy scale is 2.5% in the endcaps and 1% in the barrel of the detector [49]. It is only relevant in the $e\mu$ decay channel, where it affects the final distributions shape and yield. In all channels, the effect of the uncertainty in the muon energy scale is negligible and is not included in the systematic model.

6.7.5 b-Tagging Uncertainties

The $H \to \tau\tau$ analysis only applies a b-tagged jet selection (Section 6.1.3) criteria in the $e\mu$ channel, where the number of b-tagged jets must equal zero in the signal region. The rate uncertainty related to discarding events with a b-tagged jet in the $e\mu$ channel is up to 5% for the $t\bar{t}$ background. The uncertainty in the mistagging rate of gluon and light-flavor jets is negligible.

6.7.6 Background Estimation Uncertainties

Drell–Yan Uncertainties

The $Z \to \tau\tau$ background yield and distribution are corrected based on the agreement between data and the background prediction in a control region enriched in the $Z \to \mu\mu$ events, as explained in Section 6.6.6. The extrapolation uncertainty related to kinematic differences between the selections in the signal and control regions ranges between 3 and 10%, depending on the category. Shape uncertainties related to the uncertainties in the applied corrections are considered. They reach 20% for some ranges of m_{jj} in the VBF category, but are more modest in the $\tau_h\tau_h$ channel where the uncertainties range from 0% in the lowest m_{jj} region to 6% in the $500\,\mathrm{GeV} < m_{\mathrm{jj}} < 800\,\mathrm{GeV}$ region [114]. These uncertainties arise from the different level of agreement between data and simulation in the $Z \rightarrow \mu\mu$ control region obtained when varying the threshold on the muon p_{T}.

$W +$ jets Uncertainties

The uncertainties in the $W +$ jets event yield for the $\mu\tau_h$ and $e\tau_h$ channels are determined from their $W+$jets control regions. The uncertainties account for the statistical uncertainties of the data and $W +$ jets simulated samples. An extrapolation uncertainty is also determined which accounts for scaling from the high-m_{T} ($m_{\mathrm{T}} > 80\,\mathrm{GeV}$) control regions to the low-m_{T} ($m_{\mathrm{T}} < 50\,\mathrm{GeV}$) signal regions. The scaling uncertainty is obtained by comparing the m_{T} distributions of simulated and observed $Z \rightarrow \mu\mu$ events where one of the muons is removed and the $\vec{E}_T^{miss}$ is adjusted accordingly, to mimic $W +$ jets events. The reconstructed invariant mass of the Z in the rest frame is multiplied by the ratio of the W to the Z boson masses before removing the muon. These uncertainties range from 5 to 10% [114].

In the $\tau_h\tau_h$ and $e\mu$ channels, where the $W +$ jets background is estimated from simulation, the uncertainty on the yield of this small background is 4% and 20%, respectively. The larger value for the $e\mu$ channel includes uncertainties in the misidentification rates of jets as electrons and muons, whereas the uncertainty in the misidentification rate of jets as τ_h candidates in the $\tau_h\tau_h$ channel is accounted for by the linear uncertainty as a function of the τ_h p_{T} described in Section 6.7.4.

QCD Multijet Uncertainties

The uncertainty in the QCD multijet background yield in the $\ell\tau_h$ and $e\mu$ channels ranges from 10 to 20% [114]. They are category dependent and account for the uncertainty on the extrapolation from the same-sign measurement to the opposite-sign signal regions. In the $\ell\tau_h$ channels, statistical uncertainties in the QCD multijet control regions are automatically taken into account by the global fit with the control regions and signal regions.

In the $\tau_h\tau_h$ channel, the uncertainty in the QCD multijet background yield reflects a combination of the statistical uncertainties obtained from fitting the dedicated control regions with τ_h candidates passing relaxed isolation criteria defined in Section 6.5, and for extrapolation uncertainties to the signal region. Limited disagreement between prediction and data in signal-free regions with various loose isolation criteria is also accounted for in the QCD multijet uncertainties.

The level of agreement for the $\tau_h\tau_h$ QCD multijet estimation method is determined in signal free regions where the τ_h have loosened isolation requirements and at least one of the τ_h must not meet the signal region `Tight Tau MVA` criteria (Section 5.3.3). A schematic depiction of the signal region, QCD multijet control region, and method validation regions are shown in Figure 6.9. With both validation regions constructed to not overlap with the signal region, the method closure can be assessed.

The QCD multijet estimation method in the $\tau_h\tau_h$ channel is found to agree well when validated using closure checks in the loosened isolation control regions depicted in Figure 6.9. Closure is estimated by comparing the “predicted” QCD multijet yield in the opposite-sign `Tighter Validation Region` against the “measured” QCD multijet yield in the same region. The “measured” QCD multijet yield in the `Tighter Validation Region` is calculated from background subtracted data. The “predicted” QCD multijet yield in the same region comes from multiplying the measured yield in

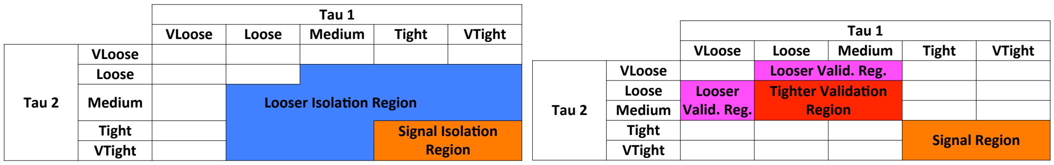

Tau 1
VLoose
Loose
Medium
Tight
VTight
Tau 2
VLoose
Loose
Medium
Tight
VTight
Looser Isolation Region
Signal Isolation Region
Tau 1
VLoose
Loose
Medium
Tight
VTight
Tau 2
VLoose
Loose
Medium
Tight
VTight
Looser Valid. Reg.
Looser Valid. Reg.
Tighter Validation Region
Signal Region

Figure 6.9: (Left) For reference, previously shown signal region and QCD multijet relaxed isolation estimation region. (Right) Schematic of the further loosened MVA isolation (defined in Section 5.3.3) selections used for the QCD multijet validation. The `Tighter Validation Region` (red) in the right plot is treated just like the `Signal Isolation Region` on the left plot. The `Looser Validation Region` (pink) in the right plot is treated just like the `Looser Isolation Region` on the left plot.

the opposite-sign `Looser Validation Region` by a loose-to-tight scale factor derived in the same-sign region just like equation 6.2. The closure values for the three analysis categories is shown in table 6.7.6 along with the associated statistical uncertainty on the measurement and the $\tau_h\tau_h$ channel QCD multijet estimation method systematic uncertainty. The systematic uncertainty on the method is added in quadrature with the statistical uncertainty on the loose-to-tight scale factor to arrive at a total per category uncertainty for QCD multijet estimation.

Category	"Estimate" / "Predicted"	Stat. Uncert. on Closure	QCD Method Syst. Uncert.	QCD Method Stat. Uncert.	QCD Total Uncert.
0-jet	0.98	0.94%	2.5%	1.0%	2.7%
VBF	0.95	7.4%	12%	9.8%	15%
Boosted	0.99	1.3%	2.3%	1.4%	2.7%

Table 6.7: QCD multijet uncertainties for the $\tau_h\tau_h$ channel for each category. The "Estimate" / "Predicted" column is the closure found for the QCD multijet estimation method validation. The statistical uncertainty on the closure value is found from adding in quadrature the statistical uncertainty of each of the four regions used in the comparison (SS vs. OS and Looser vs. Tighter Isolation). The QCD method systematic uncertainty is calculated as the deviation from unity for the closure summed with the statistical uncertainty on the closure test. The QCD method statistical uncertainty is derived from the uncertainty in the same-sign regions used to calculate the loose-to-tight scale factor. The final uncertainty values used in the analysis, the QCD total uncertainty is the sum in quadrature of the QCD method systematic uncertainty and QCD method statistical uncertainty.

$t\bar{t}$ Uncertainties

The uncertainty from the fit in the $t\bar{t}$ control region, discussed in section 6.5, is automatically propagated to the signal regions. It results in an uncertainty of about 5% on the $t\bar{t}$ cross section. Per-channel uncertainties related to the object reconstruction and identification are considered when extrapolating from the $e\mu$ final state to the others [114]. The $t\bar{t}$ simulation is corrected for differences in the top quark p_{T} distributions observed between data and simulation. An uncertainty in the correction is taken into account.

Other Background Uncertainties

The combined systematic uncertainty in the background yield arising from diboson and single top quark production processes is estimated to be 5% on the basis of recent CMS measurements [135, 136].

6.7.7 Theoretical Uncertainties for Higgs Boson

Rate and acceptance uncertainties for the Higgs boson signal processes are derived from theoretical calculations. They are largely due to uncertainties in the PDFs (Section 4.2), variations of the QCD renormalization and factorization scales [137], and uncertainties in the modelling of parton showers. The magnitude of the rate uncertainty depends on the production process and on the event category.

The inclusive uncertainty related to the PDFs amounts to 3.2, 2.1, 1.9, and 1.6%, respectively, for the ggH, VBF, WH, and ZH production modes [42]. The corresponding uncertainty for the variation of the renormalization and factorization scales is 3.9, 0.4, 0.7, and 3.8%, respectively [42]. The PDF-based acceptance uncertainties related to the particular selection criteria used in this analysis are less than 1% for the

ggH and VBF production mechanisms. The acceptance uncertainties for VBF production in the renormalization and factorization scale uncertainties are less than 1%, while the corresponding uncertainties for the ggH process are treated as shape uncertainties. The renormalization and factorization scale uncertainties increase linearly with $p_{\mathrm{T}}^{\tau\tau}$ and m_{jj}.

The p_{T} distribution of the Higgs boson in the POWHEG 2.0 simulations is tuned to match more closely the next-to-NLO (NNLO) plus next-to-next-to-leading-logarithmic (NNLL) prediction in the HRES2.1 generator [94, 138]. The signal acceptance changes with the variation of the parton shower tune in HERWIG ++ 2.6 samples [139] are considered as additional uncertainties, and amount to up to 7% in the Boosted category. The theoretical uncertainty in the branching fraction of the Higgs boson to τ leptons is equal to 2.1% [42].

The theoretical uncertainties in the signal production depend on the jet multiplicity. This effect is included by following the prescriptions in Reference [140]. These uncertainties need to be taken into account because the definitions of the three categories used in the analysis are based partially on the number of reconstructed jets. Additional shape uncertainties for boosted Higgs bosons, related to the treatment of the top quark mass in the calculations, are considered for signal events with $p_{\mathrm{T}}^{\tau\tau} > 150\,\mathrm{GeV}$. The boosted Higgs boson shape uncertainties are motivated by the treatment suggested in Reference [42].

Theory uncertainties in the signal prediction contribute an uncertainty of 10% to the measurement of the signal strength.

6.7.8 Other Uncertainties

There are additional uncertainties related to the finite number of simulated events and to the limited number of events in data control regions which are taken into account. They are considered for all bins of the distributions used to extract the results. In total, these per-bin uncertainties contribute an uncertainty of about 12% to the signal strength measurement where the majority of this uncertainty comes from sparsely populated shape templates in the VBF category.

6.7.9 Systematic Model Summary

The systematic uncertainties considered in the analysis are summarized in Table 6.7.9.

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	N/A
$\vec{E}_T^{miss}$ energy scale	Dependent upon p_T and η	N/A
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%	N/A
μ ID & isolation & trigger	2%	N/A
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
$Z \to \tau\tau/\ell\ell$ estimation	Normalization: 7–15%	3–15
	Uncertainty in $m_{\ell\ell/\tau\tau}$, $p_\mathrm{T}(\ell\ell/\tau\tau)$, and m_jj corrections	N/A
$W+\text{jets}$ estimation	Normalization ($e\mu$, $\tau_h\tau_h$): 4–20%	N/A
	Uncert. from CR ($e\tau_h$, $\mu\tau_h$): $\simeq$5–15	N/A
	Extrap. from high-m_T CR ($e\tau_h$, $\mu\tau_h$): 5–10%	N/A
QCD multijet estimation	Normalization ($e\mu$): 10–20%	5–20%
	Uncert. from CR ($e\tau_h$, $\tau_h\tau_h$, $\mu\tau_h$): $\simeq$5–15%	N/A
	Extrap. from anti-iso. CR ($e\tau_h$, $\mu\tau_h$): 20%	7–10
	Extrap. from anti-iso. CR ($\tau_h\tau_h$): 3–15%	3–10
Diboson normalization	5%	N/A
Single top quark normalization	5%	N/A
$t\bar{t}$ estimation	Normalization from CR: $\simeq$5%	N/A
	Uncertainty on top quark p_T reweighting	N/A
Integrated luminosity	2.5%	N/A
b-tagged jet rejection ($e\mu$)	3.5–5.0%	N/A
Limited number of events	Statistical uncertainty in individual bins	N/A
Signal theoretical uncertainty	Up to 20%	N/A

Table 6.8: Sources of systematic uncertainty. If the global fit to the signal and control regions, described in the next section, significantly constrains these uncertainties, the values of the uncertainties after the global fit are indicated in the third column. The acronyms CR and ID stand for control region and identification, respectively.

6.8 Results

The extraction of the results uses a global maximum likelihood fit based on visible or corrected invariant mass of the τ-pair. All channels are simultaneously fit together with the control regions for the $t\bar{t}$, QCD multijet, and $W + \text{jets}$ backgrounds. The following Section 6.8.1 shows six of the twelve distributions used in the global maximum likelihood fit. A more detailed description of the fit and fitting process is laid out afterwards in section 6.8.2. After these two sections, the rest of the $H \to \tau\tau$ results are discussed. The following Chapter 7 discusses the WH and ZH associated production $H \to \tau\tau$ analysis. A combination of this ggH and VBF targeted analysis with the associated production analysis follows in Chapter 8.

6.8.1 Two-Dimensional Distributions

A selection of the 2D distributions used in the fit, including the three $\tau_h\tau_h$ distributions and the three $\mu\tau_h$ distributions, are shown in Figures 6.10–6.15. In these distributions, the normalization of the predicted background distributions corresponds to the result of the global fit and the signal distribution is normalized to its best fit signal strength. The choice of the binning is driven by the available statistics of the background and data templates. This leads to wider $m_{\tau\tau}$ bins and fewer slices of m_{jj} in the poorly-populated VBF category, and conversely, the ability of the $\mu\tau_h$, $e\tau_h$, and $e\mu$ channels to include many slices of $p_{\text{T}}^{\tau\tau}$ in the Boosted category. The most sensitive category, VBF, is shown first and is followed by the Boosted and 0-jet categories. The signal prediction for a Higgs boson with $m_H = 125.09\,\text{GeV}$ is normalized to its best fit signal strength. The background distributions are adjusted to the results of the global maximum likelihood fit.

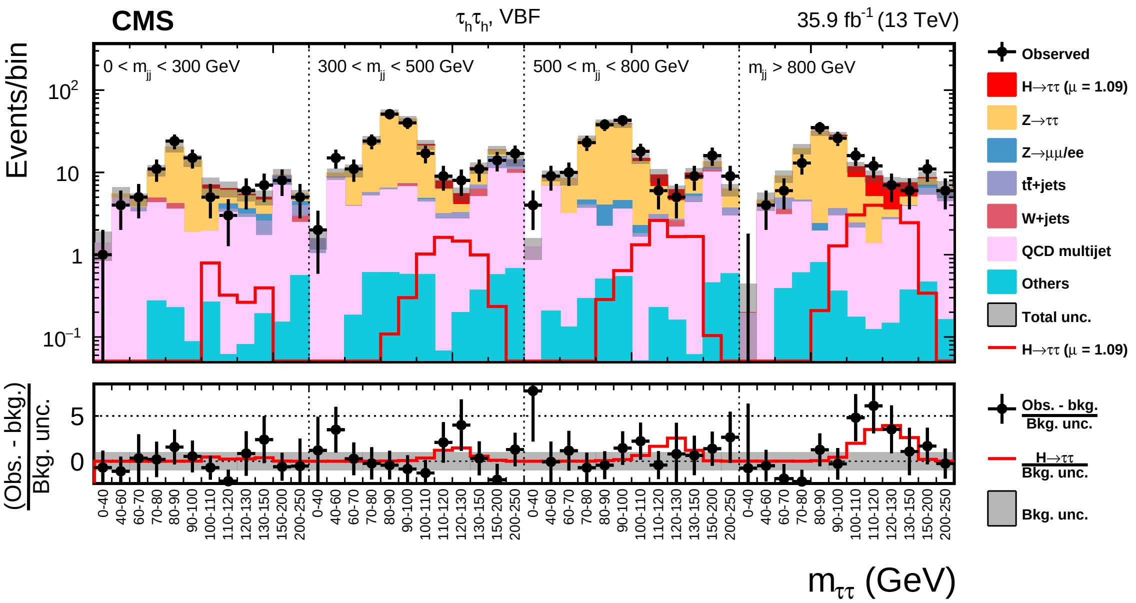


Figure 6.10: Observed and predicted 2D distributions in the VBF category of the $\tau_h\tau_h$ channel. The two variables are the mass, $m_{\tau\tau}$, and the di-jet mass, $m_{\rm jj}$. The background uncertainty band accounts for all sources of background uncertainty. The signal is shown both as a stacked filled histogram and as an open overlaid non-stacked histogram.

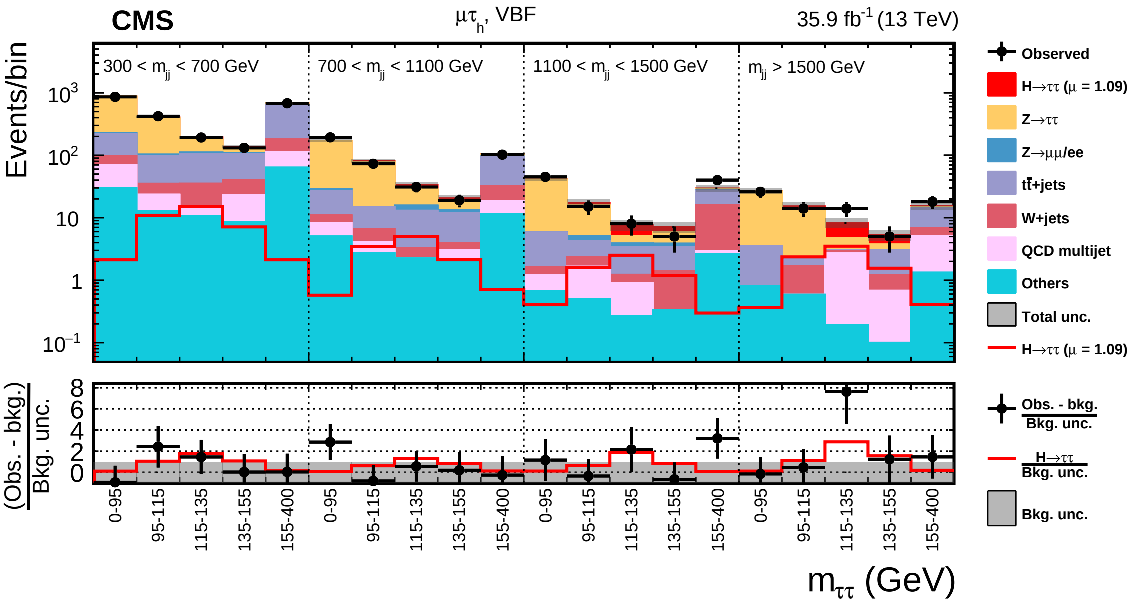


Figure 6.11: Observed and predicted 2D distributions in the VBF category of the $\mu\tau_h$ channel. The description of the histograms is the same as in Figure 6.10.

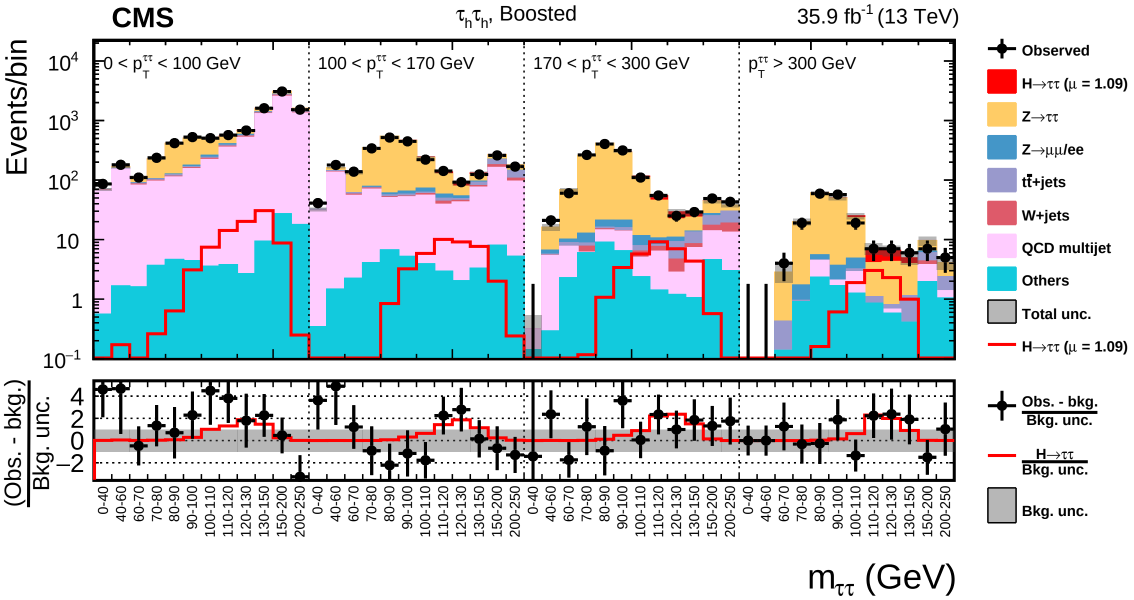


Figure 6.12: Observed and predicted 2D distributions in the Boosted category of the $\tau_h\tau_h$ channel. The description of the histograms is the same as in Figure 6.10.

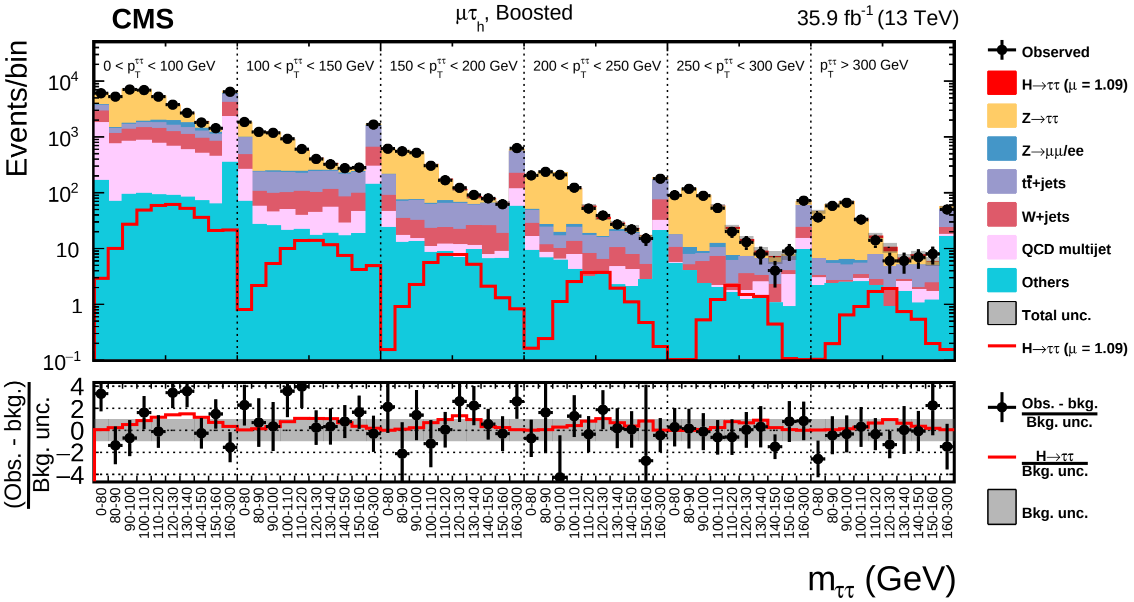


Figure 6.13: Observed and predicted 2D distributions in the Boosted category of the $\mu\tau_h$ channel. The description of the histograms is the same as in Figure 6.10.

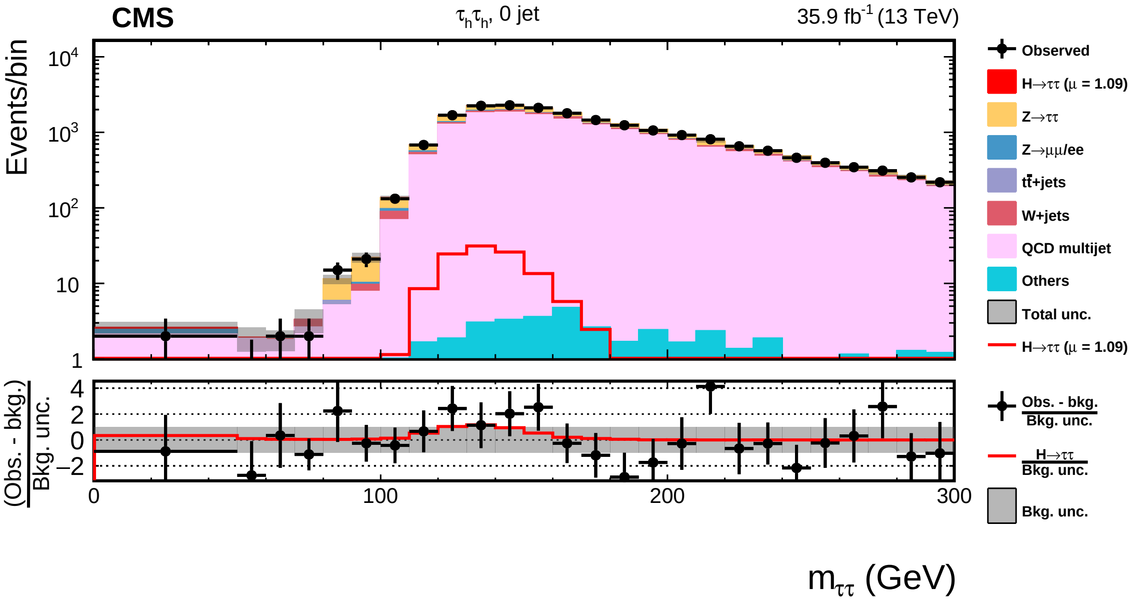


Figure 6.14: Observed and predicted distributions in the 0-jet category of the $\tau_h\tau_h$ channel. The description of the histograms is the same as in Figure 6.10.

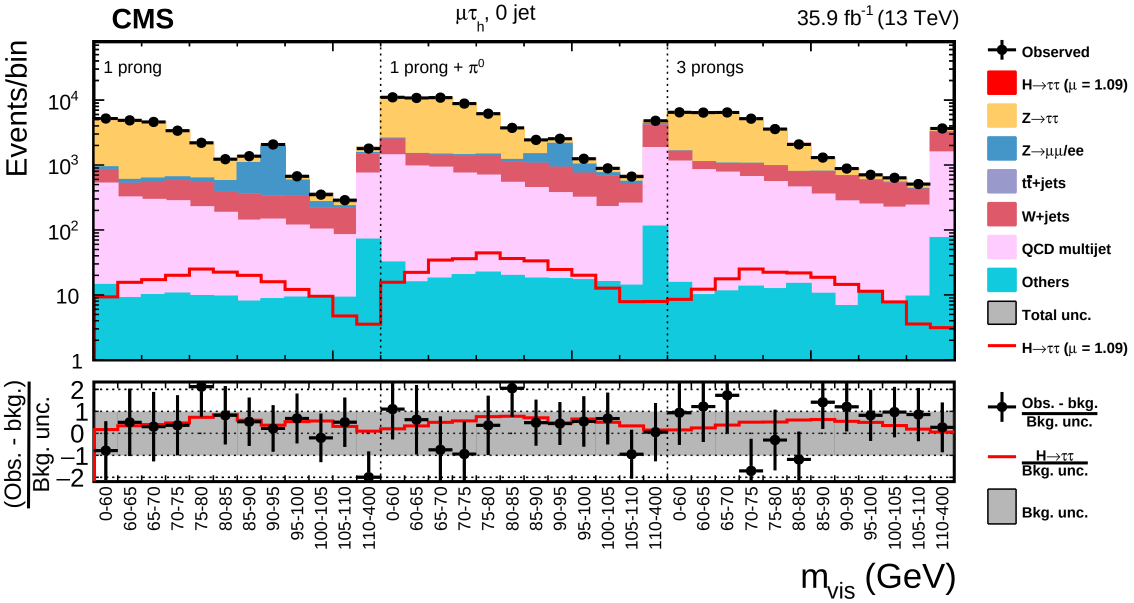


Figure 6.15: Observed and predicted 2D distributions in the 0-jet category of the $\mu\tau_h$ channel. The description of the histograms is the same as in Figure 6.10.

6.8.2 Results Extraction

The 2D distributions of the final discriminating variables, see Table 6.3, obtained for each category and each channel in the signal regions, along with the control regions, are combined in a binned likelihood fit involving the background and signal expectations and observed data in each bin. The expected number of signal events is predicted for the production of a SM Higgs boson of mass $m_H = 125.09\,\text{GeV}$ decaying into a pair of τ leptons. In the fit the number of signal events is multiplied by a signal strength modifier μ treated as a free parameter in the fit.

The systematic uncertainties are represented by nuisance parameters that are varied in the fit according to their defined probability density functions. A log-normal probability density function is assumed for the nuisance parameters affecting the event yields of the various background contributions such as theory-based cross section uncertainties and electron, muon, and τ_h reconstruction efficiencies. Systematic uncertainties that affect the shape of the distributions, such as the visible τ_h energy scale and $\vec{E}_T^{miss}$ energy scales, are represented by nuisance parameters whose variation is assumed to have a Gaussian probability density function. Overall, the statistical uncertainty in the observed event yields is the dominant source of uncertainty for the combined results.

6.8.3 Analysis Sensitivity Details

The 2D distributions shown in Figures 6.10–6.15 indicate which bins are the most sensitive to the Higgs boson signal. Each bin in these distributions can be reordered and grouped based on the sensitivity of that bin. For visualization purposes only, we regroup events in the signal region by their decimal logarithm of the ratio of the signal (S) to signal-plus-background ($S{+}B$) in each bin as shown in Figure 6.16. This method helps smooth out statistical fluctuations visible in the 2D distributions. An excess of observed events with respect to the SM background expectation is visible in the most sensitive bins of the analysis. The details of the expected background and signal contributions and their uncertainties in the most sensitive bins in this plot are broken down in table 6.8.3. These values are compared against the observed number of events in these sensitive bins. The table is defined for bins with $\log_{10}(S/(S+B)) > -0.9$. The channel that contributes the most to these highly sensitive bins is $\tau_h\tau_h$.

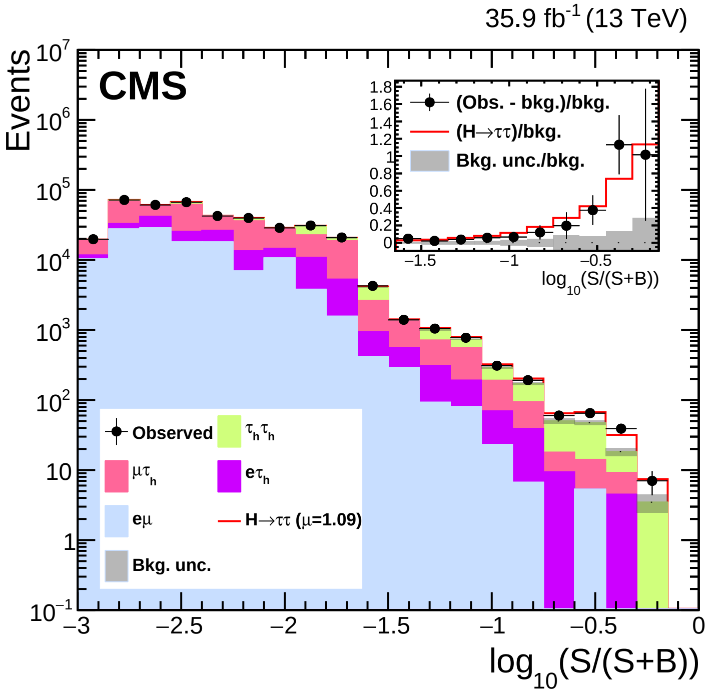


Figure 6.16: Distribution of the decimal logarithm of the ratio between the best fit signal and the sum of the best fit signal and best fit background expectations in each bin of the mass distributions used to extract the results. All signal regions and channels are included. Background contributions are broken down by channel. The inset shows the corresponding difference between the observed data and best fit background distributions divided by the best fit background expectation. The best fit signal expectation is also divided by the background expectation in the inset.

Process	$e\mu$	$e\tau_h$	$\mu\tau_h$	$\tau_h\tau_h$
$Z \to \tau\tau$	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
$t\bar{t}$+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
$\gamma\gamma H, H \to \tau\tau$	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
V$H, H \to \tau\tau$	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

Table 6.9: Best-fit background and signal expectations, together with the number of observed events, for highly sensitive bins in the signal region. High sensitivity bins are defined by $\log_{10}(S/(S+B)) > -0.9$, where S and B are the number of best fit expected signal events for a Higgs boson with a mass $m_H = 125.09\,\text{GeV}$ and of best fit expected background events. The background uncertainty accounts for all sources of background uncertainty, systematic as well as statistical, after the global fit. The contribution from "other backgrounds" includes events from diboson and single top quark production. The contribution from Higgs boson decays to a pair of W bosons is zero in these bins.

6.8.4 Mass Distributions

Another way to visualize the results of the global maximum likelihood fit is to collapse the second dimension of the 2D distributions. An excess of observed events relative to the best fit background expectation is shown in Figure 6.17, with the details of the figure's construction following. All of the $m_{\tau\tau}$ distributions with a common binning scheme are collapsed into a single $m_{\tau\tau}$ distribution. A reweighting scheme defines a weight for each $m_{\tau\tau}$ range within a slice of m_{jj} or $p_{\text{T}}^{\tau\tau}$. The reweighting scheme is designed to increase the contribution of the most sensitive distributions. The reweighting is defined by summing the best fit background expectations and summing the best fit signal contributions for a given slice of m_{jj} or $p_{\text{T}}^{\tau\tau}$; a weight of $S/(S+B)$ is applied to the these contributions when plotting them in the described distribution.

As a detail S and B are computed as the signal or background contribution in the mass distribution excluding the first and last bins, in which the amount of signal is negligible. The signal regions that use m_{vis} instead of $m_{\tau\tau}$, namely the 0-jet category of the $\mu\tau_h$, $e\tau_h$ and $e\mu$ channels, are not included. The different $m_{\tau\tau}$ binning schemes necessitate two versions of these distributions because the bin boundaries do not align so cannot be merged. The two different reweighted $m_{\tau\tau}$ distributions are in Figure 6.17 and group the $m_{\tau\tau}$ distributions which have aligning bin boundaries from Figures 6.10–6.15.

6.8.5 Signal Significance And Signal Strength

The excess in data is calculated from the corresponding local p-value using a profile likelihood ratio test statistic [141, 142, 143, 144] for each Higgs boson mass point included in the analysis. As shown in Figure 6.18, the observed significance for a SM Higgs boson with $m_H = 125.09\,\text{GeV}$ is 4.9 standard deviations, for an expected significance of 4.7 standard deviations. Testing multiple mass hypotheses shows the greatest observed significance for a SM Higgs boson with the expected SM Higgs boson value $m_H = 125.09\,\text{GeV}$.

The corresponding best fit value for the signal strength μ is $1.09^{+0.27}_{-0.26}$ at $m_H = 125.09\,\text{GeV}$. This result shows very good agreement with the best Standard Model Higgs boson calculated cross sections and $H \to \tau\tau$ branching ratio. The individual best fit signal strengths per channel and per category, using the constraints obtained on the systematic uncertainties through the global maximum likelihood fit, are given in Figure 6.19; they demonstrate consistency of the observation with the SM Higgs boson hypothesis by channel and category. The agreement for each channel and category is within 1 σ of the observed signal strength.

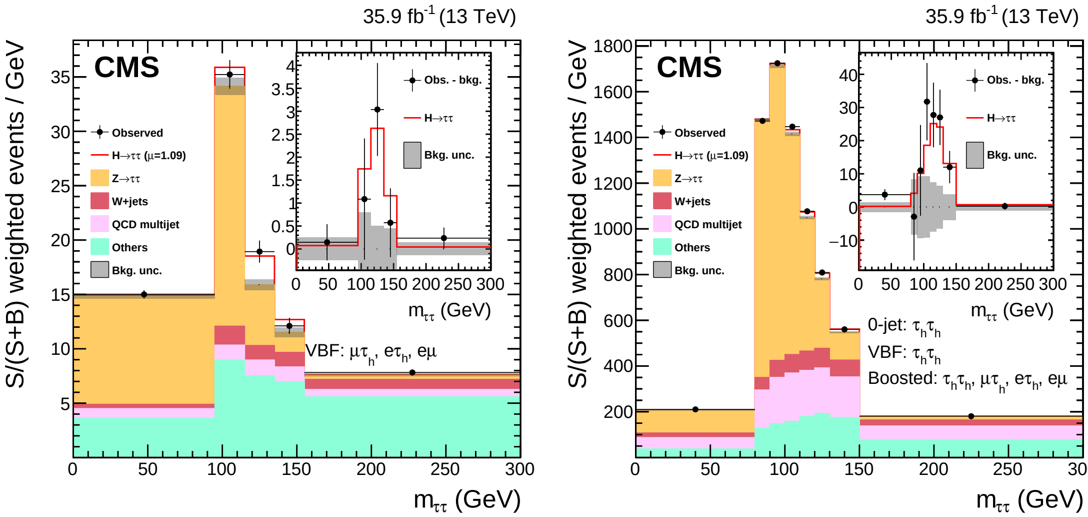


Figure 6.17: Combined observed and predicted $m_{\tau\tau}$ distributions. The left pane includes the VBF category of the $\mu\tau_h$, $e\tau_h$ and $e\mu$ channels, and the right pane includes all other channels that make use of $m_{\tau\tau}$ instead of m_{vis} for the signal strength fit. The normalization of the predicted background distributions corresponds to the result of the global fit, while the signal is normalized to its best fit signal strength. The mass distributions for a constant range of the second dimension of the signal distributions are weighted according to $S/(S+B)$, where S and B are computed as the signal or background contribution in the mass distributions. The "Others" background contribution includes events from diboson, $t\bar{t}$, and single top quark production, as well as Higgs boson decay to a pair of W bosons and Z bosons decaying to a pair of light leptons. The background uncertainty band accounts for all sources of background uncertainty, systematic as well as statistical, after the global fit. The inset shows the corresponding difference between the observed data and expected background distributions, together with the signal expectation. The signal yield is not affected by the reweighting.

6.8.6 Impact of Uncertainties

The uncertainty on the best fit signal strength, μ, can be decomposed into four components: theoretical uncertainties, bin-by-bin statistical uncertainties on the backgrounds, other systematic uncertainties, and the statistical uncertainty of the data gathered. In this format, the best fit signal strength is:

$$\mu = 1.09^{+0.15}_{-0.15}(stat.)^{+0.16}_{-0.15}(syst.)^{+0.10}_{-0.08}(theory)^{+0.13}_{-0.12}(\text{bin-by-bin}).$$

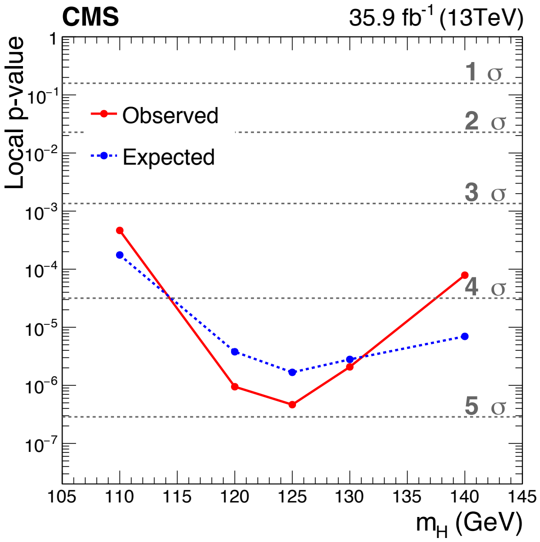


Figure 6.18: Local p-value and significance as a function of the SM Higgs boson mass hypothesis. The observation (red, solid) is compared to the expectation (blue, dashed) for a Higgs boson with a mass $m_H = 125.09\,\text{GeV}$. The background includes Higgs boson decays to pairs of W bosons, with $m_H = 125.09\,\text{GeV}$.

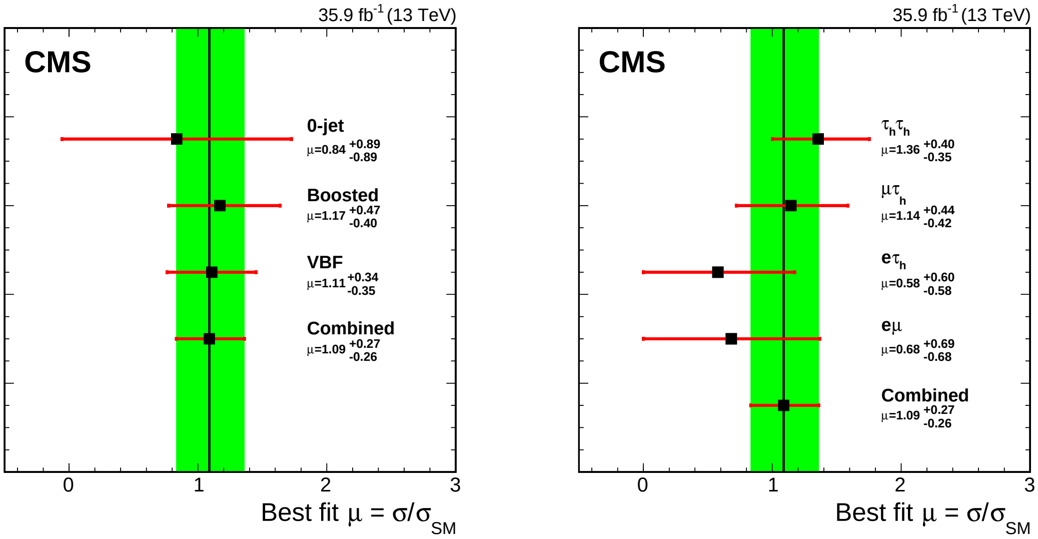


Figure 6.19: Best fit signal strength μ per category (left) and channel (right), for $m_H = 125.09\,\text{GeV}$. The constraints from the global fit are used to extract each of the individual best fit signal strengths. The combined best fit signal strength is $\mu = 1.09^{+0.27}_{-0.26}$.

The impact of these uncertainties on the signal strength was assessed by performing a 1D likelihood scan of the signal strength with multiple uncertainty hypotheses. These scans can be seen in Figure 6.20. In different versions of the likelihood scan certain uncertainty groups were frozen to their best fit values when $\mu = 1.09$. When an uncertainty group is frozen to its best fit value and not allowed to fluctuate in the likelihood scan, the impact of that uncertainty is effectively removed from the likelihood calculation process. The resulting likelihood parabola shows how the results would differ if the frozen uncertainty group did not exist. The difference between the nominal likelihood parabola and one with frozen uncertainties provides a numerical uncertainty value associated with the frozen uncertainty group.

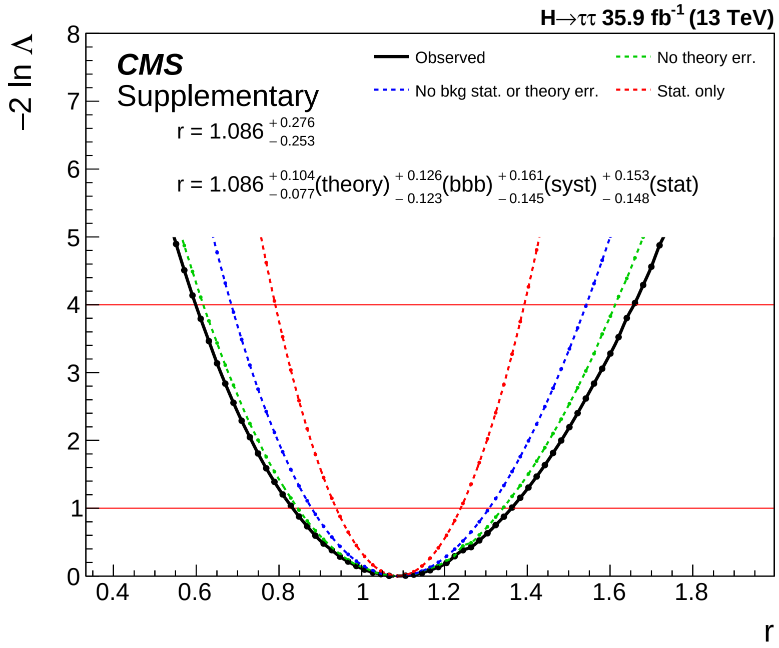


Figure 6.20: 1D likelihood scans which were used to extract the uncertainty associated with each of the four uncertainty components: theoretical uncertainties, bin-by-bin statistical uncertainties on the backgrounds, other systematic uncertainties, and the statistical uncertainty of the data gathered. The difference between the nominal scan in black and the scans represented as dotted lines show the effect of freezing out certain uncertainty components from the analysis. Freezing out an uncertainty component shows how the analysis results would change if there was zero uncertainty associated with that component. For example, the green dotted curve shows how the analysis results would change if theoretical Higgs boson uncertainties were reduced to zero.

Chapter 7

Higgs $\rightarrow \tau\tau$: WH and ZH Associated Production

Associated production of the Higgs Boson with vector bosons, WH and ZH, is suppressed by almost an order of magnitude compared to ggH production. However, these processes, when studying fermionic decays of the Higgs boson, such as $H \rightarrow \tau\tau$, provide simultaneous access to the Higgs boson couplings to fermions and vector bosons in a single measurement. By focusing on the leptonic decays of the vector bosons, the W and Z decay products are well reconstructed, including at the trigger level, resulting in highly suppressed backgrounds. This study is complementary to the previously described ggH and VBF targeted $H \rightarrow \tau\tau$ analysis in Chapter 6 and is performed using data corresponding to $35.9\,fb^{-1}$ of integrated luminosity collected at center-of-mass energy 13 TeV [4]. Combining the results from this associated production targeted analysis with the results of the ggH and VBF targeted analysis [3] we produce an observation of the $H \rightarrow \tau\tau$ process at 13 TeV, measured at the 5.5 σ confidence level.

For the ZH targeted final states, $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ decays are considered combined with four possible $H \to \tau\tau$ final states: $e\tau_h$, $\mu\tau_h$, $e\mu$ and $\tau_h\tau_h$. For the WH targeted final states, four final states are considered with the W boson decaying leptonically to an electron or muon plus a neutrino, and at least one τ_h from the Higgs boson decay: $\mu\mu\tau_h$, $e\mu\tau_h$, $e\tau_h\tau_h$ and $\mu\tau_h\tau_h$.

There are many similarities in the treatment of simulated samples, Monte Carlo corrections, and uncertainties between the ggH and VBF targeted $H \to \tau\tau$ analysis and the associated production targeted $H \to \tau\tau$ analysis. When appropriate, instead of repeating what has already been documented in other chapters, I refer back to the previous sections for full details.

7.1 Overview

This chapter specifically focuses on studying the Higgs boson produced via the associated production mechanisms, WH/ZH. In the following pages the symbol ℓ refers to light leptons (electrons and muons) and τ_h refers to hadronically decaying τ leptons. Leptons refers inclusively to electrons, muons, and τ including their decay products: τ_e, τ_μ, τ_h. We study all possible $\tau\tau$ final state combinations with the exception of two electron and two muon final states because of the low $\tau\tau \to \tau_e\tau_e/\tau_\mu\tau_\mu$ branching fractions. The $H \to \tau\tau$ final states which are studied are the same as those used in the ggH and VBF analysis and are: $\tau_e\tau_h$ ($e\tau_h$), $\tau_\mu\tau_h$ ($\mu\tau_h$), $\tau_e\tau_\mu$ ($e\mu$), and lastly, $\tau_h\tau_h$ ($\tau_h\tau_h$). This combination of $\tau\tau$ final states covers about 94% of all possibilities. We ensure uniqueness between the studied final states be applying veto criteria to events based on the number of reconstructed loosely identified electrons and muons.

7.1.1 Triggers

The W and Z bosons in the WH and ZH final states ensure the presence of one or two well-isolated leptons with sufficiently high p_{T}. Because of this, we use single or double lepton triggers to select events. The WH targeted final states rely on a set of single lepton triggers which must be fired by the W boson lepton. These single lepton triggers are the same single electron and single muon triggers which are used in the ggH and VBF targeted analysis described in Section 6.1.2. In general, the W boson lepton will have a higher p_{T} than the Higgs boson leptons, providing higher efficiency for trigger selection. Requiring that the W boson lepton fires the trigger ensures that there is negligible isolation or identification bias in the selection of the Higgs boson leptons from the trigger requirements. The lack of bias in the Higgs boson lepton selection allows us to use specific background estimation techniques.

Double electron and double muon triggers are used in the ZH targeted final states to trigger on the Z decay products. The presence of two leptons in the double lepton triggers allows for lower p_{T} thresholds online, which increases the acceptance of ZH events after offline selections are applied. Similar to the WH final states, triggering on the vector boson associated leptons removes selection bias from the Higgs boson associate leptons and enables the use of specific background estimation methods (Section 7.5). Additionally, ZH events can be selected using single lepton triggers applied to either of the Z leptons. This helps increase the overall trigger selection efficiency.

The trigger selection criteria for the WH and ZH targeted final states is detailed in Table 7.1.

WH trigger selection requirements

Final State	Trigger (p_{T}/η)	Lepton Selection: p_{T}
$e\mu\tau_h$	$\mu(22/2.1)$ or $e(25/2.1)$	$p_{\mathrm{T}}^{e} > 15, p_{\mathrm{T}}^{\mu} > 23$ or $p_{\mathrm{T}}^{e} > 26, p_{\mathrm{T}}^{\mu} > 15$
$\mu\mu\tau_h$	$\mu(22/2.1)$	$p_{\mathrm{T}}^{\mu} > 23, p_{\mathrm{T}}^{\mu} > 15$
$e\tau_h\tau_h$	$e(25/2.1)$	$p_{\mathrm{T}}^{e} > 26$
$\mu\tau_h\tau_h$	$\mu(22/2.1)$	$p_{\mathrm{T}}^{\mu} > 23$

ZH trigger selection requirements

Final State	Trigger (p_{T}/η)	Lepton Selection: p_{T}
$ee\mu\tau_h$		
$eee\tau_h$	$e(23/2.5)$ & $e(12/2.5)$,	$p_{\mathrm{T}}^{e} > 24$ & $p_{\mathrm{T}}^{e} > 13$,
$ee\tau_h\tau_h$	or $e(27/2.5)$	or $p_{\mathrm{T}}^{e} > 28$
$eee\mu$		
$\mu\mu\mu\tau_h$		
$\mu\mu e\tau_h$	$\mu(17/2.4)$ & $\mu(8/2.4)$,	$p_{\mathrm{T}}^{\mu} > 18$ & $p_{\mathrm{T}}^{\mu} > 10$,
$\mu\mu\tau_h\tau_h$	or $\mu(24/2.4)$	or $p_{\mathrm{T}}^{\mu} > 25$
$\mu\mu e\mu$		

Table 7.1: Kinematic selection requirements for WH and ZH events. The trigger requirement is defined by a combination of trigger candidates with p_{T} over a given threshold (in GeV), indicated inside parentheses. The pseudorapidity thresholds come from trigger and object reconstruction constraints. The trigger requirements for the ZH events are defined by the Z boson decay products, either $Z \to ee$ or $Z \to \mu\mu$.

7.1.2 Event Selection

In each final state, the event selection criteria was chosen to reject background events while preserving as many signal events as possible. For continuous variables such as p_{T}, the thresholds were optimized by scanning along the range of possible thresholds and maximizing the signal sensitivity. For discrete variables such as the electron, muon, and τ_h identification working points, the different working points were tested in the optimization process [4].

In the semileptonic WH associated production final states, $e\mu\tau_h$ and $\mu\mu\tau_h$, the two light leptons are required to have the same charge to reduce the $t\bar{t}$ and $Z+$ jets backgrounds where one or more jets is misidentified as a τ_h candidate. The τ_h

candidate has opposite charge to the light leptons. The leading (highest p_{T}) light lepton is considered as coming from the W boson, while the Higgs boson candidate is formed from the τ_h and the subleading (lowest p_{T}) light lepton. The correct pairing is achieved in about 75% of events. The leading light lepton is required to fire the single lepton triggers and to have a p_{T} that is 1 GeV above the online thresholds, whereas the subleading light lepton has $p_{\mathrm{T}} > 15\,\mathrm{GeV}$; this p_{T} threshold is the result of optimizing the selection for maximum signal significance. Based on optimizing for signal sensitivity, selection criteria based on three variables improve the sensitivity of the results in both final states [4]:

- $L_{\mathrm{T}} > 100\,\mathrm{GeV}$, where L_{T} is the scalar p_{T} sum of the three leptons in the final state
- $|\Delta\phi(\ell_1, H)| > 2.0$, where ℓ_1 is the leading light lepton, and H is the system formed by the subleading light lepton and the τ_h candidate
- $|\Delta\eta(\ell_1, H)| < 2.0$.

In the hadronic WH associated production final states, $e\tau_h\tau_h$ and $\mu\tau_h\tau_h$, the τ_h candidates are both assumed to be from the Higgs boson decay, and thus are required to have opposite charge. Based on optimizing for signal significance, the τ_h that has the same charge as the light lepton must have $p_{\mathrm{T}} > 35\,\mathrm{GeV}$. This is driven by the fact that the τ_h that has the same charge as the light lepton is almost always a jet misidentified as a τ_h candidate, and the jet misidentification rate strongly decreases with p_{T}. The subleading τ_h must have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$. Selection criteria based on three variables have been found to increase the sensitivity of the results in both final states:

- $L_{\mathrm{T}} > 130\,\mathrm{GeV}$, where L_{T} is the scalar p_{T} sum of the three leptons in the final state

- $|\vec{S_{\mathrm{T}}}| < 70\,\mathrm{GeV}$, where $\vec{S_{\mathrm{T}}}$ is the vectorial p_{T} sum of the three leptons in the final state and of $\vec{E}_T^{miss}$
- $|\Delta\eta(\tau_h, \tau_h)| < 2.0$.

In the ZH final states, the Z boson is reconstructed from the opposite charge, same-flavor light lepton combination which has a mass closest to the Z boson mass. Different electron and muon identification and isolation criteria are used for the leptons assigned to the Z boson than those assigned to the Higgs boson. A looser selection is applied for the Z boson leptons to increase signal acceptance. The rate of misidentification for these leptons is relatively low because of the required Z mass window cut, $60\,\mathrm{GeV} < m_{\ell\ell} < 120\,\mathrm{GeV}$, and the opposite charge criteria. In comparison, a tighter selection is applied to the leptons assigned to the Higgs boson to decrease the background contributions from $Z+$jets and other reducible backgrounds. The specific selections are detailed in Table 7.2, including those for the τ_h candidates. All identification and isolation criteria where chosen based on optimizing for the best signal sensitivity.

The signal region is split into a High-$L_{\mathrm{T}}^{\mathrm{Higgs}}$ and Low-$L_{\mathrm{T}}^{\mathrm{Higgs}}$ region; $L_{\mathrm{T}}^{\mathrm{Higgs}}$ is defined as the scalar p_{T} sum of the decay products of the Higgs boson. This splitting helps separate the $Z+$ jets background from the ZH signal which is concentrated in the High-$L_{\mathrm{T}}^{\mathrm{Higgs}}$ region. The different τ decay processes define the kinematics of the Higgs boson and the $L_{\mathrm{T}}^{\mathrm{Higgs}}$. Therefore, the $L_{\mathrm{T}}^{\mathrm{Higgs}}$ regions are defined based on the specific $H \to \tau\tau$ final states of an event. The split between the High- and Low-$L_{\mathrm{T}}^{\mathrm{Higgs}}$ regions are:

- $\ell\ell e\tau_h$: $L_{\mathrm{T}}^{\mathrm{Higgs}} = 60\,\mathrm{GeV}$
- $\ell\ell\mu\tau_h$: $L_{\mathrm{T}}^{\mathrm{Higgs}} = 60\,\mathrm{GeV}$

- $\ell\ell\tau_h\tau_h$: $L_{\mathrm{T}}^{\mathrm{Higgs}} = 75\,\mathrm{GeV}$
- $\ell\ell e\mu$: $L_{\mathrm{T}}^{\mathrm{Higgs}} = 50\,\mathrm{GeV}$

For convenience, the High- and Low-$L_{\mathrm{T}}^{\mathrm{Higgs}}$ regions are plotted side-by-side, see Figures 7.5, 7.6, and 7.7.

7.1.3 Baseline Object Selection

The expression p_{T}^{ℓ} stands for the p_{T} of the lepton. The electron and muon isolation requirements used in this analysis, based on I^{ℓ} (Section 5.1), are listed in Table 7.2. As stated above, all identification and isolation criteria where chosen based on optimizing for the best signal sensitivity.

The τ_h candidates are identified using the MVA discriminant discussed in Section 5.3.3. Three τ_h MVA working points are used in this analysis, `Very Tight Tau MVA`, `Tight Tau MVA`, and `Medium Tau MVA` ID. Each working point was selected based on optimizing the analysis for highest sensitivity to the associated production $H \rightarrow \tau\tau$ processes. In the lower statistics ZH final states the higher efficiency `Medium Tau MVA` is used. In the hadronic WH final states a combination of two working points is used. The τ_h that has same charge as the light lepton, which is likely to be a fake, has to pass the `Very Tight Tau MVA` working point. The other τ_h that has opposite charge to the light lepton is less likely to be a fake and must only pass `Medium Tau MVA`. In the semileptonic final states the `Tight Tau MVA` working points is used.

For each final state there are additional τ_h requirements which help suppress electron to τ_h and muon to τ_h misidentification. The exact working points used are tuned for each final state to suppress dominant misidentified backgrounds. These discriminants are denoted in this analysis as anti-e and anti-μ criteria and are discussed in

Section 5.3.3. The discriminants have a range of thresholds with working points being referred to as “Very Loose” (VL), “Loose” (L), “Medium” (M), and “Tight” (T).

A summary of the lepton selection details for each final state in the associated production analysis is in Table 7.2. All reconstructed leptons in the events are required to be separated from each other by $\Delta R > 0.3$. In the case of τ_h, they are required to be separated from all other leptons by $\Delta R > 0.5$. The resulting event samples are made mutually exclusive by discarding events that have additional loosely identified and isolated muons or electrons.

We reject events which have been tagged as likely including heavy flavor jet decays from b-quarks. The working point chosen gives an efficiency for identifying real b jets of about 70% for about 1% of light flavor or quark jets being misidentified, Section 5.3.2. In the ZH final states, this selection removes roughly 13% of the background events at a cost of only 2% of the signal events increasing the purity of the signal region.

7.2 Data Set

The associated production targeted $H \to \tau\tau$ study utilizes the same dataset as the ggH and VBF targeted study, the full 2016 pp dataset collected by CMS corresponding to $35.9\,fb^{-1}$ of integrated luminosity. The data were gathered at center-of-mass energy 13 TeV. For further details see Section 6.2.

7.3 Monte Carlo Samples

Signal and background processes are modeled with samples of simulated events. For details on the production of simulated events, see Section 4. For each simulated event,

WH selection requirements

τ_h baseline req: $p_T^{\tau_h} > 20$, $|\eta| < 2.3$, anti-e VL, anti-μ L

e baseline req: $p_T^e > 10$, $|\eta| < 2.5$, $I^e < 0.10$, e MVA Tight ID

μ baseline req: $p_T^\mu > 10$, $|\eta| < 2.4$, $I^\mu < 0.15$, Medium ID

Final State	Additional τ_h Criteria
$e\mu\tau_h$	`Tight Tau MVA`, anti-e VL/T, anti-μ T/L
$\mu\mu\tau_h$	`Tight Tau MVA`, anti-μ T
$e\tau_h\tau_h$	`Medium/Very Tight Tau MVA`, anti-e T
$\mu\tau_h\tau_h$	`Medium/Very Tight Tau MVA`, anti-μ T

ZH selection requirements

Z boson: opposite charge, same-flavor light leptons, $60\,\text{GeV} < m_{\ell\ell} < 120\,\text{GeV}$

τ_h baseline req: $p_T^{\tau_h} > 20$, $|\eta| < 2.3$, `Medium Tau MVA`

e baseline req: $p_T^e > 10$, $|\eta| < 2.5$, MVA Loose ID

μ baseline req: $p_T^\mu > 10$, $|\eta| < 2.4$, Loose ID, $I^\mu < 0.25$

Final State	Additional Higgs Boson Lepton Criteria
$ee\mu\tau_h$	$I^\mu < 0.15$
$eee\tau_h$	e Tight MVA ID, $I^e < 0.15$
$ee\tau_h\tau_h$	baseline selection
$eee\mu$	e Tight MVA ID, $I^e < 0.15$, $I^\mu < 0.15$
$\mu\mu\mu\tau_h$	$I^\mu < 0.15$
$\mu\mu e\tau_h$	e Tight MVA ID, $I^e < 0.15$
$\mu\mu\tau_h\tau_h$	baseline selection
$\mu\mu e\mu$	e Tight MVA ID, $I^e < 0.15$, $I^\mu < 0.15$

Table 7.2: Electron, muon, and τ_h selection criteria for each final state in the associated production $H \to \tau\tau$ analysis. In the $e\mu\tau_h$ final state there are two different working points listed for electron and muon rejection. Anti-e VL, anti-μ T applies for events where the electron and τ_h are same charge and anti-e T, anti-μ L applies for events where the muon and τ_h are same charge.

a number of additional pileup interactions is simulated and added. The number of pileup interactions added is based on best efforts to match the simulated events to the pileup in data which is estimated from the measured instantaneous luminosity for each bunch crossing. The average number of additional pileup interactions in the 2016 CMS data is approximately 27 interactions per bunch crossing.

7.4 Mass Reconstruction

The visible mass of the $\tau\tau$ system, m_{vis}, can be used to separate the $H \to \tau\tau$ signal events from the large contribution of irreducible $Z \to \tau\tau$ events. However, the neutrinos from the τ lepton decays carry a large fraction of the τ lepton energy and reduce the discriminating power of this variable. The SVFIT algorithm used in the ggH and VBF targeted analysis, discussed in Section 6.4, is used in the ZH final states. It combines the $\vec{E}_T^{miss}$ with the four-vectors of both τ candidates to calculate a more accurate estimate of the mass of the parent boson and is denoted as $m_{\tau\tau}$. The $m_{\tau\tau}$ variable is used in the ZH final state where the majority of $\vec{E}_T^{miss}$ is attributable to the τ decays. The m_{vis} is used in the WH final state because the SVFIT algorithm is not designed to account for the additional $\vec{E}_T^{miss}$ associated with the neutrino from the W boson decay [130].

7.5 Background Estimation

The simulated background processes are all scaled to their NLO cross section in Table 7.3.

The irreducible backgrounds for the associated production analysis can be split into those with four lepton final states for ZH and those with three lepton final states for WH. When a lepton escapes the fiducial volume of the detector, or is otherwise poorly reconstructed, the four lepton irreducible backgrounds can populate the WH three lepton final states. Backgrounds typically composing the irreducible background for the ZH final states are: ZZ, $t\bar{t}Z$, WWZ, WZZ, and ZZZ. The dominant irreducible backgrounds for the WH final states are: WZ and $t\bar{t}W$. The irreducible backgrounds are estimated from simulation and scaled to their theoretical cross section. Higgs boson decays to pairs of W or Z bosons are also estimated from

Background Process	Cross section (pb)
Z+jets Inclusive Jet Production	5747
$t\bar{t}$	831.8
EWK $WZ \to 3\ell\nu$	4.708
$ZZ \to 4\ell$	1.212
$gg \to ZZ \to 2\ell 2\ell$	0.005423
$gg \to ZZ \to 4\ell$	0.002703
$t\bar{t}Z + \text{jets}$	0.2529
WWW	0.2086
WWZ	0.1651
WZZ	0.05565
ZZZ	0.01398
ZH, $H \to WW$	0.02005
$W^{-}H$, $H \to WW$	0.01209
$W^{+}H$, $H \to WW$	0.01905
$gg \to H \to WW \to 2\ell 2\nu$	1.001
VBF $H \to WW \to 2\ell 2\nu$	0.08900
$gg \to H \to ZZ$	0.0121
$t\bar{t}H \to$ non-$b\bar{b}$	0.215

Table 7.3: NLO cross sections for considered backgrounds. In this table, ℓ represents all three generations of charged leptons, e,μ,τ. In some cases the production mechanism is listed: quarks (qq) versus gluons (gg).

simulation and considered as background processes. Additionally, the $t\bar{t}H$ production process with all Higgs boson decay paths is estimated from simulation and considered a background processes.

The reducible backgrounds, which have at least one jet misidentified as an electron, muon, or τ_h lepton, are estimated from data. Data events meeting specific requirements detailed below are reweighted as a function of a misidentification rate to estimate the contribution of these processes in the signal region.

In the WH final states, the misidentification rate of jets as electrons, muons, or τ_h candidates is measured in $Z + \text{jets}$ events [4]. The Z boson is reconstructed in its dielectron decay mode for measuring the jet to muon misidentification rate, and is reconstructed in its dimuon decay mode for measuring the jet to electron or τ_h misidentification rate. The rates are measured in bins of the lepton p_{T}, and are split between reconstructed decay mode for the τ_h candidates.

In the semileptonic $e\mu\tau_h$ and $\mu\mu\tau_h$ final states, events are selected for reweighting if they pass the full signal region selection except that the subleading light lepton or the τ_h do not pass the isolation or identification conditions. To remove the overlap between this method and the simulated samples, events in simulation that have a jet that is misidentified as the τ_h or as the subleading lepton, are discarded. Simulated events that have a jet misidentified as the leading W boson lepton, but two real leptons for the Higgs boson leptons, are estimated from simulation as their contribution is not taken into account with the misidentification rate method described above. These events mostly arise from $t\bar{t}$ and $Z+\text{jets}$ processes, and account for a small fraction of the total expected background in the signal region. In the hadronic $e\tau_h\tau_h$ and $\mu\tau_h\tau_h$ final states, the method is essentially the same, except that the lepton most susceptible to being misidentified, thus having the misidentification estimation method applied to it, is the τ_h candidate that has the same charge as the light lepton.

In the ZH final states, a similar misidentification rate is used to estimate the contribution of jets misidentified as electrons, muons, or τ_h in signal region events. The misidentification rate is measured in four lepton final states which is dominated by $Z+\text{jets}$ events with a small contribution from $t\bar{t}$ events [4]. Identical to the WH final states, the rates are measured in bins of the lepton p_{T}, and are split between reconstructed decay modes for the τ_h candidates.

In the four lepton final states, it is more likely that a jet will be misidentified as one of the leptons resulting from the $H \to \tau\tau$ decay compared to the $Z \to \ell\ell$ leptons. In the ZH final states, data events which pass the full signal region selection, except either or both of the Higgs boson leptons fail identification or isolation criteria, are weighted by the misidentification rate for the failing lepton. To avoid double counting events with misidentified leptons, events with both Higgs boson leptons failing have their weight subtracted from the events which only have a single lepton failing. To remove the overlap between this method and the simulated samples, events in simulation that have a jet that is misidentified as an electron, muon, or τ_h are discarded. This misidentification rate method is used to estimate the yield of the reducible backgrounds.

The shape of the reducible background contribution is taken from data in a signal-free region with same charge Higgs boson leptons. The contribution of irreducible backgrounds in the same charge region used to derive the shape template is less than 1%. This means that the data derived template is a very pure reducible background selection. A high statistics, relaxed identification and isolation selection is used to reduce statistical uncertainties for the shape template. The relaxed section is consistent with Table 7.2 except for the identification and isolation criteria listed below:

- electron: MVA Loose ID
- muon: Loose ID, $I^{\mu} < 5.0$
- τ_h: raw Tau MVA score > -0.8 (Reference [125])

Kolmogorov-Smirnov (KS) tests have been performed to validate shape compatibility between this relaxed same charge selection and the signal region reducible background distributions [4]. The KS tests indicate there is likely no shape bias.

A high level of agreement is seen in Figures 7.1, 7.2, 7.3, and 7.4 of the p_{T} distributions for the four leptons per ZH final state when using the simulated events for the irreducible background estimation and the misidentification method for the reducible jet fake backgrounds. The events estimated from these misidentification methods are labeled as “jet fakes” in the following distributions.

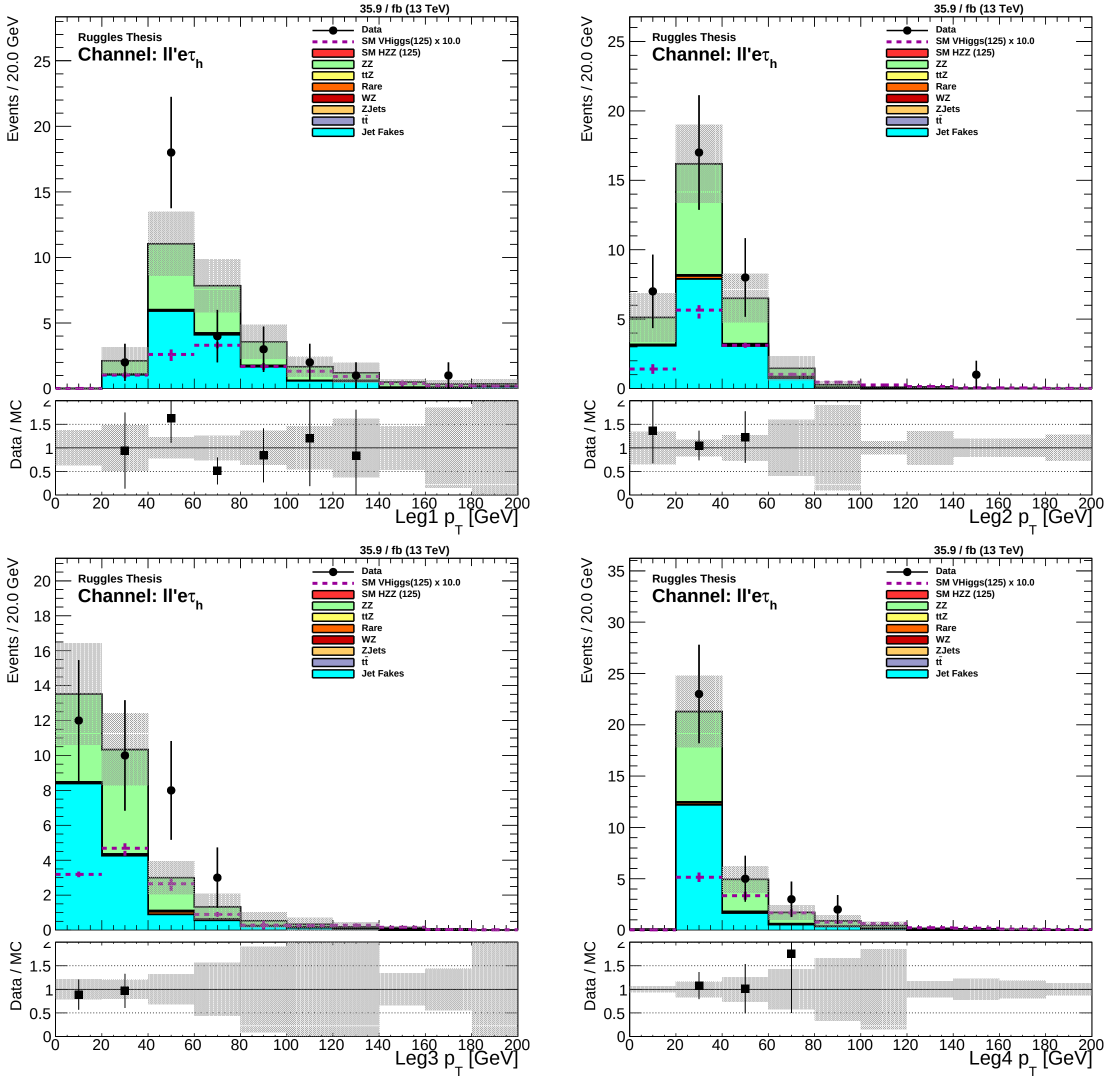


Figure 7.1: Pre-fit p_{T} distributions showing statistical uncertainties only for the four leptons in the $\ell\ell e\tau_h$ final states. (top left) and (top right) leading and subleading ℓ from Z, (bottom left) and (bottom right) e and τ_h from Higgs boson candidate. The WH and ZH signals are summed as $VHiggs$ and multiplied by a factor of 10 times their SM expected yield.

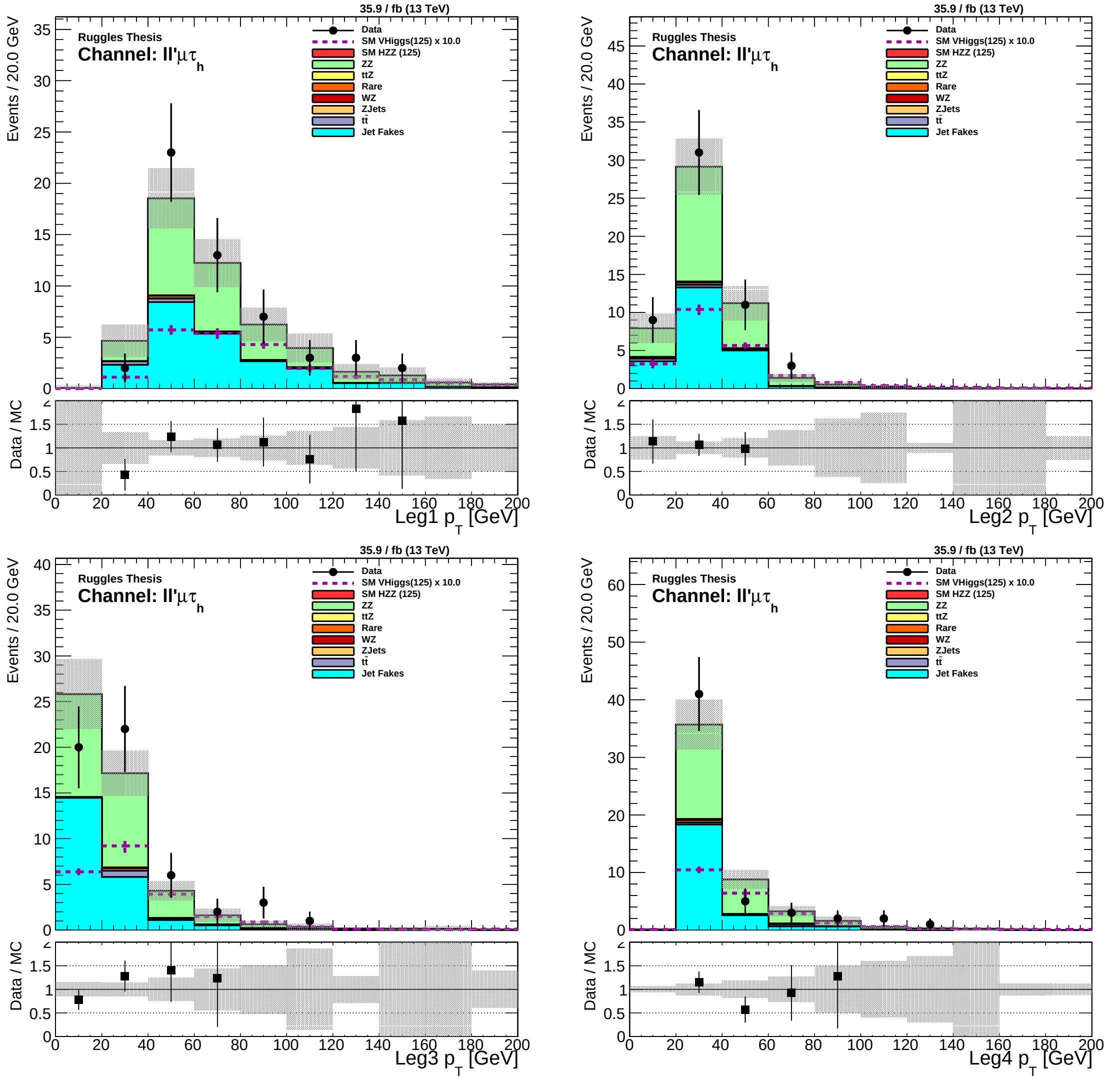


Figure 7.2: Pre-fit p_{T} distributions showing statistical uncertainties only for the four leptons in the $\ell\ell\mu\tau_h$ final states. (top left) and (top right) leading and subleading ℓ from Z, (bottom left) and (bottom right) μ and τ_h from Higgs boson candidate. The WH and ZH signals are summed as $VHiggs$ and multiplied by a factor of 10 times their SM expected yield.

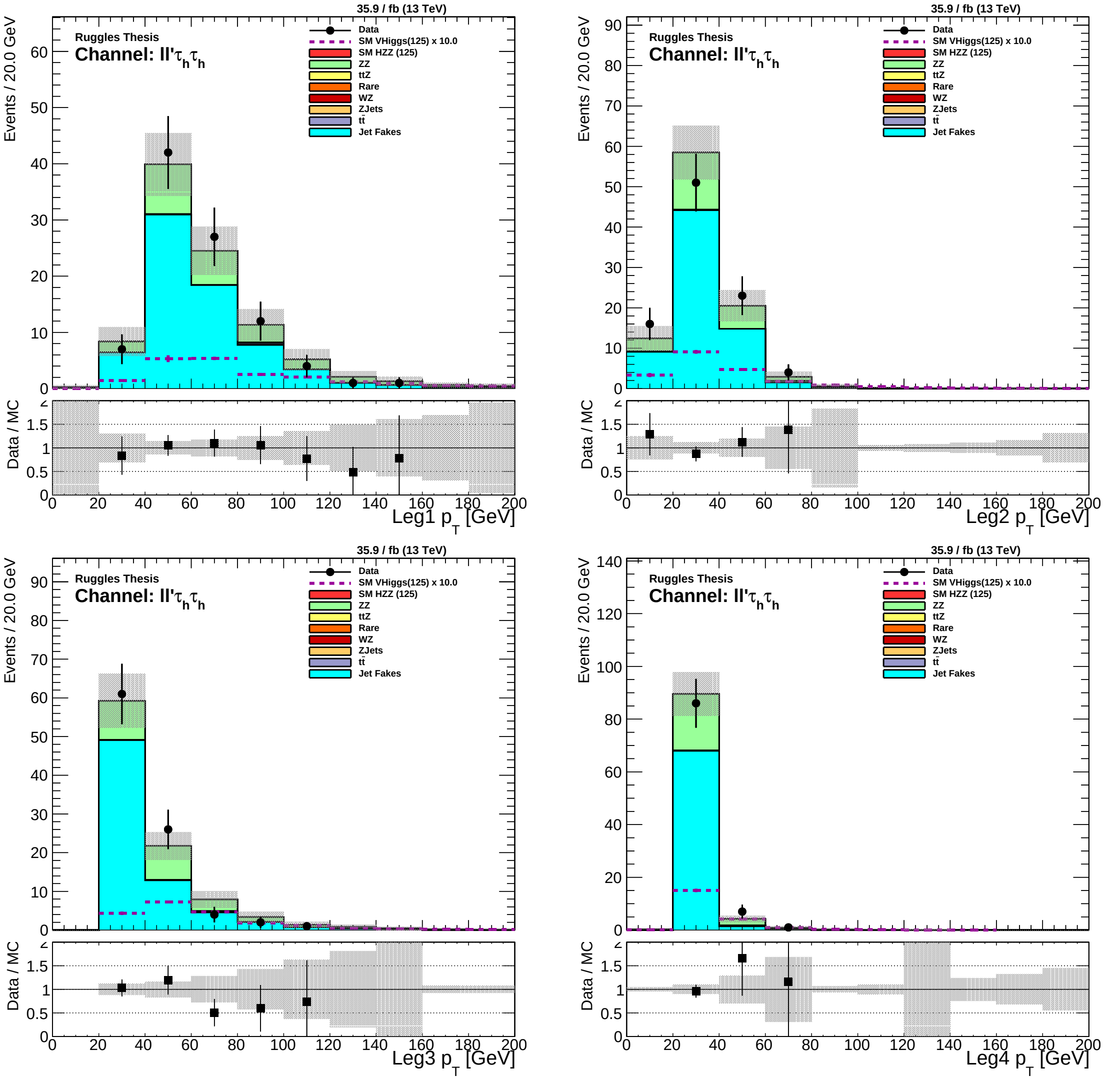


Figure 7.3: Pre-fit p_{T} distributions showing statistical uncertainties only for the four leptons in the $\ell\ell\tau_h\tau_h$ final states. (top left) and (top right) leading and subleading ℓ from Z, (bottom left) and (bottom right) leading and subleading τ_h from Higgs boson candidate. The WH and ZH signals are summed as $VHiggs$ and multiplied by a factor of 10 times their SM expected yield.

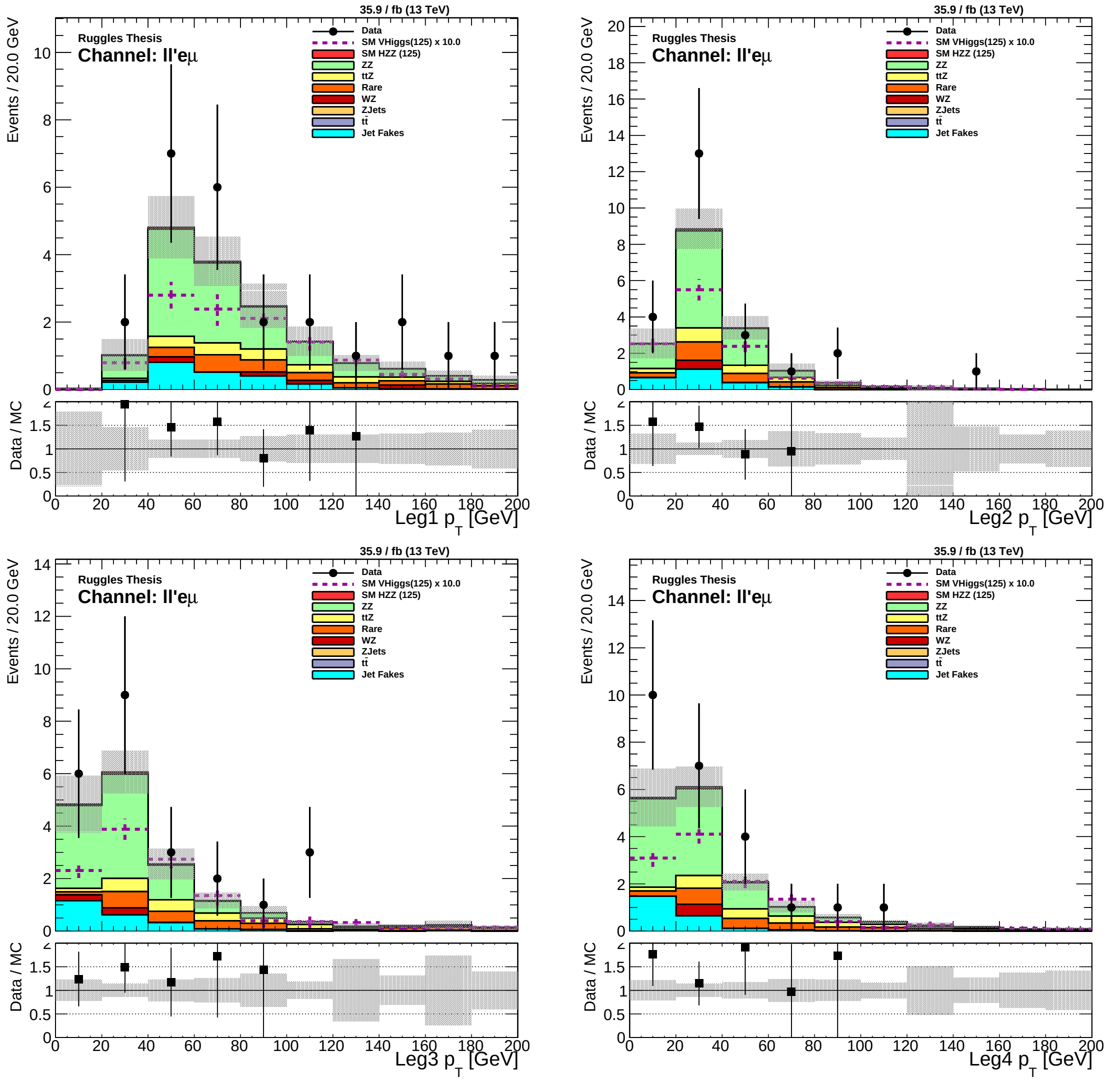


Figure 7.4: Pre-fit p_{T} distributions showing statistical uncertainties only for the four leptons in the $\ell\ell e\mu$ final states. (top left) and (top right) leading and subleading ℓ from Z, (bottom left) and (bottom right) e and μ from Higgs boson candidate. The WH and ZH signals are summed as $VHiggs$ and multiplied by a factor of 10 times their SM expected yield.

7.6 Monte Carlo Corrections

Corrections are applied to the simulated Monte Carlo samples to help correct for measured differences between observed data and expectations based on simulation. Many of these corrections are designed to correct differences in reconstruction and identification efficiencies for leptons between data and simulation. These corrections are derived in fully orthogonal regions from the associated production analysis signal regions. The Monte Carlo corrections applied in the associated production analysis are identical to those applied in the ggH and VBF targeted analysis when applicable, see Section 6.6 for details.

7.7 Systematic Uncertainties

The systematic uncertainty model used for the associated production analysis share many similarities with the uncertainty model used for the ggH and VBF targeted analysis considering the many shared object definitions, simulated backgrounds, and simulated $H \to \tau\tau$ signal samples. Specific nuisances with identical treatment are:

- Uncertainty of the τ_h identification efficiency for genuine τ_h
- Uncertainty on the visible energy scale of genuine τ_h leptons
- Uncertainties in the muon and electron identification, isolation, and trigger efficiencies
- Uncertainty related to discarding events with a b-tagged jet
- $\vec{E}_T^{miss}$ scale uncertainties; these uncertainties are skipped for the WH semileptonic final states where $\vec{E}_T^{miss}$ is not used
- Uncertainty on the finite number of simulated events

- Uncertainty in the integrated luminosity

The full details of how these identical nuisances are treated can be found in Section 6.7.

7.7.1 Simulated Background Estimation Uncertainties

Uncertainties from the renormalization and the factorization scales, and from the choice of the PDF set (Section 4.2), are taken into account for the ZZ and WZ background processes. The uncertainty from the renormalization and factorization scales is determined by varying these scales between 0.5 and 2 times their nominal value and computing the change in acceptance. This leads to yield uncertainties of $^{+3.2\%}_{-4.2\%}$ for the qq $\rightarrow ZZ$ background, and $\pm3.2\%$ for the WZ process. The uncertainty from the PDF set is determined following the PDF4LHC recommendations [145], and leads to yield uncertainties of $^{+3.1\%}_{-4.2\%}$ for the qq $\rightarrow ZZ$ background, and $\pm4.5\%$ for the WZ process. In addition, a 10% uncertainty in the k-factor used for the gg $\rightarrow ZZ$ prediction is considered. The uncertainty in the cross section of the rare $t\bar{t}W$ and $t\bar{t}Z$ processes amounts to 25%.

7.7.2 Reducible Background Estimation Uncertainties

The reducible backgrounds are estimated using the measured rates for jets to be misidentified as electrons, muons, or τ_h discussed in Section 7.5. The misidentification rates are measured in different bins of lepton p_{T}, and are further split between reconstructed decay modes for the τ_h. In the WH final states where the shape of the reducible background is estimated using the misidentification rate method, the statistical uncertainty in every bin is considered as an independent uncertainty, which is propagated to the mass distributions and to the yields of the reducible background estimate. Rate uncertainties applied in the ZH final states cover the possible fluctuation in rate from these types of uncertainties. Additionally, the shape is taken from the same charge region and has previously been validated as compatible (Section 7.5) thus no specific uncertainty is applied to the shape selection.

In both the WH and ZH final states, an additional uncertainty on the misidentification rates due to mismodeling of simulated samples is incorporated. The yields of the simulated prompt contributions, which are subtracted from data, are adjusted within their uncertainty and the effect is propagated forward. This creates a set of misidentification rates corresponding to an upwards shift in the normalization of the prompt simulated events as well as a downwards shifted set. These shifted misidentification rates are then used to estimate the reducible background yield and mass distributions corresponding to this 1σ shift in the prompt simulated background normalization.

In the WH final states, an additional uncertainty comes from potentially different misidentification rates in Z+jets events, where the rates are measured, and in W+jets or $t\bar{t}$ events, which constitute a large fraction of the reducible background in the signal region. To cover this, a 20% yield uncertainty for the reducible background is applied in each WH final state. In the ZH final states a similar uncertainty is applied based on potential differences in the measurement region versus the application region. These uncertainties range from 26% in the $\ell\ell\mu\tau_h$ final states to 100% in the $\ell\ell e\mu$ final states. The large uncertainty in the $\ell\ell e\mu$ final states results from the very low expected reducible background yields, which makes any comparison of the method susceptible to large statistical fluctuations.

7.7.3 Theoretical Uncertainties for Higgs Boson

The rate and acceptance uncertainties for the signal processes related to the theoretical calculations are due to uncertainties in the PDFs, variations of the QCD renormalization and factorization scales, and uncertainties in the modeling of parton showers. The magnitude of the rate uncertainty depends on the production process. The rate uncertainties are found to be statistically insignificant when measuring the rate of Higgs boson production. The inclusive uncertainty related to the PDFs amounts to 1.9 and 1.6%, respectively, for the WH, and ZH production modes [42]. The corresponding uncertainty for the variation of the renormalization and factorization scales is 0.7 and 3.8%, respectively [42].

The systematic uncertainties considered in the associated production targeted analysis are summarized in Table 7.4.

Source of uncertainty	Magnitude
τ_h energy scale	1.2% in energy scale
e energy scale	1–2.5% in energy scale
$\vec{E}_T^{miss}$ energy scale	Dependent upon p_{T} and η
τ_h ID & isolation	5% per τ_h
e ID & isolation & trigger	2%
μ ID & isolation & trigger	2%
Diboson normalization	5%
Integrated luminosity	2.5%
b-tagged jet rejection	4.5% heavy flavor, 0.15% light flavor
Limited number of events	Statistical uncertainty in individual bins
Signal theoretical uncertainty	Up to 20%
Reducible background uncertainties	WH: shape and yield based WH: 20% yield ZH: 26–100% yield

Table 7.4: Sources of systematic uncertainty

7.8 Results

The extraction of the results uses a global maximum likelihood fit based on a simultaneously fit of all the WH and ZH final state signal regions. Section 7.8.1 shows the twelve signal region distributions used in the global maximum likelihood fit. The global maximum likelihood fit results in a best fit signal strength for this dedicated WH and ZH associated production analysis of $\mu = 2.5^{+1.4}_{-1.3}$. This corresponds to a significance of 2.3 standard deviations while a significance of 1.0 standard deviation was expected.

7.8.1 Signal Region Details

In the ZH final states, the $m_{\tau\tau}$ distribution is used for signal extraction. The Low-$L_{\mathrm{T}}^{\mathrm{Higgs}}$ and High-$L_{\mathrm{T}}^{\mathrm{Higgs}}$ regions are plotted side-by-side in the following distributions. Figures 7.5, 7.6, and 7.7 show the $m_{\tau\tau}$ distributions for each of the ZH final states and the combined distribution for all eight ZH final states summed together. The eight ZH final states are each fit separately in the global fit; combining them together helps reduce statistical fluctuations for visualization purposes only. The distributions are post-fit and show full uncertainties. The WH and ZH signals are shown as 5x larger than their best-fit value $\mu = 2.5$.

The results in the WH final states are obtained from the distributions of the visible mass of the τ_h leptons in the hadronic $\ell\tau_h\tau_h$ final states, and of the visible mass of the τ_h and subleading light lepton in the semileptonic $\ell\ell\tau_h$ final states. The mass distributions are shown in Figure 7.8 for the four WH final states. Figure 7.9 shows all four WH final states combined together for visualization purposes only.

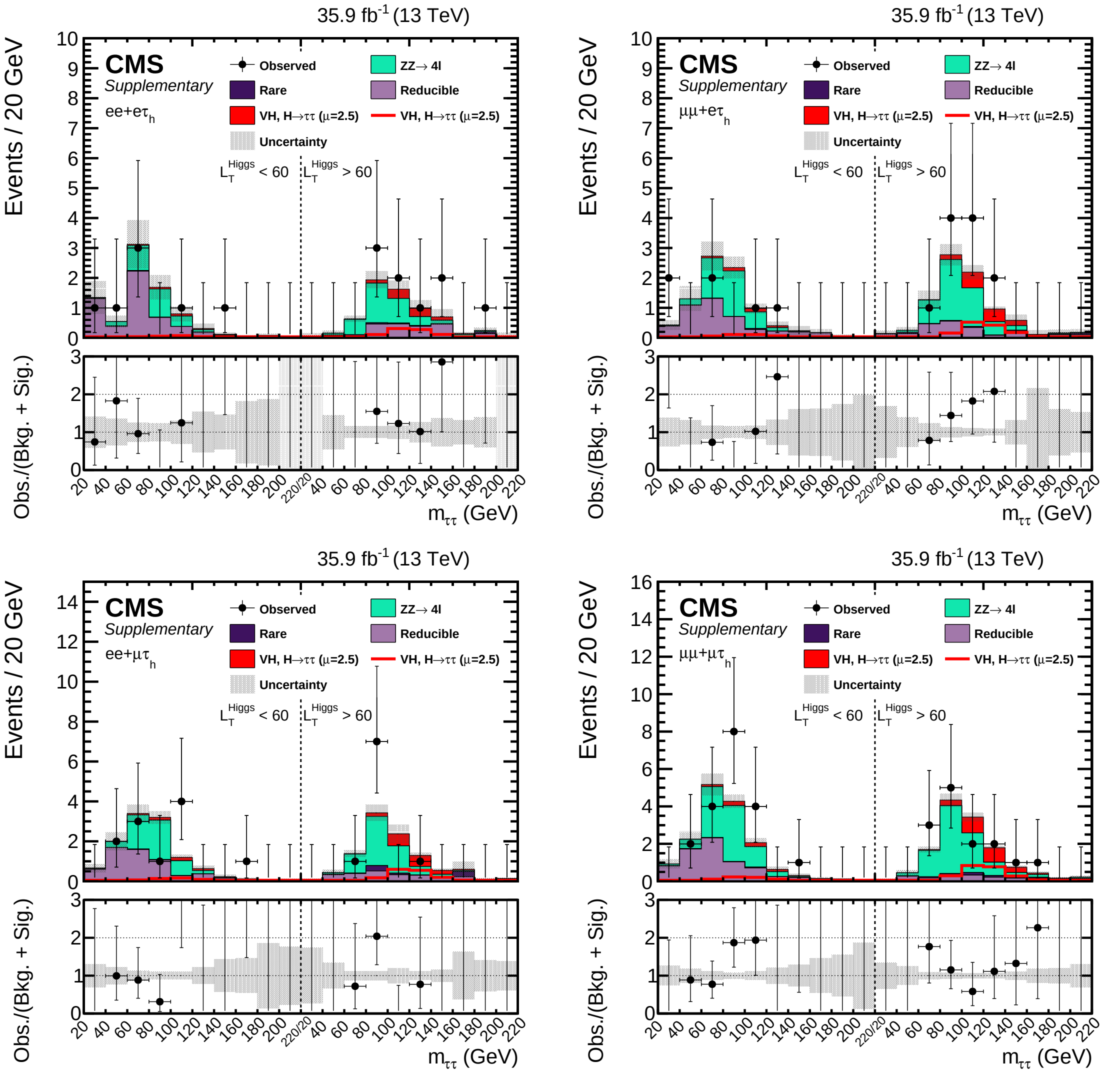


Figure 7.5: The postfit $m_{\tau\tau}$ distributions used to extract the signal shown for the (top left) $eee\tau_h$, (top right) $\mu\mu e\tau_h$, (bottom left) $ee\mu\tau_h$, and (bottom right) $\mu\mu\mu\tau_h$ final states. The distributions show full uncertainties. The WH and ZH, $H \to \tau\tau$ signal processes are summed together and shown as VH, $H \to \tau\tau$ with a best-fit $\mu = 2.5$. VH, $H \to \tau\tau$ is shown both as a stacked filled histogram and an open overlaid histogram. In these distributions the ZH, $H \to \tau\tau$ process contributes more than 99% of the total of VH, $H \to \tau\tau$.

7.8.2 Analysis Sensitivity Details

An excess of observed events with respect to the SM background expectation is visible in the most sensitive bins of the analysis, Figure 7.10. This distribution is created by

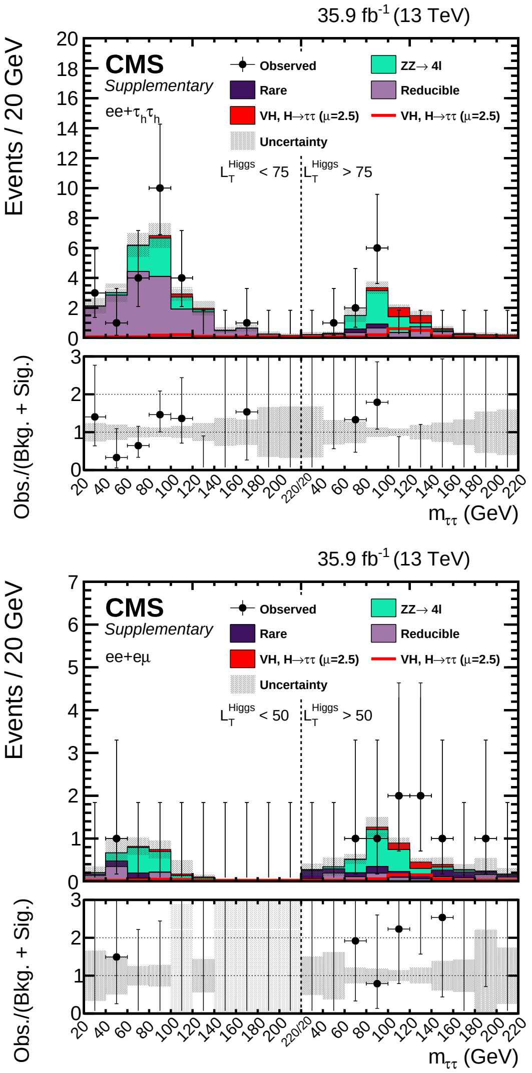


Figure 7.6: The postfit $m_{\tau\tau}$ distributions used to extract the signal shown for the (top left) $ee\tau_h\tau_h$, (top right) $\mu\mu\tau_h\tau_h$, (bottom left) $eee\mu$, and (bottom right) $\mu\mu e\mu$ final states. The distributions show full uncertainties. The WH and ZH, $H \to \tau\tau$ signal processes are summed together and shown as VH, $H \to \tau\tau$ with a best-fit $\mu = 2.5$. VH, $H \to \tau\tau$ is shown both as a stacked filled histogram and an open overlaid histogram. In these distributions the ZH, $H \to \tau\tau$ process contributes more than 99% of the total of VH, $H \to \tau\tau$.

grouping events in the signal regions by their decimal logarithm of the ratio of the signal (S) to signal-plus-background ($S + B$) in each bin

The postfit background and signal yields and the observed yields for the WH

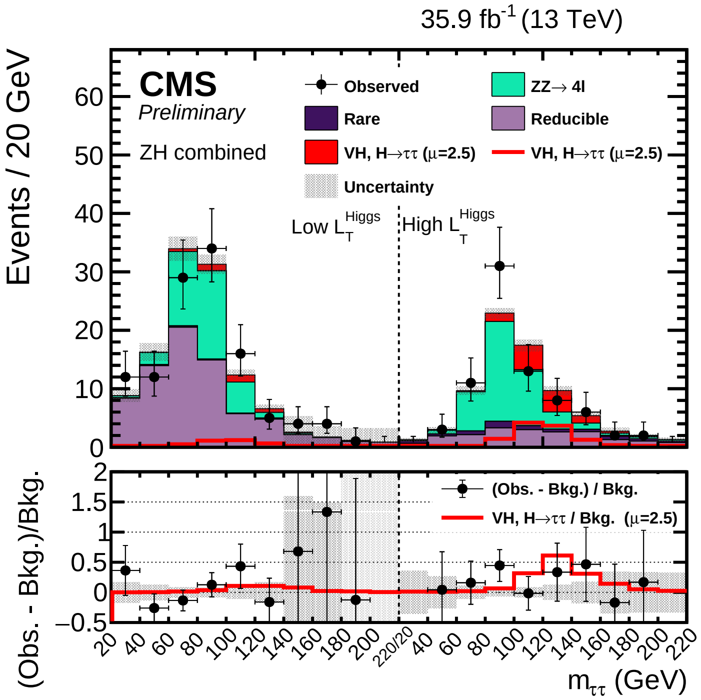


Figure 7.7: The postfit $m_{\tau\tau}$ distributions used to extract the signal shown for all 8 ZH final states combined. The distribution shows full uncertainties. The left half of the distribution is the Low-$L_{\mathrm{T}}^{\mathrm{Higgs}}$ region while the right half corresponds to the High-$L_{\mathrm{T}}^{\mathrm{Higgs}}$ region. The definitions of the $L_{\mathrm{T}}^{\mathrm{Higgs}}$ regions in this distribution are the same as those used in Figures 7.5 and 7.6 and are final state dependent. The WH and ZH, $H \to \tau\tau$ signal processes are summed together and shown as VH, $H \to \tau\tau$ with a best-fit $\mu = 2.5$. VH, $H \to \tau\tau$ is shown both as a stacked filled histogram and an open overlaid histogram. In this distribution the ZH, $H \to \tau\tau$ process contributes more than 99% of the total of VH, $H \to \tau\tau$.

final states are shown in Table 7.5 while those for the ZH final states are shown in Table 7.6. The ZH final states are grouped according to the Higgs boson decay.

Process	$WH, e\mu\tau_h$	$WH, \mu\mu\tau_h$	$WH, e\tau_h\tau_h$	$WH, \mu\tau_h\tau_h$
ZZ	1.56 ± 0.05	0.93 ± 0.03	0.82 ± 0.04	1.18 ± 0.05
WZ	7.92 ± 0.28	6.69 ± 0.24	4.83 ± 0.25	8.38 ± 0.42
Jet Fakes	10.09 ± 1.61	12.19 ± 1.72	10.68 ± 1.27	19.80 ± 1.87
Rare	2.28 ± 0.61	3.77 ± 0.84	1.71 ± 1.08	1.76 ± 0.90
Total backgrounds	21.85 ± 1.75	23.58 ± 1.92	18.04 ± 1.67	31.12 ± 2.12
$WH, H \to \tau\tau$	4.28 ± 0.72	4.25 ± 0.73	3.51 ± 0.62	5.45 ± 0.97
$ZH, H \to \tau\tau$	0.42 ± 0.07	0.40 ± 0.08	0.33 ± 0.07	0.44 ± 0.10
Total signal	4.70 ± 0.72	4.65 ± 0.73	3.84 ± 0.92	5.98 ± 0.98
Observed	28	29	23	38

Table 7.5: Background and signal expectations for the WH final states, together with the number of observed events, for the post-fit signal region distributions. The background uncertainty accounts for all sources of background uncertainty, systematic as well as statistical, after the global fit. The contribution from "Rare" includes events from triboson, $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}H$ production, and other rare processes.

Process	$\ell\ell e\tau_h$	$\ell\ell\mu\tau_h$	$\ell\ell\tau_h\tau_h$	$\ell\ell e\mu$
ZZ	14.40 ± 0.36	26.91 ± 0.55	25.58 ± 1.05	9.33 ± 0.18
Jet Fakes	14.01 ± 1.55	17.58 ± 1.17	58.05 ± 2.87	3.66 ± 4.60
Rare	0.62 ± 0.08	1.54 ± 0.61	0.81 ± 0.42	3.02 ± 0.23
Total backgrounds	29.03 ± 1.59	46.03 ± 1.43	84.44 ± 3.08	16.01 ± 4.61
$WH, H \to \tau\tau$	0.008 ± 0.002	0.01 ± 0.003	0.016 ± 0.005	0.002 ± 0.001
$ZH, H \to \tau\tau$	2.83 ± 0.39	5.31 ± 1.30	5.29 ± 1.17	1.62 ± 0.20
Total signal	2.84 ± 0.39	5.32 ± 0.70	5.31 ± 1.17	1.62 ± 0.20
Observed	33	53	87	20

Table 7.6: Background and signal expectations for the ZH final states, together with the number of observed events, for the post-fit signal region distributions. The ZH final states are each grouped according to the Higgs boson decay products. $\ell\ell$ covers both $Z \to \mu\mu$ and $Z \to ee$ events. The background uncertainty accounts for all sources of background uncertainty, systematic as well as statistical, after the global fit. The contribution from "Rare" includes events from triboson, $t\bar{t}Z$, $t\bar{t}H$ production, and other rare processes.

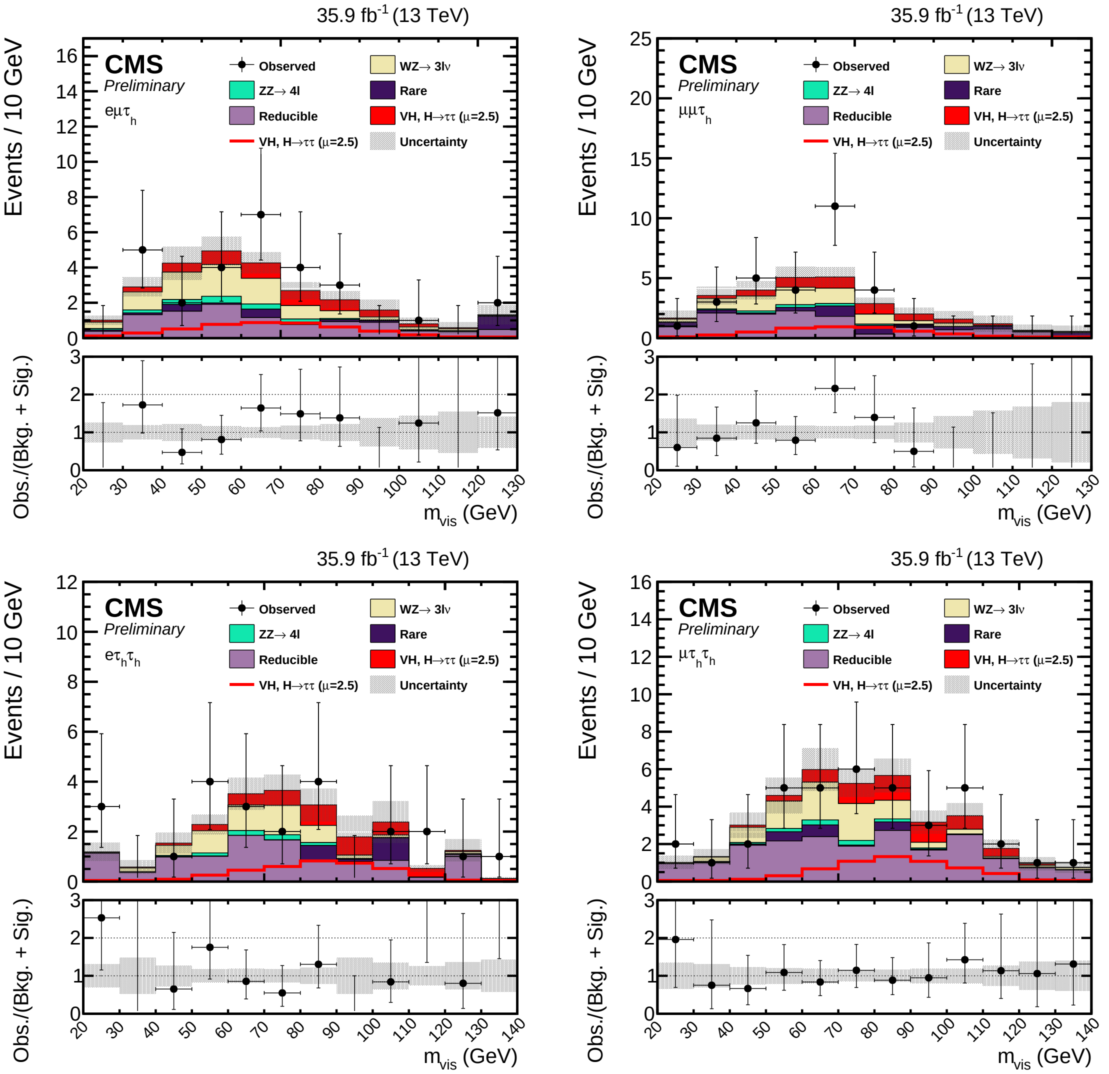


Figure 7.8: Postfit mass distributions in the $e\mu\tau_h$ (top left), $\mu\mu\tau_h$ (top right), $e\tau_h\tau_h$ (bottom left), and $\mu\tau_h\tau_h$ (bottom right) final states. The distributions show full uncertainties. The WH and ZH, $H \to \tau\tau$ signal processes are summed together and shown as VH, $H \to \tau\tau$ with a best-fit $\mu = 2.5$. VH, $H \to \tau\tau$ is shown both as a stacked filled histogram and an open overlaid histogram. In these distributions the WH, $H \to \tau\tau$ process contributes between 91%–93% of the total of VH, $H \to \tau\tau$.

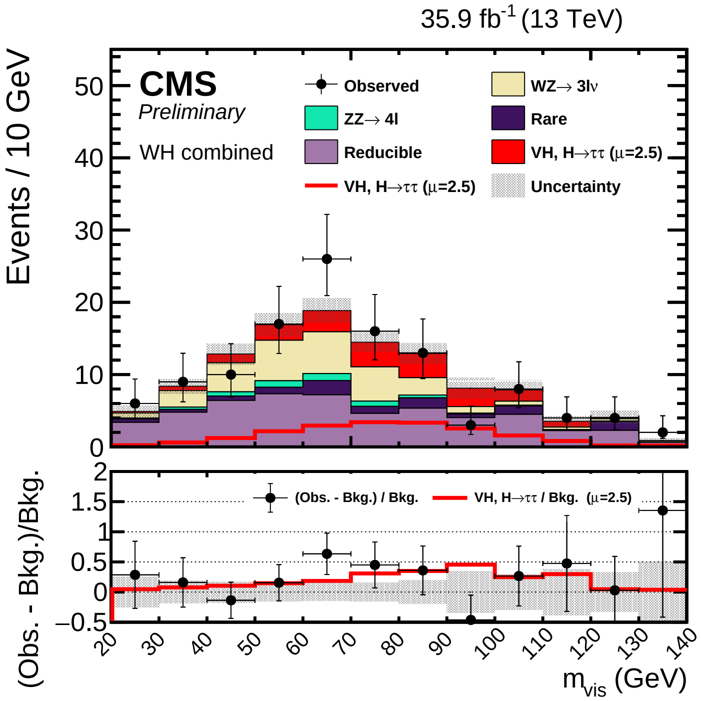


Figure 7.9: Postfit mass distributions of the four WH final states combined together. The distributions show full uncertainties. The WH and ZH, $H \rightarrow \tau\tau$ signal processes are summed together and shown as VH, $H \rightarrow \tau\tau$ with a best-fit $\mu = 2.5$. VH, $H \rightarrow \tau\tau$ is shown both as a stacked filled histogram and an open overlaid histogram. In this distribution the WH, $H \rightarrow \tau\tau$ process contributes 92% of the total of VH, $H \rightarrow \tau\tau$.

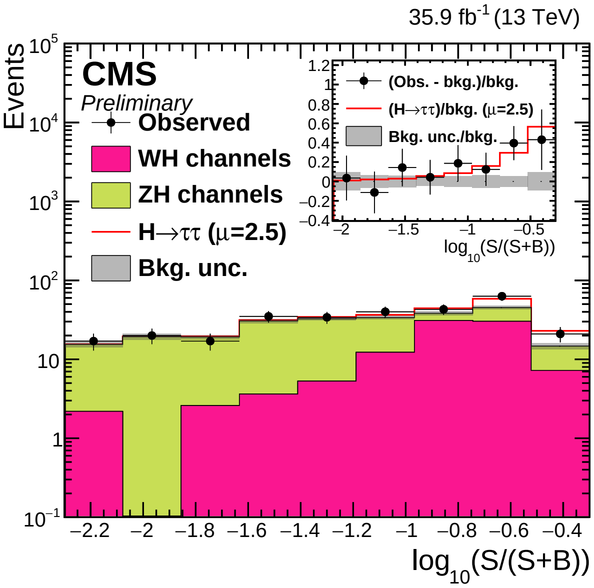


Figure 7.10: Distribution of the decimal logarithm of the ratio between the expected signal, corresponding to the best fit value $\mu = 2.5$, and the sum of expected signal and expected background in each bin of the mass distributions used to extract the results, in all signal regions. The background contributions are separated based on the final states, WH versus ZH. The inset shows the corresponding difference between the observed data and expected background distributions divided by the background expectation, as well as the signal expectation divided by the background expectation.

Chapter 8

Combined $H \to \tau\tau$ Results

The previous two Chapters, 6 and 7, discussed analyses targeted at specific Higgs boson production mechanisms. In this chapter I present results combining the ggH and VBF targeted analysis [3] with the WH and ZH associated production targeted analysis [4]. By combining these two $H \to \tau\tau$ analyses we have signal regions targeting the leading Higgs boson production processes: ggH, VBF, and WH and ZH associated production. Combining the results, signal strengths, the $H \to \tau\tau$ significance and, Higgs boson couplings can be probed with greater precision than either analysis alone.

Changes in the ggH signal modeling and uncertainties were made between publication of the ggH and VBF targeted analysis [3] and the combination here to take advantage of the most accurate, available simulations [146]. The ggH and VBF results presented in Chapters 6 uses ggH simulated with NLO accuracy (Chapter 4). In this combination, the NLO ggH samples were reweighted using the NNLOPS generator which is accurate at the next-to-next-to-leading order in the strong coupling [147]. The reweighting matches the Higgs boson p_{T} spectrum and the quantity of jets from the hard-scattering process between these generators and is defined to preserve the normalization of the inclusive ggH, $H \rightarrow \tau\tau$ simulated process [146]. Additionally, the ggH cross section uncertainty scheme has been updated to align with the one proposed in Reference [42]. This uncertainty scheme includes 9 nuisance parameters accounting for uncertainties in the cross section prediction for exclusive jet bins, the 2 jet and 3 jet VBF phase spaces, different Higgs boson p_{T} regions, and the uncertainty in the Higgs boson p_{T} distribution due to missing higher order finite top quark mass corrections.

8.1 Signal Strength and Significance

The best fit signal strength (μ) and significance can be computed for the combination and leads to a decrease in the relative uncertainty on μ and an increase in significance. The μ and significance for each analysis and the combined values are presented in Table 8.1. The slight excess in signal strength for the associated production analysis is tempered when combined with the ggH and VBF analysis, resulting in a μ fully consistent with the SM. The combination leads to an observed significance of 5.5 standard deviations (4.8 expected), surpassing the threshold for a purely 13 TeV based CMS observation of the $H \rightarrow \tau\tau$ process.

Analysis	Best Fit Signal Strength	Observed Significance
ggH and VBF	$\mu = 1.09^{+0.27}_{-0.26}$	4.9 σ
Associated production	$\mu = 2.5^{+1.4}_{-1.3}$	2.3 σ
$H \rightarrow \tau\tau$ combination	$\mu = 1.24^{+0.29}_{-0.27}$	5.5 σ

Table 8.1: Best fit signal strength and significance for three fit scenarios: ggH and VBF, associated production, and the combination with the updated ggH modeling.

The signal strength can be decomposed into the four leading Higgs boson production mechanisms. Figure 8.1 shows this decomposition for the combined results.

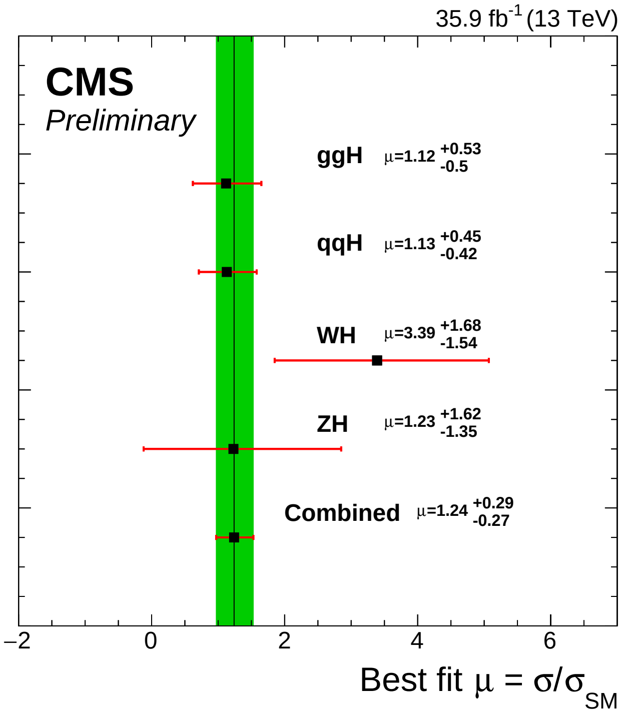


Figure 8.1: Best fit signal strength per Higgs boson production process, for $m_H = 125.09\,\mathrm{GeV}$. The constraints from the combined global fit are used to extract each of the individual best fit signal strengths. The combined best fit signal strength is $\mu = 1.24^{+0.29}_{-0.27}$.

8.2 Higgs Boson Couplings

The Higgs boson couplings to different particles can be measured in different ways. In this thesis the couplings are split into two groups, couplings to fermions (κ_f) and couplings to vector bosons (κ_V) and are measured using couplings on both the production and decay processes. κ_V and κ_f quantify the ratio between the measured and the SM value for the couplings of the Higgs boson using the methods described in Reference [51].

The $H \rightarrow \tau\tau$ decay process provides direct access to the Higgs boson fermion couplings. While all $H \rightarrow \tau\tau$ events provide access to fermionic couplings on the decay side, the different production mechanisms provide access to different couplings on that side. The ggH process provides indirect access to fermionic couplings through the top-quark loop. The VBF and associated production processes provide access to the vector boson couplings via Higgs boson production. The same combination of the dedicated ggH and VBF analysis, including the updates to the ggH modeling, with the dedicated WH and ZH analysis can place the tightest $H \rightarrow \tau\tau$ analysis limits in the (κ_{V},κ_{f}) parameter space because there is simultaneous access to processes which can constrain both couplings.

To measure the couplings, a likelihood scan is performed for $m_H = 125.09\,\mathrm{GeV}$ in the (κ_{V},κ_{f}) parameter space. For this scan only, Higgs boson decays to pairs of W or Z bosons, $H \rightarrow WW$ or $H \rightarrow ZZ$, are considered as part of the signal. They both contributed to κ_{V} on the decay side and are treated accordingly based on their production processes. All nuisance parameters are profiled for each point of the scan. As shown in Figure 8.2, the observed likelihood contour is consistent with the SM expectation of κ_{V} and κ_{f} both equal to unity showing agreement with the SM Higgs boson couplings to fermions and vector bosons.

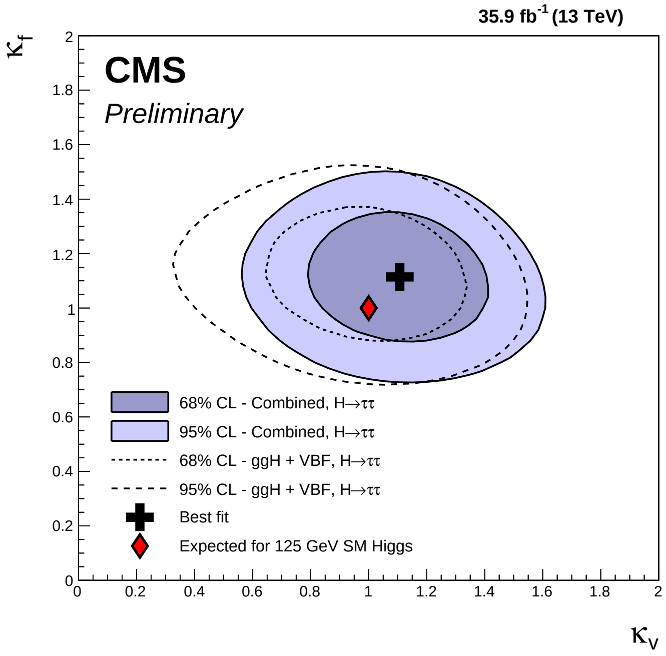


Figure 8.2: Scan of the negative log-likelihood difference as a function of κ_V and κ_f, for $m_H = 125.09\,\text{GeV}$. All nuisance parameters are profiled for each point. This scan is a combination of the ggH and VBF targeted analysis with the WH and ZH targeted analysis. For reference, the results for just the ggH and VBF targeted analysis are also presented and correspond to the updated ggH modeling. For this scan, all $H \to WW$ and $H \to ZZ$ processes are treated as signal.

Chapter 9

Conclusions

In this thesis I have presented two analyses of the standard model (SM) Higgs boson decaying to a τ lepton pair using data collected by the CMS experiment in 2016. The analyses target the leading Higgs boson production processes at the LHC: gluon fusion, vector boson fusion, and W and Z associated production. Specialized categories are used to target the unique event topologies and characteristics of each of these four production processes to maximize our sensitivity to the Higgs boson and its couplings.

These analyses studying the $H \to \tau\tau$ process provide a 5.5 standard deviation (4.8 expected) observation of the Higgs boson decaying to fermions at 13 TeV center-of-mass energy. The best fit signal strength for the $H \to \tau\tau$ process is measured to be $\mu = 1.24^{+0.29}_{-0.27}$, consistent with SM predictions. The Higgs boson couplings to fermions and vector bosons are measured and are consistent with standard model predictions within one standard deviation. This 5.5 standard deviation observation of the $H \to \tau\tau$ process and the consistency of the Higgs boson couplings with the standard model provide confirmation of the Higgs boson Yukawa couplings to fermions. This is evidence that the Higgs field provides mass for the τ lepton in addition to the vector bosons. This is an important benchmark for the CMS experiment and the high energy particle physics community.

The impressive agreement of these Higgs boson results and other previous results with SM predictions, and the lack of any observed physics signatures beyond the SM, point to the need to measure the details of the Higgs boson to increasing accuracy. The Higgs boson is the most recently discovered fundamental particle and occupies a very unique position in the SM compared to the other particles. For these reasons, the Higgs boson should continue to be used as a tool to probe the details of the SM and beyond.

9.1 Future Studies

There is still room for improvement in these measurements, which will come with including the additional 13 TeV data from the remainder of the LHC Run-II. Techniques have been developed to probe the Higgs boson properties as a function of event topology in unique phase spaces defined at the generator level using characteristics such as the number of jets in the event, the p_{T} of the Higgs boson, the m_{jj} in vector boson fusion type events, and the p_{T} of the vector boson in associated production events. In the simplified template cross section (STXS) method, the decomposition has been performed to focus on characteristics of the production topologies which are sensitive to higher order theoretical corrections as well as possible physics beyond the standard model [148]. Using only the 2016 dataset, we are too statistically limited to take full advantage of the STXS method. However, with the full Run-II dataset, the $H \to \tau\tau$ channel will be able to probe details of the Higgs boson p_{T} spectrum and more.

The vector boson fusion and associated production processes with $H \to \tau\tau$ are powerful mechanisms for studying the couplings of the Higgs boson to vector bosons (HVV couplings). This is shown in Figure 8.2 under the assumption that the HVV and fermionic couplings behave according to the SM Lagrangian. The results are consistent with the SM. However, instead of measuring the consistency of the results with the SM, we can make alternative physics assumptions and test their consistency against data. Ongoing analyses with 2016 data are studying possible anomalous HVV couplings from additional coupling terms introduced into the SM Lagrangian. These studies are expected to provide the tightest constraints at the 2σ level on possible anomalous HVV couplings.

Bibliography

[1] 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.

[2] Albert M Sirunyan et al. Measurement of the $Z\gamma^* \to \tau\tau$ cross section in pp collisions at $\sqrt{s} = 13$ TeV and validation of τ lepton analysis techniques. 2018.

[3] Observation of the higgs boson decay to a pair of leptons with the cms detector. *Physics Letters B*, 779:283 – 316, 2018.

[4] CMS Collaboration. Search for the standard model Higgs boson decaying to a pair of τ leptons produced in association with a W or a Z boson. 2018.

[5] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1, 2012.

[6] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, 716:30, 2012.

[7] Serguei Chatrchyan et al. Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV. *JHEP*, 06:081, 2013.

[8] S. L. Glashow. Partial-symmetries of weak interactions. *Nucl. Phys.*, 22:579, 1961.

[9] S. Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264, 1967.

[10] 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.

[11] E. D. Bloom, D. H. Coward, H. DeStaebler, J. Drees, G. Miller, L. W. Mo, R. E. Taylor, M. Breidenbach, J. I. Friedman, G. C. Hartmann, and H. W. Kendall. High-energy inelastic $e-p$ scattering at 6 degrees and 10 degrees. *Phys. Rev. Lett.*, 23:930–934, Oct 1969.

[12] Martin Breidenbach, Jerome I. Friedman, Henry W. Kendall, Elliott D. Bloom, D. H. Coward, H. C. DeStaebler, J. Drees, Luke W. Mo, and Richard E. Taylor. Observed Behavior of Highly Inelastic electron-Proton Scattering. *Phys. Rev. Lett.*, 23:935–939, 1969.

[13] J. J. et al. Aubert. Experimental observation of a heavy particle j. *Phys. Rev. Lett.*, 33:1404–1406, Dec 1974.

[14] J. E. et al. Augustin. Discovery of a narrow resonance in e^+e^- annihilation. *Phys. Rev. Lett.*, 33:1406–1408, Dec 1974.

[15] S. W. et al. Herb. Observation of a dimuon resonance at 9.5 gev in 400-gev proton-nucleus collisions. *Phys. Rev. Lett.*, 39:252–255, Aug 1977.

[16] F. et al. Abe. Observation of top quark production in $\overline{p}p$ collisions with the collider detector at fermilab. *Phys. Rev. Lett.*, 74:2626–2631, Apr 1995.

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

[18] D. P. et al. Barber. Discovery of three-jet events and a test of quantum chromodynamics at petra. *Phys. Rev. Lett.*, 43:830–833, Sep 1979.

[19] F.J. Hasert et al. Search for elastic muon-neutrino electron scattering. *Physics Letters B*, 46(1):121 – 124, 1973.

[20] F.J. Hasert et al. Observation of neutrino-like interactions without muon or electron in the gargamelle neutrino experiment. *Physics Letters B*, 46(1):138 – 140, 1973.

[21] J.J. Aubert et al. The ratio of the nucleon structure functions f2n for iron and deuterium. *Physics Letters B*, 123(3):275 – 278, 1983.

[22] G. Arnison et al. Experimental observation of lepton pairs of invariant mass around 95 gev/c2 at the cern sps collider. *Physics Letters B*, 126(5):398 – 410, 1983.

[23] M. L. et al. Perl. Evidence for anomalous lepton production in $e^{+} - e^{-}$ annihilation. *Phys. Rev. Lett.*, 35:1489–1492, Dec 1975.

[24] K. Kodama et al. Observation of tau neutrino interactions. *Phys. Lett.*, B504:218–224, 2001.

[25] R. Barate et al. Observation of an excess in the search for the Standard Model Higgs boson at ALEPH. *Phys. Lett. B*, 495:1, 2000.

[26] J. Abdallah et al. Final results from DELPHI on the searches for SM and MSSM neutral Higgs bosons. *Eur. Phys. J. C*, 32:145, 2004.

[27] P. Achard et al. Standard model Higgs boson with the L3 experiment at LEP. *Phys. Lett. B*, 517:319, 2001.

[28] G. Abbiendi et al. Search for the Standard Model Higgs boson in $\mathrm{e}^{+}\mathrm{e}^{-}$ collisions at $\sqrt{s} = 192$–209 GeV. *Phys. Lett. B*, 499:38, 2001.

[29] T. Aaltonen et al. 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, 2012.

[30] Victor Mukhamedovich Abazov et al. Search for the standard model Higgs boson in tau lepton final states. *Phys. Lett. B*, 714:237, 2012.

[31] P. W. Anderson. More is different. *Science*, 177:393–396, 1972.

[32] E. Noether. Invariante variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Gttingen, Mathematisch-Physikalische Klasse*, 1918:235–257, 1918.

[33] R. Slansky. Group theory for unified model building. *Physics Reports*, 79(1):1 – 128, 1981.

[34] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321, 1964.

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

[36] P. W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508, 1964.

[37] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585, 1964.

[38] P. W. Higgs. Spontaneous symmetry breakdown without massless bosons. *Phys. Rev.*, 145:1156, 1966.

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

[40] Jeffrey Goldstone, Abdus Salam, and Steven Weinberg. Broken symmetries. *Phys. Rev.*, 127:965–970, Aug 1962.

[41] J. Beringer et al. Review of particle physics. *Phys. Rev. D*, 86:010001, 2012.

[42] D. de Florian et al. Handbook of LHC Higgs cross sections: 4. deciphering the nature of the Higgs sector. CERN Report CERN-2017-002-M, 2016.

[43] M. Duhrssen, S. Heinemeyer, H. Logan, D. Rainwater, G. Weiglein, and D. Zeppenfeld. Extracting higgs boson couplings from cern lhc data. *Phys. Rev. D*, 70:113009, Dec 2004.

[44] LHC Higgs Cross Section Working Group. Handbook of LHC Higgs cross sections: 2. Differential Distributions. CERN Report CERN-2012-002, 2012.

[45] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, and M. Spira. Standard model Higgs-boson branching ratios with uncertainties. *Eur. Phys. J. C*, 71:1753, 2011.

[46] Richard D. Ball et al. Impact of heavy quark masses on parton distributions and LHC phenomenology. *Nucl. Phys. B*, 849:296, 2011.

[47] T. et al. Aaltonen. Higgs boson studies at the tevatron. *Phys. Rev. D*, 88:052014, Sep 2013.

[48] T. et al. Aaltonen. 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, Aug 2012.

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

[50] Georges Aad et al. Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector. *JHEP*, 04:117, 2015.

[51] 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.

[52] Georges Aad et al. Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment. *Eur. Phys. J. C*, 76:6, 2016.

[53] 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. C*, 75:212, 2015.

[54] Serguei Chatrchyan et al. Study of the mass and spin-parity of the Higgs boson candidate via its decays to Z boson pairs. *Phys. Rev. Lett.*, 110:081803, 2013.

[55] Georges Aad et al. Evidence for the spin-0 nature of the Higgs boson using ATLAS data. *Phys. Lett. B*, 726:120, 2013.

[56] Vardan Khachatryan et al. Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV. *Phys. Rev. D*, 92:012004, 2015.

[57] CMS Collaboration. Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$. Submitted to JHEP, 2017.

[58] Albert M Sirunyan et al. Observation of $\mathrm{t\bar{t}H}$ production. 2018.

[59] Rudiger Voss and Amos Breskin, editors. *The CERN Large Hadron Collider, accelerator and experiments*. 2009.

[60] A. Verweij et al. Retraining of the 1232 main dipole magnets in the lhc. *IEEE Transactions on Applied Superconductivity*, 26(4):1–1, June 2016.

[61] *The LHC Superconducting Magnets*, 2003.

[62] J. D. Adam et al. Status of the lhc superconducting cable mass production. *IEEE Transactions on Applied Superconductivity*, 12(1):1056–1062, Mar 2002.

[63] M. Benedikt, P. Collier, V. Mertens, J. Poole, and K. Schindl. LHC Design Report. 3. The LHC injector chain. 2004.

[64] Vardan Khachatryan et al. The CMS trigger system. *JINST*, 12:P01020, 2017.

[65] S. Chatrchyan et al. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[66] I. Dafinei, E. Auffray, P. Lecoq, and M. Schneegans. Lead tungstate for high energy calorimetry. *MRS Proceedings*, 348:99, 1994.

[67] P. Adzic et al. Energy resolution of the barrel of the CMS electromagnetic calorimeter. *JINST*, 2:P04004, 2007.

[68] CMS collaboration. The compact muon solenoid technical proposal. Technical report, CMS, 2008.

[69] 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.

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

[71] The CMS collaboration. Effect of gas composition on the performance of cathode strip chambers for the cms endcap muon system. 2004.

[72] R. Santonico and R. Cardarelli. *Nucl. Instrum. Meth.*, 187, 1981.

[73] L. Cadamuro. The CMS Level-1 trigger system for LHC Run II. *JINST*, 12(03):C03021, 2017.

[74] T. Bawej et al. The new cms daq system for run-2 of the lhc. In *2014 19th IEEE-NPSS Real Time Conference*, pages 1–1, May 2014.

[75] Torbjorn Sjostrand, 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, 2015.

[76] M. Bahr, S. Gieseke, M.A. Gigg, D. Grellscheid, K. Hamilton, et al. Herwig++ Physics and Manual. *Eur.Phys.J.*, C58:639–707, 2008.

[77] 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.

[78] 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.

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

[80] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5: going beyond. *JHEP*, 06:128, 2011.

[81] 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.

[82] E. Reya. Perturbative quantum chromodynamics. *Physics Reports*, 69(3):195 – 333, 1981.

[83] Paolo Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.

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

[85] 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.

[86] 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.

[87] J. et al. Alwall. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J. C*, 53:473, 2008.

[88] Rikkert Frederix and Stefano Frixione. Merging meets matching in MC@NLO. *JHEP*, 12:061, 2012.

[89] John M. Campbell and R. K. Ellis. MCFM for the Tevatron and the LHC. *Nucl. Phys. Proc. Suppl.*, 205-206:10, 2010.

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

[91] 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.

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

[93] 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.

[94] 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 $\to WW \to l\nu l\nu$ and H $\to ZZ \to 4l$ decay modes. *JHEP*, 06:132, 2012.

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

[96] Richard D. et al. Ball. Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO. *Nucl. Phys. B*, 855:153, 2012.

[97] Vardan Khachatryan et al. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J. C*, 76:155, 2016.

[98] Rick Field. The underlying event in hard scattering processes. 2008.

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

[100] S. Mrenna and P. Skands. Automated parton-shower variations in pythia 8. *Phys. Rev. D*, 94:074005, Oct 2016.

[101] T. Sjöstrand and P. Z. Skands. Transverse-momentum-ordered showers and interleaved multiple interactions. *The European Physical Journal C - Particles and Fields*, 39(2):129–154, Feb 2005.

[102] Stefan Hche. Introduction to parton-shower event generators. In *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders (TASI 2014): Boulder, Colorado, June 2-27, 2014*, pages 235–295, 2015.

[103] Torbjorn Sjostrand. The Lund Monte Carlo for Jet Fragmentation and e+ e- Physics: Jetset Version 6.2. *Comput. Phys. Commun.*, 39:347–407, 1986.

[104] G. S. Bali and K. Schilling. Static quark - anti-quark potential: Scaling behavior and finite size effects in SU(3) lattice gauge theory. *Phys. Rev.*, D46:2636–2646, 1992.

[105] Philip Ilten. Tau decays in pythia 8. *Nuclear Physics B - Proceedings Supplements*, 253-255:77 – 80, 2014. The Twelfth International Workshop on Tau-Lepton Physics (TAU2012).

[106] S. Agostinelli et al. GEANT4 — a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250, 2003.

[107] *Validation of Physics Models of Geant4 using Data from CMS Experiment*, number 042005 in 898, 2017.

[108] CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. Submitted to J. Instrum., 2017.

[109] ALEPH Collaboration. Performance of the aleph detector at lep. *Nuclear Instruments and Methods in Physics Research*, 360:481–506, 1995.

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

[111] Studies of Tracker Material. Technical Report CMS-PAS-TRK-10-003, 2010.

[112] A. Strandlie W. Adam, R. Fruhwirth and T. Todorov. Reconstruction of electrons with the gaussian sum filter in the cms tracker at lhc. *J. Phys. G*, 31, 2005.

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

[114] The CMS Collaboration. Observation of the sm scalar boson decaying to a pair of τ leptons with the cms experiment at the lhc. *Physics Letters B*, 779, 2017.

[115] CMS Collaboration. Performance of cms muon reconstruction in cosmic-ray events. *Journal of Instrumentation*, 5(03):T03022, 2010.

[116] 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:P06005, 2015.

[117] Vardan Khachatryan et al. Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV. *JINST*, 10:P02006, 2015.

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

[119] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet user manual. *Eur. Phys. J. C*, 72:1896, 2012.

[120] M. Cacciari and G. P. Salam. Dispelling the N^3 myth for the k_t jet-finder. *Phys. Lett. B*, 641:57, 2006.

[121] Serguei Chatrchyan et al. Determination of jet energy calibration and transverse momentum resolution in CMS. *JINST*, 6:11002, 2011.

[122] CMS Collaboration. Identification of b quark jets at the CMS Experiment in the LHC Run 2. 2016.

[123] Serguei Chatrchyan et al. Identification of b-quark jets with the CMS experiment. *JINST*, 8:P04013, 2013.

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

[125] CMS Collaboration. Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2. CMS Physics Analysis Summary CMS-PAS-TAU-16-002, 2016.

[126] Helge Voss, Andreas Höcker, Jörg Stelzer, and Fredrik Tegenfeldt. TMVA, the toolkit for multivariate data analysis with ROOT. In *XI Int. Workshop on Advanced Computing and Analysis Techniques in Physics Research*, 2007.

[127] CMS Collaboration. Pileup jet identification. CMS Physics Analysis Summary CMS-PAS-JME-13-005, 2013.

[128] CMS Collaboration. Particle–flow event reconstruction in CMS and performance for jets, taus, and $\vec{E}_T^{miss}$. CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009.

[129] Vardan Khachatryan et al. Search for neutral MSSM Higgs bosons decaying to a pair of tau leptons in pp collisions. *JHEP*, 10:160, 2014.

[130] Lorenzo Bianchini, John Conway, Evan Klose Friis, and Christian Veelken. Reconstruction of the Higgs mass in $H \to \tau\tau$ Events by Dynamical Likelihood techniques. *J. Phys. Conf. Ser.*, 513:022035, 2014.

[131] A Kolmogorov. Sulla determinazione empirica di una legge di distribuzione. *G. Ist. Ital. Attuari.*, 4:83–91, 1933.

[132] CMS Collaboration. CMS luminosity measurements for the 2016 data taking period. CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017.

[133] 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.

[134] 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.

[135] Vardan Khachatryan et al. Measurement of the WZ production cross section in pp collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 766:268, 2017.

[136] Albert M Sirunyan et al. Cross section measurement of t-channel single top quark production in pp collisions at $\sqrt{s} = 13$ TeV. In proofs, Phys. Lett. B, 2016.

[137] Charalampos et al. Anastasiou. High precision determination of the gluon fusion Higgs boson cross-section at the LHC. *JHEP*, 05:058, 2016.

[138] Massimiliano Grazzini and Hayk Sargsyan. Heavy-quark mass effects in Higgs boson production at the LHC. *JHEP*, 09:129, 2013.

[139] J. Bellm, S. Gieseke, D. Grellscheid, A. Papaefstathiou, S. Plätzer, P. Richardson, C. Röhr, T. Schuh, M. H. Seymour, A. Siódmok, A. Wilcock, and B. Zimmermann. Herwig++ 2.7 release note. 2013.

[140] Iain W. Stewart and Frank J. Tackmann. Theory uncertainties for Higgs and other searches using jet bins. *Phys. Rev. D*, 85:034011, 2012.

[141] ATLAS, 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.

[142] 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.

[143] Thomas Junk. Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth. A*, 434:435, 1999.

[144] A. L. Read. Presentation of search results: the CL_s technique. In *Durham IPPP Workshop: Advanced Statistical Techniques in Particle Physics*, page 2693, Durham, UK, March 2002. [J. Phys. G 28 (2002) 2693].

[145] Jon Butterworth et al. PDF4LHC recommendations for LHC Run II. *J. Phys. G*, 43:023001, 2016.

[146] Combined measurements of the Higgs boson's couplings at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-HIG-17-031, CERN, Geneva, 2018.

[147] Keith Hamilton, Paolo Nason, Emanuele Re, and Giulia Zanderighi. Nnlops simulation of higgs boson production. *Journal of High Energy Physics*, 2013(10):222, Oct 2013.

[148] Frank Tackmann, Kerstin Tackmann, Kerstin Tackmann, Michael Duehrssen-Debling, and Paolo Francavilla. Simplified template cross sections. Mar 2016.