SEARCH FOR NEW PHYSICS WITH TAU LEPTON FINAL STATES

by

Zaixing Mao

B.A. in Physics, Peking University, 2011

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF DOCTOR OF PHILOSOPHY

IN THE DEPARTMENT OF PHYSICS AT BROWN UNIVERSITY

Providence, Rhode Island

May 2017

© Copyright 2013 by Zaixing Mao

This dissertation by Zaixing Mao is accepted in its present form by
the Department of Physics as satisfying the
dissertation requirement for the degree of
Doctor of Philosophy

Date..... .....
Ulrich Heintz, Advisor

Recommended to the Graduate Council

Date..... .....
David Cutts, Reader

Date..... .....
Jiji Fan, Reader

Approved by the Graduate Council

Date..... .....
Andrew G. Campbell
Dean of the Graduate School

Zaixing Mao
Curriculum Vitae

Physics Department
Barus and Holley Building
Brown University
182 Hope Street
Providence, RI 02906

zaixing.mao@cern.ch

Education

- **Brown University**, Providence, RI, USA
  - Doctor of Philosophy in Physics, Sep 2011 - May 2017
  - Thesis: "Search for New Physics with Tau Lepton Final States"
  - Advisor: Prof. Ulrich Heintz
- **Peking University**, Beijing, China
  - Bachelor of Science, Physics, Sep 2007 - May 2011

Research Experience

- **Search for new physics with high-mass tau pairs**

  The Standard Model (SM) of particle physics is a successful theory that can explain many experimental observations involving weak, electromagnetic, and strong interactions. However, with the lack of explanations for gravity and dark matter, it leaves room for theories beyond the SM (BSM). Many BSM theories predict the presence of extra gauge bosons. In this analysis, I lead the search for a charge neutral gauge boson that decays into a pair of tau leptons using collision data at the center-of-mass energy of $\sqrt{s} = 13$ TeV collected by the Compact Muon Solenoid (CMS) detector. Based on the decay mode of the tau lepton, four final states: $\tau_\mu\tau_h$ (where one tau decays into a muon and the other decays hadronically), $\tau_e\tau_h$ (where one tau decays into an electron and the other decays hadronically), $\tau_h\tau_h$ (where both taus decays hadronically), and $\tau_e\tau_\mu$ (where one tau decays into an electron and the other decays into a muon), were studied. Using the data collected in 2015, this analysis has been published in JHEP. This is the first search for heavy resonances decaying to a tau lepton pair using events from proton-proton collisions at $\sqrt{s} = 13$ TeV. A similar search using data collected in 2016 is currently ongoing where I am contributing three out of the four final states.

- **Search for a heavy Higgs decaying to two 125 GeV Higgs in the 2 τ + 2 b final state using data at $\sqrt{s} = 8$ TeV**

  The inclusion of gravity requires a theory that is valid up to the Planck scale. If no new physics resides above the SM, running the SM up to the Planck scale leads to divergence

terms in the Higgs mass calculation. With the observation of the Higgs boson at 125 GeV, this can only be achieved by a large amount of fine-tuning to the SM parameters, known as the Hierarchy problem. The Minimal Supersymmetric Standard Model (MSSM) tackles the Hierarchy problem by introducing super partners to the SM particles. These new particles introduce extra correction terms in the Higgs mass calculation and automatically cancel out the diverging terms which need fine-tuning. In this analysis, a search for a massive Higgs boson predicted by the MSSM theory that decays into two SM like Higgs bosons that further decays into two tau leptons and two bottom quarks. This study was performed using collision data at the center-of-mass energy of $\sqrt{s} = 8$ TeV collected by the CMS detector. Using three tau decay channels: $\tau_\mu\tau_h$, $\tau_e\tau_h$, and $\tau_h\tau_h$, this is the first search for heavy Higgs boson through tau lepton final states and has been published in PLB.

- **Radiation Hard Studies for Phase-II Upgrade of CMS Detector**

  Between the years of 2024 to 2026, the Large Hadron Collider (LHC) is scheduled for a major upgrade, leading into the era of the High Luminosity LHC (HL-LHC), which enables it to deliver an instantaneous luminosity 5 to 7 times the current value. With the projected high radiation environment, it is very crucial to develop new detector technology to withstand such harsh environments. This project is a research into the development of radiation hard silicon sensors to be used in the CMS tracker during the HL-LHC period. Together with my colleagues at Brown University, we have built experiment setups to test key performance characteristics of prototype silicon sensors before and after irradiation. Additionally, we have also preformed annealing studies to simulate the performance of irradiated silicon sensors after many years of usage at the LHC.

- **Commissioning of CMS Sub-detector Readout System**

  With the upgrade of the back-end readout system of the Hadron Calorimeter (HCAL) in the CMS detector, I participated in many tests to ensure the new hardware performs as expected. The HCAL back-end readout system is responsible for receiving data from the front-end system, calculate relevant information for the level 1 trigger system, and transferring the buffered data to the Data Acquisition system (DAQ) upon acknowledgement from the trigger system. To commission all of the connections to the different systems, my colleagues and I developed multiple tests, ranging from automatic checks on fiber connections to standalone stress tests, to make sure the new back-end system is ready to be installed.

List of Publications

- CMS Collaboration, Z. Mao, "Search for new physics with tau final states at 13 TeV", Nuclear and Particle Physics Proceedings, S2405-6014(17)30120-7.
- CMS Collaboration, "Search for heavy resonances decaying to tau lepton pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV", J. High Energ. Phys. (2017) 2017: 48.
- CMS Collaboration, "Searches for a heavy scalar boson H decaying to a pair of 125 GeV Higgs bosons hh or for a heavy pseudoscalar boson A decaying to Zh, in the final states with h $\rightarrow \tau\tau$", Phys. Lett. B 755 (2016) 217.

List of Presentations

- "LPC CMSDAS long exercise $Z \rightarrow \tau\tau$ production cross section measurement," in *CMS Data Analysis Schools*, Fermilab, Chicago, USA, Jan 2017.
- "Search for new physics with tau final states at 13 TeV," in *the 14th International Workshop on Tau Lepton Physics*, Beijing, China, Sep 2016.

- "Tau Performance at CMS," in *Tau Hands-on Advanced Tutorial Sessions*, Fermilab, Chicago, USA, Jun 2016.
- "Search for heavy resonances decaying to tau lepton pairs in proton-proton collisions at $\sqrt{s} =$ 13 TeV," in *EXO-16-008 Pre-approval Presentation*, CERN, Switzerland, Feb 2016.
- "Phase I Upgrade of the HCAL Backend Electronics," in *US-CMS Annual Meeting*, Cornell University, Ithaca, NY, USA, May 2015.
- "AMC13 readout kit," in *CMS Upgrade Week at CERN*, CERN, Switzerland, Oct 2013.
- Several related presentations at various CMS internal meetings and workshops.

Awards

- **The LHC Physics Center Guests and Visitors program**, Fermilab, Chicago, Jun 2015 - Aug 2015.
- **Dissertation Fellow**, Brown University, Aug 2016 - May 2017.

Mentorship

- **CMS Data Analysis Schools**, Facilitator, Fermilab, Chicago, Jan 2017 and Jan 2016.
- **Tau Hands-on Advanced Tutorial Sessions**, Facilitator, Fermilab, Chicago, Sep 2016.
- **Teaching Assistant**, Brown University, Providence, Sep 2011 to May 2012.

Skills

- **Computing**: C++, Python, Unix, GitHub, ROOT, CMSSW.
- **Statistics**: Hypothesis testing with frequentist and Bayesian methods, machine learning (regularization, sampling methods etc).
- **Communication/collaboration**: Worked in a large collaboration of more than 2000 scientists, participated in weekly discussions with $\sim$10 physicists, and gave tutorial and seminars to undergrads and grad students.

Preface and Acknowledgments

My years at Brown have been an amazing ride filled with fond memories and I could not have asked for more. Many people have contributed in their own unique ways to this journey and I do regret not being able to write down my gratitude for each and every one of you. However, some of you are just too special to miss.

First and foremost, I would like to express my greatest gratitude to my advisor Professor Ulrich Heintz for your support in academia as well as in personal life. Thank you for you knowledge, the opportunities you've fought for me, your understanding of my concerns, and your faith in me. To my thesis readers Professor David Cutts and Professor Jiji Fan, thank you for your advise and help in improving this dissertation. To Professor Meenakshi Narain, thank you for all of your supports throughout the years, it has been a great feeling to know that you'll aways have our backs. To Professor Greg Landsberg, thank you for introducing me to the field and all the support while I was at CERN. To Professor Edward Laird, who has been a great mentor and friend, thank you for everything.

To my colleagues at Brown, I want to thank Dr. Stefan Piperov for managing the computing facility and making my work a breeze. To all of my fellow post-graduate students, in particular, Dr. Sinan Sagir, for the countless discussions and advise. To my friends for all the laughs and joys we had shared together.

I want to reserve a special thanks to my parents and family. I would have never reached here if it were not for you. Finally, to my wife and best friend Yang Hu. Thank you for your support and love, especially at times when I needed them the most.

Contents

Curriculum Vitae **iv**

Preface and Acknowledgments **vii**

Contents **viii**

List of Figures **xiii**

List of Tables **xx**

1 The Standard Model **1**

1.1 Elementary Particles of the Standard Model 1

1.1.1 Fermions 1

1.1.2 Bosons 3

1.2 Standard Model Forces 4

1.2.1 The Quantum Electrodynamics 4

1.2.2 Yang-Mills Theory 6

1.2.3 Quantum Chromodynamics 6

1.2.4 The Higgs Mechanism 7

1.2.5 The Eelectroweak Unification 8

1.3 The Hierarchy Problem 11

1.4 Beyond the Standard Model 11

1.4.1 Minimal Supersymmetric Standard Model 11

1.4.2 Little Higgs Model 13

1.4.3 Left-Right Symmetric Model 13

1.4.4 E_6 Grand Unified Theory 14

1.4.5 Sequential Standard Model 15

2 The Compact Muon Solenoid Experiment at the Large Hadron Collider 16

2.1 The Large Hadron Collider . . . 16

2.2 The Compact Muon Solenoid . . . 19

2.2.1 Silicon Trackers . . . 20

2.2.2 Electromagnetic Calorimeter . . . 21

2.2.3 Hadronic Calorimeter . . . 23

2.2.4 Muon System . . . 24

3 Object Reconstruction at CMS 26

3.1 Primary Vertex . . . 26

3.2 Electron Reconstruction and Identification . . . 27

3.3 Muon Reconstruction and Identification . . . 28

3.4 Jet Reconstruction and Identification . . . 29

3.4.1 b-Jet Tagging . . . 30

3.5 Tau Reconstruction and Identification . . . 31

3.5.1 Tau Reconstruction . . . 33

3.5.2 Tau Isolation Discriminants . . . 34

3.5.2.1 Cut-based Discriminants . . . 34

3.5.2.2 MVA-based Discriminators Against Jets . . . 35

3.5.2.3 MVA-based Discriminators Against Electrons . . . 35

3.5.2.4 Cut-based Discriminators Against Muons . . . 35

3.6 Missing Transverse Momentum . . . 36

4 Radiation Damage Studies for the CMS Strip Tracker 38

4.1 LHC Upgrades . . . 38

4.2 Basic Features of Silicon Detectors . . . 39

4.2.1 P-N Junctions . . . 39

4.2.2 Characteristics of a Reverse Biased P-N Junction . . . 40

4.3 Test Structures . . . 41

4.4 Experimental Setup . . . 42

4.5 Measurement . . . 44

4.6 Radiation Damage . . . 46

4.6.1 Effect of Radiation Damage on Leakage Current . . . 46

4.6.2 Radiation Damage on the Effective Doping Concentration . . . 47

4.7 Annealing 48

5 Search for a heavy Higgs decaying to two 125 GeV Higgs in the 2 τ + 2 b final state using data at $\sqrt{s}$ = 8 TeV 50

5.1 Introduction 51

5.2 Strategy 51

5.3 Datasets and Monte Carlo samples 52

5.4 Event Selection 53

5.5 Background Estimation 55

5.5.1 $t\bar{t}$ background 56

5.5.2 QCD background 56

5.5.3 Drell-Yan $Z \to \tau\tau$ 58

5.5.4 W + jets 59

5.5.5 Other small backgrounds 59

5.6 Event yields and control plots 59

5.7 Signal extraction 61

5.7.1 Kinematic Fit 61

5.7.2 BDT Approach 61

5.8 Systematics 62

5.8.1 Normalization Uncertainties 68

5.8.2 Shape Uncertainties 69

5.9 Results and Interpretation 69

5.10 Summary 73

6 Search for new physics with high-mass tau pairs with 2015 data 74

6.1 Introduction 74

6.2 Strategy 75

6.3 Data and Monte Carlo Samples 75

6.3.1 Corrections for Pile-Up 76

6.4 Electron + Hadronic Tau Channel 78

6.4.1 Event selection 78

6.4.2 Genuine dilepton events 79

6.4.2.1 Drell-Yan process 79

6.4.2.2 $t\bar{t}$ and single top processes 79

6.4.2.3 Di-boson process . . . 81
6.4.3 QCD background . . . 82
6.4.4 W+jets background . . . 83
6.5 Electron + Muon Channel . . . 84
6.5.1 Event selection . . . 84
6.5.2 Genuine di-lepton events . . . 85
6.5.2.1 Drell-Yan process . . . 85
6.5.2.2 $t\bar{t}$ and single top processes . . . 87
6.5.2.3 Di-boson process . . . 88
6.5.3 QCD background . . . 88
6.5.4 W+jets background . . . 89
6.6 Systematics . . . 89
6.7 Results and Interpretation . . . 91
6.8 Summary . . . 93

7 Search for new physics with high-mass tau pairs with 2016 data 94
7.1 Introduction . . . 94
7.2 Strategy . . . 95
7.3 Triggers . . . 96
7.3.1 Single Lepton Trigger Efficiency . . . 96
7.3.2 Single Electron Trigger Efficiency . . . 97
7.3.3 Single Muon Trigger Efficiency . . . 98
7.4 Data and Monte Carlo Samples . . . 98
7.4.1 Corrections for Pile-Up . . . 101
7.5 Muon + Hadronic Tau Channel . . . 103
7.5.1 Event selection . . . 103
7.5.2 Signal region selections . . . 104
7.5.3 Genuine dilepton events . . . 106
7.5.3.1 Drell-Yan process . . . 106
7.5.3.2 $t\bar{t}$ and single top processes . . . 107
7.5.3.3 Di-boson process . . . 107
7.5.4 QCD background . . . 108
7.5.5 W+jets background . . . 110

7.5.6 Validation of Background Estimations 112
7.6 Electron + Hadronic Tau Channel 112
7.6.1 Event selection 112
7.6.2 Genuine dilepton events 116
7.6.2.1 Drell-Yan process 116
7.6.2.2 $t\bar{t}$ and single top processes 116
7.6.2.3 Di-boson process 117
7.6.3 QCD background 117
7.6.4 W+jets background 118
7.6.5 Validation of Background Estimations 120
7.7 Electron + Muon Channel 120
7.7.1 Event selection 120
7.7.2 Genuine di-lepton events 122
7.7.2.1 Drell-Yan process 123
7.7.2.2 $t\bar{t}$ and single top processes 123
7.7.2.3 Di-boson process 124
7.7.2.4 Di-boson cross check 126
7.7.3 QCD background 127
7.7.4 W+jets background 127
7.7.5 Validation of Background Estimations 128
7.8 Systematics 131
7.9 Results and Interpretation 132
7.10 Summary 132

Bibliography **134**

A Selection Optimization **143**

List of Figures

1.1.1 The Standard Model of elementary particles [1] 2

2.1.1 The Large Hadron Collider [2] 17

2.1.2 The CERN Accelerator Complex [3] 18

2.1.3 Delivered Luminosity versus time for 2010, 2011, 2012, 2015 and 2016 (p-p data only) [4] 18

2.2.1 A cutaway view of the CMS detector [5] 20

2.2.2 Schematic view of a CMS quadrant [6] 21

2.2.3 Schematic cross section through the CMS tracker. Each line represents a detector module. Double lines indicate back-to-back modules that deliver stereo hits. [7] . . . 22

2.2.4 Layout of the CMS electromagnetic calorimeter showing the arrangement of crystal modules, supermodules and endcaps, with the preshower in front. [7] 23

2.2.5 Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters. [7] 24

2.2.6 An cross section of a quadrant of the CMS detector. The 4 drift tube (DT, in light orange) stations are labeled MB ("muon barrel") and the cathode strip chambers (CSC, in green) are labeled ME ("muon endcap"). Resistive plate chambers (RPC, in blue) are in both the barrel and the endcaps of CMS, where they are labeled RB and RE, respectively. [8] 25

3.4.1 Discriminator values for the CSVv2 algorithm. Underflow and overflow are added to the first and last bins, respectively. [9] 31

3.4.2 Performance of the b jet identification efficiency algorithms demonstrating the probability for non-b jets to be misidentified as b jet as a function of the efficiency to correctly identify b jets. The curves are obtained on simulated $t\bar{t}$ events using jets with $p_T > 30$ GeV, b jets from gluon splitting to a pair of b quarks are considered as b jets. [9] 32

4.2.1 A p-n junction in thermal equilibrium with zero-bias voltage applied [10]. 39
4.3.1 The schematics cross section (left) and top view (right) of the test structure used [11]. 42
4.4.1 The overview of the CV-IV measurement apparatus in the Brown Silicon Lab. 43
4.4.2 A close up shot of the measurement (left) and the light-tight box (right). 44
4.5.1 Illustration of the two methods used to determine the depletion voltage from CV measurements. On the left panel, $1/C^2$ versus V. On the right panel, log(C) versus log(V). At each voltage step, the capacitance measurement is performed twice with 1kHz (top plots) and 10 kHz (bottom plots). 45
4.6.1 Figure taken from Ref. [12] as the results of a collaborational work. Left: the increase in volume leakage current (corresponding to the depletion voltage) as a function of fluence after 800 MeV proton irradiation. Right: the effective doping concentration as a function of fluence after 800 MeV proton irradiation. 46
4.7.1 Figure taken from Ref. [12] as the results of a collaborational work. Hamburg model parameterization of FZ and DOFZ sensors. 49

5.5.1 Estimates of the rate of events with two fake τ_h's, after the mass windows cuts of 90 GeV $< m_{\tau\tau} <$ 150 GeV and 70 GeV $< m_{bb} <$ 150 GeV Left: 2 jet - 1 tag category; Right: 2 jet - 2 tag category. Each point represents the expected QCD yield predicted by a non-isolation region definition in events per 19.7/fb. Uncertainties shown are from control sample statistics. The rate estimate is independent of the choice of non-isolation region definition, within the statistical uncertainties. 58
5.6.1 The p_T distributions of the selected τ_h candidates with baseline selections. Overflows are included in the last bin. Top: leading τ_h's p_T; Bottom: subleading τ_h's p_T. Left: 2 jet - 0 tag category; middle: 2 jet - 1 tag category; right: 2 jet - 2 tag category. . . 60

5.7.1 Distributions of the reconstructed four-body mass with the kinematic fit after applying mass selections on $m_{\tau\tau}$ and m_{bb} in the $\tau_h\tau_h$ channel. The plots are shown for events in the 2 jet - 0 tag (top), 2 jet - 1 tag (middle), and 2 jet - 2 tag (bottom) categories. The expected signal scaled by a factor 10 is shown superimposed as an open dashed histogram for tanβ = 2 and m_H = 300 GeV in the low tanβ scenario of the MSSM. Expected background contributions are shown for the values of nuisance parameters (systematic uncertainties) obtained after fitting the background only hypothesis to the data. [13] 62

5.7.2 Distribution of the BDT discriminant for signal and background from the testing and training samples for m_H = 260 to 350 GeV in 10 GeV steps. 63

5.7.3 Background efficiency for BDTs trained using different sets of input variables for a fixed signal efficiency of 80%. 64

5.7.4 Distributions of the BDT discriminants from collider data and estimated backgrounds (histogram) with overflows included in the last bin. Left: 0-tag category; middle: 1-tag category; right: 2-tag category. The BDTs were trained at top row: m_H = 260 GeV, middle row: m_H = 270 GeV, bottom row: m_H = 280 GeV. The expected signal scaled by a factor 10 is shown superimposed as an open dashed histogram in the low tanβ scenario of the MSSM. Expected background contributions are shown for the values of nuisance parameters (systematic uncertainties) obtained after fitting the background only hypothesis to the data. 65

5.7.5 Distributions of the BDT discriminants from collider data and estimated backgrounds (histogram) with overflows included in the last bin. Left: 0-tag category; middle: 1-tag category; right: 2-tag category. The BDTs were trained at top row: m_H = 290 GeV, middle row: m_H = 300 GeV, bottom row: m_H = 310 GeV. The expected signal scaled by a factor 10 is shown superimposed as an open dashed histogram in the low tanβ scenario of the MSSM. Expected background contributions are shown for the values of nuisance parameters (systematic uncertainties) obtained after fitting the background only hypothesis to the data. 66

5.7.6 Distributions of the BDT discriminants from collider data and estimated backgrounds (histogram) with overflows included in the last bin. Left: 0-tag category; middle: 1-tag category; right: 2-tag category. The BDTs were trained at first row: $m_H = 320$ GeV, second row: $m_H = 330$ GeV, third row: $m_H = 340$ GeV, fourth row: $m_H = 350$ GeV. The expected signal scaled by a factor 10 is shown superimposed as an open dashed histogram in the low tanβ scenario of the MSSM. Expected background contributions are shown for the values of nuisance parameters (systematic uncertainties) obtained after fitting the background only hypothesis to the data. 67

5.9.1 Expected and observed upper limits comparison between kinematic fit approach (referred as CutBased) and BDT analyses, using the 0-tag, 1-tag and 2-tag categories. . 71

5.9.2 95% CL upper limits on cross section times branching fraction on $H \to hh \to \tau\tau bb$ in the combination of $\tau_h\tau_h$, $e\tau_h$ and $\mu\tau_h$ channels. 72

6.4.1 Distributions of the variables used for eτ_h signal selection, after all other signal selection requirements on variables other than the one plotted: $p_{\mathrm{T}}^{\mathrm{miss}}$ (top left), "ζ" (top right), $\cos\Delta\phi(e, \tau_h)$ (bottom left), and n_b (bottom right). 80

6.4.2 (Data driven QCD with only statistical uncertainties) Distributions of $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$. Left: validation region with $p_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV, $n_b = 0$ and $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right) < 200$ GeV. Right: validation region with $n_b \geq 1$. 81

6.4.3 QCD estimation and validation strategy for the $e\tau_h$ channel. 82

6.4.4 The distribution of reconstructed parent mass, $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$, in the same-charge, tightly-isolated sample: $e\tau_h$ channel. 83

6.4.5 Left: comparison of the simulated distributions of $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$, using tight τ_h isolation and with relaxed τ_h isolation, in the control region described in Section 6.4.4. Center: analogous comparison for the signal region. Right: comparison in the signal region of the "tighter" and "looser" halves of the relaxed isolation region. 84

6.5.1 (Data driven QCD with only statistical uncertainties) Distributions of the variables used for $e\mu$ signal selection, after all other signal selection requirements on variables other than the one plotted: $p_{\mathrm{T}}^{\mathrm{miss}}$ (top left), "ζ" (top right), $\cos\Delta\phi(e, \mu)$ (bottom left), and n_b (bottom right). 86

6.5.2 (Data driven QCD with only statistical uncertainties) Distributions of $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$. Left: validation region with $p_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV, $n_b = 0$ and $60 <$ GeV $m\left(e, \mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right) <$ 150 GeV. Right: validation region with $n_b \geq 1$. 87

6.5.3 The distribution of reconstructed parent mass, $m\left(e, \mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$, in the same-charge, tightly-isolated sample: $e\mu$ channel. 88

6.7.1 Observed $m\left(\tau_1, \tau_2, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ distribution in the signal region compared to the expected SM backgrounds for the (left) $e\tau_h$, and (right) $e\mu$ channels. The dashed histogram shows the distribution expected for a Z' boson with mass 1.5 TeV. The rightmost bins also include events with $m\left(\tau_1, \tau_2, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$, and are normalized to the displayed bin width. The lower panel shows the ratio of the observed number of events to the total background prediction. The shaded bands represent the total uncertainty in the background prediction [14]. 91

6.7.2 The observed 95% CL upper limits on the product of the cross section and branching fraction into τ lepton pairs as a function of the Z' mass m(Z') (solid black lines), for the (top left) $e\tau_h$, (top right) $e\mu$, and (bottom) for the combination of the four channels. The expected limits (dash-dotted lines) with one and two standard deviation (s.d.) uncertainty bands are also shown. The predictions of the NLO theory cross sections in the SSM and topcolor-assisted technicolor (TAT) models are represented by the red (lighter) and blue (darker) solid curves, respectively [14]. 92

7.3.1 The efficiency vs. p_T curves of the single-electron triggers used. Left column: $|\eta| < 1.479$. Right column: $|\eta| > 1.479$. 98

7.3.2 The efficiency vs. p_T curves of the single-muon triggers used. Top left: $|\eta| < 0.8$. Top right: $0.8 < |\eta| < 1.24$. Bottom: $|\eta| > 1.24$. 99

7.4.1 Distributions of the number of reconstructed primary vertices in the $e\mu$ channel. Left: before reweighting. Right: after reweighting. 102

7.5.1 (Data driven QCD and MC based W+jets with only statistical uncertainties) Left: Distribution of $\cos\Delta\phi(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$. Right: Distribution of $\cos\Delta\phi(\tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$ 105

7.5.2 With $\cos\Delta\phi(\tau_h, p_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ selection. Left: (Data driven QCD and MC based W+jets with only statistical uncertainties) Distribution of $m_{\mathrm{T}}(\mu, p_{\mathrm{T}}^{\mathrm{miss}})$ with $Z'(3000)$. Right: $m_{\mathrm{T}}(\mu, p_{\mathrm{T}}^{\mathrm{miss}})$ of $Z'(750)$, $Z'(1750)$ and $Z'(3000)$ 105

7.5.3 (Data driven W+jets and QCD with only statistical uncertainties) Distributions of the variables used for $\mu\tau_h$ signal selection, after all other signal selection requirements on variables other than the one plotted: $p_{\mathrm{T}}^{\mathrm{miss}}$ (left), $\cos\Delta\phi(\mu, \tau_h)$ (middle, with $\cos\Delta\phi(\mu, \tau_h) > 0$) and the number of b-tagged (CSV loose) jets, n_b (right). 106

7.5.4 (Data driven W+jets and QCD with only statistical uncertainties) Distributions of $m\left(\mu, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$. Left: validation region with $p_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV, $n_b = 0$ and $60 <$ GeV $m\left(\mu, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right) < 150$ GeV. Right: validation region with $n_b \geq 1$. 108

7.5.5 Data driven QCD and W+jets estimation and validation strategy for the $e\tau_h$, $\mu\tau_h$ channels. 109

7.5.6 (Only statistical uncertainties are included) Distributions of $m\left(\mu, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$. Left: in region B with signal region selections (data driven QCD and MC based W+jets). Right: in region A' with signal region like selections (data driven QCD and W+jets). 111

7.5.7 (Data driven W+jets and QCD with only statistical uncertainties) Distributions with $-0.95 < \cos\Delta\phi(\mu, \tau_h) < 0$ selection. Top: left: $m\left(\mu, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ right: $p_{\mathrm{T}}^{\mathrm{miss}}$. Bottom: left: muon p_T right: τ_h p_T. 113

7.5.8 (Data driven W+jets and QCD with only statistical uncertainties) $m\left(\mu, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ distribution with signal region selections. 114

7.6.1 (Data driven QCD with only statistical uncertainties) Distributions of the variables used for $e\tau_h$ signal selection, after all other signal selection requirements on variables other than the one plotted. Top row plots are made with data driven W+jets in distributions of: $p_{\mathrm{T}}^{\mathrm{miss}}$ (top left), $\cos\Delta\phi(e, \tau_h)$ (top middle, with $\cos\Delta\phi(e, \tau_h) > 0$) and n_b (top right). Bottom row plots are made with MC W+jets in distributions of: $\cos\Delta\phi(e, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$ (bottom left) and $\cos\Delta\phi(\tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$ (bottom right). 115

7.6.2 (Data driven W+jets and QCD with only statistical uncertainties) Distributions of $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$. Left: validation region with $p_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV, $n_b = 0$ and 60 GeV $<$ $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right) < 150$ GeV. Right: validation region with $n_b \geq 1$. 117

7.6.3 (Only statistical uncertainties are included) Distributions of $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$. Left: in region B with signal region selections (data driven QCD and MC based W+jets). Right: in region A' with signal region like selections (data driven QCD and W+jets). 119

7.6.4 (Data driven W+jets and QCD with only statistical uncertainties) Distributions with $-0.95 < \cos\Delta\phi(e, \tau_h) < 0$ selection. Top: left: $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ right: $p_{\mathrm{T}}^{\mathrm{miss}}$. Bottom: left: electron p_T right: τ_h p_T. 121

7.6.5 (Data driven W+jets and QCD with only statistical uncertainties) $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ distribution with signal region selections. 122

7.7.1 (Data driven QCD with only statistical uncertainties) Distributions of the variables used for $e\mu$ signal selection, after all other signal selection requirements on variables other than the one plotted: $p_{\mathrm{T}}^{\mathrm{miss}}$ (left), $\cos\Delta\phi(e,\mu)$ (middle, with $\cos\Delta\phi(e,\mu) > 0$) and n_b (right). 123

7.7.2 (Data driven QCD with only statistical uncertainties) Distributions of $m\left(e,\tau_h,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$. Left: validation region with $p_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV, $n_b = 0$ and $60 <$ GeV $m\left(e,\mu,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right) <$ 150 GeV. Right: validation region with $n_b \geq 1$. 124

7.7.3 (Data driven QCD with only statistical uncertainties) Left: $P_\zeta - 3.1 \times P_\zeta^{vis}$ distribution after requiring preselection, $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV and no jet with $p_T > 20$ GeV. Middle: $m\left(e,\mu,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ distribution after all the requirements mentioned in this section. Right: $m\left(e,\mu,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ distribution after all the requirements mentioned in this section. . . . 126

7.7.4 (Data driven QCD with only statistical uncertainties) Left: muon relative isolation with QCD estimated from same-sign region. Right: W+jets MC closure test between isolated and anti-isolated events. 128

7.7.5 (Data driven QCD with only statistical uncertainties) Distributions with $-0.95 < \cos\Delta\phi(e,\mu) < 0$ selection. Top: left: $m\left(e,\mu,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ right: $p_{\mathrm{T}}^{\mathrm{miss}}$. Bottom: left: electron p_T right: muon p_T. 129

7.7.6 (Data driven QCD with only statistical uncertainties) $m\left(e,\mu,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ distribution with signal region selections. 130

7.9.1 Expected limits for the $\mu\tau_h$, $e\tau_h$ and $e\mu$ channels. The leading order signal cross-section has been used. 133

A.0.1 (Data driven QCD and MC based W+Jets with only statistical uncertainties) Left: Distribution of $\cos\Delta\phi(\mu,p_{\mathrm{T}}^{\mathrm{miss}})$. Right: Distribution of $\cos\Delta\phi(\tau_h,\vec{p}_{\mathrm{T}}^{\mathrm{miss}})$ 144

A.0.2 Total background and signal distributions with different selections after requiring: $\cos\Delta\phi(e,\tau_h) < -0.95$; $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV; and no jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose). Top left: The total background distribution of $m\left(\mu,\tau_h,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$. Top right: $Z'(750)$ distribution of $m\left(\mu,\tau_h,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$. Bottom left: $Z'(1750)$ distribution of $m\left(\mu,\tau_h,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$. Bottom right: $Z'(3000)$ distribution of $m\left(\mu,\tau_h,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ 146

List of Tables

1.1 Basis modes and fit values for the 2016 fit to τ branching fraction data [15]. 3

1.2 Chiral supermultiplet fields in the MSSM 12

1.3 Gauge supermultiplet fields in the MSSM 12

3.1 Loose Jet-ID Selections. 30

3.2 Approximate branching fractions of different τ decay modes. Here, h^- represents a charged hadron (either a pion or a kaon) [16]. 32

3.3 Reconstructed Tau Decay Modes 33

3.4 MET Filters. 37

4.1 Design specifications of the LHC and the HL-LHC. 39

4.2 Basic properties of the test structures. 41

4.3 Fluence achieved for the test sensors with 800 MeV protons. 42

4.4 Values of fitting parameters for N_{eff} vs fluence. 47

5.1 List of dataset analyzed in the $\tau_h\tau_h$ channel. 53

5.2 List of Monte Carlo samples used to model $gg \to H \to hh \to \tau\tau bb$ signal as well as $Z/\gamma^* \to ee$, $Z/\gamma^* \to \mu\mu$, W + jets, $t\bar{t}$, single top and di-boson (WW, WZ, ZZ) backgrounds. 54

5.3 List of the embedded samples. 54

5.4 Trigger used in the $\tau_h\tau_h$ channel. 55

5.5 Expected event yields from standard model processes, and the observed event yield for the baseline selection and an integrated luminosity of 19.7 fb^{-1}. 59

5.6 Yields after the baseline selection and mass window requirements, per 19.7 fb^{-1}. . . 61

5.7 List of variables considered for multivariate analysis in order of decreasing discrimination power, as determined by the BDT for $m_H = 300$ GeV. For other masses the ranking varies somewhat. Those with an "x" in the second column were used in the final training. 64
5.8 Systematic uncertainties considered for $\tau_h\tau_h$ channel 68
5.9 Background bin by bin uncertainties. 70
5.10 Expected limits (in pb) at 95% CL, using the 0-tag, 1-tag and 2-tag categories. . . . 71
5.11 Observed limits (in pb), using the 0-tag, 1-tag and 2-tag categories. 71

6.1 Collision Data Samples 76
6.2 MC Signal Samples 76
6.3 MC Background Samples 77
6.4 Summary of systematic uncertainties. Values are given in percent. "s" indicates template variations ("shape" uncertainties). "p_T" indicates the τ_h p_T dependent uncertainty of 6% + 20%·(p_T/1 TeV) 91

7.1 The trigger paths used. 96
7.2 Collision Data Samples 100
7.3 MC Signal Samples 100
7.4 MC Background Samples 101
7.5 Event yields in regions B and F used for the calculation of $f_{\mathrm{LT}}^{\mathrm{QCD}}$. 110
7.6 Event yields in regions B and F used for the calculation of $f_{\mathrm{LT}}^{\mathrm{QCD}}$. 118
7.7 Event yields in isolated and anti-isolated regions used for the calculation of the "Loose-to-Tight" scale factor. 127
7.8 Summary of systematic uncertainties. Values are given in percent. "s" indicates template variations ("shape" uncertainties). "p_T" indicates the τ_h p_T dependent uncertainty of 6% + 20%·(p_T/1 TeV) 133

A.1 selections considered 145

Chapter 1

The Standard Model

The Standard Model (SM) is a collection of related theories which provides a mathematical framework in describing the nature of all known elementary particles. Based on one general principle, the requirement of local gauge invariance, the SM provides at the present time the most accurate description of the properties and interactions of elementary particles and has led to the discovery of the Higgs boson. In this chapter, we shall review: the elementary particles we have directly or indirectly detected; the gauge theories of the three forces described by the SM; and the Higgs mechanism that explains why some gauge bosons are allowed to be massive under the local gauge invariance assumption.

1.1 Elementary Particles of the Standard Model

Throughout centuries and across cultures humans have pondered the basic elements that formed our universe. From the five elements of the Chinese to the "Atom" of ancient Greek, theories were formed in hopes to capture the essence of nature. The first scientific elementary particles breakthrough came in the year 1897, with J.J. Thomson's discovery of the electron [17]. Since then, throughout the course of a little over 100 years, the "particle zoo" of our current understanding came to be organized into two groups, as summarized in Fig. 1.1.1: the fermions that make up matter and the bosons that mediate interactions.

1.1.1 Fermions

The fermions of the SM carries half integer spins and obey the Fermi-Dirac statistics. Based on their quantum characteristics, three groups, known as *generations*, of fermions have been discovered where

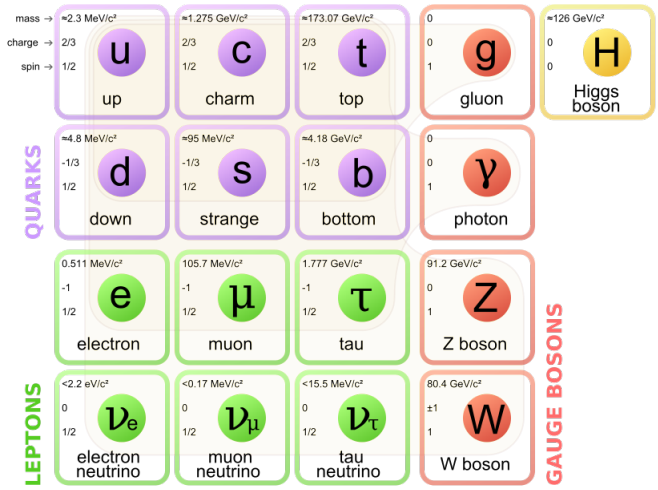

Figure 1.1.1: The Standard Model of elementary particles [1]

each generation is a heavier copy of the previous generation with almost all quantum numbers the same. Each fermion also has an antiparticle partner, not shown in Fig. 1.1.1, with all characteristics but charge the same. According to their interactions with the strong force, fermions are further separated into quarks, who feel the strong force, and leptons, who don't.

In 1962, Lev B. Okun introduced the term *hadron* to describe strongly interacting particles [18]. Three years later, in 1964, Gell-Mann and Zweig independently proposed that all hadrons are in fact composed of even more elementary constituents, which Gell-Mann called *quarks* [19]. There are two types of quarks in each generation: the up-type quark which carries $+2/3$ electric charge and the down-type quark which carries -1/3 electric charge. With only the quantum characteristics mentioned above came a dilemma: for example, the Σ^{++} is supposed to consist of three identical up quarks, which clearly violates the Pauli exclusion principle for fermions. O. W. Greenberg resolved this problem by introducing a new degree of freedom, later recognized as the *color charge*, for quarks [20]. Each quark is allowed to have one of the three colors charges (red, green and blue) while their antiparticles carry the anti-color charge such that the combination, e.g red + anti-red, is colorless. Hence, the dilemma of Σ^{++} is solved as each up quark can carry a different color charge. Up to this date, all bound state particles discovered either have zero color charge or the presence of all three colors in equal amounts. The most common quark bound states are: quark and anti-quark pairs (the mesons) and quark/antiquark triplets (the baryons and antibaryons). Exotic bound states such as pentaquarks have been recently discovered [21] but are much rarer compared to mesons and baryons.

Within each generation, there are two leptons: one massive carrying electric charge and one

massless, in the context of SM despite recent discoveries of neutrino oscillations [22], which only experiences weak interactions. The lightest of the generations of leptons is the electron and electron-neutrino pair which guaranties their stability and high abundances. The muon and muon-neutrino is next in the list with the muon being roughly 200 times the mass of the electron. Despite the muon being more massive than the up, down and even the strange quark, due to color confinement, the lightest possible hadronic decay product ($\pi^{\pm}$) is more massive than the muon. Thus, muons only exhibit decays into electrons + neutrinos or electron + multiple electron-positron pairs + neutrinos. Finally, the tau lepton from the heaviest generation is more than 16 times the mass of the muon and is the only lepton heavy enough to decay hadronically into mesons.

τ^- Decay Modes	Branching Fraction
$\mu\bar{\nu}_\mu\nu_\tau$	(17.39 ± 0.04) %
$e\bar{\nu}_e\nu_\tau$	(17.82 ± 0.04) %
hadronically	(64.79 ± 0.06) %

Table 1.1: Basis modes and fit values for the 2016 fit to τ branching fraction data [15].

As summarized in Table 1.1, approximately 1/3 of the tau leptons decay to an electron or muon with equal probability and the other 2/3 of the tau leptons decay into one or multiple mesons. This makes searches of physics involving tau leptons more complicated than those only looking for electrons or muons and will be discussed more in detail in Section 3.5.

1.1.2 Bosons

The SM describes three of the four known forces through the non-Abelian group:

$$\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y \quad , \tag{1.1.1}$$

where $\mathrm{SU}(3)_C$ is the color charge group describing strong interactions through quantum chromodynamics (QCD), and $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ combined describe the unified electroweak interactions.

Gauge bosons arising from these groups are spin 1 particles following the Bose-Einstein statistics. Eight massless strong interaction mediators, the *gluons*, emerge from the $\mathrm{SU}(3)_C$ group with no electric charge and a superposition of color charge. In addition to their interaction with quarks, gluons can also self-interact to form *glue-balls*. As the electroweak symmetry breaks under the Higgs

mechanism (as will be discussed in detail in Section 1.2.4):

$$\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y \to \mathrm{U}(1)_{\mathrm{EM}} \quad , \tag{1.1.2}$$

the force mediators of weak interactions, $W^{\pm}$ and Z, gain mass. The $W^{\pm}$ and Z bosons carry weak hypercharge and therefore interact with themselves and each other. The photon, mediator of the electromagnetic (EM) interaction, is a massless neutral particle that does not self-interact.

Finally, the Higgs boson, discovered in 2012 [23, 24], is a spin 0 scalar field needed by the SM to explain the spontaneous symmetry breaking and how particles obtained mass.

1.2 Standard Model Forces

The SM describes three of the four known forces of nature through the language of gauge theories, which originated from the existence of degrees of freedom in the description of elementary particle states that are indeterminate and have no effect on the predicted outcomes of any experiment [25]. In this section, we shall start with the abelian gauge theory with the symmetry group U(1), which describes the quantum electrodynamics (QED), followed by the non-Abelian gauge theory which describe the strong and electroweak forces .

1.2.1 The Quantum Electrodynamics

Consider the Dirac Lagrangian for a free fermion field $\psi(x)$:

$$\mathscr{L}_0 = \bar{\psi}(i\hbar c\partial\!\!\!/ - m)\psi \quad , \tag{1.2.1}$$

where $\partial\!\!\!/ = \gamma^\nu \partial_\nu$ and γ^ν are the gamma matrices. By imposing local gauge invariance, we hope to have $\mathscr{L} \to \mathscr{L}$, under transformation: $\psi \to e^{i\theta(x)}\psi$. It's easy to see that equation 1.2.1 does not satisfy this requirement, as it transforms like: $\mathscr{L}_0 \to \mathscr{L}_0 - \hbar c(\partial_\nu \theta)\bar{\psi}\gamma^\nu\psi$. The local gauge invariance can be achieved by adding a new term to the Lagrangian such that:

$$\mathscr{L}_{\text{improved}} = \bar{\psi}(i\hbar c\partial\!\!\!/ - m)\psi - (q\bar{\psi}\gamma^\nu\psi)A_\nu \quad , \tag{1.2.2}$$

where A_ν is some new field that transforms under local transformation like:

$$A_\nu \to A_\nu + \frac{\hbar c}{q}\partial_\nu \theta. \tag{1.2.3}$$

The introduction of parameter q may seem trivial here but we will soon discover it actually corresponds to the coupling constant, in this case the electric charge, between the fermion and new field A_ν.

As constructed, this improved Lagrangian $\mathscr{L}_{\text{improved}}$ is now locally invariant with the price of introducing a new vector field A_ν. However, $\mathscr{L}_{\text{improved}}$ only includes the interaction term between the fermion and the new field and lacks the free field term for A_ν. Thus, we turn to the Proca Lagrangian, which describes a free spin-1 vector field, for help:

$$\mathscr{L}_{\text{Proca}} = -\frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} + \frac{1}{8\pi}(\frac{m_A c}{\hbar})A^\nu A_\nu \quad , \tag{1.2.4}$$

where $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$. It's easy to see that $F^{\mu\nu}$ is locally invariant:

$$\begin{aligned} F^{\mu\nu} \rightarrow & \partial^\mu(A^\nu + \frac{\hbar c}{q}\partial^\nu\theta) - \partial^\nu(A^\mu + \frac{\hbar c}{q}\partial^\mu\theta) \\ &= F^{\mu\nu} + \frac{\hbar c}{q}(\partial^\mu\partial^\nu\theta - \partial^\nu\partial^\mu\theta) \\ &= F^{\mu\nu} \quad , \end{aligned} \tag{1.2.5}$$

while in the second term, $A^\nu A_\nu$, is not:

$$A^\nu A_\nu \rightarrow A^\nu A_\nu + \frac{\hbar c}{q}(A^\nu\partial_\nu + A_\nu\partial^\nu)\theta + \frac{\hbar^2 c^2}{q^2}\partial^\nu\partial_\nu\theta \quad . \tag{1.2.6}$$

Thus, in order to add the Proca Lagrangian for the description of field A_ν while maintaining local invariance, the field A_ν must be massless ($m_A = 0$). Finally, the complete Lagrangian for a fermion with local gauge invariance becomes:

$$\mathscr{L}_{\text{QED}} = \bar{\psi}(i\hbar c\partial\!\!\!/ - m)\psi - \frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} - (q\bar{\psi}\gamma^\nu\psi)A_\nu \quad . \tag{1.2.7}$$

We can take a step back and investigate the physical meanings within equation 1.2.7 and immediately notice the followings:

- Two interacting fields are described within this Lagrangian with the vector field required to be massless, which turns out to be the photon.
- The interaction strength is proportional to q, the electric charge of our fermion.
- This Lagrangian has no self-coupling terms of A^ν as photons do not carry charge.
- The last two terms reproduce the Maxwell Lagrangian with the current density $J^\nu = cq(\bar{\psi}\gamma^\nu\psi)A_\nu$.

Hence, the classical electromagnetism has been reproduced by imposing a local U(1) symmetry on the Dirac Lagrangian, since $\theta(x)$ is 1-dimensional. This formulation is also known as an Abelian gauge theory. As we will see in the next two sections, by taking similar approaches, the strong and electroweak interactions can be explained through similar, non-Abelian, gauge theories.

1.2.2 Yang-Mills Theory

In 1954, Yang and Mills extended the gauge principles of QED to the SU(2) group [26]. The basic idea follows closely our previous QED discussion but starts differently: Instead of a single Dirac field, we start with a Dirac field doublet such that:

$$\psi \equiv \begin{pmatrix} \psi_1(x) \\ \psi_2(x) \end{pmatrix} \quad . \tag{1.2.8}$$

The complete Yang-Mills Lagrangian of equal mass Dirac fields is [26]:

$$\mathscr{L} = \bar{\psi}(i\hbar c\partial\!\!\!/ - m)\psi - \frac{1}{16\pi}\mathbf{F}^{\mu\nu}\mathbf{F}_{\mu\nu} - (q\bar{\psi}\gamma^{\nu}\boldsymbol{\tau}\psi)\cdot\mathbf{A}_{\nu} \quad , \tag{1.2.9}$$

which looks identical to equation 1.2.7 but has a few nontrivial differences:

- $\boldsymbol{\tau} = (\sigma_1,\ \sigma_2,\ \sigma_3)$, where σ_i are 2×2 Pauli matrices.
- $\mathbf{A}^{\nu} = (A_1^{\nu},\ A_2^{\nu},\ A_3^{\nu})^{\mathrm{T}}$ is now a vector.
- $\mathbf{F}_{\mu\nu} = \partial^{\mu}\mathbf{A}^{\nu} - \partial^{\nu}\mathbf{A}^{\mu} - \frac{2q}{\hbar c}(\mathbf{A}^{\mu}\times\mathbf{A}^{\nu})$.

The additional term in the $\mathbf{F}_{\mu\nu}$ compared with QED is caused by the fact that SU(2) is non-Abelian (2×2 matrices do not commute, where as 1×1 matrices do). We will soon see that this term naturally gives rise to the self-coupling features of gluons, $W^{\pm}$ and Z bosons.

1.2.3 Quantum Chromodynamics

The elementary strong interaction, also known as quantum chromodynamics (QCD), is a gauge interaction between color-charged quarks. The symmetry group of the gauge transformation is given by $\mathrm{SU}(3)_C$, where C stands for color [25]. Similar to the discussion of SU(2) groups, given the

three possible colors for each quark, the triplets of the SU(3) group are [25]:

$$\begin{pmatrix} u_{\mathrm{r}} \\ u_{\mathrm{g}} \\ u_{\mathrm{b}} \end{pmatrix}, \begin{pmatrix} d_{\mathrm{r}} \\ d_{\mathrm{g}} \\ d_{\mathrm{b}} \end{pmatrix}, \begin{pmatrix} c_{\mathrm{r}} \\ c_{\mathrm{g}} \\ c_{\mathrm{b}} \end{pmatrix}, \begin{pmatrix} s_{\mathrm{r}} \\ s_{\mathrm{g}} \\ s_{\mathrm{b}} \end{pmatrix}, \begin{pmatrix} t_{\mathrm{r}} \\ t_{\mathrm{g}} \\ t_{\mathrm{b}} \end{pmatrix}, \begin{pmatrix} b_{\mathrm{r}} \\ b_{\mathrm{g}} \\ b_{\mathrm{b}} \end{pmatrix}, \tag{1.2.10}$$

where r = red, g = green and b = blue. Using the Yang-Mills theory, the complete Lagrangian for QCD for each quark is:

$$\mathscr{L} = \bar{\psi}(i\hbar c\partial\!\!\!/ - m)\psi - \frac{1}{16\pi}\mathbf{F}^{\mu\nu}\mathbf{F}_{\mu\nu} - (q\bar{\psi}\gamma^{\nu}\boldsymbol{\lambda}\psi)\cdot\mathbf{A}_{\nu} \quad , \tag{1.2.11}$$

where $\boldsymbol{\lambda} = (\lambda_1,\ \lambda_2,\ ...,\ \lambda_8)$ and λ_i are the 3×3 Gell-Mann matrices. Hence, $\mathbf{A}_{\nu}$ corresponds to the eight gluons and q represents the color-charge. $\mathbf{F}_{\mu\nu}$ has the same definition as in Section 1.2.2 and the term $-\frac{2q}{\hbar c}(\mathbf{A}^{\mu}\times\mathbf{A}^{\nu})$ indicates the self-coupling feature of gluons.

1.2.4 The Higgs Mechanism

A key ingredient when imposing the local gauge invariance is the requirement that newly introduced gauge fields are massless. While this is true for gluons and photons, the $W^{\pm}$ and Z bosons are certainly not massless. This issue was resolved with the idea of *Electroweak Symmetry Breaking* (EWSB) published almost simultaneously by three independent groups in 1964: by Robert Brout and François Englert [27]; by Peter Higgs [28]; and by Gerald Guralnik, C. R. Hagen, and Tom Kibble [29]. By introducing a scalar field with a non-zero vacuum expectation value into the Lagrangian, known as the Higgs mechanism, symmetries of the electroweak group are broken into the U(1) group of electromagnetism, while giving mass to the $W^{\pm}$ and Z bosons.

For simplicity, we will first demonstrate the spontaneous symmetry breaking through the Higgs mechanism under the U(1) local gauge invariance, and then expand the results for the EWSB in the next section.

The Klein-Gordon Lagrangian for a scalar field, $\psi = \psi_1 + i\psi_2$, with real fields ψ_1 and ψ_2 and mass m, is:

$$\mathscr{L} = \frac{1}{2}\partial_{\mu}\psi^{*}\partial^{\mu}\psi - \frac{1}{2}\left(\frac{mc}{\hbar}\right)^{2}(\psi^{*}\psi) \quad . \tag{1.2.12}$$

Thus, it's easy to see that the potential energy is a minimum when $\psi = 0$. Now, let's assume this

scalar field has a non-zero vacuum expectation value, thus the Lagrangian can be written as:

$$\mathscr{L} = \frac{1}{2}\partial_\mu\psi^*\partial^\mu\psi + \frac{1}{2}\mu^2(\psi^*\psi) - \frac{1}{4}\lambda^2(\psi^*\psi) \quad , \tag{1.2.13}$$

where the minima of the potential energy takes place at $|\psi| = \pm\mu/\lambda$.

Imposing local gauge invariance and performing similar tricks as in Section 1.2.1, a new vector field A^μ is introduced and the Lagrangian takes the form:

$$\mathscr{L} = \frac{1}{2}D^*_\mu\psi^* D^\mu\psi - \frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}\mu^2(\psi^*\psi) - \frac{1}{4}\lambda^2(\psi^*\psi)^2 \quad , \tag{1.2.14}$$

where $D_\mu = \partial_\mu + i\frac{q}{\hbar c}A_\mu$ and $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$.

Since the Lagrangian is locally gauge invariant, we can simplify our problem and look at the special gauge transformation such that $\psi \to \psi' \equiv \psi'_1$, or ψ' is a real field. Given the new potential minima being $\pm\mu/\lambda$, we can also rewrite the Lagrangian in terms of a perturbation around the minima, $\eta = \psi'_1 - \mu/\lambda$. The final Lagrangian is:

$$\begin{aligned}\mathscr{L} = &\left[\frac{1}{2}(\partial_\mu\eta)(\partial^\mu\eta) - \mu^2\eta^2\right] + \left[-\frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}\left(\frac{q}{\hbar c}\frac{\mu}{\lambda}\right)^2 A_\mu A^\mu\right] \\ &+ \left[\frac{\mu}{\lambda}\left(\frac{q}{\hbar c}\right)^2\eta(A_\mu A^\mu) + \frac{1}{2}\left(\frac{q}{\hbar c}\right)^2\eta^2(A_\mu A^\mu) - \lambda\mu\eta^3 - \frac{1}{4}\lambda^2\eta^4\right] \\ &+ \left(\frac{\mu^2}{2\lambda}\right)^2 \quad .\end{aligned} \tag{1.2.15}$$

The first square bracket term describes the massive scalar field η, which is the Higgs field in U(1) gauge symmetry. The second square bracket term describes the new vector field A^μ, which is introduced by the local gauge invariance requirement, this time with mass: $m_A = 2\sqrt{\pi}(q\mu)/(\lambda c^2)$. The third square bracket term describes the interactions between the scalar field η and vector field A^ν. The final term is just a constant.

Thus, with the help of a scalar field with non-zero minima, the vector fields introduced by local gauge invariance obtain mass while maintaining the required invariance.

1.2.5 The Eelectroweak Unification

The weak force is a short range force mediated by the $W^\pm$ and Z bosons. The fact that $W^\pm$ carries electric charge suggests the QED and the weak force are somehow related. The idea of unifying the electroweak interaction was first realized by Glashow in 1961 [30], who proposed an extended model with a larger symmetry group, SU(2)×U(1), and a fourth gauge boson Z [31]. However, within this

model, the $W^{\pm}$ and Z bosons are still required to be massless, which contradicts the fact that the weak force is a short range force. The full story of electroweak interactions and the massiveness of $W^{\pm}$ and Z bosons was finally explained in 1967-68 when the Galshow-Salam-Weinberg theory applied the symmetry breaking mechanism, as described earlier, to the SU(2)×U(1) group.

Electroweak interactions are described through the $\mathrm{SU}_L(2) \times \mathrm{U}_Y(1)$ model with the Lagrangian written as:

$$\mathscr{L}_{\text{Electroweak}} = \mathscr{L}_{\text{gauge}} + \mathscr{L}_{\text{scalar}} + \mathscr{L}_{\text{fermion}} + \mathscr{L}_{\text{Yukawa}} \quad . \tag{1.2.16}$$

The gauge term is:

$$\mathscr{L}_{\text{gauge}} = -\frac{1}{4} W^{\mu\nu i} W^i_{\mu\nu} - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} \quad , \tag{1.2.17}$$

which describes the kinematics and self interactions of the gauge fields $W^i_\nu (i = 1, 2, 3)$ and B_ν of gauge group $\mathrm{SU}_L(2)$ and $\mathrm{U}_Y(1)$. The field strength tensors are defined as:

$$\begin{aligned} W^i_{\mu\nu} &= \partial_\mu W^i_\nu - \partial_\nu W^i_\mu - g\epsilon_{ijk} W^j_\mu W^k_\nu \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \quad , \end{aligned} \tag{1.2.18}$$

where g is the $\mathrm{SU}_L(2)$ gauge coupling and ϵ_{ijk} is the antisymmetric symbol. B is a $\mathrm{U}_Y(1)$ field associated with the weak hypercharge $Y = Q - T^3$, where Q and T^3 are respectively the electric charge operator and the third component of the $\mathrm{SU}_L(2)$. It has no self-interactions. As will be shown, the B and W^i fields will eventually mix to form the photon, the $W^{\pm}$ and the Z bosons [32].

The fermion term is:

$$\mathscr{L}_{\text{fermion}} = \sum_{m=1}^{\#\text{families}} i(\bar{q}_{mL} \not{D} q_{mL} + \bar{l}_{mL} \not{D} l_{mL} + \bar{u}_{mR} \not{D} u_{mR} + \bar{d}_{mR} \not{D} d_{mR} + \bar{e}_{mR} \not{D} e_{mR}) \quad , \tag{1.2.19}$$

which describes the kinematics of the fermions with left (right) chiral projections. The left-handed quarks and leptons (q_{mL}, l_{mL}) transform as a SU(2) doublet while the right-handed fermions (u_{mR}, d_{mR} and e_{mR}) are singlets.

The Yukawa term describes the interactions between the scalar field, the Higgs field, and the fermions.

Finally, the scalar term, which is responsible for the spontaneous symmetry breaking, takes the form:

$$\mathscr{L}_{\text{scalar}} = (D^\mu \phi)^\dagger D^\mu \phi - V(\phi) \quad , \tag{1.2.20}$$

where ϕ is a complex scalar field and a doublet under the SU(2) group with the form:

$$\phi = \begin{bmatrix} \phi^+ \\ \phi^0 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \phi_1 - i\phi_2 \\ \phi_3 - i\phi_4 \end{bmatrix} \quad . \tag{1.2.21}$$

Due to local gauge invariance requirements, the covariant derivatives are:

$$D_\mu = \partial_\mu + \frac{ig}{2}\boldsymbol{\tau} \cdot W_\mu + \frac{ig'}{2} B_\mu \quad , \tag{1.2.22}$$

where $\boldsymbol{\tau} = (\sigma_1,\ \sigma_2,\ \sigma_3)$, with σ_i as 2×2 Pauli matrices, and g' is the U(1) gauge coupling.

$V(\phi)$ is the potential of the scalar field and takes the form:

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \quad . \tag{1.2.23}$$

Following a similar approach as in Section 1.2.4, we apply a special gauge transformation such that:

$$\phi \to \phi' = \begin{bmatrix} 0 \\ \nu + \mathrm{H} \end{bmatrix} \quad , \tag{1.2.24}$$

where the H is a Hermitian filed which will turn out to the be physical Higgs scalar. Thus, the covariant kinetic energy term becomes:

$$(D^\mu \phi')^\dagger D^\mu \phi' = M_W^2 W^{+\mu} W_\mu^- + \frac{M_Z^2}{2} Z^\mu Z_\mu + \text{ H terms} \quad , \tag{1.2.25}$$

where $M_W = g\nu/2$, $M_Z = \nu\sqrt{g^2 + g'^2}/2$, $W^\pm = (W^1 \mp iW^2)/\sqrt{2}$ and $Z = (-g'B + gW^3)/\sqrt{g^2 + g'^2}$. Thus, the symmetry spontaneously breaks and the $W^\pm$ and Z bosons obtain mass.

After the symmetry breaking, the scalar potential term becomes:

$$V(\phi) = -\frac{\mu^4}{4\lambda} - \mu^2 H^2 + \lambda \nu H^3 + \frac{\lambda}{4} H^4 \quad . \tag{1.2.26}$$

The mass of the scalar field, the Higgs mass, can be obtained as:

$$M_h = \sqrt{2\lambda}\nu \quad . \tag{1.2.27}$$

The weak scale ν can be estimated from the masses of the $W^\pm$ and Z bosons but the quadratic Higgs coupling λ is unknown. Thus, the Higgs mass is not predicted by the SM.

1.3 The Hierarchy Problem

In 2012, a Higgs boson was discovered by ATLAS [24] and CMS [23] with a mass of 125.09 ± 0.21 (stat) ± 0.11 (syst) GeV [33]. For the first time all relevant SM parameters are determined with remarkable accuracy [34]. However, if we were to extend the SM theories up to the Planck scale ($O(10^{19})$ GeV) where the quantum field theories meet general relativity, some SM parameters must be fine-tuned to specific values to match the experimental observed values. This problem, sometimes leading to requiring parameters with an accuracy of one part in 10^{34}, is known as the hierarchy problem.

An example of the hierarchy problem can be seen in the Higgs mass calculation. Additional to the bare mass term we calculated in equation 1.2.27, the Higgs also receives quantum loop corrections from other massive particles. The largest of these corrections comes from the quadratic coupling with the most massive SM particle, the top quark.

$$M_h^2 \sim 2\lambda\nu^2 + O(\lambda_t)\Lambda^2 \quad , \tag{1.3.1}$$

where λ_t is the coupling constant between the Higgs and the top quark and $\Lambda^2 \propto \int (1/k) d^4k$, which integrates up to the cut off energy Λ. If no new physics exists beyond the SM, Λ is free to run all the way up to the Planck scale where a great deal a fine-tuning, due to $\Lambda_{\mathrm{p}}^2/M_h^2 \sim 10^{34}$, is required.

1.4 Beyond the Standard Model

In order to resolve the hierarchy problem, many theories beyond the SM have been proposed. For example, Supersymmetry (SUSY) theories, Little Higgs (LH) model, the Left-Right (LR) symmetric models, the Grand Unified Theory (GUT) and the Sequential Standard Model (SSM).

1.4.1 Minimal Supersymmetric Standard Model

One solution to the hierarchy problem comes in the form of SUSY [35, 36]. As fermions and bosons loops carry opposite signs in the Higgs mass calculation, the divergence term would be automatically canceled if there exists a boson partner of the top quark with a similar coupling to the Higgs. In SUSY theories, each SM particle has a *superpartner* of similar quantum numbers but different spin statistics denoted with an additional " ˜ ". The two superpartners are joined together to form *supermultiplets*. For each fermion, there exists a scalar fermion (*sfermion*) with spin-0. For each gauge boson, there exists a *gaugino* of spin-1/2. One simple form of SUSY is known as the minimal

supersymmetric extension to the standard model (MSSM) [37,38], which predicts two Higgs doublets opposed to the single Higgs doublet expected in the SM.

Table 1.2: Chiral supermultiplet fields in the MSSM

Name		spin-0	spin-1/2
squarks, quarks	Q	$(\tilde{u}_{\rm L}, \tilde{d}_{\rm L})$	$(\tilde{u}_{\rm L}, \tilde{d}_{\rm L})$
($\times$ 3 families)	$\bar{u}$	$\tilde{u}^*$	$u^\dagger_{\rm R}$
	$\bar{d}$	$\tilde{d}^*$	$d^\dagger_{\rm R}$
sleptons, leptons	L	$(\tilde{\nu}_{e\rm L}, \tilde{e}_{\rm L})$	$(\tilde{\nu}_{e\rm L}, \tilde{e}_{\rm L})$
($\times$ 3 families)	$\bar{e}$	$\tilde{e}^*$	$e^\dagger_{\rm R}$
higgs, higgsinos	$H_{\rm u}$	$(H^+_{\rm u}, H^0_{\rm u})$	$(\tilde{H}^+_{\rm u}, \tilde{H}^0_{\rm u})$
	$H_{\rm d}$	$(H^0_{\rm d}, H^-_{\rm d})$	$(\tilde{H}^0_{\rm d}, \tilde{H}^-_{\rm d})$

Table 1.3: Gauge supermultiplet fields in the MSSM

Name	spin-1/2	spin-1
gluinos, gluons	$\tilde{g}$	g
winos, W bosons	$\tilde{W}^\pm\ \tilde{W}^0$	$W^\pm\ W^0$
binos, B bosons	$\tilde{B}^\pm\ \tilde{B}^0$	$B^\pm\ B^0$

The chiral and gauge supermultiplets of the MSSM are listed in Table 1.2 and 1.3. After the breaking of electroweak symmetry, the W^0, B^0 gauge eigenstates mix to give mass eigenstates Z^0 and γ. Similarly, their superpartners $\tilde{W}^0$ and $\tilde{B}^0$ mix to give rise to the zino ($\tilde{Z}^0$) and photino ($\tilde{\gamma}^0$).

Five physical Higgs bosons arise from the two Higgs doublets, two of which are charged $H^\pm$, three neutral Higgs bosons are a light scalar Higgs h, a heavy scalar Higgs H and a pseudo-scalar Higgs A. All the parameters in the Higgs sector at tree level are determined by two free parameters, conventionally chosen to be the mass of the pseudo-scalar Higgs, m_A, and the ratio of vacuum expectation values of the two Higgs doublets, $\tan\beta \equiv v_u/v_d$. Thus, the mass of the Higgs at tree level are given as [39]:

$$\begin{aligned} m^2_{H^\pm} &= m^2_A + m^2_W \\ m^2_{h^0} &= \frac{2m^2_Z m^2_A \cos^2 2\beta}{m^2_A + m^2_Z + \sqrt{(m^2_A + m^2_Z)^2 - 4m^2_Z m^2_A \cos^2 2\beta}} \\ m^2_{H^0} &= \frac{2m^2_Z m^2_A \cos^2 2\beta}{m^2_A + m^2_Z - \sqrt{(m^2_A + m^2_Z)^2 - 4m^2_Z m^2_A \cos^2 2\beta}} \quad . \end{aligned} \tag{1.4.1}$$

The mixing angle α between two neutral scalar fields h^0 and H^0 is expressed as:

$$\cos^2(\beta - \alpha) = \frac{m^2_{h^0}(m^2_Z - m^2_{h^0})}{m^2_A(m^2_{H^0} - m^2_{h^0})} \quad . \tag{1.4.2}$$

In the "decoupling limit" $m_A \gg m_Z, \sin(\beta-\alpha) \sim 1, \cos(\beta-\alpha) \sim 0$. Then h^0 is light and SM-like, while all the other Higgs bosons are heavy, nearly degenerate, and the H^0 coupling to W^+W^-, ZZ is highly suppressed [39].

1.4.2 Little Higgs Model

The LH theories are based on an old idea to stabilize the Higgs mass by making the Higgs a pseudo-Goldstone boson resulting from a spontaneously broken approximate symmetry [40]. In the simplest model, the SM $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ gauge group is enlarged to $\mathrm{SU}(3) \times \mathrm{U}(1)$. The symmetry breaking of $\mathrm{SU}(3) \times \mathrm{U}(1) \rightarrow \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ gives rise to the SM Higgs. Accompanying the SU(3) gauge group, new top partners are introduced that directly cancel the top divergence term in the Higgs mass calculation. With the additional symmetry, the new model predicts extra gauge fields that consist of the Z' boson and a complex SU(2) doublet (Y^0, X^-).

1.4.3 Left-Right Symmetric Model

The asymmetry between left and right chirality in the SM has puzzled physicists for more than half a century. One idea to restore parity at high energies is provided by LR symmetric theories. The minimal LR symmetric theory is based on the following gauge group (suppressing color) [41]:

$$G_{LR} = SU(2)_L \times SU(2)_R \times U(1)_{\tilde{Y}} \quad , \tag{1.4.3}$$

where $\tilde{Y} = B - L$ with B as the baryon number and L the overall lepton number.

The quarks and leptons are now completely LR symmetric and the Higgs sector consists of a bi-doublet Φ and $SU(2)_{L,R}$ triplets Δ_L and Δ_R [41]:

$$\Phi = \begin{bmatrix} \phi_1^0 & \phi_2^+ \\ \phi_1^- & \phi_2^0 \end{bmatrix}, \quad \Delta_{L,R} = \begin{bmatrix} \Delta^+/\sqrt{2} & \Delta^{++} \\ \Delta^0 & -\Delta^+/\sqrt{2} \end{bmatrix}_{L,R} \quad . \tag{1.4.4}$$

The symmetry breaking from G_{LR} to the SM model takes the following form [41]:

$$\langle \Delta_L \rangle = 0, \langle \Delta_R \rangle = \begin{bmatrix} 0 & 0 \\ v_R & 0 \end{bmatrix} \quad . \tag{1.4.5}$$

The next step takes place when the symmetry of $SU(2)_L \times U(1)_Y$ breaks down to $U(1)_{em}$ as the

neutral component Φ develops a vacuum expectation value (VEC) [41].

$$\langle\Phi\rangle = \begin{bmatrix} v_1 & 0 \\ 0 & v_2 e^{i\alpha} \end{bmatrix} \quad , \tag{1.4.6}$$

where $M_W^2 = g^2(v_1^2 + v_2^2)$ and v_1 and v_2 are real and positive. This model predicts that there are more than one Higgs boson or the Higgs boson is actually a composite particle.

Additional gauge bosons arise from the symmetry breaking of G_{LR} with their masses [41]:

$$\begin{aligned} M_{W_R}^2 &\simeq g^2 v_R^2 \\ M_{Z_R}^2 &\simeq 2(g^2 + g_{\tilde{Y}}^2) v_R^2 \quad , \end{aligned} \tag{1.4.7}$$

when neglecting the mixings of v_1 and v_2.

1.4.4 E_6 Grand Unified Theory

With the unification of electromagnetism and weak interactions through electroweak theory, physisits have pondered ways to further unify the strong interactions. A plethora of models has been proposed under the name of the grand unified theory (GUT). The E_6 model stands out of the bunch with its connections to the $E_8 \otimes E_8'$ string theory, which is much more ambitious at unifying all forces including gravity. In the E_6 model, each family of fermions is put into a fundamental representation, a 27-plet [42]. Each family includes the SM fermions: $u\bar{u}, d, \bar{d}, e^{\pm}$ and ν_e, as well as some new particles: a right-handed neutrino $\bar{N}$, an isosinglet charge -1/3 quark h and its antiparticle, two lepton isodoublets (ν_E, E^-) and ($\bar{N}_E$, E^+) and a neutral isosinglet lepton n.

Considering the following symmetry breaking from E_6 to the SM [42].:

$$\begin{aligned} E_6 &\to \mathrm{SO}(10) \times \mathrm{U}(1)_\psi \\ &\to \mathrm{SU}(5) \times \mathrm{U}(1)_\chi \times \mathrm{U}(1)_\psi \\ &\to \mathrm{SU}(3)_c \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_\psi \times \mathrm{U}(1)_\chi \times \mathrm{U}(1)_\psi \quad . \end{aligned} \tag{1.4.8}$$

Extra neutral gauge bosons arise through the combination of Z_χ and Z_ψ as [42].:

$$Z'(\theta) = Z_\psi \cos\theta + Z_\chi \sin\theta \quad , \tag{1.4.9}$$

where θ is the mixing angle. There are three particular cases which people are mostly interested in:

- $Z' = Z'(0) = Z_\psi \quad ,$
- $Z' = Z'(\pi/2) = Z_\chi \quad ,$
- $Z' = Z'(\tan^- 1(-\sqrt{5/3})) = Z_\eta \quad .$

1.4.5 Sequential Standard Model

As described in the sections above, many extensions to the SM predict new gauge bosons. The sequential standard model (SSM) [43] is a minimal extension to the SM by assuming the new massive $V^\pm$ and V^0 gauge bosons with the same couplings terms as $W^\pm$ and Z in the SM. This not only refers to the coupling to quarks and leptons but also includes $V^0W^+W^-$ and $V^\pm W^\mp Z$ vertices. The decay width into WW, WZ and fermions are given as [43]:

$$\begin{aligned}
\Gamma(V^\pm \to ZW^\pm) =& \frac{\alpha}{48}\cot^2\theta_w M_{V^\pm}\frac{M_{V^\pm}^4}{M_Z^2 M_W^2}\left[\left(1-\frac{M_Z^2-M_W^2}{M_{V^\pm}^2}\right)^2 - 4\frac{M_W^2}{M_{V^\pm}^2}\right]^{3/2} \\
&\cdot\left[1+10\left(\frac{M_Z^2+M_W^2}{M_{V^\pm}^2}\right)+\frac{M_W^4+M_Z^4+10M_W^2M_Z^2}{M_{V^\pm}^4}\right] \\
\Gamma(V^0 \to W^+W^-) =& \frac{\alpha}{48}\cot^2\theta_w M_{V^0}\left(\frac{M_{V^0}}{M_W}\right)^4\left(1-4\frac{M_W^2}{M_{V^0}^2}\right)^{3/2} \\
&\cdot\left[1+20\left(\frac{M_{V^0}}{M_W}\right)^2+12\left(\frac{M_{V^0}}{M_W}\right)^4\right] \\
\Gamma(V^\pm \to f\bar{f}') =& \frac{\alpha}{12}N_c\frac{M_{V^\pm}}{\sin^2\theta_w} \\
\Gamma(V^0 \to f\bar{f}) =& \frac{\alpha}{48}N_c\frac{M_{V^0}}{\sin^2\theta_w\cos^2\theta_w}\left[1+(1-4|Q_f|\sin^2\theta_w)^2\right] \quad ,
\end{aligned} \tag{1.4.10}$$

where $N_c = 1$ for fermions, $N_c = 3$ for quarks and Q_f is the electric charge of the fermion.

Compared to other extended gauge models, with the assumption of having the same coupling to fermion as SM bosons, the SSM model generally has a larger production cross-section. Despite the fact the SSM may not be gauge invariant, the SSM serves as a benchmark for most of the extra gauge boson searches.

Chapter 2

The Compact Muon Solenoid Experiment at the Large Hadron Collider

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the world's largest and most powerful particle accelerator built in search of the now found Higgs boson and new physics beyond the standard model. Located roughly 100 meters beneath the France-Switzerland border, as shown in Fig 2.1.1, the LHC consists of a 27 km circumference ring housing over a thousand superconducting magnets and a number of acceleration structures. Additional to the magnets used to direct the beams along the accelerator, hundreds of multipole magnets are used to "squeeze" the particles in the beam closer together to increase the chance of collision. This task demands such accuracy akin to firing two needles 10 km apart and hoping that they meet halfway [44].

The LHC is designed to accelerate two opposite circulating proton beams up to 7 TeV and collide them at a center-of-mass energy of up to 14 TeV. Protons in each beam are grouped together into 2808 bunches, with around one hundred billion protons in each bunch. Bunches are spaced apart by 25 ns, resulting a collision rate of 40 MHz. The LHC's design instantaneous luminosity is 10^{34} cm^{-2}s^{-1} and was first reached in June 2016.

Prior to entering the LHC acceleration ring, protons are accelerated through a series of accel-

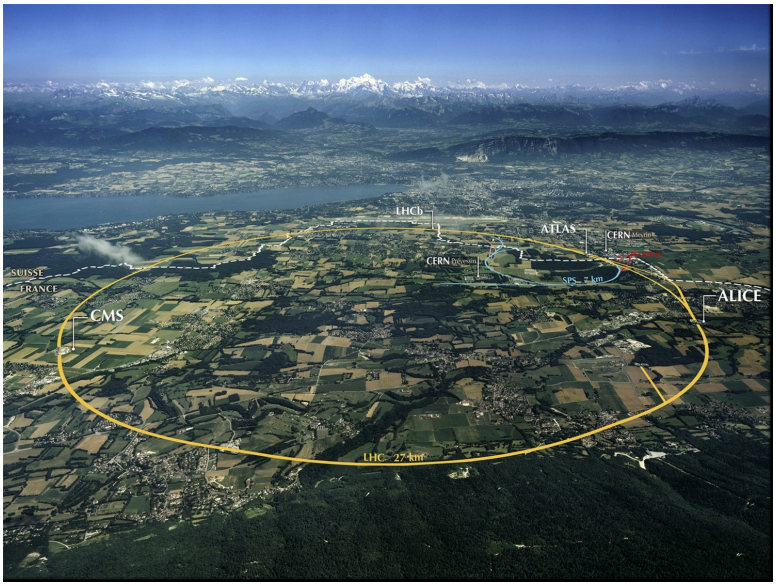

Figure 2.1.1: The Large Hadron Collider [2]

erators shown in Fig. 2.1.2. Protons are produced by removing electrons off hydrogen atoms with an electric field. They then enter the Linear accelerator 2 (Linac 2) where they're accelerated up to 50 MeV. These protons are then fed to the Proton Synchrotron Booster (labeled BOOSTER in the figure) where they're further boosted to 1.4 GeV before being injected into the Proton Synchrotron (PS). The PS then accelerates the protons up to 25 GeV with its 628 meter ring. The final acceleration before entering the LHC ring comes with the help from the Super Proton Synchrotron (SPS). With nearly 7 km circumference and the second-largest machine in CERN's accelerator complex, the SPS further pushes the protons to 460 GeV. Bunches of these ∼0.5 TeV protons finally enter the LHC ring, where it then takes of around 20 mins for them to reach 6.5 TeV. The two beams are brought into collision at four detectors (labeled by yellow circles) : the Compact Muon Solenoid (CMS), A Toroidal LHC Apparatus (ATLAS), A Large Ion Collider Experiment (ALICE), and Large Hadron Collider beauty (LHCb) [44]. CMS and ATLAS are two general purpose detectors with almost full detector coverage around the collision point. They are designed for various physics studies at the highest LHC luminosity. ALICE on the other hand is optimized for heavy-ion collisions and focuses on the physics of strongly interacting matter at extreme energy densities. The LHCb, unlike CMS, ATLAS and ALICE experiments, uses a series of sub-detectors to detect particles along the beam line and focuses on B mesons physics.

The LHC went live in September 2008 but soon experienced a magnet quench incident that delayed the project by more than a year. Following the accident, a decision was made to start the

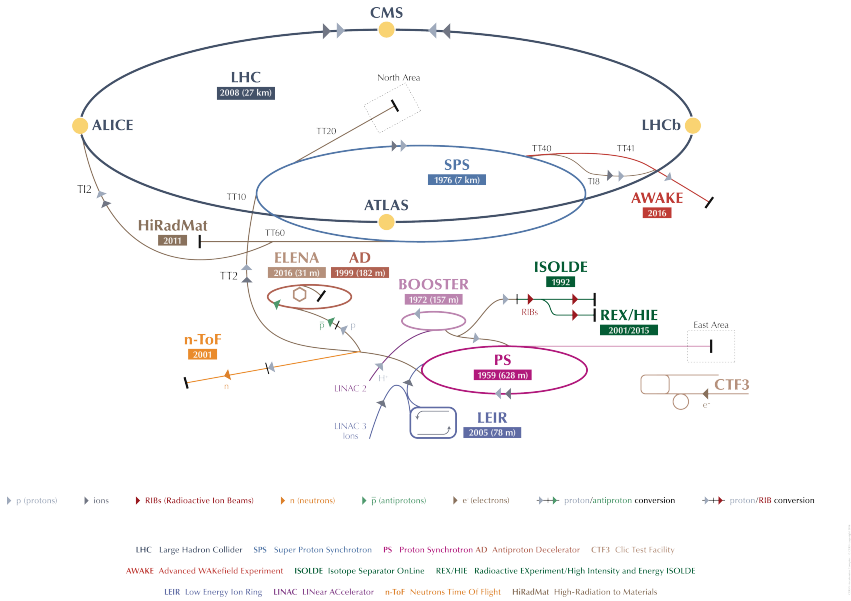

Figure 2.1.2: The CERN Accelerator Complex [3]

LHC at a lower collision energy and to proceed to the designed 14 TeV only after operating for several years at lower energies. As shown in Fig. 2.1.3, the first physics run of the LHC took place between 2010 to 2012 collecting a total of 6.146 fb^{-1} with 7 TeV collision data and 23.3 fb^{-1} with 8 TeV collisions. During a shutdown between 2013 to 2015, detectors and the accelerator complex received repairs and safety improvements to enable collision at 13 TeV. The second run began in 2015 with protons racing to 6.5 TeV in each beam and reaching the center-of-mass energy of 13 TeV. Since then, a combined 45.3 fb^{-1} of p-p collision data has been delivered by the LHC.

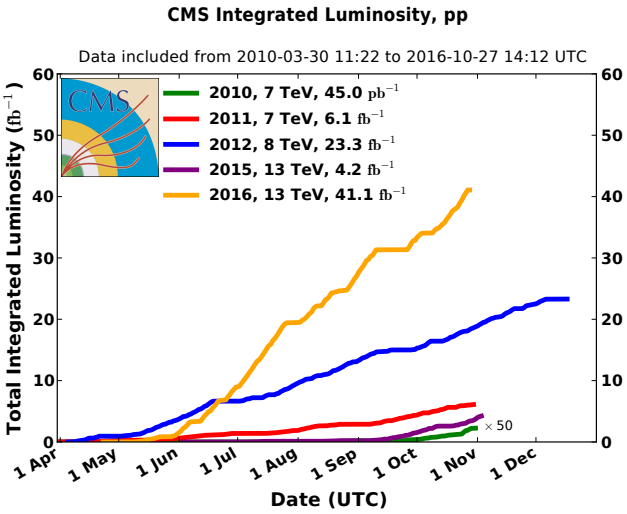

Figure 2.1.3: Delivered Luminosity versus time for 2010, 2011, 2012, 2015 and 2016 (p-p data only) [4]

2.2 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) detector is a general-purpose detector operating at the LHC at CERN [7]. To meet the goals of LHC physics, CMS was designed with the following requirements in mind: good muon property measurement (energy, momentum, charge, etc); good charged particle momentum measurement, high reconstruction efficiency and good energy resolution measurements. Thus, the key component in the design of CMS is the choice of magnetic field configuration. As shown in Fig. 2.2.1, weighing 14,000 tonnes, CMS is constructed with multiple layers of sub-detectors and a superconducting solenoid at its heart. The solenoid delivers 3.8 T of magnetic field when operating with 18,160 A of current.

Upon collision, exiting particles pass through a series of detectors build to measure different features of these particles. Detectors that sit closest to the collision point are the silicon trackers, which specialize in measuring the momentum of charged particles. Further from the intersection point is the electromagnetic calorimeter (ECAL), where most of the photons and electrons stop and deposit their energies. The hadron calorimeter (HCAL) awaits those particles, mostly hadrons, that penetrate through the ECAL and measures their energies. Finally, for muons, which do not interact much with the ECAL and HCAL, an additional set of muon chambers, resting outside the superconducting solenoid, are built to measure their trajectories. We will dive into the details of these detectors in the following sections.

The coordinate system adopted by CMS, which will be used throughout this thesis, has the origin centered at the nominal collision point. The z-axis points along the beam direction toward the Jura mountains from LHC Point 5 [7]. The y-axis points vertically upward and the x-axis points radially inwards toward the center of the LHC ring. The azimuthal angle ϕ is measured in the x-y plane, also known as the transverse plane, from the x-axis. The polar angle θ is measured from the z-axis. Despite all efforts on focusing the beams before collision, on an event level, the rest frames of each collision will likely have different boosts along the z-axis. To remove the longitudinal boost dependence of θ, particle physicists use rapidity y, defined as:

$$y \equiv \frac{1}{2} \ln \left(\frac{E + p_{\mathrm{L}}}{E - p_{\mathrm{L}}} \right) \tag{2.2.1}$$

For particles with momentum in the transverse plane (p_{T}) much larger than their invariant mass, $y \sim \eta$, with $\eta \equiv -\ln[\tan(\theta/2)]$. Hence, due to its simplicity in calculation and the fact that most of the outgoing particles are relativistic, we use η to denote the angular separation between the

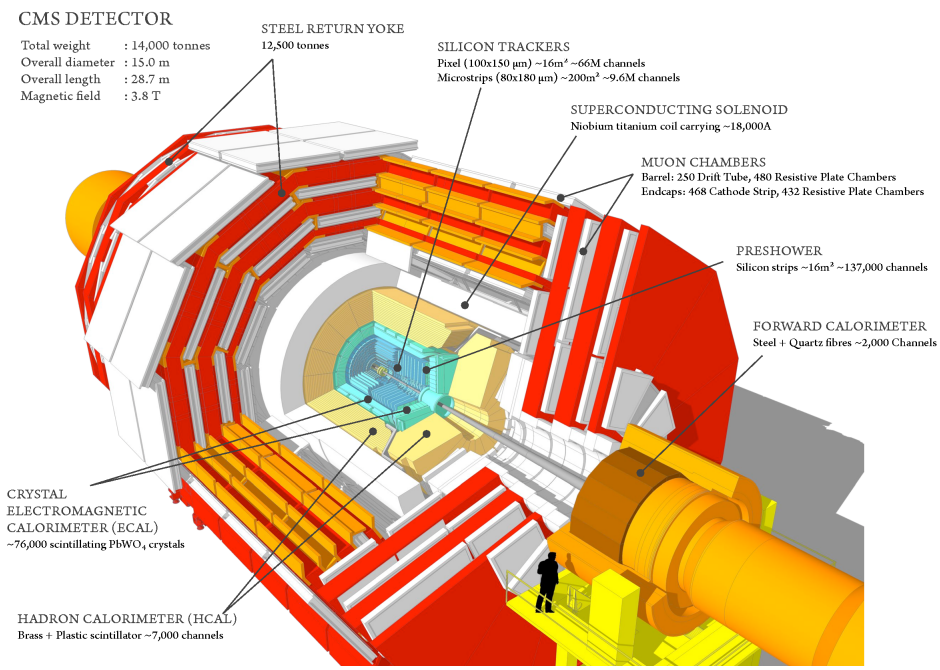

Figure 2.2.1: A cutaway view of the CMS detector [5]

particle's and the direction of the z-axis. Fig. 2.2.2 shows the relation of θ and η with respect to the CMS detector.

2.2.1 Silicon Trackers

The silicon tracker system consists of multiple layers of silicon sensors that are reversely biased p-n junctions. When applied with high enough voltage, the bulk of the silicon is depleted of free charge carriers. As charged particles pass through the depletion region, electron-hole pairs are produced and are accelerated due to the strong electric field. Thus, a current is triggered and the position of the particle can be identified.

With 5.8 m in length and 2.5 m in diameter, a schematic drawing of the CMS tracker is shown in Fig 2.2.3. Geometrically, two types of silicon sensors are deployed in CMS: pixel sensors that can measure 2-dimensional information in each layer but are very expensive to make; and strip sensors that measure only 1-dimensional information of the hit but are much cheaper.

The pixel detector consists of three cylindrical layers and four disks of pixel modules covering an area of about 1 m^2 with 66 million pixels. Each pixel is of size $100 \times 150 \mu\text{m}^2$. The pixel detector covers a pseudorapidity range of $-2.5 < \eta < 2.5$ and plays a crucial role in vertex reconstruction.

The silicon strip detector surrounds the pixel detector in the radial region of 20 cm to 116 cm.

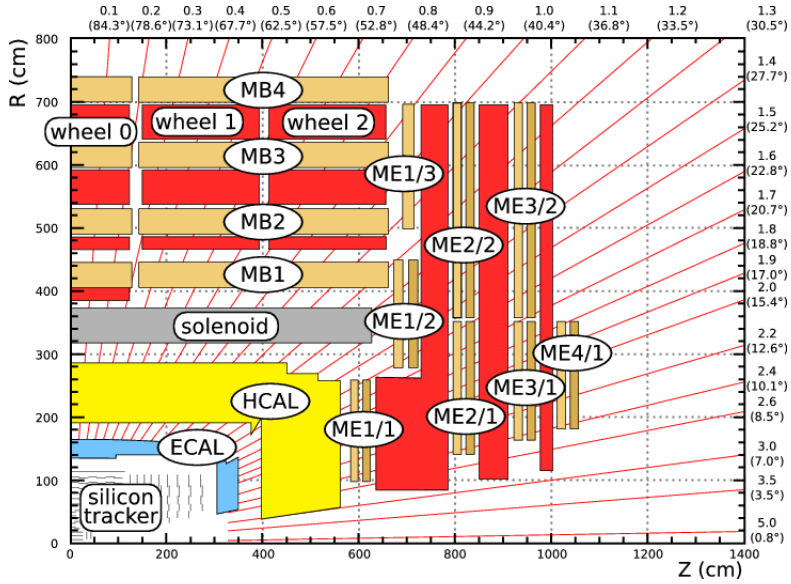

Figure 2.2.2: Schematic view of a CMS quadrant [6]

It is composed of four subsystems: the Tracker Inner Barrel (TIB); the Tracker Inner Disk (TID); the Tracker Outer Barrel (TOB); and the Tracker EndCaps (TEC). Starting closest to the collision point, the TIB consists of four layers of silicon strip sensors with strip pitch ranging from 80 μm to 120 μm. The TID consists of three layers of disks with the mean pitch varying between 100 μm to 141 μm. Surrounding the TIB/TID is the TOB with 6 layers of silicon strips pitched from 122 μm to 183 μm. Finally, two 9-disk TECs encapsulate them all along the z direction with radial strips of 97 μm to 184 μm average pitch. With about 200 m^2 of active silicon area, the CMS tracker is the largest silicon tracker ever built [7].

2.2.2 Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECAL) is a hermetic, homogeneous, finely granular and compact calorimeter designed to measure the energy of electrons and photons. As shown in Fig 2.2.4, the ECAL consists of 61,200 lead tungstate ($PbWO_4$) crystals in the central barrel and 7,324 crystals in each of the two endcaps. Two additional preshower detectors are placed before the endcap crystals to help identify neutral pion decays and to improve position determination of electrons and photons in the endcap region. $PbWO_4$ crystals are chosen for their high density (8.28 g/cm^3), short radiation length (0.89 cm) and small Molière radius (2.2 cm). As electrons and photons travel into these crystals, they loose energy through bremsstrahlung and photon conversion. These energies are then absorbed by atoms in the lattice and emitted as blue-green scintillation light of 420-430 nm

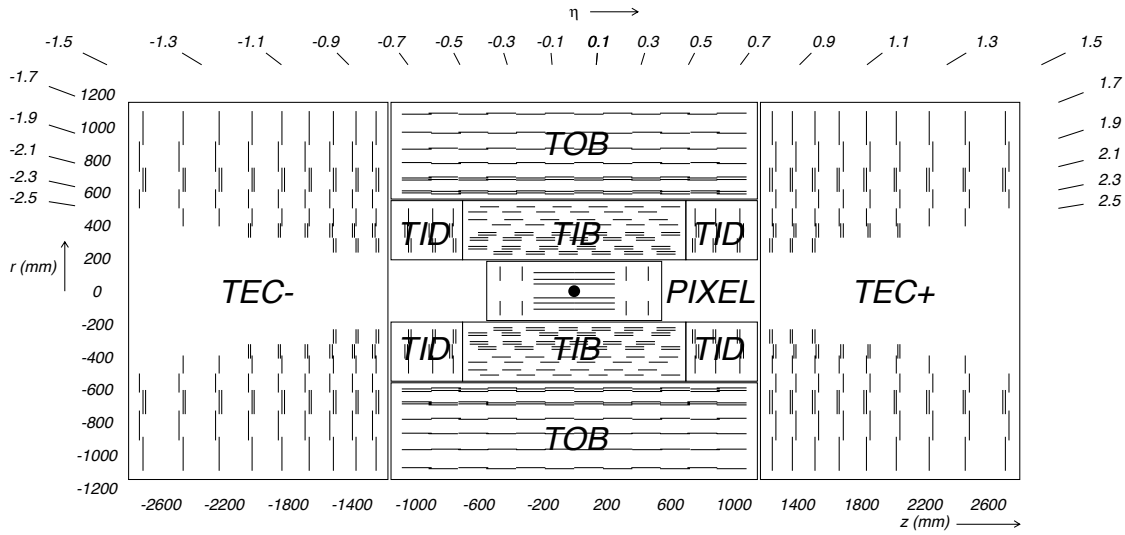

Figure 2.2.3: Schematic cross section through the CMS tracker. Each line represents a detector module. Double lines indicate back-to-back modules that deliver stereo hits. [7]

wavelength. The scintillation light is collected by photodetectors and converted into electric pulses.

The 61,200 crystals in the barrel part of the ECAL (EB) cover the pseudorapidity range $|\eta| < 1.479$. To avoid crystal boundaries aligned with particle trajectories, the tapered shaped crystals are mounted with a 3° angle with respect to the vector from the nominal collision point in both the ϕ and η projections. Each crystal is 230 mm long with a cross-section of 22×22 mm^2 at the front face and 26×26 mm^2 at the rear face. Silicon avalanche photodiodes (APDs) are used in the barrel region of the ECAL operated at a gain of 50×.

The endcaps of the ECAL (EE) cover the pseudorapidity range $1.479 < |\eta| < 3.0$. Similar to the arrangement of crystals in EB, crystals in EE are aligned to point at a focus 1,300 mm beyond the interaction point, giving an off-pointing angle between 2° to 8°. Crystals in EB are 220 mm long with a cross-section of 28.62×28.62 mm^2 at the front face and 30×30 mm^2 at the rear face. Vacuum phototriodes (VPTs) are used as photodetectors in EE.

The 20 cm preshower detector sits before EE, covering the pseudorapidity range $1.653 < |\eta| < 2.6$. It consists of two layers: lead radiators that initiate electromagnetic showers from incoming electrons and photons, and silicon strip sensors behind each radiator to measure the deposited energy and the transverse shower profiles.

A typical energy resolution measured in EB was found to be [7]:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{12\%}{E}\right)^2 + (0.3\%)^2 \quad , \tag{2.2.2}$$

where E is in GeV. The first term is the stochastic term which has three main sources: event-to-event fluctuations in the lateral shower containment; a photostatistics contribution of 2.1%; and fluctuations in the energy deposited in the preshower absorber (where present) with respect to what is measured in the preshower silicon detector. The second term is the noise term caused by electronics, digitization and pileup noise. The last constant term arise from non-uniformity of the longitudinal light collection, intercalibration errors, and leakage of energy from the back of the crystal [7].

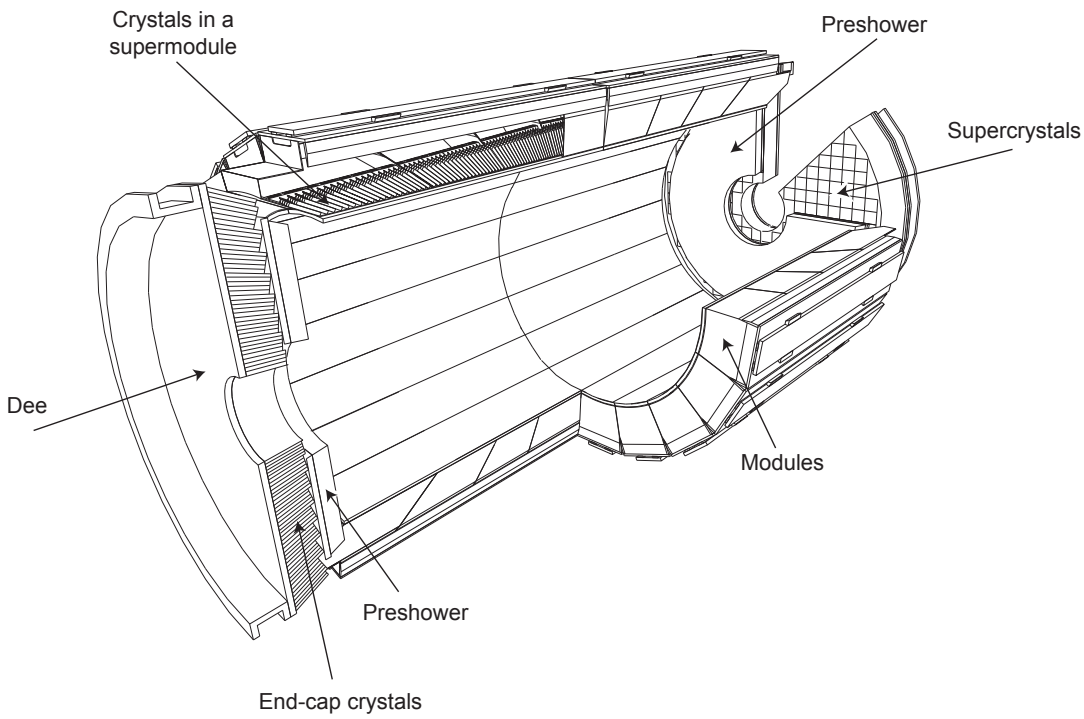

Figure 2.2.4: Layout of the CMS electromagnetic calorimeter showing the arrangement of crystal modules, supermodules and endcaps, with the preshower in front. [7]

2.2.3 Hadronic Calorimeter

The hadronic calorimeter (HCAL) mostly sits in between the ECAL and the magnet coil covering a wide pseudorapidity range $|\eta| < 5.2$. Designed to measure the energy and direction of hadronic particles, the HCAL is a sampling calorimeter made of alternating layers of absorbers and active layers.

As shown in Fig. 2.2.5, HCAL consists of four sub-detectors: the barrel and endcap of HCAL (HB, HE), which cover the pseudorapidity range $|\eta| < 3$, use brass and steel plates as absorbers and plastic scintillators as the active material; the forward calorimeter (HF), covering the pseudorapidity

range $3 < |\eta| < 5$, uses steel absorbers and Cerenkov radiating quartz fibers as active material to withstand the high radiation in the large $|\eta|$ regions; and the outer calorimeter (HO), is sited outside the central magnet uses the solenoid coil as an additional absorber, adding additional containment for hadron showers in the pseudorapidity range $|\eta| < 1.3$.

The HB consists of 36 identical azimuthal absorber wedges with plastic scintillator divided into 16 η sectors, resulting a granularity of $\Delta\phi \times \Delta\eta = 0.087 \times 0.087$. Following a similar design, the HE has a granularity of $\Delta\phi \times \Delta\eta = 0.087 \times 0.087$ for $|\eta| < 1.6$ and $\Delta\phi \times \Delta\eta \approx 0.17 \times 0.17$ for $|\eta| \geq 1.6$. HO roughly maps the layers of HB to make granularity 0.087×0.087 in η and ϕ.

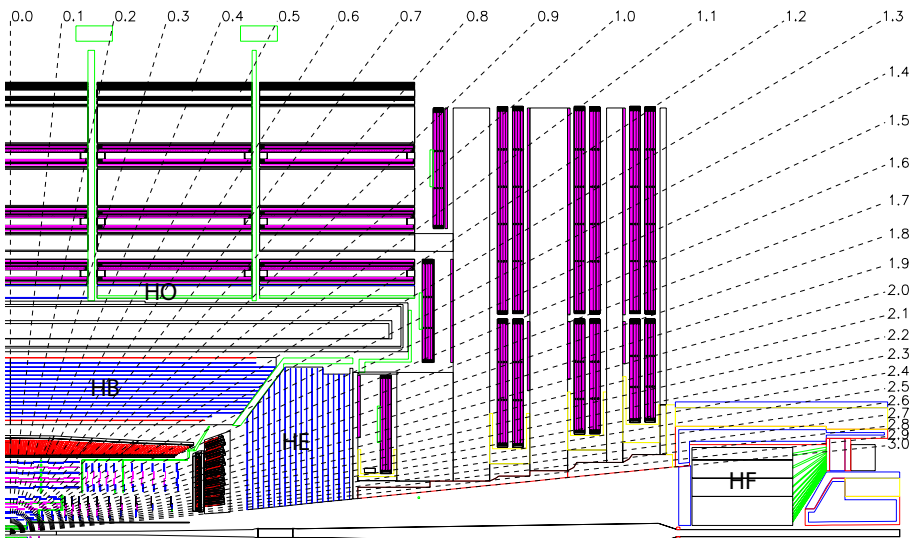

Figure 2.2.5: Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters. [7]

2.2.4 Muon System

Muon detection is a powerful tool in separating signatures of interesting processes from the very high background rate expected at the LHC. Therefore, as the experiment's middle name suggests, the detection of muons is of central importance to CMS. The basic detector utilized in the CMS muon systems are gas ionization chambers. As muons travel through these chambers, they knock electrons off the atoms of the gas. These electrons will then follow the external electric field and register as electric pulses in the readout system.

Due to the geometry of the magnet and performance of the detector, as shown in Fig. 2.2.6, three types of gaseous particle detectors are used in the CMS muon system: the drift tube chambers

(DT), which cover the barrel region in pseudorapidity range $|\eta| < 1.2$; the cathode strip chambers (CSC), which cover the endcap region in pseudorapidity range $0.9 < |\eta| < 2.4$; and the resistive plate chambers (RPC), which are interspersed in both the barrel and endcap regions.

The DT system has four stations, namely MB1 to MB4, where each station consists of 8 layers of tubes measuring the position in the transverse plane and 4 layers in the longitudinal plane (except MB4). The basic element of the DT system is the drift cell of transverse size 42×13 mm^2 operating at a voltage of 3.6 kV. The gas mixture (85%/15% of Ar/CO_2) provides good quenching properties and a maximum drift time of almost 400 ns.

The CSC system also has four stations, ME1 to ME4, where each consists of 6 layers, each of which measures the muon's position in 2 coordinates. All chambers use a gas mixture of 50% CO_2, 40% Ar, and 10% tetrafluoride (CF_4) and operate between a voltage of 2.9 kV to 3.6 kV.

RPCs are fast gaseous parallel-plate detectors that combine adequate spatial resolution with a time resolution of just 1 ns. Integrated alongside with the DTs and CSCs, the RPC provides a muon trigger system parallel with those of the DTs and CSCs. The RPC has four stations in the barrel region, RB1 to RB4, and three stations in the endcaps, RE1 to RE3. RPCs operate at a voltage of about 9.6 kV and have a gas mixture that consists of 95.2% Freon ($C_2H_2F_4$), 4.5% isobutane (i-C_4H_{10}), and 0.3% sulphur hexafluoride (SF_6).

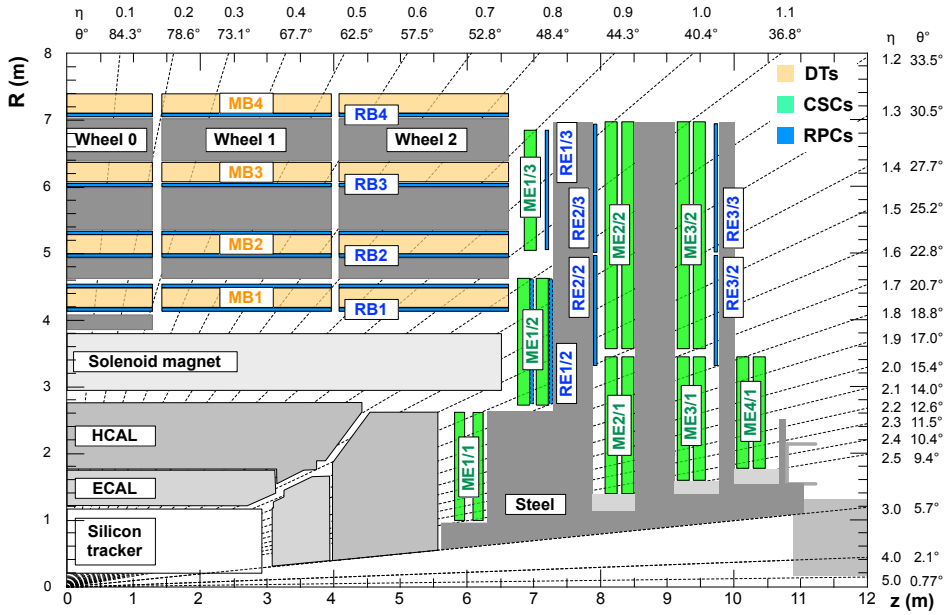

Figure 2.2.6: An cross section of a quadrant of the CMS detector. The 4 drift tube (DT, in light orange) stations are labeled MB ("muon barrel") and the cathode strip chambers (CSC, in green) are labeled ME ("muon endcap"). Resistive plate chambers (RPC, in blue) are in both the barrel and the endcaps of CMS, where they are labeled RB and RE, respectively. [8]

Chapter 3

Object Reconstruction at CMS

As described in Section 2.2, the CMS detector is a complex apparatus containing multiple sub-detectors each designed to measure different properties of particles. Typically, in CMS, particles are reconstructed by the particle-flow (PF) algorithm [45] that combines the raw information from all sub-detectors. By utilizing the information from all sub-detectors, the particle-flow algorithm aims to identify particles, including electrons, muons, tau leptons, photons, charged and neutral hadrons, in the event with an optimal determination of their direction, energy and type. In this chapter, we will go over the reconstruction of the objects that are most important to my analysis.

3.1 Primary Vertex

The precise reconstruction of the p-p collision vertex is the foundation of all particle reconstruction. It is especially important for the identification of particles that have a spacial displacement between their decay vertices and the primary vertex (PV).

As mentioned in Section 2.1, collisions take place between two bunches of protons where each bunch contains over one hundred billion protons. Thus, during each such collision, there will likely be multiple p-p collisions taking place at the same time. Out of the subset of inelastic interactions recorded by CMS, we consider the most active (largest $\sum p_{\mathrm{T}}^2$) PV as the *signal* PV and the other PVs as *pile-up*.

The reconstruction of PV consists of three stages: selection of the tracks, where tracks are required to be produced promptly in the primary interaction region; clustering of the tracks that appear to originate from the same interaction vertex, where the deterministic annealing (DA) algorithm is used to determine the most probable set of PVs; and fitting for the position of each vertex,

using its associated tracks [46]. A good indicator of the success of the fit is the number of degrees of freedom for the vertex. Each track in the vertex in is assigned a weight w_i between 0 and 1, which reflects the likelihood that it belongs to the vertex. The number of degrees of freedom in the fit is defined as:

$$n_{\mathrm{dof}} = -3 + 2 \sum_{i=1}^{\#\,\mathrm{tracks}} w_i \quad . \tag{3.1.1}$$

Finally, these reconstructed PVs are required to have $n_{\mathrm{dof}} > 4$, and to be within 24 cm and 2 cm of the beam spot along the z axis and in the transverse plane, respectivley.

3.2 Electron Reconstruction and Identification

Electrons are reconstructed by associating a track reconstructed in the silicon detector with a cluster of energy in the ECAL. One of the main challenges for electron reconstruction in the CMS detector is from bremsstrahlung radiation in the silicon tracker. Additional to the energy deposit of the electrons in the ECAL, photons radiated off electrons are spread over several crystals of the ECAL detector along the electron trajectory, mostly in the ϕ direction.

To measure the initial energy of the electron accurately, two algorithms based on energy clustering, "Hybrid" for the barrel and "multi-5×5" for the endcaps, are used to measure the energy of electrons and photons. Electron tracks are reconstructed by matching hits in the silicon strip tracker to seed hits in the pixel detector, based on the combinatorial Kalman filter method. A pixel seed is composed of two or three pixel hits compatible with the beam spot. Once the hits are collected, a Gaussian Sum Filter (GSF) fit is performed to estimate the track parameters. For each GSF track, several PF clusters, corresponding to the electron at the ECAL surface and the bremsstrahlung photons emitted along its trajectory, are grouped together. Most of the bremsstrahlung photons are recovered in this way. In order to minimize the many possible trajectories due to different combinations of hits, the track that best matches to the energy deposit in the ECAL is chosen to be the reconstructed track [47].

Electron selections have two main components, electron identification and electron isolation. In this analysis we use the multivariate (MVA) electron identification provided by the Electron Gamma Physics Object Group (POG) [48]. Two working points, loose and tight, are chosen which corresponds to an efficiency of 90% and 80% respectively.

The electron relative combined isolation is defined as below:

$$\text{relative combined isolation} = \frac{\Sigma p_T^{\text{charged}} + \max[0, \Sigma p_T^{\text{neutral had}} + \Sigma p_T^{\gamma} - 0.5 \times \Sigma p_T^{\text{charged hadrons from PU}}]}{p_T^{\text{electron}}} , \tag{3.2.1}$$

where the sum runs over charged PF candidates, neutral hadrons and photons, within a $\Delta R < 0.3$ (defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$) around the electron direction. Charged candidates are required to originate from the signal PV, and $p_T^{\text{charged hadrons from PU}}$ is a correction related to event pileup and is defined as: $p_T^{\text{charged hadrons from PU}} = 0.5 \times \sum_{\text{not coming from PV}, \Delta R < 0.3} p_{\text{T}}^{\text{charged hadrons}}$.

3.3 Muon Reconstruction and Identification

Muon reconstruction is a multistep process that begins with track reconstruction independently in the tracker and muon system. In the muon system, standalone muon tracks are reconstructed by building straight-line track segments from the hits in the DT and CSC systems. Based on these tracks, two reconstruction algorithms are used: the Global Muon reconstruction and the Tracker Muon reconstruction. The Global Muon reconstruction starts by matching the standalone muon tracks to the tracker tracks. Once a match is found, hits from the tracker track and the standalone muon tracks are combined and fitted as the Global Muon track using the Kalman-filter technique. On the contrary, the Tracker Muon reconstruction method starts from the tracker tracks and extrapolates them to the muon system taking into consideration the magnetic field, the average expected energy loss, and multiple Coulomb scattering in the detector material. The tracker track is considered as a Tracker Muon if there is at least one muon segment in the muon system that matches to the extrapolated track [49]. With the requirements in both the tracker and muon system, the remaining major source of objects faking prompt muons are punch-throughs and non-prompt muons from hadron decays. Punch-throughs happen when charged hadrons penetrate through the hadronic calorimeter and leave hits in the muon system. They often occur due to pions from the fragmentation of quarks and gluons. These punch-throughs can often be discriminated by comparing the energy deposit in HCAL and the muon system. Non-prompt muons from jet decays are expected to be within the jet and can be discriminated against by imposing an isolation requirement.

Similar to the selection of electrons, muon selections have two main components, muon identification and muon isolation. Muons are selected with the "medium" μ identification (ID) criteria suggested by the muon POG [50].

Isolated muons are required to have minimal energy from PF neutral and charged hadron candi-

dates in a cone of $\Delta R = 0.3$ around the muon trajectory. PF charged candidates considered in the calculation of isolation are required to be near the signal PV. Similar to electrons, relative isolation for muons is defined as:

$$\text{relative combined isolation} = \frac{\Sigma p_T^{\text{charged}} + \max[0, \Sigma p_T^{\text{neutral had}} + \Sigma p_T^{\gamma} - 0.5 \times \Sigma p_T^{\text{charged hadrons from PU}}]}{p_T^{\text{muon}}} \quad , \tag{3.3.1}$$

where the sum runs over charged PF candidates, neutral hadrons and photons, within $\Delta R < 0.3$ around the muon direction. Charged candidates are required to originate from the signal PV, and $p_T^{\text{charged hadrons from PU}}$ is a correction related to event pileup.

3.4 Jet Reconstruction and Identification

Quarks and gluons produced in high-energy processes manifest themselves as jets with typically three components: photons, charged hadrons and neutral hadrons. The energy of photons can be measured directly from ECAL, while the energy of charged hadrons is derived by combining the information from the tracker, ECAL and HCAL. Neutral hadrons leave information only in HCAL but they normally only account for 10% of the total jet's energy [51].

Jets are then clustered from the four-momenta of these reconstructed particles using the anti-k_{T} jet clustering-algorithm [52] with a distance parameter of $\Delta R = 0.4$ in the η-ϕ plane. By introducing the distance d_{ij} defined between particles i and j, the anti-k_{T} algorithm clusters particles with the smallest distance d together. Then, distances are recalculated and the procedure repeated until no particles are left. The distance d_{ij} is defined as:

$$d_{ij} = \min(k_{ti}^{-2}, k_{tj}^{-2}) \frac{\Delta_{ij}^2}{R^2} \quad , \tag{3.4.1}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and k_{ti}, y_i and ϕ_i are respectively the transverse momentum, rapidity and azimuth angle of particle i.

Thus, the key feature of the anti-k_{T} algorithm is that less energetic particles tend to cluster with energetic neighbors before clustering with other less energetic particles. If an energetic particle has no energetic neighbors within a distance 2R, it will simply cluster all the less energetic particles within a circle of radius R, resulting in a perfectly conical jet [52].

As suggested and provided by the Jet/MET Physics Object Group [53], jets with $|\eta| < 2.4$ are required to pass the "loose" ID requirements summarized in Table 3.1 with an overall efficiency of

above 99%.

Table 3.1: Loose Jet-ID Selections.

Selection	Cut
Neutral Hadron Fraction	< 0.99
Neutral EM Fraction	< 0.99
Number of Constituents	> 1
And for $\|\eta\| < 2.4$ in addition apply	
Charged Hadron Fraction	> 0
Charged Multiplicity	> 0
Charged EM Fraction	< 0.99

To correct the energy of the reconstructed jets and their energy at particle level, which is independent of the detector response, a jet energy correction (JEC) is applied for both data and simulated Monte Carlo (MC) events. The JECs are applied at the following levels: L1 corrections; L2L3 MC-truth corrections and L2L3 residuals corrections (applied to data only).

L1 corrections are designed to remove the energy coming from pile-up events. L2L3 MC-truth corrections improve the energy response as a function of jet p_T and η. L2L3 residuals corrections, applied only to data, corrects the remaining difference within the jet response between data and MC.

3.4.1 b-Jet Tagging

The relatively long lifetime of bottom quarks ($\sim$ 1.5 ps), corresponding to a flight distance of a few centimeters that is observable with high resolution tracking detectors, can lead to a secondary vertex displaced from the primary event vertex. With this special property, several algorithms has been developed to identify b quark decays. In the analysis of this thesis, we use the Combined Secondary Vertex (CSV) algorithm [54], which combines the information of the secondary vertices with track-based lifetime information. Two likelihood ratios are built from the input variables used to discriminate between b and c jets and between b and light-parton jets. The final discriminator is a combination of the two with prior weights of 0.25 and 0.75, respectively.

Since data reconstruction in 2015, the CSV algorithm was further optimized and the new version is referred to as CSVv2 [9]. The main differences with the original CSV algorithm are the different vertex reconstruction algorithm used, the number of input variables, and the way they are combined. The new vertex reconstruction algorithm, the Inclusive Vertex Finder (IVF), is seeded with the collection of reconstructed tracks in the event. Compared with the original vertex reconstruction algorithm used in CSV, the efficiency to reconstruct a secondary vertex for b (c) jets using the

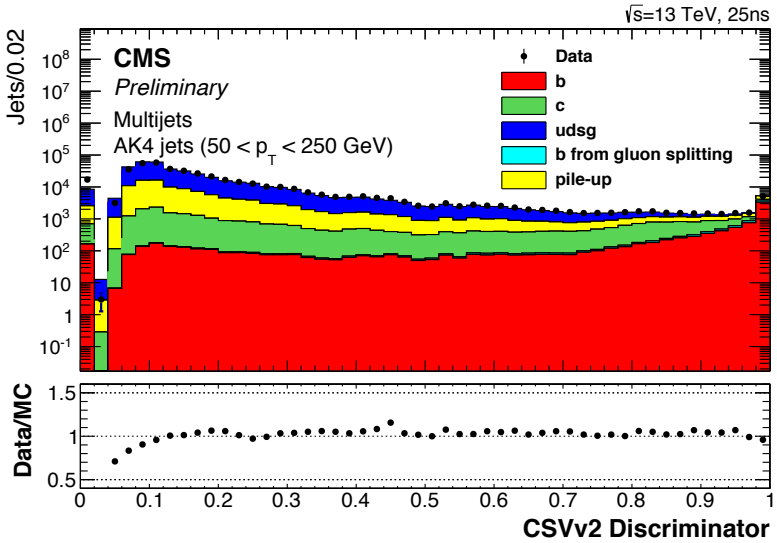

Figure 3.4.1: Discriminator values for the CSVv2 algorithm. Underflow and overflow are added to the first and last bins, respectively. [9]

IVF algorithm is about 10% (15%) higher. In the CSV algorithm, input variables are combined with a likelihood ratio. In the CSVv2 algorithm, they are combined using a multilayer perceptron. Examples of newly added variables include: the number of secondary vertices; the angle between the secondary vertex and the jet axis; the ratio of the transverse momentum of the summed track four-momenta and the jet; the track decay length; and the angle between the track and the jet [9]. Figure 3.4.1 shows the distributions of the discriminator values for the CSVv2 algorithm.

The performance of the various b-tagging algorithms, including CSV and CSVv2, is presented in Figure 3.4.2 as the b jet identification efficiency versus the misidentification probability for jets in simulated $t\bar{t}$ events requiring a jet with $p_T > 30$ GeV. The improvement of the CSVv2 algorithm with respect to the CSV algorithm is shown.

The minimum thresholds on the CSV discriminator define loose, medium, and tight working points with a misidentification probability (efficiency) of about 10% (83%), 1% (69%), and 0.1% (49%), respectively, for light-flavour jets with a p_T above 30 GeV.

3.5 Tau Reconstruction and Identification

As discussed in Section 1.1.1, with a invariant mass of $m_\tau = 1.777$ GeV, the τ lepton is the only lepton heavy enough to decay into hadrons. As shown in Table 3.2, in one third of the cases, τ's decay into an electron or muon with two neutrinos and are reconstructed with the electron and muon

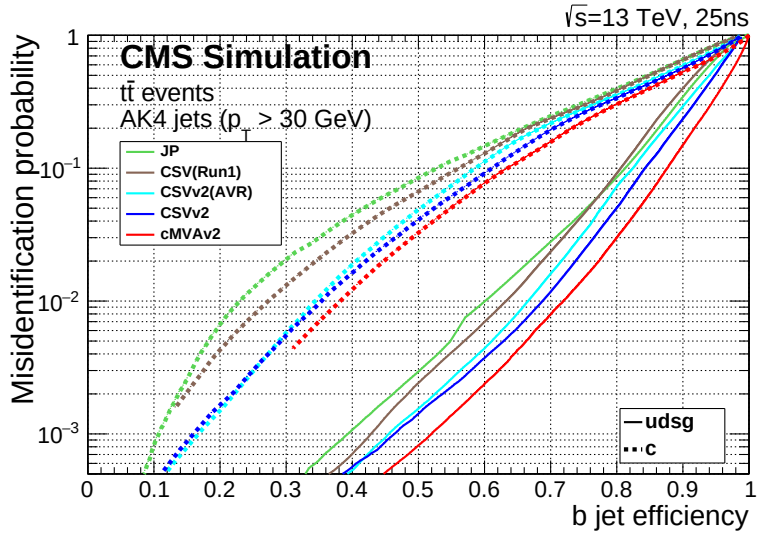

Figure 3.4.2: Performance of the b jet identification efficiency algorithms demonstrating the probability for non-b jets to be misidentified as b jet as a function of the efficiency to correctly identify b jets. The curves are obtained on simulated $t\bar{t}$ events using jets with $p_T > 30$ GeV, b jets from gluon splitting to a pair of b quarks are considered as b jets. [9]

techniques mentioned above. The remaining τ leptons decay hadronically (τ_h) into a combination of charged and neutral mesons with a τ neutrino. In CMS, hadronically decaying τ leptons are reconstructed with the hadrons-plus-strips (HPS) algorithm [55] which pays special attention to photon conversions in the CMS tracker material.

Table 3.2: Approximate branching fractions of different τ decay modes. Here, h^- represents a charged hadron (either a pion or a kaon) [16].

Decay mode	Meson resonance	Branching ratio
$\tau^- \to e^- \bar{\nu}_e \nu_\tau$		17.8
$\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$		17.4
$\tau^- \to h^- \nu_\tau$		11.5
$\tau^- \to h^- \pi^0 \nu_\tau$	$\rho(770)$	26.0
$\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$	9.5
$\tau^- \to h^- h^+ h^- \nu_\tau$	$a_1(1260)$	9.8
$\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$		4.8
Other modes with hadrons		3.2
All modes containing hadrons		64.8

The HPS algorithm for τ_h reconstruction and identification is performed in two stages:

- **Reconstruction**: τ_h candidates are reconstructed by combining charged and neutral particles that are compatible with specific τ_h decays.

- **Identification**: discriminators that separate τ_h decays from jets originating from quarks and

gluons, electrons, and muons, are computed. This helps to reduce the misidentification rate of non-τ_h particles as τ_h candidates.

3.5.1 Tau Reconstruction

Many photons are produced, due to the high $\pi^0 \to \gamma\gamma$ branching ratio, when τs decay hadronically. As photons scatter off a nucleus, they can be converted into an electron-positron pair. Under the influence of the magnetic field, the calorimeter signature of the converted photon broadens in the ϕ direction. This effect is taken into account by the HPS algorithm by clustering the photon and electron constituents of the jet that seeds tau reconstruction into $\eta-\phi$ strips. The strip reconstruction starts with the most energetic electromagnetic particle within the PF jet and then searches for other electromagnetic particles within a window of $\Delta\eta \times \Delta\phi = 0.05 \times 0.20$. If other electromagnetic particles are found within that window, the most energetic one is associated with the strip and the strip four-momentum is recalculated. With no further particles that can be associated with the strip, strips with $p_T > 1$ GeV are combined with the charged hadrons in attempts to reconstruct the possible tau decay modes outlined in Table 3.3.

Table 3.3: Reconstructed Tau Decay Modes

HPS Tau Decay Modes
Single Charged Hadron + Zero Strip
Single Charged Hadron + One Strip
Single Charged Hadron + Two Strips
Two Charged Hadrons
Three Hadrons

Assuming all charged hadrons to be pions, the four-momenta of charged hadrons and strips are reconstructed according to the τ_h decay topologies. The mass of the reconstructed π^0 from the strips is required to be between 50-200 MeV. Additionally, the mass of the reconstructed tau is required to be consistent with the masses of the intermediate meson resonances listed in Table 3.5.1. If there are more than one topology satisfying the mass constraints, the hypothesis giving the highest $p_{\mathrm{T}}^{\tau_h}$ is chosen.

Decay mode	Resonance	Mass (MeV)	Allowed Mass Window
$\tau^- \to h^-\pi^0\nu_\tau$	ρ^-	770	300 - 1300
$\tau^- \to h^-\pi^0\pi^0\nu_\tau$	a_1^-	1200	800 - 1500
$\tau^- \to h^-h^+h^-\nu_\tau$	a_1^-	1200	800 - 1500

In Run-2, based on the HPS algorithm, the dynamic strip reconstruction algorithm [56] has been deployed for τ_h reconstruction. Compared with HPS, the dynamic strip algorithm changes the

previously fixed window of $\Delta\eta \times \Delta\phi = 0.05 \times 0.20$ to an energy dependent one that better matches to the different curvature of charged particle tracks at different energies. As shown in studies performed to optimize the strip size, particles with low p_{T}, due to nuclear interaction with tracker material or bremsstrahlung, can easily escape the previously defined strip window. The dynamic strip algorithm revised the definition of the strip size as $\Delta\eta = 0.2 \cdot p_{\mathrm{T}}^{-0.66}$ and $\Delta\phi = 0.35 \cdot p_{\mathrm{T}}^{-0.71}$ with the upper limit set to 0.15 in $\Delta\eta$ and 0.3 in $\Delta\phi$.

As discussed, the key elements in the HPS algorithm are the reconstructed charged hadrons and photons. Thus, the reconstruction efficiency of the tracks and photons sets an upper limit on the tau reconstruction efficiency. On average, track-reconstruction efficiency for promptly-produced charged particles with $p_T > 0.9$ GeV is 94% for $|\eta| < 0.9$ and 85% for $0.9 < |\eta| < 2.5$ [57]. The typical photon reconstruction efficiency is around 90% [58].

3.5.2 Tau Isolation Discriminants

The most important handle to suppress the jet $\rightarrow \tau$ misidentification is through isolation requirements. Two types of τ_h isolation discriminants have been developed: one uses simple cut-based selections and the other taking a Multi-Variate Analysis (MVA) approach.

3.5.2.1 Cut-based Discriminants

The combined isolation of τ_h candidates is computed by the scalar sum of the p_{T} of charged particles and photons with $p_T > 0.5$ GeV and within an isolation cone of size $\Delta R = 0.5$. Charged hadrons and photons used to form the τ_h candidate are excluded from the summation. To reduce the effects of pileup on charged particles, tracks associated with charged particles are required to be compatible with the production vertex of the τ_h candidate within a distance of 0.2 cm in z-direction. The effect of pileup on photons is compensated on a statistical basis through the $\Delta\beta$ corrections with the combined isolation given as [16]:

$$I_\tau = \sum_{d_z < 0.2\mathrm{cm},\ \Delta R = 0.5} p_{\mathrm{T}}^{\mathrm{charged}} + \max(0, \sum_{\Delta R = 0.3} p_{\mathrm{T}}^{\gamma} - \Delta\beta) \quad , \tag{3.5.1}$$

where $\Delta\beta$ is computed by summing the p_T of charged particles that are within a cone of $\Delta R = 0.8$ around the τ_h direction, whose associated tracks are required not originating from the τ_h production vertex, $d_z > 0.2$ cm. The sum is scaled by a factor of 0.2 to make the τ_h identification efficiency insensitive to pileup [56]:

$$\Delta\beta = 0.2 \sum_{d_z > 0.2\text{cm},\ \Delta R = 0.8} p_{\mathrm{T}}^{\mathrm{charged}} \quad , \tag{3.5.2}$$

Three working points of the cut-based discriminator are defined with a cut on I_τ which is less than: 2.5 GeV (loose), 1.5 GeV (medium) and 0.8 GeV (tight). The thresholds have been chosen such that the signal efficiency will be equidistance between the three working points [56]. The identification efficiencies (mis-ID probabilities) are 62% (0.4%), 54% (0.2%), and 45% (0.1%) for loose, medium, and tight working points respectively.

3.5.2.2 MVA-based Discriminators Against Jets

The MVA tau ID combines the isolation and shape variables with variables sensitive to τ-lifetime information and was trained with a Boosted Decision Tree (BDT) to achieve a strong rejection of quark and gluon jets. Detail information of the variables used to train the BDT can be found in [56]. Among the various working points, the tight working point is chosen, which balances the real τ_h efficiency (60%) and hadron jet rejection rate (greater than 99.8% for jets with $p_T > 50$ GeV).

3.5.2.3 MVA-based Discriminators Against Electrons

Aside from jets, electrons are another source of fake τ_hs. To discriminate against electrons, τ_hs are required to pass the anti-electron MVA discriminator. The anti-electron MVA algorithm utilizes observables that quantify the distribution in energy depositions in the ECAL, in combination with observables sensitive to the amount of bremsstrahlung emitted along the leading track, and observables sensitive to the overall particle multiplicity, to distinguish electromagnetic from hadronic showers. Finally, τ_h must not reside in the ECAL cracks. Additional information of the against-electron MVA discriminator can be found in Ref. [56]. Typically, for the medium working point of this discriminator, the τ_h efficiency is around 80% with a electron misidentification rate of around 0.1%.

3.5.2.4 Cut-based Discriminators Against Muons

In order to reject rare cases where a muon fakes a τ_h, τ_hs are required to pass the muon rejection discriminator. Two working points are provided as the following [16]:

- **Loose**: τ_h candidates are rejected when their track segments are found in at least two muon stations within a cone size of $\Delta R = 0.3$ around the τ_h direction, or when the energy deposits

in the ECAL and HCAL correspond to less than 20% of the momentum of the leading track of the τ_h candidate.

- **Tight**: additional to the requirements in the loose working point, τ_h candidates are rejected if there are hits, within a cone of $\Delta R = 0.3$ around the τ_h direction, in the CSC, DT or RPC detectors located in the two outermost muon stations.

Two working points, loose and tight, are defined for the anti-muon discriminator with identification efficiencies (mis-ID probabilities) as 96.4% (0.18%) and 95% (0.08%).

3.6 Missing Transverse Momentum

The presence of neutrinos and hypothetical neutral weakly interacting particles causes imbalance of total momentum as they escape from the detector without producing any direct response. The missing transverse momentum vector $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$, is defined as the negative vector sum of the visible transverse momentum [59].

However, minimum energy thresholds in the calorimeters, inefficiencies in the tracker, nonlinearity of the response of the calorimeter for hadronic particles could lead to overestimated or underestimated values of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$. This misestimation can be reduced by correcting the energy of the jets to the particle level jet energy using the JEC, as mentioned in Section 3.4, according to:

$$\vec{p}_{\mathrm{T}}^{\mathrm{miss\ corr}} = \vec{p}_{\mathrm{T}}^{\mathrm{miss}} - \sum_{\mathrm{jets}} (\vec{p}_{\mathrm{T,\ jet}}^{\mathrm{corr}} - \vec{p}_{\mathrm{T,\ jet}}), \tag{3.6.1}$$

where the superscript "corr" refers to the corrected values. This correction for $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$, also referred as the "type-I" correction, uses jet energy scale corrections for all corrected jets with $p_{\mathrm{T}} > 15\mathrm{GeV}$ that have less than 90% of their energy deposited in ECAL. Furthermore, if a muon is found in a jet, its 4-momentum is subtracted from the 4-momentum of the jet when performing the correction and is then added back to the corrected object [60].

Various phenomena can lead to anomalous high-$\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ events. For example, in the ECAL, spurious deposits may appear due to particles striking sensors in the ECAL photodetectors, and in the HCAL spurious energy can arise due to noise in the hybrid photodiode (HPD) and readout box (RBX) electronics [60]. In order to identify and suppress such high-$\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ events, the following MET POG [61] recommended filters are applied to the analysis:

Table 3.4: MET Filters.

primary vertex filter
beam halo filter
HBHE noise filter
HBHEiso noise filter
ECAL TP filter
ee badSC noise filter
badMuon filter
badCharged hadron filter

Chapter 4

Radiation Damage Studies for the CMS Strip Tracker

In this chapter, a silicon radiation damage study performed at the Brown Silicon Lab for the Phase-II upgrade of the CMS strip tracker is presented. We will first take a brief look at the road map of the LHC and the harsh conditions silicon trackers will be facing after the Phase-II upgrade. Then we will study the effects of radiation on the electrical characteristics of silicon sensors and how the annealing procedure further changes these characteristics.

4.1 LHC Upgrades

Between the years of 2024 to 2026, the LHC experiments will install many upgrades during the Long Shut-down 3 (LS3). It is planned to increase the instantaneous luminosity to 5 to 7 times the nominal value. Hence, enters the era of the High Luminosity LHC (HL-LHC) that is planned to operate for 10 more years delivering an integrated luminosity of 3000 fb^{-1}.

Table 4.1 shows a brief comparison between the design specifications of the LHC and HL-LHC [62, 63]. The increase in luminosity and pile-up demands the CMS silicon trackers to meet new requirements:

- Improved radiation hardness to withstand the much higher instantaneous luminosity.
- Higher granularity to separate the collision of interest from a higher pile-up.

The study presented below tackles the first problem by measuring the effects of radiation on several types of silicon diode sensors.

Parameters	LHC	HL-LHC
Peak luminosity [$cm^{-2}s^{-1}$]	10^{34}	$5-7\times10^{34}$
Maximum integrated yearly luminosity [fb^{-1}]	40	>250
Protons per bunch	1.15×10^{11}	2×10^{11}
Average pile-up	27	140

Table 4.1: Design specifications of the LHC and the HL-LHC.

4.2 Basic Features of Silicon Detectors

4.2.1 P-N Junctions

The basic element of silicon detectors at CMS are p-n junctions. By doping silicon with different substances, p-type and n-type silicons are created respectively with an excess of holes or an excess of electrons. P-n junctions are formed by connecting a p-type and n-type silicon as shown in Fig 4.2.1. With no external voltage, electrons from the n-type region will diffuse into the p-type region and recombine with holes. Thus part of the p-type region will have an excess of negative electric charge while part of the n-type region will have an excess of positive electric charge. This non-neutral region is known as the depletion region while the voltage build by the charges is known as the built it voltage V_{bi}.

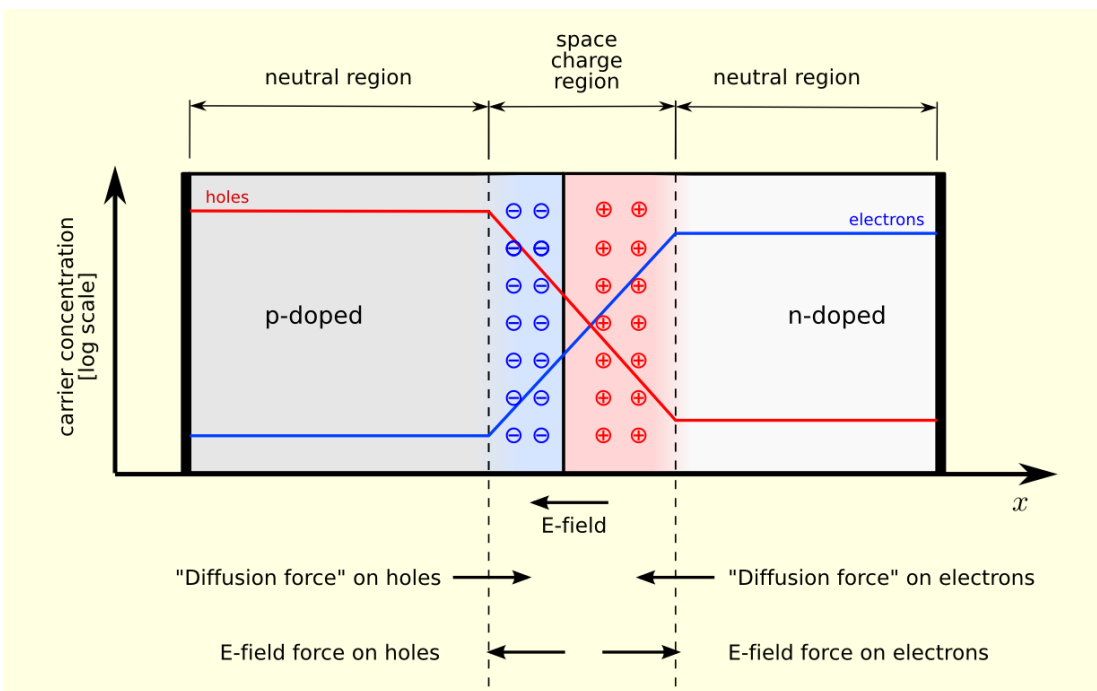

Figure 4.2.1: A p-n junction in thermal equilibrium with zero-bias voltage applied [10].

4.2.2 Characteristics of a Reverse Biased P-N Junction

Reverse bias corresponds to when an external voltage is applied where the positive terminal is connected to the n-type silicon while the negative terminal is connected to the p-type silicon. Thus, additional electrons and holes would be moving away from the p-n junction surface creating a larger depletion region. By Poisson's equation, the electric field potential Φ in the depletion region is given as [11]:

$$-\frac{d^2\Phi(x)}{dx^2} = \frac{\rho_e}{\epsilon\epsilon_0} = \frac{q_0 N_{eff}}{\epsilon\epsilon_0} \quad , \tag{4.2.1}$$

where N_{eff} is the effective doping concentration given by the difference between the concentration of ionized donors and acceptors in the depletion region and $\epsilon\epsilon_0$ stands for the permittivity of silicon.

Solving Eq 4.2.1 with boundary conditions $\Phi(x = W) = 0$ and $\Phi(x = 0) = -V_{bi} - V$, the depletion depth W is:

$$W(V) = \sqrt{\frac{2\epsilon\epsilon_0}{q_0|N_{eff}|}(V + V_{bi})} \quad , \tag{4.2.2}$$

where V is the reverse bias voltage. As the depletion voltage increases the depletion region expands until no neutral region is left ($W = d$). Thus, the silicon is referred as fully depleted with the bias voltage referred as the depletion voltage V_{dep}. As V_{dep} is normally orders of magnitudes larger than V_{bi}, V_{bi} is often ignored and:

$$V_{dep} = \frac{q_0}{2\epsilon\epsilon_0}|N_{eff}|d^2 \quad . \tag{4.2.3}$$

The junction capacitance is defined as $C = dQ/dV$, where $Q = q_0 N_{eff} A W$ with A being the area of the junction. Hence, the junction capacitance can be expressed as:

$$\begin{aligned} C &= \frac{dQ}{dV} = \frac{dQ}{dW}\frac{dW}{dV} = q_0 N_{eff} A \cdot \frac{1}{2}\sqrt{\frac{2\epsilon\epsilon_0}{q_0|N_{eff}|(V + V_{bi})}} \\ &= A\sqrt{\frac{\epsilon\epsilon_0|q_0 N_{eff}|}{2(V + V_{bi})}} \\ &\propto \frac{1}{\sqrt{V}} \quad \text{for } V \gg V_{bi} \quad . \end{aligned} \tag{4.2.4}$$

At full depletion, $C(V_{dep}) = \epsilon\epsilon_0 A/d = C_{geo}$ is called the geometrical capacitance, as it only depends on the geometrical size of the junction.

The current of a reverse biased p-n junction is known as the leakage current. For un-irradiated sensors, the leakage current consists of diffusion currents and currents from defects (from the manufacture process) in the silicon. After irradiation, additional leakage current ΔI is introduced by the

defects caused by radiation in the form of [11]:

$$\Delta I = \alpha \Phi_{eq} A W \quad , \tag{4.2.5}$$

where α is the current related damage rate and Φ_{eq} is the 1 MeV neutron equivalent fluence. Φ_{eq} is related to the 800 MeV proton fluence Φ_{proton}, the radiation used in this study, by $\Phi_{eq} = \kappa \Phi_{proton}$ with $\kappa = 0.72$ [64] as the hardness factor.

The leakage current has a strong temperature dependence and is normally scaled to a reference temperature of $T_{\mathrm{R}} = 20°\mathrm{C}$ by [11]:

$$I(T_{\mathrm{R}}) = I(T) \cdot \left(\frac{T_{\mathrm{R}}}{T}\right)^2 \exp\left[-\frac{E_g}{2k_B}\left(\frac{1}{T_{\mathrm{R}}} - \frac{1}{T}\right)\right] \quad , \tag{4.2.6}$$

where the band gap $E_g = 1.12$ eV is used.

4.3 Test Structures

The studied test structures are silicon diode sensors provided by the RD50 collaboration. Fig 4.3.1 shows a cross section view and top view of the test structure used. Depending on their manufacturing technique, four types of silicon are studied including: Float Zone (FZ), Oxygen Enriched FZ (DOFZ), Magnetic Czochalski (MCz) and Epitaxial (Epi). As shown on the left panel of Fig 4.3.1, depending on the doping material in the bulk region (ν), two types of silicon sensors can be constructed including: n-type bulk silicon when the ν region is doped to be n-type, and p-type bulk silicon when the ν region is doped to be p-type. All of the silicon sensors provided by the RD50 collaboration and used in this study are n-type sensors. The test structures have a variety of geometric size with their thickness ranging from 50 μm to 300 μm and an active area of 6.25 mm^2 or 25 mm^2. Table 4.2 lists the exact details on the silicon sensors as well as their measured initial (before irradiation) depletion voltage V_{dep} and effective doping concentration N_{eff} calculated with Eq 4.2.3.

Material	FZ	DOFZ	MCz	Epi	Epi
Thickness [μm]	285	285	300	150	50
Active Area [mm^2]	25	25	6.25	6.25	25
Initial V_{dep} [V]	46	76	281	146	125
Initial N_{eff} [$\times 10^{11}$cm^{-3}]	12.4	7.5	41.3	80.5	657.3

Table 4.2: Basic properties of the test structures.

After measuring the initial electrical properties of the silicon sensors, they were irradiated with

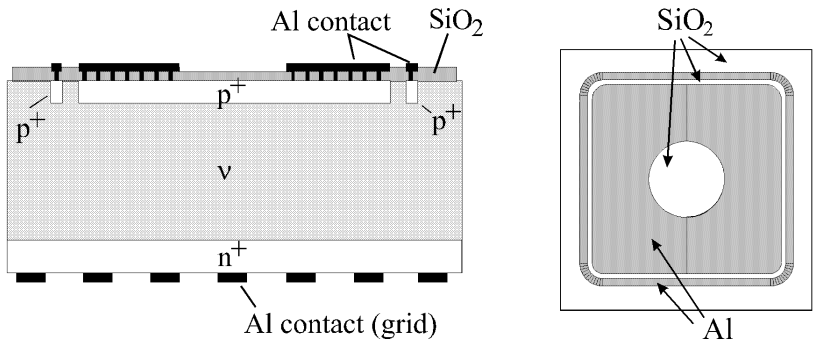

Figure 4.3.1: The schematics cross section (left) and top view (right) of the test structure used [11].

800 MeV protons at the LANSCE proton facility at Los Alamos, NM. Table 4.3 lists the amount of radiation fluence the different types of sensors were exposed to where only MCz and Epi sensors achieved a fluence of 46 $\times 10^3$ $n_{eq} \cdot \mathrm{cm}^{-3}$.

Fluence [$\times 10^3$ $n_{eq} \cdot \mathrm{cm}^{-3}$]	FZ	DOFZ	MCz	Epi	Epi
1.8	✓	✓	✓	✓	✓
4.3	✓	✓	✓	✓	✓
13	✓	✓	✓	✓	✓
46	×	×	✓	✓	✓

Table 4.3: Fluence achieved for the test sensors with 800 MeV protons.

4.4 Experimental Setup

As discussed in Section 4.2, the key characteristics of a reverse biased p-n junction are its depletion voltage V_{dep}, capacitance C and leakage current $I_{leakage}$. In order to measure their values as well as their correlations, we have built a Capacitance vs Voltage and Current vs Voltage (CV-IV) measurement apparatus in the Brown Silicon Lab.

Fig 4.4.1 shows an overview of the CV-IV measurement setup where the function of the different instruments are highlighted on the left. The tested sensor is placed on the aluminum platform (known as the chuck) in the right, as shown in the left panel of Fig 4.4.2. In order to improve the electrical contact between the chuck and the back of the sensor, the chuck is designed like an air hockey table but with a reversed air flow provided by a vacuum pump. For maximum suction, additional switches allow us to enable only the holes beneath the silicon sensor. The bias voltage is applied directly to the chuck by a Keithley 6517B HV (±1000 V) source which also measures the leakage current.

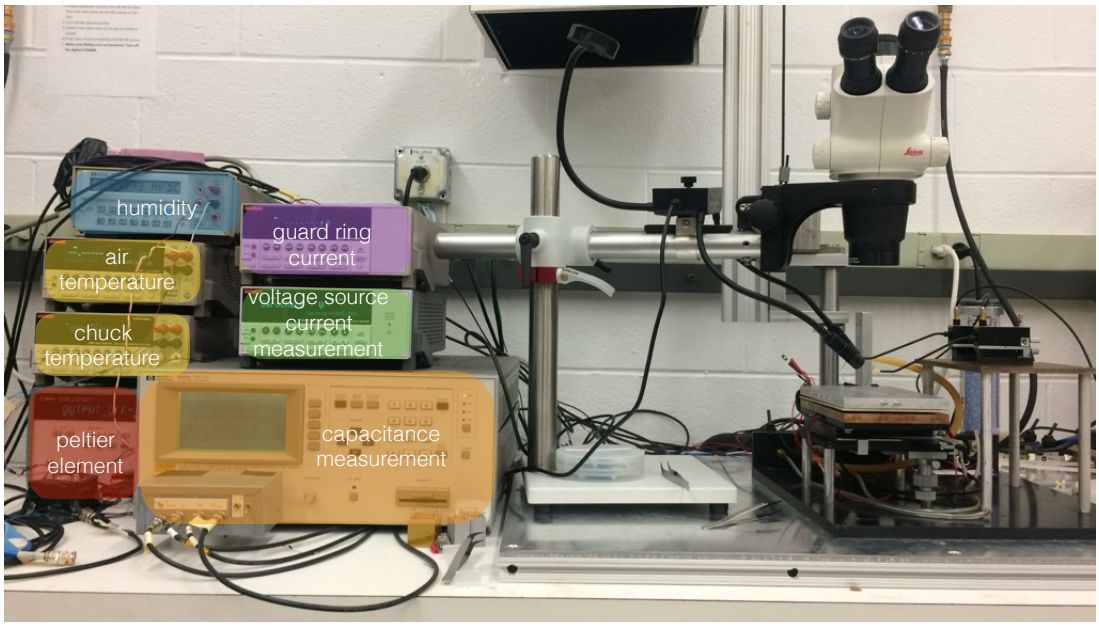

Figure 4.4.1: The overview of the CV-IV measurement apparatus in the Brown Silicon Lab.

In order to perform measurements at different temperatures, four Peltier elements are used and placed right under the chuck (not shown in picture). Due to the limited temperature difference achievable by Peltier elements, a bronze heat sink is place underneath them. When taking measurements at low temperatures, 0°C and −20°C, the bronze heat sink is first cooled to +5°C or +10°C of the desired temperature by liquid coolant, then the Peltier elements kick in to further reduce the temperature and stabilize it to be within ±0.1°C.

The connections to the front of the sensor are made with probes with tip sizes of 7 μm. These probes are controlled using micro manipulators, shown in the left panel of Fig 4.4.2, with the help of a Leica optical microscope.

The capacitance is measured with an HP 4284A LCR meter at 1 kHz and 10 kHz to check for frequency dependence. A specially designed switch box is attached to the LCR meter which enables fast switching between C-V and I-V measurements.

All the measurement instruments shown on the left side of Fig 4.4.1 are connected to the PC via a General Purpose Interface Bus (GPIB). The measurements are controlled and recorded by a dedicated LabView program.

During measurements, as shown in the right panel of Fig 4.4.2, a light-tight aluminum box encloses the measurement setup to prevent currents caused by external photons. In order to avoid condensation during low temperature measurements, dry air is flushed into the aluminum box. The temperature and humidity of the air in the box is constantly monitored to ensure that the air

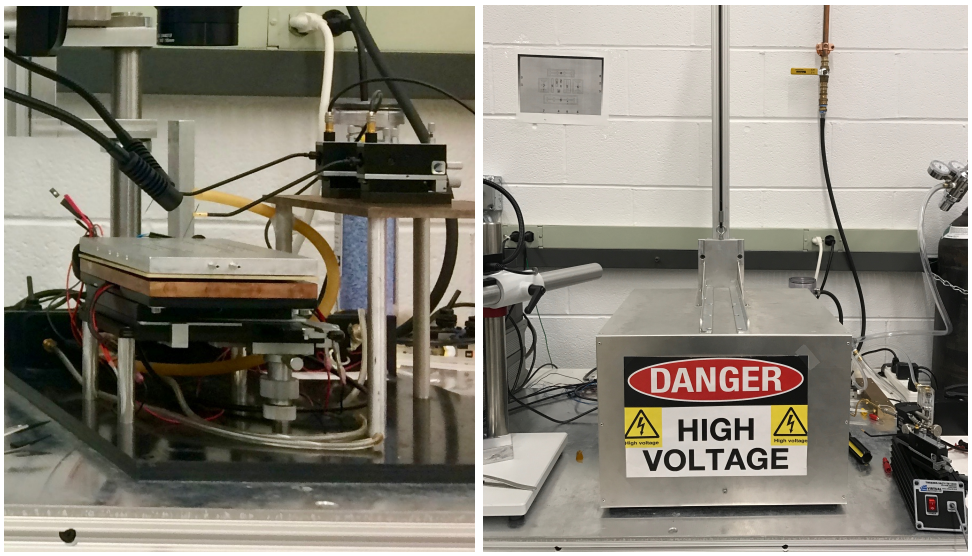

Figure 4.4.2: A close up shot of the measurement (left) and the light-tight box (right).

temperature never reaches below the dew point.

4.5 Measurement

CV and IV curves are measured by ramping up the voltage in representative steps from 0 up to 500 V or 800 V depending on the sensor and the exposed fluence. At each voltage step, the current and capacitance are measured four times and their average values are stored.

As shown in Eq 4.2.4, $1/C^2 \propto V$. Thus, the C-V correlation can be plotted in two ways: $1/C^2$ versus V and $\log(C)$ versus $\log(V)$. Fig 4.5.1 shows an example of the CV-IV measurement at room temperature. In each plot, a slope is used to fit the dependency of $C(V)$ before full depletion and a flat line is used to fit the plateau after full depletion. The intersection of the two fitted lines, labeled by the green dashed line, is taken as the depletion voltage.

As shown comparing the left $(1/C^2 - V)$ and right $(\log(C) - \log(V))$ panels of Fig 4.5.1, the depletion voltage measurement for the two methods agree within statistical uncertainties. In each measurement, the capacitance is measured twice with 1 kHz (shown in the top row) and 10 kHz (shown in the middle row) as a check for frequency dependence. Overall, when comparing the results from the two frequencies, the difference in the measured depletion voltage is negligible.

The measured leakage current is shown as blue dots in the top and middle rows with the gray dots indicating the rescaled leakage current at 20°C with Eq 4.2.6.

The bottom row in Fig 4.5.1 shows the temperature of the chuck during the whole measurement.

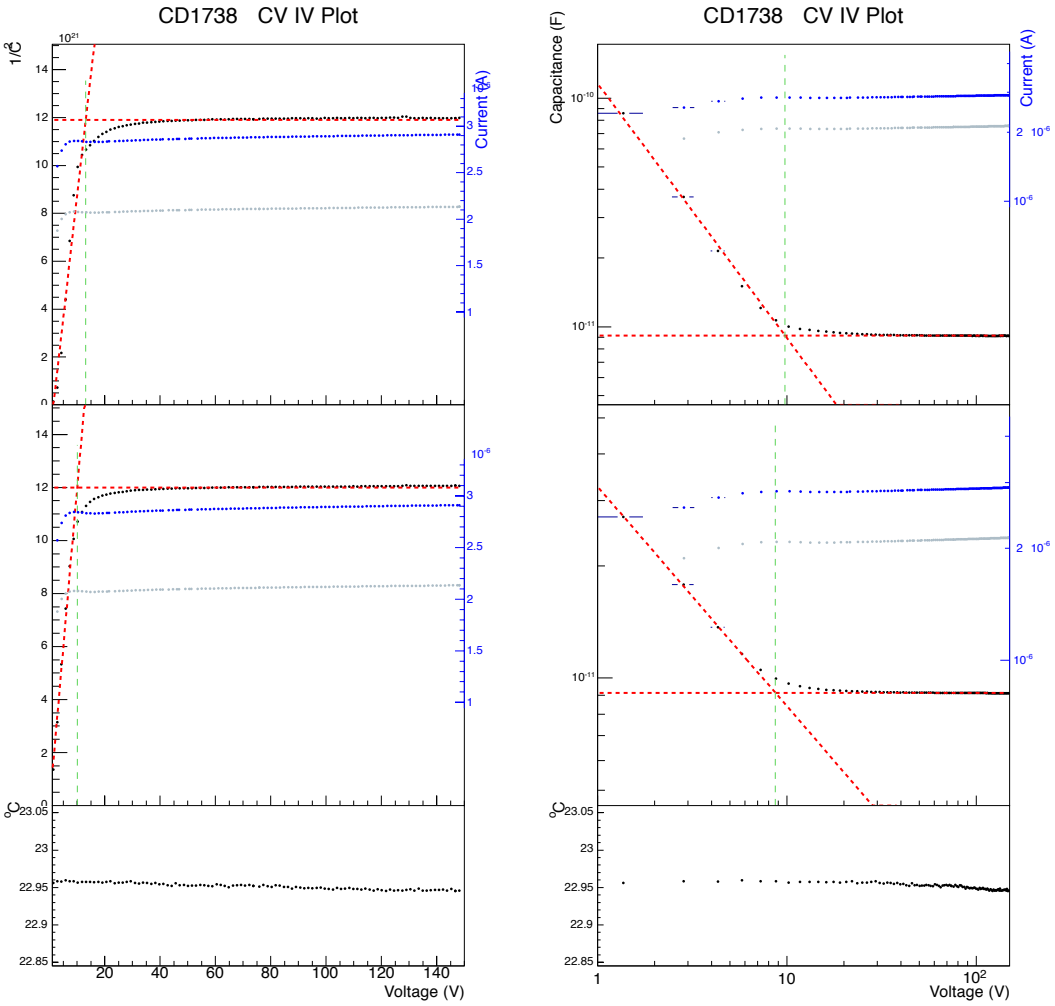

Figure 4.5.1: Illustration of the two methods used to determine the depletion voltage from CV measurements. On the left panel, $1/C^2$ versus V. On the right panel, log(C) versus log(V). At each voltage step, the capacitance measurement is performed twice with 1kHz (top plots) and 10 kHz (bottom plots).

As shown, the temperature is very stable with fluctuations within ± 0.1°C.

4.6 Radiation Damage

The bulk damage in silicon detectors caused by energetic particles is primarily due to displacing a Primary Knock on Atom (PKA) out of its lattice site [65]. Such displacements result in Frankel pairs consisting of a silicon interstitial and a vacancy. Depending on the recoil energy of the PKA, two types of defects are formed. For low-energy recoils, PKA usually forms a point defect. For recoil energies above 5 keV [65], a dense agglomeration of defects is formed at the end of the PKA tracks and these disordered regions are referred as defect clusters.

Both point defects and defect clusters can have significant effects on the detector performance. Depending on their energy levels, defects could act as recombination/generation centers thus increasing the leakage current of the detector, as shown in Eq 4.2.5. The effective doping concentration is also changed as a result of dopant removals by the formation of PKA tracks as well as the generation of charged centers.

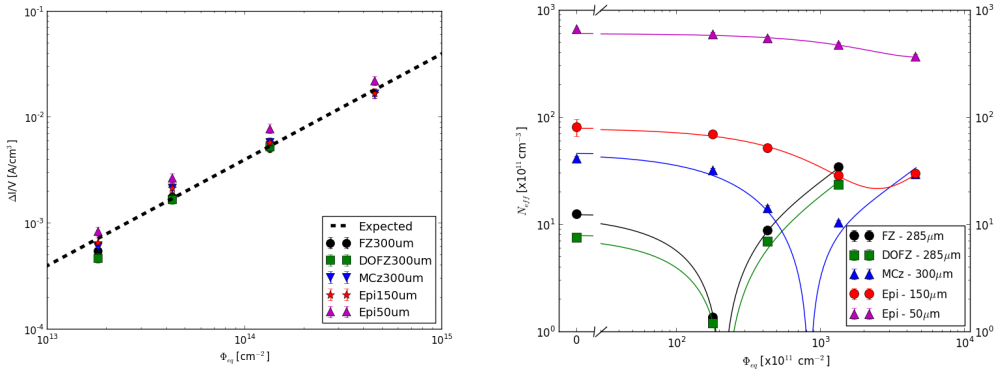

Figure 4.6.1: Figure taken from Ref. [12] as the results of a collaborational work. Left: the increase in volume leakage current (corresponding to the depletion voltage) as a function of fluence after 800 MeV proton irradiation. Right: the effective doping concentration as a function of fluence after 800 MeV proton irradiation.

4.6.1 Effect of Radiation Damage on Leakage Current

After irradiation, silicon sensors first go through a short term annealing stage for 8 mins at 80°C. This step is crucial for it removes the unstable defects and gives more reliable results for further analysis. The leakage current of irradiated sensors is then measured after the first short term annealing stage. The left panel of Fig 4.6.1 shows the increase in volume leakage current, corresponding to the

depletion voltage, with respect to the leakage current before irradiation for the various sensor as a function of fluence after 800 MeV proton irradiation. The leakage currents shown have been scaled to a reference value at 20°C using Eq. 4.2.6. The black dashed line is drawn by transforming Eq 4.2.5 to $\Delta I/(AW) = \alpha\kappa\Phi_{eq}$ with $\alpha = 4.0 \times 10^{-17}$ A/cm [65] and $\kappa = 0.72$ [64]. As shown, the results are in good agreement with the expected leakage current behavior. Additionally, these results also confirm that, in terms of the change in the leakage current, different materials react similarly to 800 MeV proton irradiation.

4.6.2 Radiation Damage on the Effective Doping Concentration

Similar to the leakage current, the effective doping concentration of irradiated sensors is also measured after the first stage of annealing for 8 mins at 80°C. The effective doping concentration is calculated with Eq 4.2.3 after measuring the depletion voltage. The right panel of Fig 4.6.1 shows the measured effective doping concentration as a function of the fluence after 800 MeV proton irradiation. The fit of the data points uses the following function [11]:

$$N_{eff}(\Phi_{eq}) = N_{eff,0} \exp(-c\Phi_{eq}) + \beta_{eff}\Phi_{eq} \quad , \tag{4.6.1}$$

where $N_{eff,0}$ is the effective doping concentration before irradiation, c is the donor removal cross-section and β_{eff} is the probability to create an acceptor state. This is referred to below as the stable damage component $N_C(\Phi_{eq})$ in Eq. 4.7.1. As shown, all sensors start as n-type with their N_{eff} decreasing as the fluence increases. For FZ, DOFZ and MCz sensors, radiation creates so much acceptors that it inverts the sensor type to p-type. This type-inversion happens at around 2×10^{13} MeV n_{eq}/cm^2 for FZ and DOFZ sensors while at around 9×10^{13} MeV n_{eq}/cm^2 for MCz sensors. The values of the fit parameters are shown in Table 4.4.

Sensor	c [10^{-14}cm^{-2}]	$\beta eff[10^{-2}\text{cm}^{-1}]$
FZ	4.07 ± 0.04	-2.54 ± 0.01
DOFZ	3.01 ± 1.14	-1.90 ± 2.66
MCz	2.41 ± 0.62	-0.74 ± 0.12
Epi (150 μm)	1.04 ± 0.06	0.63 ± 0.02
Epi (50 μm)	0.29 ± 0.04	4.41 ± 0.68

Table 4.4: Values of fitting parameters for N_{eff} vs fluence.

4.7 Annealing

It has been known that the radiation damage mentioned in the previous section may be subject to changes after long-term storage at room temperature. Such annealing affects are likely caused by the dissolution of clusters that releases mitigating vacancies and interstitials [65]. In order to study the effects of annealing on the LHC operational scenario (around 10 years of projected operation before changing the silicon detectors) while not to actually wait for 10 years, the effects of annealing was expedited by heating the sensors up to 80°C. Multiple representative annealing steps were taken with an integrated annealing time up to 10,000 minutes. After each annealing step, the CV and IV curves of the sensor are measured and then stored in a freezer at $-20°$ C, to avoid undesired annealing, before the next controlled annealing step.

The N_{eff} dependency on the annealing time t has been well modeled by the Hamburg model as the following [11]:

$$\begin{aligned} \Delta N_{eff}(\Phi_{eq}, t) &= N_{eff,0} - N_{eff}(\Phi_{eq}, t) \\ &= N_C(\Phi_{eq}) + N_A(\Phi_{eq}, t) + N_Y(\Phi_{eq}, t) \quad , \end{aligned} \tag{4.7.1}$$

where $N_C(\Phi_{eq})$ is the stable damage component mentioned in Section 4.6.2 that is independent of the annealing time, $N_A(\Phi_{eq}, t)$ is the short term annealing component and $N_Y(\Phi_{eq}, t)$ is the reverse annealing term.

The mathematical representation of the short term annealing is given as [11]:

$$N_A = \Phi_{eq} g_a \exp\left(-\frac{t}{\tau_a}\right) \quad , \tag{4.7.2}$$

where $g_a = 1.81 \times 10^{-2}$ cm^{-1} is the average introduction rate and the time constant $\tau_a = 2$ min at 80°C.

The mathematical representation of the reverse annealing term can be taken as [11]:

$$N_Y(t) = N_{Y,\infty}\left(1 - \frac{1}{1 + kN_{Y,\infty}t}\right) \quad , \tag{4.7.3}$$

where the $N_{Y,\infty}$ is the reverse annealing amplitude.

Fig 4.7.1 shows the measured results of the change in the effective doping concentration as a function of the annealing time at 80°C for the FZ and DOFZ sensors. The solid lines are fits using the Hamburg model in Eq 4.7.1. As shown, it is obvious that the N_{eff} degradation is proportional

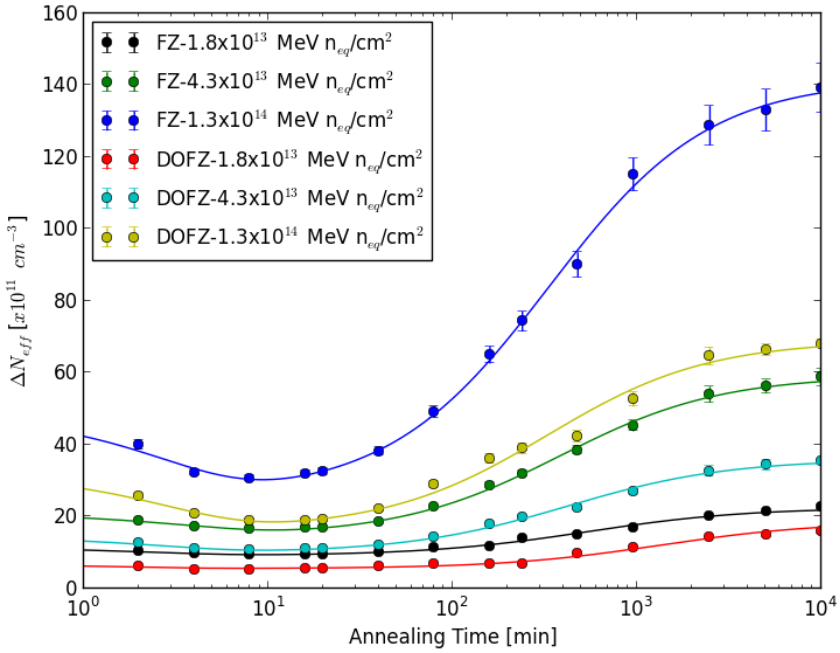

Figure 4.7.1: Figure taken from Ref. [12] as the results of a collaborational work. Hamburg model parameterization of FZ and DOFZ sensors.

to the fluence exposure. Right after the irradiation, there is a beneficial period of annealing where N_{eff} increases. However, in long terms, N_{eff} decreases with annealing and this degradation is more pronounced for heavily irradiated sensors.

Chapter 5

Search for a heavy Higgs decaying to two 125 GeV Higgs in the 2 τ + 2 b final state using data at $\sqrt{s} = 8$ TeV

In this chapter a search for a heavy Higgs boson (H) decaying to two 125 GeV Higgs bosons (h) using the final state of 2 taus and 2 bottom quarks is presented. The analysis uses data collected during 2012 by the CMS experiment with a centre-of-mass energy $\sqrt{s} = 8$ TeV and was published in 2016 [13]. The mass of H is studied in a range from $m_H = 260$ GeV to 350 GeV, between the range of the kinematic threshold for $H \to h(125)h(125)$ and before $H \to t\bar{t}$ begins to dominate. No excess is found above the standard model expectation and a model-independent limit on the cross-section times branching ratio for the $H \to hh \to \tau\tau bb$ process is set.

This chapter is organized as following: Section 5.1 provides a brief introduction to the motivations of the search. Section 5.2 provides an overview of the analysis along with the key background estimation methods and signal extraction techniques. Section 5.3 describes the dataset and Monte Carlo samples used. Section 5.4 describes in details how events are selected in the $\tau_h\tau_h$ channel. The estimation of background contributions is described in Section 5.5, followed by a discussion on the the different techniques used for signal extraction in Section 5.7. Section 5.8 discusses the systematic uncertainties. The statistical procedure used to set limits and their results are detailed in Section 5.9. Finally, a summary is given in Section 5.10 which concludes this chapter.

5.1 Introduction

As discussed in Section 1.4.1, one way to resolve the hierarchy problem is through MSSM theories. This leads to the prediction of five Higgs particles: two charged Higgs $H^{\pm}$, a light scalar Higgs h, a heavy scalar Higgs H and a pseudo-scalar Higgs A. At tree level, the masses and coupling of these bosons can be described by the mass of A, m_A, and the ratio of vacuum expectation values of the two Higgs doublets, $\tan\beta \equiv v_u/v_d$. However, when radiative corrections [37, 38, 66, 67] are considered one would also need to consider other parameters: the mass of the top quark m_t, the scale of the soft supersymmetry breaking M_{SUSY}, the higgsino mass parameter μ, the wino mass parameter M_2, the third-generation trilinear couplings A_{t}, A_{b} and A_τ, the mass of the gluino $m_{\tilde{g}}$, and the third-generation slepton mass parameter $M_{\tilde{l}_3}$.

Some MSSM scenarios [68] with high M_{SUSY} allow the light Higgs h to have a mass consistent with the 125 GeV Higgs discovered at low $\tan\beta$. At these low $\tan\beta$ values, the branching ratio of heavy Higgs H decaying into a pair of light Higgs h is greatly enhanced. Thus, with such models, we propose the existence of a heavy Higgs with mass between 260 GeV to 350 GeV that decays to a pair of SM-like Higgs bosons with the mass of 125 GeV. The lower bound of 260 GeV is chosen such that it allows kinematically the decay into two h bosons. The upper bound of 350 GeV is chosen such that it is smaller than the mass of two top quarks. The decay of H to two top quarks would become the dominant decay mode, if m_H is large enough, due to the large mass of the top quark.

Searches for $H \to hh$ have been performed by the ATLAS [69–71] and CMS collaborations [72–74] in diphoton + multi-lepton, diphoton + bb, and four b final states. In this analysis, we present a search of the heavy Higgs H decaying into a pair of SM Higgs h where one of the h decays into a pair of τ leptons while the other h decays into a pair of bottom quarks.

5.2 Strategy

In this analysis, similar techniques used in SM Higgs to $\tau\tau$ searches [75] are used while additionally requiring the presence of two b-jets. Out of the possible decay modes of the tau lepton, the three most sensitive decay channels, $e\tau_h$, $\mu\tau_h$ and $\tau_h\tau_h$, are studied. In this thesis, only the $\tau_h\tau_h$ will be presented in detail as the channel I've personally worked on.

The dominant background in the $\tau_h\tau_h$ channel are QCD multi-jets with jets misidentified as τ_h. In order to properly estimate their contributions, a data-driven technique is used where the QCD shape and yields are estimated with data from a signal depleted region. Events with a W boson and

additional jets (later referred as W+jets), where at least one jet is misidentified as τ_h, is a minor background for the $\tau_h\tau_h$ channel and is mostly estimated by MC simulations. However, for some regions where the MC statistics is too low, W+jets backgrounds are estimated together with the QCD component as they share similar mechanism of jets faking τ_h. The rest of the backgrounds are estimated directly from MC simulations. Details about the background estimations can be found in Section 5.5.

In order to better separate the $h \to \tau\tau$ events from $Z \to \tau\tau$ events, the di-tau mass $m_{\tau\tau}$ is calculated using a likelihood approach with the SVFit algorithm [75]. The model makes a prediction for the probability density $p(\vec{x}|\vec{y},\vec{a})$ to observe the values $\vec{x} = (E_x^{\text{miss}}, E_y^{\text{miss}})$ measured in an event, given that the unknown parameters specifying the kinematics of the tau pair decay have values $\vec{a} = (f_1, \Phi_1, m_1^{\nu\nu}, f_2, \Phi_2, m_2^{\nu\nu})$ and the momenta of the visible decay products are equal to the observed $\vec{y} = (p_1^{vis}, p_2^{vis})$. Here, 1 and 2 refers to the two reconstructed lepton candidates, in this channel the two τ_h. f is the fraction of the tau lepton energy, in the laboratory frame, carried by visible decay products. Φ is the azimuthal angle of the tau lepton in the laboratory frame. $m^{\nu\nu}$ is the mass of the neutrino system from tau decays which is 0 for the τ_h case and a combination of $\nu_{e/\mu}$ and ν_τ for leptonically decaying taus [76]. The $m_{\tau\tau}$ giving the maximum probability is taken and referred as the svMass.

Two approaches were studied for signal and background separation and upper limit extraction. One method attempts to reconstruct the invariant mass of the heavy Higgs H by performing a likelihood fit using the information of the two τ_h and 2 b-jets. The other method uses Boosted Decision Trees (BDT) to maximize the signal-background discrimination power. In the end, both of these methods yield similar results and will be discussed in detail in Section 5.7

5.3 Datasets and Monte Carlo samples

In this analysis, 8 TeV data collected during 2012 with the CMS detector is used and shown in Table 5.2. The recommended file (`Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON.txt`) containing only certified events is used to filter out bad events for all of the data.

The list of MC samples with their corresponding theoretical cross-sections are summarized in Table 5.2. Samples of $gg \to H \to hh \to \tau\tau bb$ signal events have been simulated using PYTHIA [77] with invariant mass of m_H in 10 GeV steps from 260 GeV to 350 GeV. For background samples, $Z/\gamma^* \to ee$, $Z/\gamma^* \to \mu\mu, t\bar{t}$ and diboson (WW, WZ, and ZZ) samples are generated by MADGRAPH [78] while the single top samples are generated by POWHEG [79]. In order to improve background

Dataset Name	Integrated Luminosity [fb^{-1}]
/Tau/Run2012A-22Jan2013-v1/AOD	0.887
/TauParked/Run2012B-22Jan2013-v1/AOD	4.446
/TauParked/Run2012C-22Jan2013-v1/AOD	7.153
/TauParked/Run2012D-22Jan2013-v1/AOD	7.318

Table 5.1: List of dataset analyzed in the $\tau_h\tau_h$ channel.

event statistics in the signal region, samples binned in jet multiplicity at parton level are used for $Z/\gamma^* \to ll$ ($l = e, \mu$) and W + jets backgrounds. Taus are decayed by TAUOLA [80]. As listed, $Z/\gamma^* \to ll$, single-top and di-boson events are normalized to their corresponding theoretical cross-sections. The $t\bar{t}$ events are normalized to the CMS and ATLAS measured cross-section [81].

As discussed in Section 3.1, in an event, many collisions take place resulting in dozens of pile-up interactions. These pile-up interactions may affect the calculation of quantities such as $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ and isolation, where energy depositions are summed up over some range in the detector. In order to simulate this inefficiency MC samples are generated with pile-up simulations. To further correct the difference in the pile-up distribution between the simulated profile and results from data, an event based pile-up reweighting is performed.

In order to have the most accurate modeling of $Z/\gamma^* \to \tau\tau$ events, a special technique referred as embedding is used. In short, embedded samples are produced by selecting $Z/\gamma^* \to \mu\mu$ events in data and replacing the reconstructed muons by generator level tau leptons [82]. These taus are then decayed using TAUOLA with their polarization effects modeled by TAUSPINNER [83]. The detector responses simulation for the tau decay production is modeled by GEANT4 [84]. After removing all of the detector response of the two muons originating from Z/γ^* the tracker hits and energy deposits of the simulated taus are embedded. In order to maximize event statistics, separate sets of embedded samples are produced for each tau pair decay channel. Table 5.3 shows the list of embedded samples used for the $\tau_h\tau_h$ channel. In order to remove the small amount of contamination from $t\bar{t}$ events in the embedded samples, a separate $t\bar{t}$ MC sample with embedded τ_hs, also shown in Table 5.3, is used to estimate this contamination and subtract it from the total $Z/\gamma^* \to \tau_h\tau_h$ prediction.

5.4 Event Selection

Following the recommendation of the Tau Physics Object Group (POG), events in the $\tau_h\tau_h$ channel are selected by requiring:

Signal Monte Carlo samples		
Dataset Description	Dataset Name	Cross-section [pb]
$gg \to H \to hh \to \tau\tau bb$	/GluGluToHTohhTo2Tau2B_mH-*_mh-125_8TeV-pythia6-tauola[1]	–
Background Monte Carlo samples		
Dataset Description	Dataset Name	Cross-Section [pb]
$t\bar{t}$	/TTJets_MassiveBinDECAYTTJets_TuneZ2star_8TeV-madgraph-tauola[1]	241.5
Single top	/T_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola[1]	22.2
	/Tbar_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola[1]	22.2
$Z \to ll$	/DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball[1]	3503.7
$Z + 1$jet	/DY1JetsToLL_M-50_TuneZ2Star_8TeV-madgraph[1]	666.3
$Z + 2$jets	/DY2JetsToLL_M-50_TuneZ2Star_8TeV-madgraph[1]	215.0
$Z + 3$jets	/DY3JetsToLL_M-50_TuneZ2Star_8TeV-madgraph[1]	60.7
$Z + 4$jets	/DY4JetsToLL_M-50_TuneZ2Star_8TeV-madgraph[1]	27.3
W+jets	/WJetsToLNu_TuneZ2Star_8TeV-madgraph-tarball[1]	36257.2
$W + 1$jet	/W2JetsToLNu_TuneZ2Star_8TeV-madgraph-tarball[1]	6381.2
$W + 2$jets	/W2JetsToLNu_TuneZ2Star_8TeV-madgraph-tarball[1]	2039.8
$W + 3$jets	/W3JetsToLNu_TuneZ2Star_8TeV-madgraph-tarball[1]	612.5
$W + 4$jets	/W4JetsToLNu_TuneZ2Star_8TeV-madgraph-tarball[1]	251.0
WW	/WWJetsTo2L2Nu_TuneZ2star_8TeV-madgraph-tauola[1]	5.8
WZ	/WZJetsTo3LNu_TuneZ2_8TeV-madgraph-tauola[1]	1.1
WZ	/WZJetsTo2L2Q_TuneZ2star_8TeV-madgraph-tauola[1]	2.2
ZZ	/ZZJetsTo4L_TuneZ2star_8TeV-madgraph-tauola[1]	0.2
ZZ	/ZZJetsTo2L2Nu_TuneZ2star_8TeV-madgraph-tauola[1]	0.7
ZZ	/ZZJetsTo2L2Q_TuneZ2star_8TeV-madgraph-tauola[1]	2.5

[1] `/Summer12-DR53X-PU_S10_START53_V7A-v1/AODSIM`

Table 5.2: List of Monte Carlo samples used to model $gg \to H \to hh \to \tau\tau bb$ signal as well as $Z/\gamma^* \to ee$, $Z/\gamma^* \to \mu\mu$, W + jets, $t\bar{t}$, single top and di-boson (WW, WZ, ZZ) backgrounds.

$Z/\gamma^* \to \tau_h\tau_h$ embedded samples
/DoubleMu/StoreResults-Run2012A_22Jan2013_v1_[1]
/DoubleMu/StoreResults-Run2012B_22Jan2013_v1_[1]
/DoubleMu/StoreResults-Run2012C_22Jan2013_v1_[1]
/DoubleMu/StoreResults-Run2012D_22Jan2013_v1_[1]
$t\bar{t}$ embedded samples
/TTJets_FullLeptMGDecays_8TeV-madgraph-tauola/
StoreResults-Summer12_TTJets_FullLeptMGDecays_DR53X_PU_S10_START53_V7C_v2_[1]

[1] `PFembedded_trans1_tau132_pthad1_30had2_30_v1-5ef1c0fd428eb740081f19333520fdc8/USER`

Table 5.3: List of the embedded samples.

- event to pass the $\tau_h\tau_h$ trigger listed in Table 5.4;
- two reconstructed τ_h each with:
  - $p_T > 45$ GeV, $|\eta| < 2.1$;
  - matched to HLT tau objects within ΔR < 0.5;
  - pass the "Medium" working point of the HPS combined isolation discriminator;

- the τ_h with lower p_T is required to pass the "Loose" working point of the MVA3 anti-electron discriminator;
- in case of more than two τ_hs passing the selections above:
  - the pair with the lowest HPS combined isolation discriminator raw outputs is chosen;
  - in case there are still more than one pair, the pair with the highest scalar sum p_T is chosen;
- the selected τ_hs are required to have opposite electric charge;
- at least two jets with $p_T > 20$ GeV, $|\eta| < 2.4$, are present in the event;
- events are rejected if there is an electron or muon;

Sample	Trigger
data and MC	two τ_hs passing the "Medium" working point of the HPS combined isolation discriminator with $|\eta| < 2.1$

Table 5.4: Trigger used in the $\tau_h\tau_h$ channel.

Events are further categorized by the presence of b-jets. In this analysis, the CSV algorithm discussed in Section 3.4.1 is used. In order to enhance selection of signal over background, jets are ordered by their CSV discriminator value such that the leading jet is the most b-jet like in the event. Then, events are separated in the following categories:

- 2 jet - 0 tag: Events in this category have no jets passing the "Medium" working point of the CSV discriminator. This category contains a very small amount of signal and is dominated by backgrounds. Thus, it mostly serves as a control region to help better model the various backgrounds.
- 2 jet - 1 tag: Events in this category have only the leading jet passing the "Medium" working point of the CSV discriminator.
- 2 jet - 2 tag: Events in this category have at least two jets passing the "Medium" working point of the CSV discriminator. This is the most signal-sensitive category.

5.5 Background Estimation

The dominant background in the $\tau_h\tau_h$ channel, especially in the 2 jet 2 tag category, is from $t\bar{t}$ decays followed by QCD multi-jet and Drell-Yan $Z \to \tau_h\tau_h$ backgrounds. Whenever possible, a data-driven

estimation of the background is applied. This section gives a short description of the background estimation for each source.

5.5.1 $t\bar{t}$ background

The $t\bar{t}$ contribution is estimated using MC samples. A validation of the $t\bar{t}$ modeling is performed in the $e\mu$ channel for its high $t\bar{t}$ purity.

In the $e\mu$ channel, events are selected by requiring two opposite charged leptons, one electron and one muon, where the highest (second-highest) p_T lepton is required to have $p_T > 20$ (10) GeV. In order to improve the $t\bar{t}$ acceptance and reduce signal contamination, events are further required to contain at least 2 jets with at least one of them passing the "Medium" CSV working point and $E_{\mathrm{T}}^{\mathrm{miss}} > 80$ GeV. $t\bar{t}$ background in this control region achieved a purity of 91%, with the remaining 9% of the background dominated by di-boson events.

By comparing the observed data and MC estimations, a scale factor was derived for the $t\bar{t}$ MC samples to account for the differences. It was estimated to be $1.033 \pm 0.013(\mathrm{stat.}) \pm 0.088(\mathrm{syst.})$ which agrees well with unity.

Thus, $t\bar{t}$ backgrounds are estimated directly with MC samples without additional correction on the $t\bar{t}$ cross-section.

5.5.2 QCD background

QCD events pass the event selections if two jets are misidentified as τ_h's. Thus, if one knows the jet to τ_h fake rate, QCD can be estimated with data in a loose tau requirement region where QCD is dominant. Based on the fact that QCD events consists a collection of jets with random electric charge, the two jets faking the two jets faking τ_h's are as likely to have the same electric charges as opposite electric charges, events in each category are split into the following four regions:

- Opposite-sign (OS) isolated: Events in this region have two τ_h candidates with opposite electric charge and both of them passing the "Medium" working point of the HPS combined isolation 3-hits discriminator.

- Same-sign (SS) isolated: Events in this region have two τ_h candidates with same electric charge and both of them passing the "Medium" working point of the HPS combined isolation 3-hits discriminator.

- OS non-isolated: Events in this region have two τ_h candidates with opposite electric charge

and one or both of them satisfies the relaxed τ_h isolation requirements, to be discussed in more detail, in order to enhance the QCD purity.

- SS non-isolated: Events in this region have two τ_h candidates with same electric charge and one or both of them satisfies the relaxed τ_h isolation requirements.

The first region is the signal region and the latter three regions serve as control regions.

Thus, the template for the QCD distribution is estimated from the OS non-isolated region by subtracting observed events with the MC simulated contributions from other non-QCD processes. For normalization, the yield estimated in the OS non-isolated region is multiplied by a jet to τ_h fake rate estimated among the SS regions. This is summarized as the following:

$$\begin{aligned} \text{Shape} &= \text{Histogram(OS non-isolated data)} - \text{Histogram(OS non-isolated non-QCD MC)} \\ \text{Yield} &= (N^{\text{Observed}}_{\text{OS non-isolated}} - N^{\text{non-QCD MC}}_{\text{OS non-isolated}}) \times \frac{N^{\text{Observed}}_{\text{SS isolated}} - N^{\text{non-QCD MC}}_{\text{SS isolated}}}{N^{\text{Observed}}_{\text{SS non-isolated}} - N^{\text{non-QCD MC}}_{\text{SS non-isolated}}} \end{aligned} \tag{5.5.1}$$

Regarding the relaxed τ_h isolation requirements, eight variations were studied based on:

- the working point for the b-tagging, "Medium" or "Loose", labeled as "M" and "L";
- the number of τ_h's to be non-isolated, "one" or "both";
- the definition of the non-isolation range: 1 - 4 GeV or 3 - 10 GeV for the raw value of HPS combined isolation 3-hits.

Results of this study are shown in Fig 5.5.1 for the 2 jet -1 tag and 2 jet - 2 tag categories as they are the most signal sensitive. The y-axis of Fig 5.5.1 is the estimated QCD yield in the OS isolated region based on the eight variations of non-isolated sideband definitions. As shown, aside from the "M; one 1-4" case in the 2 jet - 2 tag category, referring to "Medium" b-tag working point with both τ_h having their raw HPS combined isolation 3-hits value to be between 1 to 4 GeV, all the predictions of the various non-isolation sideband definitions agree within statistics. In the end, the non-isolation region of:

- "Loose" working point for the b-tagging;
- only one τ_h is required to be non-isolated while the other one is required to pass the "Medium" working point for the HPS combined isolation 3-hits discriminator;
- the non-isolated requirement is set to be between 1 - 4 GeV;

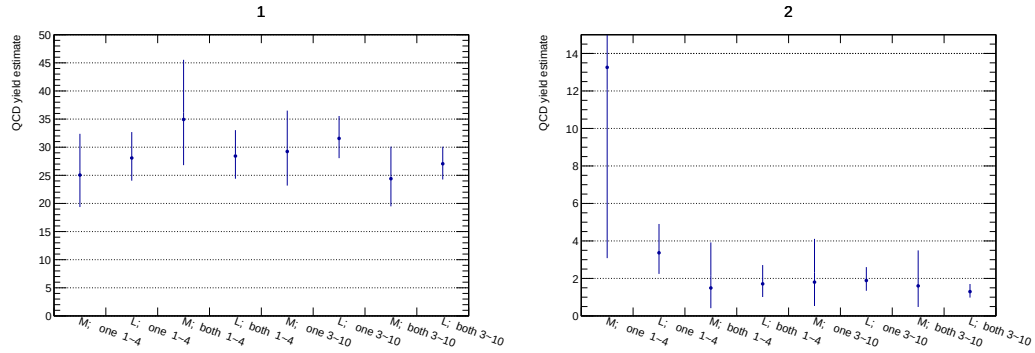

Figure 5.5.1: Estimates of the rate of events with two fake τ_h's, after the mass windows cuts of 90 GeV $< m_{\tau\tau} <$ 150 GeV and 70 GeV $< m_{bb} <$ 150 GeV Left: 2 jet - 1 tag category; Right: 2 jet - 2 tag category. Each point represents the expected QCD yield predicted by a non-isolation region definition in events per 19.7/fb. Uncertainties shown are from control sample statistics. The rate estimate is independent of the choice of non-isolation region definition, within the statistical uncertainties.

is chosen to agree with previous analyses while increasing the event statistics of the QCD prediction.

One underlying assumption for this data-driven method requires that the jet to τ_h fake rate is the same in OS and SS regions. This assumption is verified in the 2 jet - 0 tag category, as it is dominated by QCD backgrounds and the signal yield is expected to be negligible. By comparing:

$$f_{\text{fake rate}} = \frac{N^{\text{Observed}}_{\text{isolated}} - N^{\text{non-QCD MC}}_{\text{isolated}}}{N^{\text{Observed}}_{\text{non-isolated}} - N^{\text{non-QCD MC}}_{\text{non-isolated}}} \tag{5.5.2}$$

estimated in OS and SS regions, the $f_{\text{fake rate}}(\text{OS})$ estimated in the OS regions is found to be 0.25 ± 0.02 (stat.) while the $f_{\text{fake rate}}(\text{SS})$ estimated in the SS regions is 0.23 ± 0.02 (stat.).

5.5.3 Drell-Yan $Z \to \tau\tau$

The Drell-Yan $Z \to \tau\tau$ background is often referred as an "irreducible" background as it contains two genuine τ_h's. Thus, this is a non-trivial background for the 2 jet - 0 tag category but much suppressed by the requirements of b-jets in the 2 jet - 1 tag and 2 jet - 2 tag categories. As described in Section 5.3, embedded samples are used to better model the contribution of $Z \to \tau\tau$. Thus, shape template is taken from the embedded samples. The overall yield is estimated by calculating the inclusive (with requirement on jets) yield using Drell-Yan $Z \to \tau\tau$ MC, and then scaling it by the efficiency for inclusive events to pass the category selection, which is estimated using the embedded sample.

Table 5.5: Expected event yields from standard model processes, and the observed event yield for the baseline selection and an integrated luminosity of 19.7 fb^{-1}.

Processes	Categories		
gg→H→hh	2jet-0tag	2jet-1tag	2jet-2tag
Z/γ*→$\tau\tau$	756.2	96.4	1.5
Z/γ*→ll (l = e, μ)	34.6	3.3	0.49
QCD + WJets	3103.2	303.1	20.8
$t\bar{t}$	33.1	66.5	39.8
Single top + di-boson	30.3	12.7	3.9
Total background	3957.4	482.0	66.5
Observed	4022	468	69

5.5.4 W + jets

Events with a W boson and additional jets represent a small background in the $\tau_h\tau_h$ channel as it requires at least one jet to be misidentified as a τ_h. In the 2 jet - 0 tag category, the W + jets background is estimated by MC simulations. Due to low event statistics in the 2 jet -1 tag and 2 jet - 2 tag categories and a similar jet to τ_h faking mechanism as QCD, W + jets contribution in these categories are estimated together with QCD.

5.5.5 Other small backgrounds

The remaining small backgrounds are from $Z \to ll$ ($l = e, \mu$), di-boson and single top processes. Single top events are combined with di-boson events in the plots. MC samples are used to estimate their shape and yields.

5.6 Event yields and control plots

The expected event yields from SM processes and the observed event yield after the baseline selections described in Section 5.4 are shown in Table 5.5. This forms the basis of the signal extraction methods to be described in Section 5.7. As a first check for background modeling, Fig 5.6.1 shows the p_T distribution of the two τ_h candidates with the signal sample of $m_H = 260$ GeV, in the categories: 2 jet - 0 tag, 2 jet - 1 tag and 2 jet - 2 tag. The error bars in the ratio plots contain statistical uncertainties from both the observed data as well as the estimated backgrounds.

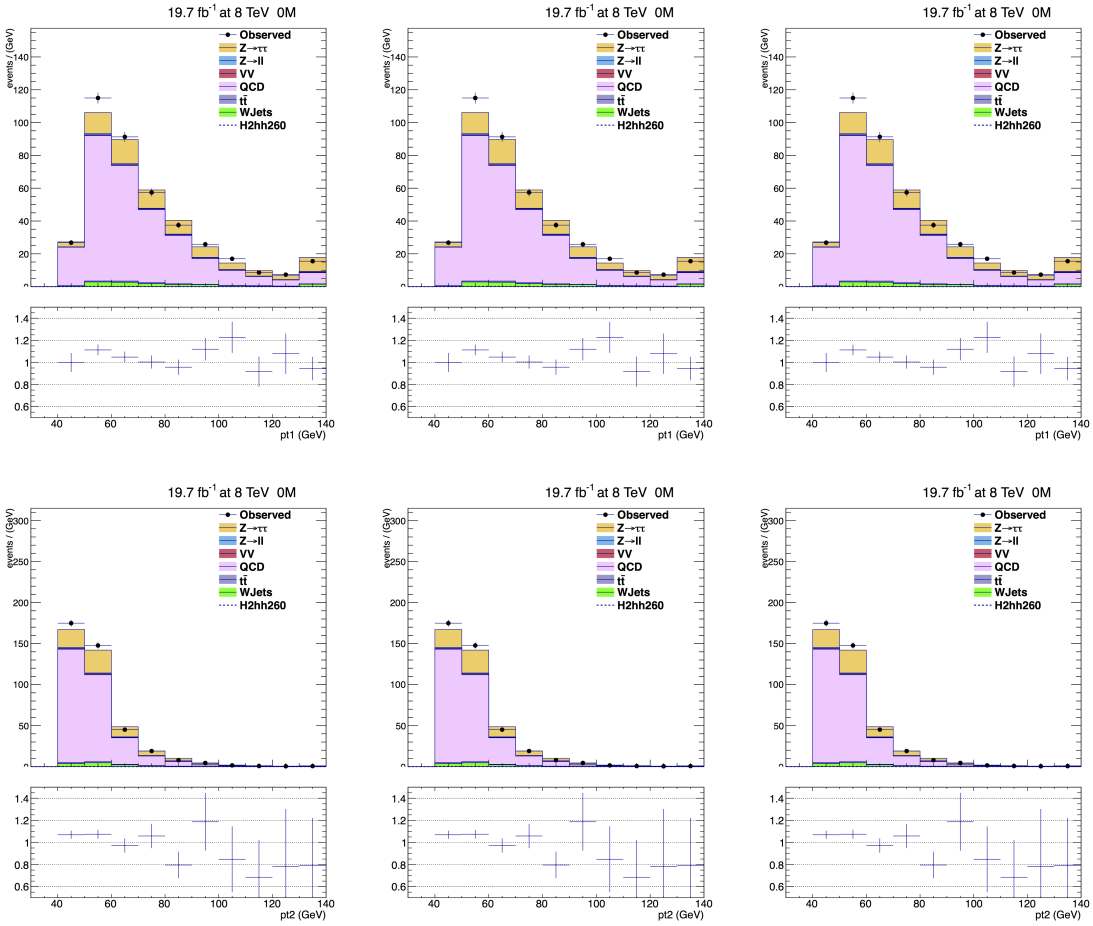

Figure 5.6.1: The p_T distributions of the selected τ_h candidates with baseline selections. Overflows are included in the last bin. Top: leading τ_h's p_T; Bottom: subleading τ_h's p_T. Left: 2 jet - 0 tag category; middle: 2 jet - 1 tag category; right: 2 jet - 2 tag category.

Table 5.6: Yields after the baseline selection and mass window requirements, per 19.7 fb^{-1}.

Processes	Categories		
gg→H→hh	2jet-0tag	2jet-1tag	2jet-2tag
Z/γ*→$\tau\tau$	92.4	8.0	0.55
Z/γ*→ll (l = e, μ)	3.4	0.31	0.19
QCD + WJets	278.7	27.8	4.2
$t\bar{t}$	3.1	6.7	3.3
Single top + di-boson	3.7	1.1	0.72
Total background	381.3	43.9	9.0
Observed	411	45	69

5.7 Signal extraction

As mentioned in Section 5.2, two approaches were studied to improve the separation of signal and background. One method, referred as the kinematic fit, attempts to reconstruct the invariant mass of the heavy Higgs H by performing a likelihood fit using the information of the two τ_h and 2 b-jets. The other method uses BDT to maximize the signal-background discrimination power.

5.7.1 Kinematic Fit

Kinematic fit uses the constraint that b-jets and τ's in signal events originate from SM Higgs, such that $m_{bb} = m_{\tau\tau} = 125$ GeV. It then shifts the energy of each object within their known resolution to bring the mass of each individual pair as close as possible to 125 GeV. Detailed description of the kinematic fit tool can be found in the supplemental material of [13].

To further enhance signal and background separation a mass window requirement of $70 < m_{bb} < 150$ GeV and $90 < m_{\tau\tau} < 150$ GeV is imposed and the expected event yields from SM processes, and the observed event yield are shown in Table 5.6.

The kinematic fit distributions with the mass window selection for the different categories are shown in Fig 5.7.1.

5.7.2 BDT Approach

In hopes to further enhance the separation of signal and background, a machine learning study was conducted using the TMVA package [85]. After optimization studies, the Boosted Decision Tree (BDT) method with 150 trees, each tree with a depth of 3, and the AdaBoost method with a boosting parameter of 0.5 is used. In order to maximize performance, a separate BDT training

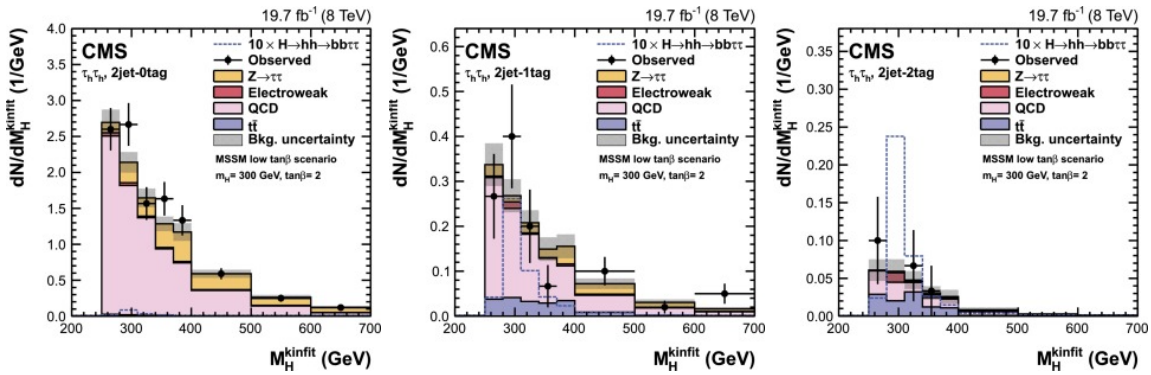

Figure 5.7.1: Distributions of the reconstructed four-body mass with the kinematic fit after applying mass selections on $m_{\tau\tau}$ and m_{bb} in the $\tau_h\tau_h$ channel. The plots are shown for events in the 2 jet - 0 tag (top), 2 jet - 1 tag (middle), and 2 jet - 2 tag (bottom) categories. The expected signal scaled by a factor 10 is shown superimposed as an open dashed histogram for tanβ = 2 and m_H = 300 GeV in the low tanβ scenario of the MSSM. Expected background contributions are shown for the values of nuisance parameters (systematic uncertainties) obtained after fitting the background only hypothesis to the data. [13]

was conducted for each signal mass point ranging from 260 to 350 GeV in 10 GeV steps. For each mass point, we randomly selected half of our samples for training the BDT and used the other half for testing and limit extraction. Fig 5.7.2 shows the comparison of the BDT distribution in training and testing samples for the various mass points. The good agreement in the BDT distribution for background samples shows that we are most likely safe from overtraining. Due to the low number of events, the shapes of training and testing sample for signal are slightly different.

Initially, 18 input variables, as listed in Table 5.7, were considered for the BDT. An optimization study was then conducted to reduce the number of input variables, especially variables with high correlation, while maintaining a good discrimination power. An example of this study is shown in Fig 5.7.3 where we compare the background acceptance rate for BDTs trained using different sets of input variables at the signal efficiency of 80%. In the end, 7 input variables, variables with an "x" mark in Table 5.7, were chosen that have almost the same performance as the original 18 input variables.

The final BDT distributions of the expected background, signal and observed data for each mass point are shown in Figs. 5.7.4–5.7.6.

5.8 Systematics

The systematic uncertainties considered in this analysis consist of two types:

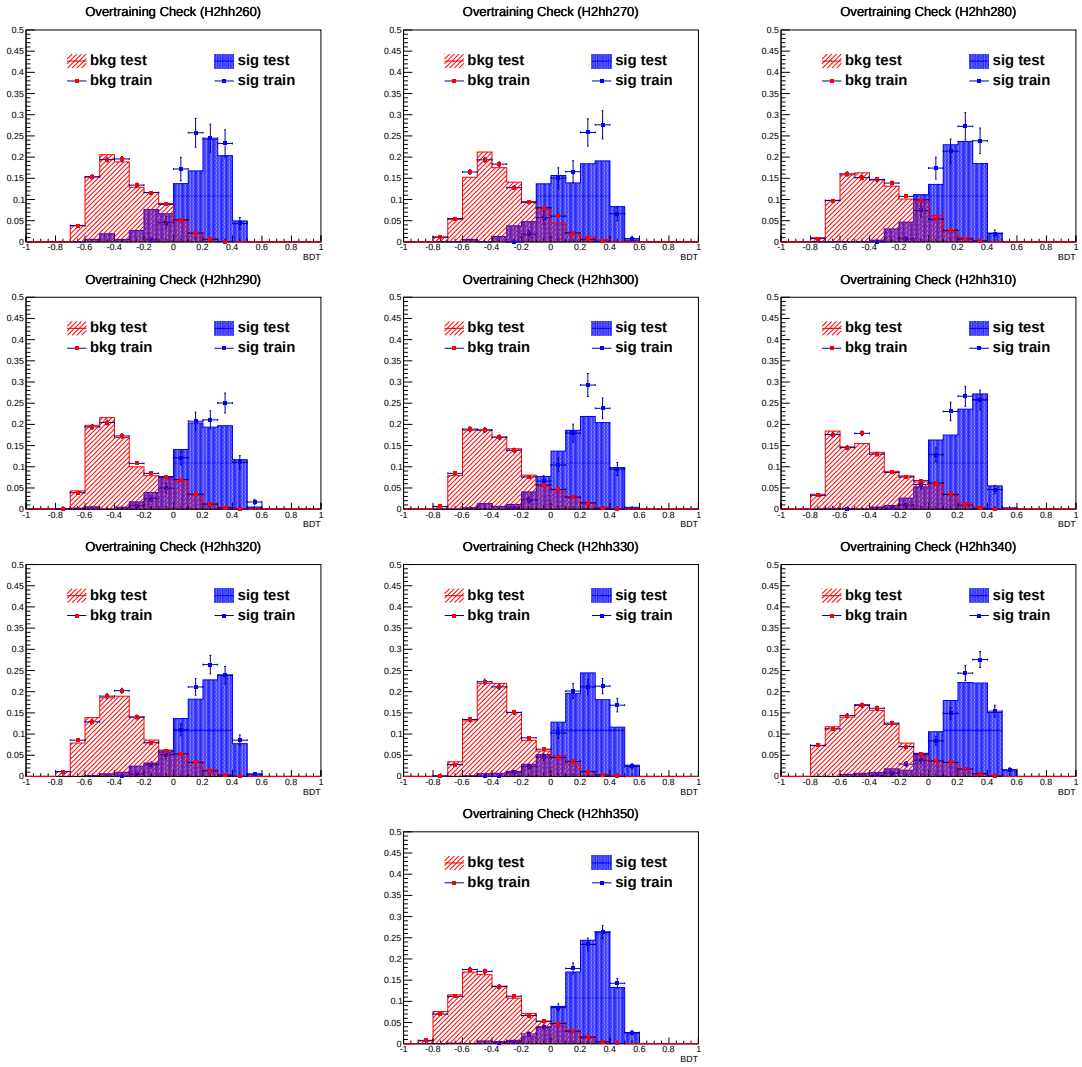

Figure 5.7.2: Distribution of the BDT discriminant for signal and background from the testing and training samples for $m_H = 260$ to 350 GeV in 10 GeV steps.

rank	variable	subset
1	svMass	x
2	$\Delta R(\tau_1, \tau_2)$	x
3	4-body mass from kinematic fit	x
4	$E_{\mathrm{T}}^{\mathrm{miss}}$	x
5	$b\bar{b}$ mass	x
6	$\Delta\phi(\vec{E}_{\mathrm{T}}^{\mathrm{miss}},$ jet 2)	
7	χ^2 from kinematic fit	x
8	subleading CSV discriminant	
9	leading CSV discriminant	
10	$\Delta\phi(\vec{E}_{\mathrm{T}}^{\mathrm{miss}},$ di-jet system)	
11	ΔR(jet 1, jet 2)	x
12	$\Delta\phi(\vec{E}_{\mathrm{T}}^{\mathrm{miss}},$ p_T sub-leading τ)	
13	$\Delta\phi(\vec{E}_{\mathrm{T}}^{\mathrm{miss}},$ di-τ system from SV fit)	
14	$\Delta R(\tau\tau, b\bar{b})$	
15	$\Delta\phi(\vec{E}_{\mathrm{T}}^{\mathrm{miss}},$ jet 1)	
16	$\Delta\phi(\vec{E}_{\mathrm{T}}^{\mathrm{miss}},$ di-τ system)	
17	$\Delta\phi(\vec{E}_{\mathrm{T}}^{\mathrm{miss}}, p_T$ leading τ)	
18	di-τ p_T from SV fit	

Table 5.7: List of variables considered for multivariate analysis in order of decreasing discrimination power, as determined by the BDT for $m_H = 300$ GeV. For other masses the ranking varies somewhat. Those with an "x" in the second column were used in the final training.

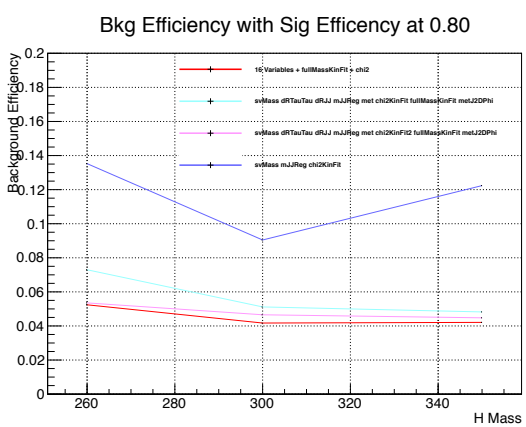

Figure 5.7.3: Background efficiency for BDTs trained using different sets of input variables for a fixed signal efficiency of 80%.

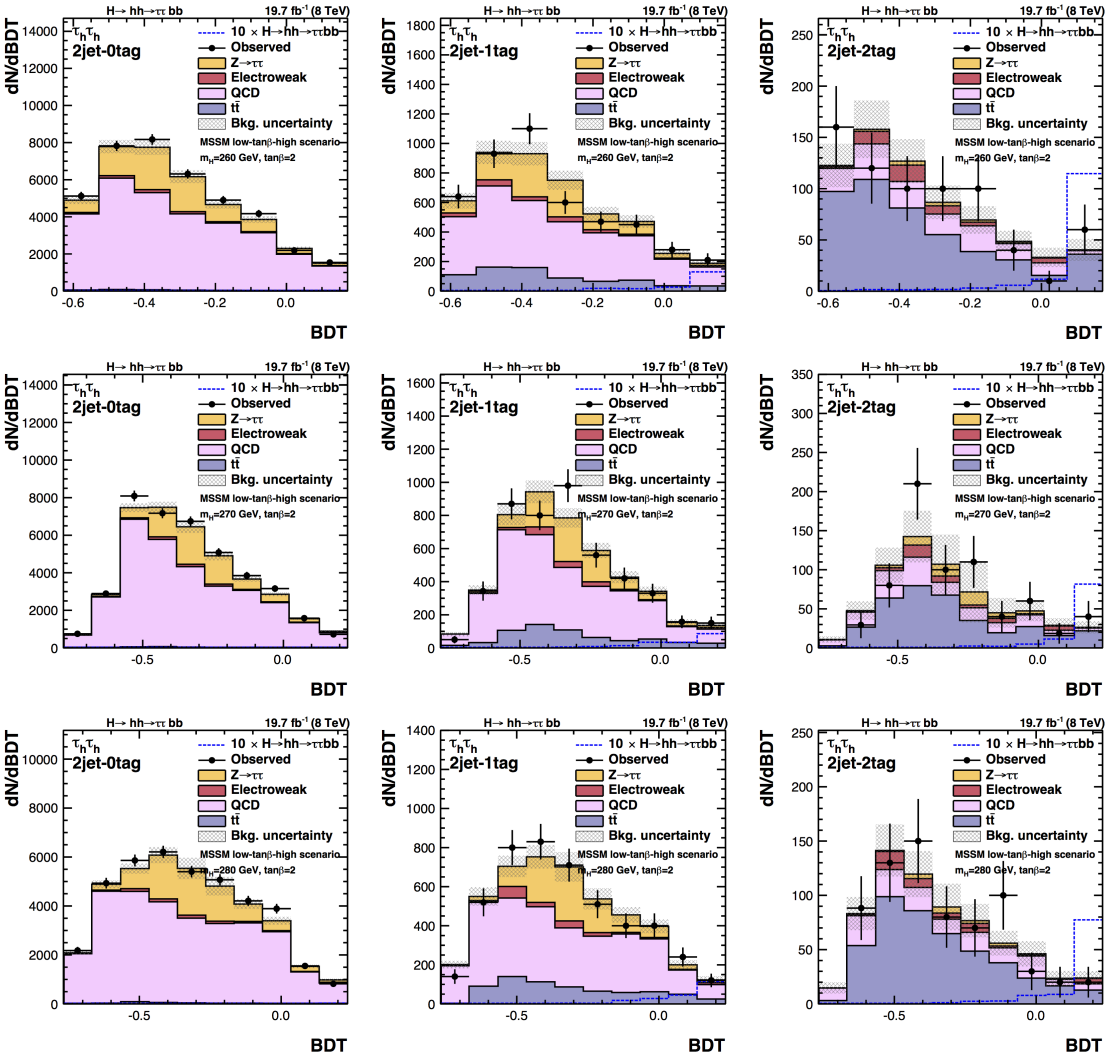

Figure 5.7.4: Distributions of the BDT discriminants from collider data and estimated backgrounds (histogram) with overflows included in the last bin. Left: 0-tag category; middle: 1-tag category; right: 2-tag category. The BDTs were trained at top row: $m_H = 260$ GeV, middle row: $m_H = 270$ GeV, bottom row: $m_H = 280$ GeV. The expected signal scaled by a factor 10 is shown superimposed as an open dashed histogram in the low tanβ scenario of the MSSM. Expected background contributions are shown for the values of nuisance parameters (systematic uncertainties) obtained after fitting the background only hypothesis to the data.

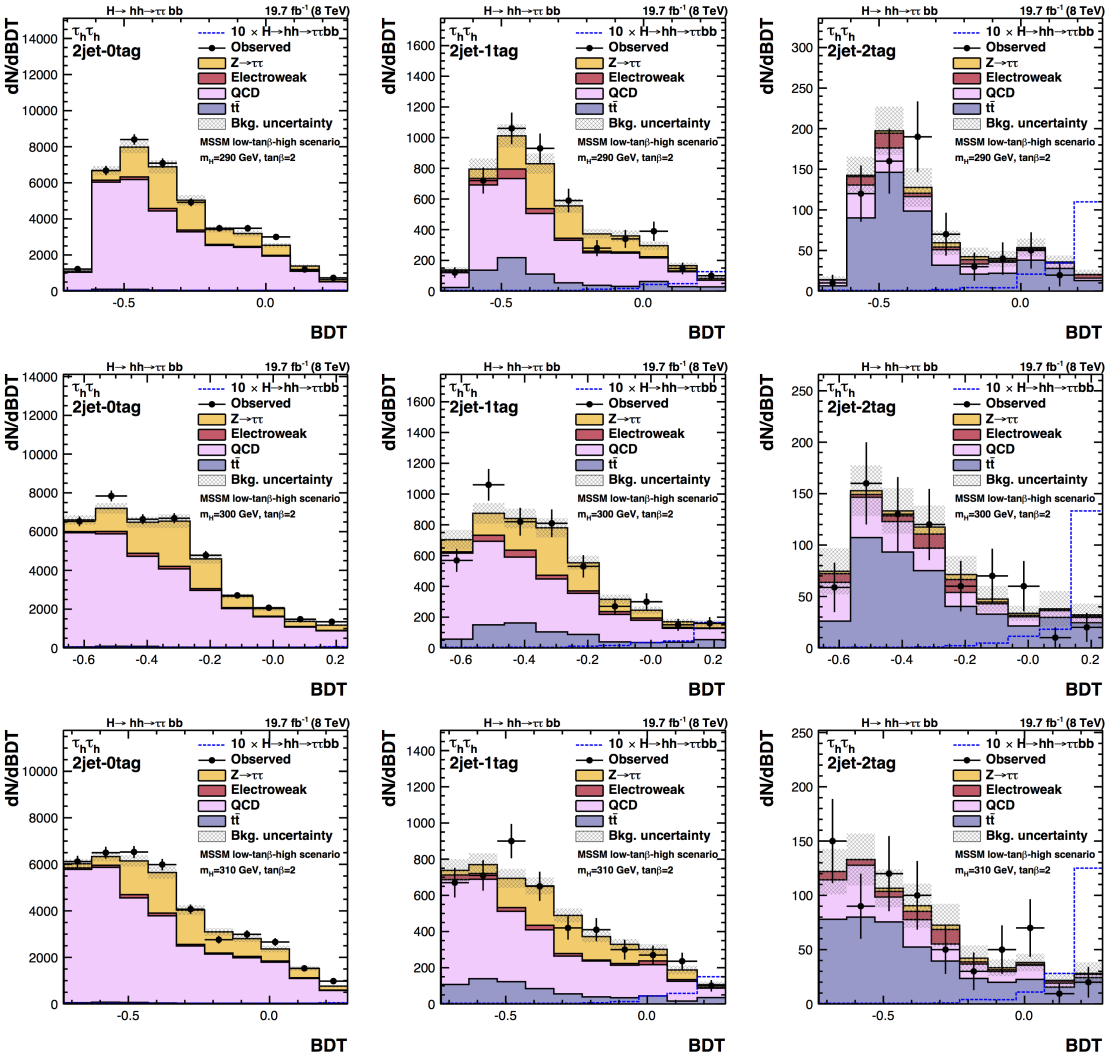

Figure 5.7.5: Distributions of the BDT discriminants from collider data and estimated backgrounds (histogram) with overflows included in the last bin. Left: 0-tag category; middle: 1-tag category; right: 2-tag category. The BDTs were trained at top row: $m_H = 290$ GeV, middle row: $m_H = 300$ GeV, bottom row: $m_H = 310$ GeV. The expected signal scaled by a factor 10 is shown superimposed as an open dashed histogram in the low tanβ scenario of the MSSM. Expected background contributions are shown for the values of nuisance parameters (systematic uncertainties) obtained after fitting the background only hypothesis to the data.

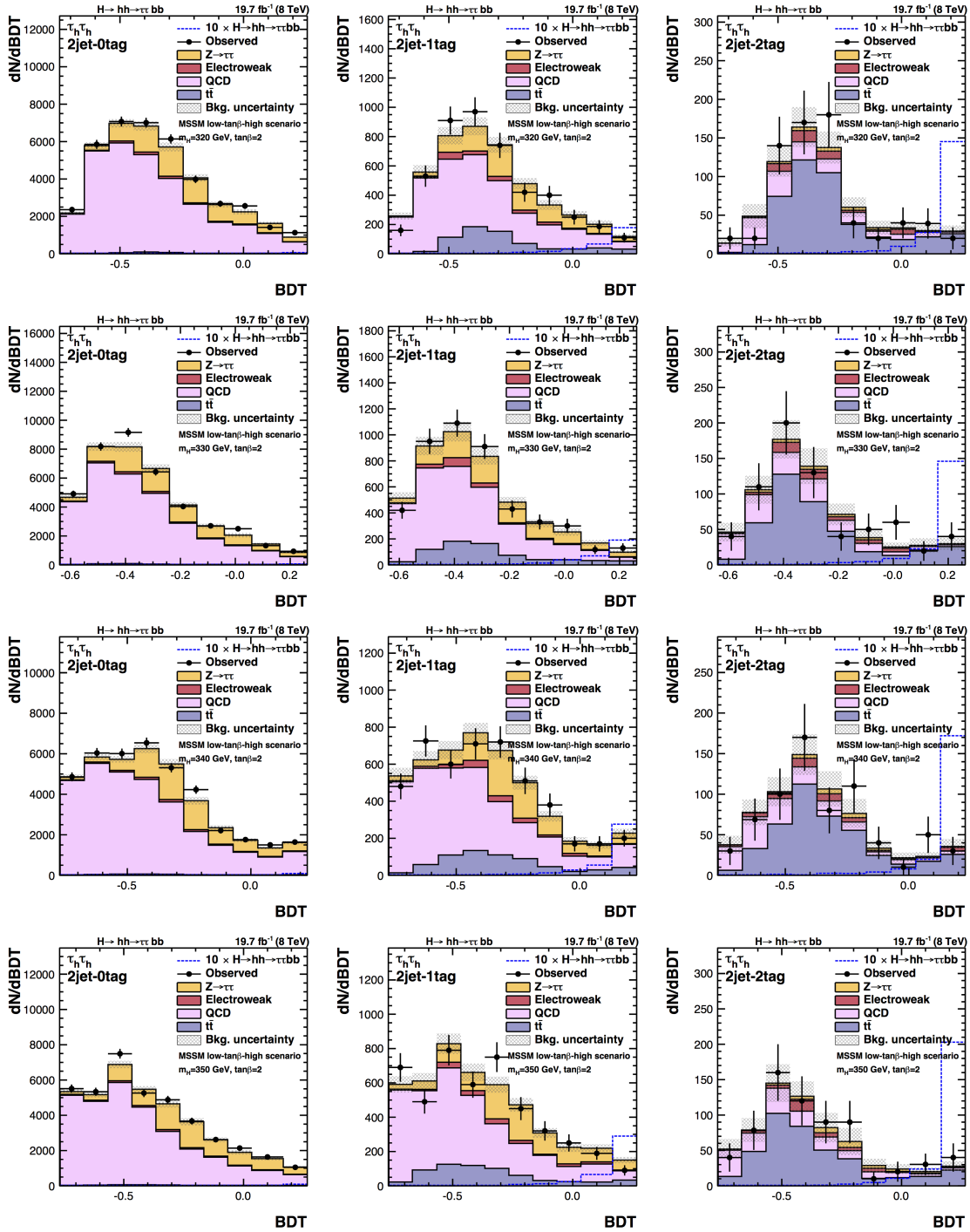

Figure 5.7.6: Distributions of the BDT discriminants from collider data and estimated backgrounds (histogram) with overflows included in the last bin. Left: 0-tag category; middle: 1-tag category; right: 2-tag category. The BDTs were trained at first row: $m_H = 320$ GeV, second row: $m_H = 330$ GeV, third row: $m_H = 340$ GeV, fourth row: $m_H = 350$ GeV. The expected signal scaled by a factor 10 is shown superimposed as an open dashed histogram in the low tanβ scenario of the MSSM. Expected background contributions are shown for the values of nuisance parameters (systematic uncertainties) obtained after fitting the background only hypothesis to the data.

Table 5.8: Systematic uncertainties considered for $\tau_h\tau_h$ channel

Experimental uncertainties	2 jet - 0 tag	2 jet - 1 tag	2 jet - 2 tag
Integrated luminosity 8 TeV	2.6%	2.6%	2.6%
Tau identification and trigger	19%	19%	19%
Normalization, QCD Multijet	10%	20%	40%
Normalization, Z production	3.3%	3.3%	3.3%
$Z \to \tau\tau$: category selection	5.7%	25.2%	175%
$Z \to \tau\tau$ due to $t\bar{t}$ embedded	–	5%	49%
Normalization, $Z \to ll$: jet and l misidentification	30.1%	51.5%	67.1%
Normalization, W+jets	20%	–	–
Normalization, $t\bar{t}$	10%	10%	10%
Normalization, di-boson	15%	15%	15%
b-tagging	shape	shape	shape
Tau energy scale	shape	shape	shape
Jet energy scale	shape	shape	shape

- Normalization uncertainties: these affect the overall yield of a process and are uniform among different bins in the final discriminator.
- Shape uncertainties: these affect the shape as well as the yield of a process.

Table 5.8 lists an overview of the systematics uncertainties used in the $\tau_h\tau_h$ channel. Detailed descriptions of the different sources of the uncertainties can be found in the following sections.

5.8.1 Normalization Uncertainties

- **Luminosity.** For 2012 data, the uncertainty of the luminosity amounts to 2.6%. This is used on all MC based backgrounds as well as $Z \to \tau\tau$ given that the overall yield is estimated with MC.
- **ID, isolation and trigger efficiencies for τ_h.** Combining the τ_h ID, isolation and trigger efficiencies, provided by the tau POG [86], and added in quadrature for both τ_h legs, the overall uncertainty for τ_h identification amounts to 19% for the $\tau_h\tau_h$ channel. This is applied to all MC based backgrounds like the luminosity.
- **QCD.** The QCD normalization systematic uncertainty is estimated propagating the statistical errors from all used sideband regions.

- **Z production.** The normalization of the $Z/\gamma^* \to \tau\tau$ embedded samples, as obtained using MC samples, is attributed an uncertainty of 3%.

- **$Z \to \tau\tau$ category selection.** The uncertainty is the result of extrapolation from the inclusive selection to the category selection and is driven by statistical error in the $Z/\gamma^* \to \tau\tau$ embedded samples.

- **$Z \to \tau\tau$ due to $t\bar{t}$ embedded.** This uncertainty is the result of statistical error in the $t\bar{t}$ embedded samples.

- **$Z \to ll,\ l = e, \mu$: jet and l misidentification.** The $Z \to ll$ background is very small after the requirement of 2 jets, and is estimated from statistical uncertainty of the MC sample.

- **Other production cross-section uncertainties.** The W+jets background is very small in the $\tau_h\tau_h$ channel and data-driven in 2 jet - 1 tag and 2 jet - 2 tag categories. A 20% uncertainty is assigned to W+jets MC backgrounds in the 2 jet - 0 tag category considering W+jets MC production cross-section uncertainty. The uncertainty of the $t\bar{t}$ cross-section amounts to 10%, and uncertainty of the single top and di-boson production cross-sections amounts to 15%.

5.8.2 Shape Uncertainties

- **b-tagging.** Uncertainties of b-tagging efficiencies and mistag rates as function of jet p_T and η are provided by the official CMS values [87]. For MC based samples, the b-tag scale factors are varied within their recommended uncertainties to obtain the shape variations.

- **Tau energy scale.** In order to account for the sensitivity of the final discriminator to the energy of reconstructed τ_h, the energy of MC τ_h's is varied by $\pm 3\%$, as recommended by the official CMS values [86].

- **Jet energy scale.** Similar to the energy of τ_h's, the final discriminator is also sensitive to the energy of jets. Thus, MC reconstructed jet's energies are shifted up and down based on the uncertainties as function of jet p_T and η provided by the official CMS values [88].

5.9 Results and Interpretation

As shown in Fig. 5.7.1 and Figs. 5.7.4–5.7.6, no significant excess over the expected background was found in neither the kinematic fit or the BDT spectra. A model independent upper limit at

Table 5.9: Background bin by bin uncertainties.

Bin	1	2	3	4	5	6	7	8	9	10
kinematic fit	16.7%	24.7%	24.5%	21.1%	26.7%	20.7%	35.3%	63.4%		
BDT (260)	11.5%	10.6%	11.0%	13.8%	12.9%	15.3%	21.3%	25.0%		
BDT (270)	21.0%	14.3%	11.1%	10.6%	12.1%	14.2%	18.7%	18.1%	25.0%	24.7%
BDT (280)	19.3%	12.6%	11.0%	12.8%	13.3%	14.9%	15.9%	13.1%	20.5%	24.5%
BDT (290)	21.0%	9.5%	9.7%	12.7%	12.5%	14.9%	20.4%	18.1%	19.2%	24.4%
BDT (300)	15.4%	9.7%	10.4%	12.9%	18.9%	21.3%	21.5%	20.9%	24.7%	
BDT (310)	11.3%	9.9%	10.6%	13.6%	16.0%	17.3%	21.6%	17.0%	24.1%	25.0%
BDT (320)	16.7%	13.3%	11.3%	9.7%	11.5%	13.6%	18.4%	21.4%	21.7%	24.7%
BDT (330)	11.6%	11.0%	9.3%	10.9%	15.8%	20.7%	20.1%	23.3%	24.8%	
BDT (340)	15.0%	10.8%	11.2%	10.1%	12.0%	14.8%	22.6%	23.7%	21.9%	24.3%
BDT (350)	11.8%	11.7%	9.6%	11.1%	12.4%	16.3%	21.8%	20.5%	24.9%	24.7%

95% confidence level (CL) on the cross-section times branching ratio of $H \to hh \to \tau\tau bb$ is set using a modified frequentist approach, the CL_s method [89], with implementation provided by the Higgs Combined tool [90].

The binning of the BDT discriminant, shown in Figs. 5.7.4–5.7.6, is configured such that the background uncertainty in each bin falls below a maximum threshold of 25%, which is the background uncertainty of the most sensitive bins for the kinematic fit distribution. Starting from the right-most edge of the BDT distribution, events are collected until the background uncertainty in that bin falls below the maximum threshold, while the bin width is at least 0.1 (to prevent formation of narrow bins in the background dominated region). Then, we start from the left edge of the newly formed bin and continue leftwards until all events are collected. Table 5.9 shows a comparison of the background bin-by-bin uncertainty between the kinematic fit and BDT distributions.

The expected and observed 95% CL upper limits on the cross section times branching ratio on $H \to hh \to \tau\tau bb$ is shown in Table 5.10 and 5.11 as well as in Fig 5.9.1. All results are comparable within 1 σ. Compared with the kinematic fit method, the BDT method provides a stronger or comparable limit for $m_H > 280$ GeV. For low masses of m_H, the BDT analysis suffers from low number of signal MC events.

For publication, the kinematic fit discriminator is chosen as the final discriminator to synchronize with the $e\tau_h$ and $\mu\tau_h$ channels. The combined limit of the $\tau_h\tau_h$, $e\tau_h$ and $\mu\tau_h$ channels is shown in Fig 5.9.2. As shown, the $l + \tau_h$ channels drives the sensitivity for $m_H < 310$ GeV while the $\tau_h\tau_h$ channel is the most sensitive at $m_H > 310$ GeV.

Table 5.10: Expected limits (in pb) at 95% CL, using the 0-tag, 1-tag and 2-tag categories.

m_H [GeV]	260	270	280	290	300	310	320	330	340	350
kinematic fit	0.705	0.793	0.824	0.545	0.498	0.439	0.317	0.294	0.235	0.19
BDT	0.785	0.824	0.654	0.42	0.378	0.339	0.271	0.243	0.224	0.164

Table 5.11: Observed limits (in pb), using the 0-tag, 1-tag and 2-tag categories.

m_H [GeV]	260	270	280	290	300	310	320	330	340	350
kinematic fit	0.714	0.801	0.883	0.485	0.444	0.438	0.295	0.244	0.174	0.123
BDT	0.989	1.02	0.594	0.296	0.318	0.312	0.239	0.326	0.213	0.149

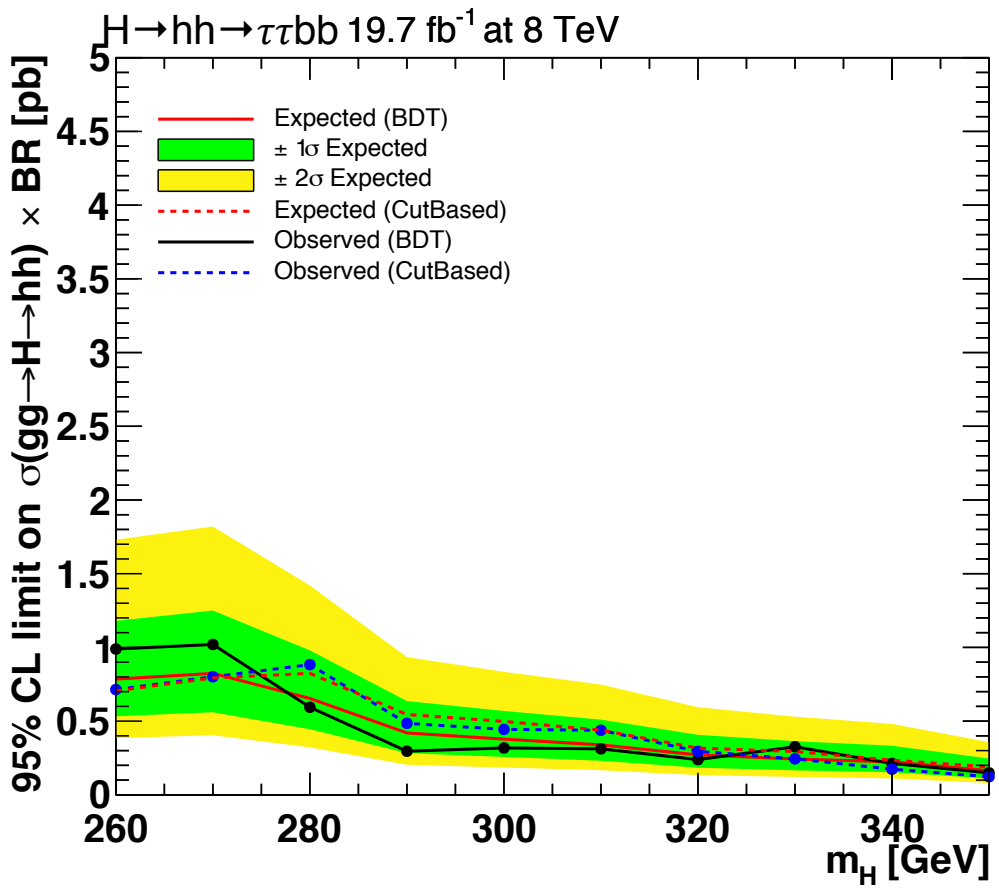

Figure 5.9.1: Expected and observed upper limits comparison between kinematic fit approach (referred as CutBased) and BDT analyses, using the 0-tag, 1-tag and 2-tag categories.

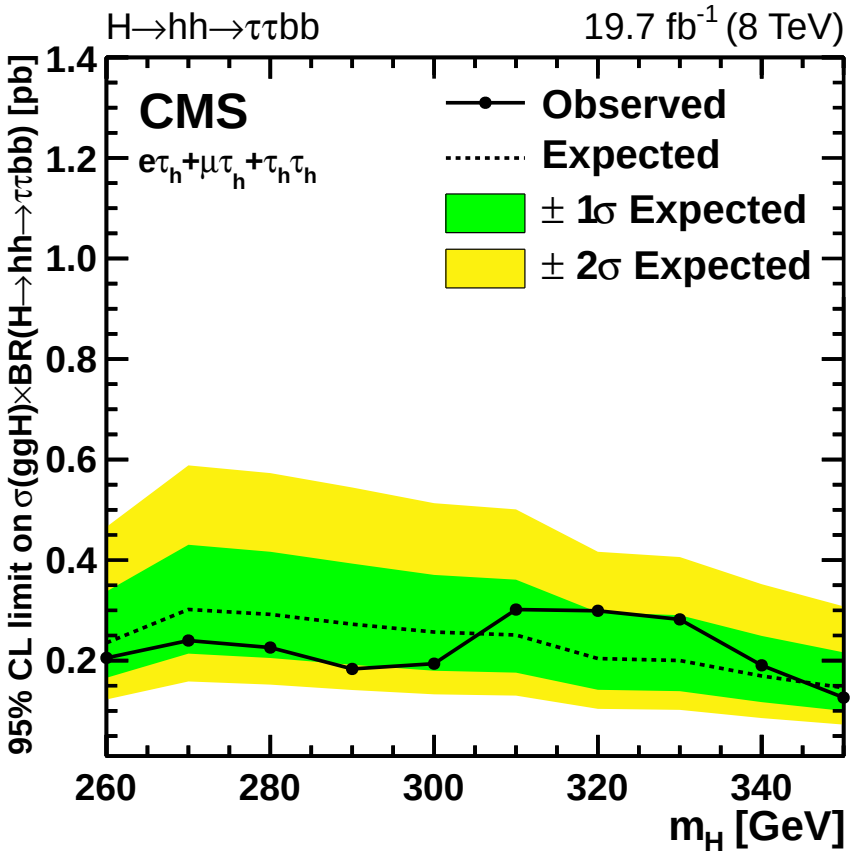

Figure 5.9.2: 95% CL upper limits on cross section times branching fraction on $H \to hh \to \tau\tau bb$ in the combination of $\tau_h\tau_h$, $e\tau_h$ and $\mu\tau_h$ channels.

5.10 Summary

The results of a search for a heavy Higgs boson (h) decaying to two 125 GeV Higgs bosons (hh) in the final state of $\tau\tau bb$, detailed in $\tau_h\tau_h$ and combined with $e\tau_h$ and $\mu\tau_h$, have been presented. No evidence for a signal is found in the 8 TeV pp collision data recorded by in 2012, corresponding to an integrated luminosity of 19.7 fb^{-1}. Limits on production cross section times branching fraction for the $H \to hh \to \tau\tau bb$ process have been set.

Chapter 6

Search for new physics with high-mass tau pairs with 2015 data

In this chapter, a search for an extra neutral gauge boson that decays to a pair of τ leptons is presented. This analysis uses data collected during 2015 by the CMS experiment with a centre of mass energy $\sqrt{s} = 13$ TeV and an integrated luminosity of 2.2 fb^{-1} and has been published [14].

This chapter is organized as follows: Section 6.1 provides a brief introduction to the motivations of the search. Section 6.2 provides an overview of the analysis along with the key background estimation methods. The sample sets used in this analysis are given in details in Section 6.3. Section 6.4–6.5 describes in details the event selection and backgrounds estimation methods used in each of the considered decay channels. Section 6.6 discusses the considered systematic uncertainties. The statistical procedure used to set limits and their results are detailed in Section 6.7. Finally, a summary is given in Section 6.8 which concludes this chapter.

Out of the four decay channels studied: $\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$, and $e\mu$, only the channels of $e\tau_h$, and $e\mu$ will be presented in detail as they are the channels I personally worked on.

6.1 Introduction

As discussed in Section 1.4, many models, ranging from the Little Higgs model to E_6 GUTs, predict the existence of extra gauge bosons. Of particular interest for this analysis are models that include an extra massive neutral gauge boson, referred to as Z', that decays to a pair of τ leptons. Due to its simplicity, the Z' boson of the Sequential Standard Model (SSM) is commonly chosen as the

benchmark for Z' searches. Here, SSM predicts a Z' boson with the same coupling to quarks and leptons as the SM Z boson.

The most stringent mass limits on SSM Z' production are set, using the e^+e^- and $\mu^+\mu^-$ decay channels by CMS and ATLAS collaborations, to be 3.2 [91] and 3.4 [92] TeV respectively.

This study uses data samples of pp collisions at $\sqrt{s} = 13$ TeV recorded in 2015 by the CMS detector and corresponds to an integrated luminosity of 2.2 fb^{-1}.

6.2 Strategy

As discussed in Section 1.1.1, τ leptons have a rather short life time and decay within the detector. Around 1/3 of all τ leptons decay to an electron or muon while the rest decay into hadronic jets, referred to as τ_h. For hadronically decaying τ leptons, due to charge conservation, their decay products mostly consist of one or three charged mesons and are respectively referred to as one-prong and three-prong decay modes. Out of the six possible decay channels of a pair of τ leptons, four channels are chosen ($\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$ and $e\mu$) as ee and $\mu\mu$ channels are hard to separate from prompt $Z' \to ee/\mu\mu$ decays. SSM Z' MC samples between the mass of 500 GeV to 3000 GeV, in steps of 500 GeV, are produced and used in this study.

The modeling of SM backgrounds consists of a mixture of MC based and data-driven methods. In general, backgrounds with genuine τ_h are modeled with MC simulations. Dedicated signal depleted control regions are used to estimate the MC modeling of such backgrounds. For backgrounds with jets misidentified as τ_h or l, mostly QCD, a data-driven background estimation technique is used.

The invariant mass of the two reconstructed objects + $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ system, referred to as the effective mass, provides a good separation between signal and background. To quantify the significance of any possible excess, a fit is performed with the effective mass distribution using a modified frequentist approach, the CL_s method [89], for each channel. Finally, a joint limit combining all studied channels is obtained by combining the posterior probability density functions and taking into account correlation of systematic uncertainties within and across all channels.

6.3 Data and Monte Carlo Samples

The 13 TeV collision data collected by the CMS detector in year 2015 is used in this analysis. Table 6.1 shows the collision datasets used. The total integrated luminosity of the collision data samples is 2.2 fb^{-1}.

The official Spring 2015 miniAODv2 MC samples are used to model most of the SM processes. The leading order generators, PYTHIA8 and MADGRAPH, are mainly used for signal and background MC production. The predicted background yields in simulation are determined using NLO or NNLO cross-sections, while the signal yields and distributions in all plots shown are normalized using the LO cross-sections shown in Table 6.2. Table 6.3 shows the entire list of the MC samples used for this analysis.

Table 6.1: Collision Data Samples

Run Period	Official CMS Datasets
Run 2015C SingleElectron 05 Oct ReMiniAOD	*/SingleElectron/Run2015C_ 25ns-05Oct2015-v1/MINIAOD*
Run 2015D SingleElectron 05 Oct ReMiniAOD	*/SingleElectron/Run2015D-05Oct2015-v1/MINIAOD*
Run 2012D SingleElectron PromptReco v4	*/SingleElectron/Run2015D-PromptReco-v4/MINIAOD*

Table 6.2: MC Signal Samples

Process (mass [GeV])	width[GeV]	cross-section [pb]	Official CMS Datasets (MINIAODSIM)
SSM Z'(500)	14.73	5.751	*/ZprimeToTauTau_ M-500_ TuneCUETP8M1_ 13TeV-pythia8-tauola* [1]
SSM Z'(1000)	30.97	0.3865	*/ZprimeToTauTau_ M-1000_ TuneCUETP8M1_ 13TeV-pythia8-tauola* [1]
SSM Z'(1500)	47.05	0.06479	*/ZprimeToTauTau_ M-1500_ TuneCUETP8M1_ 13TeV-pythia8-tauola* [1]
SSM Z'(2000)	63.10	0.01583	*/ZprimeToTauTau_ M-2000_ TuneCUETP8M1_ 13TeV-pythia8-tauola* [1]
SSM Z'(2500)	79.16	0.004684	*/ZprimeToTauTau_ M-2500_ TuneCUETP8M1_ 13TeV-pythia8-tauola* [1]
SSM Z'(3000)	95.24	0.001592	*/ZprimeToTauTau_ M-3000_ TuneCUETP8M1_ 13TeV-pythia8-tauola* [1]

[1] `/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1`

6.3.1 Corrections for Pile-Up

In order to match the pile-up distribution used to generate MC samples to that of the observed data, an event based reweighting is deployed for the MC samples with the weight:

$$w_{PU}(n) = \frac{P_{data}(n)}{P_{MC}(n)} \quad , \tag{6.3.1}$$

where $P_{data}(n)$ and $P_{MC}(n)$ are the number of PVs in data and MC.

The recommended min-bias cross-section of 69 mb is used to determine these weights.

Table 6.3: MC Background Samples

Process	cross-section [pb] best theoretical calculation or $\sigma_{\mathrm{LO}}\times$ k-factor, where k-factor $= \frac{\sigma_{\mathrm{NLO}}}{\sigma_{\mathrm{LO}}}$	Official CMS Datasets (MINIAODSIM)
$Z \to ll$	5765.4	/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 [1]
mass binned	7.67×1.031	/DYJetsToLL_M-200to400_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8 [1]
LO samples	0.423×1.008	/DYJetsToLL_M-400to500_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8 [1]
	0.24×0.996	/DYJetsToLL_M-500to700_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8 [2]
	0.035×0.973	/DYJetsToLL_M-700to800_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8 [1]
	0.03×0.961	/DYJetsToLL_M-800to1000_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8 [1]
	0.016×0.938	/DYJetsToLL_M-1000to1500_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8 [1]
$W + jets$	61526.7	/WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 [1]
HT binned	1345×1.21	/WJetsToLNu_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 [1]
LO samples	359.7×1.21	/WJetsToLNu_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 [1]
	48.91×1.21	/WJetsToLNu_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 [1]
	12.05×1.21	/WJetsToLNu_HT-600ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 [1]
$t\bar{t}$	831.76	/TT_TuneCUETP8M2T4_13TeV-powheg-pythia8 [1]
single Top samples	35.6	/ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1 [1]
	35.6	/ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1 [3]
	136.02	/ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1 [1]
	26.23	/ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1 [1]
$VV \to 2l2\nu$	11.95	/VVTo2L2Nu_13TeV_amcatnloFXFX_madspin_pythia8 [1]
$ZZ \to 2l2q$	3.22	/ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8 [1]
$ZZ \to 4l$	1.256	/ZZTo4L_13TeV-amcatnloFXFX-pythia8 [1]
$WW \to l\nu 2q$	1.212	/WWTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8 [1]
$WZ \to 2l2q$	5.595	/WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8 [1]
$WZ \to 3l\nu$	4.708	/WZJets_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8 [1]
$WZ \to l3\nu$	3.05	/WZTo1L3Nu_13TeV_amcatnloFXFX_madspin_pythia8 [1]
$WZ \to l\nu 2q$	10.71	/WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8 [1]

[1] `/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v1`
[2] `/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v3`
[3] `/RunIISpring15MiniAODv2-74X_mcRun2_asymptotic_v2-v2`

6.4 Electron + Hadronic Tau Channel

6.4.1 Event selection

Events in the $e\tau_h$ channel are required to fire the single-electron trigger (which requires an electron with $p_T > 27$ GeV, $|\eta| < 2.1$, and pass the "Loose" working point of the electron ID). Then reconstructed electrons are selected satisfying:

- $p_T > 35$ GeV, $|\eta| < 2.1$, $d_{xy} < 0.045$ cm and $d_z < 0.2$ cm;
- the "Tight" electron MVA ID requirements;
- having no matched conversion nor missing hits;
- within $\Delta R < 0.5$ of the HLT electron that fired the trigger.

Offline τ_h are required to:

- have $p_T > 20$ GeV, $|\eta| < 2.1$, and $d_z < 0.2$ cm;
- pass the HPS tau reconstruction requirement while containing one or three charged hadrons;
- pass the "Tight" working point for anti-electron MVA ID and pass the "Loose" working point for anti-muon ID.

Electron and τ_h pairs are constructed by choosing the pair which the electron and τ_h are separated by at least $\Delta R > 0.5$. In events with more than one such pair, we select the pair with the two most isolated leptons, considering first the electron, and then the τ_h. After a pair has been chosen for an event, we apply the following isolation requirements on the leptons, for an event to enter the signal region: electron relative isolation < 0.15; τ_h passing the "Tight" working point of the cut-based tau isolation ID. In order to keep the different channels exclusive, an event is rejected if there is an additional electron satisfying the above identification requirements with relative isolation < 0.3, or a muon satisfying the identification requirements described in Section 6.5.1 with relative isolation < 0.3. To further reduce backgrounds, we require the electron and τ_h to have opposite electric charge. The selection set mentioned above is defined as the preselection.

Following the preselection, the signal region is defined as:

- $\cos\Delta\phi(e, \tau_h) < -0.95$;
- $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;

- $P_\zeta - 3.1 \times P_\zeta^{vis} < -50$ GeV;
- no jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose).

Here, $\hat{\zeta}$ is defined as the unit vector along the bisector of visible tau decay products and two the projection variables are defined as:

$$P_\zeta^{vis} = (\vec{p}_e^{\,vis} + \vec{p}_\mu^{\,vis}) \cdot \hat{\zeta} \quad . \tag{6.4.1}$$

$$P_\zeta = (\vec{p}_e^{\,vis} + \vec{p}_\mu^{\,vis} + \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}) \cdot \hat{\zeta} \quad . \tag{6.4.2}$$

The distributions of these variables after preselection, and after selection requirements on the other variables, are shown in Figure 6.4.1.

6.4.2 Genuine dilepton events

Studies of simulated events indicate that for Drell-Yan process, top quark single and pair production, and di-boson production, the reconstructed and selected electrons and hadronic taus are typically associated with genuine simulated leptons. The nominal expected event rates are estimated by scaling the simulated samples by the best available cross sections, listed in Table 6.3, and by the integrated luminosity of the data samples.

6.4.2.1 Drell-Yan process

Due to large W+jets and QCD contamination in the $e\tau_h$ channel, as shown in the left panel of Fig 6.4.2 with the following selections:

- pass preselection;
- $p_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV;
- no jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose);
- $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}\right) < 200$ GeV,

we use the Drell-Yan rate systematic uncertainty (12%) estimated from the $e\mu$ final state in 6.5.2.1.

6.4.2.2 $t\bar{t}$ and single top processes

The $t\bar{t}$ and single top production rate systematics (8%), are estimated from the $e\mu$ channel as described in Section 6.5.2.2, for its higher top purity. A check of the $t\bar{t}$ modeling in the $e\tau_h$ channel

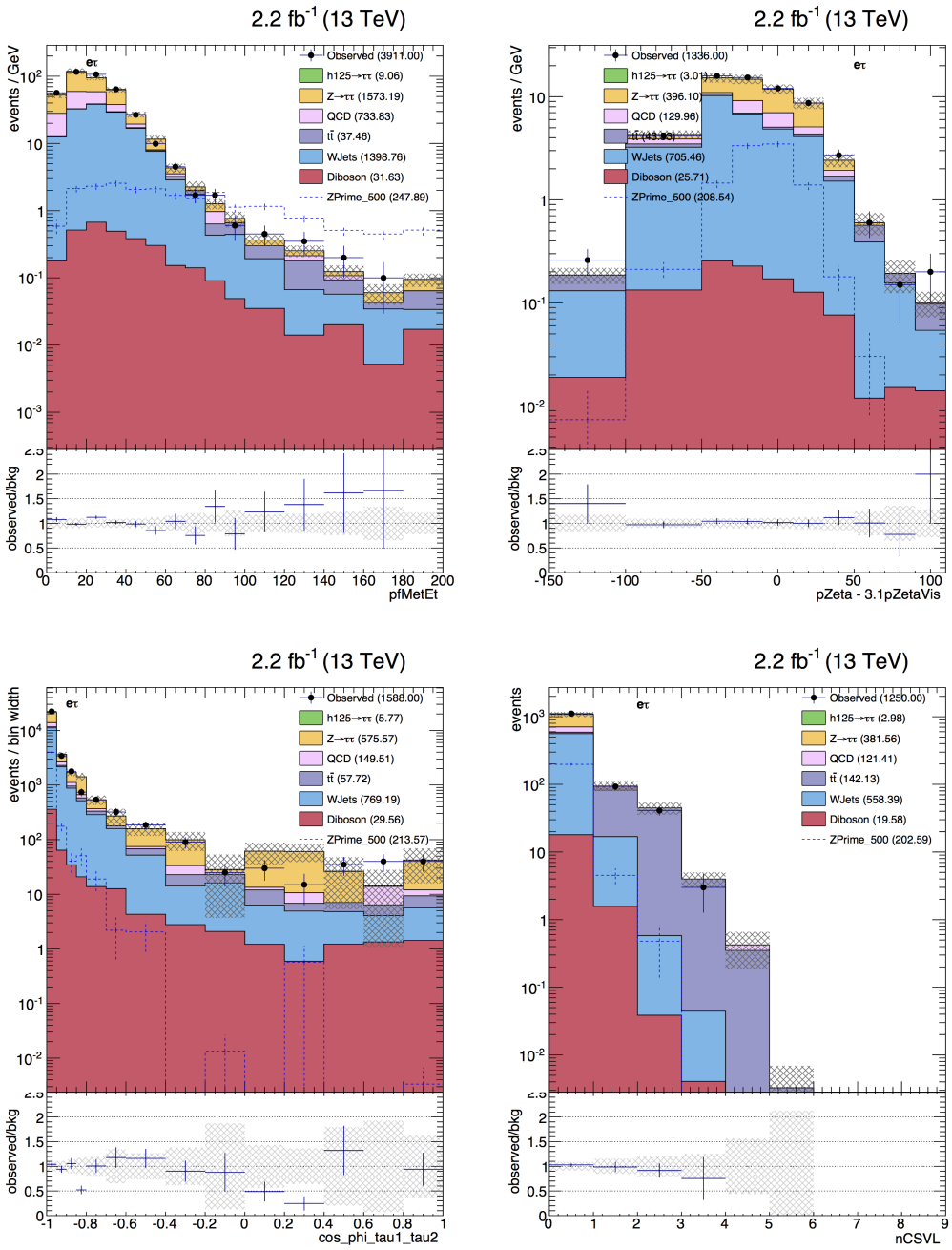

Figure 6.4.1: Distributions of the variables used for eτ_h signal selection, after all other signal selection requirements on variables other than the one plotted: $p_{\mathrm{T}}^{\mathrm{miss}}$ (top left), "ζ" (top right), $\cos\Delta\phi(e,\tau_h)$ (bottom left), and n_b (bottom right).

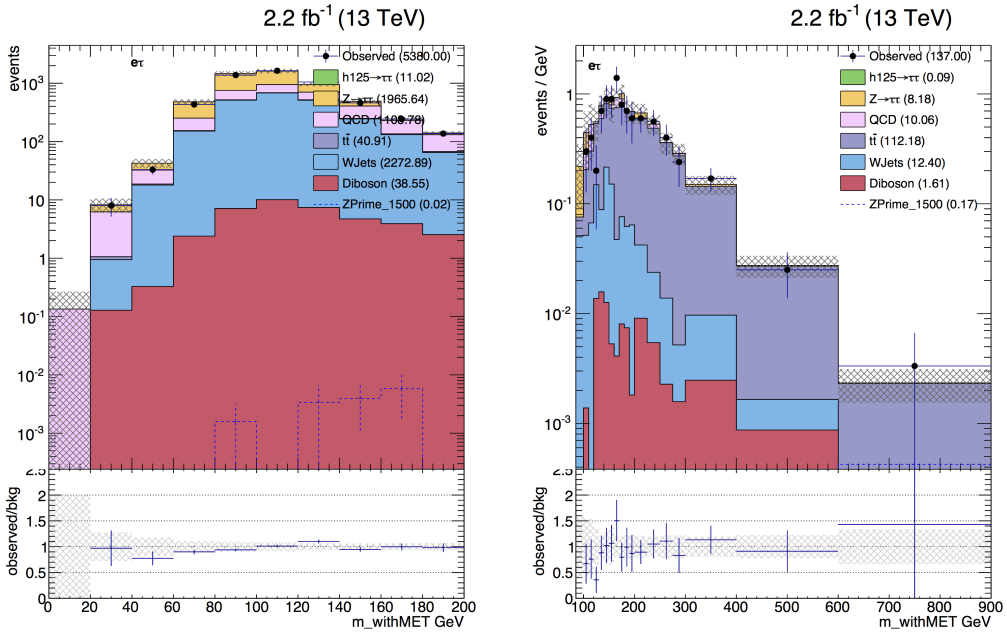

Figure 6.4.2: (Data driven QCD with only statistical uncertainties) Distributions of $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$. Left: validation region with $p_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV, $n_b = 0$ and $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right) < 200$ GeV. Right: validation region with $n_b \geq 1$.

is performed in a top-rich region, defined by the following selections and shown in the right panel of Fig. 6.4.2:

- pass preselection;
- $\cos\Delta\phi(e, \tau_h) < -0.95$;
- $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;
- $P_\zeta - 3.1 \times P_\zeta^{vis} < -50$ GeV;
- at least one jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose).

Overall, the expected $t\bar{t}$ distribution from MC simulations agrees well with the observation.

6.4.2.3 Di-boson process

Di-boson processes are a relatively small background in the $e\tau_h$ channel. They are estimated directly from simulation with a 15% systematic uncertainty.

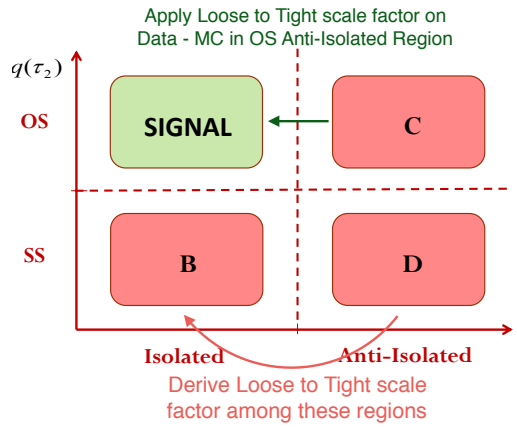

Figure 6.4.3: QCD estimation and validation strategy for the $e\tau_h$ channel.

6.4.3 QCD background

For a given variable and binning, e.g. the effective mass variable used for signal extraction, we construct a data-driven template for the shape of the QCD multi-jet background, i.e. the processes lacking prompt leptons. Based on the charge of the final state and τ_h isolation, we split the events into four regions shown in Figure 6.4.3 and described below:

- A (Signal) Region: electron and τ_h have opposite charge and τ_h pass "Tight" isolation requirement.
- B Region: electron and τ_h have same charge and τ_h pass "Tight" isolation requirement.
- C Region: electron and τ_h have opposite charge and τ_h pass anti-isolation requirement.
- D Region: electron and τ_h have same charge and τ_h pass anti-isolation requirement.

In each region (B, C, D), QCD events are estimated by subtracting events with genuine leptons (estimated by simulation) bin-by-bin from data. The two jets faking τ_h in QCD events are assumed to have the same probability of having the same or opposite electric charge, thus, the amount of QCD events in region B should be comparable to that of in the signal region. However, with the freedom to define the anti-isolation region, we choose an anti-isolation definition such that there are much more QCD in region C compared to the signal region. Taking the QCD shape from region C will provide us a much smoother template for QCD estimation.

Hence, QCD events in the signal region are estimated with the shape from region C and multi-

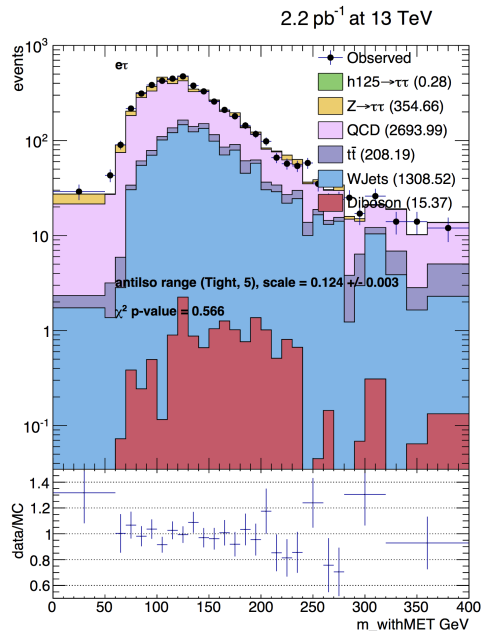

Figure 6.4.4: The distribution of reconstructed parent mass, $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$, in the same-charge, tightly-isolated sample: $e\tau_h$ channel.

plying a scale factor derived from regions B and D. The factor is defined as:

$$f_{\mathrm{LT}}^{\mathrm{QCD}} = \left(N_{\mathrm{data}}^{\mathrm{B}} - N_{\mathrm{MC}}^{\mathrm{B}}\right) / \left(N_{\mathrm{data}}^{\mathrm{D}} - N_{\mathrm{MC}}^{\mathrm{D}}\right) \quad . \tag{6.4.3}$$

This QCD estimation method is valid only if the QCD shape in anti-isolated correctly models the QCD shape in the isolated region. The check is done by comparing the observation and background estimation in region B with the QCD shape taking from region D and normalized to the QCD in region B. This test is shown in the right panel of Figure 6.4.4 with the anti-isolation definition as: τ_h isolation failing the "Tight" working point but below 5.0 GeV.

After the signal region selection the "Loose-to-Tight" scale factor is estimated to be: 0.13 ± 0.02 where an additional 15% uncertainty is added to the QCD systematics on top of the bin-by-bin systematics.

6.4.4 W+jets background

The simulated W+jets samples were not generated with large MC statistics. Using the samples directly would have two disadvantages: (a) avoiding non-smooth templates in the signal extraction would restrict somewhat the choice of signal selection and binning; (b) the poorly sampled shapes

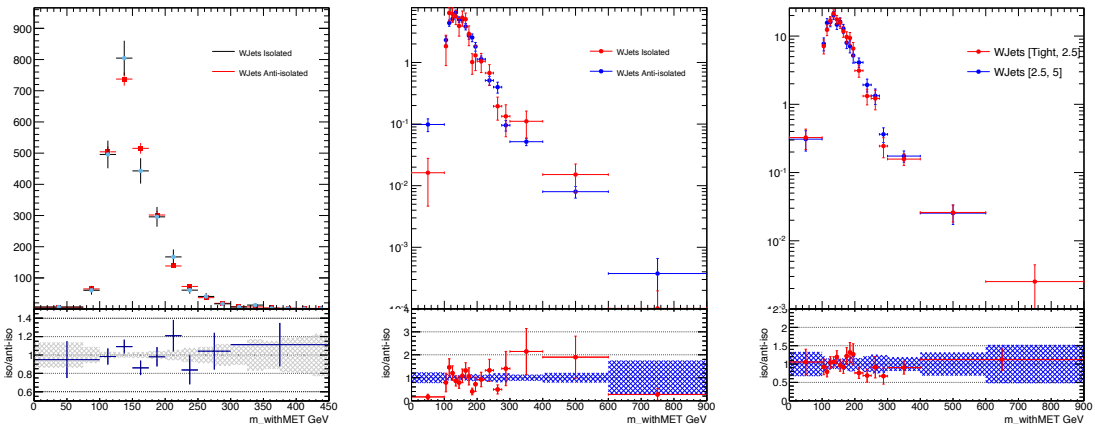

Figure 6.4.5: Left: comparison of the simulated distributions of $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$, using tight τ_h isolation and with relaxed τ_h isolation, in the control region described in Section 6.4.4. Center: analogous comparison for the signal region. Right: comparison in the signal region of the "tighter" and "looser" halves of the relaxed isolation region.

degrade the ability to validate the background estimates. As a workaround, we use a region of relaxed τ_h isolation to obtain the simulated shape: τ_h isolation is required to fail the "tight" working point but be less than 5 GeV. This shape template is then normalized to the integrated MC yield when requiring tight τ_h isolation.

Figure 6.4.5 compares the direct and relaxed predictions in the signal region and in the control region discussed below. It also shows the compatibility of the predicted shapes when the W+jets MC events are split into to regions: τ_h isolation failing "Tight" working point but < 2.5 GeV; 2.5 GeV$<$ τ_h isolation < 5 GeV.

6.5 Electron + Muon Channel

6.5.1 Event selection

In the $e\mu$ channel, the requirement of trigger and the selection of electrons are identical to that described in Section 6.4.1. Muons are required to have:

- $p_T > 10$ GeV, $|\eta| < 2.1$, $d_{xy} < 0.045$ cm and $d_z < 0.2$ cm;
- satisfy the medium muon requirement.

We build pairs of electrons and muons in which the electron and muon are separated by at least $\Delta R > 0.5$. In events with more than one such pair, we select the pair with the two most isolated

leptons, considering first the muon, and then the electron. In the rare case of multiple such pairs having identical isolation values, the reconstructed p_T's are considered, preferring higher values.

After a pair has been chosen for an event, it enters the signal region only when both the electron and muon relative isolations are < 0.15. To reduce Drell-Yan backgrounds, events are rejected if there is an additional electron satisfying the requirements described in Section 6.4.1 with relative isolation < 0.3, or an additional muon satisfying the above identification requirements with relative isolation < 0.3. To further reduce backgrounds, we require the electron and muon to have opposite electric charge. The selection set mentioned above is defined as preselection.

Following the preselection, based on the optimization studies performed in the $\tau_h\tau_h$ channel, signal region is defined with the following requirements:

- $\cos\Delta\phi(e,\mu) < -0.95$;
- $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;
- $P_\zeta - 3.1 \times P_\zeta^{vis} > -50$ GeV;
- no jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose).

The distributions of these variables after preselection, and after selection requirements on the other variables, are shown in Fig 6.5.1.

6.5.2 Genuine di-lepton events

Similar as to the $e\tau_h$ channel, backgrounds of Drell-Yan process, top quark single and pair production, W+jets, and di-boson production are estimated by MC simulations and normalized to the best available cross sections, listed in Table 6.3.

6.5.2.1 Drell-Yan process

Systematics for Drell-Yan process are estimated in a Drell-Yan rich region with the following selections and the distributions of data and estimated backgrounds are shown in the left panel of Fig 6.5.2:

- pass preselection;
- $p_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV;
- no jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose);

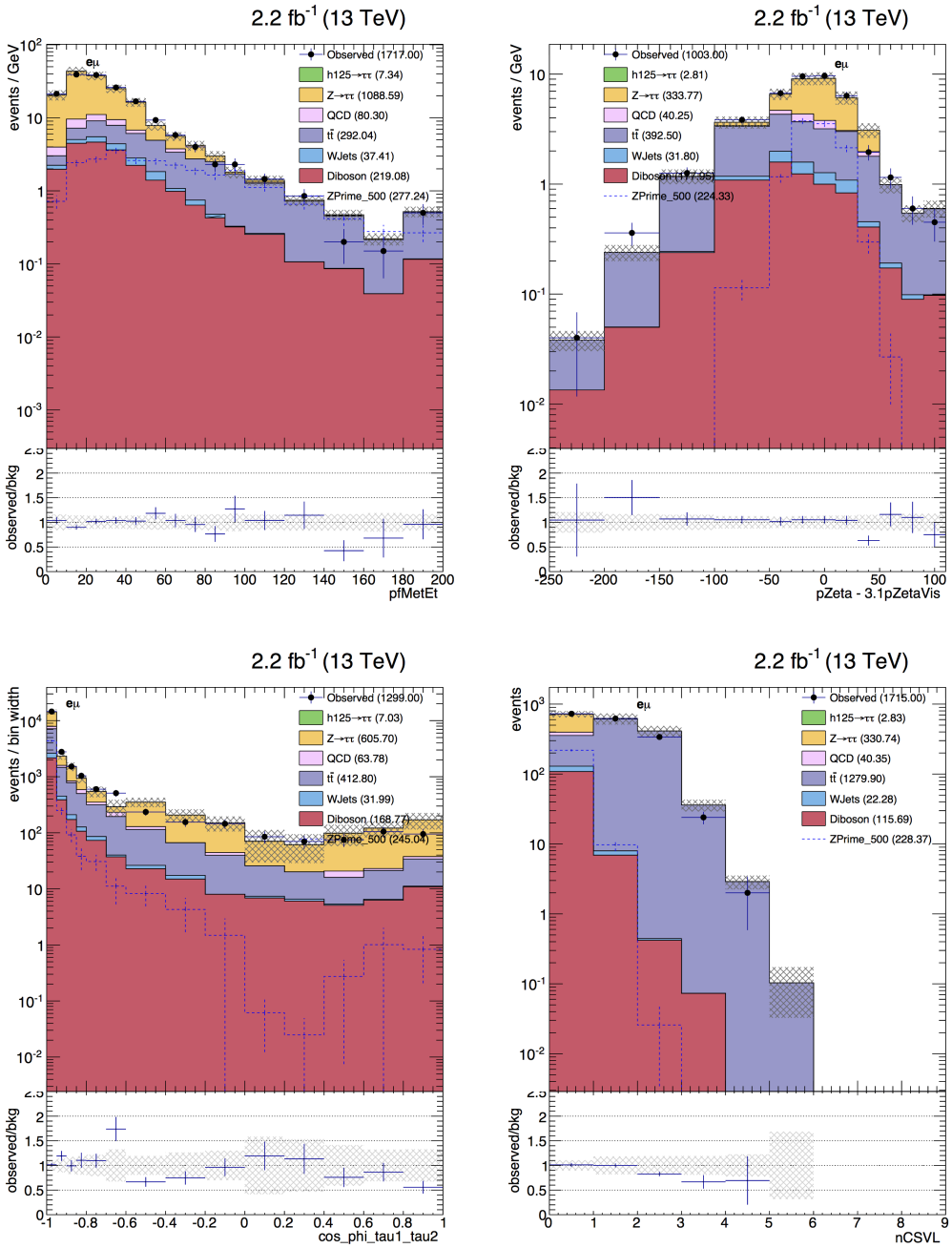

Figure 6.5.1: (Data driven QCD with only statistical uncertainties) Distributions of the variables used for $e\mu$ signal selection, after all other signal selection requirements on variables other than the one plotted: $p_{\mathrm{T}}^{\mathrm{miss}}$ (top left), "ζ" (top right), $\cos\Delta\phi(e,\mu)$ (bottom left), and n_b (bottom right).

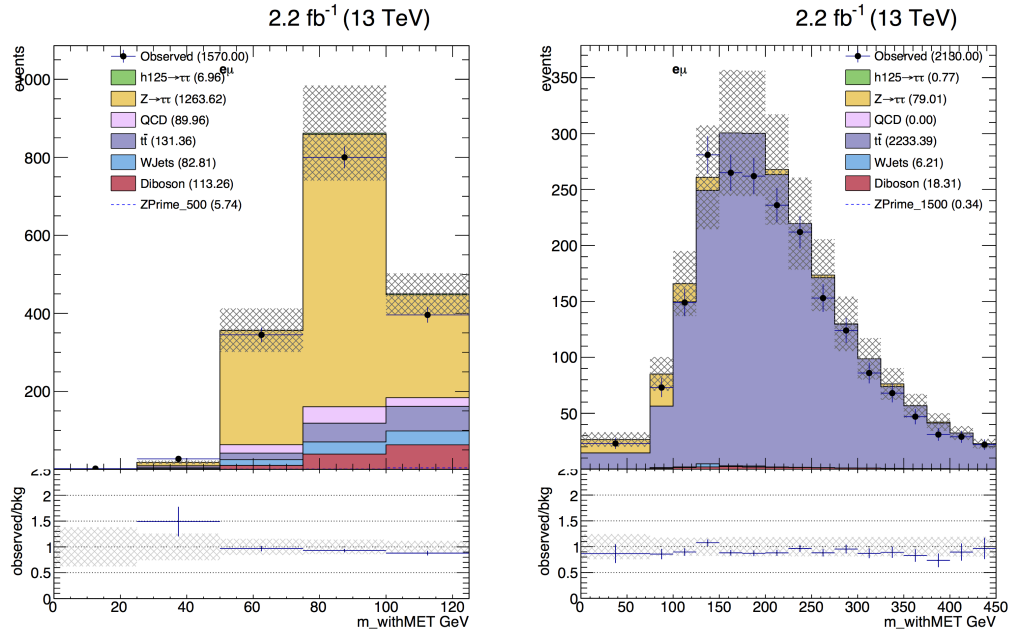

Figure 6.5.2: (Data driven QCD with only statistical uncertainties) Distributions of $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$. Left: validation region with $p_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV, $n_b = 0$ and $60 <$ GeV $m\left(e, \mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right) < 150$ GeV. Right: validation region with $n_b \geq 1$.

- $m\left(e, \mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right) < 125$ GeV .

The Drell-Yan production rate systematic uncertainty is estimated to be:

$$\text{Drell-Yan systematics} = \left|1 - \frac{\text{Drell-Yan}}{\text{Data - other backgrounds}}\right| = 12\% \tag{6.5.1}$$

which we apply both to $e\tau_h$ and $e\mu$ final states.

6.5.2.2 $t\bar{t}$ and single top processes

Systematics for $t\bar{t}$ and single top processes are estimated in a top quark rich region with the following selections and the distributions of data and estimated backgrounds are shown in the right panel of Figure 6.5.2:

- pass preselection;
- $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;
- $P_\zeta - 3.1 \times P_\zeta^{vis} > -50$ GeV;
- at least one jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose).

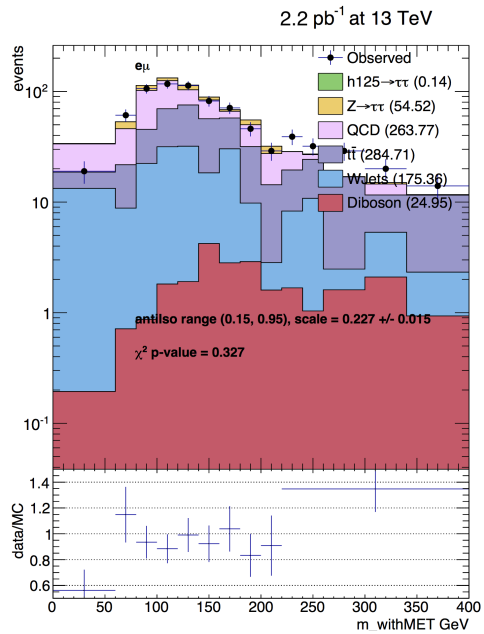

Figure 6.5.3: The distribution of reconstructed parent mass, $m\left(e, \mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$, in the same-charge, tightly-isolated sample: $e\mu$ channel.

The $t\bar{t}$ + single top production rate systematics estimated to be:

$$t\bar{t} + \text{single top systematics} = \left|1 - \frac{t\bar{t} + \text{single top}}{\text{Data - other backgrounds}}\right| = 8\% \tag{6.5.2}$$

which we apply both to $e\tau_h$ and $e\mu$ final states.

6.5.2.3 Di-boson process

We take di-boson processes directly from simulation with a 15% production uncertainty.

6.5.3 QCD background

The estimation of the QCD background for the $e\mu$ channel is directly analogous to that in the $e\tau_h$ channels, except that the relaxed isolation region is defined by the muon isolation. As shown in the Fig. 6.5.3, in order to achieve good QCD purity, the relative isolation sideband of 0.15 to 0.95 was chosen. After the signal region selection the "Loose-to-Tight" scale factor is estimated to be: 0.20 ± 0.08 where this 40% rate uncertainty is applied to the QCD process (in addition to the bin-by-bin systematic uncertainties).

6.5.4 W+jets background

W+jets background in the $e\mu$ channel is very small. However, the W+jets simulated sample was not generated with high statistics. As a workaround, the W+jets shape is taken from the simulated sample with the muon relative isolation from 0.15 to 0.95, and scaled to match the simulated yield where the muon relative isolation < 0.15. The "loose-to-tight" factor is 0.07 ± 0.03. Thus, we improve the statistics of the W+jets template by $1/0.07 \sim 13$ times.

6.6 Systematics

The systematic uncertainties considered in this analysis are detailed below and summarized in Table 6.4):

- **Luminosity:** As suggested by the official CMS recommendation [93], a 2.7% uncertainty on the measured luminosity is considered for all MC based events.
- **Electron Energy Scale:** Due to potential mis-modeling of the electron energy, a systematic variation of 1% (2.5%) shift on the electron energy scale in the barrel (endcap) regions is applied. The resultant systematic uncertainty on MC based events is $< 1\%$.
- **Electron Identification + Trigger:** A 6% uncertainty on the combination of electron reconstruction, identification and the single electron trigger is considered as measured with a data-driven method [94].
- **Muon Momentum Scale:** Due to potential mis-modeling of the muon momentum, a systematic variation of 1% shift on the muon momentum is applied. The resultant systematic uncertainty on MC based events is $< 1\%$.
- **Muon Identification + Trigger:** A 7% uncertainty on the combination of muon reconstruction, identification and the single muon trigger is considered as measured with a data-driven method [94].
- **Tau Energy Scale:** As suggested by the official CMS recommendation, for MC based samples, the tau energy is varied by $\pm 5\%$ and propagated to the final discriminant. In the end, the signal and MC based backgrounds fluctuates by up to $\sim 11\%$.
- **Tau Identification:** As suggested by the official CMS recommendation a 6% uncertainty per τ_h on the τ_h identification is considered. Additionally, due to low statistics of high-p_T τ_h

validations, an uncertainty per τ_h of 20% $\cdot$ (p_T/ 1 TeV) is considered.

- **Jet Energy Scale:** The effects of the jet energy measurement is estimated by varying the jet energy by $\pm 5\%$ (as suggested by the official CMS recommendation) and propagating it to the final discriminant.

- ***b*-Tagging Efficiency:** Uncertainties of CSV b-tagging is estimated by varying the b-tagged status of each jet based on the b-tagging efficiency scale factors (SF $= \epsilon_{\text{Data}}/\epsilon_{\text{MC}}$, where ϵ_{Data} and ϵ_{MC} are the b-tagging efficiencies in data and MC) provided by the official CMS recommendation [95]. Details of how the jet's b-tag status is varied can be found at [96], a brief summary is given below. For jets with SF < 1, one would only need to "demote" b-tagged jets to a "not b-tagged" state. The percentage of demoted jets, α, should be such that $\epsilon_{\text{MC}} \cdot \alpha = \epsilon_{\text{Data}}$. Thus, a random number c is generated between 0 and 1 and a b-tagged jet is demoted if $c >$ SF. On the other hand, if SF > 1, not b-tagged jets should be "promoted" as b-tagged. The percentage of promoted jets, β, should be such that $(1-\epsilon_{\text{MC}}) \cdot \beta + \epsilon_{\text{MC}} = \epsilon_{\text{Data}}$. Thus, a random number c between 0 and 1 is generated and a jet is "promoted" to be b-tagged if $c < \frac{1-\text{SF}}{1-1/\epsilon_{\text{MC}}}$.

- **PDF Systematics Uncertainty:** Following the "PDF4LHC recommendations for LHC Run II" [97], the PDF uncertainties are computed from the 68% confidence level with the PDF4LHC15_mc sets. The PDF uncertainties for our main backgrounds, $t\bar{t}$, W+Jets and DY, are much smaller than their bin-by-bin statistical uncertainties thus is neglected. The PDF uncertainties on the signal acceptance is estimated to range from 0.7% for Z' at 500 GeV up to 12% for Z' at 3 TeV [14].

- **Background Estimates:** The uncertainty on the data-driven background estimations are driven by the statistics in data in the various control regions. There is also a mostly negligible contribution from the contamination of other MC based backgrounds in the control regions. In cases where MC based backgrounds must be subtracted off, the statistical uncertainties in the MC backgrounds are propagated throughout the subtraction and used to assign a systematic uncertainty on the background prediction.

Table 6.4: Summary of systematic uncertainties. Values are given in percent. "s" indicates template variations ("shape" uncertainties). "p_T" indicates the τ_h p_T dependent uncertainty of 6% + 20%·(p_T/1 TeV)

Source channels:	QCD $e\tau_h$, $e\mu$	W $e\tau_h$, $e\mu$	DY $e\tau_h$, $e\mu$	$t\bar{t}$ $e\tau_h$, $e\mu$	di-boson $e\tau_h$, $e\mu$	Signal $e\tau_h$, $e\mu$
Lumi	–,–	2.7,2.7	2.7,2.7	2.7,2.7	2.7,2.7	2.7,2.7
μ ID + Trig	–,–	–,7	–,7	–,7	–,7	–,7
e ID + Trig	–,–	–,6	6,6	6,6	6,6	6,6
τ_h ID	–,–	–,–	p_T,–	p_T,–	p_T,–	p_T,–
b ID	–,–	s,s	s,s	s,s	s,s	s,s
JES	–,–	s,s	s,s	s,s	s,s	s,s
TES	–,–	s,s	s,s	s,s	s,s	s,s
MMS	–,–	–,1	–,1	–,1	–,1	–,1
EES	–,–	1,1	1,1	1,1	1,1	1,1
top p_T	–,–	–,–	–,–	–,s	–,–	–,–
pdf	–,–	–,–	–,–	–,–	–,–	(1–12)
bin-by-bin stat.	s,s	s,s	s,s	s,s	s,s	s,s
Closure+Norm.	18,37	9,41	10,10	8,8	15,15	–,–

6.7 Results and Interpretation

The distributions of $m\left(\tau_1, \tau_2, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ are shown in Figure 6.7.1 for $e\tau_h$ and $e\mu$ channels. As shown, no significant excess is found above the SM backgrounds.

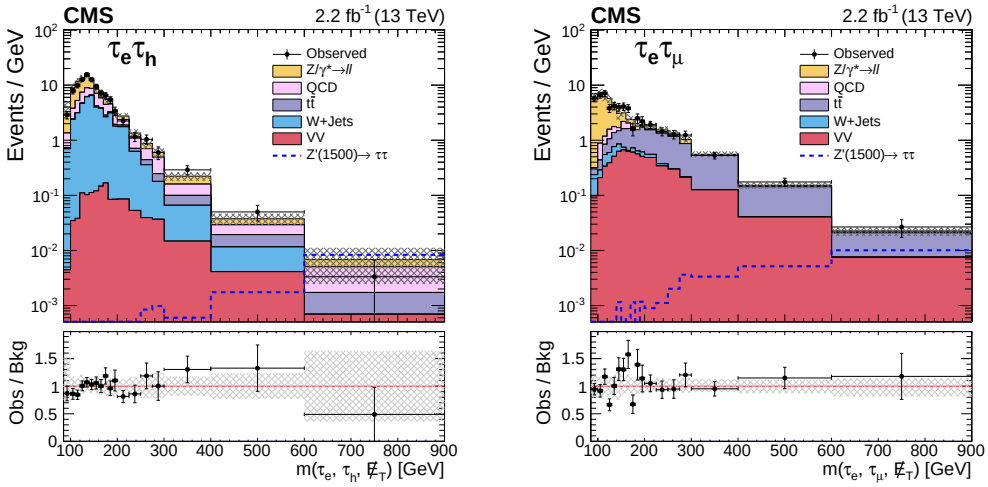

Figure 6.7.1: Observed $m\left(\tau_1, \tau_2, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ distribution in the signal region compared to the expected SM backgrounds for the (left) $e\tau_h$, and (right) $e\mu$ channels. The dashed histogram shows the distribution expected for a Z' boson with mass 1.5 TeV. The rightmost bins also include events with $m\left(\tau_1, \tau_2, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$, and are normalized to the displayed bin width. The lower panel shows the ratio of the observed number of events to the total background prediction. The shaded bands represent the total uncertainty in the background prediction [14].

At 95% confidence level (CL), upper limits on the product of the cross section and branching fraction into τ lepton pairs as a function of the Z' mass for $e\tau_h$, $e\mu$, and the combination of the four

channels are shown in Figure 6.7.2. Additional to the SSM Z' model, the topcolor-assisted technicolor (TAT) model [98,99], which predicts an enhanced coupling for Z' bosons to third-generation fermions, is also tested here. Z'_{TAT} predictions are estimated by scaling the Z'_{SSM} production cross section and branching fraction to those of Z'_{TAT}. Combining the four final states, we exclude Z'_{SSM} and Z'_{TAT} models with masses less than 2.1 TeV and 1.7 TeV, respectively, at 95% CL.

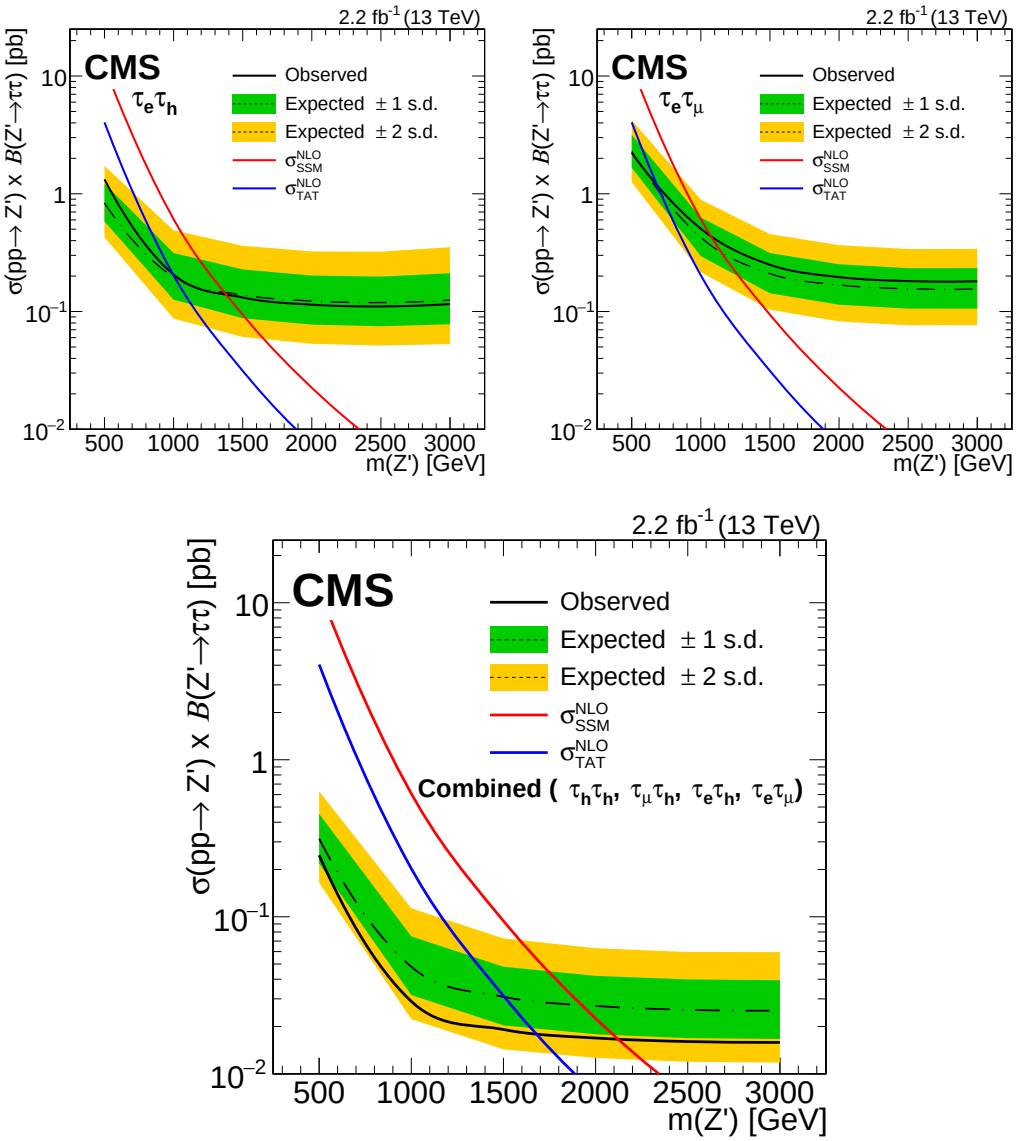

Figure 6.7.2: The observed 95% CL upper limits on the product of the cross section and branching fraction into τ lepton pairs as a function of the Z' mass m(Z') (solid black lines), for the (top left) $e\tau_h$, (top right) $e\mu$, and (bottom) for the combination of the four channels. The expected limits (dash-dotted lines) with one and two standard deviation (s.d.) uncertainty bands are also shown. The predictions of the NLO theory cross sections in the SSM and topcolor-assisted technicolor (TAT) models are represented by the red (lighter) and blue (darker) solid curves, respectively [14].

6.8 Summary

A search for heavy resonances decaying to a tau lepton pair has been performed by the CMS experiment, using a data sample of proton-proton collisions at $\sqrt{s} = 13TeV$ collected in 2015, corresponding to an integrated luminosity of 2.2 fb^{-1}. The observed invariant mass spectra in the $\mu\tau_h$, $e\tau_h$, $\tau_h\tau_h$ and $e\mu$ channels are measured and are found to be consistent with expectations from the SM. Upper limits at 95% CL are derived for the product of the cross section and branching fraction of a Z' boson decaying to a tau lepton pair, as a function of the Z' mass. Z' masses below 2.1 TeV and 1.7 TeV are excluded for the sequential standard and the topcolor-assisted technicolor models respectively.

Chapter 7

Search for new physics with high-mass tau pairs with 2016 data

In this chapter, a search for an extra neutral gauge boson that decays to a pair of τ leptons is presented. The analysis uses data collected during 2016 by the CMS experiment with a centre of mass energy $\sqrt{s} = 13$ TeV and an integrated luminosity of 35.9 fb^{-1}.

This chapter is organized as follows: Section 7.1 provides a brief introduction to the motivations of the search. Section 7.2 provides an overview of the analysis along with the key background estimation methods. Section 7.3 describes the triggers used in this analysis. The sample sets used in this analysis are given in details in Section 7.4. Section 7.5–7.7 describes in details the event selection and backgrounds estimation methods used in each of the considered decay channels. Section 7.8 discusses the considered systematic uncertainties. The statistical procedure used to set limits and their results are detailed in Section 7.9. Finally, a summary is given in Section 7.10 which concludes this chapter.

Out of the four decay channels studied: $\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$, and $e\mu$, only the channels of $\mu\tau_h$, $e\tau_h$, and $e\mu$ will be presented in detail as they are the channels I personally worked on.

7.1 Introduction

As discussed in Section 1.4, many models, ranging from the Little Higgs model to E_6 GUTs, predict the existence of extra gauge bosons. Of particular interest for this analysis are models that include an extra massive neutral gauge boson, referred to as Z', that decays to a pair of τ leptons. Due to its

simplicity, the Z' boson of Sequential Standard Model (SSM) is commonly chosen as the benchmark for Z' searches. Here, SSM predicts a Z' boson with the same coupling to quarks and leptons as the SM Z boson.

The latest results of searches for $Z' \to \tau\tau$ in pp collisions at $\sqrt{s} = 13$ TeV have been reported by the CMS collaboration [14] which excludes SSM Z' masses below 2.1 TeV. The most stringent mass limits on SSM Z' production are set, using the e^+e^- and $\mu^+\mu^-$ decay channels by CMS and ATLAS collaborations, to be 3.2 [91] and 3.4 [92] TeV respectively.

In this analysis, we follow closely to the $Z' \to \tau\tau$ search conducted using data collected in 2015 [14] with optimizations in event selection and background estimation methods. This study uses data samples of pp collisions at $\sqrt{s} = 13$ TeV recorded in 2016 by the CMS detector and corresponds to an integrated luminosity of 35.9 fb^{-1}.

7.2 Strategy

As discussed in Section 1.1.1, τ leptons have a rather short life time and decay within the detector. Around 1/3 of all τ leptons decay to an electron or muon while the rest decay into hadronic jets, referred to as τ_h. For hadronically decaying τ leptons, due to charge conservation, their decay products mostly consist of one or three charged mesons and are respectively referred to as one-prong and three-prong decay modes. Out of the six possible decay channels of a pair of τ leptons, four channels are chosen ($\tau_h\tau_h$, $\mu\tau_h$, $e\tau_h$ and $e\mu$) as ee and $\mu\mu$ channels are hard to separate from prompt $Z' \to ee/\mu\mu$ decays. SSM Z' MC samples between the mass of 500 GeV to 4000 GeV, in steps of 50 GeV to 500 GeV, are produced and used in this study.

Given the large mass of the Z' of our interest, τ's coming from Z' are very energetic and most likely to be traveling in opposite directions. Due to the large energy of these τ leptons, their decay products are most likely collinear with the original τ. Thus, in general, we selected events with two back-to-back objects that have opposite electric charge. Due to the neutrinos from τ decays, we also expect the presence of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ in the event. For signal events in $\mu\tau_h$ and $e\tau_h$ channels, due to the extra neutrino when $\tau \to l + \nu_\tau + \nu_l$ with $l = e, \mu$, $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is more likely to be aligned with the lepton direction than the τ_h direction. For signal events in $\tau_h\tau_h$ and $e\mu$ channels, with the balanced amount of neutrinos in either directions, $p_{\mathrm{T}}^{\mathrm{miss}}$ is more likely aligned with the direction of the reconstructed particle which has less energy. Based on these topological features, as will be discussed in more detail in sections 7.5–7.7, selections are chosen to suppress SM backgrounds while maintaining a high acceptance efficiency for signal events.

The modeling of SM backgrounds consists of a mixture of MC based and data-driven methods. In general, backgrounds with genuine τ_h are modeled with MC simulations. Dedicated signal depleted control regions are used to estimate the MC modeling of such backgrounds. For backgrounds with jets misidentified as τ_h or l, mostly QCD and events with a W boson and accompanying jets, a data-driven background estimation technique is used.

The invariant mass of the two reconstructed objects + $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ system, referred to as the effective mass, provides a good separation between signal and background. To quantify the significance of any possible excess, a fit is performed with the effective mass distribution using a modified frequentist approach, the CL_s method [89], for each channel.

7.3 Triggers

For the $l+\tau_h$ final states, we use single lepton triggers instead of $l \times \tau_h$ cross-triggers which allows us to use the different τ_h isolation values to construct control and validation regions. For the $e\mu$ final state, any trigger with iso-muon requirement, either a e-μ cross-trigger or single-muon trigger, would eliminate the isolation regions needed for QCD estimation. Hence, we use the same single-electron trigger as for the $e\tau_h$ final state. Trigger paths are summarized in Table 7.1. This constrains the object ID and phase space that we are studying. For example, the use of these triggers requires p_T $>$ 35 GeV for the leading light leptons in the $\mu\tau_h$, $e\tau$, and $e\mu$ channels. However it doesn't affect our analysis since our signal is characterized by high-p_T leptons.

Table 7.1: The trigger paths used.

Channel	Trigger Path
$\mu\tau_h$ (data and MC)	one isolated muon with $p_T > 24$ GeV
$e\tau_h$, $e\mu$ (data and MC)	one electron passing tight working point with $p_T > 27$ GeV

7.3.1 Single Lepton Trigger Efficiency

The single electron and muon trigger efficiencies are measured using a tag and probe method where we selected events with at least one electron and one muon satisfying the following requirements:

For electrons:

- $p_T > 13$ GeV, $|\eta| < 2.1$, isolation < 0.15, transverse distance of closest approach to the leading sum-p_T^2 primary vertex (d_{xy}) is less than 0.045 cm, longitudinal distance of closest approach

to the leading sum-p_T^2 primary vertex (d_z) is less than 0.2 cm;

- passing 'Loose' working point of MVA electron ID;
- no matched conversions;
- number of missing hits = 0.

For muons:

- $p_T > 10$ GeV, $|\eta| < 2.1$, isolation < 0.15, $d_{xy} < 0.045$ cm, $d_z < 0.2$ cm;
- passing "Medium" muon ID;

For the pair:

- $\Delta\mathrm{R}(e, \mu) > 0.3$;
- selected the pair containing the most isolated muon, if there are more than one of such pair, select the pair containing the most isolated electron;
- reject events if the pair has the same electric charge;
- reject events if there is an additional electron with isolation < 0.3, or an additional muon with isolation < 0.3.

7.3.2 Single Electron Trigger Efficiency

To measure the single-electron trigger efficiency, after the selections mentioned above, we select (tag) events with a single-muon trigger while requiring the muon $p_T > 24$ GeV and matching to the HLT muon that fired the trigger. Then, the trigger efficiency is defined as the fraction of events which also pass (probe) the single-electron trigger.

The efficiency curves of the single-electron triggers versus electron p_T, are shown in Figure 7.3.1. A $p_T > 35$ GeV cut is motivated to achieve a flat data/MC ratio versus electron p_T. The difference in trigger efficiency between data and MC is measured in different detector sections: barrel ($|\eta| < 1.479$) and endcap ($1.479 < |\eta| < 2.1$). As shown, this difference has little dependence on the electron p_T. Thus, a flat line is used to fit the difference and we obtained a correction of 1.002 (0.957) for simulated events when the electron is in the barrel (endcap) region. Overall, we assign a 5% systematic uncertainty for the single electron trigger efficiency correction due to the outliers in the fit.

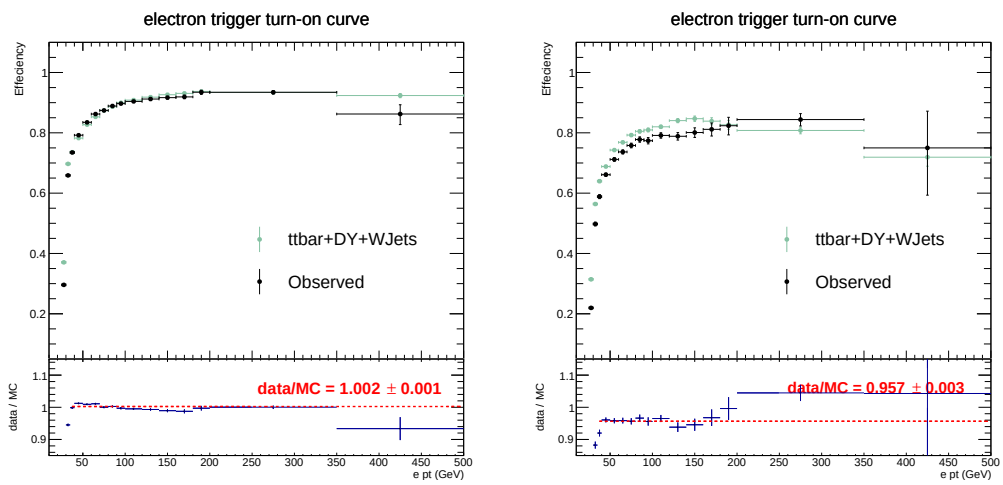

Figure 7.3.1: The efficiency vs. p_T curves of the single-electron triggers used. Left column: $|\eta| < 1.479$. Right column: $|\eta| > 1.479$.

7.3.3 Single Muon Trigger Efficiency

To measure the single-muon trigger efficiency, after the selections mentioned above, we select (tag) events with a single-electron trigger while requiring the electron $p_T > 28$ GeV and matching to the HLT electron that fired the trigger. Then, the trigger efficiency is defined as the fraction of events which also pass (probe) the single-muon trigger (HLT_IsoMu24).

The efficiency curves of the single-muon triggers vs. muon p_T, are shown in Figure 7.3.2. A $p_T >$ 35 GeV cut is motivated to achieve a flat data/MC ratio versus muon p_T and to be synchronized with the $e\tau_h$ channel. Similar to the single electron trigger, we measured the difference in trigger efficiency between data and simulated events in several regions: barrel ($|\eta| < 0.8$); middle ($0.8 < |\eta| < 1.24$); and endcap ($1.24 < |\eta| < 2.1$). From the flat fit, the corrections for simulated events are found to be 0.973, 0.956 and 0.974 for events with muons in the η regions mentioned above. A 5% single muon trigger efficiency correction systematics is assigned to accommodate the few outliers in the fit.

7.4 Data and Monte Carlo Samples

The 13 TeV collision data collected by the CMS detector in year 2016 is used in this analysis. Table 7.2 shows the collision datasets used. The recommended file (Cert_271036-284044_13TeV _03Feb2017ReReco_Collisions16_JSON.txt) containing only certified events is used to filter out bad events for all of the data. The total integrated luminosity of the collision data samples is 35.9 fb^{-1}.

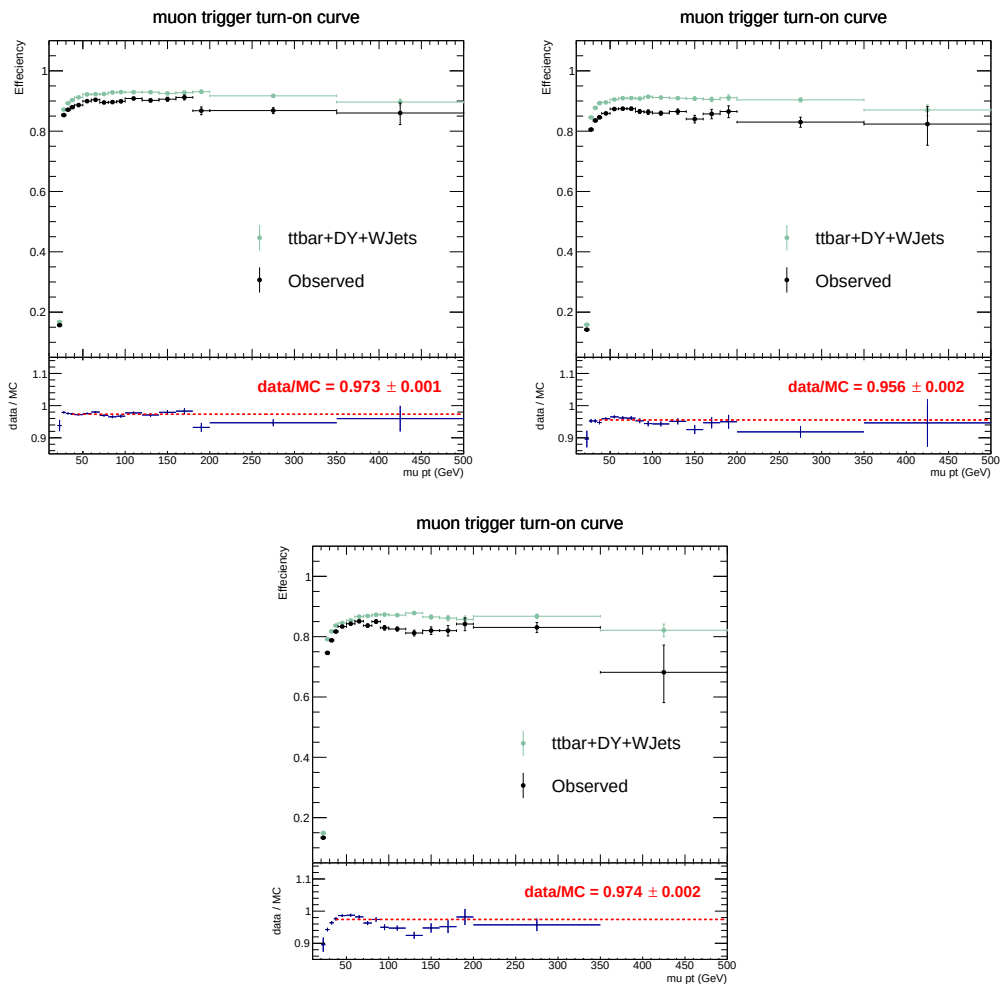

Figure 7.3.2: The efficiency vs. p_T curves of the single-muon triggers used. Top left: $|\eta| < 0.8$. Top right: $0.8 < |\eta| < 1.24$. Bottom: $|\eta| > 1.24$.

The official Summer 2016 miniAODv2 MC samples are used to model most of the SM processes. The leading order generators, PYTHIA8 and MADGRAPH, are mainly used for signal and background MC production. The predicted background yields in simulation are determined using NLO or NNLO cross-sections, while the signal yields and distributions in all plots shown are normalized using the LO cross-sections shown in Table 7.3. Table 7.4 shows the entire list of the MC samples used for this analysis.

Table 7.2: Collision Data Samples

Physics Sample	Run Period	Official CMS Datasets
SingleMu	Run2016B	*/SingleMuon/Run2016B-03Feb2017_ ver2-v2/MINIAOD*
SingleMu	Run2016C	*/SingleMuon/Run2016C-03Feb2017-v1/MINIAOD*
SingleMu	Run2016D	*/SingleMuon/Run2016D-03Feb2017-v1/MINIAOD*
SingleMu	Run2016E	*/SingleMuon/Run2016E-03Feb2017-v1/MINIAOD*
SingleMu	Run2016F	*/SingleMuon/Run2016F-03Feb2017-v1/MINIAOD*
SingleMu	Run2016G	*/SingleMuon/Run2016G-03Feb2017-v1/MINIAOD*
SingleMu	Run2016H	*/SingleMuon/Run2016H-03Feb2017_ ver2-v1/MINIAOD*
SingleMu	Run2016H	*/SingleMuon/Run2016H-03Feb2017_ ver3-v1/MINIAOD*
SingleElectron	Run2016B	*/SingleElectron/Run2016B-03Feb2017_ ver2-v2/MINIAOD*
SingleElectron	Run2016C	*/SingleElectron/Run2016C-03Feb2017-v1/MINIAOD*
SingleElectron	Run2016D	*/SingleElectron/Run2016D-03Feb2017-v1/MINIAOD*
SingleElectron	Run2016E	*/SingleElectron/Run2016E-03Feb2017-v1/MINIAOD*
SingleElectron	Run2016F	*/SingleElectron/Run2016F-03Feb2017-v1/MINIAOD*
SingleElectron	Run2016G	*/SingleElectron/Run2016G-03Feb2017-v1/MINIAOD*
SingleElectron	Run2016H	*/SingleElectron/Run2016H-03Feb2017_ ver2-v1/MINIAOD*
SingleElectron	Run2016H	*/SingleElectron/Run2016H-03Feb2017_ ver3-v1/MINIAOD*

Table 7.3: MC Signal Samples

Process (mass [GeV])	width[GeV]	cross-section [pb]	Official CMS Datasets (MINIAODSIM)
SSM Z'(500)	14.73	5.751	*/ZprimeToTauTau_ M-500_ TuneCUETP8M1_ 13TeV-pythia8-tauola* [1]
SSM Z'(750)	22.89	1.236	*/ZprimeToTauTau_ M-750_ TuneCUETP8M1_ 13TeV-pythia8-tauola* [1]
SSM Z'(1000)	30.97	0.3865	*/ZprimeToTauTau_ M-1000_ TuneCUETP8M1_ 13TeV-pythia8-tauola* [1]
SSM Z'(1250)	39.02	0.149	*/ZprimeToTauTau_ M-1250_ TuneCUETP8M1_ 13TeV-pythia8-tauola* [1]
SSM Z'(1500)	47.05	0.06479	*/ZprimeToTauTau_ M-1500_ TuneCUETP8M1_ 13TeV-pythia8-tauola* [1]
SSM Z'(1750)	55.07	0.03104	*/ZprimeToTauTau_ M-1750_ TuneCUETP8M1_ 13TeV-pythia8-tauola* [1]
SSM Z'(2000)	63.10	0.01583	*/ZprimeToTauTau_ M-2000_ TuneCUETP8M1_ 13TeV-pythia8-tauola* [1]
SSM Z'(2500)	79.16	0.004684	*/ZprimeToTauTau_ M-2500_ TuneCUETP8M1_ 13TeV-pythia8-tauola* [1]
SSM Z'(3000)	95.24	0.001592	*/ZprimeToTauTau_ M-3000_ TuneCUETP8M1_ 13TeV-pythia8-tauola* [1]
SSM Z'(3500)	111.3	0.0005972	*/ZprimeToTauTau_ M-3500_ TuneCUETP8M1_ 13TeV-pythia8-tauola* [1]
SSM Z'(4000)	127.5	0.0002447	*/ZprimeToTauTau_ M-4000_ TuneCUETP8M1_ 13TeV-pythia8-tauola* [1]

[1] `/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1`

Table 7.4: MC Background Samples

Process	cross-section [pb] best theoretical calculation or $\sigma_{\rm LO}\times$ k-factor, where k-factor $= \frac{\sigma_{\rm NLO}}{\sigma_{\rm LO}}$	Official CMS Datasets (MINIAODSIM)
$Z \to ll$	5765.4	/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 [3]
mass binned	7.67×1.031	/DYJetsToLL_M-200to400_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8 [5]
LO samples	0.423×1.008	/DYJetsToLL_M-400to500_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8 [2]
	0.24×0.996	/DYJetsToLL_M-500to700_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8 [2]
	0.035×0.973	/DYJetsToLL_M-700to800_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8 [2]
	0.03×0.961	/DYJetsToLL_M-800to1000_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8 [2]
	0.016×0.938	/DYJetsToLL_M-1000to1500_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8 [2]
	0.002×0.882	/DYJetsToLL_M-1500to2000_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8 [2]
	0.00054×0.825	/DYJetsToLL_M-2000to3000_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8 [2]
$W + jets$	61526.7	/WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 [4]
HT binned	1345×1.21	/WJetsToLNu_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 [3]
LO samples	359.7×1.21	/WJetsToLNu_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 [3]
	48.91×1.21	/WJetsToLNu_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 [2]
	12.05×1.21	/WJetsToLNu_HT-600To800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 [2]
	5.501×1.21	/WJetsToLNu_HT-800To1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 [2]
	1.329×1.21	/WJetsToLNu_HT-1200To2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 [2]
	0.03216×1.21	/WJetsToLNu_HT-2500ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 [2]
$t\bar{t}$	831.76	/TT_TuneCUETP8M2T4_13TeV-powheg-pythia8 [1]
single Top samples	35.6	/ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1 [2]
	35.6	/ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1 [2]
	136.02	/ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1 [1]
	26.23	/ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1 [1]
	3.344	/ST_s-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8_TuneCUETP8M1 [1]
$VV \to 2l2\nu$	11.95	/VVTo2L2Nu_13TeV_amcatnloFXFX_madspin_pythia8 [2]
$ZZ \to 2l2q$	3.22	/ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8 [1]
$ZZ \to 4l$	1.256	/ZZTo4L_13TeV_powheg_pythia8 [1]
$WW \to l\nu 2q$	1.212	/WWToLNuQQ_13TeV-powheg [2]
$WZ \to 2l2q$	5.595	/WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8 [1]
$WZ \to 3l\nu$	4.708	/WZJToLLLNu_TuneCUETP8M1_13TeV-amcnlo-pythia8 [1]
$WZ \to l3\nu$	3.05	/WZTo1L3Nu_13TeV_amcatnloFXFX_madspin_pythia8 [1]
$WZ \to l\nu 2q$	10.71	/WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8 [6]

[1] /RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1
[2] /RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1
[3] /RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v2
[4] /RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext2-v1
[5] /RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext2-v2
[6] /RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v3

7.4.1 Corrections for Pile-Up

In order to match the pile-up distribution used to generate MC samples to that of the observed data, an event based reweighting is deployed for the MC samples with the weight:

$$w_{PU}(n) = \frac{P_{data}(n)}{P_{MC}(n)} \quad , \tag{7.4.1}$$

where $P_{data}(n)$ and $P_{MC}(n)$ are the number of PVs in data and MC.

The recommended min-bias cross-section of 69.2 mb is used to determine these weights. Figure 7.4.1 shows the distributions of the number of reconstructed vertices, before and after applying the pile-up weights.

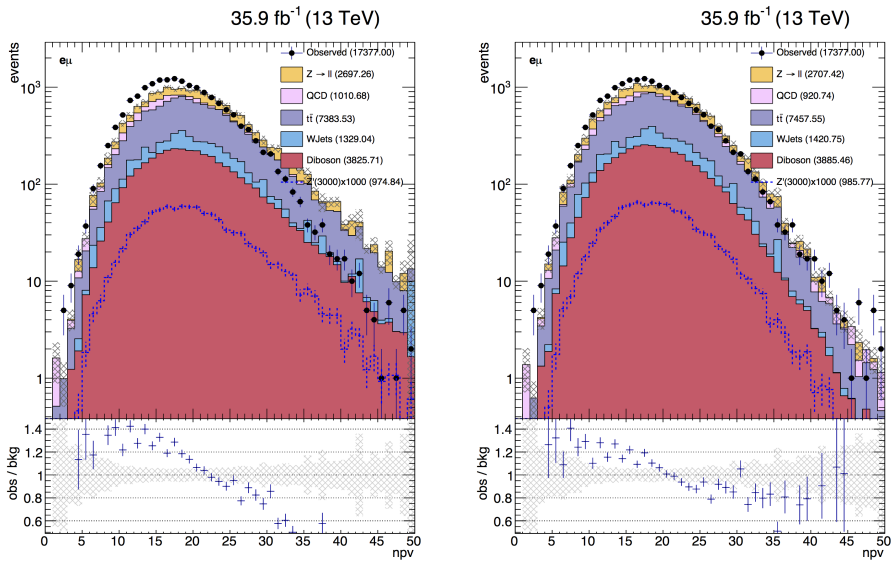

Figure 7.4.1: Distributions of the number of reconstructed primary vertices in the $e\mu$ channel. Left: before reweighting. Right: after reweighting.

7.5 Muon + Hadronic Tau Channel

7.5.1 Event selection

Events are first required to fire the single-muon trigger described in Section 7.3. Muons are required to:

- have $p_T > 35$ GeV, $|\eta| < 2.1$, $d_{xy} < 0.045$ cm and $d_z < 0.2$ cm;
- satisfy the "Medium" muon requirement;
- have no matched conversion nor missing hits;
- be within $\Delta R < 0.5$ of the HLT muon that fired the trigger.

Offline τ_h's are required to:

- have $p_T > 20$ GeV, $|\eta| < 2.1$ and $d_z < 0.02$ cm;
- pass the HPS tau reconstruction requirement while containing one or three charged hadrons;
- pass the "Very Loose" working point for anti-electron MVA ID and pass the "Tight" working point for anti-muon ID.

Pairs of muons and τ_h's are built when the muon and τ_h are separated by at least $\Delta R > 0.5$. For events with more than one such pair, we select the pair with the two most isolated leptons, considering first the muon, and then the τ_h. This criterion was seen to have good efficiency for signal samples. In the rare case of multiple such pairs having identical isolation values, the pair with the highest scalar sum p_T is chosen.

After a pair has been chosen for an event, we apply the following isolation requirements on the leptons, for an event to enter the signal region: muon relative isolation < 0.15; τ_h isolation pass the "Tight" working point of MVA tau isolation ID. In order to keep the different final states exclusive, an event is rejected if there is an additional muon satisfying the above identification requirements and with relative isolation < 0.3, or an electron satisfying the identification requirements described in Section 7.7.1 with relative isolation < 0.3. To further reduce backgrounds, we require the muon and τ_h to have opposite electric charge. The selection set mentioned above is referred to as the preselection.

7.5.2 Signal region selections

The topology of a massive $Z' \to \tau\tau$ decay consists of two energetic back-to-back τ's. As τ's decay in the detector, undetectable neutrinos are produced leaving imbalance in the total transverse momentum. Backgrounds without such topology are reduced by requiring $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV and $\cos\Delta\phi(\mu, \tau_h) < -0.95$. To further reduce $t\bar{t}$ backgrounds, events with one or more jets passing the CSV loose b-tag requirement are removed from the signal region.

The selections mentioned above follow closely the ones used in the 2015 $Z' \to \tau\tau$ search [14]. Since then, an optimization study was performed, detailed in appendix A, in attempt to further separate signal from background. The following selections are the results of this study.

W+jets is one of the most dominant backgrounds in the $l + \tau_h$ channels, where the lepton comes from a leptonically decaying W and a random jet fakes the τ_h. The neutrino from the W boson decay is likely to be back-to-back with respect to the lepton. Thus, with $\cos\Delta\phi(\mu, \tau_h) < -0.95$, the $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ (due to neutrinos from W decays) is most likely in the same direction as the τ_h. However, for $Z' \to \tau\tau$ decays, two neutrinos are expected in the lepton direction while only one neutrino is expected in the τ_h direction. Thus, the $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ in $Z' \to \tau\tau$ decays is more likely to be aligned with the reconstructed lepton. Figure 7.5.1 shows, with $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV, $\cos\Delta\phi(\mu, \tau_h) < -0.95$ and 0 b-jet requirements, the separation in ϕ between the muon and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ in the left panel and τ_h and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ in the right panel.

As shown, most of the W+jets background can be rejected by $\cos\Delta\phi(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$. However, there are roughly 20% of signal events in the $\cos\Delta\phi(\tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ category. Figure 7.5.2 shows the transverse mass distribution, $m_{\mathrm{T}}(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) = \sqrt{2p_T^{\mu} p_{\mathrm{T}}^{\mathrm{miss}}(1 - \cos\theta(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}))}$, of events whose $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is aligned with τ_h. By requiring $m_{\mathrm{T}}(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 150$ GeV, most of the W+jets background are rejected while having little to no effect on signal events.

Hence, following the preselection, the signal region is defined as having:

- $\cos\Delta\phi(\mu, \tau_h) < -0.95$;
- $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;
- $\cos\Delta\phi(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ or ($\cos\Delta\phi(\tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ and $m_{\mathrm{T}}(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 150$ GeV);
- no jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose) .

The distributions of these variables after preselection, and after selection requirements on the other variables, are shown in Figure 7.5.3.

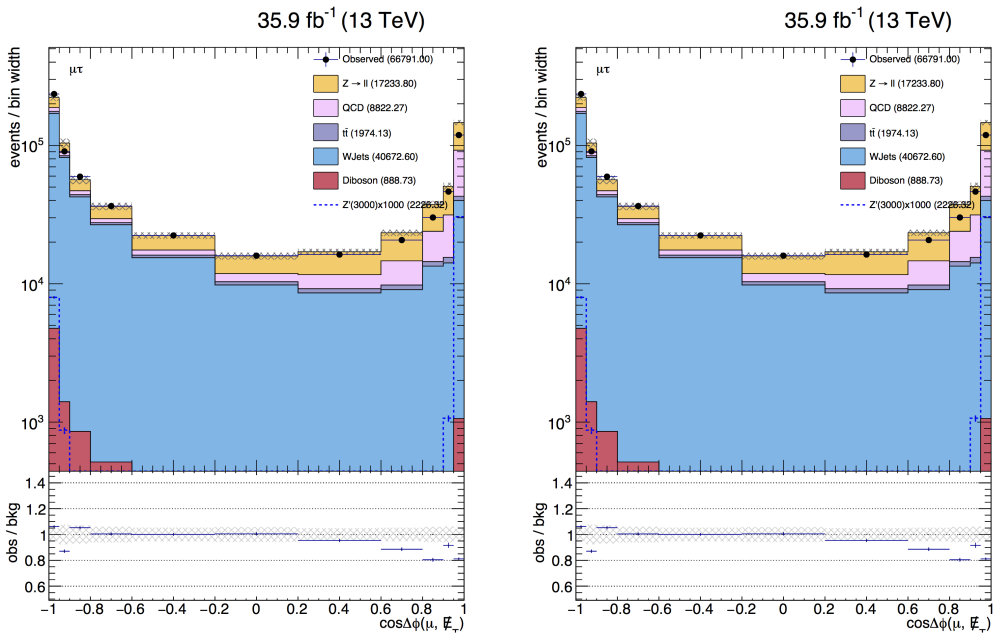

Figure 7.5.1: (Data driven QCD and MC based W+jets with only statistical uncertainties) Left: Distribution of $\cos\Delta\phi(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$. Right: Distribution of $\cos\Delta\phi(\tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$

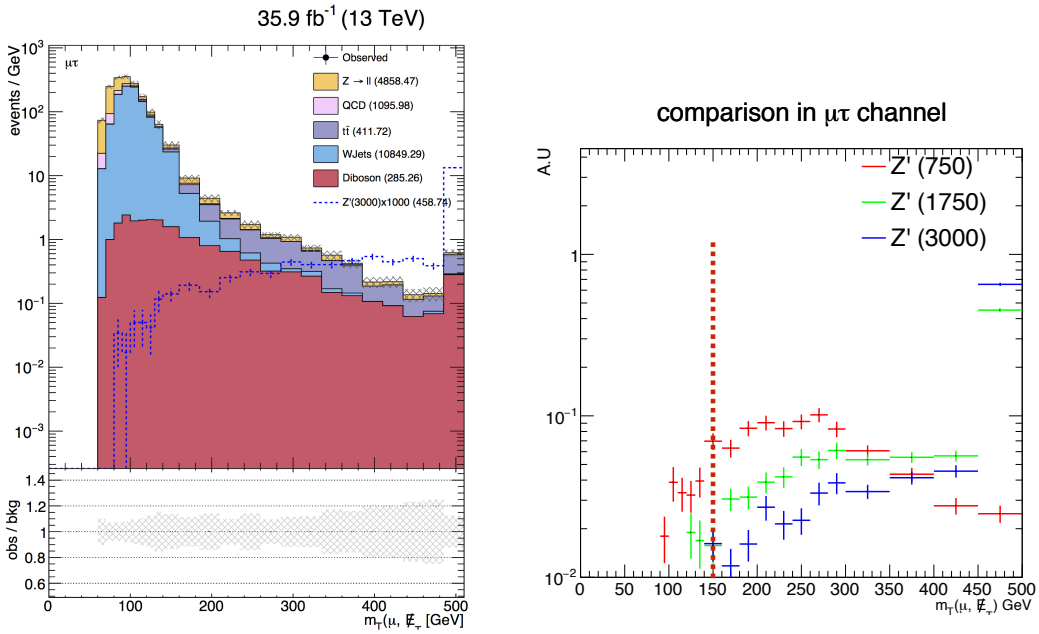

Figure 7.5.2: With $\cos\Delta\phi(\tau_h, p_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ selection. Left: (Data driven QCD and MC based W+jets with only statistical uncertainties) Distribution of $m_{\mathrm{T}}(\mu, p_{\mathrm{T}}^{\mathrm{miss}})$ with $Z'(3000)$. Right: $m_{\mathrm{T}}(\mu, p_{\mathrm{T}}^{\mathrm{miss}})$ of $Z'(750)$, $Z'(1750)$ and $Z'(3000)$

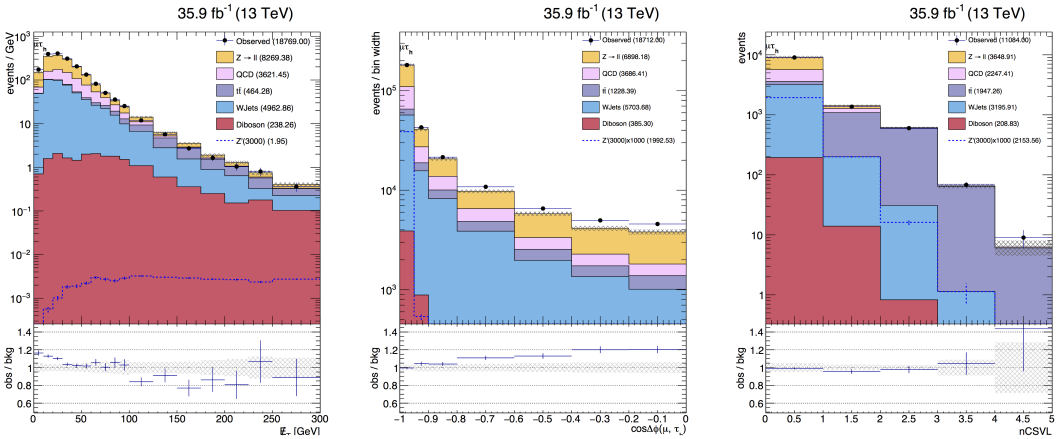

Figure 7.5.3: (Data driven W+jets and QCD with only statistical uncertainties) Distributions of the variables used for $\mu\tau_h$ signal selection, after all other signal selection requirements on variables other than the one plotted: $p_{\mathrm{T}}^{\mathrm{miss}}$ (left), $\cos\Delta\phi(\mu,\tau_h)$ (middle, with $\cos\Delta\phi(\mu,\tau_h) > 0$) and the number of b-tagged (CSV loose) jets, n_b (right).

The Standard Model processes considered as backgrounds are Drell-Yan, di-boson production, top quark single and pair production, W+jets production, and QCD multi-jet production.

7.5.3 Genuine dilepton events

Studies of simulated events indicate that for Drell-Yan process, top quark single and pair production, and di-boson production, the reconstructed and selected electrons and hadronic taus are typically associated with genuine simulated leptons. The nominal expected event rates are estimated by scaling the simulated samples by the best available cross sections, listed in Table 7.4, and by the integrated luminosity of the data samples.

7.5.3.1 Drell-Yan process

Due to large W+jets and QCD contamination in the $\mu\tau_h$ channel, as shown in the left panel of Fig 7.5.4 with the following selections:

- pass preselection;
- $p_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV;
- no jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose);
- $\cos\Delta\phi(\mu,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ or $\cos\Delta\phi(\tau_h,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$;
- $60\ \mathrm{GeV} < m\left(\mu,\tau_h,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right) < 150\ \mathrm{GeV}$;

we use the Drell-Yan yield systematic uncertainty (5%) estimated from the $e\mu$ channel in 7.7.2.1 for its higher Drell-Yan purity. This value is cross-checked in the $\mu\tau_h$ channel by comparing the Drell-Yan rate between data and MC:

$$\frac{\text{data - non-Drell-Yan backgrounds}}{\text{MC Drell-Yan}} = 1.04 \pm 0.03 \quad , \tag{7.5.1}$$

which agrees with the measurement from the $e\mu$ channel.

7.5.3.2 $t\bar{t}$ and single top processes

For the $\mu\tau_h$ channel, the $t\bar{t}$ and single top production rate systematics (1%) are estimated from the $e\mu$ channel, as described in 7.7.2.2, for its higher top purity. To accommodate the difference between data and theoretical calculations in $t\bar{t}$ differential cross section, $t\bar{t}$ events are reweighted based on the momentum of the top quarks and the latest scale factors provided by the official CMS recommendation [100].

This is cross-checked for the $\mu\tau_h$ channel in a top-rich region, defined by the following selections and shown in the right panel of Figure 7.5.4:

- pass preselection;
- $\cos\Delta\phi(\mu, \tau_h) < -0.95$;
- $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;
- $\cos\Delta\phi(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ or ($\cos\Delta\phi(\tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ and $m_{\mathrm{T}}(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 150$GeV)
- at least one jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose).

The $t\bar{t}$ + single top data/MC overall agreement is estimated to be:

$$\frac{\text{data - non } t\bar{t} + \text{single top backgrounds}}{\text{MC } t\bar{t} + \text{single top}} = 1.02 \pm 0.04 \quad , \tag{7.5.2}$$

which agrees with the measurement from the $e\mu$ channel.

7.5.3.3 Di-boson process

Di-boson processes are a relatively small background in the $\mu\tau_h$ channel. They are estimated directly from simulation with a 20% systematic uncertainty measured from the $e\mu$ channel in Section 7.7.2.3.

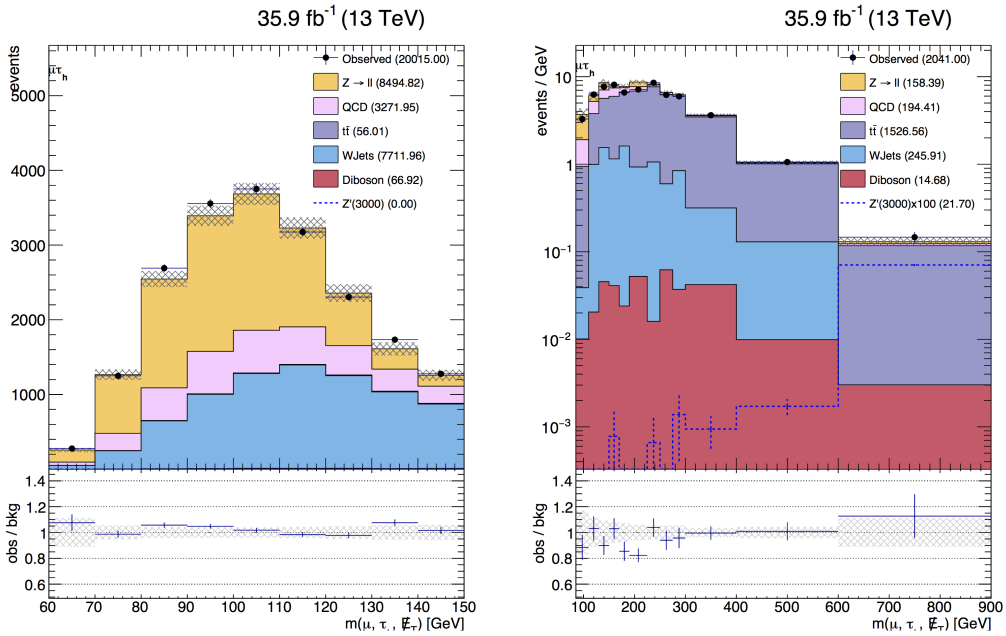

Figure 7.5.4: (Data driven W+jets and QCD with only statistical uncertainties) Distributions of $m\left(\mu, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$. Left: validation region with $p_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV, $n_b = 0$ and $60 <$ GeV $m\left(\mu, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right) < 150$ GeV. Right: validation region with $n_b \geq 1$.

7.5.4 QCD background

For a given variable and binning, e.g. the effective mass variable used for signal extraction, we construct a data-driven template for the shape of the QCD background. Based on the electric charge of the reconstructed objects, the muon relative isolation and τ_h isolation, events are split into four regions shown in Figure 7.5.5 and described below:

- A (Signal) Region: μ and τ_h have opposite charge, μ relative isolation is less than 0.15, and τ_h pass "Tight" isolation requirement.
- B Region: μ and τ_h have same charge, μ relative isolation is less than 0.15, and τ_h pass "Tight" isolation requirement.
- E Region: μ and τ_h have opposite charge, μ relative isolation is between 0.15 and 1.0, and τ_h fail "Tight" isolation requirement.
- F Region: μ and τ_h have same charge, μ relative isolation is between 0.15 and 1.0, and τ_h fail "Tight" isolation requirement.

In region B, QCD events are estimated by subtracting events with genuine leptons (estimated by MC) bin-by-bin from data. In regions E and F, due to very high purity of QCD events, QCD

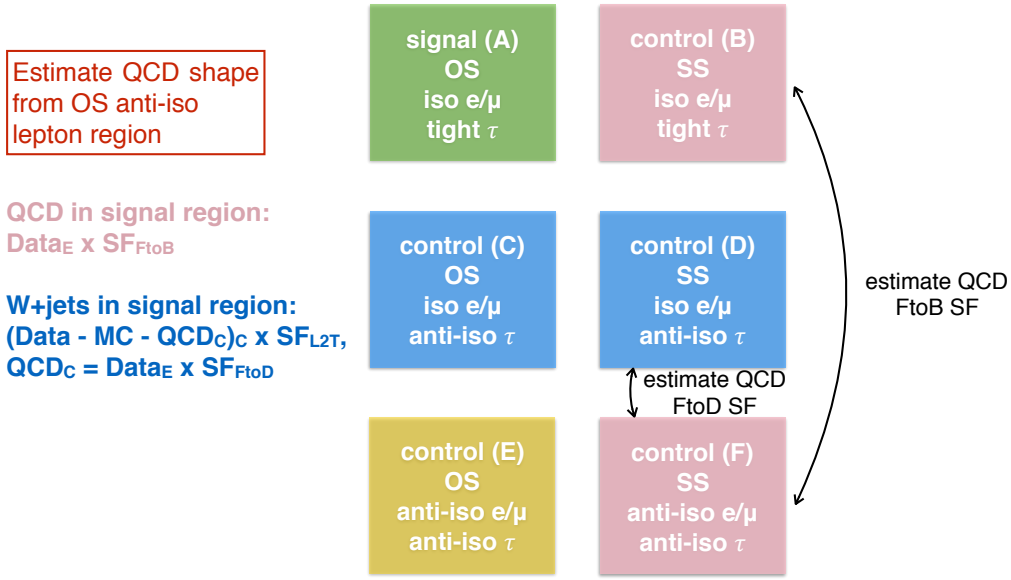

Figure 7.5.5: Data driven QCD and W+jets estimation and validation strategy for the $e\tau_h$, $\mu\tau_h$ channels.

is estimated directly as the observed events. The two jets faking τ_h leptons are expected to have the same probability of carrying the same and opposite electric charges, thus, the amount of QCD events in region B should be comparable to that of in the signal region. However, with the freedom to define the anti-isolation region, we can choose an anti-isolation definition such that QCD purity (98%) is much higher compared to the signal region. Taking the QCD shape from region E will help us reduce the effects of potential mis-modeling of non-QCD backgrounds on the QCD estimation.

Hence, QCD events in the signal region are estimated with the shape from region E and multiplied by a "QCD loose to tight" scale factor derived between regions B and F. Given the high purity (98%) of QCD in region F, QCD is taken straight from data in region F. The factor is defined as:

$$f_{\mathrm{LT}}^{\mathrm{QCD}} = \left(N_{\mathrm{data}}^{\mathrm{B}} - N_{\mathrm{non-QCD\ MC}}^{\mathrm{B}}\right) / N_{\mathrm{data}}^{\mathrm{F}} \quad . \tag{7.5.3}$$

As a summary, the QCD distribution in the signal region is estimated as the following:

$$(\text{QCD distribution})_{\mathrm{A}} = f_{\mathrm{LT}}^{\mathrm{QCD}} \times (\text{data distribution})_{\mathrm{E}} \quad . \tag{7.5.4}$$

Table 7.5.4 shows the yields of data and MC samples in regions B and F used for the calculation of $f_{\mathrm{LT}}^{\mathrm{QCD}}$.

This QCD estimation method is valid only if the QCD shape in the anti-isolated region correctly

Table 7.5: Event yields in regions B and F used for the calculation of $f_{\mathrm{LT}}^{\mathrm{QCD}}$.

Process	region B	region F
Z + jets	189 ± 28	13 ± 8
$t\bar{t}$	67 ± 5	16 ± 2
W + jets	745 ± 134	76 ± 35
DiBoson	19 ± 2	0.6 ± 0.3
non-QCD background	1020 ± 137	106 ± 36
Data	2197	6452
Data - non-QCD background	1177 ± 137	6452
$f_{\mathrm{LT}}^{\mathrm{QCD}}$	0.18 ± 0.02	

models the QCD shape in the isolated region. This check is performed by comparing the observation and background estimation in region B with the QCD shape taken from region F and normalized to QCD in region B. An example of this test is shown in the left panel of Fig. 7.5.6. Here, W+jets is estimated directly from MC simulations. Overall, the data and the estimated background agree reasonably. Due to low statistics of the W+jets MC sample, some bins disagree by 20%.

7.5.5 W+jets background

The simulated W+jets samples were not generated with large enough MC statistics. If used directly, avoiding non-smooth templates restricts somewhat the choice of signal selection and binning. It also complicates the validation of the background estimates. As a workaround, similar to the QCD estimation, we construct a data-driven template for the shape of the W+jets background in a control region with high W+jets purity. Based on the electric charge of the reconstructed objects and τ_h isolation, events are split into four regions shown in Fig. 7.5.5 and described below:

- Regions A and B are the same as in Section 7.5.4.
- C Region: μ and τ_h have opposite charge, μ relative isolation is less than 0.15, and τ_h pass "Very Loose" isolation requirement but fail "Tight" isolation requirement.
- D Region: μ and τ_h have same charge, μ relative isolation is less than 0.15, and τ_h pass "Very Loose" isolation requirement but fail "Tight" isolation requirement.

In most of the cases, W+jets backgrounds pass our signal region selection by having a jet faking the reconstructed τ_h. Thus, by relaxing the τ_h ID, in region C and D, one would greatly increase the acceptance of W+jets events. W+jets events in the signal region is estimated by applying a "jet to tau fake rate" to the estimated W+jets events in region C as the following:

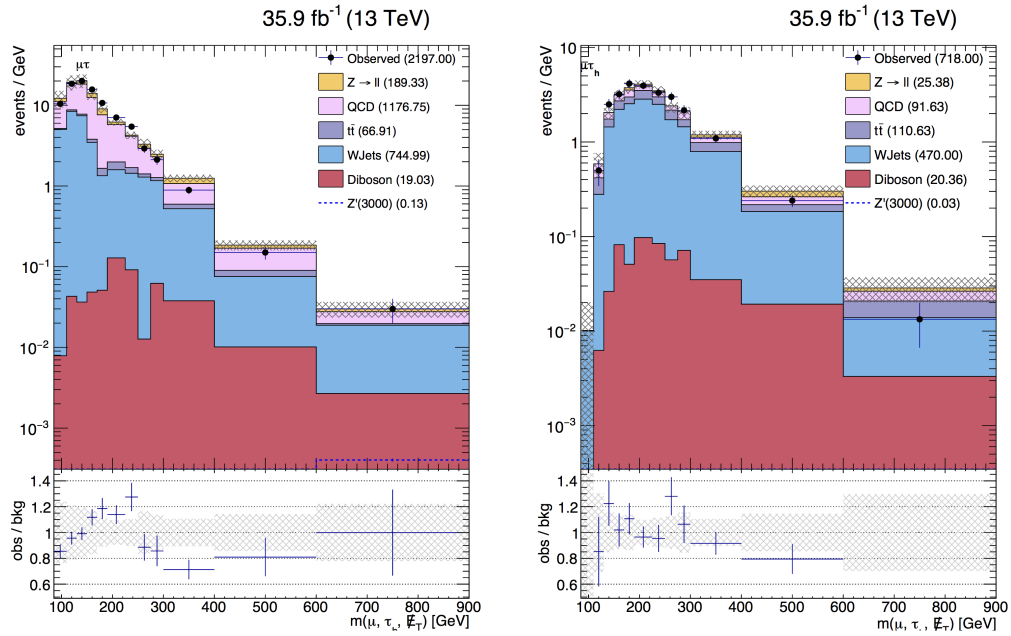

Figure 7.5.6: (Only statistical uncertainties are included) Distributions of $m\left(\mu, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$. Left: in region B with signal region selections (data driven QCD and MC based W+jets). Right: in region A' with signal region like selections (data driven QCD and W+jets).

$$(\text{W+jets distribution})_{\mathrm{A}} = f_{\mathrm{jet}\to\tau} \times (\text{W+jets distribution})_{\mathrm{C}} \tag{7.5.5}$$

where the W+jets distribution in region C is estimated by subtracting non W+jets events with genuine leptons (estimated by MC) and data-driven QCD events bin-by-bin from data. The data-driven QCD events in region C is estimated from the previously mentioned region E with a scale factor ($f_{\mathrm{FtoD}}^{\mathrm{QCD}}$) to properly transfer the yield from region E to C. The data-driven QCD estimation in region C can be summarized as:

$$(\text{QCD distribution})_{\mathrm{C}} = f_{\mathrm{FtoD}}^{\mathrm{QCD}} \times (\text{data distribution})_{\mathrm{E}} \quad , \tag{7.5.6}$$

where $f_{\mathrm{FtoD}}^{\mathrm{QCD}} = \left(N_{\mathrm{data}}^{\mathrm{D}} - N_{\mathrm{non-QCD\ MC}}^{\mathrm{D}}\right) / N_{\mathrm{data}}^{\mathrm{F}}$.

To properly estimate the most relevant (closest to our signal region as possible) "jet to tau fake rate" from data, we construct a W+jets rich region with the following selections:

- pass preselection;
- $\cos\Delta\phi(\mu, \tau_h) < -0.95$;
- $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;

- $0.5 < \cos\Delta\phi(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) < 0.9$ and 55 GeV $< m_{\mathrm{T}}(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) < 120$ GeV;
- no jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose)

Events are further separated in to six regions (A', B', C', D', E' and F') by the same methods as shown in Fig. 7.5.5. The "jet to tau fake rate" is estimated by the following:

$$f_{\mathrm{jet}\to\tau} = \frac{N_{\mathrm{data}}^{\mathrm{A'}} - N_{\mathrm{non\text{-}W\ MC}}^{\mathrm{A'}} - N_{\mathrm{data}}^{\mathrm{E'}} \times f_{\mathrm{LT}}^{\mathrm{QCD}}}{N_{\mathrm{data}}^{\mathrm{C'}} - N_{\mathrm{non\text{-}W\ MC}}^{\mathrm{C'}} - N_{\mathrm{data}}^{\mathrm{E'}} \times f_{\mathrm{FtoD}}^{\mathrm{QCD}}} \quad . \tag{7.5.7}$$

Region A' is shown in the right panel of Fig. 7.5.6 with the "jet to tau fake rate" estimated as: $f_{\mathrm{jet}\to\tau} = 0.25 \pm 0.02$.

7.5.6 Validation of Background Estimations

Additional to the validation tests shown in Fig. 7.5.4, a test to simultaneously check the data driven QCD and W+jets estimations is performed by reverting the $\cos\Delta\phi(\mu, \tau_h)$ cut in the following configuration:

- pass preselection;
- $-0.95 < \cos\Delta\phi(\mu, \tau_h) < 0$;
- $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;
- $\cos\Delta\phi(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ or ($\cos\Delta\phi(\tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ and $m_{\mathrm{T}}(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 150$ GeV);
- no jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose).

Figure 7.5.7 shows the distributions of $m\left(\mu, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$, $p_{\mathrm{T}}^{\mathrm{miss}}$, μ p_T and τ_h p_T in region A with the above set of selections.

Figure 7.5.8 shows the distributions of $m\left(\mu, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ in the signal region with data blinded.

7.6 Electron + Hadronic Tau Channel

7.6.1 Event selection

Selections in the $e\tau_h$ channel closely resemble that of the $\mu\tau_h$ channel with electrons replacing muons. First, events must fire the single-electron trigger described in Section 7.3. Then reconstructed electrons are selected satisfying:

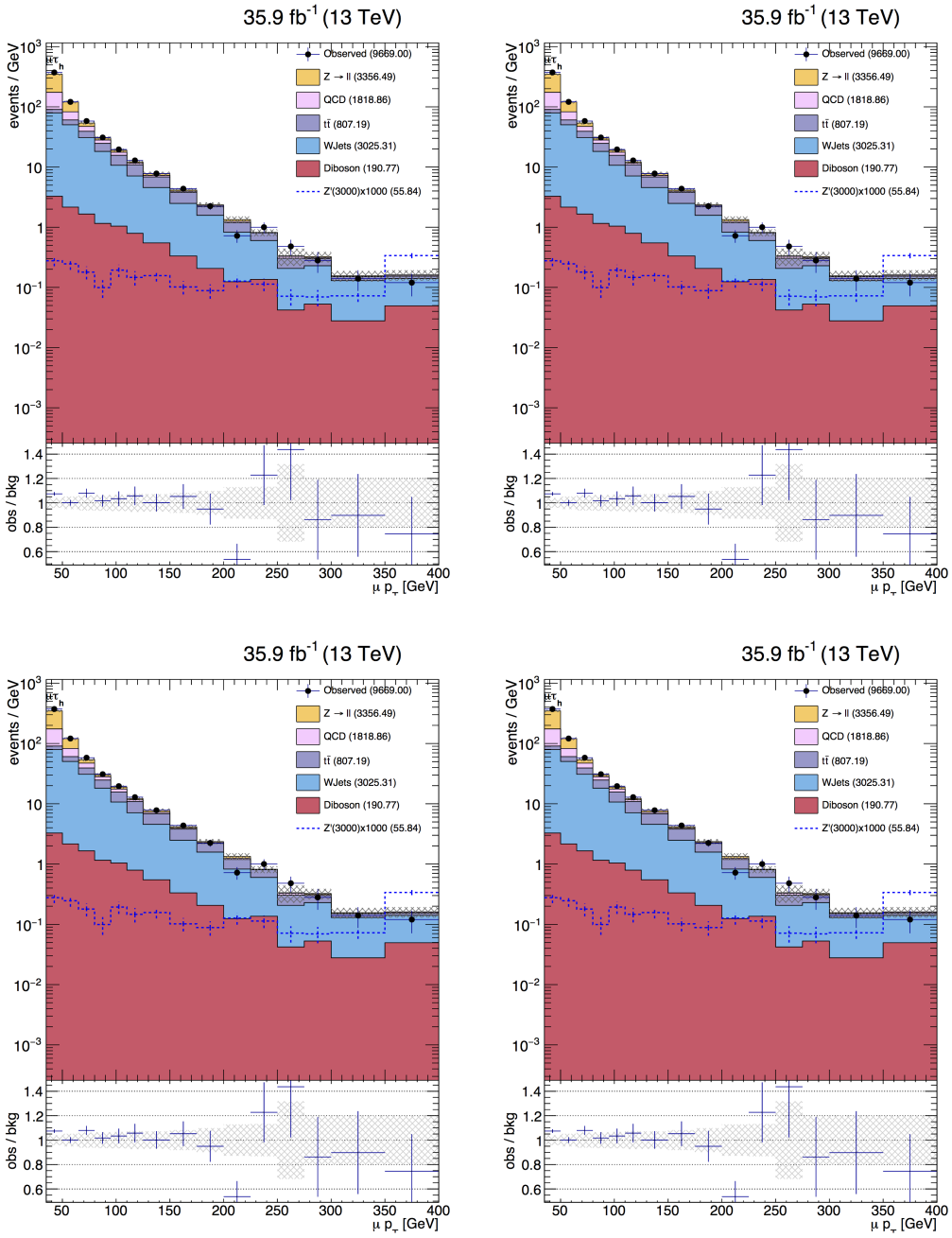

Figure 7.5.7: (Data driven W+jets and QCD with only statistical uncertainties) Distributions with $-0.95 < \cos\Delta\phi(\mu, \tau_h) < 0$ selection. Top: left: $m\left(\mu, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ right: $p_{\mathrm{T}}^{\mathrm{miss}}$. Bottom: left: muon p_T right: τ_h p_T.

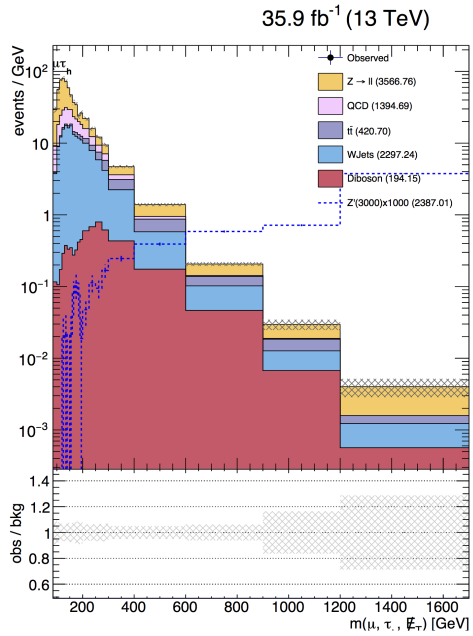

Figure 7.5.8: (Data driven W+jets and QCD with only statistical uncertainties) $m\left(\mu, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ distribution with signal region selections.

- $p_T > 35$ GeV, $|\eta| < 2.1$, $d_{xy} < 0.045$ cm and $d_z < 0.2$ cm;
- the "Loose" electron MVA ID requirements;
- having no matched conversion nor missing hits;
- within $\Delta R < 0.5$ of the HLT electron that fired the trigger.

Offline τ_h requirements are almost the same as in Section 7.6.1 but with a different working point of the lepton veto ID as “Tight” for anti-electron and “Loose” for anti-muon.

Following the same method as in the $\mu\tau_h$ channel, electron and τ_h pairs are constructed by choosing the most isolated pair which the electron and τ_h are separated by at least $\Delta R > 0.5$.

After a pair has been chosen for an event, we apply the following isolation requirements on the leptons, for an event to enter the signal region: electron relative isolation < 0.15; τ_h passing the "Tight" working point of MVA tau isolation ID. In order to keep the different channels exclusive, an event is rejected if there is an additional electron satisfying the above identification requirements with relative isolation < 0.3, or a muon satisfying the identification requirements described in Section 7.7.1 with relative isolation < 0.3. To further reduce backgrounds, we require the electron and τ_h to have opposite electric charge. The selection set mentioned above is defined as the preselection.

Following the preselection, as for the $\mu\tau_h$ channel, the signal region is defined as:

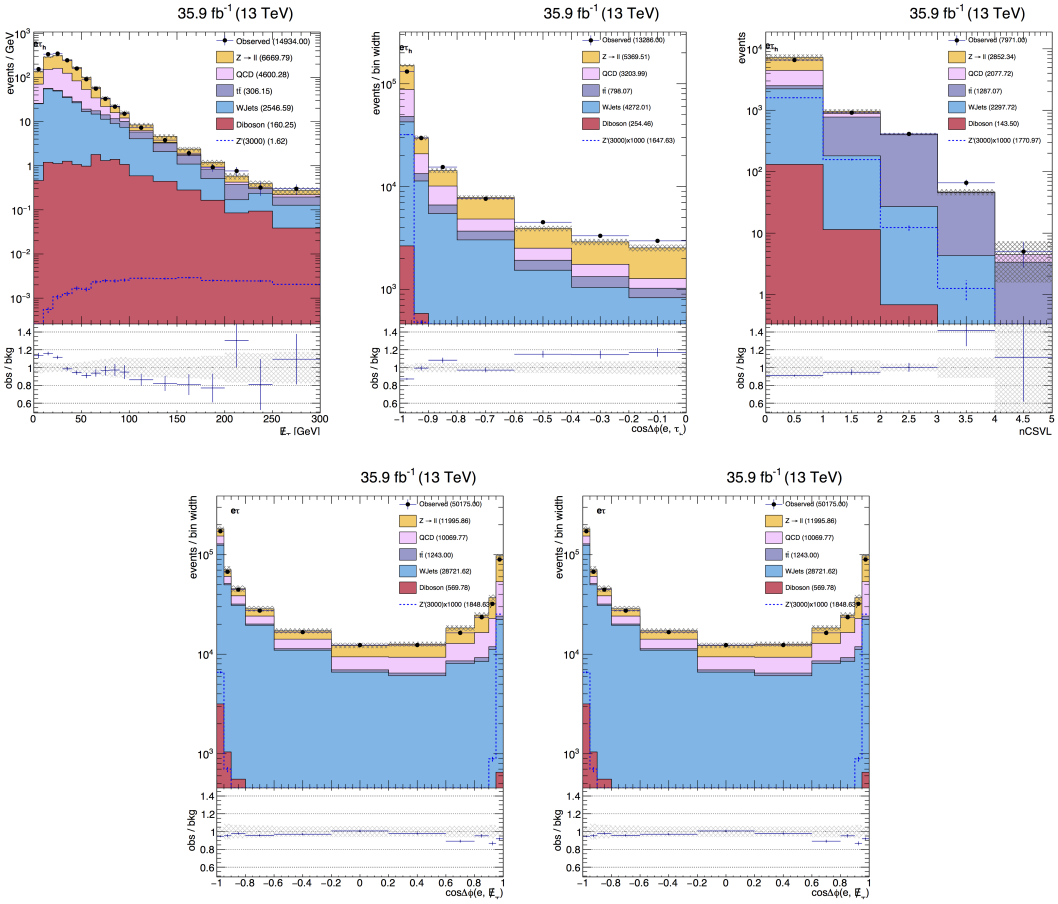

Figure 7.6.1: (Data driven QCD with only statistical uncertainties) Distributions of the variables used for $e\tau_h$ signal selection, after all other signal selection requirements on variables other than the one plotted. Top row plots are made with data driven W+jets in distributions of: $p_{\mathrm{T}}^{\mathrm{miss}}$ (top left), $\cos\Delta\phi(e,\tau_h)$ (top middle, with $\cos\Delta\phi(e,\tau_h) > 0$) and n_b (top right). Bottom row plots are made with MC W+jets in distributions of: $\cos\Delta\phi(e,\vec{p}_{\mathrm{T}}^{\mathrm{miss}})$ (bottom left) and $\cos\Delta\phi(\tau_h,\vec{p}_{\mathrm{T}}^{\mathrm{miss}})$ (bottom right).

- $\cos\Delta\phi(e,\tau_h) < -0.95$;
- $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;
- $\cos\Delta\phi(e,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ or ($\cos\Delta\phi(\tau_h,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ and $m_{\mathrm{T}}(e,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 150$ GeV);
- no jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose).

The distributions of these variables after preselection, and after selection requirements on the other variables, are shown in Fig 7.6.1.

7.6.2 Genuine dilepton events

Same as the $\mu\tau_h$ channel, backgrounds of Drell-Yan process, top quark single and pair production, and di-boson production are estimated by MC simulations and normalized to the best available cross sections, listed in Table 7.4.

7.6.2.1 Drell-Yan process

Due to large W+jets and QCD contamination in the $e\tau_h$ channel, as shown in the left panel of Fig 7.6.2 with the following selections:

- pass preselection;
- $p_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV;
- no jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose);
- $\cos\Delta\phi(e, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ or $\cos\Delta\phi(\tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$;
- 60 GeV $< m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right) <$ 150 GeV,

the Drell-Yan yield systematic uncertainty (5%) is estimated from the $e\mu$ channel in Section 7.7.2.1, for its higher Drell-Yan purity. A cross check of this number is performed in the $e\tau_h$ channel by comparing the Drell-Yan rate between data and MC as the following:

$$\frac{\text{data - non-Drell-Yan backgrounds}}{\text{MC Drell-Yan}} = 0.95 \pm 0.05 \quad , \tag{7.6.1}$$

which agrees with the measurements from the $e\mu$ channel.

7.6.2.2 $t\bar{t}$ and single top processes

The $t\bar{t}$ and single top production rate systematics (1%), with the top pt reweighting as described in Section 7.5.3.2, are estimated from the $e\mu$ channel as described in Section 7.7.2.2, for its higher top purity. A cross check of this number in the $e\tau_h$ channel is performed in a top-rich region, defined by the following selections and shown in the right panel of Fig. 7.6.2:

- pass preselection;
- $\cos\Delta\phi(e, \tau_h) < -0.95$;
- $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;

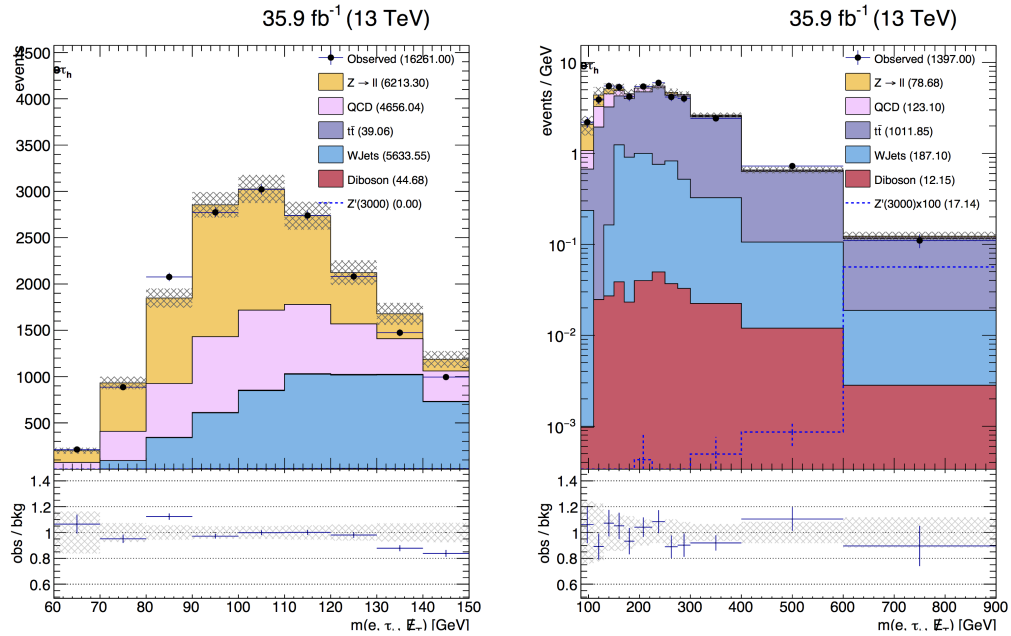

Figure 7.6.2: (Data driven W+jets and QCD with only statistical uncertainties) Distributions of $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$. Left: validation region with $p_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV, $n_b = 0$ and 60 GeV $< m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ < 150 GeV. Right: validation region with $n_b \geq 1$.

- $\cos\Delta\phi(e, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ or $(\cos\Delta\phi(\tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ and $m_{\mathrm{T}}(e, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 150$ GeV);
- at least one jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose).

The $t\bar{t}$ + single top data/MC overall agreement is estimated to be:

$$\frac{\text{data - non } t\bar{t} + \text{single top backgrounds}}{\text{MC } t\bar{t} + \text{single top}} = 0.96 \pm 0.06 \quad , \tag{7.6.2}$$

which agrees with the measurement in the $e\mu$ channel.

7.6.2.3 Di-boson process

Di-boson processes are a relatively small background in the $e\tau_h$ channel. They are estimated directly from simulation with a 20% systematic uncertainty measured from the $e\mu$ channel in Section 7.7.2.3.

7.6.3 QCD background

Similar as the $\mu\tau_h$ channel, we construct a data-driven template for the shape of the QCD background. Based on the electric charge of the reconstructed objects, the electron relative isolation and τ_h isolation, events are split into four regions shown in Fig 7.5.5 and described below:

Table 7.6: Event yields in regions B and F used for the calculation of $f_{\mathrm{LT}}^{\mathrm{QCD}}$.

Process	region B	region F
Z + jets	151 ± 26	0 ± 0
$t\bar{t}$	40 ± 4	2.0 ± 0.8
W + jets	418 ± 94	3 ± 2
DiBoson	15 ± 2	0.1 ± 0.0
non-QCD background	624 ± 98	5 ± 2
Data	2265	1539
Data - non-QCD background	1641 ± 98	1539
$f_{\mathrm{LT}}^{\mathrm{QCD}}$	1.07 ± 0.08	

- A (Signal) Region: e and τ_h have opposite charge, e relative isolation is less than 0.15, and τ_h pass "Tight" isolation requirement.
- B Region: e and τ_h have same charge, e relative isolation is less than 0.15, and τ_h pass "Tight" isolation requirement.
- E Region: e and τ_h have opposite charge, e relative isolation is between 0.15 and 1.0, and τ_h fail "Tight" isolation requirement.
- F Region: e and τ_h have same charge, e relative isolation is between 0.15 and 1.0, and τ_h fail "Tight" isolation requirement.

Table 7.6.3 shows the yields of data and MC samples in regions B and F used for the calculation of $f_{\mathrm{LT}}^{\mathrm{QCD}}$.

To check whether if the QCD shape in the anti-isolated region correctly models the QCD shape in the isolated region, we compare the observation and background estimation in region B with the QCD shape taken from region F and normalized to data - non QCD background in region B. An example of this test is shown in the left panel of Fig. 7.6.3. Here, W+jets events are estimated directly from MC simulations. Overall, the data and estimated background agree reasonably. Due to low statistics of the W+jets MC sample, some bins disagree by more than 20%.

7.6.4 W+jets background

Similar as the $\mu\tau_h$ channel, we construct a data-driven template for the shape of the W+jets background in a control region with high W+jets purity. Swapping muons with electrons, events are split into four regions shown in Fig. 7.5.5 and described below:

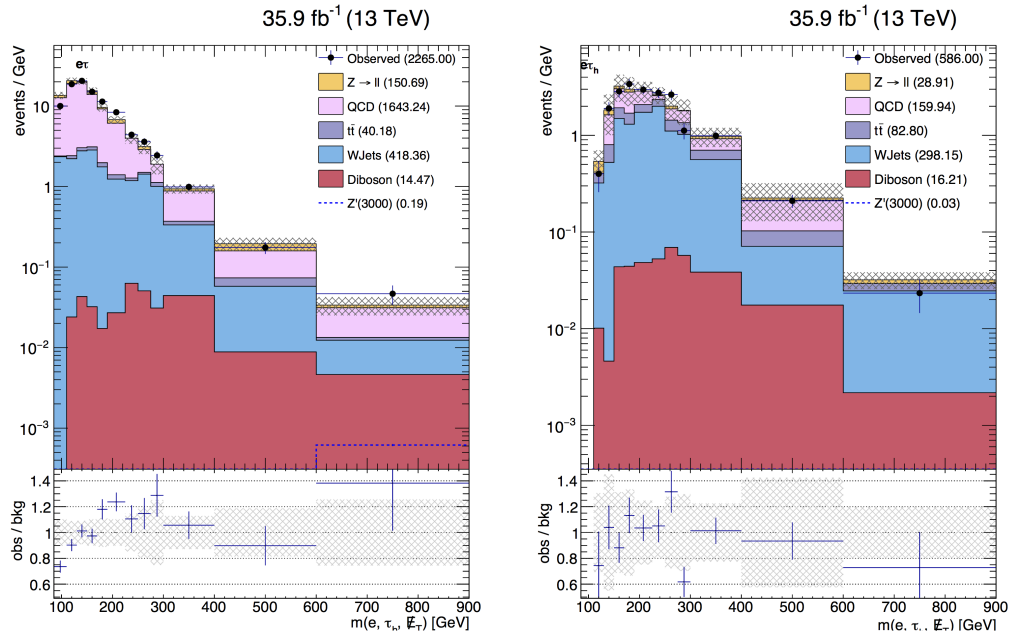

Figure 7.6.3: (Only statistical uncertainties are included) Distributions of $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$. Left: in region B with signal region selections (data driven QCD and MC based W+jets). Right: in region A' with signal region like selections (data driven QCD and W+jets).

- Regions A and B are the same as in Section 7.6.3.
- C Region: e and τ_h have opposite charge, e relative isolation is less than 0.15, and τ_h pass "VeryLoose" isolation requirement but fail "Tight" isolation requirement.
- D Region: e and τ_h have same charge, e relative isolation is less than 0.15, and τ_h pass "VeryLoose" isolation requirement but fail "Tight" isolation requirement.

To properly estimate the most relevant (closest to our signal region as possible) "jet to tau fake rate" from data, we construct a W+jets rich region with the following selections:

- pass preselection;
- $\cos\Delta\phi(e, \tau_h) < -0.95$;
- $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;
- $0.5 < \cos\Delta\phi(e, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) < 0.9$ and 55 GeV $< m_{\mathrm{T}}(e, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) < 120$ GeV;
- no jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose).

The "jet to tau fake rate" in the $e\tau_h$ channel is estimated the same way as in Eq. 7.5.7. Region A' is shown in the right panel of Fig. 7.6.3 with the "jet to tau fake rate" estimated as: $f_{\mathrm{jet}\to\tau} = 0.20 \pm 0.03$.

7.6.5 Validation of Background Estimations

Additional to the validation tests shown in Fig. 7.6.2, a test to simultaneously check the data driven QCD and W+jets estimations is performed by reverting the $\cos\Delta\phi(e,\tau_h)$ cut in the following configuration:

- pass preselection;
- $-0.95 < \cos\Delta\phi(e,\tau_h) < 0$;
- $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;
- $\cos\Delta\phi(e,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ or ($\cos\Delta\phi(\tau_h,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ and $m_{\mathrm{T}}(e,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 150$ GeV);
- no jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose).

Figure 7.6.4 shows the distributions of $m\left(e,\tau_h,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$, $p_{\mathrm{T}}^{\mathrm{miss}}$, electron p_T and τ_h p_T in region A with the above set of selections.

Figure 7.6.5 shows the distributions of $m\left(e,\tau_h,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ in the signal region with data blinded.

7.7 Electron + Muon Channel

7.7.1 Event selection

In the $e\mu$ channel, the selection of electrons is identical to that described in Section 7.6.1. Muons are required to have:

- $p_T > 10$ GeV, $|\eta| < 2.1$, $d_{xy} < 0.045$ cm and $d_z < 0.2$ cm;
- satisfy the medium muon requirement.

We build pairs of electrons and muons in which the electron and muon are separated by at least $\Delta R > 0.5$. In events with more than one such pair, we select the pair with the two most isolated leptons, considering first the muon, and then the electron. This criterion was seen to have good efficiency for signal samples. In the rare case of multiple such pairs having identical isolation values, the reconstructed p_T's are considered, preferring higher values.

After a pair has been chosen for an event, it enters the signal region only when both the electron and muon relative isolations are < 0.15. To reduce Drell-Yan backgrounds, events are rejected if there is an additional electron satisfying the requirements described in Section 7.6.1 with relative isolation < 0.3, or an additional muon satisfying the above identification requirements with relative

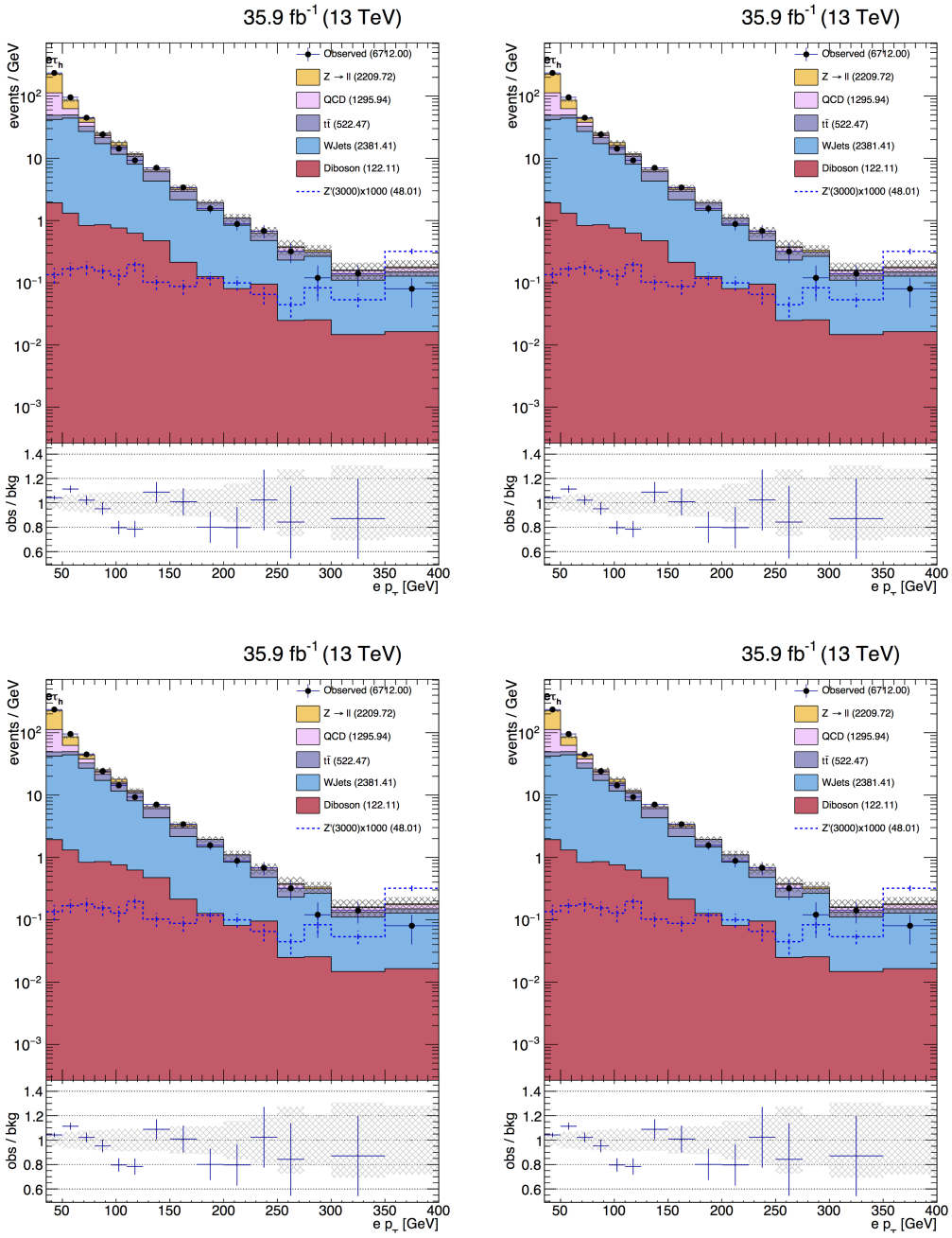

Figure 7.6.4: (Data driven W+jets and QCD with only statistical uncertainties) Distributions with $-0.95 < \cos\Delta\phi(e, \tau_h) < 0$ selection. Top: left: $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ right: $p_{\mathrm{T}}^{\mathrm{miss}}$. Bottom: left: electron p_T right: τ_h p_T.

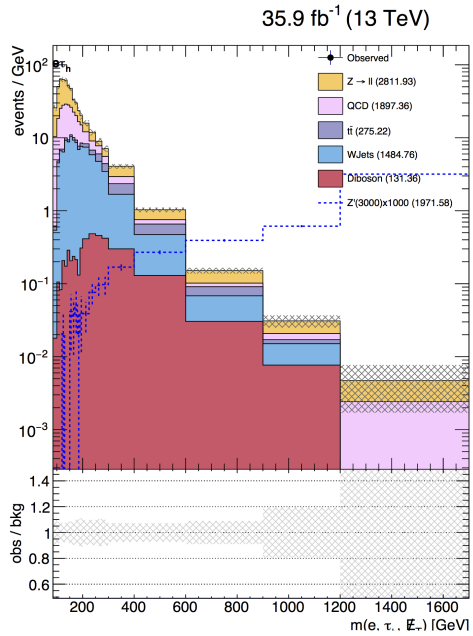

Figure 7.6.5: (Data driven W+jets and QCD with only statistical uncertainties) $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ distribution with signal region selections.

isolation < 0.3. To further reduce backgrounds, we require the electron and muon to have opposite electric charge. The selection set mentioned above is defined as preselection.

Following the preselection, based on the optimization studies performed in the $\tau_h\tau_h$ channel, signal region is defined with the following requirements:

- $\cos\Delta\phi(e, \mu) < -0.95$;
- $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;
- $\cos\Delta\phi(p_T \text{ leading lepton}, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) < -0.9$;
- no jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose).

The distributions of these variables after preselection, and after selection requirements on the other variables, are shown in Fig 7.7.1.

7.7.2 Genuine di-lepton events

Similar as to the $\mu\tau_h$ channel, backgrounds of Drell-Yan process, top quark single and pair production, W+jets, and di-boson production are estimated by MC simulations and normalized to the best available cross sections, listed in Table 7.4.

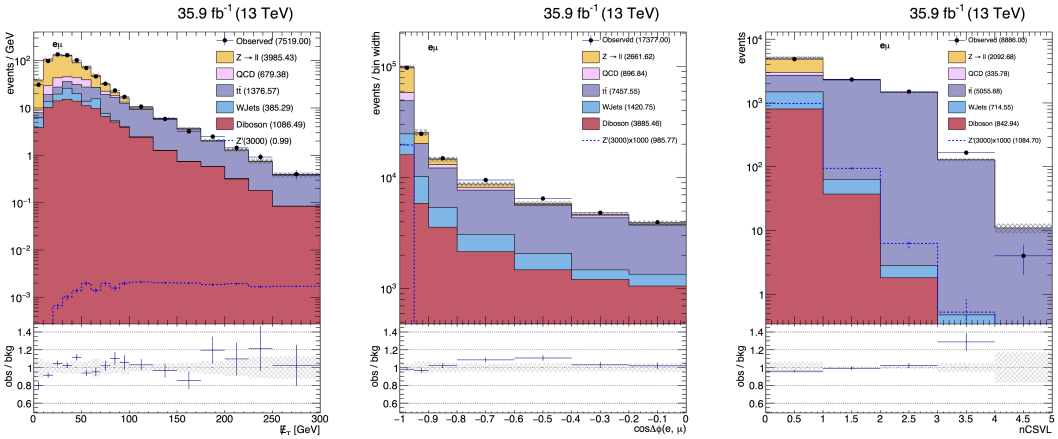

Figure 7.7.1: (Data driven QCD with only statistical uncertainties) Distributions of the variables used for $e\mu$ signal selection, after all other signal selection requirements on variables other than the one plotted: $p_{\mathrm{T}}^{\mathrm{miss}}$ (left), $\cos\Delta\phi(e,\mu)$ (middle, with $\cos\Delta\phi(e,\mu) > 0$) and n_b (right).

7.7.2.1 Drell-Yan process

Systematics for Drell-Yan process are estimated in a Drell-Yan rich region with the following selections and the distributions of data and estimated backgrounds are shown in the left panel of Fig 7.7.2:

- pass preselection;
- $p_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV;
- no jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose);
- 60 GeV $< m\left(e, \mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right) <$ 150 GeV .

The Drell-Yan production rate difference between data and MC is estimated to be:

$$\frac{\text{Data - non-Drell-Yan backgrounds}}{\text{MC Drell-Yan}} = 0.98 \pm 0.05 \quad . \tag{7.7.1}$$

Thus, the Drell-Yan production rate systematic uncertainty is estimated to be 5%, which we apply to all final states.

7.7.2.2 $t\bar{t}$ and single top processes

Systematics for $t\bar{t}$ and single top processes are estimated in a top quark rich region with the following selections and the distributions of data and estimated backgrounds are shown in the right panel of Figure 7.7.2:

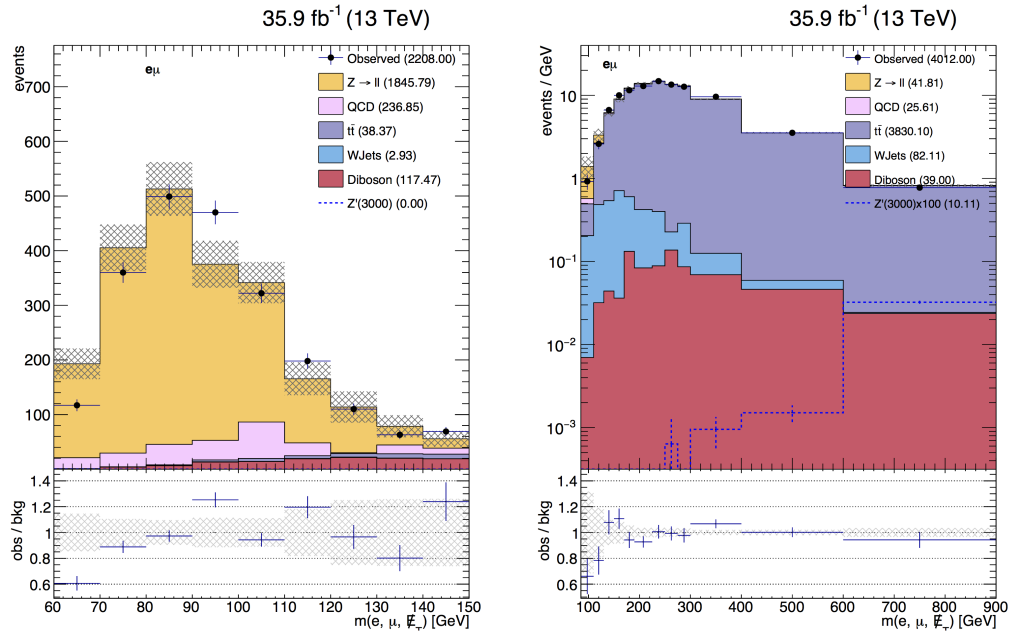

Figure 7.7.2: (Data driven QCD with only statistical uncertainties) Distributions of $m\left(e, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$. Left: validation region with $p_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV, $n_b = 0$ and $60 <$ GeV $m\left(e, \mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right) < 150$ GeV. Right: validation region with $n_b \geq 1$.

- pass preselection;
- $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;
- $\cos \Delta\phi(p_T$ leading lepton$, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) < -0.9$;
- at least one jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose).

As described in Section 7.5.3.2, top pt reweighting is applied to simulated $t\bar{t}$ events. Good agreement between data and background predictions is observed. The $t\bar{t}$ + single top data/MC overall agreement is estimated to be:

$$\frac{\text{Data - non } t\bar{t} + \text{single top backgrounds}}{\text{MC } t\bar{t} + \text{single top}} = 1.00 \pm 0.01 \quad . \tag{7.7.2}$$

Thus, the $t\bar{t}$ and single top production rate systematic uncertainty is estimated to be 1%, which we apply to all final states.

7.7.2.3 Di-boson process

Di-boson processes are estimated using simulated MC samples. In order to quantify the agreement between simulated and observed di-boson events, a di-boson rich region is constructed with the

following selections:

- pass preselections;
- $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;
- no jet with $p_T > 20$ GeV;
- $P_\zeta - 3.1 \times P_\zeta^{vis} < -50$;
- $\cos\Delta\phi(p_T \text{ leading lepton}, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > -0.9$;
- $m\left(e, \mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right) > 200$ GeV.

Here $\hat{\zeta}$ is defined as the unit vector along the bisector of visible tau decay products and two the projection variables are defined as:

$$P_\zeta^{vis} = (\overrightarrow{p}_e^{vis} + \overrightarrow{p}_\mu^{vis}) \cdot \hat{\zeta} \quad . \tag{7.7.3}$$

$$P_\zeta = (\overrightarrow{p}_e^{vis} + \overrightarrow{p}_\mu^{vis} + \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) \cdot \hat{\zeta} \quad . \tag{7.7.4}$$

Additional to the usual electric charge and $p_{\mathrm{T}}^{\mathrm{miss}}$ requirements, events containing jets with $p_T > 20$ GeV and $|\eta| < 2.1$ are rejected to reduce $t\bar{t}$ contamination. This serves as a tighter $t\bar{t}$ reduction as opposed to the b-jet veto. The motivation of $P_\zeta - 3.1 \times P_\zeta^{vis} < -50$ GeV can be seen in the left panel of Fig 7.7.3 which shows the distribution of $P_\zeta - 3.1 \times P_\zeta^{vis}$. By cutting at -50, we remove almost all the signal events while keeping 50% of the di-boson events. The $\cos\Delta\phi(p_T \text{ leading lepton}, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > -0.9$ cut ensures this control region is orthogonal to the signal region. Finally, the $m\left(e, \mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right) > 200$ GeV cut is chosen to reject Drell-Yan, QCD and W+jets events and improve di-boson purity (63%).

The middle panel of Fig 7.7.3 shows the $m\left(e, \mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ distribution after all the requirements mentioned in this section. Overall, we see a reasonable agreement between data and the expected backgrounds. The ratio between data and MC for di-boson processes is estimated to be:

$$\frac{\text{Data - non di-boson backgrounds}}{\text{MC di-boson}} = 1.01 \pm 0.20 \quad . \tag{7.7.5}$$

Taking the shape variation into consideration, we estimate a 20% systematic uncertainty for di-boson processes.

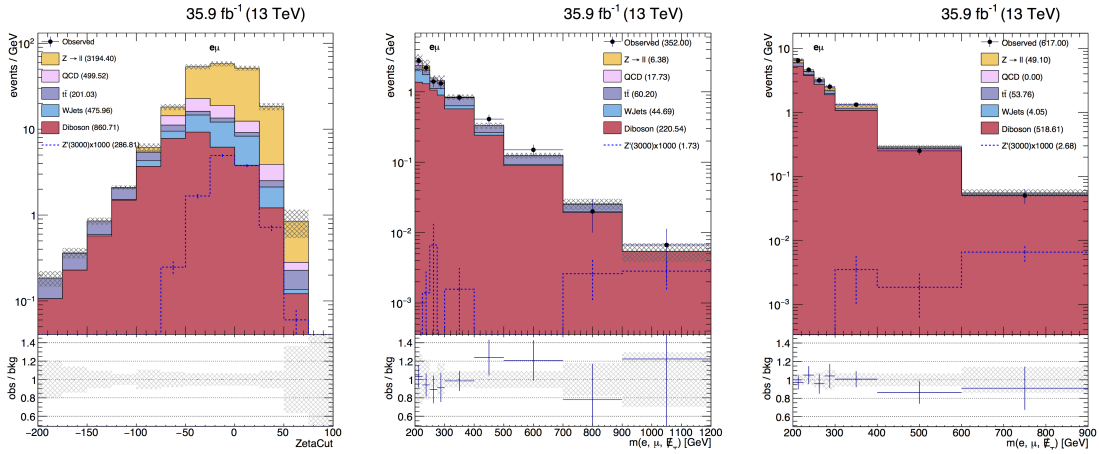

Figure 7.7.3: (Data driven QCD with only statistical uncertainties) Left: $P_\zeta - 3.1 \times P_\zeta^{vis}$ distribution after requiring preselection, $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV and no jet with $p_T > 20$ GeV. Middle: $m\left(e, \mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ distribution after all the requirements mentioned in this section. Right: $m\left(e, \mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ distribution after all the requirements mentioned in this section.

7.7.2.4 Di-boson cross check

To cross check the di-boson study in Section 7.7.2.3, a WZ rich region is constructed with the following requirements:

- pass preselection but with $Q(e) \times Q(\mu) > 0$;
- $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;
- at least three leptons with at least one electron and one muon;
- no jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose);
- $m\left(e, \mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right) > 200$ GeV.

By requesting at least three leptons, the di-boson purity is enhanced to 84%, as Drell-Yan, W+jets and $t\bar{t}$ backgrounds contain at most two leptons. The right panel of Fig. 7.7.3 shows the $m\left(e, \mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ distribution after all the requirements mentioned in this section where we see a good agreement between data and the expected backgrounds. The ratio between data and MC for di-boson processes in the WZ enriched region is estimated to be:

$$\frac{\text{Data - non di-boson backgrounds}}{\text{MC di-boson}} = 0.98 \pm 0.04 \quad . \tag{7.7.6}$$

which agrees well with the results in equation 7.7.5.

Table 7.7: Event yields in isolated and anti-isolated regions used for the calculation of the "Loose-to-Tight" scale factor.

Process	muon relative iso < 0.15	0.15 < muon relative iso < 1.0
Z + jets	27 ± 10	49 ± 14
$t\bar{t}$	26 ± 3	241 ± 8
W + jets	52 ± 39	209 ± 75
DiBoson	41 ± 3	6 ± 1
non-QCD background	146 ± 40	505 ± 77
Data	370	1490
Data - non-QCD background	224 ± 40	985.0 ± 77
"Loose-to-Tight" scale factor	0.23 ± 0.05	

7.7.3 QCD background

The estimation of the QCD background for the $e\mu$ channel is directly analogous to that in the $\mu\tau_h$ and $e\tau_h$ channels, except that the relaxed isolation region is defined by the muon isolation. As shown in the left panel of Fig. 7.7.4, in order to achieve good QCD purity, the relative isolation sideband of 0.15 to 1.0 was chosen. After the signal region selection the "Loose-to-Tight" scale factor is estimated to be: 0.23 ± 0.05 where this 23% rate uncertainty is applied to the QCD process (in addition to the bin-by-bin systematic uncertainties).

Table 7.7.3 shows the yields of data and MC samples in isolated (muon iso < 0.15) and anti-isolated (0.15 < muon iso < 1.0) regions used for the calculation of the "Loose-to-Tight" scale factor.

7.7.4 W+jets background

W+jets background in the $e\mu$ channel is very small. However, the W+jets simulated sample was not generated with high statistics. As a workaround, the W+jets shape is taken from the simulated sample with the muon relative isolation from 0.15 to 1.0, and scaled to match the simulated yield where the muon relative isolation < 0.15. The "loose-to-tight" factor is 0.21 ± 0.07. Thus, we improve the statistics of the W+jets template by $1/0.21 \sim 5$ times. A shape comparison of the $m\left(e, \mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ distribution between the isolated and anti-isolated W+jets MC events is shown in the right panel of Fig. 7.7.4. As shown, the W+jets shape from the relaxed isolation region does a reasonable job at modeling the shape of the isolated region.

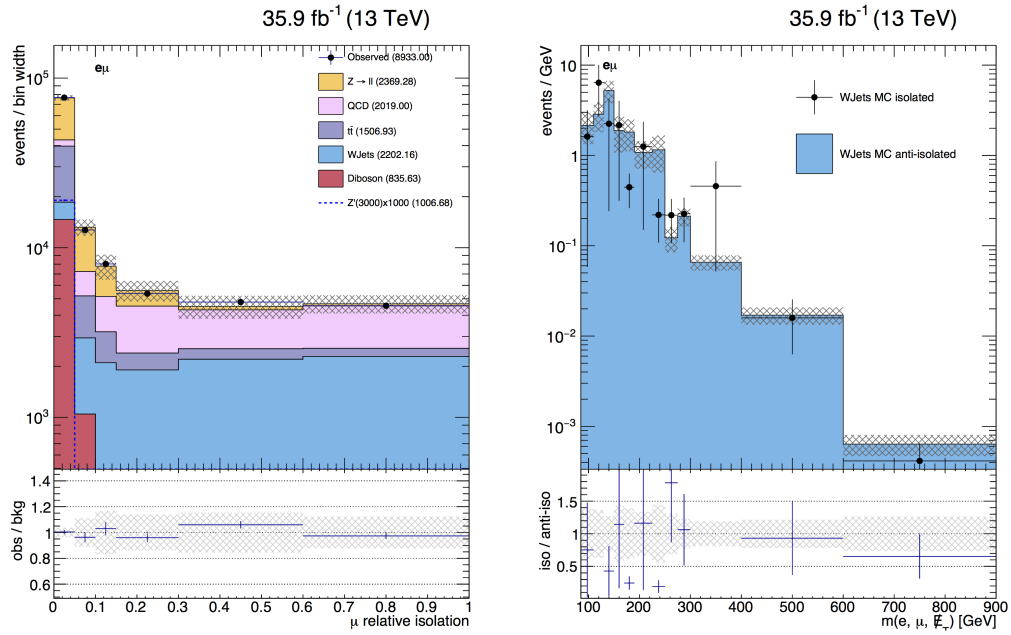

Figure 7.7.4: (Data driven QCD with only statistical uncertainties) Left: muon relative isolation with QCD estimated from same-sign region. Right: W+jets MC closure test between isolated and anti-isolated events.

7.7.5 Validation of Background Estimations

Additional to the validation tests shown in Fig. 7.7.2, a test to check the overall performance of the background estimations is constructed by reverting the $\cos\Delta\phi(e,\mu)$ cut as the following:

- pass preselection;
- $-0.95 < \cos\Delta\phi(e,\mu) < 0$
- $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;
- $\cos\Delta\phi(p_T \text{ leading lepton}, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) < -0.9$;
- no jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose).

Figure 7.7.5 shows the distributions of $m\left(e,\mu,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$, $p_{\mathrm{T}}^{\mathrm{miss}}$, electron p_T and muon p_T in the signal like region with the above set of selections.

Figure 7.7.6 shows the distributions of $m\left(e,\mu,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ in the signal region with data blinded.

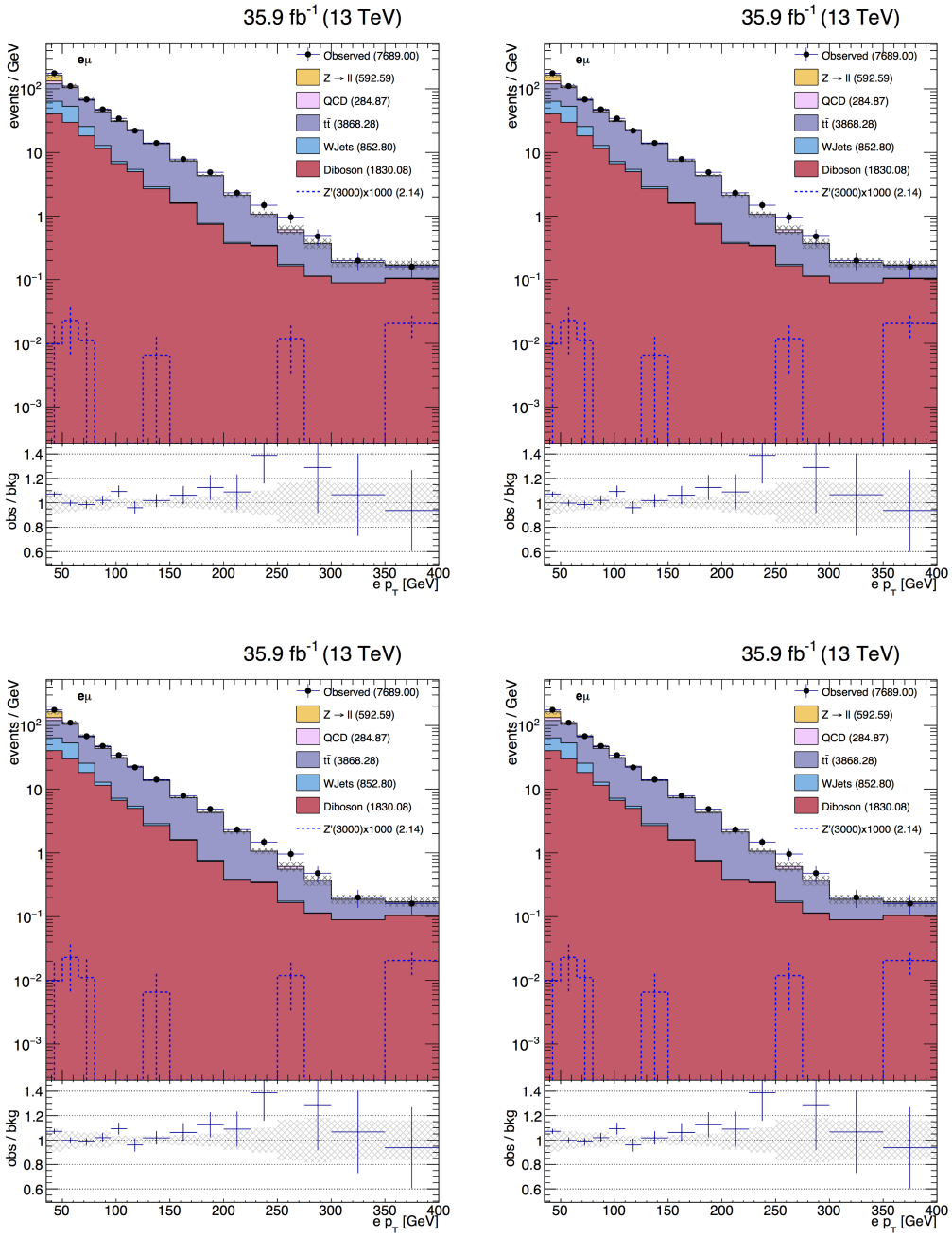

Figure 7.7.5: (Data driven QCD with only statistical uncertainties) Distributions with $-0.95 < \cos\Delta\phi(e,\mu) < 0$ selection. Top: left: $m\left(e,\mu,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ right: $p_{\mathrm{T}}^{\mathrm{miss}}$. Bottom: left: electron p_T right: muon p_T.

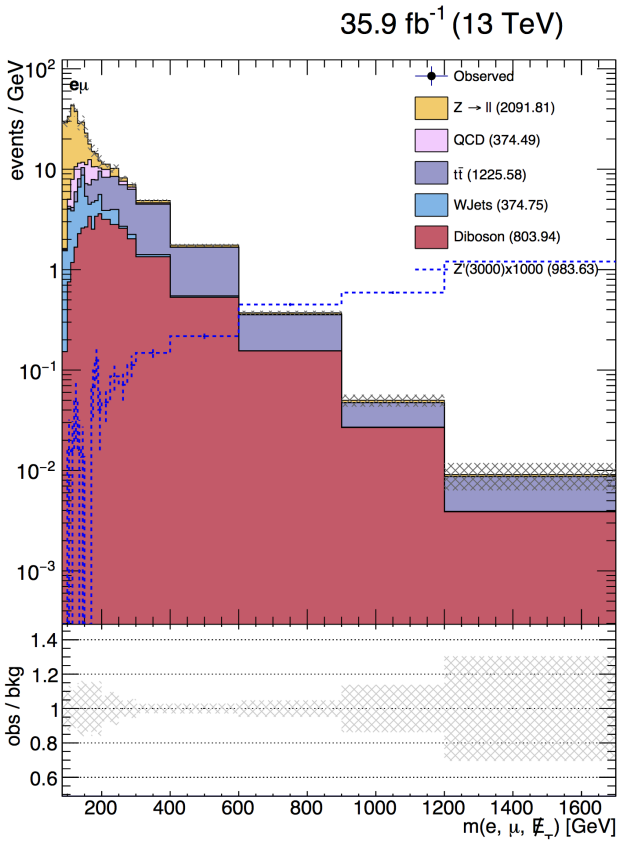

Figure 7.7.6: (Data driven QCD with only statistical uncertainties) $m\left(e, \mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ distribution with signal region selections.

7.8 Systematics

The systematic uncertainties considered in this analysis are detailed below and summarized in Table 7.8):

- **Luminosity:** As suggested by the official CMS recommendation [93], a 2.6% uncertainty on the measured luminosity is considered for all MC based events.

- **Electron Energy Scale:** Due to potential mis-modeling of the electron energy, a systematic variation of 1% (2.5%) shift on the electron energy scale in the barrel (endcap) regions is applied. The resultant systematic uncertainty on MC based events is $< 1\%$.

- **Electron Identification + Trigger:** A 6% uncertainty on the combination of electron reconstruction, identification and the single electron trigger is considered as measured with a data-driven method [94].

- **Muon Momentum Scale:** Due to potential mis-modeling of the muon momentum, a systematic variation of 1% shift on the muon momentum is applied. The resultant systematic uncertainty on MC based events is $< 1\%$.

- **Muon Identification + Trigger:** A 7% uncertainty on the combination of muon reconstruction, identification and the single muon trigger is considered as measured with a data-driven method [94].

- **Tau Energy Scale:** As suggested by the official CMS recommendation, for MC based samples, the tau energy is varied by $\pm 5\%$ and propagated to the final discriminant. In the end, the signal and MC based backgrounds fluctuates by up to $\sim 11\%$.

- **Tau Identification:** As suggested by the official CMS recommendation a 6% uncertainty per τ_h on the τ_h identification is considered. Additionally, due to low statistics of high-p_T τ_h validations, an uncertainty per τ_h of $20\% \cdot (p_T/\ 1\ \text{TeV})$ is considered.

- **Jet Energy Scale:** The effects of the jet energy measurement is estimated by varying the jet energy by $\pm 5\%$ (as suggested by the official CMS recommendation) and propagating it to the final discriminant.

- ***b*-Tagging Efficiency:** Uncertainties of CSV b-tagging is estimated by varying the b-tagged status of each jet based on the b-tagging efficiency scale factors (SF $= \epsilon_{\text{Data}}/\epsilon_{\text{MC}}$, where ϵ_{Data} and ϵ_{MC} are the b-tagging efficiencies in data and MC) provided by the official CMS

recommendation [95]. Details of how the jet's b-tag status is varied can be found at [96], a brief summary is given below. For jets with SF < 1, one would only need to "demote" b-tagged jets to a "not b-tagged" state. The percentage of demoted jets, α, should be such that $\epsilon_{\mathrm{MC}} \cdot \alpha = \epsilon_{\mathrm{Data}}$. Thus, a random number c is generated between 0 and 1 and a b-tagged jet is demoted if $c >$ SF. On the other hand, if SF > 1, not b-tagged jets should be "promoted" as b-tagged. The percentage of promoted jets, β, should be such that $(1-\epsilon_{\mathrm{MC}}) \cdot \beta + \epsilon_{\mathrm{MC}} = \epsilon_{\mathrm{Data}}$. Thus, a random number c between 0 and 1 is generated and a jet is "promoted" to be b-tagged if $c < \frac{1-\mathrm{SF}}{1-1/\epsilon_{\mathrm{MC}}}$.

- **PDF Systematics Uncertainty:** Following the "PDF4LHC recommendations for LHC Run II" [97], the PDF uncertainties are computed from the 68% confidence level with the PDF4LHC15_mc sets. The PDF uncertainties for our main backgrounds, $t\bar{t}$, W+Jets and DY, are much smaller than their bin-by-bin statistical uncertainties thus is neglected. The PDF uncertainties on the signal acceptance is estimated to range from 0.7% for Z' at 500 GeV up to 12% for Z' at 3 TeV [14].

- **Background Estimates:** The uncertainty on the data-driven background estimations are driven by the statistics in data in the various control regions. There is also a mostly negligible contribution from the contamination of other MC based backgrounds in the control regions. In cases where MC based backgrounds must be subtracted off, the statistical uncertainties in the MC backgrounds are propagated throughout the subtraction and used to assign a systematic uncertainty on the background prediction.

7.9 Results and Interpretation

As shown in Fig. 7.9.1 the expected upper limit at 95% confidence level (CL) on the cross-section times branching ratio of SSM $Z' \to \tau\tau$ is set using a modified frequentist approach, the CL_s method [89], with interpretation provided by the Higgs Combined tool [90].

7.10 Summary

An ongoing search for an extra neutral gauge boson, Z', that decays to a pair of τ leptons, detailed in $\mu\tau_h$, $e\tau_h$ and $e\mu$ final states, have been presented. Based on the techniques used in previous Z' searches [14], this new study improves the W+jets background suppression and proposes a new data

Table 7.8: Summary of systematic uncertainties. Values are given in percent. "s" indicates template variations ("shape" uncertainties). "p_T" indicates the τ_h p_T dependent uncertainty of 6% + 20%·(p_T/1 TeV)

Source channels:	QCD $\mu\tau_h$, $e\tau_h$, $e\mu$	W $\mu\tau_h$, $e\tau_h$, $e\mu$	DY $\mu\tau_h$, $e\tau_h$, $e\mu$	$t\bar{t}$ $\mu\tau_h$, $e\tau_h$, $e\mu$	di-boson $\mu\tau_h$, $e\tau_h$, $e\mu$	Signal $\mu\tau_h$, $e\tau_h$, $e\mu$
Lumi	–,–,–	–,–,2.6	2.6,2.6,2.6	2.6,2.6,2.6	2.6,2.6,2.6	2.6,2.6,2.6
μ ID + Trig	–,–,–	–,–,7	7,–,7	7,–,7	7,–,7	7,–,7
e ID + Trig	–,–,–	–,–,6	–,6,6	–,6,6	–,6,6	–,6,6
τ_h ID	–,–,–	–,–,–	p_T,p_T,–	p_T,p_T,–	p_T,p_T,–	p_T,p_T,–
b ID	–,–,–	–,–,s	s,s,s	s,s,s	s,s,s	s,s,s
JES	–,–,–	–,–,s	s,s,s	s,s,s	s,s,s	s,s,s
TES	–,–,–	–,–,–	s,s,s	s,s,s	s,s,s	s,s,s
MMS	–,–,–	–,–,1	1,–,1	1,–,1	1,–,1	1,–,1
EES	–,–,–	–,–,1	–,1,1	–,1,1	–,1,1	–,1,1
top p_T	–,–,–	–,–,–	–,–,–	–,–,s	–,–,–	–,–,–
pdf	–,–,–	–,–,–	–,–,–	–,–,–	–,–,–	(1–12)
bin-by-bin stat.	s,s,s	s,s,s	s,s,s	s,s,s,s	s,s,s	s,s,s
Closure+Norm.	11,7,22	8,15,33	5,5,5	1,1,1,1	20,20,20	–,–,–

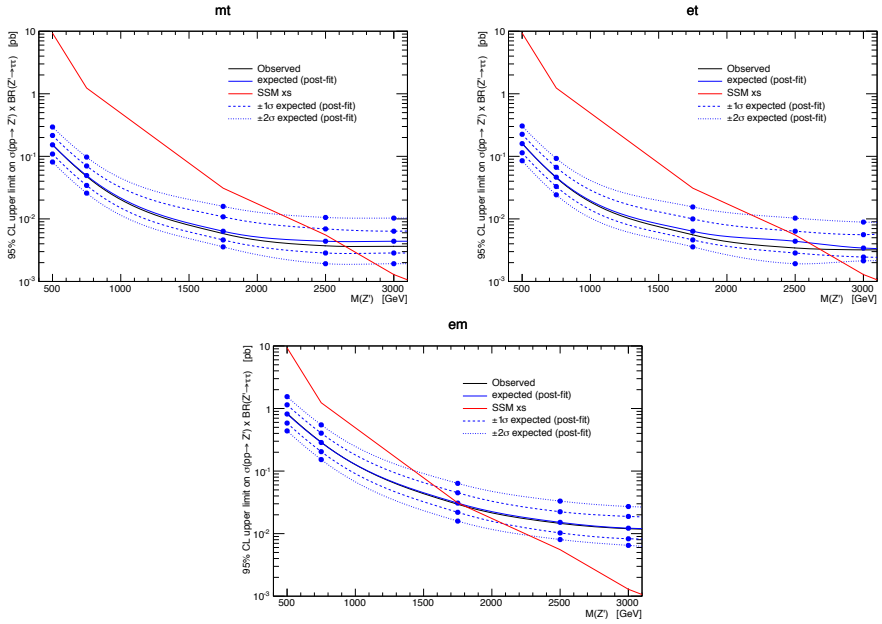

Figure 7.9.1: Expected limits for the $\mu\tau_h$, $e\tau_h$ and $e\mu$ channels. The leading order signal cross-section has been used.

driven technique for W+jets background estimation. This analysis is currently blinded for internal reviews, but early expected limit calculations shows the sensitivity in the $\mu\tau_h$ channel alone already exceeds the combined sensitivity of all four channels from Z' searches using data collected in 2015.

Bibliography

[1] "Standard Model — Wikipedia." [link].

[2] CERN, "The Large Hadron Collider." [link]. [Online; accessed Jan 25, 2017].

[3] C. De Melis, "The CERN accelerator complex. Complexe des accelerateurs du CERN,". [link]. General Photo.

[4] CMS, "CMS Integrated Luminosity." [link]. [Online; accessed Jan 25, 2017].

[5] T. Sakuma and T. McCauley, "Detector and Event Visualization with SketchUp at the CMS Experiment," *J. Phys. Conf. Ser.* **513** (2014) 022032. arXiv:1311.4942 [physics.ins-det].

[6] **CMS** Collaboration, S. Chatrchyan *et al.*, "Aligning the CMS Muon Chambers with the Muon Alignment System during an Extended Cosmic Ray Run," *JINST* **5** (2010) T03019. arXiv:0911.4770 [physics.ins-det].

[7] **CMS** Collaboration, S. Chatrchyan *et al.*, "The CMS experiment at the CERN LHC," *JINST* **3** (2008) S08004. [doi].

[8] T. C. collaboration, "The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC," *Journal of Instrumentation* **8** (2013) no. 11, P11002. [link].

[9] **CMS** Collaboration, "Identification of b quark jets at the CMS Experiment in the LHC Run 2," Tech. Rep. CMS-PAS-BTV-15-001, CERN, Geneva, 2015. [link].

[10] Wikipedia, "A p-n junction in thermal equilibrium with zero-bias voltage applied.." [link].

[11] M. Moll, *Radiation damage in silicon particle detectors: Microscopic defects and macroscopic properties.* PhD thesis, Hamburg U., 1999. [link].

[12] S. Sagir and M. Narain, *Search for Heavy Top Quark Partners with Charge 5/3 and Anomalous Higgs ($\to b\bar{b}$) Couplings to Vector Bosons*. PhD thesis, Brown U., 2016. [link]. Presented 15 Aug 2016.

[13] **CMS** Collaboration, "Searches for a heavy scalar boson H decaying to a pair of 125 GeV Higgs bosons hh or for a heavy pseudoscalar boson A decaying to Zh, in the final states with h $\to \tau\tau$," *Physics Letters B* **755** (2016) 217 – 244. [link].

[14] **CMS** Collaboration, "Search for heavy resonances decaying to tau lepton pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV," *Journal of High Energy Physics* **2017** (2017) no. 2, 48. [link].

[15] **Particle Data Group** Collaboration, C. Patrignani *et al.*, "Review of Particle Physics," *Chin. Phys.* **C40** (2016) no. 10, 100001. [doi].

[16] **CMS** Collaboration, "Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS," *Journal of Instrumentation* **11** (2016) no. 01, P01019. [link].

[17] J. J. Thomson, "Cathode Rays," *Philosophical Magazine* (1897) no. 44, 293.

[18] L. B. Okun, "The theory of weak interaction," in *1962 International Conference on High-Energy Physics at CERN*, J. Prentki, ed., p. 845. 1962.

[19] D. Griffiths, *Introduction to Elementary Particles*. Physics textbook. Wiley, 2008.

[20] O. W. Greenberg, "Spin and Unitary-Spin Independence in a Paraquark Model of Baryons and Mesons,"*Phys. Rev. Lett.* **13** (Nov, 1964) 598–602. [link].

[21] **LHCb Collaboration** Collaboration, R. Aaij *et al.*, "Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays,"*Phys. Rev. Lett.* **115** (Aug, 2015) 072001. [link].

[22] **Super-Kamiokande Collaboration** Collaboration, Y. Fukuda *et al.*, "Evidence for Oscillation of Atmospheric Neutrinos,"*Phys. Rev. Lett.* **81** (Aug, 1998) 1562–1567. [link].

[23] C. Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Physics Letters B* **716** (2012) no. 1, 30 – 61. [link].

[24] A. Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Physics Letters B* **716** (2012) no. 1, 1 – 29. [link].

[25] C. G. Tully, *Elementary Particle Physics in a Nutshell.* Princeton University Press, 2011.

[26] C. N. Yang and R. L. Mills, "Conservation of Isotopic Spin and Isotopic Gauge Invariance,"*Phys. Rev.* **96** (Oct, 1954) 191–195. [link].

[27] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons,"*Phys. Rev. Lett.* **13** (Aug, 1964) 321–323. [link].

[28] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons,"*Phys. Rev. Lett.* **13** (Oct, 1964) 508–509. [link].

[29] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global Conservation Laws and Massless Particles,"*Phys. Rev. Lett.* **13** (Nov, 1964) 585–587. [link].

[30] S. L. Glashow, "Partial Symmetries of Weak Interactions," *Nucl. Phys.* **22** (1961) 579–588. [doi].

[31] T. W. B. Kibble, "History of electroweak symmetry breaking," *J. Phys. Conf. Ser.* **626** (2015) no. 1, 012001. arXiv:1502.06276 [physics.hist-ph].

[32] P. Langacker, "Introduction to the Standard Model and Electroweak Physics," in *Proceedings of Theoretical Advanced Study Institute in Elementary Particle Physics on The dawn of the LHC era (TASI 2008): Boulder, USA, June 2-27, 2008*, pp. 3–48. 2010. arXiv:0901.0241 [hep-ph].

[33] **ATLAS Collaboration and CMS Collaboration** Collaboration, "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** (May, 2015) 191803. [link].

[34] F. Jegerlehner, "The hierarchy problem of the electroweak Standard Model revisited,". arXiv:1305.6652 [hep-ph].

[35] Yu. A. Golfand and E. P. Likhtman, "Extension of the Algebra of Poincare Group Generators and Violation of p Invariance," *JETP Lett.* **13** (1971) 323–326. [Pisma Zh. Eksp. Teor. Fiz.13,452(1971)].

[36] J. Wess and B. Zumino, "Supergauge transformations in four dimensions," *Nuclear Physics B* **70** (1974) no. 1, 39 – 50. [link].

[37] P. Fayet, "Supergauge invariant extension of the Higgs mechanism and a model for the electron and its neutrino," *Nuclear Physics B* **90** (1975) 104 – 124. [link].

[38] P. Fayet, "Spontaneously broken supersymmetric theories of weak, electromagnetic and strong interactions," *Physics Letters B* **69** (1977) no. 4, 489 – 494. [link].

[39] N. Christensen, T. Han, and S. Su, "MSSM Higgs bosons at the LHC," *Phys. Rev. D* **85** (Jun, 2012) 115018. [link].

[40] M. Schmaltz, "Physics beyond the standard model (Theory): Introducing the Little Higgs," *Nuclear Physics B - Proceedings Supplements* **117** (2003) 40 – 49. [doi].

[41] A. Maiezza, M. Nemevšek, F. Nesti, and G. Senjanović, "Left-right symmetry at LHC," *Phys. Rev. D* **82** (Sep, 2010) 055022. [link].

[42] D. London and J. L. Rosner, "Extra Gauge Bosons in E(6)," *Phys. Rev.* **D34** (1986) 1530. [doi].

[43] G. Altarelli, B. Mele, and M. Ruiz-Altaba, "Searching for New Heavy Vector Bosons in $p\bar{p}$ Colliders," *Z. Phys.* **C45** (1989) 109. [doi]. [Erratum: Z. Phys.C47,676(1990)].

[44] "The Large Hadron Collider,". [link].

[45] **CMS** Collaboration, "Particle Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET," CMS Note 2009/039, 2009. [link].

[46] **CMS** Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker," *Journal of Instrumentation* **9** (2014) no. 10, P10009. [link].

[47] **CMS** Collaboration, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV," *JINST* **10** (2015) P06005. arXiv:1502.02701.

[48] **CMS** Collaboration, "Multivariate Electron Identification for Run2." [link].

[49] **CMS** Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt(s) = 7$ TeV," *JINST* **7** (2012) P10002. [doi].

[50] **CMS** Collaboration, "Baseline muon selections for Run-II." [link].

[51] M. Schröder and the CMS collaboration, "Performance of jets at CMS," *Journal of Physics: Conference Series* **587** (2015) no. 1, 012004. [link].

[52] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *Journal of High Energy Physics* **2008** (2008) no. 04, 063. [link].

[53] **CMS** Collaboration, "Jet Identification for the 13 TeV data Run2016." [link].

[54] **CMS** Collaboration, "Identification of b-quark jets with the CMS experiment," *JINST* **8** (2013) 04013.

[55] **CMS** Collaboration, "Performance of τ-lepton reconstruction and identification in CMS," *JINST* **7** (2012) no. 01, P01001. [link].

[56] **CMS** Collaboration, "Performance of tau identification with 2016 data at $\sqrt{s} = 13$ TeV," CMS Detector Performance Summaries CMS-DP-2017-006, 2017. [link].

[57] T. C. Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker," *Journal of Instrumentation* **9** (2014) no. 10, P10009. [link].

[58] "Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV," *Journal of Instrumentation* **10** (2015) no. 08, P08010. [link].

[59] **CMS** Collaboration, "Missing transverse energy performance of the CMS detector," *JINST* **6** (2011) 09001.

[60] **CMS** Collaboration, CMS Collaboration, "Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector," CMS Physics Analysis Summary CMS-PAS-JME-16-004, 2016. [link].

[61] **CMS** Collaboration, "MET Filter Recommendations for Run II." [link].

[62] **The project is partially supported by the EC as FP7 HiLumi LHC Design Study under grant no. 284404 Collaboration** Collaboration, L. Rossi and O. Brüning, "High Luminosity Large Hadron Collider A description for the European Strategy Preparatory Group," Tech. Rep. CERN-ATS-2012-236, CERN, Geneva, Aug, 2012. [link].

[63] C. García-Argos, C. Lacasta Llácer, and S. J. McMahon, *A Silicon Strip Detector for the Phase II High Luminosity Upgrade of the ATLAS Detector at the Large Hadron Collider.* PhD thesis, Valencia U., Nov, 2015. [link]. Presented 08 Jan 2015.

[64] J. Metcalfe, M. Hoeferkamp, and S. Seidel, "Annealing effects on depletion voltage and capacitance of Float Zone and Magnetic Czochralski silicon diodes after 800 MeV proton exposure," in *IEEE Nuclear Science Symposuim Medical Imaging Conference*, pp. 608–611. Oct, 2010. [doi].

[65] G. Lindstrom, "Radiation damage in silicon detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **512** (2003) no. 1-2, 30 – 43.

[66] H. E. Haber and R. Hempfling, "Can the mass of the lightest Higgs boson of the minimal supersymmetric model be larger than m_Z?,"*Phys. Rev. Lett.* **66** (Apr, 1991) 1815–1818. [link].

[67] M. Carena, J. Espinosa, M. Quirós, and C. Wagner, "Analytical expressions for radiatively corrected Higgs masses and couplings in the MSSM," *Physics Letters B* **355** (1995) no. 1, 209 – 221. [link].

[68] A. Djouadi and J. Quevillon, "The MSSM Higgs sector at a high M_{SUSY}: reopening the low tan β regime and heavy Higgs searches," *Journal of High Energy Physics* **2013** (2013) no. 10, 28. [link].

[69] **ATLAS** Collaboration, "Search for Higgs Boson Pair Production in the $\gamma\gamma b\bar{b}$ Final State Using pp Collision Data at $\sqrt{s} = 8$ TeV from the ATLAS Detector,"*Phys. Rev. Lett.* **114** (Feb, 2015) 081802. [link].

[70] **ATLAS** Collaboration, "Search for Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state from pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector," *The European Physical Journal C* **75** (2015) no. 9, 412. [link].

[71] **ATLAS** Collaboration, "Searches for Higgs boson pair production in the $hh \to bb\tau\tau$, $\gamma\gamma WW^*$, $\gamma\gamma bb$, $bbbb$ channels with the ATLAS detector,"*Phys. Rev. D* **92** (Nov, 2015) 092004. [link].

[72] **CMS** Collaboration, "2HDM scenario, H to hh and A to Zh," Tech. Rep. CMS-PAS-HIG-13-025, CERN, Geneva, 2013. [link].

[73] **CMS** Collaboration, "Search for resonant HH production in 2gamma+2b channel," Tech. Rep. CMS-PAS-HIG-13-032, CERN, Geneva, 2014. [link].

[74] **CMS** Collaboration, "Search for di-Higgs resonances decaying to 4 bottom quarks," Tech. Rep. CMS-PAS-HIG-14-013, CERN, Geneva, 2014. [link].

[75] **CMS** Collaboration, "Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons," *Journal of High Energy Physics* **2014** (2014) no. 5, 104. [link].

[76] **CMS** Collaboration, "Search for neutral MSSM Higgs bosons decaying to a pair of tau leptons in pp collisions," *Journal of High Energy Physics* **2014** (2014) no. 10, 160. [link].

[77] T. Sjöstrand, L. Lönnblad, and S. Mrenna, "PYTHIA 6.2 Physics and Manual," Tech. Rep. hep-ph/0108264. LUTP-2001-21, Lunds Univ. Dept. Theor. Phys., Lund, Aug, 2001. [link].

[78] F. Maltoni and T. Stelzer, "MadEvent: automatic event generation with MadGraph," *Journal of High Energy Physics* **2003** (2003) no. 02, 027. [link].

[79] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method," *Journal of High Energy Physics* **2007** (2007) no. 11, 070. [link].

[80] S. Jadach, Z. Was, R. Decker, and J. H. Kuhn, "The tau decay library TAUOLA: Version 2.4," *Comput. Phys. Commun.* **76** (1993) 361–380. [doi].

[81] **CMS** Collaboration, "Combination of ATLAS and CMS top quark pair cross section measurements in the emu final state using proton-proton collisions at 8 TeV,".

[82] **CMS** Collaboration, "Modelling of $\tau\tau$ final states by embedding τ pairs in $Z \to \mu\mu$ events," CMS Note 2011/020, 2011. [link].

[83] Z. Czyczula, T. Przedzinski, and Z. Was, "TauSpinner program for studies on spin effect in tau production at the LHC," *The European Physical Journal C* **72** (2012) no. 4, 1988. [link].

[84] S. A. et al., "Geant4-a simulation toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003) no. 3, 250–303. [link].

[85] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, "TMVA: Toolkit for Multivariate Data Analysis," *PoS* **ACAT** (2007) 040. arXiv:physics/0703039.

[86] **CMS** Collaboration, "Tau ID recommendation." [link].

[87] **CMS** Collaboration, "Recommendation for b/c-tagging and mistagging Data/MC Scale Factors." [link].

[88] **CMS** Collaboration, "Jet energy correction recommendations." [link].

[89] A. L. Read, "Presentation of search results: the CL_s technique," *Journal of Physics G: Nuclear and Particle Physics* **28** (2002) no. 10, 2693. [link].

[90] **CMS** Collaboration, "Documentation of the RooStats based statistics tools for Higgs PAG." [link].

[91] **CMS** Collaboration, "Search for narrow resonances in dilepton mass spectra in proton-proton collisions at $\sqrt{s} = 13$ TeV and combination with 8 TeV data," *Physics Letters B* **768** (2017) 57 – 80. [link].

[92] **ATLAS** Collaboration, "A search for high-mass resonances decaying to τ^{τ}_{-} in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector," *Journal of High Energy Physics* **2015** (2015) no. 7, 157. [link].

[93] **CMS** Collaboration, "Current recommendations for luminosity estimations." [link].

[94] **CMS** Collaboration, "Search for a Narrow Resonance Produced in 13 TeV pp Collisions Decaying to Electron Pair or Muon Pair Final States," CMS Physics Analysis Summary CMS-PAS-EXO-15-005, 2015. [link].

[95] **CMS** Collaboration, "Usage of b/c Tag Objects for 13 TeV Data in 2016 and 80X MC." [link].

[96] **CMS** Collaboration, "Jet-by-jet updating of the b-tagging status." [link].

[97] J. Butterworth, S. Carrazza, A. Cooper-Sarkar, A. D. Roeck, J. Feltesse, S. Forte, J. Gao, S. Glazov, J. Huston, Z. Kassabov, R. McNulty, A. Morsch, P. Nadolsky, V. Radescu, J. Rojo, and R. Thorne, "PDF4LHC recommendations for LHC Run II," *Journal of Physics G: Nuclear and Particle Physics* **43** (2016) no. 2, 023001. [link].

[98] C. T. Hill, "Topcolor assisted technicolor," *Physics Letters B* **345** (1995) no. 4, 483 – 489. [link].

[99] K. Lane, "A new model of topcolor-assisted technicolor," *Physics Letters B* **433** (1998) no. 1-2, 96 – 101. [link].

[100] **CMS** Collaboration, "pt(top-quark) based reweighting of ttbar MC." [link].

Appendix A

Selection Optimization

Based on the set of signal region selections used in the $Z' \to \tau\tau$ search performed with data collected in 2015 [14]:

- $\cos\Delta\phi(e, \tau_h) < -0.95$;
- $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV;
- no jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose);
- $P_\zeta - 3.1 \times P_\zeta^{vis} > -50$ GeV,

optimizations were performed in attempts to improve the separation of signal and background. Here $\hat{\zeta}$ is defined as the unit vector along the bisector of visible tau decay products and two the projection variables are defined as:

$$P_\zeta^{vis} = (\overrightarrow{p}_{\tau_1}^{vis} + \overrightarrow{p}_{\tau_2}^{vis}) \cdot \hat{\zeta} \quad . \tag{A.0.1}$$

$$P_\zeta = (\overrightarrow{p}_{\tau_1}^{vis} + \overrightarrow{p}_{\tau_2}^{vis} + \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) \cdot \hat{\zeta} \quad . \tag{A.0.2}$$

With Z' excluded up to 3 TeV, the τ's from hypothetical Z' decays are most likely to be highly boosted and traveling in opposite directions. Neutrinos from these highly boosted τ decays will also be energetic thus resulting a non-trivial $p_{\mathrm{T}}^{\mathrm{miss}}$. In order to reject $t\bar{t}$ and single top events, the b-jet veto requirement is also crucial. Thus, keeping three of the four signal regions selections untouched, we focused on improving the remaining selection, $P_\zeta - 3.1 \times P_\zeta^{vis} > -50$ GeV.

Neutrinos from highly boosted τ's are likely to be co-linear with the visible products of the τ decay. For channels where one τ decays leptonically and the other τ decays hadronically, the vectorial sum of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is more likely to be aligned with the reconstructed lepton since there are two neutrinos

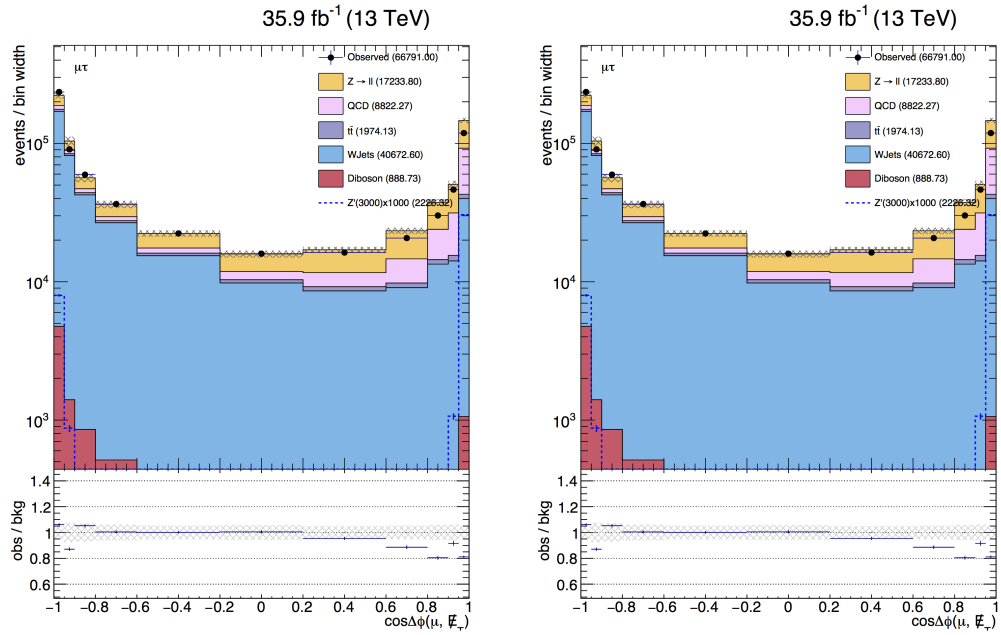

Figure A.0.1: (Data driven QCD and MC based W+Jets with only statistical uncertainties) Left: Distribution of $\cos\Delta\phi(\mu, p_{\mathrm{T}}^{\mathrm{miss}})$. Right: Distribution of $\cos\Delta\phi(\tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$

in this direction. As an example, figure A.0.1 shows, with the $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV, $\cos\Delta\phi(\mu, \tau_h) < -0.95$ and 0 b-jet requirements, the separation in ϕ between the muon and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ in the left panel and τ_h and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ in the right panel.

As expected, most of the signal events have their $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ aligned with the direction of the reconstructed lepton. A requirement of $\cos\Delta\phi(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ would be sufficient in separating most of the background from signal. However, for a small fraction of the signal events, the $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is aligned with the direction of the reconstructed τ_h. In order to recover these signal events, one would need to also accept events with $\cos\Delta\phi(\tau_h, p_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ at the cost of including many W+jets background events. One potential workaround, as demonstrated in Section 7.5.2, is to apply an addition $m_{\mathrm{T}}(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) = \sqrt{2p_T^{\mu}p_{\mathrm{T}}^{\mathrm{miss}}(1-\cos\theta(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}))} > 150$ GeV requirements for events with $\cos\Delta\phi(\tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$. This additional m_{T} requirement guaranties a good signal acceptance while rejects most of the W+jets background.

Using the correlation between $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ and the reconstructed objects, $\cos\Delta\phi(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$ and $\cos\Delta\phi(\tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$, the following selections in Table A.1 were constructed and compared.

Fig A.0.2 shows the background and signal $m\left(\mu, \tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$ distributions after requiring $\cos\Delta\phi(e, \tau_h) < -0.95$, $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV, no jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose) and the set of considered selections mentioned in Table A.1.

Table A.1: selections considered

Label in plot	Color in plot	Selections
pZeta cut	Gray	$P_\zeta - 3.1 \times P_\zeta^{vis} > -50$
orCosDPhi_Met_lep cut	Red	$\cos\Delta\phi(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ or $\cos\Delta\phi(\tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$
orCosDPhi_Met_delta cut	Green	$\cos\Delta\phi(\mu, (\vec{p}_{\mathrm{T}}^{\mu} + \vec{p}_{\mathrm{T}}^{\tau_h})) > 0.9$ or $\cos\Delta\phi(\tau_h, (\vec{p}_{\mathrm{T}}^{\mu} + \vec{p}_{\mathrm{T}}^{\tau_h})) > 0.9$
cosDPhi_Met_r cut	Blue	$-1.05 < \cos\Delta\phi(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) / \cos\Delta\phi(\tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) < -0.95$
orCosDPhi_Met_lowerPt cut	Yellow	$\cos\Delta\phi(\text{sub-leading } p_{\mathrm{T}} \text{ leg}, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$
orCosDPhi_Met_withMt cut	Purple	$\cos\Delta\phi(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ or $\cos\Delta\phi(\tau_h, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 0.9$ and $m_{\mathrm{T}}(\mu, \vec{p}_{\mathrm{T}}^{\mathrm{miss}}) > 150$

Overall, the "orCosDPhi_Met_withMt" cut (colored purple) out performs the other selections in rejecting background while maintaining a relatively high signal acceptance efficiency. Thus, this cut is chosen for the $l + \tau_h$ channels.

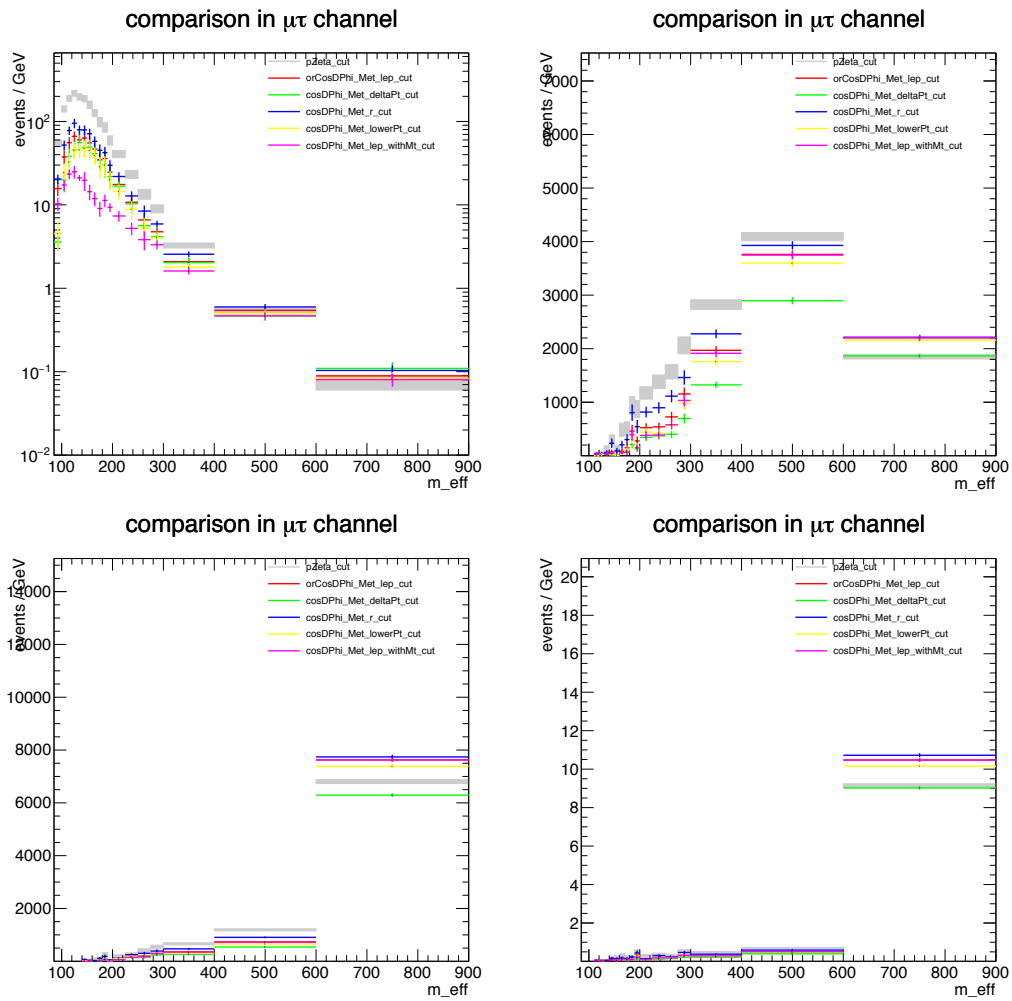

Figure A.0.2: Total background and signal distributions with different selections after requiring: $\cos\Delta\phi(e,\tau_h) < -0.95$; $p_{\mathrm{T}}^{\mathrm{miss}} > 30$ GeV; and no jet with $p_T > 30$ GeV tagged as a b-jet (CSV loose). Top left: The total background distribution of $m\left(\mu,\tau_h,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$. Top right: $Z'(750)$ distribution of $m\left(\mu,\tau_h,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$. Bottom left: $Z'(1750)$ distribution of $m\left(\mu,\tau_h,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$. Bottom right: $Z'(3000)$ distribution of $m\left(\mu,\tau_h,\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right)$