This Dissertation

entitled

Search for lepton flavor violating decays of the Higgs boson

typeset with NDdiss2$_\varepsilon$ v3.0 (2005/07/27) on March 9, 2018 for

Fanbo Meng

This LaTeX 2$_\varepsilon$ classfile conforms to the University of Notre Dame style guidelines established in Spring 2004. However it is still possible to generate a non-conformant document if the instructions in the class file documentation are not followed!

> Be sure to refer to the published Graduate School guidelines at `http://graduateschool.nd.edu` as well. Those guidelines override everything mentioned about formatting in the documentation for this NDdiss2$_\varepsilon$ class file.

It is YOUR responsibility to ensure that the Chapter titles and Table caption titles are put in CAPS LETTERS. This classfile does *NOT* do that!

*This page can be disabled by specifying the "*`noinfo`*" option to the class invocation.* (i.e.,`\documentclass[...,noinfo]{nddiss2e}` )

**This page is *NOT* part of the dissertation/thesis, but MUST be turned in to the proofreader(s) or the reviwer(s)!**

NDdiss2$_\varepsilon$ documentation can be found at these locations:

`http://www.gsu.nd.edu`
`http://graduateschool.nd.edu`

Search for lepton flavor violating decays of the Higgs boson

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Fanbo Meng,

Colin Jessop, Director

, Director

Graduate Program in Physics
Notre Dame, Indiana
March 2018

Search for lepton flavor violating decays of the Higgs boson

Abstract

by

Fanbo Meng

this going to put with the abstract, a summary of the whole analysis

NEW DEDICATION NAME

To be written

CONTENTS

FIGURES . . . . . v

TABLES . . . . . vii

PREFACE . . . . . ix

ACKNOWLEDGMENTS . . . . . x

CHAPTER 1: Introduction . . . . . 1
1.1 Pre-LHC era on lepton Flavor violation(LFV) search and results . 1
1.2 Overview of CMS Run1 and Run2 direct search on LFV . . . . . 1

CHAPTER 2: Theory . . . . . 2
2.1 Standard model . . . . . 2
2.2 LFV in beyond standard model theories . . . . . 2

CHAPTER 3: LHC and CMS experiment . . . . . 3
3.1 LHC accelerator . . . . . 3
3.2 CMS experiment . . . . . 3
3.2.1 Tracker . . . . . 3
3.2.2 ECAL . . . . . 3
3.2.3 HCAL . . . . . 3
3.2.4 Muon System . . . . . 3
3.2.5 Trigger . . . . . 3
3.2.5.1 Level 1 Trigger . . . . . 3
3.2.5.2 High Level Trigger . . . . . 3

CHAPTER 4: Datasets . . . . . 4
4.1 Datasets used in LFV analysis . . . . . 4
4.2 Event reconstruction . . . . . 4
4.2.1 Muon reconstruction and selection criteria . . . . . 4

4.2.2 Tau lepton reconstruction . . . . . 4
4.3 Electron reconstruction . . . . . 13
4.4 Event simulation . . . . . 13

CHAPTER 5: LFV event selection . . . . . 14
5.1 $H \to \mu\tau_h$ . . . . . 14
5.1.1 Loose selection . . . . . 14
5.1.2 Cut-based analysis . . . . . 16
5.1.3 Multivariate analysis . . . . . 19
5.2 $H \to e\tau_h$ . . . . . 20
5.2.1 Loose selection . . . . . 20
5.2.2 Cut-based analysis . . . . . 25

CHAPTER 6: LFV analysis background estimation . . . . . 27
6.0.1 Background estimates for H $\to \mu\tau_h$ . . . . . 27
6.0.1.1 Misidentified Leptons . . . . . 27
6.0.1.2 Other backgrounds . . . . . 32
6.0.2 Background estimates for H $\to e\tau_h$ . . . . . 33
6.0.2.1 Misidentified Leptons . . . . . 33
6.0.2.2 $Z \to \tau\tau$ . . . . . 35
6.0.2.3 Other backgrounds . . . . . 37

CHAPTER 7: Signal extraction and systematics . . . . . 38
7.1 Boosted decision trees method . . . . . 38
7.2 statistical methods . . . . . 41
7.3 Systematics . . . . . 41
7.3.1 Systematics used in H $\to \mu\tau_h$ . . . . . 42
7.3.2 Systematic uncertainties in H $\to e\tau_h$ . . . . . 43

CHAPTER 8: LFV Higgs decay searching results . . . . . 49
8.0.1 Results in H $\to \mu\tau_h$ search . . . . . 49
8.0.2 Results in H $\to e\tau_h$ search . . . . . 57

CHAPTER 9: Conclusion . . . . . 62

BIBLIOGRAPHY . . . . . 63

FIGURES


5.1 Expected limits based on an Asimov dataset as a function of $M_T(\tau, MET)$ for the different categories. . . . . 17

5.2 Expected limits based on an Asimov dataset as a function of $M_{jj}$ for the 2 jet categories. . . . . 19

5.3 Distributions of the input variables to the BDT for the H $\rightarrow \mu\tau_h$ channel. . . . . 21

5.4 Expected limits based on an Asimov dataset as a function of $M_T(\tau, MET)$ for the different categories. . . . . 22

5.5 Overtraining checking for the BDT training in the TMVA package. 22

5.6 With loose selection conditions, the comparison of the observed collinear mass distributions with background from prediction. The shaded grey bands indicate the total background uncertainty. The open histograms correspond to the expected signal distributions for $\mathcal{B}(\mathrm{H} \rightarrow e\tau_h) = 100\%$ in the 0-jet, 1-jet and 2-jet categories, respectively. . . . . 24

6.1 $f_\tau$ ratio for the $\tau$ fake rate calculation shows in term of tau decay modes, $p_T$, ECAL barrel, ECAL endcap, and $|\eta|$ . . . . . 30

6.2 $f_\tau$ ratio for the $\mu$ fake rate calculation shows in term of muon $p_T$ and $|\eta|$. . . . . 31

6.3 The validation of full data driven method for the misidentified lepton background. 30% of uncertainty is assumed in both cases. . . 32

6.4 $f_\tau$ and $f_\tau$ ratio for the fake rate calculation shows in term of $p_T$ and $|\eta|$ . . . . . 36

6.5 Data driven method validation in Region II with e and $\tau$ leptons in same sign of charge . . . . . 37

7.1 Tree structure example in BDT . . . . . 40

8.1 $M_{col}$ distribution after the selection. The signal and backgrounds in the plots have been normalized to the best fit values. The gray bands shows the total uncertainties in each bin and the signal is plotted with the branching ratio of 5% of the higgs decay branching ratio for visualization purpose. . . . . 51

8.2 BDT discriminator distribution after the selection. The signal and backgrounds in the plots has been normalized to the best fit values. The gray bands shows the total uncertainties in each bin and the signal is plotted with the branching ratio of 5% of the higgs decay branching ratio for visualization purpose. . . . . 52

8.3 The expected and observed 95% CL limits on branching ratio of total $H \to \mu\tau$ which shows the results from both H $\to \mu\tau_h$ and $H \to \mu\tau_e$ in the $M_{col}$ fit analysis and BDT fit analysis. . . . . 55

8.4 Upper limit of flavour violating Yukawa coupling $|Y_{\mu\tau}|, |Y_{\tau\mu}|$ from the BDT fit analysis. The red solid line represents the expected limit and the black solid line is the observed limit. 68% and 95% range of containing the observed limit is shown in the green and yellow band. The result from $\tau \to 3\mu$ is shown in dark green [18, 19, 27] and the one from $\tau \to \mu\gamma$ is shown in lighter green [18, 27]. The theoretical naturalness limit $Y_{ij}Y_{ji} \leq m_i m_j/v^2$ is shown by purple diagonal line [18]. . . . . 56

8.5 $M_{col}$ distribution after the selection. The signal and backgrounds in the plots have been normalized to the best fit values. The gray bands shows the total uncertainties in each bin and the signal is plotted with the branching ratio of 0.69% of the higgs decay branching ratio for visualization purpose. . . . . 58

8.6 95% CL upper limit on branching ratio on $H \to e\tau$ with both the results from H $\to e\tau_h$ and $H \to e\tau_\mu$ and the combined. . . . . 59

8.7 Upper limit of flavour violating Yukawa coupling $|Y_{e\tau}|, |Y_{\tau e}|$ from the combined $H \to e\tau$ result. The red solid line represents the expected limit and the black solid line is the observed limit. 68% and 95% range of containing the observed limit is shown in the green and yellow band. The result from $\tau \to 3e$ is shown in gray, the one from $\tau \to e\gamma$ is shown in dark green and presented analysis is in light blue. The theoretical naturalness limit $Y_{ij}Y_{ji} \leq m_i m_j/v^2$ is shown by purple diagonal line [18] . . . . . 61

TABLES


4.1 Monte Carlo samples used in the search, together with their respective cross sections. The k-factors for the W+jets and Z+jets samples are 1.21 and 1.16, respectively. The NNLO+NNLL QCD cross sections for the Higgs boson samples in gluon-gluon fusion production mode are considered in this version of the analysis, and will be replaced by N3LO cross sections before the approval. . . . 5

4.2 Muon ID used in the analysis, for the LHC data 2016, running period BCDEF. . . . . 6

4.3 Muon ID used in the analysis, for the LHC data 2016, running period G and H, also the monte Carlo samples. . . . . 7

4.4 Dominant hadronic $\tau$ lepton decays branching fractions and the associated intermediate resonance. The h stands for both $\pi$ and K. The table is symmetric under charge conjugation. . . . . 10

4.5 $\tau$ hadronic decay mode hypothesis signatures compatibility tests. $m_\tau$ is required to be in the mass window . . . . . 12

5.1 Selection criteria in each category with the optimization of the H $\rightarrow$ $\mu\tau_h$ analysis . . . . . 18

5.2 Selection criteria for each event category after cut optimization, for the H $\rightarrow$ $e\tau_h$ channel . . . . . 26

6.1 Definition of the samples used to estimate the misidentified lepton background. . . . . 28

6.2 Scale factor to the events in which a muon is misidentified as a tau 33

7.1 Part one of the systematic uncertainties considered in H $\rightarrow \mu\tau_h$ analysis. All of the uncertainties listed in the systematic tables are correlated between categories besides the ones after the sign $\oplus$. These are the uncertainties correlated in within each category but independent between categories. The theoretical uncertainties related to the acceptance and migration of events are listed in a range. The negative or positive values indicate a anticorrelated or correlated between categories . . . . . 44

7.2 Part two of the systematic uncertainties considered in H $\rightarrow \mu\tau_h$ analysis . . . . . 45

7.3 b tagging veto systematic uncertainty in each category . . . . . 45

7.4 The normalization systematic uncertainties considered in H $\rightarrow e\tau_h$ analysis. The uncertainties are correlated between categories besides the ones after $\oplus$. These uncertainies are uncorrelated between categories. . . . . 47

7.5 The systematic uncertainties that affect the shape of $M_{col}$ distribution. . . . . 48

7.6 Theoretical uncertainties that affects the Higgs boson production cross section. These uncertainties are correlated or anticorrelated(with minus sign superscript) between all of the categories. . . . . 48

8.1 $M_{col}$ fit analysis expected and observed upper limits at 95% CL and the best fit branching fractions in each of the categories. . . . . 53

8.2 BDT fit analysis expected and observed upper limits at 95% CL and the best fit branching fractions in each of the categories. . . . 54

8.3 Events yields of the signal and backgrounds are shown in the mass range 100 GeV $< M_{col} <$ 150 GeV. The branching ratio of H $\rightarrow e\tau_h$ is assumed to be 0.69% of the SM higgs production cross section. The numbers in the table are normalized to the integrated luminosity of 19.7 $fb^{-1}$. . . . . 59

8.4 $M_{col}$ fit analysis expected and observed upper limits at 95% CL and the best fit branching fractions in each of the categories of both H $\rightarrow e\tau_h$ and $H \rightarrow e\tau_\mu$ analysises. The asymmetric one standard-deviation uncertainties around the expected limits are shown in parentheses. . . . . 60

# PREFACE

probably I will write some preface, but it depends $H \to \mu\tau_h$

# ACKNOWLEDGMENTS

this part is going to be filled with acknowledge

# CHAPTER 1

# Introduction

## 1.1 Pre-LHC era on lepton Flavor violation(LFV) search and results

## 1.2 Overview of CMS Run1 and Run2 direct search on LFV

# CHAPTER 2

# Theory

## 2.1 Standard model

## 2.2 LFV in beyond standard model theories

# CHAPTER 3

# LHC and CMS experiment

## 3.1 LHC accelerator

## 3.2 CMS experiment

### 3.2.1 Tracker

### 3.2.2 ECAL

### 3.2.3 HCAL

### 3.2.4 Muon System

### 3.2.5 Trigger

probably I have to put something there

#### 3.2.5.1 Level 1 Trigger

probably I have to put something there2

#### 3.2.5.2 High Level Trigger

# CHAPTER 4

# Datasets

## 4.1 Datasets used in LFV analysis

## 4.2 Event reconstruction

### 4.2.1 Muon reconstruction and selection criteria

With the PF muon as the input, muons used in the analysis are further categorized into different identification, isolation categories. Muon ID suggested in CMS RunII, for the LHC data 2016, running period BCDEF, ICHEP medium muon ID is applied(table 4.2), for running period G and H, also the monte Carlo samples, standard medium muon ID(table 4.3)are applied to achieve the best performance for muon identification.

### 4.2.2 Tau lepton reconstruction

In Run I CMS experiment, tau lepton are constructed with hadrons plus strips(HPS) algorithm. In general, HPS starts with PF jets which are reconstructed with $anti - k_T$, as the initial seeds. $\pi_0$ components from the $\tau$ hadronic decays are first constructed and combined with the charge hadrons parts, to iden-

| MC simulations | Cross section |
|---|---|
| /DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 4954.0 pb (LO) |
| /DY1JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 1012.5 pb (LO) |
| /DY2JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 332.8 pb (LO) |
| /DY3JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 101.8 pb (LO) |
| /DY4JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 54.8 pb (LO) |
| /DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 18610 pb (LO) |
| /TT_TuneCUETP8M1_13TeV-powheg-pythia8 | 831.76 pb |
| /WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 50380 pb (LO) |
| /W1JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 9644.5 pb (LO) |
| /W2JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 3144.5 pb (LO) |
| /W3JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 954.8 pb (LO) |
| /W4JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/ | 485.6 pb (LO) |
| /WZTo1L3Nu_13TeV_amcatnloFXFX_madspin_pythia8 | 3.05 pb |
| /WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8 | 10.71 pb |
| /WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8 | 5.595 pb |
| /ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1 | 35.85 pb |
| /ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1 | 35.85 pb |
| /ST_t-channel_antitop_4f_leptonDecays_13TeV-powheg-pythia8_TuneCUETP8M1 | 80.95 pb |
| /ST_t-channel_top_4f_leptonDecays_13TeV-powheg-pythia8_TuneCUETP8M1 | 136.02 pb |
| /WWTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8 | 1.212 pb |
| /ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8 | 3.22 pb |
| /VVTo2L2Nu_13TeV_amcatnloFXFX_madspin_pythia8 | 11.95 pb |
| /WGToLNuG_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8 | 489.0 pb |
| /WGstarToLNuMuMu_012Jets_13TeV-madgraph | 2.793 pb |
| /WGstarToLNuEE_012Jets_13TeV-madgraph | 3.526 pb |
| /GluGluHToTauTau_M125_13TeV_powheg_pythia8 | 48.58 pb X 0.0627 |
| /VBFHToTauTau_M125_13TeV_powheg_pythia8 | 3.782 pb X 0.0627 |
| /GluGluHToWWTo2L2Nu_M125_13TeV_powheg_pythia8 | 48.58 pb X 0.0227 |
| /VBFHToWWTo2L2Nu_M125_13TeV_powheg_pythia8 | 3.782 pb X 0.0227 |
| /ZHToTauTau_M125_13TeV_powheg_pythia8 | 0.8839 pb X 0.0627 |
| /WminusHToTauTau_M125_13TeV_powheg_pythia8 | 1.373 pb X 0.0627 |
| /WplusHToTauTau_M125_13TeV_powheg_pythia8 | 1.373 pb X 0.0627 |
| /ttHJetToTT_M125_13TeV_amcatnloFXFX_madspin_pythia8 | 0.507 pb X 0.0627 |



TABLE 4.1

TABLE 4.2

Muon ID used in the analysis, for the LHC data 2016, running period BCDEF.

| ICHEP mediumID description | Technical description |
|---|---|
| Loose muon ID | PFLoose Muon |
| Fraction of valid tracker hits | $> 0.49$ |
| 1.Good Global muon | Global muon |
| | Normalized global-track $\chi^2 < 3$ |
| | Tracker-Standalone position match $< 12$ |
| | kick finder $< 20$ |
| | Segment compatibility $> 0.303$ |
| 2. Tight segment compatibility | Segment compatibility $> 0.451$ |

TABLE 4.3

Muon ID used in the analysis, for the LHC data 2016, running period G and H, also the monte Carlo samples.

| Standard mediumID description | Technical description |
|---|---|
| Loose muon ID | PFLoose Muon |
| Fraction of valid tracker hits | $> 0.8$ |
| 1.Good Global muon | Global muon |
| | Normalized global-track $\chi^2 < 3$ |
| | Tracker-Standalone position match $< 12$ |
| | kick finder $< 20$ |
| | Segment compatibility $> 0.303$ |
| 2. Tight segment compatibility | Segment compatibility $> 0.451$ |

tify different $\tau$ decay modes and calculate $\tau$ four-momentum and other quantities [10].

Photon conversions and the bremsstrahlung of electron/positron when traveling inside the CMS detector are well treated by the HPS algorithm. These phenomenons broaden the signature of the tau decay. With PF jets as input, the algorithm constructs strips out of electromagnetic particles and starts by taking the strip in which contains the most energetic electromagnetic particle as the center one. With the center strip, a window of the size $\Delta\eta = 0.05$ and $\Delta\phi = 0.2$ is taken. Within this window, if other charged particles are found, they are associated with the strip. The position of the strip is taken and four momentum of the strip is calculated. This procedure is repeated, until no strips can be constructed. The selected strips are required to have $P_T^{strip} > 1GeV$. The following decay topologies are taking into account by HPS:

- one charged particle without any strip, $h^{\pm}$ and the case when $\pi^0$ is not energetic enough to form a strip
- one charged particle plus one strip
- one charged particle plus two strips
- three charged partibles.

All of the charged hadrons and strips are required to be contained in the $\Delta R = 2.8/P_T^{\tau_h}$ core, where the $P_T^{\tau_h}$ is the reconstructed $\tau_h$ transverse momentum and $\Delta R$ is defined as $\Delta R = \sqrt{(}\Delta\phi^2 + \Delta\eta^2)$. The $\tau_h$ candidate is also required to match the direction of the seed PF jet within $\Delta R = 0.1$. Assuming all of the charged hadrons to be pions and taking in the associated strips, the HPS

algorithm requires that different decay topologies meet the intermediate meson mass as listed in Table. 4.4.

The cut based $\tau_h$ isolation discriminant required that the PF charged particles and photons to be considered in the isolation variable have $p_T > 0.5$ GeV and within an isolation cone $\Delta R = 0.5$ in $\tau_h$ direction. The particles that constituent $\tau_h$ are excluded from the summation. The effect of charged particle from pileup is eliminated by considering on the charged particle oriented from the $\tau_h$ production vertex with in $D_z = 0.2$ cm and $\Delta r = 0.03$ cm. The effect of pileup on the isolation of the photons on the strips is estimated by summing the charged particles that are not oriented from $\tau_h$ decay primary vertex, within $\Delta R = 0.8$ cm in the direction of and have the impact parameter $D_z > 0.2$ cm. Then a factor $\Delta\beta$ is multiplied to the $p_T$ sum. The isolation variable is defined as in Equation. 4.1.

$$I_\tau = \sum p_T^{\text{charged}}(d_z < 0.2 \text{ cm}) + \max(0, \sum p_T^\gamma - \Delta\beta \sum p_T^{\text{charged}}(d_z > 0.2 \text{ cm})) \tag{4.1}$$

Tight, loose, medium working points(WP) for the tau isolation discriminants. The exact of the energy selection is suggested by the study of QCD dijet events, by requiring the $I_\tau$ in Equation. eq:taucutiso1 to have different values. The loose cut brings in approximate 1% of fake $\tau$ from jets.

In CMS Run II, Tau reconstruction algorithm HPS has been improved [12]. The major improvement lies in Dynamic strip instead of fix size strip. Tau decay products can also affect the isolation. Charged pions in tau decay products experience nuclear interaction with tracker materials, which can results in low $P_T$ secondary particles. Photons from the neutral pion decay can also go through

TABLE 4.4

Dominant hadronic $\tau$ lepton decays branching fractions and the associated intermediate resonance. The h stands for both $\pi$ and K. The table is symmetric under charge conjugation.

| Decay mode | Resonance | Mass ($MeV/c^2$) | Branching fraction(%) |
|---|---|---|---|
| $\tau^- \to h^- v_\tau$ | | | 11.6% |
| $\tau^- \to h^- \pi^0 v_\tau$ | $\rho^-$ | 770 | 26.0% |
| $\tau^- \to h^- \pi^0 \pi^0 v_\tau$ | $\alpha_1^-$ | 1200 | 9.5% |
| $\tau^- \to h^- h^+ h^- v_\tau$ | $\alpha_1^-$ | 1200 | 9.8% |
| $\tau^- \to h^- h^+ h^- \pi^0 v_\tau$ | | | 4.8% |

pair production into $e^+e-$, which further spread because of bremsstrahlung and the magnetic field. Broadening the strip is need in these cases in order to better cover the tau decay production. On the other hand, if the tau is boosted, high $P_T$ decay products tends to be more concentrate and smaller strip size will be better. Similar to RunI tau reconstruction, the algorithm starts with hightest $p_T$ charged particle as seeds for the strip. Starting from the seed strip, a window in $\eta$ and $\phi$ direction is set.

$$\delta\eta = f(P_T^\gamma) + f(P_T^{strip}) \qquad f(P_T) = 0.2 \cdot P_T^{-0.66}$$

$$\delta\phi = g(P_T^\gamma) + g(P_T^{strip}) \qquad g(P_T) = 0.35 \cdot P_T^{-0.71}$$

The window is determined from single $\tau$ gun MC simulation. 95% of the decay product will be covered in that range. The upward and downward limits for $\eta$ is 0.15 and 0.05, for $\phi$ the range is 0.3 and 0.05. The position of strip is set as $p_T$ weighted average against all of the objects.

$$\eta_{strip} = \frac{1}{P_T^{strip}} \cdot \sum P_T^{\gamma} \cdot \eta_{\gamma}$$
$$\phi_{strip} = \frac{1}{P_T^{strip}} \cdot \sum P_T^{\gamma} \cdot \phi_{\gamma}$$

Construct the strip until no seed strip can be found. After the construction of the $\tau$ lepton, for different decay mode, $m_\tau$ is required to lie in different mass windows [13]. The conditions of different hadronic decay mode mass window are listed in the Table. 4.5. With respect to RunI, the difference in mass window is $\delta m$, which originates from dynamic clustering. $\delta m$ is calculated as:

$$\delta m = \sqrt{\left(\frac{\partial m_\tau}{\partial \eta_{\text{strip}}} \cdot f(p_T^{\text{strip}})\right)^2 + \left(\frac{\partial m_\tau}{\partial \phi_{\text{strip}}} \cdot g(p_T^{\text{strip}})\right)^2}$$

with:

$$\frac{\partial m_\tau}{\partial \eta_{strip}} = \frac{P_z^{strip} \cdot E_\tau - E_{strip} \cdot P_z^\tau}{m_\tau}$$
$$\frac{\partial m_\tau}{\partial \phi_{\text{strip}}} = \frac{-(P_y^\tau - P_y^{\text{strip}}) \cdot P_x^{\text{strip}} + (P_x^\tau - P_x^{\text{strip}}) \cdot P_y^{\text{strip}}}{m_\tau}$$

TABLE 4.5

$\tau$ hadronic decay mode hypothesis signatures compatibility tests. $m_\tau$ is required to be in the mass window

| Decay mode | Mass window |
|---|---|
| $\tau^- \to h^-\pi^0 v_\tau$ | $0.3 - \delta m_\tau < m_\tau < 1.3 \cdot \sqrt{p_T/100} + \delta m_\tau$ |
| $\tau^- \to h^-\pi^0\pi^0 v_\tau$ | $0.4 - \delta m_\tau < m_\tau < 1.2 \cdot \sqrt{p_T/100} + \delta m_\tau$ |
| $\tau^- \to h^-h^+h^- v_\tau$ | $0.8 - \delta m_\tau < m_\tau < 1.5 + \delta m_\tau$ |

In current algorithm, $\tau^- \to h^-h^+h^- v_\tau$ is not included, because of the jets contamination. This hadronic $\tau$ decay mode composed of 4.8% of total branching fraction. The $h^-\pi^0$ and $h^-\pi^0\pi^0$ are analyzed together, which is referred as $h^-\pi^0$.

The analysis with 2016 datasets, MVA based $\tau$ isolation criteria is used, which keeps high identification efficiency while maintains relatively low fake rate compared with cut based discriminator. A Boosted Decision Tree(BDT) has been used in the training of the isolation variable. With BDT, the isolation variable shows a good distinguishing power against jets. Various variables have been used as BDT inputs. The variables are isolation variable($I_\tau$), impact parameter from highest $p_T$ track of $\tau_h$ candidate, $\tau_h$ decay mode information, shape variables like $\Delta R$, $\Delta\eta$, $\tau-$lifetime information and photon electron multiplicity, more of the exact variables used are discussed in [12, 13]. BDT uses these variables to distinguishing $\tau_h$ dacay($H \to \tau\tau$) from jets, which can be the decay products from quarks and gluons(QCD MC).

BDT method also used in the tau discriminating again electron training. The algorithm utilizes the variables that sensitive to the energy deposit in Ecal and Hcal, the electron bremsstrahlung, overall particle multiplicity and difference in electromagnetic and hadronic showers. Detail list of variable can be find in [12, 13].

Tau signals can be faked by muons, especially in $\tau_h$ decay mode $h^{\pm}$. Tau against muon cut based muon discriminant is set by checking if there are signals in the muon system within $\Delta R = 0.3$ of the $\tau_h$ direction or if the energy sum from Ecal and Hcal is less than 20% of the total $\tau$ energy. If less than two hits are found in the muon system, then it passes the loose working point. If no hits are found in the muon system, then this is the tight working point.

## 4.3 Electron reconstruction

## 4.4 Event simulation

# CHAPTER 5

## LFV event selection

In both 8 TeV $H \to e\tau_h$ and 13 TeV $H \to \mu\tau_h$ analyses, events are selected in several steps. A loose selection selects on the different IDs, energy, geometry parameters of the analysis related objects and followed by a tighter set of selection criteria in which selection requirements are placed on the kinematics variables and fits on variable $M_{col}$ in both $H \to e\tau_h$ and $H \to \mu\tau_h$ analysis. This selection sequence is referred as $M_{col}$ fit analysis. In $H \to \mu\tau_h$, there is another selection sequence. After placing the loose selection, a multivariate analysis with Boosted decision tree (BDT) is exploited and fits on the BDT discriminator. This is referred as the BDT fit analysis and provides more sensitive results in H $\to \mu\tau_h$ analysis.

### 5.1 $H \to \mu\tau_h$

#### 5.1.1 Loose selection

Tau leptons from signal events decay hadronically. SM higgs is much heavier than its decay products $\mu$ and $\tau$. $\mu$ and $\tau$ leptons are expected to have high $P_T$. The decay products from signal events are boosted, a cut on $\Delta R = \sqrt{(}\Delta phi^2 + \Delta eta^2)$, $\Delta R > 0.3$ is applied. $\mu$ and $\tau$ candidates are required to have opposite sign of charges as Higgs has no electric charge. Further, events with additional $\mu$ and

$\tau$ that pass a loose selection, events with jets that are identified by the combined secondary vertex(CSVv2) b-tagging algorithm [14] as a b quark jets are discarded. The trigger HLT_IsoMu24 or HLT_IsoTkMu24 used in the analysis selects isolated muons that have energy higher than 24 GeV at HLT level. An further $P_T$ cut on reconstructed $\mu$, $P_T > 26$ GeV and $|\eta| < 2.4$ are required. Muons are required to pass the recommended Medium muon ID and tight cut based isolation $I^{\mu}_{rel} < 0.15$.

Hadronic taus are required to have $P_T > 30$ GeV,$|\eta| < 2.3$, passing old tau decay mode finding, the MVA based tight tau isolation ID and tau discriminators against electrons and muons.

Events in the analysis are divided into four categories based on the number of jets. In 2-jets category, it is furthered divided into 2 categories, 2-jet gluon gluon fusion higgs production(ggH) category and 2-jet vector boson fusion(VBF) category based on the value of 2 jets invariant mass($M_{jj}$) . The 0-jet category enhances ggH production mode. In 1-jet category, the dominant signal production mode is also ggH, but with a boosted jet associated with the production and VBF higgs production also contributes this category. In the 2-jet ggH category, signal events mainly come from ggH, while in 2-jet VBF, VBF higgs is the dominant production channel. The following is a more detailed list of the selection condition in each categories.

**0-jet:** No events have jets pass the loose PF ID and with jet $P_T > 30$ GeV, $|\eta| < 4.7$.

**1-jet:** Events with one jet passes losse PF ID and jet $P_T > 30$ GeV, $|\eta| < 4.7$.

**2-jets ggH:** Events have two jets passing loose PF ID, $P_T > 30$ GeV, $|\eta| < 4.7$ and a requirement on the invariant mass of the two jets, $M_{jj} < 550$ GeV.

**2 jets VBF:** Events with two jets pass loose PF ID. Jets $P_T > 30$ GeV, $|\eta| < 4.7$ and

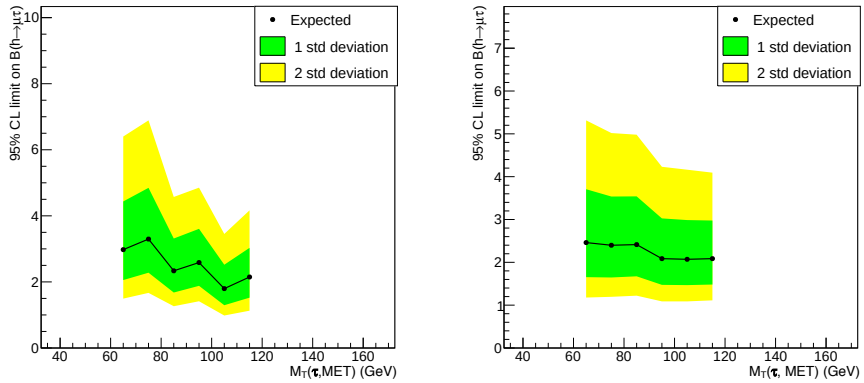
$M_{jj} > 550$ GeV are required.

The threshold on $M_{jj}$ has been optimized to give the best expected exclusion limits.

### 5.1.2 Cut-based analysis

After the loose selection and the categorization, a further cut-based selection is applied. Variables that can help distinguish signal from background are $P_T^{\mu}$, $P_T^{\tau}$ and $M_T(\tau_h)$. The lepton $P_T$ variables are very powerful background discriminant variables, but it will also cause the problem that signal picks under backgrounds. Leptons from signal process incline to have higher $P_T$ values, by cutting tighter on the lepton $P_T$, more background events can be removed. However this will also reshape some of the backgrounds, making them peak closer under the signal so that signal processes will be affected more by the background statistics fluctuation. In the $H \to e\tau_h$ analysis, the effect of cutting tighter on lepton $P_T$ will be shown. In $H \to \mu\tau_h$ search, lepton $P_T$ variables are kept at loose values and optimizes on other variables to achieve tighter expected limits.

Optimization is proceeded with only Monte Carlo samples so as not to double use data. In $H \to \mu\tau_h$ channel, the variables tuned are $M_{jj}$ and $M_T(\tau_h)$. Cuts have been optimized to have the most stringent expected limits with the Asimov dataset. If looser values of the cuts give the same expected limits as the tighter ones, then the looser cut values are chosen to have more statistics. Examples of optimization by checking the limits are shown in Figure. 5.1 and 5.2, in which the optimization of $M_T(\tau_h)$ and $M_{jj}$ are shown. A full summary of the cuts optimized for $H \to \mu\tau_h$ is shown in Table. 5.1




(a) 0 jet

(b) 1 jet

(c) 2 jets, gg-enriched

(d) 2 jets, VBF-enriched

Figure 5.1. Expected limits based on an Asimov dataset as a function of $M_T(\tau, MET)$ for the different categories.

TABLE 5.1

Selection criteria in each category with the optimization of the H $\rightarrow \mu\tau_h$ analysis

| |
|---|
| **0-jet category** |
| • $p_T^{\mu} > 26$ GeV, $p_T^{\tau} > 30$ GeV |
| • $M_T(\tau) < 105$ GeV |
| • No jets with $p_T^{jet} > 30$ GeV, $\lvert\eta\rvert < 4.7$, LooseID |
| **1-jet category** |
| • $p_T^{\mu} > 26$ GeV, $p_T^{\tau} > 30$ GeV |
| • $M_T(\tau) < 105$ GeV |
| • One jet with $p_T^{jet} > 30$ GeV, $\lvert\eta\rvert < 4.7$, LooseID |
| **2-jet, gg-enriched category** |
| • $p_T^{\mu} > 26$ GeV, $p_T^{\tau} > 30$ GeV |
| • $M_T(\tau) < 105$ GeV |
| • $p_T^{jet1} > 30$ GeV,$p_T^{jet2} > 30$ GeV $\lvert\eta_{jet1}\rvert < 4.7$,$\lvert\eta_{jet2}\rvert < 4.7$, LooseID |
| • $M_{jj} < 550$ GeV |
| • Two jets with $p_T^{jet} > 30$ GeV, $\lvert\eta\rvert < 4.7$, LooseID |
| **2-jet, vbf-enriched category** |
| • $p_T^{\mu} > 26$ GeV, $p_T^{\tau} > 30$ GeV |
| • $M_T(\tau) < 85$ GeV |
| • $p_T^{jet1} > 30$ GeV,$p_T^{jet2} > 30$ GeV $\lvert\eta_{jet1}\rvert < 4.7$,$\lvert\eta_{jet2}\rvert < 4.7$, LooseID |
| • $M_{jj} > 550$ GeV |
| • Two jets with $p_T^{jet} > 30$ GeV, $\lvert\eta\rvert < 4.7$, LooseID |

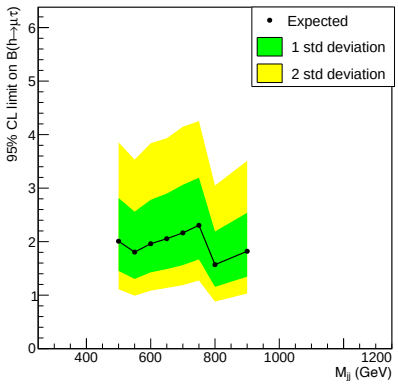



Figure 5.2. Expected limits based on an Asimov dataset as a function of $M_{jj}$ for the 2 jet categories.

### 5.1.3 Multivariate analysis

A Boost decision trees(BDT) method is used as the multivariate analysis method in $H \to \mu\tau_h$ search which is more sensitive than the $M_{col}$ fit analysis. BDT algorithm used in this search is implemented in the TMVA package [20]. BDT takes in signal and background datasets with a selected set of input variables. Input variables are the ones that show distinguishing power between signal and backgrounds. The training output is a weight file, which contains a list of weights to indicate how likely an event is signal like with a give set of input variables. A more detail description of the BDT method is available in section 7.1. In this analysis, signal and background events are required to pass the loose selection criteria. All of the categories are combined. The signal events from ggH and VBF higgs production mode are mixed by the weight with respect to their production cross section. The background sample used in the training is misidentified lepton background from the like sign region(Region II as in Table. 6.1). The list of BDT input variables is in the following list and the distribution of the variables are

shown in Fig. 5.3.

- Transverse mass between the $\tau_h$ and $\mathrm{E}_T^{miss}$, $M_T(\tau_h)$.
- Missing transverse energy, $\mathrm{E}_T^{miss}$.
- Pseudorapidity difference between the $\mu$ and the $\tau_h$ candidate, $\Delta\eta(\mu, \tau_h)$.
- Azimutal angle between the $\mu$ and the $\tau_h$, $\Delta\phi(\mu, \tau_h)$.
- Azimutal angle between the $\tau_h$ and the $\mathrm{E}_T^{miss}$, $\Delta\phi(\tau_h, \mathrm{E}_T^{miss})$.
- Collinear mass, $M_{col}$.
- Muon $p_T$.
- $\tau_h$ $p_T$.

The chosen input variables show low correlation in both samples as shown in Figure. 5.4. The TMVA overtraining checks is performed by checking with the testing samples to see if similar distinguishing power can be achieved with the output weight file from the training samples. Training and testing samples show good agreement and the training process exempts from overtraining as indicated in Figure. 5.5.

## 5.2 $H \to e\tau_h$

### 5.2.1 Loose selection

In H $\to$ $e\tau_h$ channel, trigger $HLT\_Ele27\_WP80$ is used, which applies an electron $p_T$ cut at 27 GeV at HLT level. A further cut on electron $P_T > 30$ GeV is applied. Electrons are also required to have $|\eta_e < 2.3|$ and $D_z < 0.2cm$.

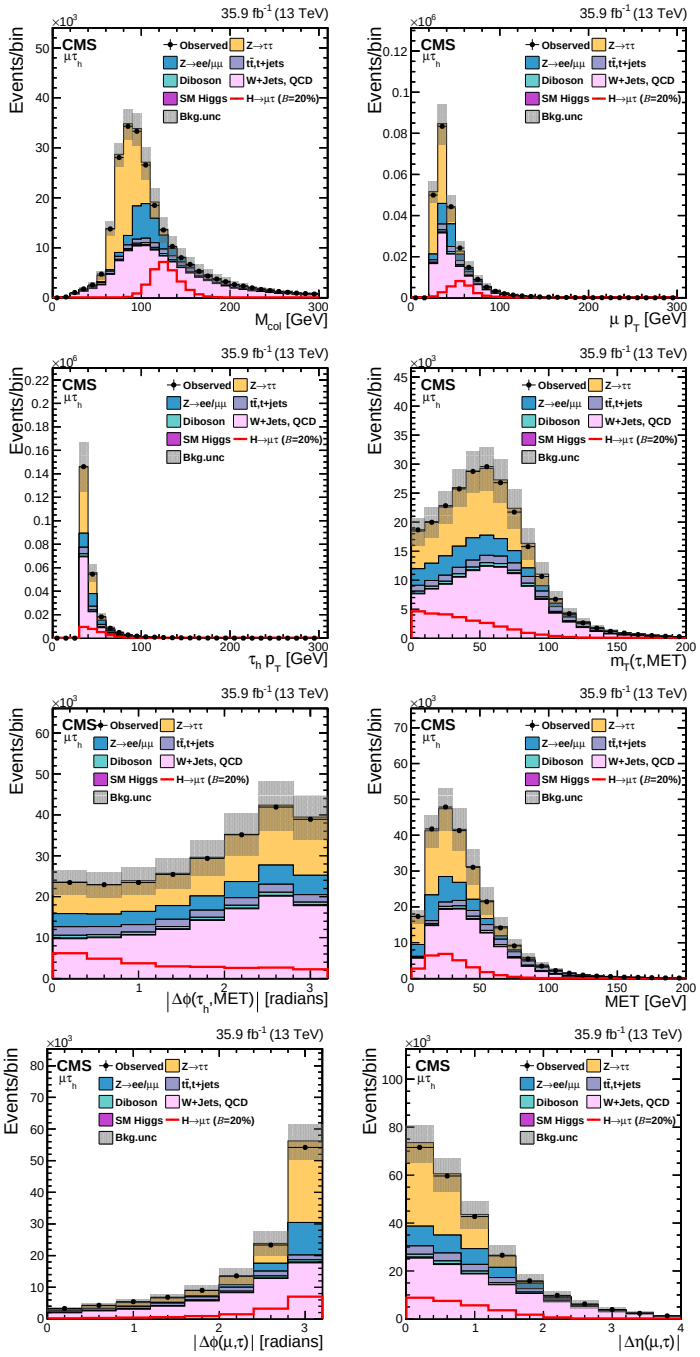
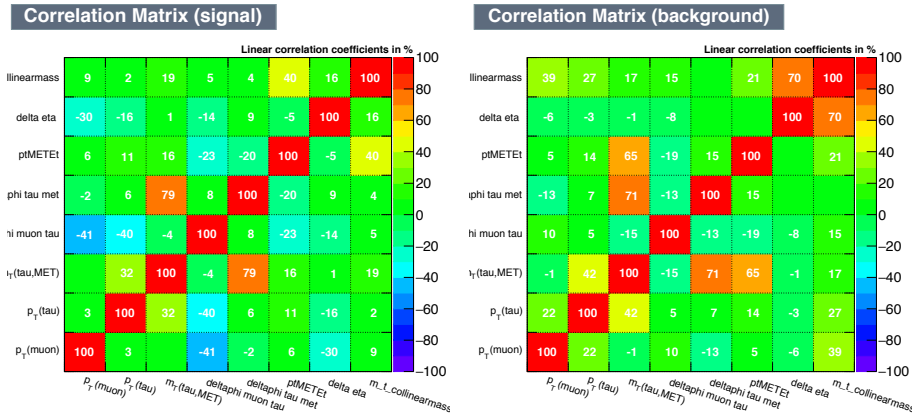



Figure 5.3. Distributions of the input variables to the BDT for the $H \to \mu\tau_h$ channel.




(a) Signal sample variables correlation (b) Background sample variable correlation

Figure 5.4. Expected limits based on an Asimov dataset as a function of $M_T(\tau, MET)$ for the different categories.

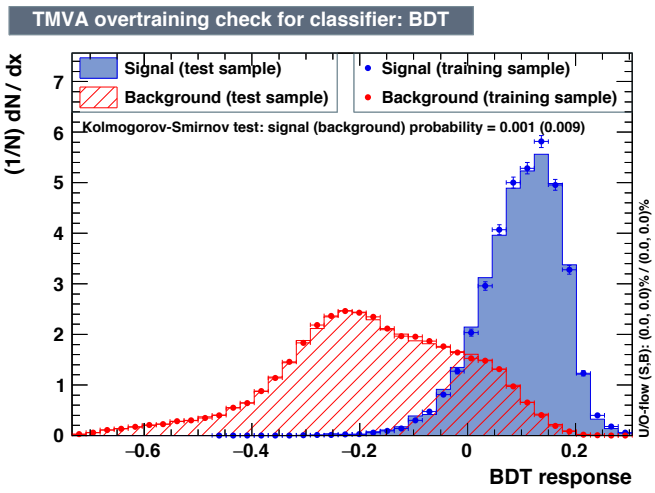



Figure 5.5. Overtraining checking for the BDT training in the TMVA package.

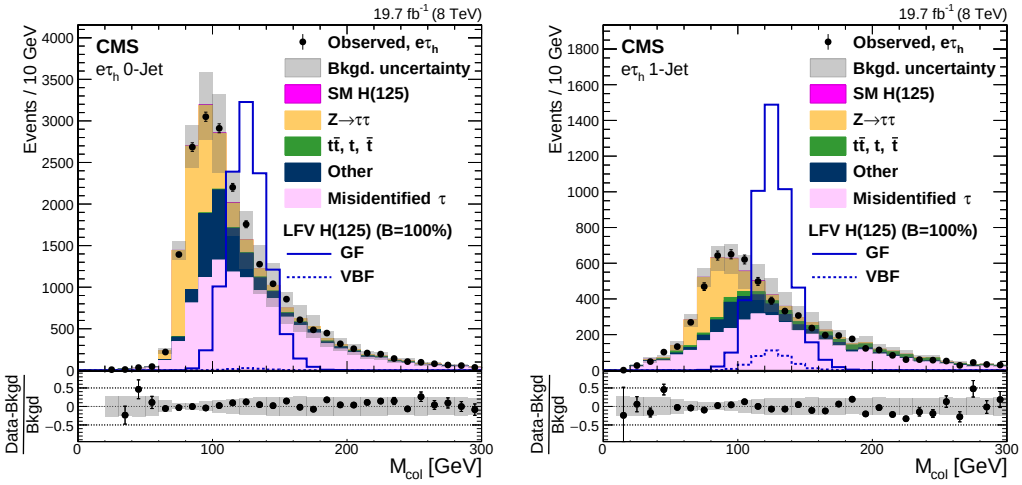
$D_z$ is the longitudinal impact parameter that shows the displacement between primary vertex and track path. Electrons are required to pass the MVA based tight ID and cut based PF tight isolation $I^e_{rel} < 0.1$. Tau candidates are required to have $P_T > 30$ GeV, pseudorapidity $|\eta^\tau < 2.3|$ and the longitudinal impact parameter $D_z < 0.2$ cm. Tau isolation used is the cut based tight tau isolation. In addition, tau candidates are required to pass the tau Decay mode finding and tau discriminator against electrons and muons. The analysis also requires no extra isolated electrons with $p_T > 10$ GeV and extra taus with $p_T > 20$ GeV. Tau and electron candidates are required to have opposite sign of charges and separate with $\Delta R > 0.4$ from any jets in the events with $p_T > 30$ GeV. All of the requirements contribute to the selections of good qualities candidates. The datasets are binned into three categories according to the number of jets in the events:

**0-jet:** No events have jets pass the loose PF ID and with jet $P_T > 30$ GeV, $|\eta| < 4.7$. This category enhances the gluon-gluon fusion contribution.

**1-jet:** Events with one jet passes loose PF ID and jet $P_T > 30$ GeV , $|\eta| < 4.7$. This category enhances the gluon-gluon fusion production with initial state radiation.

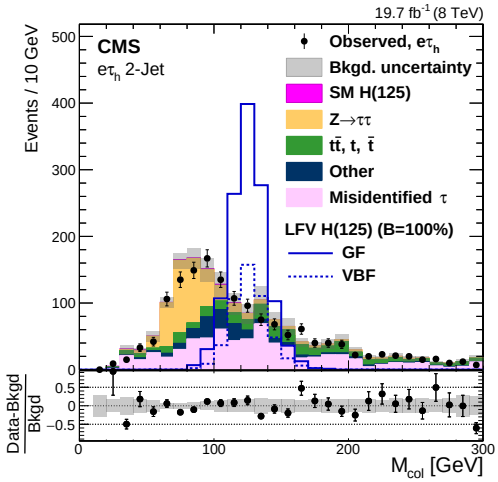
**2 jets:** Events with two jets pass loose PF ID and with jet $P_T > 30$ GeV and $|\eta| < 4.7$, This category contains both Higgs production mode and with an enhancement in VBF production mode.

With the preselection and binning of the jets numbers, the $M_{col}$ distribution of H $\rightarrow e\tau_h$ in different categories are shown in Figure. 5.6




(a) 0 jet

(b) 1 jet




(c) 2 jets

Figure 5.6. With loose selection conditions, the comparison of the observed collinear mass distributions with background from prediction. The shaded grey bands indicate the total background uncertainty. The open histograms correspond to the expected signal distributions for $\mathcal{B}(\mathrm{H} \to e\tau_h) = 100\%$ in the 0-jet, 1-jet and 2-jet categories, respectively.

### 5.2.2 Cut-based analysis

A set of kinematics variables are used to further select signal events. In H $\rightarrow$ $e\tau_h$ channel, similar to the H $\rightarrow$ $\mu\tau_h$, muon and tau leptons from the signal events are highly boosted so as $p_T$ variables plays an important role in distinguishing from background events and the separation in $\phi$ direction is bigger between $\mu$ and $\tau$ of signal events. The variable $M_T(\tau_h)$ which is the transverse mass form by the reconstructed $\tau$ and MET is also used in the selection. In two jets category, the invariant mass of two jets $M_{jj}$ is used as a cut variable. The cuts have been optimized to have the most stringent expected limits with the Asimov dataset. The detailed cuts used for H $\rightarrow$ $e\tau_h$ is shown in table. 5.2.

TABLE 5.2

Selection criteria for each event category after cut optimization, for the $H \to e\tau_h$ channel

| |
|---|
| **0-jet category** |
| • $p_T^e > 45$ GeV, $p_T^\tau > 30$ GeV, $\Delta\phi_{e\tau} > 2.3$ |
| • $M_T(\tau) < 70$ GeV |
| • No jets with $p_T^{jet} > 30$ GeV, $\|\eta\| < 4.7$, LooseID |
| **1-jet category** |
| • $p_T^e > 35$GeV, $p_T^\tau > 40$ GeV |
| • One jet with $p_T^{jet} > 30$ GeV, $\|\eta\| < 4.7$, LooseID |
| **2-jet category** |
| • $p_T^e > 35$ GeV, $p_T^\tau > 30$ GeV |
| • $M_T(\tau) < 50$ GeV |
| • $p_T^{jet1} > 30$ GeV,$p_T^{jet2} > 30$ GeV $\|\eta_{jet1}\| < 4.7$,$\|\eta_{jet2}\| < 4.7$, LooseID |
| • $\Delta\eta(jet1, jet2) > 2.3$ |
| • $M_{jj} > 400$ GeV |
| • Two jets with $p_T^{jet} > 30$ GeV, $\|\eta\| < 4.7$, LooseID |

# CHAPTER 6

## LFV analysis background estimation

### 6.0.1 Background estimates for H $\rightarrow \mu\tau_h$

#### 6.0.1.1 Misidentified Leptons

The misidentified lepton backgrounds are estimated with full data driven method from the collision data. In H $\rightarrow \mu\tau_h$ channel, the misidentification $\tau$ and $\mu$ lepton rate is obtained from independent Z + jets data sets and then apply the rate to a control region that orthogonal to the signal region to estimate the misidentified lepton backgrounds. The control regions for the estimation and validation of the misidentified lepton background are shown in Table. 6.1. Region I is the signal region with full selection applied. Region III is the background enriched region, in which one or two leptons have loose isolation but not tight isolation. Region II is the same to Region I besides the requirement that $\mu$ and $\tau$ lepton pairs have the same sign of charge, so as Region IV to Region III. Region III is used to estimate the misidentified background in Region I. Region II and IV are used in the validation of the method.

$Z \rightarrow \mu\mu$ is used in Z + jets data sets to estimate both misidentified $\tau$ and $\mu$ lepton rates. Z bosons are selected in the invariant mass window $70 < M_{ll} < 110$ GeV and trigger HLT_IsoMu24 or HLT_IsoTkMu24 is used in the selection. Two muons that are composed of Z bosons are required to have $p_T > 26$ GeV, $|\eta| < 2.4$,

TABLE 6.1

Definition of the samples used to estimate the misidentified lepton background.

| **Region I** | **Region II** |
|---|---|
| $\ell_1^{\pm}$(isolated) | $\ell_1^{\pm}$(isolated) |
| $\ell_2^{\mp}$(isolated) | $\ell_2^{\pm}$(isolated) |
| **Region III** | **Region IV** |
| $\ell_1^{\pm}$(isolated) | $\ell_1^{\pm}$(isolated) |
| $\ell_2^{\mp}$(not-isolated ) | $\ell_2^{\pm}$(not-isolated) |

cut based tight muon isolation($I_{rel}^{\mu} < 0.15$) and passing the muon medium ID. In the Z+jets samples, with the selected Z boson in an event, the remaining jets are checked if they pass $\tau$ or $\mu$ selection.

To estimate the misidentified $\tau$ lepton rate, The jets in the sample Z + jets are checked if they pass the $\tau$ isolation ID. Misidentified $\tau$ lepton ratio $\tau(f_\tau)$ is calculated as in Equation . 6.1, together with one related ratio $f_\tau$. In Equation . 6.1, $\ell$ stands for leptons in general.

$$\ell(f_\ell) = \frac{f_\ell}{1 - f_\ell} \tag{6.1}$$

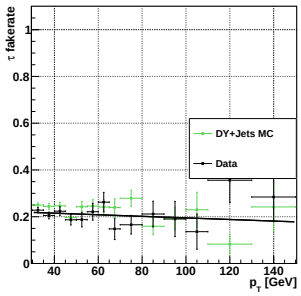
$$f_\ell = \frac{N_\ell(Z + jets\ \ell\ tight\ Iso)}{N_\tau(Z + jets\ \ell\ loose\ Iso)} \tag{6.2}$$

$f_\tau$ is the ratio between number of jets that pass tight $\tau$ MVA isolation ID and number of jets that pass very loose $\tau$ MVA isolation ID. The contribution from Diboson sample is subtracted with simulation, in which the third leptons are genuine leptons. With the ratio $f_\tau$ , misidentified rate $\tau(f_\tau)$(Equation. 6.2)

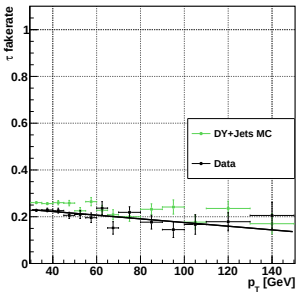
is applied as a weight to Region III. When estimating misidentified *tau* leptons, Region III is defined in the same selection condition as the signal region Region I , but $\tau$ leptons are required to pass the very loose MVA isolation and not pass the tight MVA isolation. The misidentified $\tau$ lepton rate $\tau(f_\tau)$ shows dependence on tau $p_T$, tau decay mode and ECAL geometry. Thus $\tau(f_\tau)$ is applied in term of tau $p_T$, tau decay modes and ECAL barrel and endcap seperately. The distribution of $f_\tau$ is shown in Fig. 6.1, which shows a comparison of the estimated $f_\tau$ between data and simulated Z+jets samples. The distribution of $f_\tau$ vs $|\eta|$ are also shown, which show less dependency.

The estimation of $\mu$ misidentified rate the same as the $\tau$ misidentified rate, but checks if the jets in Z + jets samples pass the muon isolation. As in Equation. 6.1, $f_\mu$ is defined as the ratio between tight cut based muon isolation($I^\mu_{rel} < 0.15$) and loose cut based muon isolation($I^\mu_{rel} < 0.25$). The misidentified $\mu$ lepton rate $\mu(f_\mu)$ is calculated by Equation. 6.2. $\mu(f_\mu)$ is a weight to apply to Region III. In the case of estimating the misidentified $\mu$, Region III is defined as Regin I but with $\mu$ passing the loose cut based muon isolation and not passing tight cut based muon isolation. The ratio $f_\mu$ is shown in Figure. 6.2 in term of muon $p_T$ and $|\eta|$ distribution.

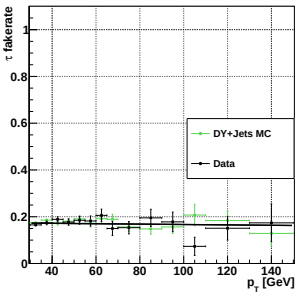
There are overlap between the estimated misidentified muon and tau lepton when both of the leptons are fake ones. The overlap events are estimated as in Equation. 6.4. In the equation, the N(Region III) stands for the events in Region III. In this case, Region III is defined as muon leptons passing the lose isolation, but not passing the tight isolation, while tau leptons pass the very loose isolation but not pass the tight isolation. The final misidentified lepton background obtained by adding the misidentified muon and tau background and deducts the overlap



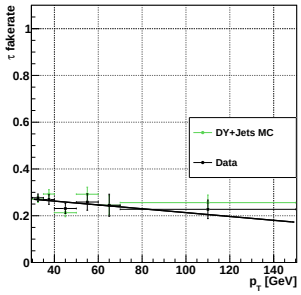

(a) EB tau decay mode 0



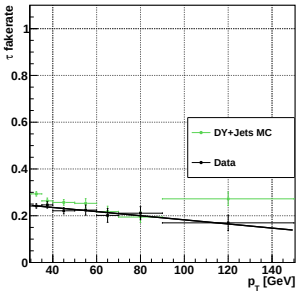

(b) EB tau decay mode 1



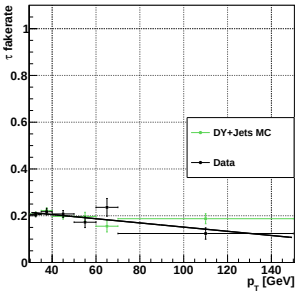

(c) EB tau decay mode 10



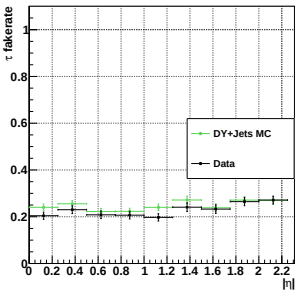

(d) EE tau decay mode 0



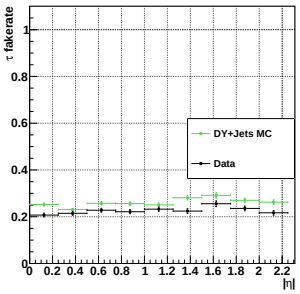

(e) EE tau decay mode 1



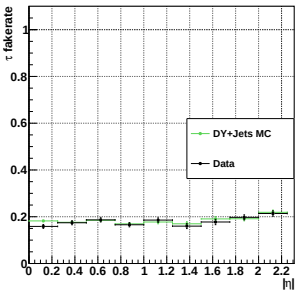

(f) EE tau decay mode 10




(g) tau decay mode 0 vs $|\eta|$

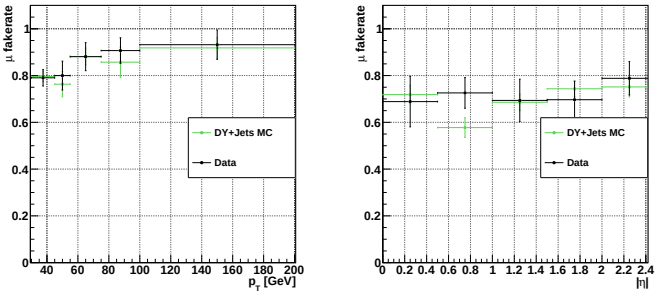



(h) tau decay mode 0 vs $|\eta|$




(i) tau decay mode 0 vs $|\eta|$

Figure 6.1. $f_\tau$ ratio for the $\tau$ fake rate calculation shows in term of tau decay modes, $p_T$, ECAL barrel, ECAL endcap, and $|\eta|$




(a) $f_\mu$ vs $p_T$ (b) $f_\mu$ vs $|\eta|$

Figure 6.2. $f_\tau$ ratio for the $\mu$ fake rate calculation shows in term of muon $p_T$ and $|\eta|$.

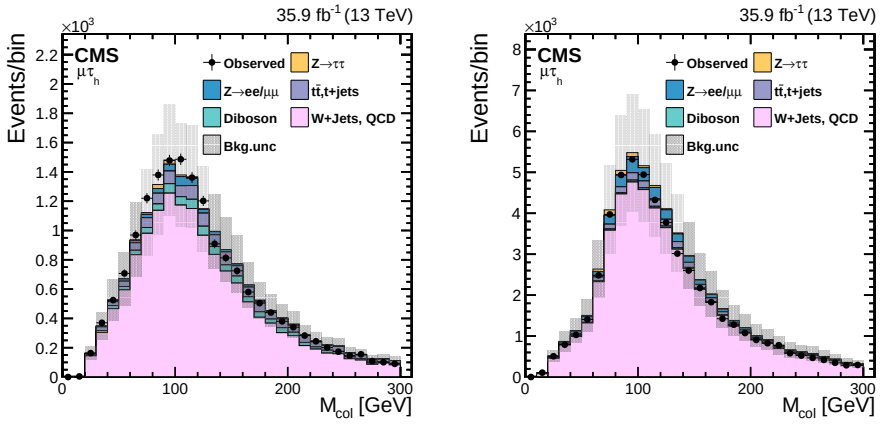
events between these two.

$$N(\text{overlap}) = \frac{f_\tau \cdot f_\tau}{(1 - f_\tau) \cdot (1 - f_\mu)} \cdot N(\text{Region III}) \tag{6.3}$$

$$\tag{6.4}$$

In H $\to \mu\tau_h$ channel, the validation of the misidentified lepton background with full data driven method is performed in a like-sign sample and W+jets enriched control region. The like-sign sample validation requires the events pass the loose selection, the same as the signal region, but with inverted charge. Misidentified background is enriched with the requirement of inverting charge. In the W+jets enriched control region, events that pass the loose selection are further selected with $M_T(\tau) > 60$ GeV and $M_T(\mu) > 80$ GeV. In both of the control regions, misidentified background is the dominant one, as shown in Fig. 6.3. With full data driven method, applied with the $\tau(f_\tau)$ as weight, the like-sign and W+jets

control regions show good data and MC agreement.




(a) Wjets enriched control region (b) Like-sign control region

Figure 6.3. The validation of full data driven method for the misidentified lepton background. 30% of uncertainty is assumed in both cases.

#### 6.0.1.2 Other backgrounds

Besides the misidentified background, which is composed by W+jets and QCD multi-jets background and estimated with data driven method, $Z \to \tau\tau$ and $Z \to \mu\mu$ are also major background in this analysis. In $Z \to \tau\tau$, the muons arise from a $\tau$ decay, while in $Z \to \mu\mu$, one of the $\mu$ can be misidentified as a $\tau$. In other smaller backgrounds, the $\mu\tau$ pairs can be produced from the weak decays of quarks and vector bosons. These smaller backgrounds include standard model Higgs production($H \to \tau\tau, WW$),$t\bar{t}$ ,single top quark, di-boson(WW,WZ,ZZ). A full list

of background used is in the Table. 4.1. These backgrounds are all estimated with MC simulation.

In the $Z \rightarrow \mu\mu$ background, an additional scale factor is applied to the tau leptons that do not come from real taus. This is checked in the MC generator level. The scale factor is applied in term of $|\eta|$ as shown in Table. 6.2.

TABLE 6.2

Scale factor to the events in which a muon is misidentified as a tau

| $\|\eta\|$ ranges | Correction values |
|---|---|
| 0-0.4 | 1.263 |
| 0.4-0.8 | 1.364 |
| 0.8-1.2 | 0.854 |
| 1.2-1.7 | 1.712 |
| 1.7-2.3 | 2.324 |

### 6.0.2 Background estimates for H $\rightarrow e\tau_h$

#### 6.0.2.1 Misidentified Leptons

Misidentified lepton background is a very important one in H $\rightarrow e\tau_h$ channel. This background mainly raises from W+jets and QCD multijets. A background

event passes the signal selection, may have one or both leptons are misidentified, which are faked by other objects. In H $\rightarrow e\tau_h$ 8 TeV analysis, both misidentified electron and tau leptons are considered.

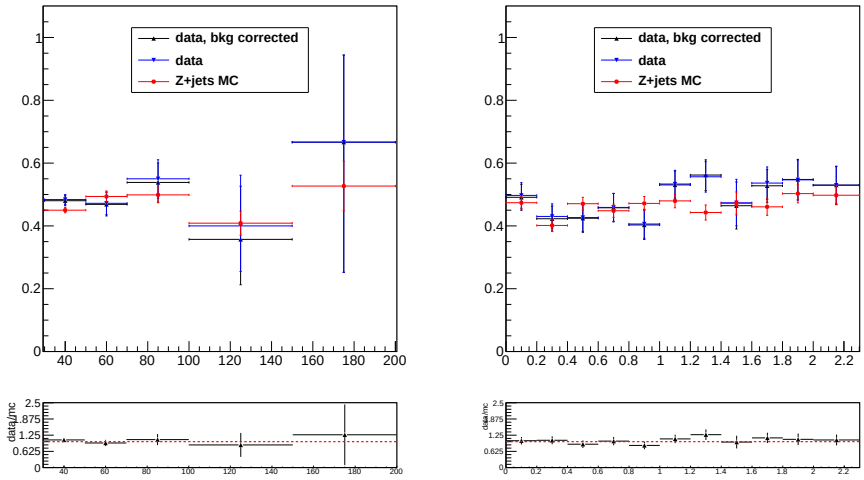
Similar to H $\rightarrow \mu\tau_h$ channel, the misidentified background in H $\rightarrow e\tau_h$ is estimated with full data driven method, with independent data sets Z+jets, in which Z bosons decay into two electrons. Data sets are divided into four control regions for this estimation as in Table. 6.1. Region II is used to extra misidentified background in signal region, Region I, while Region II and Region IV are used for the validation of the fake rate. In the Z+jets data set selection, trigger HLT_Ele27_WP80 is used. The two electrons in Z+jets are required to have $p_T > 30$ GeV, $|\eta| < 2.3$, tight electron MVA ID and cut based tight PF isolation($I^e_{rel} < 0.1$). The invariant mass formed by the two electrons are required to be in the Z boson mass window. Then the jets in the events are checked if they pass $\tau$ or e selection. Tau leptons should have $p_T > 30$ GeV, $|\eta| < 2.3$, passing tau decay mode finder and discriminator against elections and muons. Electrons pass the same conditions as the elections used in forming the Z boson mass peak, but without the requirement of isolation. The isolation conditions are required in different conditions in the fake ratio estimation. Tau leptons use the cut based tight or loose isolation, while loose or tight MVA based electron isolation have the value $I^e_{rel} < 0.1$ and $I^e_{rel} < 0.2$ respectively. After the selection, the fake ratio $f_\tau$ is defined as the number of tau leptons passing the tight isolation divided the number of tau leptons passing the loose isolation. Similarly the fake ratio $f_e$ is defined as the number of elections passing the loose isolation divided by the number of electron passing the tight isolation as shown in Equation. 6.2. With the fake ratio $f_\tau$ and $f_e$, the electron and tau fake rate $\tau(f_\tau)$ and $e(f_e)$ are estimated according to Equation. 6.1. The

fake ratio for tau leptons and electrons are shown in Figure. 6.4.

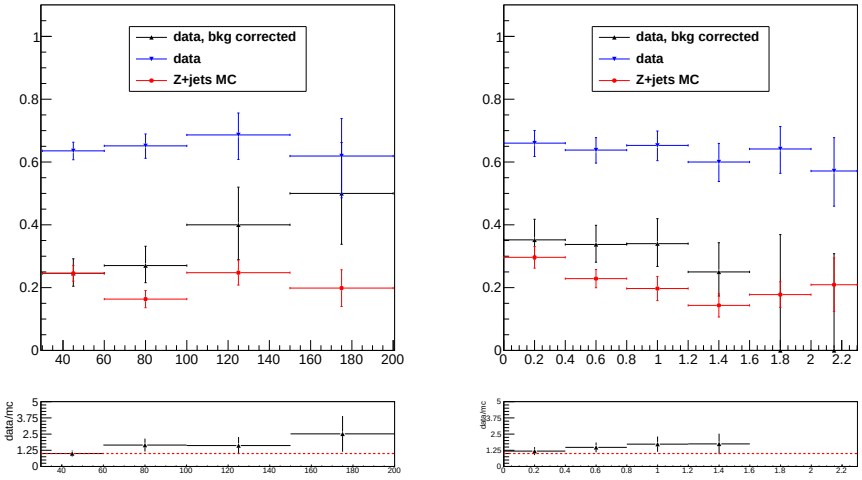
The tau and electron fake ratio is applied in term of $|\eta|$. In the estimation, the background source of three prompt leptons from Diboson samples are deducted. The tau and electron misidentified backgrounds are estimated separately, the overlap events in which both leptons are misidentified objects are estimated with Equation. 6.4. In the equation, N(Region III) in H $\rightarrow e\tau_h$ channel satisfies the condition both leptons are loose isolated but not tight isolated. The validation of full data driven method is done with Region II and Region IV. The distribution of the $M_{col}$ is checked in the same sign region of the two leptons, shown in Figure. 6.5, in which observed data agrees well with the events from MC and misidentified background from data driven method.

#### 6.0.2.2 $Z \rightarrow \tau\tau$

$Z \rightarrow \tau\tau$ in another significant background in H $\rightarrow e\tau_h$ channel. In the case one of the tau leptons decay to a electron and the other hadronically decay. The background is estimated with the embedding technique[9], which starts by selecting the $Z \rightarrow \mu\mu$ events from collision data. The muons pass loose muon selection and then are replaced by simulated $\tau$ leptons and reconstructed with PF algorithm. The key advantage of this technique is that the major topologies like missing energy, jet multiplicity, underlying events are directly from data. Besides tau, the other event records and detector responds are from real environment. When added to the background, the embedded data is normalized to MC prediction. Compared with MC prediction, the visible mass peak from embedding sample shifts 2%. This is mainly because of the difference in final-state radiation of photons between muon and tau and is corrected for. Identification and isolation




(a) $f_\tau$ vs $p_T$ (b) $f_\tau$ vs $|\eta|$




(c) $f_e$ vs $p_T$ (d) $f_e$ vs $|\eta|$

Figure 6.4. $f_\tau$ and $f_\tau$ ratio for the fake rate calculation shows in term of $p_T$ and $|\eta|$

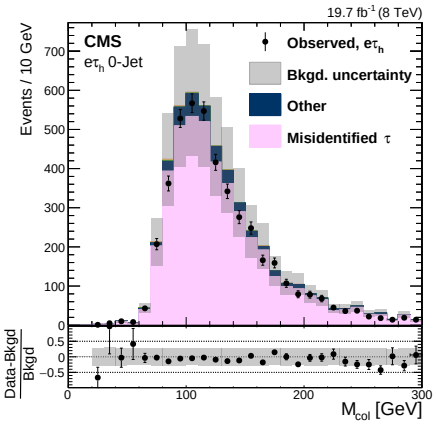



Figure 6.5. Data driven method validation in Region II with e and $\tau$ leptons in same sign of charge

correction for this shift to embedded sample are obtained by comparing with MC simulations.

#### 6.0.2.3 Other backgrounds

$t\bar{t}$ is one of the main background. Leptonic decay W bosons can produce opposite sign lepton pair and missing energy. The shape of this background is obtained from MC samples and normalized to $t\bar{t}$ control region sample. The evens in $t\bar{t}$ control samples are required to pass the 2-jet base selection in H $\rightarrow$ $e\tau_h$ channel and at lease one the jets is identified as a b jet.

The other backgrounds presented in the samples are Z+jets in which one of the electrons is misidentified as a $\tau$ lepton. Di-boson(WW,WZ,ZZ) events, single Top, SM Higgs boson production($H \rightarrow \tau\tau$) and W$\gamma^{(\star)}$. These backgrounds are all estimated from MC simulation.

# CHAPTER 7

## Signal extraction and systematics

### 7.1 Boosted decision trees method

Boosted decision trees(BDT) is used in the H $\rightarrow \mu\tau_h$ analysis. With this MVA method, the final result, which is referred as BDT fit analysis, improved a factor of two with respect to the $M_{col}$ fit analysis. There are two parts within BDT, the boosting algorithm and the decision tree. AdaBoost(adaptive boost) is the main boosting algorithm focused on.

Decision tree is a simple two dimensional tree structure. BDT takes in one signal and one background data sets, which are further divided into training and test samples. A set of selected variables are used the training. At the initial node of the tree structure, events are ordered by the value of variables. Trying with every variable, the node is split into two branches by a binary selection of the variable that gives the best separation between signal and background in each of the branch. The splitting continues until the training reaches the number of layers or the number of events in the branch is smaller than the set limit. The splitting stops and the last notes are called leaves. Depending on the main population inside each leaf, the leafs are categorized as signal leaves or background leaves. An example of the tree structure is shown in Figure. 7.1. A criterion is needed for the selection of a variable and the exact point of the cut at the node or branch.

Purity is defined as in Equation. 7.1, in which the superscript s stands of signal events, b stands of background events and W stands of event weight. With the definition of purity, a couple of criteria can be configured. Here the one called Gini, defined in Equation. 7.2, is selected also it is the default one for BDT in TMVA package. In every splitting, the variable value that maximize the criterion in Equation. 7.3 is selected.

$$P = \frac{\sum_s W_s}{\sum_s W_s + \sum_b W_b} \tag{7.1}$$

$$Gini = (\sum_{i=1}^{n} W_i) P(1-P) \tag{7.2}$$

$$Criterion = Gini_{n-1} - Gini_{n\ left} - Gini_{n\ right} \tag{7.3}$$

Decision tree method is powerful, but the performance suffers from fluctuations. The selection of two variables with similar selection quality may be affected by a small change in the training sample. One of the solution is introducing the boosting algorithm. The boosting algorithm not only help stabilize decision tree training with respect to the small fluctuations inside the training data, but also enhance the classification. AdaBoost is one of the boosting algorithms. With AdaBoost, after the first decision tree training, the misclassified events get higher even weights and as input into the second tree. Typically the AdaBoost training 1000 to 2000 trees. Misclassified event weights depend on the training error of

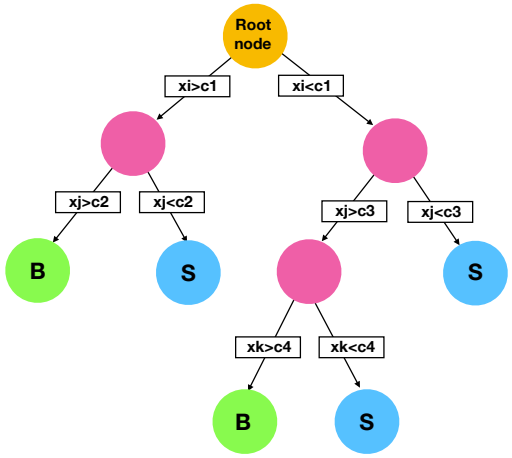



Figure 7.1. Tree structure example in BDT

each decision tree. Training error is calculated as in Equation. 7.4, in which the superscript m, w are the tree label and event weight respectively. $y_i$ indicates the ith event type, 1 for signal and -1 for background. $T_m(x_i)$ indicate the what kind of leaf event i with variable x lands on, 1 for signal leaf and -1 for background leaf. Variable $I(y_i \neq T_m(x_i))$ inside the Equation. 7.4 constructs on the variables $y_i$ and $T_m$ equal 1 if $y_i \neq T_m(x_i)$ or 0 if $y_i = T_m(x_i)$. As shown in Equation. 7.5, with the intermediate quantity $\alpha_m$ for the tree m, the weight for event i is updated to its new value. While $\text{error}_m$ is required to be less than 0.5 so as the event weights are updated to the right direction. Learning rate parameter $\beta$ can be used to adjust the step of each re-weighting. The default value of $\beta$ is 1. Event weights in each tree is renormalized to keep the summed weights constant. The final score of each event after the boosting and training processes is obtained with Equation. 7.8. High score indicates a signal like event while low score indicates a background ground like event.

$$\text{error}_m = \frac{\sum_{i=1}^{N} w_i I(y_i \neq T_m(x_i))}{\sum_{i=1}^{N} w_i} \tag{7.4}$$

$$\alpha_m = \beta \times \ln((1 - \text{error}_m)/\text{error}_m) \tag{7.5}$$

$$w_i \rightarrow w_i \times e^{\alpha I(y_i \neq T_m(x_i))} \tag{7.6}$$

$$w_i \rightarrow w_i / \sum_{i=1}^{N} w_i \tag{7.7}$$

$$T(x) = \sum_{m=1}^{N_{tree}} \alpha_m T_m(x) \tag{7.8}$$

## 7.2 statistical methods

## 7.3 Systematics

Systematic uncertainties originated from different sources that experimentally or theoretically affects the physics results. These uncertainties are considered by the effects on the normalization and shape of the distribution of different processes.

### 7.3.1 Systematics used in H $\rightarrow \mu\tau_h$

The systematic uncertainties considered in this analysis are summarized in Table. 7.1 and Table. 7.2. The uncertainty on muon trigger, identification and isolation together amounts to 2% and the hadronic tau lepton efficiency amounts to 5%. The uncertainties on lepton selections that include trigger, ID, isolation efficiencies

are estimated Tag and probe method with Z boson data sets [6, 11, 21, 22, 24]. The systematic uncertainty of b tagging veto is taken from the uncertainty of b tagging veto efficiency measurement, which adjusts b tagging veto performance in simulation to match data samples. The exact values used in each category of $t\bar{t}$ and single top samples are in Table. 7.3. The uncertainties on $Z \to \tau\tau$, WW, WZ, ZZ, $t\bar{t}$, single top backgrounds mainly come from the uncertainties on the cross section measurement of each process. The normalization uncertainty of $\mu$ fake $\tau$ scale factor measurement and the uncertainty on $Z \to \mu\mu$ process together amounts to 25%. An additional shape uncertainty of muon misidentifying tau is extracted from the measurement of the scale factor and treats independently of hadronic tau decay modes. $Z \to \mu\mu$ is the main source that contributes to $\mu$ fake $\tau$. The uncertainty of the misidentified lepton background is estimated from the matching of the same sign control region which is defined as region II in Table. 6.1. The 30% on normalization uncorrelated between each category and additional 10% partially correlated within each category is a conservative estimation. Only the tau shape uncertainties are considered, which are taken from the variations of the misidentified tau parameters in fitting. Muon misidentified shape uncertainty is omitted as it would be negligible compared with the bin-by-bin uncertainty applied. The shape uncertainties of jet energy scale is estimated by varying the fitting parameters of different sources that affect this scale by one $\sigma$. Tau energy scale is treated as shape uncertainties and each of the tau decay modes is considered as independently. Unclustered energy refers to the energy loses from the jets $p_T < 10$ GeV and the PF candidates that are not included in the jets reconstruction. The unclustered energy uncertainty is considered independently for charged particles, photons, neutral hadrons and very forward particles.

The uncertainties on Higgs boson cross section is affected by the renormalization scales, factorization, parton distribution functions(PDF) and the strong coupling constant($\alpha_s$). These affections in normalization are taken from [16].The uncertainties on Higgs cross section also affect the acceptance and result in the migration of events between categories. The bin-by-bin uncertainty is applied to account for the statistics uncertainties within each bin. These uncertainties are uncorrected between different bins and categories. The uncertainty of integrated luminosity [1] amounts to 2.5%.

### 7.3.2 Systematic uncertainties in H $\to e\tau_h$

The systematic uncertainties affect the H $\to e\tau_h$ analysis are summarized in Table. 7.4, 7.6 and 7.5, which includes the ones affect the normalization and the ones affects the shape of $M_{col}$ distribution. The uncertainty of the electron measurements, which include the trigger, ID and isolation come from the Tag and Probe measurement with Z boson datasets [5, 24]. The uncertainties from the Z $\to \tau\tau$ is from the uncertainty of the cross section measurement [8] and the $\tau$ identification in the embedded technique. The normalization uncertainty on Z $\to \mu\mu$ amounts to 30% which is the from the measurement of cross section and statistics uncertainty in the yields. An extra 5% shape uncertainty due to the mismeasured energy of the electron reconstructed as $\tau$ in Z $\to$ ee background. The shift in $M_{col}$ distribution is measured by comparison between data and MC simulation. The uncertainty in misidentified tau lepton background is 30% normalization uncertainty correlated between categories and shape uncertainties. The shape uncertainties is obtained by varying the parameters from fitting the fake ra-

TABLE 7.1

Part one of the systematic uncertainties considered in $H \to \mu\tau_h$ analysis. All of the uncertainties listed in the systematic tables are correlated between categories besides the ones after the sign $\oplus$. These are the uncertainties correlated in within each category but independent between categories. The theoretical uncertainties related to the acceptance and migration of events are listed in a range. The negative or positive values indicate a anticorrelated or correlated between categories

| Systematic uncertainty | $H \to \mu\tau_h$ |
|---|---|
| Muon trigger/identification/isolation | 2% |
| Hadronic tau lepton efficiency | 5% |
| b tagging veto | 2.0–4.5% |
| $Z \to \tau\tau$ + jets background | 10%$\oplus$5% |
| $WW, ZZ$ background | 5%$\oplus$5% |
| $t\bar{t}$ background | 10%$\oplus$5% |
| Single top quark background | 5%$\oplus$5% |
| $\mu \to \tau_h$ background | 25% |
| Jet $\to \tau_h, \mu$ background | 30%$\oplus$10% |
| Jet energy scale | 3–20% |
| $\tau_h$ energy scale | 1.2% |

TABLE 7.2

Part two of the systematic uncertainties considered in H $\to \mu\tau_h$ analysis

| Systematic uncertainty | H $\to \mu\tau_h$ |
|---|---|
| $\mu \to \tau_h$ energy scale | 1.5% |
| $\mu$ energy scale | 0.2% |
| Unclustered energy scale | $\pm 1\sigma$ |
| Renorm./fact. scales (g gH) [16] | 3.9% |
| Renorm./fact. scales (VBF and VH) [16] | 0.4% |
| PDF + $\alpha_s$ (g gH) [16] | 3.2% |
| PDF + $\alpha_s$ (VBF and VH) [16] | 2.1% |
| Renorm./fact. acceptance (g gH) | $-3.0\%$ – $+2.0\%$ |
| Renorm./fact. acceptance (VBF and VH) | $-0.3\%$ – $+1.0\%$ |
| PDF + $\alpha_s$ acceptance (g gH) | $-1.5\%$ – $+0.5\%$ |
| PDF + $\alpha_s$ acceptance (VBF and VH) | $-1.5\%$ – $+1.0\%$ |
| Integrated luminosity | 2.5% |

TABLE 7.3

b tagging veto systematic uncertainty in each category

| | 0 jet | 1 jet | 2 jets gg-enriched | 2 jets VBF-enriched |
|---|---|---|---|---|
| $t\bar{t}$ | – | 2.45% | 4.37% | 2.59% |
| $t$ | – | 2.11% | 3.07% | 1.98% |

tio one standard deviation. The uncertainty from the pileup process are estimated by varying the total inelastic cross section by ±5 percentage [7]. The uncertainty from Diboson, single top quark background is taken from the measurement of the cross section. In $t\bar{t}$ background, the uncertainty in 0 jet, 1 jet category are taken from the cross section measurement, in 2 jets category, an extra 33% uncertainty uncorrelated between categories is added due to the statistical uncertainty. The luminosity uncertainty amounts to 2.6%. Jet energy scale(JES) and resolution is measured with $\gamma/\mathrm{Z}+jets$ and dijet data [4]. The uncertainty from JES is applied as a function of $p_T$ and $\eta$. Jet energy resolution(JER) uncertainty is obtained by smearing jets energy as a function of $p_T$ and $\eta$. Both JES and JER affect the shape of $M_{col}$ distribution and are taken as shape uncertainties. The energy of the jets below 10 GeV and PF objets that are not clustered are accounted as the unclustered energy scale, which is also taken as shape uncertainties. The uncertainty of $\tau_h$ energy scale is from the comparing $\mathrm{Z} \to \tau\tau$ distruction between data and MC sample and applied as a shape uncertainty. The theoretical uncertainties considered in the $\mathrm{H} \to e\tau_h$ analysis are listed in Table. 7.6. There are several sources that contribute to the theoretical uncertainties and considered fully correlated between LFV Higgs and SM Higgs production. The uncertainty in Parton distribution function is obtained by counting the yields with different PDFs, CT10 [26], MSTW [25], NNPDF [2] from the recommendation in PDF4LHC [3]. The uncertainty in renormalization and factorization scales are from scaling up and down by a factor of two to their normal values($\mu_R = \mu_F = M_H/2$). The uncertainty in underlying events and parton shower is estimated by checking with different PYTHIA tunes. The theoretical uncertainties are all correlated or anticorrelated which have a minus superscript and are the results of the events migration.

TABLE 7.4: The normalization systematic uncertainties considered in H $\to e\tau_h$ analysis. The uncertainties are correlated between categories besides the ones after $\oplus$. These uncertainies are uncorrelated between categories.

| Systematic uncertainty | H $\to$ e$\tau_h$ | | |
|---|---|---|---|
| | 0-jet | 1-jet | 2-jet |
| Electron trigger/ID/isolation | 1% | 1% | 2% |
| Efficiency of $\tau_h$ | 6.7% | 6.7% | 6.7% |
| Z $\to \tau\tau$ background | 3%$\oplus$5% | 3%$\oplus$5% | 3%$\oplus$10% |
| Z $\to$ ee background | 30% | 30% | 30% |
| Misidentified leptons background | 30% | 30% | 30% |
| Pileup | 4% | 4% | 2% |
| WW, WZ, ZZ+jets background | 15% | 15% | 15% |
| $t\bar{t}$ background | 10% | 10% | 10%$\oplus$33% |
| Single top quark background | 25% | 25% | 25% |
| Luminosity | 2.6% | 2.6% | 2.6% |

TABLE 7.5

The systematic uncertainties that affect the shape of $M_{col}$ distribution.

| Systematic Uncertainty | H $\to$ e$\tau_h$ |
|---|---|
| $Z \to$ ee bias | 5% |
| Jet energy scale | 3%–7% |
| Jet energy resolution | 1%–10% |
| Unclustered energy scale | 10% |
| $\tau_h$ energy scale | 3% |

TABLE 7.6: Theoretical uncertainties that affects the Higgs boson production cross section. These uncertainties are correlated or anticorrelated(with minus sign superscript) between all of the categories.

| Systematic uncertainty | Gluon fusion | | | Vector boson fusion | | |
|---|---|---|---|---|---|---|
| | 0-jet | 1-jet | 2-jet | 0-jet | 1-jet | 2-jet |
| Parton distribution function | 9.7% | 9.7% | 9.7% | 3.6% | 3.6% | 3.6% |
| Renormalization/factorization scale | 8% | 10% | $30^{-}$% | 4% | 1.5% | 2% |
| Underlying event/parton shower | 4% | $5^{-}$% | $10^{-}$% | 10% | <1% | $1^{-}$% |

# CHAPTER 8

## LFV Higgs decay searching results

Results can be extracted after applying the selection criteria and with the full consideration of the relevant systematics. In both H $\rightarrow \mu\tau_h$ and H $\rightarrow e\tau_h$ analysis, a maximum likelihood fit is performed to derive the expected and observed limits. Each category in the analysis is fitted separately and then combined. Systematic uncertainties are used as nuisance parameter in the fittings. The upper limits on signal branching faction with $CL_s$ criterion are set with asymptotic formula.

### 8.0.1 Results in H $\rightarrow \mu\tau_h$ search

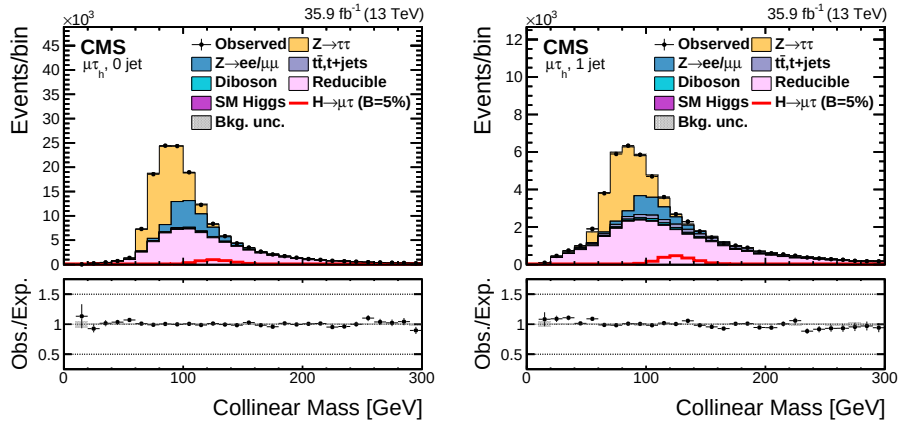
In Lepton flavour violation H $\rightarrow \mu\tau_h$ search at central of mass energy of 13 TeV, the analysis is performed in two parallel searching method, the $M_{col}$ fit analysis and BDT fit analysis. After the selection in $M_{col}$ fit analysis and adjusting all of distribution by the fit, the distribution of the signal and background discriminant variable $M_{col}$ is shown in Figure. 8.3. In the BDT fit analysis, in which the BDT discriminator is used as the variable to distinguish signal from background. After fitting to all of the distributions, the observed data versers the expected backgrounds is shown in Figure. 8.2. No excess over the backgrounds is observed in both of the methods. The expected and observed median upper limits at 95% CL and the best fit branching fraction in the H $\rightarrow \mu\tau_h$ search with the $M_{col}$ fit

analysis is shown in Table. 8.1 and the result of the BDT fit analysis is shown in Table. 8.2.

The search in $H \to \mu\tau$ with CMS 13 TeV, 36 fb$^{-1}$ data is motivated by the search CMS 8 TeV 19.7 fb$^{-1}$ $H \to \mu\tau$, in which a significance of 2.4 standard deviation excess of data with respect to the SM background-only hypothesis was observed. This thesis is written in H $\to \mu\tau_h$ search, together with another search $H \to \mu\tau_e$ [15], where the tau lepton decays electronically to an electron, instead of tau hadronically decay, gives the most recent and update check for the best fit branching ratio $H \to \mu\tau$. This combined result is shown with BDT fit analysis and $M_{col}$ fit analysis separately in Figure. 8.3. No evidence of excess is found in this combined result. Compared with the 8 TeV result, the new upper limits on the signal branching fraction improved a factor of 5. The BDT fit analysis is more sensitive to the $M_{col}$ fit analysis. The new and tighter limit result comes from the BDT fit analysis.

The constrains on Yukawa couplings, $|Y_{\mu\tau}|, |Y_{\tau\mu}|$, can be derived from the limits on branching ratio $\mathcal{B}(H \to \mu\tau)$, which is the combined result between H $\to \mu\tau_h$ search and $H \to \mu\tau_e$ search. Lepton flavour violating decay can rise from the tree level by assuming the Yukawa interactions. In $H \to \mu\tau$ decays, the Yukawa interactions is related to the decay width $\Gamma(H \to \ell^\alpha \ell^\beta)$, $\ell^\alpha$,$\ell^\beta$ can be $\mu$ or $\tau$ and $\ell^\alpha \neq \ell^\beta$. $\Gamma$ in terms of the Yukawa couplings is given by:

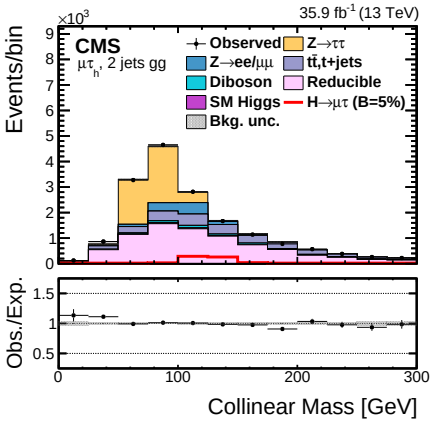
$$\Gamma(H \to \ell^\alpha \ell^\beta) = \frac{m_H}{8\pi}\left(|Y_{\ell^\beta \ell^\alpha}|^2 + |Y_{\ell^\alpha \ell^\beta}|^2\right), \tag{8.1}$$



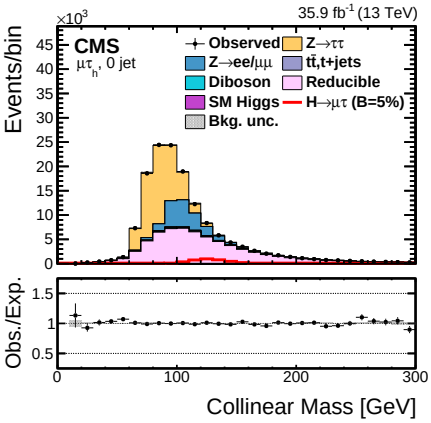

(a) 0 jet




(b) 1 jet

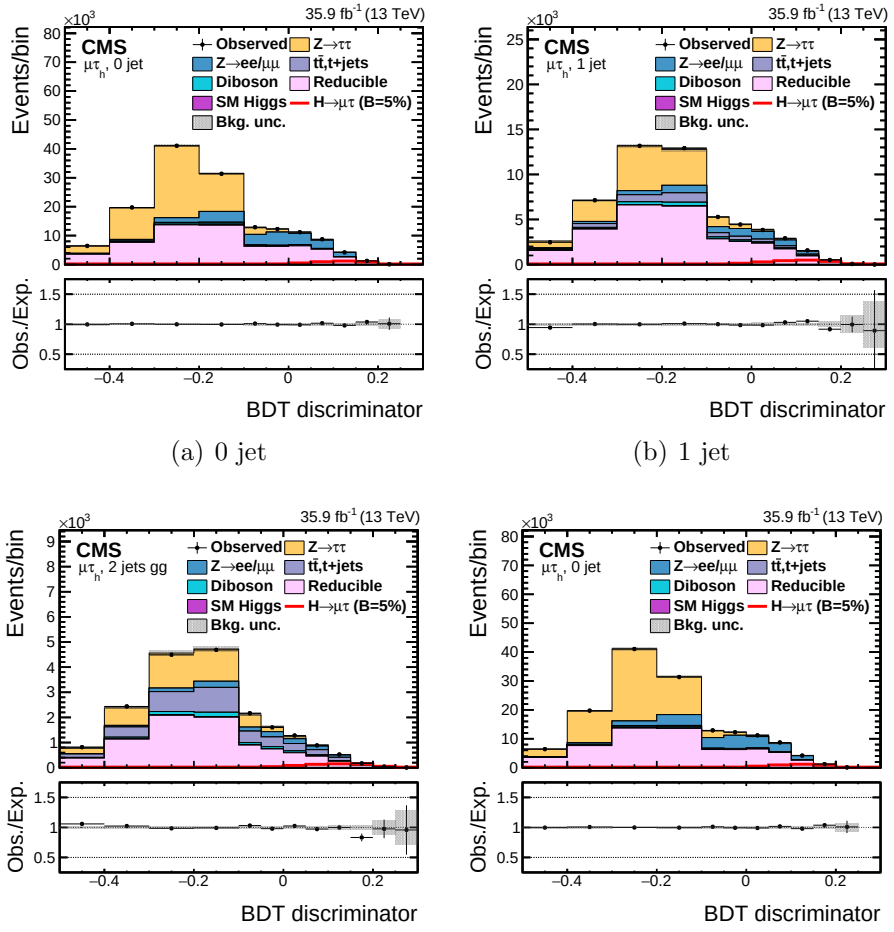



(c) 2 jets, gg-enriched




(d) 2 jets, VBF-enriched

Figure 8.1. $M_{col}$ distribution after the selection. The signal and backgrounds in the plots have been normalized to the best fit values. The gray bands shows the total uncertainties in each bin and the signal is plotted with the branching ratio of 5% of the higgs decay branching ratio for visualization purpose.




(a) 0 jet

(b) 1 jet

(c) 2 jets, gg-enriched

(d) 2 jets, VBF-enriched

Figure 8.2. BDT discriminator distribution after the selection. The signal and backgrounds in the plots has been normalized to the best fit values. The gray bands shows the total uncertainties in each bin and the signal is plotted with the branching ratio of 5% of the higgs decay branching ratio for visualization purpose.

TABLE 8.1

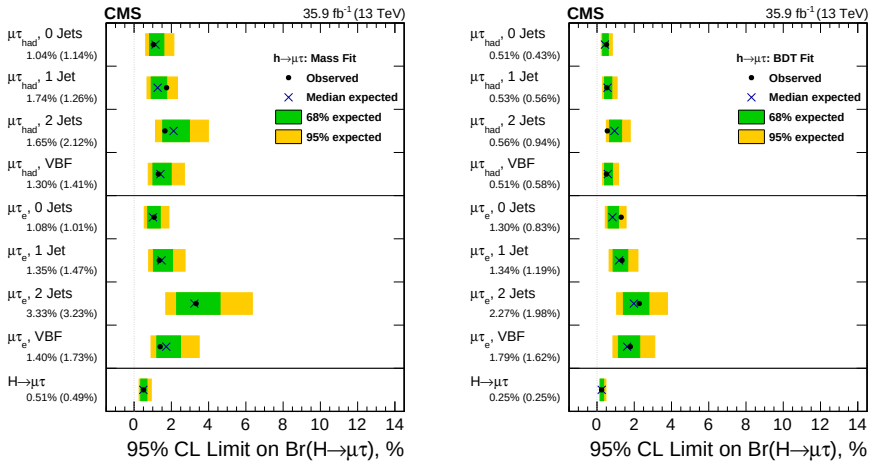
$M_{col}$ fit analysis expected and observed upper limits at 95% CL and the best fit branching fractions in each of the categories.

| Expected limits (%) | | | | | |
|---|---|---|---|---|---|
| | 0-jet | 1-jet | 2-jets | VBF | Combined |
| $\mu\tau_h$ | $< 1.14$ | $< 1.26$ | $< 2.12$ | $< 1.41$ | $< 0.71$ |
| **Observed limits (%)** | | | | | |
| | 0-jet | 1-jet | 2-jets | VBF | Combined |
| $\mu\tau_h$ | $< 1.04$ | $< 1.74$ | $< 1.65$ | $< 1.30$ | $< 0.66$ |
| **Best fit branching fractions (%)** | | | | | |
| | 0-jet | 1-jet | 2-jets | VBF | Combined |
| $\mu\tau_h$ | -0.30 ± 0.45 | 0.68 ± 0.56 | -1.23 ± 1.04 | -0.23 ± 0.66 | -0.08 ± 0.34 |

TABLE 8.2

BDT fit analysis expected and observed upper limits at 95% CL and the best fit branching fractions in each of the categories.

| | Expected limits (%) | | | | |
|---|---|---|---|---|---|
| | 0-jet | 1-jet | 2-jets | VBF | Combined |
| $\mu\tau_h$ | <0.43 | <0.56 | <0.94 | <0.58 | <0.29 |
| | Observed limits (%) | | | | |
| | 0-jet | 1-jet | 2-jets | VBF | Combined |
| $\mu\tau_h$ | <0.51 | <0.53 | <0.56 | <0.51 | <0.27 |
| | Best fit branching fractions (%) | | | | |
| | 0-jet | 1-jet | 2-jets | VBF | Combined |
| $\mu\tau_h$ | $0.12 \pm 0.20$ | $-0.05 \pm 0.25$ | $-0.72 \pm 0.43$ | $-0.22 \pm 0.31$ | $-0.04 \pm 0.14$ |




(a) $M_{col}$ fit analysis (b) BDT fit analysis

Figure 8.3. The expected and observed 95% CL limits on branching ratio of total $H \to \mu\tau$ which shows the results from both $H \to \mu\tau_h$ and $H \to \mu\tau_e$ in the $M_{col}$ fit analysis and BDT fit analysis.

and the branching fraction by:

$$\mathcal{B}(H \to \ell^\alpha \ell^\beta) = \frac{\Gamma(H \to \ell^\alpha \ell^\beta)}{\Gamma(H \to \ell^\alpha \ell^\beta) + \Gamma_{\rm SM}}. \tag{8.2}$$

The 125 GeV Standard Model Higgs decay width is taken as $\Gamma_{\rm SM} = 4.1$MeV [17].In BDT fit analysis, the 95% CL upper limit on the Yukawa coupling is $\sqrt{Y_{\mu\tau}|^2 + |Y_{\tau\mu}^2|} < 1.43 \times 10^{-3}$ and in $M_{col}$ fit analysis, $\sqrt{Y_{\mu\tau}|^2 + |Y_{\tau\mu}|^2} < 2.05 \times 10^{-3}$. The upper limit on Yukawa couplings of BDT fit analysis is shown in Figure. 8.4. The comparison with the previous experimental search and theoretical naturalness results is also shown in the plot.

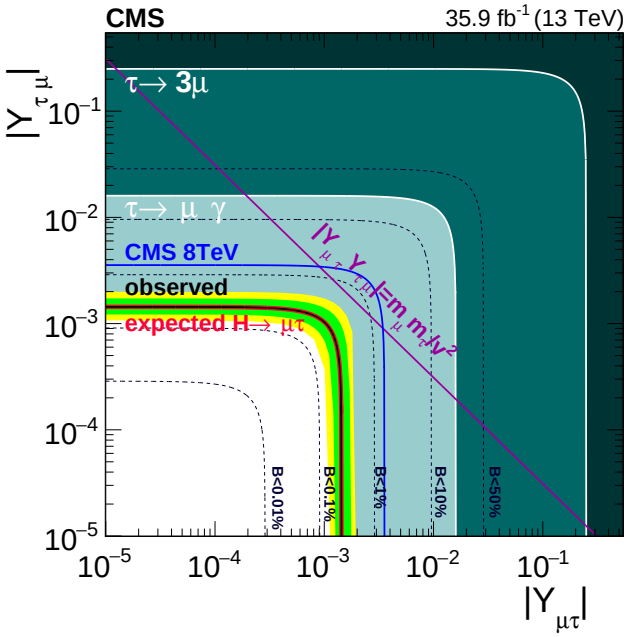
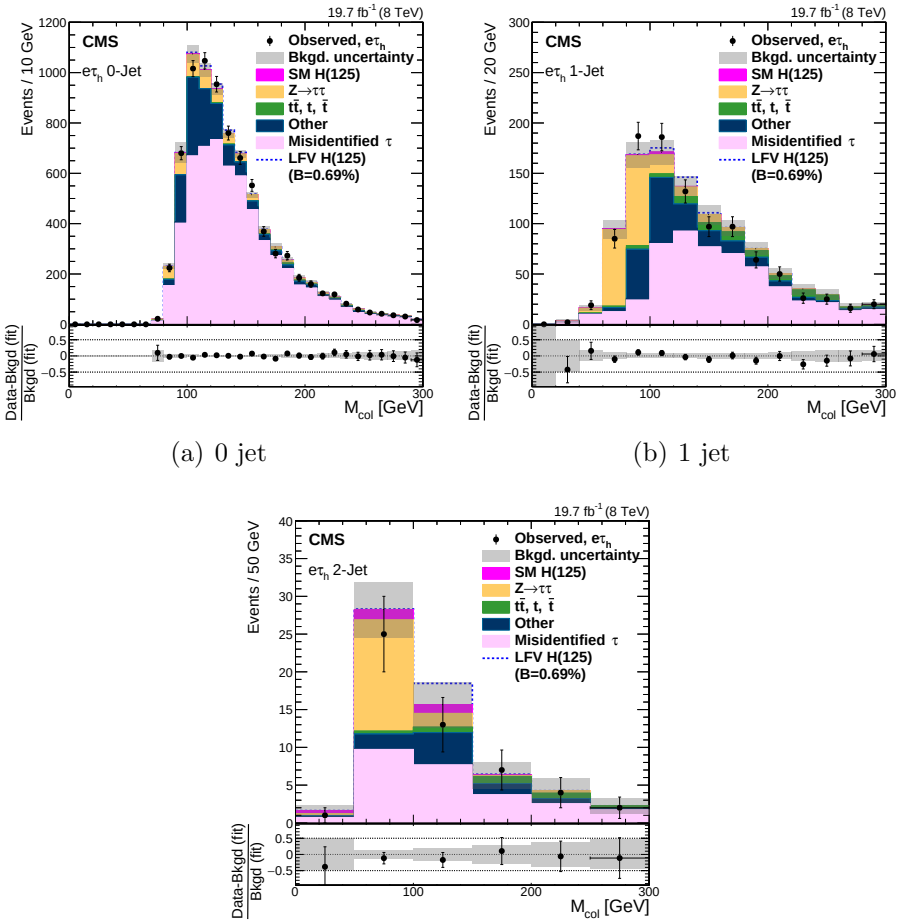



Figure 8.4. Upper limit of flavour violating Yukawa coupling $|Y_{\mu\tau}|, |Y_{\tau\mu}|$ from the BDT fit analysis. The red solid line represents the expected limit and the black solid line is the observed limit. 68% and 95% range of containing the observed limit is shown in the green and yellow band. The result from $\tau \to 3\mu$ is shown in dark green [18, 19, 27] and the one from $\tau \to \mu\gamma$ is shown in lighter green [18, 27]. The theoretical naturalness limit $Y_{ij}Y_{ji} \leq m_i m_j / v^2$ is shown by purple diagonal line [18].

### 8.0.2 Results in H $\rightarrow e\tau_h$ search

The search of H $\rightarrow e\tau_h$ is performed in $M_{col}$ fit analysis with CMS 8 TeV 19.7 $fb^{-1}$. The analysis is binned with 3 categories. After applying the selection cuts and adjusting the signal and background processes by the fit, the fitting variable $M_{col}$ distribution is shown in Figure. 8.5. The number of event yields of both signal and background processes are shown in Table. 8.3. The yields are normalized to the integrated luminosity of 19.7 $fb^{-1}$ and the signal branching ratio is assumed to be 0.69% of the SM higgs production cross section. This thesis is focused and written on the analysis H $\rightarrow e\tau_h$ which is one part of the whole search on lepton flavour violating higgs decay to $e\tau$. Another closely related search is H $\rightarrow e\tau_\mu$ [23]. The results of each categories and combined ones of the two analysis H $\rightarrow e\tau_h$ and H $\rightarrow e\tau_\mu$ on the expected and observed upper signal branching fraction $\mathcal{B}(H \rightarrow e\tau)$ limits is shown in Table. 8.4 and summarized in Figure. 8.6. The combined upper limit on $\mathcal{B}(H \rightarrow e\tau) < 0.69$ is the first direct search on this channel and the tightest limit by the time the result was published. Yukawa coupling between electron and tau leptons is related to decay width $\Gamma$ and branching fraction $\mathcal{B}$ as shown in Equation. 8.1 and 8.2. The 95% CL upper limit on Yukawa coupling from the combined result in H $\rightarrow e\tau$ is $\sqrt{|Y_{e\tau}|^2 + |Y_{\tau e}|^2} < 2.4 \times 10^{-3}$. The limit on $e\tau$ Yukawa coupling from this search and other previous search results are shown in Figure. 8.0.2.




(a) 0 jet

(b) 1 jet

(c) 2 jets

Figure 8.5. $M_{col}$ distribution after the selection. The signal and backgrounds in the plots have been normalized to the best fit values. The gray bands shows the total uncertainties in each bin and the signal is plotted with the branching ratio of 0.69% of the higgs decay branching ratio for visualization purpose.

TABLE 8.3: Events yields of the signal and backgrounds are shown in the mass range 100 GeV $< M_{col} <$ 150 GeV. The branching ratio of H $\rightarrow e\tau_h$ is assumed to be 0.69% of the SM higgs production cross section. The numbers in the table are normalized to the integrated luminosity of 19.7 $fb^{-1}$.

| Jet category: | 0-jet | 1-jet | 2-jet |
|---|---|---|---|
| Misidentified leptons | 3366±25 | 223±11 | 8.7± 2.2 |
| $Z \rightarrow ee, \mu\mu$ | 714±30 | 85±4 | 3.2± 0.2 |
| $Z \rightarrow \tau\tau$ | 270±10 | 32±3 | 1.6± 0.3 |
| $t\bar{t}, t, \bar{t}$ | 10±2 | 13±2 | 0.5± 0.2 |
| $ZZ, WZ, WW$ | 53±2 | 6±1 | 0.3± 0.1 |
| SM H background | 12±1 | 3±1 | 1.0± 0.1 |
| Sum of background | 4425±28 | 363±11 | 15.3±2.3 |
| Observed | 4438 | 375 | 13 |
| LFV H signal | 61±4 | 15±1 | 2.8±0.5 |

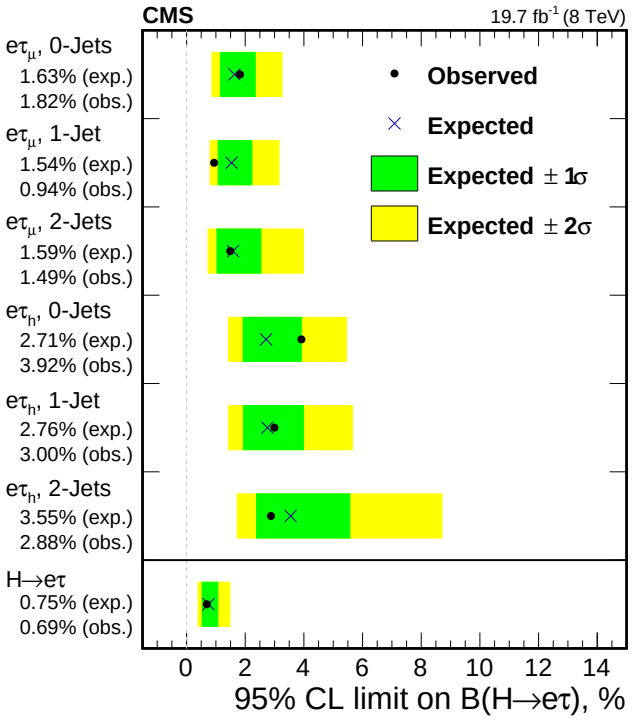



Figure 8.6. 95% CL upper limit on branching ratio on $H \rightarrow e\tau$ with both the results from H $\rightarrow e\tau_h$ and $H \rightarrow e\tau_\mu$ and the combined.

TABLE 8.4

$M_{col}$ fit analysis expected and observed upper limits at 95% CL and the best fit branching fractions in each of the categories of both H $\to e\tau_h$ and $H \to e\tau_\mu$ analysises. The asymmetric one standard-deviation uncertainties around the expected limits are shown in parentheses.

| | 0-jet | 1-jet | 2-jet |
|---|---|---|---|
| Expected limits at 95% CL (%) | | | |
| $e\tau_\mu$ | $<1.63\ \left(^{+0.66}_{-0.44}\right)$ | $<1.54\ \left(^{+0.71}_{-0.47}\right)$ | $<1.59\ \left(^{+0.93}_{-0.55}\right)$ |
| $e\tau_h$ | $<2.71\ \left(^{+1.05}_{-0.75}\right)$ | $<2.76\ \left(^{+1.07}_{-0.77}\right)$ | $<3.55\ \left(^{+1.38}_{-0.99}\right)$ |
| $e\tau$ | $<0.75\ \left(^{+0.32}_{-0.22}\right)$ | | |
| Observed limits at 95% CL (%) | | | |
| $e\tau_\mu$ | $<1.83$ | $<0.94$ | $<1.49$ |
| $e\tau_h$ | $<3.92$ | $<3.00$ | $<2.88$ |
| $e\tau$ | $<0.69$ | | |

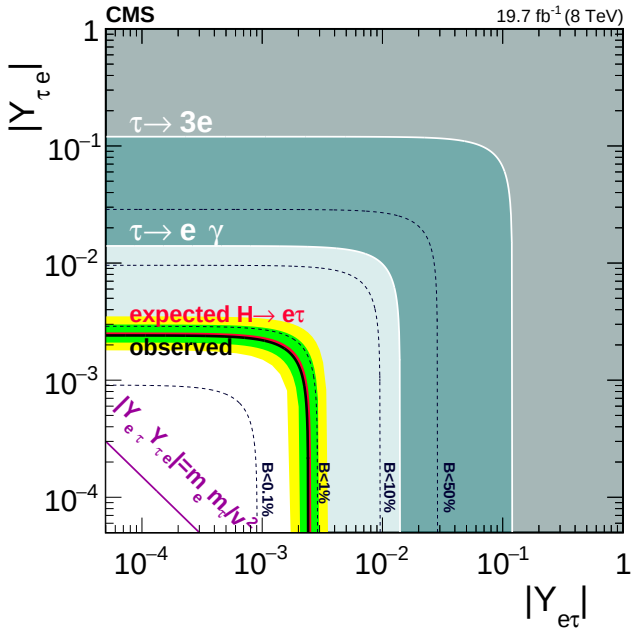



Figure 8.7. Upper limit of flavour violating Yukawa coupling $|Y_{e\tau}|, |Y_{\tau e}|$ from the combined $H \to e\tau$ result. The red solid line represents the expected limit and the black solid line is the observed limit. 68% and 95% range of containing the observed limit is shown in the green and yellow band. The result from $\tau \to 3e$ is shown in gray, the one from $\tau \to e\gamma$ is shown in dark green and presented analysis is in light blue. The theoretical naturalness limit $Y_{ij}Y_{ji} \leq m_i m_j/v^2$ is shown by purple diagonal line [18]

.

# CHAPTER 9

# Conclusion

# BIBLIOGRAPHY


1. CMS luminosity measurements for the 2016 data taking period. CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017. URL `https://cds.cern.ch/record/2257069`.

2. R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali. A first unbiased global NLO determination of parton distributions and their uncertainties. *Nucl. Phys. B*, 838:136, 2010. doi: 10.1016/j.nuclphysb.2010.05.008.

3. M. Botje, J. Butterworth, A. Cooper-Sarkar, A. de Roeck, J. Feltesse, S. Forte, A. Glazov, J. Huston, R. McNulty, T. Sjöstrand, and R. S. Thorne. The PDF4LHC Working Group Interim Recommendations. 2011.

4. S. Chatrchyan et al. Determination of jet energy calibration and transverse momentum resolution in CMS. *JINST*, 6:11002, 2011. doi: 10.1088/1748-0221/6/11/P11002.

5. S. Chatrchyan et al. Measurement of the inclusive W and Z production cross sections in pp collisions at $\sqrt{s} = 7$ TeV with the CMS experiment. *JHEP*, 10:132, 2011. doi: 10.1007/JHEP10(2011)132.

6. S. Chatrchyan et al. Performance of CMS muon reconstruction in pp collision

events at $\sqrt{s} = 7$ TeV. *JINST*, 7:P10002, 2012. doi: 10.1088/1748-0221/7/10/P10002.

7. S. Chatrchyan et al. Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 7TeV$. *Phys. Lett. B*, 722:5, 2013. doi: 10.1016/j.physletb.2013.03.024.

8. S. Chatrchyan et al. Measurement of inclusive $W$ and $Z$ boson production cross sections in $pp$ collisions at $\sqrt{s} = 8$ TeV. *Phys. Rev. Lett.*, 112:191802, 2014. doi: 10.1103/PhysRevLett.112.191802.

9. S. Chatrchyan et al. Evidence for the 125 GeV Higgs boson decaying to a pair of $\tau$ leptons. 2014.

10. C. Collaboration". Performance of $\tau$-lepton reconstruction and identification in CMS. *JINST*, 7(arXiv:1109.6034. CMS-TAU-11-001. CERN-PH-EP-2011-137):P01001. 33 p, Sep 2011. URL `http://cds.cern.ch/record/1385560`.

11. C. Collaboration. Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2. CMS Physics Analysis Summary CMS-PAS-TAU-16-002, 2016. URL `https://cds.cern.ch/record/2196972`.

12. C. Collaboration. Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2. 2016.

13. C. Collaboration. Reconstruction and identification of tau lepton decays to hadrons and tau neutrino at CMS. *JINST 11 (2016) P01019*, 11, 2016. doi: 10.1088/1748-0221/11/01/P01019.

14. C. Collaboration. Identification of b quark jets at the CMS Experiment in the LHC Run 2. Technical Report CMS-PAS-BTV-15-001, CERN, Geneva, 2016. URL `https://cds.cern.ch/record/2138504`.

15. C. Collaboration. Search for lepton flavour violating decays of the Higgs boson to  and e in proton-proton collisions at $\sqrt{s}$= 13 TeV. *JHEP*, 2017.

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

17. A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, and M. Spira. Standard model Higgs-boson branching ratios with uncertainties. *Eur. Phys. J. C*, 71: 1753, 2011. doi: 10.1140/epjc/s10052-011-1753-8.

18. R. Harnik, J. Kopp, and J. Zupan. Flavor violating higgs decays. *JHEP*, 03: 26, 2013. doi: 10.1007/JHEP03(2013)026.

19. K. Hayasaka et al. Search for lepton flavor violating $\tau$ decays into three leptons with 719 million produced $\tau^+\tau^-$ pairs. *Phys. Lett. B*, 687:139, 2010. doi: 10.1016/j.physletb.2010.03.037.

20. A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, H. Voss, M. Backes, T. Carli, O. Cohen, A. Christov, D. Dannheim, K. Danielowski, S. Henrot-Versille, M. Jachowski, K. Kraszewski, A. Krasznahorkay, Jr., M. Kruk, Y. Mahalalel, R. Ospanov, X. Prudent, A. Robert, D. Schouten, F. Tegenfeldt, A. Voigt, K. Voss, M. Wolter, and A. Zemla. TMVA - Toolkit for Multivariate Data Analysis. *ArXiv Physics e-prints*, Mar. 2007.

21. V. Khachatryan et al. Measurements of inclusive W and Z cross sections

in pp collisions at $\sqrt{s} = 7$ TeV. *J. High Energy Phys.*, 01:080, 2011. doi: 10.1007/JHEP01(2011)080.

22. V. Khachatryan et al. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. *JINST*, 10:P06005, 2015. doi: 10.1088/1748-0221/10/06/P06005.

23. V. Khachatryan et al. Search for lepton flavour violating decays of the higgs boson to e$\tau$ and e$\mu$ in proton-proton collisions at $\sqrt{s} = 8$ TeV. *Phys. Lett. B*, 763:472, 2016. doi: 10.1016/j.physletb.2016.09.062.

24. V. Khachatryan et al. Reconstruction and identification of tau lepton decays to hadrons and $\nu_\tau$ at CMS. *JINST*, 11:P01019, 2016. doi: 10.1088/1748-0221/11/01/P01019.

25. A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur. Phys. J. C*, 63:189, 2009. doi: 10.1140/epjc/s10052-009-1072-5.

26. P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, D. Stump, W.-K. Tung, and C.-P. Yuan. Implications of CTEQ global analysis for collider observables. *Phys. Rev. D*, 78:013004, 2008. doi: 10.1103/PhysRevD.78.013004.

27. Particle Data Group, C. Patrignani, et al. Review of Particle Physics. *Chin. Phys. C*, 40:100001, 2016. doi: 10.1088/1674-1137/40/10/100001.