

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

Graduate Program in Physics

Notre Dame, Indiana

August 2018

© Copyright by

Fanbo Meng

2017

All Rights Reserved

# Search for lepton flavor violating decays of the Higgs boson

Abstract

by

Fanbo Meng

A search for lepton flavour violation Higgs decay in the  $\mu\tau$  and  $e\tau$  in which tau leptons decay hadronically are presented. Results are presented for both tau lepton hadronic decay channels and the full results combined with tau lepton leptonic decays. The  $H \rightarrow e\tau_h$  search utilizes the 2012 pp collision datasets at LHC with an integrated luminosity of  $19.7 \text{ fb}^{-1}$  at center-of-mass energy of 8 TeV collected by CMS experiment. No significant excess was observed. The upper limits on branching fraction( $H \rightarrow e\tau$ ) and corresponding Yukawa coupling were set at 95% CL. For the  $H \rightarrow \mu\tau_h$  search, full 2016 pp collision datasets with integrated luminosity of  $35.9 \text{ fb}^{-1}$  at center-of-mass energy of 13 TeV was used. No significant excess was observed and upper limits on branching fraction  $H \rightarrow \mu\tau$  and corresponding Yukawa coupling were set.

To my family

## CONTENTS

FIGURES . . . . .	v
TABLES . . . . .	viii
ACKNOWLEDGMENTS . . . . .	x
CHAPTER 1: Introduction . . . . .	1
CHAPTER 2: Theory . . . . .	4
2.1 Standard model . . . . .	4
2.1.1 Spontaneous symmetry breaking and Higgs mechanism . . . . .	7
2.2 Lepton flavour violation in beyond stand model theories . . . . .	12
2.2.1 Lepton flavour violation Higgs decay in effective field theory	12
2.2.2 LFV in two Higgs models . . . . .	15
CHAPTER 3: LHC and CMS experiment . . . . .	18
3.1 LHC accelerator . . . . .	18
3.2 Compact Muon Solenoid . . . . .	20
3.2.1 Tracker . . . . .	23
3.2.2 Electromagnetic calorimeter . . . . .	25
3.2.3 The hadron calorimeter . . . . .	26
3.2.4 Muon detector . . . . .	28
3.2.5 Trigger . . . . .	29
CHAPTER 4: Datasets . . . . .	32
4.1 Datasets used in LFV analysis . . . . .	32
4.1.1 $H \rightarrow \mu\tau_h$ . . . . .	32
4.1.2 $H \rightarrow e\tau_h$ . . . . .	33
4.2 Event reconstruction . . . . .	33
4.2.1 Particle flow algorithm . . . . .	37
4.2.1.1 Tracking and Calorimeter algorithm . . . . .	37

4.2.2	Muon reconstruction and selection criteria . . . . .	42
4.2.3	Electron identification . . . . .	43
4.2.4	Tau lepton reconstruction . . . . .	47
4.2.5	Jet reconstruction . . . . .	53
4.2.6	Missing transverse momentum . . . . .	56
4.3	Event simulation . . . . .	58
CHAPTER 5: LFV event selection . . . . .		60
5.1	$H \rightarrow \mu\tau_h$ . . . . .	60
5.1.1	Loose selection . . . . .	60
5.1.2	Cut-based analysis . . . . .	62
5.1.3	Multivariate analysis . . . . .	63
5.2	$H \rightarrow e\tau_h$ . . . . .	68
5.2.1	Loose selection . . . . .	68
5.2.2	Cut-based analysis . . . . .	70
CHAPTER 6: LFV analysis background estimation . . . . .		73
6.1	Background estimates for $H \rightarrow \mu\tau_h$ . . . . .	73
6.1.1	Misidentified Leptons . . . . .	73
6.1.2	Other backgrounds . . . . .	78
6.2	Background estimates for $H \rightarrow e\tau_h$ . . . . .	79
6.2.1	Misidentified Leptons . . . . .	79
6.2.2	$Z \rightarrow \tau\tau$ . . . . .	81
6.2.3	Other backgrounds . . . . .	83
CHAPTER 7: Signal extraction and systematics . . . . .		84
7.1	Boosted decision trees method . . . . .	84
7.2	Statistical methods . . . . .	87
7.3	Systematics . . . . .	93
7.3.1	Systematics used in $H \rightarrow \mu\tau_h$ . . . . .	93
7.3.2	Systematic uncertainties in $H \rightarrow e\tau_h$ . . . . .	97
CHAPTER 8: LFV Higgs decay searching results . . . . .		101
8.0.1	Results in $H \rightarrow \mu\tau_h$ search . . . . .	101
8.0.2	Results in $H \rightarrow e\tau_h$ search . . . . .	107
CHAPTER 9: Conclusion . . . . .		114
BIBLIOGRAPHY . . . . .		116

## FIGURES

2.1	Standard model particle group[8]	5
2.2	Production models in LHC	10
3.1	A sketch view LHC injection chain and main experiments associated [72].	21
3.2	A sketch view of CMS detector [29]	22
3.3	The structure of tracker in CMS [29]	24
3.4	ECAL geometrical configuration [47]	26
3.5	A longitudinal view of CMS HCAL sub-detector [29]	27
3.6	A overview of muon chamber configuration in CMS [30]	30
4.1	A sketch view of CMS detecter. Examples are given to show how the particles interact with different sub-detectors.	41
4.2	Electron BDT-based ID shows good discriminating power against background in both EB and EE [69]	46
4.3	PF electron isolation shows better performance in both EB and EE with respect to detector based isolation variable [69]	47
5.1	Expected limits based on an Asimov dataset as a function of $M_T(\tau, MET)$ for the different categories.	64
5.2	Expected limits based on an Asimov dataset as a function of $M_{jj}$ for the 2 jet categories.	66
5.3	Distributions of the input variables to the BDT for the $H \rightarrow \mu\tau_h$ channel.	67
5.4	Expected limits based on an Asimov dataset as a function of $M_T(\tau, MET)$ for the different categories.	68
5.5	Overtraining checking for the BDT training in the TMVA package.	69

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 $\mathcal{B}(H \rightarrow e\tau_h) = 100\%$ in the 0-jet, 1-jet and 2-jet categories, respectively. . . . .	71
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 $ . . . . .	76
6.2	$f_\tau$ ratio for the $\mu$ fake rate calculation shows in term of muon $p_T$ and $ \eta $ . . . . .	77
6.3	The validation of the fully data driven method for the misidentified lepton background. The uncertainties in the distributions include both statistics and systematics uncertainties. . . . .	78
6.4	$f_\tau$ and $f_\tau$ ratio for the mididentification calculation shows in term of $p_T$ and $ \eta $ . . . . .	82
6.5	Data driven method validation in Region II with e and $\tau$ leptons in same sign of charge . . . . .	83
7.1	Tree structure example in BDT . . . . .	86
7.2	Test statistics distribution of signal+background and background only PDFs and the observed valued shown in arrow. . . . .	90
7.3	Signal strength modifier $\mu$ at 95% $CL_s$ distribution from the MC pseudo-data. The right plot is the cumulative distribution of $\mu^{95\%}$ with $\pm 1\sigma$ and $\pm 2\sigma$ bands. . . . .	91
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. . . . .	103
8.2	The 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. . . . .	104
8.3	The expected and observed 95% CL limits on branching ratio of total $H \rightarrow \mu\tau$ which shows the results from both $H \rightarrow \mu\tau_h$ and $H \rightarrow \mu\tau_e$ in the $M_{col}$ fit analysis and BDT fit analysis. . . . .	107

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 $\tau \rightarrow 3\mu$ is shown in dark green [64, 65, 78] and the one from $\tau \rightarrow \mu\gamma$ is shown in lighter green [64, 78]. The theoretical naturalness limit $Y_{ij}Y_{ji} \leq m_i m_j / v^2$ is shown by purple diagonal line [64]. . . . .	108
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. . . . .	110
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. . . . .	112
8.7	Upper limit of flavour violating Yukawa coupling $ Y_{e\tau} ,  Y_{\tau e} $ from the combined $H \rightarrow e\tau$ 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 $\tau \rightarrow 3e$ is shown in gray, the one from $\tau \rightarrow e\gamma$ 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 [64] . . . . .	113

## TABLES

2.1	Mediators in standard model [62] . . . . .	5
2.2	SM Higgs( $m=125$ GeV) production cross sections at $\sqrt{s} = 8$ TeV and 13 TeV. The ttH production is at NLO QCD and NLO electro-weak accuracies while other production models listed here are at NNLO QCD and NLO electro-weak accuracies [1] . . . . .	12
2.3	Four types of 2HDM models differs by the coupling to Higgs doublet fields . . . . .	15
4.1	Monte Carlo samples used in the search, together with their respective cross sections. . . . .	34
4.2	Continue with MC samples used in the analysis. . . . .	35
4.3	Signal and background MC samples . . . . .	36
4.4	Iterative tracking steps take in CMS . . . . .	39
4.5	Muon ID used in the analysis, for the LHC data 2016, running period BCDEF. . . . .	43
4.6	Muon ID used in the analysis, for the LHC data 2016, running period G and H, also the monte Carlo samples. . . . .	44
4.7	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. . . . .	50
4.8	$\tau$ hadronic decay mode hypothesis signatures compatibility tests. $m_\tau$ is required to be in the mass window . . . . .	52
5.1	Selection criteria in each category with the optimization of the $H \rightarrow \mu\tau_h$ analysis . . . . .	65
5.2	Selection criteria for each event category after cut optimization, for the $H \rightarrow e\tau_h$ channel . . . . .	72
6.1	Definition of the samples used to estimate the misidentified lepton background. . . . .	74

6.2	Scale factor to the events in which a muon is misidentified as a tau	79
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 . . . . .	95
7.2	Part two of the systematic uncertainties considered in $H \rightarrow \mu\tau_h$ analysis . . . . .	96
7.3	b tagging veto systematic uncertainty in each category . . . . .	96
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 uncertainties are uncorrelated between categories. . . . .	99
7.5	The systematic uncertainties that affect the shape of $M_{col}$ distribution. . . . .	99
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. . . . .	100
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. . . . .	105
8.2	BDT fit analysis expected and observed upper limits at 95% CL and the best fit branching fractions in each of the categories. . . . .	106
8.3	Events yields of the signal and backgrounds are shown in the mass range $100 \text{ GeV} < M_{col} < 150 \text{ GeV}$ . 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 $19.7 \text{ fb}^{-1}$ . . . . .	109
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 $H \rightarrow e\tau_\mu$ analyses. The asymmetric one standard-deviation uncertainties around the expected limits are shown in parentheses. . . . .	111

## ACKNOWLEDGMENTS

I would first like to thank my adviser Prof.Colin Jessop for the constant guidance, help and support in the past years in both research and everyday life. Thanks for Colin's advises step by step towards being a qualified researcher and great helps and understanding when I was in difficulties. Together with choosing University of Notre Dame, I will never regret. I would like also thank the help and accompanies from my friends at Notre Dame and CERN. Thanks for making my life colorful. Thanks for the high energy group provides me the opportunity working at CERN. CMS is a great collaboration with great research environments, researchers and collaborators, where I have never experienced anywhere better.

Finally I would like to thank the love and supports from my family.

## CHAPTER 1

### Introduction

The goal of elementary particle physics is to understand the fundamental particles and the interactions among them. Based in the microscope world, particle physics attempts to construct theories that try to understand phenomena, make predictions and finally come to a consistent and unified theory.

The Standard Model(SM) provides the modern understanding of particle physics. It incorporates the knowledge from decades of research. Besides gravity, the SM describes interactions among elementary particles that includes three out of four fundamental forces. The last missing piece, predicted by the SM, the Higgs boson was discovered in the Large Hadron Collider by both of the general purpose detectors, the CMS and ATLAS detector in 2012. The Higgs boson is a unique and crucial component of the SM. The Higgs boson is incorporated in the SM through the Higgs mechanism, which provides masses to fermions while keeping the theory renormalizable under local gauge symmetry. If viewing the Higgs mechanism within a bigger picture, the spontaneous symmetry breaking of  $SU(2)$  group, it is also related to providing the masses to other gauge bosons that mediate the interactions. The discovery of the Higgs boson is a great achievement both theoretically and experimentally. Further studies about the properties of the newly discovered boson show good agreement with the predictions from the SM. Besides the great success of the SM, there are still a number of questions that beyond the

scope of the SM, for example, failing to include gravity and describing dark matter that shows its signature in the astronomy observations. New physics models are needed to understand these questions beyond the SM.

The newly discovered Higgs boson can be a great probe to the physics beyond the SM, which in turn can test and inspire new physics models. In the SM, each flavour of leptons in the three generations is assigned a unique lepton number that is conserved within each family. The violation of this conservation which is referred as lepton flavour violation(LFV) is forbidden in the SM but allowed in many new physics scenarios. Thus, any observation of LFV will be a clear signature of new physics. The discovery of the Higgs boson can open a new window towards this topic. Before LHC era, direct searches for LFV Higgs decay, for example  $H \rightarrow \mu\tau$  instead of the SM process  $H \rightarrow \tau\tau$  or  $H \rightarrow \mu\mu$ , was impossible. The upper limits on  $B(H \rightarrow \mu\tau)$  and  $B(H \rightarrow e\tau)$  of  $O(10\%)$  [20, 63] from the indirect searches like  $\tau \rightarrow \mu$  and  $\tau \rightarrow e$  [28]. In these indirect searches, the LFV Higgs decays can only occur in the processes involving loop diagrams and the LFV Higgs decays to  $\mu\tau$  and  $e\tau$  from indirect searches leave plenty of room to further probe this topic. While for  $H \rightarrow e\mu$ , there is already a strong constrain from  $\mu \rightarrow e\gamma$  at  $B(H \rightarrow e\mu) < O(10^{-9})$  [18]. The CMS experiment published the first direct search on  $H \rightarrow \mu\tau$  with integrated luminosity of  $19.7 \text{ fb}^{-1}$  at center-of-mass energy of 8 TeV. A  $2.4\sigma$  excess was observed and the best fit branching ratio was found to be  $B(H \rightarrow \mu\tau_h) = (0.84^{+0.39}_{-0.37})\%$  at 95% CL [5]. A following up search from ATLAS experiment showed the upper limits in  $H \rightarrow \mu\tau$  channel as  $B(H \rightarrow \mu\tau) = (1.24^{+0.50}_{-0.35})\%$  at 95% CL [9].

This dissertation focuses on the first direct search in  $H \rightarrow e\tau_h$  with integrated luminosity of  $19.7 \text{ fb}^{-1}$  at 8 TeV and the search in  $H \rightarrow \mu\tau_h$  with integrated

luminosity of  $35.9\text{ }fb^{-1}$  at 13 TeV with the CMS detector. The dissertation is organized as following. A brief summary of the SM, specifically the electro-weak sector, the Higgs mechanism and LFV in beyond the SM theories are presented in Chapter 2. Following the theory chapter is the one that describes LHC and CMS experiment in general and sub-detectors in CMS. Chapter 4 includes the information about datasets used in the analyses, which incorporate the data collected from the experiment and Monte Carlo(MC) simulations, also the re-construction procedure of the basic objects used in the analysis, like muon, tau and jet. The selections used in  $H \rightarrow \mu\tau_h$  and  $H \rightarrow e\tau_h$  analysis are described in Chapter 5. In Chapter 6, a detailed description of the background processes considered in analyses is presented. In chapter 7, the event selection and signal extraction is described. In Chapter 8, the results from  $H \rightarrow \mu\tau_h$  and  $H \rightarrow e\tau_h$  analysis are summarized and presented. In chapter 9, the results are interpreted in terms of LFV Yukawa couplings.

## CHAPTER 2

### Theory

#### 2.1 Standard model

The discovery of the Higgs boson in 2012 was the last missing piece in SM. The particles in the SM are shown in Figure. 2.1. Besides the Higgs boson, there are three generations of leptons and quarks and four gauge bosons. Leptons and quarks are fermions that compose the observable matter currently known in the universe, while gauge bosons serve as the mediators of the interactions.

The SM is a gauge theory, which describes three natural forces, the strong, electromagnetic and weak force by symmetry group  $SU(3)_C \times SU(2)_L \times U(1)_Y$ . The electroweak sector is described by the  $SU(2)_L \times U(1)_Y$ . The L in the subscript of  $SU(2)_L$  stands for the left chiral signature of the weak interaction and the Y in  $U(1)_Y$  stands for the weak hypercharge. The strong interaction is described by  $SU(3)_C$ , in which the C stands for the colors of quarks. Photons, gluons and  $W^\pm$ , Z bosons are the mediators of electromagnetic, strong and weak force respectively. The force of gravity is not included in the SM. Each particle in Figure. 2.1 has its anti-particle or the anti-particle is the particle itself. The characters of the forces and mediators are summarized in Table. 2.1

The fermion fields in SM can be further categorized. In weak interaction, left-handed fermions are isodoublets, while right-handed fermions are isosinglets.

# Standard Model of Elementary Particles

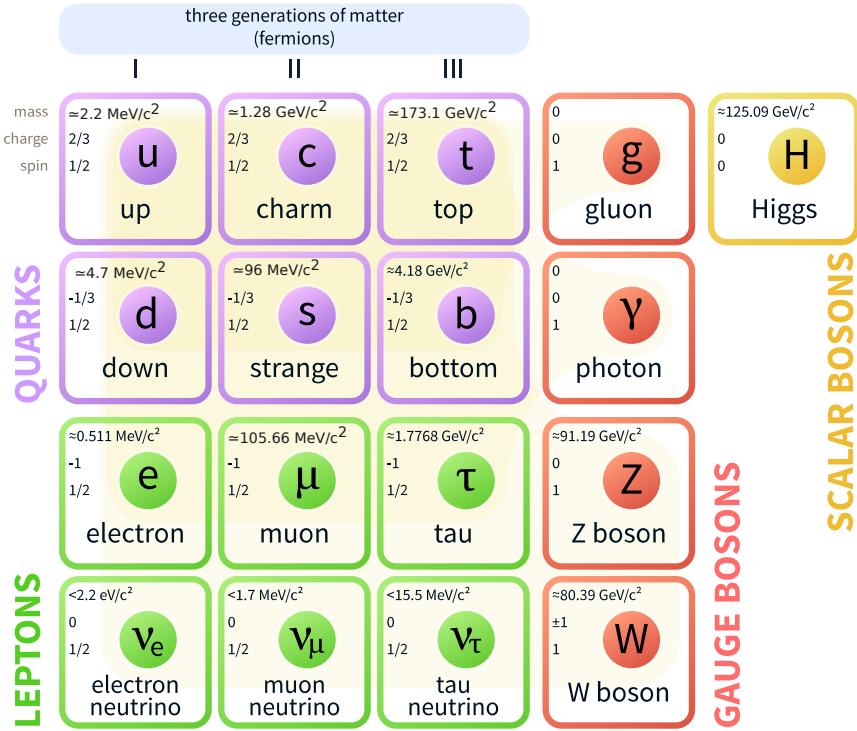


Figure 2.1. Standard model particle group[8]

TABLE 2.1

Mediators in standard model [62]

Force	Mediator	Charge	Mass(GeV)	Range(m)
Strong	g(8 gluons)	0	0	$10^{-15}$
Electromagnetic	$\gamma$ (photon)	0	0	$\infty$
Weak	$W^\pm$	$\pm 1$	80.379	$10^{-18}$
	Z	0	91.1876	$10^{-18}$

Only left-handed fermions or right-handed anti-fermions participate in the weak interaction. Weak hypercharge is a combined quantity, which is defined by the electric charge  $Q$  and the third component of the weak isospin  $I_f^3$ ,  $Y = Q - I_f^3$ . The weak interaction, together with Electromagnetic interaction is best described and understood by the electro-weak theory, which will be further discussed in the next section. Quarks are described by  $SU(3)$ , which are color triplets, while leptons are color singlets.

The mediators in the interaction are represented by the gauge boson fields.  $B_\mu$  is associated with symmetry  $U(1)_Y$  and the corresponding generator is  $Y$ .  $W_\mu^{1,2,3}$  are associated to  $SU(2)_L$  symmetry with the generators  $T^a$  ( $a=1,2,3$ ). The  $T^a$  are the  $2 \times 2$  Pauli matrices, while the  $G_\mu^{1,\dots,8}$  are associated with the  $SU(3)_c$  symmetry and the corresponding generators are the Gell-Mann matrices. The field strengths are expressed as

$$\begin{aligned} G_{\mu\nu}^a &= \partial_\mu G_v^a - \partial_v G_\mu^a + g_s f^{abc} G_\mu^b G_v^c \\ W_{\mu\nu}^a &= \partial_\mu W_v^a - \partial_v W_\mu^a + g_2 \epsilon^{abc} W_\mu^b W_v^c \\ B_{\mu\nu} &= \partial_\mu B_v - \partial_v B_\mu \end{aligned} \quad (2.1)$$

The  $g_s$  and  $g_2$  are the coupling constant of  $SU(3)_C$  and  $SU(2)_L$  respectively. With the requirement of local gauge invariable, covariant derivatives are widely used. The covariant derivative in the Lagrangian is a substitution of normal derivative. It specifies the interaction between matter and gauge fields while keeps the Lagrangian invariant under local transformation. An example of a covariant derivative acting on left-handed quarks in Lagrangian is expressed as

$$D_\mu \psi = (\partial_\mu - ig_s T_a G_\mu^a - ig_2 T_a W_\mu^a - ig_1 \frac{Y_q}{2} B_\mu) \psi \quad (2.2)$$

The masses of gauge bosons are generated through spontaneous symmetry

breaking by applying the Higgs mechanism, which also enables one to introduce the fermion masses. If boson mass terms are directly added in the SM Lagrangian, the local symmetry  $SU(2) \times U(1)$  will be destroyed [55]. Local symmetry here refers to the symmetry that smoothly depends on space and time. In SM, the interaction between fermion field and scalars field is through Yukawa interaction. The Lagrangian of SM in  $SU(3)_C \times SU(2)_L \times U(1)_Y$  symmetry without the mass terms and Yukawa interactions is given by

$$L_{SM} = -\frac{1}{4}G_{\mu\nu}^a G_a^{\mu\nu} - \frac{1}{4}W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4}B_{\mu\nu} B^{\mu\nu} + \bar{L}_i iD_\mu \gamma^\mu L_i \\ + \bar{e}_{R_i} iD_\mu \gamma^\mu e_{R_i} + \bar{Q}_i iD_\mu \gamma^\mu Q_i + \bar{\mu}_{R_i} iD_\mu \gamma^\mu \mu_{R_i} + \bar{R}_i iD_\mu \gamma^\mu d_{R_i} \quad (2.3)$$

### 2.1.1 Spontaneous symmetry breaking and Higgs mechanism

In the SM, the electroweak sector follows the symmetry  $SU(2)_L \times U(1)_Y$ , which is spontaneously broken in the  $SU(2)_L$  part. This mechanism plays an important role in giving masses to the mediator bosons in weak interaction and introducing the Higgs boson(Higgs). The interactions of the scalar doublet with fermion fields through the Yukawa interactions are responsible for generating the masses of the fermions in SM. A good example of showing the concept of spontaneous symmetry breaking is with the  $\phi^4$  theory which can be found in [79].

In the SM, spontaneous symmetry breaking is introduced through the scalar doublet:

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

The corresponding terms of the scalar doublet in the SM Lagrangian are expressed as

$$L_s = (D^\mu \phi)^\dagger (D_\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 \quad (2.4)$$

If  $\mu^2 < 0$  and  $\lambda > 0$ , the doublet field  $\phi$  will have a none zero minimum value, which is called the vacuum expectation value(VEV). In SM, the VEV is shown in the following equation and measured to be 246 GeV.

$$\langle 0|\phi|0\rangle = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} \quad \text{with} \quad v = \left( -\frac{\mu^2}{\lambda} \right)^{1/2} \quad (2.5)$$

When the SU(2) symmetry is spontaneously broken, the scalar doublet field  $\phi$  can expand around the VEV at first order together with the Higgs field:

$$\phi = \begin{pmatrix} \theta_2 + i\theta_1 \\ \frac{1}{\sqrt{2}}(v + H) - i\theta_3 \end{pmatrix} = e^{i\theta_a(x)T^a(x)/v} \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}(v + H) \end{pmatrix} \quad (2.6)$$

Taking the unitary gauge, the scalar field transforms as  $\phi \rightarrow e^{-i\theta_a(x)T^a(x)/v}\phi$  and the term  $(D^\mu\phi)^\dagger(D_\mu\phi)$  can be expended as:

$$\begin{aligned} (D^\mu\phi)^\dagger(D_\mu\phi) &= |(\partial_\mu - ig_2 \frac{T_a}{2} W_\mu^a - \frac{i}{2}g_1 B_\mu)\phi|^2 \\ &= \frac{1}{2}(\partial_\mu H)^2 + \frac{1}{8}g_2^2(v + H)^2|W_\mu^1 + iW_\mu^2|^2 \\ &\quad + \frac{1}{8}(v + H)^2|g_2 W_\mu^3 - g_1 B_\mu|^2 \end{aligned} \quad (2.7)$$

$W^\pm$  and Z can be expressed as following by re-grouping the terms in the above equation:

$$W^\pm = \frac{1}{\sqrt{2}}(W_1 \mp iW_2), Z_\mu = \frac{g_2 W_\mu^3 - g_1 B_\mu}{\sqrt{g_2^2 + g_1^2}}, A_\mu = \frac{g_2 W_\mu^3 + g_1 B_\mu}{\sqrt{g_2^2 + g_1^2}}, \quad (2.8)$$

From the terms in Equation 2.7 that involving the mass of  $W^\pm$  and Z boson, these

vector bosons acquire the mass as

$$M_W = \frac{1}{2}vg_2 \quad \text{and} \quad M_Z = \frac{1}{2}v\sqrt{g_2^2 + g_1^2} \quad (2.9)$$

while photon  $A_\mu$  remains massless. The mixing of electromagnetic and weak interaction is often expressed in terms of Weinberg angle or weak mixing angle. The angle  $\theta_W$  is defined as following:

$$\begin{pmatrix} \gamma \\ Z^0 \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix}$$

$$M_Z = \frac{M_W}{\cos\theta_W} \quad \text{and} \quad \cos\theta_W = \frac{g_2}{\sqrt{g_2^2 + g_1^2}} \quad (2.10)$$

The mass of the fermion can also be generated with the interaction between the scalar doublet  $\phi$  and fermion fields. Taking the muon as an example, the interaction term and the term that gives mass to muon in the Lagrangian is shown as following:

$$L_\mu = -\lambda_\mu(\bar{v}_\mu, \bar{\mu}_L)\phi\mu_R = -\frac{1}{\sqrt{2}}\lambda_\mu(v + H)\bar{\mu}_L\mu_R + \dots \quad (2.11)$$

The mass of muon can be expressed as  $M_\mu = \frac{\lambda_\mu v}{\sqrt{2}}$ . Other fermions acquire masses in the same way. The coupling of Higgs and fermions can also be derived from Equation. 2.11 [55].

The production modes of the Higgs boson at LHC are the gluon-gluon fusion(ggH), the vector boson fusion(VBF), the associated production with a vector boson(VH) and the production in association with a pair of top quarks( $t\bar{t}H$ ). The Feynman diagrams for these production modes are shown in Figure. 2.2. In LHC,

proton proton collision, the real collisions are among quarks and gluons. ggH holds the biggest Higgs production cross section, while the VBF follows. The other two are relatively small compared with ggH and VBF. VBF and ggH are the signal production modes considered in the analyses that are presented in the later chapters.

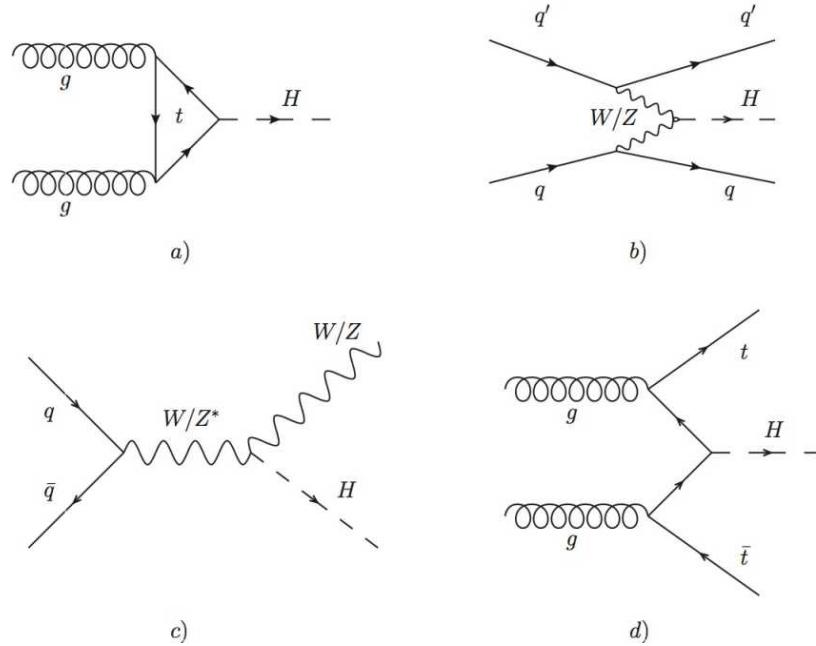


Figure 2.2. Production models in LHC

The cross section of the Higgs boson production modes in LHC have been calculated accurately. Besides the tree level diagrams that are illustrated in the Figure 2.2, high order processes are needed. The inclusive cross sections of the Higgs boson production modes that are used in the 8 TeV and 13 TeV analysis are shown

in Table. 2.2. These cross sections are calculated at NNLO at QCD and NLO electroweak accuracies for all of the production modes, besides ggH at 13 TeV, which is calculated at N3LO QCD and NLO electroweak accuracies. The accuracies here refers to the correction in high order terms, especially the QCD radiative corrections and electroweak correction [53]. The high order correction in ggH and VBF production modes are discussed in the following. In the ggH production mode, the Higgs boson couples to gluons through quark loops and the dominant contributions are from heavy quarks. The corrections in this mode includes vertex corrections, real gluon radiations, initial state rescattering and etc. Dealing with these corrections, the technique like large- $m_t$  limits and full consideration of top and bottom quark mass dependence have been used in the NLO corrections [56]. In the same manner, NNLO correction has been calculated and further improved by resumming the soft-gluon contributions [27]. In 13 TeV, N3LO correction in the ggH inclusive production mode has been computed in the effective theory in which the top-quark is integrated out [48]. The next important correction in the ggH production mode is the electroweak correction with two loop diagrams and the mixed QCD-EW effects have been computed [15]. A most update evaluation of the EW correction can be found in [48]. The uncertainties in the ggH cross section computation has two primary origins, the missing terms in the cross section calculation and the limited knowledge of PDF. In the VBF production mode, the NNLO QCD correction is computed by the structure function approach, which view the process as a double deep-inelastic scattering process and the two vector boson are emitted and merged into Higgs boson [21], while the EW correction is with the Monte Carlo programs [54].

TABLE 2.2

SM Higgs( $m=125$  GeV) production cross sections at  $\sqrt{s} = 8$  TeV and 13 TeV. The ttH production is at NLO QCD and NLO electro-weak accuracies while other production models listed here are at NNLO QCD and NLO electro-weak accuracies [1]

	ggH(pb)	VBF(pb)	WH(pb)	ZH(pb)	ttH(pb)
8 TeV	1.947E+01	1.601E+00	7.026E-01	4.208E-01	1.330E-01
13 TeV	4.414E+01	3.783E+00	1.373E+00	8.839E-01	5.071E-01

## 2.2 Lepton flavour violation in beyond stand model theories

Lepton flavour violating Higgs decays are forbidden in SM. The relevant terms of LFV in SM Lagrangian will not be compatible with renormalization under local gauge symmetry. But beyond Standard Model, lepton flavour conservation does not necessarily hold. In the following, two methods that can introduce LFV are presented, the model independent effective field approach and the two Higgs doublet model. In both cases, the focus is how the LFV Higgs decay shows up in the new theories.

### 2.2.1 Lepton flavour violation Higgs decay in effective field theory

The Effective Lagrangian approach is widely used to explore the new physics with higher dimensional operators in a model independent way. The effective

Lagrangian in the theory can be written as up to dimension six operators:

$$L_{eff} = L_{SM} + \sum_{n,i,j} \frac{a_n^{ij}}{\Lambda^2} O_n^{ij} \quad (2.12)$$

where  $a_n^{ij}$  is the coefficients,  $i, j, (= 1, 2, 3)$  are flavour indices,  $n$  holds the number of the operators and  $\Lambda$  is the new physics scale.

Various effective field operators can introduce processes that violate lepton flavour conservation [52]. As an example, the Yukawa-type operators that generate LFV Higgs decays will be discussed. The Yukawa-type operators hold the following terms:

$$O_{L\phi}^{ij} = (\Phi^\dagger \Phi)(\bar{L}_{L_i} l_{R_j} \Phi) \quad (2.13)$$

While, inside the Lagrangian, the change to SM Lagrangian takes the form:

$$\Delta L_Y = -\frac{\lambda'_{ij}}{\Lambda^2} (\Phi^\dagger \Phi)(\bar{L}_L^i l_R^j \Phi) + h.c... \quad (2.14)$$

Taking in the scalar doublet expansion around the VEV as shown in Equation. 2.6, the  $O(6)$  Yukawa-type operators have the following form:

$$\begin{aligned} \Delta L_Y &= -\frac{\lambda'_{ij}}{\Lambda^2} (\Phi^\dagger \Phi)(\bar{L}_L^i l_R^j \Phi) + h.c... \\ &= -\frac{\lambda'_{ij}}{2\sqrt{2}\Lambda^2} l_L^i l_R^j (v + H)^3 + h.c... \\ &= -\frac{\lambda'_{ij} v^3}{2\sqrt{2}\Lambda^2} l_L^i l_R^j - \frac{\lambda'_{ij} 3v^2}{2\sqrt{2}\Lambda^2} l_L^i l_R^j + h.c... \end{aligned} \quad (2.15)$$

The SM Higgs and lepton coupling components in Equation. 2.11 is obtained through the diagonalization of mass matrices [64]. The total Lagrangian in the effective field theory is a combination of SM Lagrangian  $L_{SM}$  and the effective

field Lagrangian  $\Delta L_Y$ . Thus the combined fermion mass and Yukawa interaction terms are following:

$$\sqrt{2}m = V_L \left[ \lambda + \frac{v^2}{2\Lambda^2} \lambda' \right] V_R^\dagger, \quad \sqrt{2}Y = V_L \left[ \lambda + 3 \frac{v^2}{2\Lambda^2} \lambda' \right] V_R^\dagger \quad (2.16)$$

The Yukawa couplings of Higgs and leptons mixed in the contributions from dimension six operators and have the following form, in which  $\hat{\lambda} = V_L \lambda' V_R$

$$Y_{ij} = \frac{m_i}{v} \delta_{ij} + \frac{v^2}{\sqrt{2}\Lambda^2} \hat{\lambda}_{ij} \quad (2.17)$$

In the limit  $\Lambda \rightarrow \infty$ , the SM results can be recovered, but in general case, like in the electro-weak scale and a arbitrary non-diagonal matrix  $\hat{\lambda}$ , LFV Higgs decays can be introduced into the theory through effective fields.

In the search of LFV Higgs decays, for example the  $H \rightarrow \mu\tau$  channel, the constrains on Yukawa couplings,  $|Y_{\mu\tau}|, |Y_{\tau\mu}|$ , can be derived from the limits on branching fraction  $\mathcal{B}(H \rightarrow \mu\tau)$ . In  $H \rightarrow \mu\tau$  decays, the Yukawa interactions is related to the decay width  $\Gamma(H \rightarrow \ell^\alpha \ell^\beta)$ ,  $\ell^\alpha, \ell^\beta$  can be  $\mu$  or  $\tau$  and  $\ell^\alpha \neq \ell^\beta$ .  $\Gamma$  in terms of the Yukawa couplings is given by:

$$\Gamma(H \rightarrow \ell^\alpha \ell^\beta) = \frac{m_H}{8\pi} (|Y_{\ell^\beta \ell^\alpha}|^2 + |Y_{\ell^\alpha \ell^\beta}|^2), \quad (2.18)$$

and the branching fraction by:

$$\mathcal{B}(H \rightarrow \ell^\alpha \ell^\beta) = \frac{\Gamma(H \rightarrow \ell^\alpha \ell^\beta)}{\Gamma(H \rightarrow \ell^\alpha \ell^\beta) + \Gamma_{\text{SM}}}. \quad (2.19)$$

### 2.2.2 LFV in two Higgs models

In the model with two Higgs doublets(2HDM)  $\Phi_1$  and  $\Phi_2$ , similar to the SM, the Yukawa interaction can be written as:

$$L = y_1 \bar{L} \Phi_1 E + y_2 \bar{L} \Phi_2 E + h.c, \quad (2.20)$$

Here,  $y_1$  and  $y_2$  are Yukawa couplings. If there is not a parity symmetry to distinguish the two Higgs doublets, there can be a coupling of the two doublets to the leptons in the tree level. In general it is impossible to diagonalize  $y_1$  and  $y_2$  simultaneously thus the LFV Higgs decay can be presented in the renormalizable Lagrangian [50].

In 2HDM, there are in general four sub-type of models, the main difference comes from the coupling of Higgs doublets to the quarks and leptons as shown in Table. 2.3.

TABLE 2.3

Four types of 2HDM models differs by the coupling to Higgs doublet fields

Model	$u_R^i$	$d_R^i$	$e_R^i$
Type I	$\Phi_2$	$\Phi_2$	$\Phi_2$
Type II	$\Phi_2$	$\Phi_1$	$\Phi_1$
Type III (lepton specific)	$\Phi_2$	$\Phi_2$	$\Phi_1$
Type IV (Flipped)	$\Phi_2$	$\Phi_1$	$\Phi_2$

The type III 2HDM is more relevant to this dissertation which can have the tree level LFV Higgs decay. The Higgs doublets can have the general form:

$$\Phi_j = \begin{pmatrix} \phi_j^+ \\ (v_j + \phi_j + i\eta_j/\sqrt{2}) \end{pmatrix} \quad (2.21)$$

Under the common assumption of CP conservation in the Higgs sector and not spontaneously broken, the quartic odd terms are eliminated in the potential, then the scalar potential can be expressed as

$$V = m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - m_{12}^2 (\Phi_1^\dagger \Phi_2 + \Phi_2^\dagger \Phi_1) + \frac{\lambda_1}{2} (\Phi_1^\dagger \Phi_1)^2 + \frac{\lambda_2}{2} (\Phi_2^\dagger \Phi_2)^2 + \lambda_3 \Phi_1^\dagger \Phi_1 \Phi_2^\dagger \Phi_2 + \lambda_4 \Phi_1^\dagger \Phi_2 \Phi_2^\dagger \Phi_1 + \frac{\lambda_5}{2} [(\Phi_1^\dagger \Phi_2)^2 + (\Phi_2^\dagger \Phi_1)^2] \quad (2.22)$$

The scalar doublets fields  $\Phi_1$  and  $\Phi_2$  are not physical observables but the mass eigenstates are. So any combination of the scalar doublet fields, as long as it preserves CP and gauge symmetry in SM, produces the same physics results [23].

The following basis is referred to as Higgs basis for the Higgs doublets:

$$H_1 = \begin{pmatrix} G^+ \\ \frac{1}{\sqrt{(2)}}(v + \phi_1^0 + iG^0) \end{pmatrix}, \quad H_2 = \begin{pmatrix} H^+ \\ \frac{1}{\sqrt{(2)}}(\phi_2^0 + iA) \end{pmatrix} \quad (2.23)$$

The relationship between scalar field  $\phi_1^0$  and  $\phi_2^0$  and neutral Higgs mass eigenstates  $h$  and  $H$  is the following:

$$\begin{aligned} h &= \sin(\alpha - \beta)\phi_1^0 + \cos(\alpha - \beta)\phi_2^0 \\ H &= \cos(\alpha - \beta)\phi_1^0 - \sin(\alpha - \beta)\phi_2^0 \end{aligned} \quad (2.24)$$

The angle  $\alpha - \beta$  is the mixing angle between these two groups of scalars. The in-

teractions of Higgs and fermions are through Yukawa coupling. In the Higgs base, the Yukawa interaction terms in the Lagrangian of the 2HDM can be expressed as:

$$\begin{aligned} -L_Y = & \sqrt{2} \left( \bar{q}_{L_j} \tilde{H}_1 \frac{K_{ij}^* m_i^U}{v} u_{R_i} + \bar{q}_{L_i} H_1 \frac{m_i^D}{v} d_{R_i} + \bar{l}_{L_i} H_1 \frac{m_i^E}{v} e_{R_i} \right) \\ & + \bar{q}_{L_i} \tilde{H}_2 \rho_{ij}^U u_{R_j} + \bar{q}_{L_i} H_2 \rho_{ij}^D d_{R_j} + \bar{l}_{L_i} H_2 \rho_{ij}^E e_{R_j} + h.c.. \end{aligned} \quad (2.25)$$

Inside the Lagrangian, the terms denote as following,  $\tilde{H}_i = i\sigma_2 H^{*i}$ ,  $K_{ij}$  as the CKM matrix and  $\rho^{U,D,E}$  are complex matrices in flavor space. Taking in Equation. 2.23 and 2.24, the lepton Yukawa interaction terms are collected as:

$$\begin{aligned} -L_Y = & \bar{e}_i \left( \frac{m_i^E}{v} \delta_{ij} s_{\beta-\alpha} + \frac{1}{\sqrt{2}} \rho_{ij}^E c_{\beta-\alpha} \right) e_j h \\ & + \bar{e}_i \left( \frac{m_i^E}{v} \delta_{ij} c_{\beta-\alpha} - \frac{1}{\sqrt{2}} \rho_{ij}^E S_{\beta-\alpha} \right) e_j H + ... \end{aligned} \quad (2.26)$$

Term  $s_{\beta-\alpha}$  and  $c_{\beta-\alpha}$  stand for  $\sin(\beta-\alpha)$  and  $\cos(\beta-\alpha)$  respectively. The coupling of leptons and Higgs in the 2HDM model typeIII are expressed as

$$\begin{aligned} g_{hff'} &= \frac{m_f}{v} s_{\beta-\alpha} \delta_{ff'} + \frac{\rho_{ff'}}{\sqrt{2}} c_{\beta-\alpha} \\ g_{Hff'} &= \frac{m_f}{v} s_{\beta-\alpha} \delta_{ff'} - \frac{\rho_{ff'}}{\sqrt{2}} c_{\beta-\alpha} \end{aligned} \quad (2.27)$$

So this shows the possibility of LFV Higgs decay at tree level [16].

## CHAPTER 3

### LHC and CMS experiment

The Large Hadron Collider [58] is the most powerful hadron collider ever built. The circumference of this circle superconducting collider is 26.7 km and the designed full operation energy is 14 TeV. In 2012, the LHC operated on 8 TeV and in 2016, it boosted up to 13 TeV. There are four collision locations on the LHC ring which hold four particle detectors, ALICE, ATLAS, CMS and LHCb. The ATLAS and CMS are general purpose detectors, aiming for the high luminosity operation . The LHCb focuses on the b physics study, while ALICE studies the lead-lead collision.

#### 3.1 LHC accelerator

A sketch view of the proton accelerating process is shown in Figure. 3.1. The LHC is the most powerful accelerator in the accelerating chain. Before the beams are injected into the LHC, a series of steps are taken. In the proton-proton(p-p) collision, protons are from the source Duoplasmatron, which uses electric field to break down hydrogen gas into protons and electrons, then the protons are accelerated by a 90 kV supply. Leaving the source, the protons are focused and accelerated to 750 keV by the radio frequency quadrupole. Then the protons are further accelerated by linear accelerator Linac2 to 50 MeV. The proton synchrotron booster further accelerates the protons from 50 MeV to 1.4 GeV and

injects the protons into the proton synchrotron(PS). The PS accelerates the protons to 25 GeV followed by the super proton synchrotron which boosts the protons to 450 GeV. The LHC is the last ring in the whole accelerating process and accelerates the protons to its current final energy 6.5 GeV. In a normal fill of protons in LHC, the ring holds 2808 bunches with an approximation of  $10^{11}$  protons.

There are thousands of superconducting magnets along the LHC ring to bend and focus the beam. Among the magnets, there are 1232 main dipoles which are used to bend the beam with a magnetic field above 8 T. Other types of magnets, for example, the quadrupole magnet can tight the beam either vertically or horizontally, while the sextupole, octupole and decapole can help fine tuning the magnetic field. The radiofrequency cavities(RF) in the LHC are used to accelerate the protons from 450 GeV to 6.5 GeV, keep the bunches in the beam pipe compact and restore the energy loss from synchrotron radiation. Eight RFs per beam, provide 16 MV longitudinal oscillating voltage with the 400 MHz superconducting cavity system.

The machine luminosity is an important parameter of the collider. For a process under study, the number of events created per-second  $N_{\text{event}}$  is shown in Equation. 3.1, in which  $L$  is the machine luminosity and  $\sigma_{\text{event}}$  is the cross section of that process.

$$N_{\text{event}} = L\sigma_{\text{event}} \quad (3.1)$$

The machine luminosity is determined by a number of factors as shown in Equation. 3.2. The  $n_b$  and  $N_b$  are the number of bunches per-beam and the number of protons per-bunch respectively. The  $f_{\text{rev}}$  and  $\gamma_\gamma$  are the revolution

frequency and relativistic gamma factor respectively. The  $\beta*$  is the amplitude function of the beam at the collision point while the function F describes the reduction of luminosity because of the crossing angle. This machine luminosity is also called instantaneous luminosity which is the luminosity at a unite time. The integrated luminosity which later is referred as luminosity is the instantaneous luminosity that integrates over time 3.1.

$$L = \frac{N_b^2 n_b f_{fev} \gamma_\gamma}{4\pi\epsilon\beta*} F \quad (3.2)$$

### 3.2 Compact Muon Solenoid

The compact muon solenoid(CMS) is a general purpose detector. It is a high performance detector that is designed to observe any new physics produced by LHC. It covers a broad range of physic studies like the standard model physics and the search for supersymmetry and dark matter candidates. The CMS detector is designed to have good muon momentum and position resolution over a large range of energy and angles, good charged particle momentum resolution and high identification efficiency within inner tracker, good electromagnetic energy and position resolution, high photon and lepton isolation efficiency in high luminosity condition and good missing transverse momentum and jet energy resolution.

An general view of the CMS detector is shown in Figure. 3.2. The detector is composed of a set of sub-detectors from the inside and out in a ring structure. The main sub-detectors are the followings:

- The inner tracker consists of two parts, the pixel detector and silicon strip

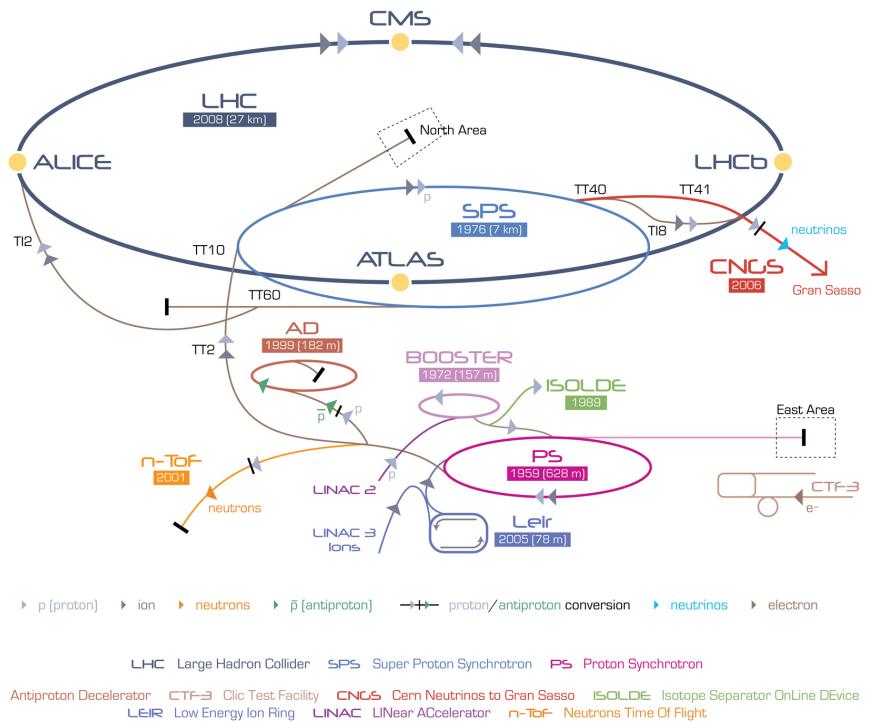


Figure 3.1. A sketch view LHC injection chain and main experiments associated [72].

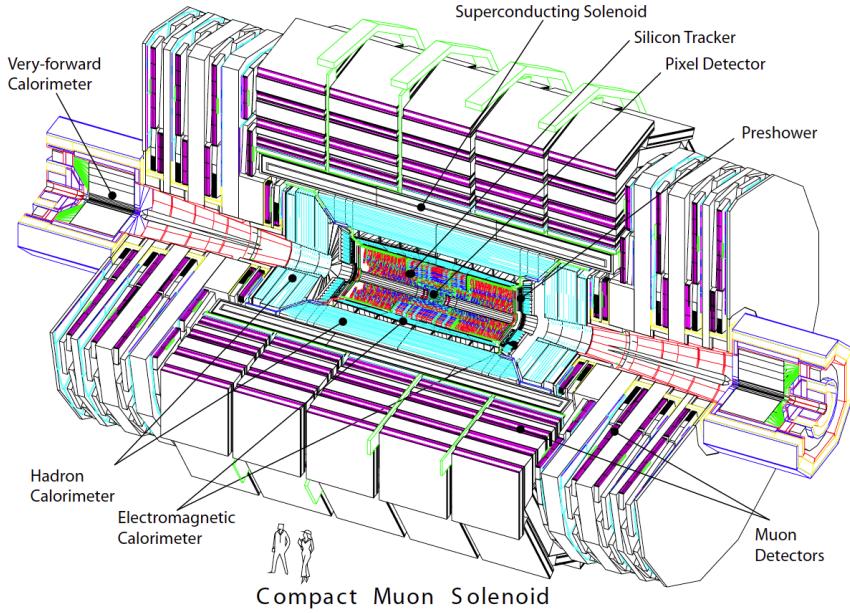


Figure 3.2. A sketch view of CMS detector [29]

tracker which are used measure the momentum and tracks of the charged particles.

- The electromagnetic calorimeter mainly measures the energy and the position of electrons and photons. Other particles will leave some percentage of energy inside while passing through.
- The hadron calorimeter measures the energy of hadrons.
- The muon detector measures the tracks and momentum of the muons.

Another outstanding feature of CMS is the superconducting magnet system, which provides a 3.8 T magnetic field. The configuration of the magnet system drives the design and layout of the detector. Besides the sub-detectors listed above, trigger system is also crucial for the success of the whole program. The

trigger system consists of two parts, the hardware based level one trigger and the software based high level trigger. The trigger system does the initial selection of interesting events from a huge flux of events per-collision, which makes it possible for data-acquisition and recording. The details of the sub-detectors and other systems mentioned will be further discussed later.

In general, the CMS detector is 21.6 m long, 14.6 m in diameter and weighs 12500 tonnes in total. The coordinate system adopted by CMS sets the center at the collision point. The x-axis points towards the center of LHC ring and the z-axis points along the beam direction. The azimuthal angle  $\phi$  measures from the x-axis in the x-y plane, polar angle  $\theta$  measures from the zenith direction and the r is the radial coordinate. Another variable called the pseudorapidity  $\eta$ , defined as  $\eta = -\ln[\tan(\theta/2)]$  is also frequently used in the measurement. In the case, a particle with  $E \gg m$ , pseudorapidity can be approximated as  $\eta = -\ln\left(\frac{E+p_z}{E-p_z}\right)$ , where the  $p_z$  is the longitudinal component of momentum [29].

### 3.2.1 Tracker

The inner tracking system of CMS is designed to measure the trajectories of charged particles. The efficient and precise measurement is crucial for the reconstruction and identification of particles. The LHC operates with the instantaneous luminosity in the order of  $10^{34} cm^{-2}s^{-1}$ , in average producing more than 20 p-p interactions and 1000 particles per-bunch crossing collision. High granularity and fast response are the primary features of the tracker system. To measure the trajectories precisely, low interactions of tracker materials with incoming particles, like multi-scattering, photon conversion and nuclear interaction are also important. In the long operation period, radiation hardness of the tracker material is

needed. All of these factors drive the design of CMS tracker system.

The tracker system of CMS surrounds the collision point and has the dimension of 5.8 m in length and 2.5 m in diameter, with a coverage up to pseudorapidity  $|\eta| < 2.5$ . A overview of the tracker system layout is shown in Figure. 3.3. Three layers of silicon pixel detector modules surround the interaction point and two additional disks of pixel modules on each side, in all 66 million pixels of the size  $100 \times 150 \mu\text{m}$  each. Following the pixel detector is the silicon strip tracker. There are four components of the strip tracker, tracker inner barrel(TIB), tracker inner Discs(TID), tracker outer barrel(TOB) and tracker end caps(TEC). The arrangement of the strip tracker components is shown in Figure. 3.3 which consist of 10 layers and about 10 million strips.

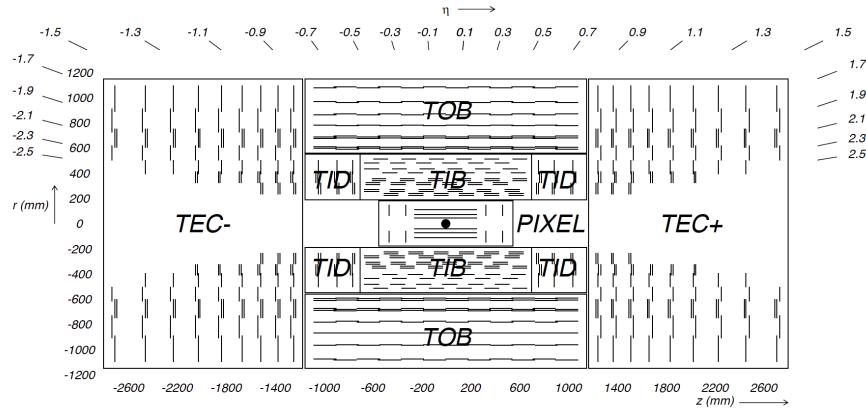


Figure 3.3. The structure of tracker in CMS [29]

### 3.2.2 Electromagnetic calorimeter

The electromagnetic calorimeter(ECAL) in CMS is a hermetic homogeneous lead tungstate( $\text{PbWO}_4$ ) detector. The whole sub-detector is composed of two parts, the central barrel(EB) covering the pseudorapidity range  $|\eta| < 1.479$  and the endcap disks(EE) covering the range  $1.479 < |\eta| < 3.0$ . The EB is made up of 61200  $\text{PbWO}_4$  crystals with  $22 \times 22 \text{ mm}^2$  in the front face, 23 cm in length(25.8 in radiation lengths). The EE is made up of 7324 crystals per disk with front face  $29 \times 29 \text{ mm}^2$ , 22 in length(24.7 in radiation lengths) and a preshower detector(ES) [47]. The geometrical configuration of ECAL is shown in Figure. 3.4.

The  $\text{PbWO}_4$  crystals used in ECAL have high density( $8.28 \text{ g/cm}^3$ ), short radiation length(0.89 cm) and small Molière radius(2.2 cm), together with the specific geometrical parameters used, rendering ECAL good energy resolution, fast response, fine granularity and high radiation resistant. Photodetectors are placed at the end of the crystals. In EB, avalanche photodiodes(APDs) are used, while vacuum phototriodes(VPTs) are used in EE. Both types of the photodiode show good performance in the environment with hard radiation and 4 T magnetic field. The ES is in front of the EE in the pseudorapidity range  $1.653 < |\eta| < 2.6$ . The ES is a sampling detector with silicon strip sensors placed behind the lead radiator to measure the energy and position of the incoming particles.

Each half of the EB is composed of 18 supermodules that each supermodule contains 1700 crystals. The relative energy resolution of  $\text{PbWO}_4$  crystal refers to the resolution that is measured with ECAL supermdules directly exposed to electron beam without considering the materials in front as the case in reality. The relative energy resolution as a function of electron energy can be expressed

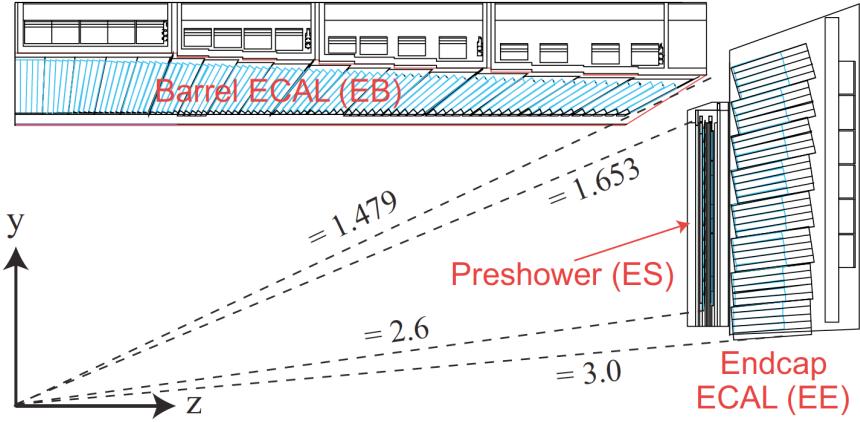


Figure 3.4. ECAL geometrical configuration [47]

as

$$\frac{\sigma}{E} = \frac{2.8\%}{\sqrt{E/\text{GeV}}} \oplus \frac{12\%}{E/\text{GeV}} \oplus 0.3\%$$

The first term stands for the contribution from stochastic factors, like the number fluctuation in production of the secondary particles. The second term is the noise contribution coming from electronics and digitization, while the last is a constant term that covers the other contribution factors [13].

### 3.2.3 The hadron calorimeter

The CMS hadron calorimeter(HCAL) is a hermetic sampling detector, which is important for the measurement of the energy and momentum of hadrons, also the missing energy that caused by the non-interacting particles. The HCAL is composed of three parts, the HCAL barrel(HB), HCAL endcaps(HE) and forward calorimeter(HF). The mechanical structure of HCAL is shown in Figure. 3.5.

The HB is composed of several layer of brass absorber plates, between which

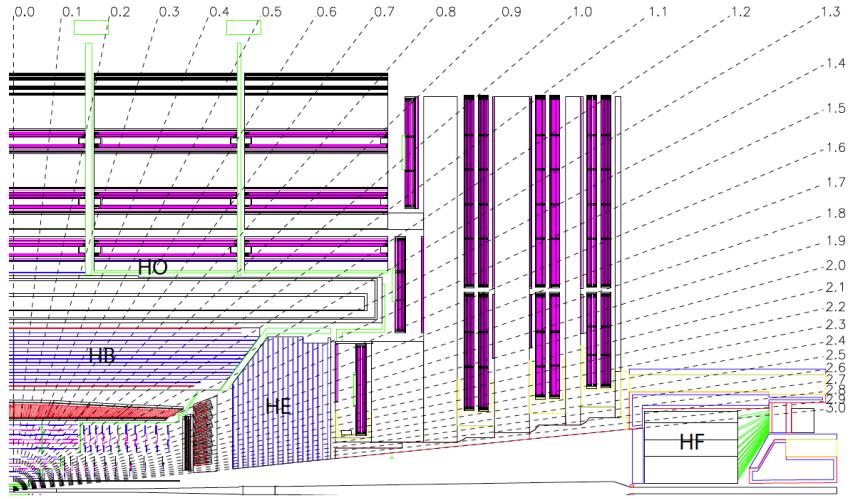


Figure 3.5. A longitudinal view of CMS HCAL sub-detector [29]

are the plastic scintillator tiles. The innermost and outermost layer plates are made of stainless steel to gain structural strength. When hadrons hit the absorber, secondary particles are produced in showers. As the showers develop, the alternating layers of scintillators are activated and emit blue-violet light. The lights are collected as signals. The HB covers the central pseudorapidity range  $|\eta| < 1.3$  with an individual read out unit  $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ . Because of the limited spaces between ECAL and muon detector, to have enough sampling depth in HCAL central region, an extra outer calorimeter(HO) is installed. The HO utilizes the outside solenoid coil as additional absorber to insert plastic scintillators. Similar to the HB, the HE is also made of brass absorber and plastic scintillator tiles layers, which covers the range  $1.3 < |\eta| < 3.0$ . The read out units have the geometry  $\Delta\eta \times \Delta\phi = 0.17 \times 0.17$  in HE [29]. A combined ECAL and HCAL energy resolution [86] measured in test beam with pions is

$$\frac{\sigma}{E} = \frac{110\%}{\sqrt{E/\text{GeV}}} \oplus 0.9\%$$

The HF situates  $\pm 11$  m from the interaction point to complement the large pseudorapidity measurement of HE in the range  $3.0 < |\eta| < 5.0$ . The HF is made of grooved steel plates with quartz fibers. Charged shower particles generate Cherenkov light, which is collected by the quartz fibers as signals. Radiation hardness is critical for the operation of HF.

### 3.2.4 Muon detector

The CMS muon detector is designed to measure the momentum and charge of muon. Three types of gas detectors are used in CMS, the barrel drift tube(DT) chambers, the cathode strip chambers(CSC) in the endcaps and the resistive plate chambers(RPC) in both barrel and endcap regions.

In the muon detector barrel(MB), DT chambers and RPCs are used which covers the pseudorapidity region  $|\eta| < 1.2$ . The MB is composed of 250 chambers. The chambers are located in 4 stations inside the magnet return yoke. The yoke is further divided into 5 wheels, each of which is composed of 12 sectors. As shown in Figure. 3.6, the stations are named MB1 to MB4, which are composed of one DT chamber and varied number of RPCs that depend on the exact location. DT chambers measure the position of the incoming muons which knock off the electrons in the gas atoms of the chambers and are collected by a large numbers of charged wires inside.

In muon detector endcaps(ME), 468 CSCs which covers the range  $0.9 < |\eta| < 2.4$  are used. The MEs are also composed of 4 stations of chambers in each of the

discs. The CSCs are in a wedge shape and composed of 6 gas gaps. Each gap is filled in with a cathodes strip and anode wires which run perpendicularly to the strip. Incoming muons knock off the electrons and create avalanches which are received by the positive charged wires.

Both DT chambers and CSCs measure the position and trigger on the  $p_T$  of muons. To better deal with the high luminosity and improve the  $p_T$  resolution of the triggered muon, a dedicated trigger system, the RPCs are added to both MB and ME. RPCs are double-gap bakelite chambers which operate with the avalanches that caused by the muons. RPCs can provide additional fast triggering and sharp  $p_T$  triggering thresholds. Four RPCs layers are used in the MB first two stations(two each) and another two layers(one each) are used in the last two stations. In the ME, one RPC layer in each of the first three stations.

### 3.2.5 Trigger

The LHC is a high luminosity collider in the order of  $10^{34} \text{cm}^2\text{s}^{-1}$ . In the p-p collision, the bunch crossing time is 25 ns and the corresponding frequency is 40 MHz. In each of the 25 ns, there are approximately 20 collisions. This high rate makes it impossible to transmit and record all of the events, also it is not necessary, since most of the events are not of current physics research interests. The CMS use a two level trigger system to select events and reduce the event recording rate, the Level-1(L1) trigger and High-level trigger(HLT) systems.

The L1 trigger system is based on custom designed and programmable electronics which are situated inside the detector. It utilizes the information from calorimeters and muon system, performing simplified but effective reconstructions, corrections and selections. The L1 reduces the event rate from 40 MHz to 100

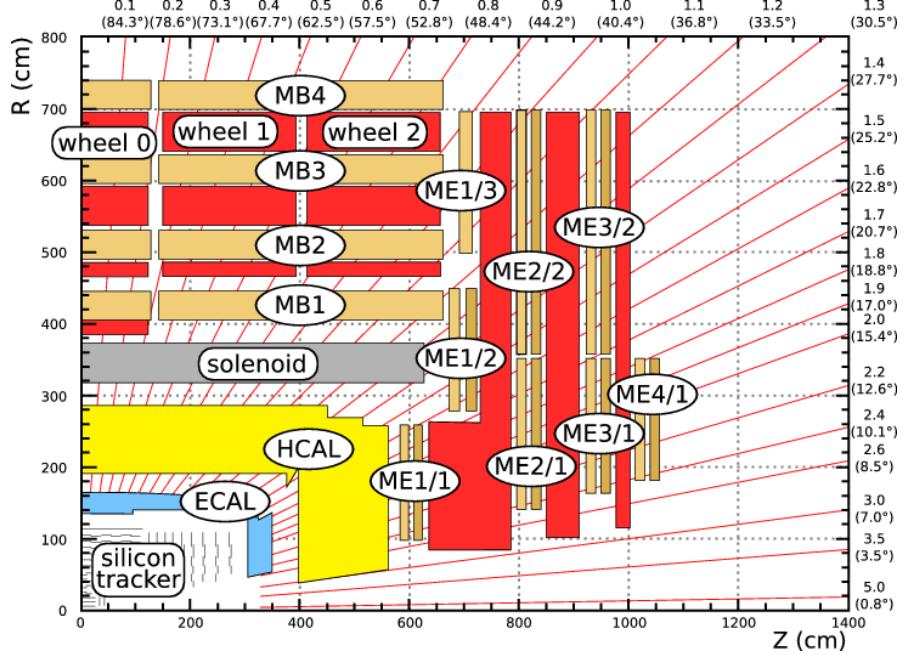


Figure 3.6. A overview of muon chamber configuration in CMS [30]

kHz. In LHC Run II, the L1 trigger system has been upgraded to deal with the increasing luminosity and improves the performance. The L1 calorimeter trigger accesses the information from the whole ECAL and HCAL in the granularity of trigger tower level, which approximately corresponds to a region  $0.087 \times 0.087$  in  $\eta$  and  $\phi$ . The L1 trigger reconstructs the  $e/\gamma$ , jets,  $\tau$  and the sum energy of the candidates with the algorithms implemented in the time multiplexed trigger architecture [17]. The algorithms are implemented at hardware trigger level, with dynamic clustering of trigger towers, pile-up migration and innovated tau and jet reconstruction with various look-up table for the calibrations and corrections [88]. The L1 muon trigger system fully utilizes three muon detectors in the track reconstruction. In general, the track reconstruction of muons are divided into three regions, the barrel, the overlap and the endcap, based on the geometry.

Through dedicated construction algorithm, tracks are built, so as computing the muon qualities. These informations are used in the triggering processes [81].

The HLT in CMS further reduces the event rate from 100 kHz(after L1 trigger selection) down to 1 kHz. The HLT is software based trigger system and utilized the streamlined version of the CMS soft-ware for event reconstruction on the large computer farm [10]. Maximizing the trigger efficiency and keeping acceptable CPU-time is crucial for the HLT system. The HLC accesses to the full granularity and sub-detectors of CMS, including the tracker. The dedicated algorithms used in the HLT is the very closed to the ones used in the off-line reconstruction and analysis, besides some of the parameter configurations [61].

## CHAPTER 4

### Datasets

#### 4.1 Datasets used in LFV analysis

##### 4.1.1 $H \rightarrow \mu\tau_h$

The data sample used in the analysis is from LHC 2016 Runs, recorded by CMS detector. Total integrated luminosity of the analyzed data is  $35.9 \text{ fb}^{-1}$  with center-of-mass energy of  $\sqrt{s} = 13 \text{ TeV}$ . The data sample labeled in the CMS experiment as SingleMuon\_Run2016B,C,D,E,F,G,H. The name SingleMuon in the data sample represents the trigger used in the analysis, which selects the events with at least a muon passed the trigger selection. Gluon gluon fusion Higgs(ggH) [60] production and vector boson fusion Higgs(VBF) [26] are the main Higgs production channels considered in the analysis. For the background samples, besides the misidentified background which is talked about in detail in Chapter6, the other background samples are all generated with Monte Carlo(MC) simulation. POWHEG [77] or MadGraph [14] generator is used for the generation and all of the MC samples, including the signal samples, the parton showering, fragmentation, and decays are performed by Pythia8 [85]. Pileup effect is taking into account in the generator by generate minimum bias events simultaneously and a correction is applied by comparing the data sample. The average number of pileup interaction per bunch crossing is 27. CMS detector environment is simulated by GEANT4 [45]. The

details of the MC samples used in the analysis are summarized in Table. 4.1 and Table. 4.2.

#### 4.1.2 $H \rightarrow e\tau_h$

The search of lepton flavour violation Higgs decay  $H \rightarrow e\tau_h$  is performed with CMS 2012 RunI dataset at center-of-mass energy  $\sqrt{s} = 8$  TeV. The dataset is named as SingleElectron\_Run2012A,B,C,D with an integrated luminosity of  $19.7 \text{ fb}^{-1}$ . Single electron HLT trigger is used which is talked more in details in Chapter 5. A detail list of simulation samples used in the analysis is listed in Table. 4.3. For the signal samples, ggH and VBF Higgs production channels are the main channels considered. For background samples, besides  $Z \rightarrow \tau\tau$  which is produced with embedding technique and misidentified background which is estimated with data-driven method are produced with MC simulation. Various simulation packages are used. Signal samples are produced with PYTHIA8, which uses sophisticated  $\tau$ -lepton decay machinery. Tauola [76] is also used for the simulation of  $\tau$  lepton decay in some of samples. CMS detector environment is simulated by GEANT4.

## 4.2 Event reconstruction

In this section, a general presentation of the event reconstruction algorithm used in CMS, particle-flow(PF) reconstruction is shown. A more detail information about the reconstruction of tracks and the objects related to the lepton flavour violation higgs decay are in the following section.

TABLE 4.1

Monte Carlo samples used in the search, together with their respective cross sections.

Processes	Generator	Cross section [pb]
DYJets $\rightarrow \ell\ell, m_{\ell\ell} > 50$ GeV	MadGraph+pythia8	4954.0
DY1Jets $\rightarrow \ell\ell, m_{\ell\ell} > 50$ GeV	MadGraph+pythia8	1012.5
DY2Jets $\rightarrow \ell\ell, m_{\ell\ell} > 50$ GeV	MadGraph+pythia8	332.8
DY3Jets $\rightarrow \ell\ell, m_{\ell\ell} > 50$ GeV	MadGraph+pythia8	101.8
DY4Jets $\rightarrow \ell\ell, m_{\ell\ell} > 50$ GeV	MadGraph+pythia8	54.8
DYJets $\rightarrow \ell\ell, m_{\ell\ell} < 50$ GeV	MadGraph+pythia8	1861.0
$t\bar{t}$	powheg+PYTHIA	831.76
$t\backslash\bar{t} \rightarrow tw$	powheg+PYTHIA8	35.85
$WZ \rightarrow \ell 3v$	MadGraph+PYTHIA8	3.05
$WZ \rightarrow \ell v 2q$	MadGraph+PYTHIA8	10.71
$WZ \rightarrow 2\ell 2q$	MadGraph+PYTHIA8	5.595
$t \rightarrow 4f$	POWHEG+PYTHIA8	136.02
$\bar{t} \rightarrow 4f$	POWHEG+PYTHIA8	80.95
$WW \rightarrow \ell v 2q$	MadGraph+PYTHIA8	1.212
$ZZ \rightarrow 2\ell 2q$	MadGraph+PYTHIA8	3.22
$VV \rightarrow 2\ell 2q$	MadGraph+PYTHIA8	11.95

TABLE 4.2

Continue with MC samples used in the analysis.

MC simulations	Generator	Cross section [pb]
$VV \rightarrow 2\ell 2q$	MadGraph+PYTHIA8	11.95
$ggH \rightarrow \tau\tau$	POWHEG+PYTHIA8	3.046
$VBFH \rightarrow \tau\tau$	POWHEG+PYTHIA8	0.237
$ggH \rightarrow WW \rightarrow 2\ell 2v$	POWHEG+PYTHIA8	1.103
$VBFH \rightarrow WW \rightarrow 2\ell 2v$	POWHEG+PYTHIA8	0.086
$ZH \rightarrow \tau\tau$	POWHEG+PYTHIA8	0.055
$W^- \backslash W^+ H \rightarrow \tau\tau$	POWHEG+PYTHIA8	0.086
$t\bar{t}H Jet \rightarrow \tau\tau$	MadGraph+PYTHIA8	0.32

TABLE 4.3  
Signal and background MC samples

Processes	Generator	Cross section [pb]
$ggH \rightarrow e\tau$	PYTHIA8	19.27
$VBH \rightarrow e\tau$	PYTHIA8	1.58
$ggH \rightarrow \tau\tau$	POWHEG+PYTHIA6	19.27
$VBF \rightarrow \tau\tau$	POWHEG+PYTHIA6	1.58
$t\bar{t} + \text{jets full leptonic}$	MadGraph+Tauola	26.20
$t\bar{t} + \text{jets Semi leptonic}$	MadGraph+Tauola	109.28
$t \rightarrow tw$	POWHEG+Tauola	56.4
$\bar{t} \rightarrow tw$	POWHEG+Tauola	30.7
$t, \bar{t}(\text{T channel})$	POWHEG+Tauola	11.1 (11.1)
$WW \rightarrow 2l2\nu + \text{jets}$	PYTHIA6+Tauola	5.824
$ZZ \rightarrow 4l$	MadGraph+Tauola	0.18
$ZZ \rightarrow 2l2Q$	MadGraph+Tauola	2.502
$ZZ \rightarrow 2l2\nu$	MadGraph+Tauola	0.716
$WZ \rightarrow 2l2Q$	MadGraph+Tauola	2.21
$WZ \rightarrow 3l\nu$	MadGraph+Tauola	1.06

### 4.2.1 Particle flow algorithm

Particle flow event reconstruction algorithm is the main algorithm used in CMS. PF performs a global event reconstruction [82], which aim to utilize the information from the whole detector to identify individual particles in each event.

#### 4.2.1.1 Tracking and Calorimeter algorithm

In the PF algorithm, the reconstruction of charged particles in the inner tracker is a crucial part. In this section, the reconstruction of the trajectories of charged particles, especially the electron and muon reconstruction is discussed.

Inner tracker is aimed at measuring the track of energetic charged particle. The track finder is based on the Kalman Filtering(KF) [12]. This reconstruction comes in a couple steps. First an initial seed is generated from a couple hits that compatible with a track in the tracker, then a trajectory is builded with the seed and other hits from the tracker along this track. At last a fit is perform on the track to determine the properties of this particle candidate, like the momentum, the charge and the direction. The qualities that can affect the performance of the reconstruction are like the number of hits in the pixel detector, the total number of hits in the tracker, the distance from the cylinder and the energy of this charged particle. This is discussed in more details later.

The performance of the track reconstruction is measured in reconstruction efficiency and misreconstruction rate. The reconstruction efficiency is defined as the ratio of track reconstructed with more than 50% hits from simulated hits and the total simulated tracks. The misreconstruction rate is defined as the fraction of the tracks can not be associated with simulated tracks of the whole simulated tracks. If a charged hadron is not identified by the tracking algorithm, then the hadron

is take as a neutral hadron and measured by the calorimeters. This will affect the jet energy and position resolution. Improving the track reconstruction efficiency while keeping the misreconstructed rate low is critical for PF reconstruction.

In CMS, an iterative tracking is perform, in which the reconstruction of tracks is done is a couple steps. Each step is aimed for a moderate efficiency but with a high purity. After one step, the hits that used to form tracks are masked and a next step is perform. This iterative tracking is down in ten steps if necessary and the detail information is shown in Table. 4.4 and more can be found in [82]. In the table, the name column points out the processes that that step of iteration is aiming at. The seeding column shows the requirement on the seeding, while the targeted tracks column shows the characters of the tracks.

Electron is one of the main particle understudy in  $H \rightarrow e\tau_h$  search. In CMS, electron track reconstruction is taken as a merge of ECAL based and tracker based strategy. The tracker based seeding strategy is as described above, besides the case when energetic photons are radiated, then a preselection based on the number of hits and  $\chi^2$  of the fit is set with the Gaussian-sum filter [11]. The ECAL based electron seeding strategy building the superclusters(SC) to gathering the bremsstrahlung photons. The energy of the SC is taken as the energy sum of the cell crystals inside and the position is evaluated with energy weight. The electron trajectory in the first layer of the the tracker is estimated and the seed from the track is selected. This works in the case when there is not much bremsstrahlung photons, then most of the energy is deposited in ECAL. In the case when soft photons are radiated mostly, ECAL based seeding may still performs well. But when the electrons are in the jets or low energy, either electron contributions are overlapping with other particle or the radiation and bending is too much, it is hard

TABLE 4.4  
Iterative tracking steps take in CMS

Iteration	Name	Seeding	Targeted Tracks
1	InitialStep	pixel triplets	prompt, high $p_t$
2	DetachedTriplet	pixel triplets	from b hadron decays, $R \lesssim 5$ cm
3	LowPtTriplet	pixel triplets	prompt, low $p_t$
4	PixelPair	pixel pairs	recover high $p_t$
5	MixedTriplet	pixel+strip triplets	displaced, $R \lesssim 7$ cm
6	PixelLess	strip triplets / pairs	very displaced, $R \lesssim 25$ cm
7	TobTec	strip triplets / pairs	very displaced, $R \lesssim 60$ cm
8	JetCoreRegional	pixel+strip pairs	inside high $p_t$ jets
9	MuonSeededInOut	muon-tagged tracks	muons
10	MuonSeededOutIn	muon detectors	muons

to recover these electrons with ECAL based algorithm only. The ECAL based and tracker based electron seeding is merged into one collection, which significantly improves the reconstruction efficiency.

Calorimeters are crucial components for the PF algorithm in CMS. The clustering algorithm in calorimeters is used to identify neutral stable particles like phone and neutral hadron, together with tracker in the identification of charged particles, reconstructing the energy of electrons and the possible associated bremsstrahlung photons and measuring the energy of charge particles that are missed by the

tracker.

The clustering algorithm is perform separately in each sub-detector system beside the HF in which each cell directly raise a cluster. The algorithm starts by finding a cluster seed. Then a topological walking around the neighbouring cells is performed. Both seed cells and neighbouring cells are required to pass certain thresholds to construct high quality candidates and suppress the contribution from noise. In ECAL endcaps, additional requirements on  $E_T$  because of the high noise level. In each of the topological clusters, the energy for the neighbouring cells are assumed from the seed cells. A Gaussian-mixture model is used in the construction of the topological clusters to evaluate the contribution of each cells. The finally parameters of the model is obtained by analytical fitting to the Gaussian model which give the maximum expectation.

To accurately measure the energy of particles like photons and neutral hadrons, the calibration of calorimeters is indispensable. The calibration of calorimeter also affects identification efficiency and misidentification rate of the particles measured in the calorimeters. The calibration of ECAL is done with a couple source, like the test beam, the radioactive source and the cosmic ray measurements and refined with the collision data. There are thresholds used in the formation of topological clusters, which results in the energy measured is smaller than the incoming particle energy. A residual energy calibration is applied to all ECAL clusters to account for these affects with simulation samples. The correction is applied as a function cluster energy and position,  $f(E, \eta) = g(E)h(\eta)$ . In the endcaps, the energy is taken as a liner combination of ECAL and preshower energy, the parameters in the combination is optimized by the  $\chi^2$  method. Hadrons generally leave energy in both ECAL and HCAL. The calibration of ECAL mentioned above is for the

photon, electron calibration. The behavior of hadron is different and a consequent calibration involving both ECAL and HCAL is needed. Simulated single neutral hadrons is used. The relationship involving both the energy measured by ECAL and HCAL and the calibrated energy  $E_{calib}$  is expressed as

$$E_{calib} = a + b(E)f(\eta)E_{ECAL} + c(E)g(\eta)E_{HCAL}$$

In the equation,  $a$  is independent of  $E$  and accounts for the effects of thresholds in the clustering algorithm. The other coefficients are determined by the  $\chi^2$  optimization with the  $E_{calib}$  and true energy  $E$  from simulation.

A detail description of particle flow algorithm can be found in [83].

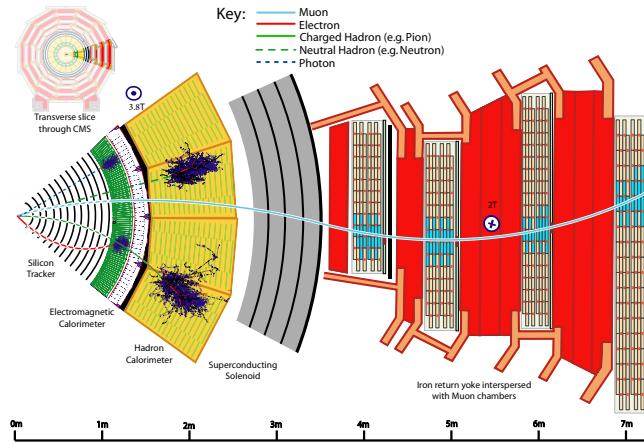


Figure 4.1. A sketch view of CMS detector. Examples are given to show how the particles interact with different sub-detectors.

#### 4.2.2 Muon reconstruction and selection criteria

PF muon is used in the  $H \rightarrow \mu\tau_h$  analysis. In CMS, the event reconstruction starts from building the tracks in tracker(tracker track) and muon system(standalone-muon track) separately. Global muon reconstruction and tracker muon reconstruction are based on these tracks. [34]. Global muon reconstruction starts from the standalone-muon tracks and requires at least two muon station in the muon system. For each of the standalone-muon track, a propagating is done to find a matching tracker track on the common surface. Kalman-filter technique [59] is used in the fitting to combine the hits in standalone-muon track and tracker track. The tracker muon starts from the tracks with  $p_T > 0.5\text{GeV}$  and total momentum  $p > 2.5\text{ GeV}$ . The tracks are then extrapolated to match the tracks in the muon system with at least one muon segment. Within the geometry acceptance of muon system, the muon reconstruction efficiency is high, especially high momentum muons, constructed as either global muon or tracker muon or both has the efficiency about 99%. PF reconstruction utilizes the whole detector information to identify and reconstruct all of the particle. The muon PF algorithm applies a series of selections to the global muon and tracker muon to select out the PF muon. The selection is optimized to identify the muons in the jets with high efficiency and low misidentify rate. The details of the selection is in [2].

With the PF muon as the input, muons used in the analysis are further categorized into different identification(ID), isolation(Iso) categories. Loose muon ID is referred to the muons selected by the PF algorithm. The medium Muon ID was suggested in CMS RunII. For the LHC data 2016, running period BCDEF, ICHEP medium muon ID is applied(table 4.5), for running period G and H, also the monte Carlo samples, standard medium muon ID(table 4.6)are applied to

TABLE 4.5

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
	Global muon
	Normalized global-track $\chi^2 < 3$
1. Good Global muon	Tracker-Standalone position match < 12
	kick finder < 20
	Segment compatibility > 0.303
2. Tight segment compatibility	Segment compatibility > 0.451

achieve the best performance for muon identification. The muon isolation used in the analysis is calculated the same way as the electron isolation which is shown in the following electron section.

#### 4.2.3 Electron identification

Electrons in CMS is constructed with the information from tracker and calorimeters. One of the main difficulties is the bremsstrahlung emitted by the electron during the traveling among the detector materials. The conversion of the photons from the bremsstrahlung affect the reconstruction of tracks in the track and these

TABLE 4.6

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
	Global muon
	Normalized global-track $\chi^2 < 3$
1. Good Global muon	Tracker-Standalone position match < 12
	kick finder < 20
	Segment compatibility > 0.303
2. Tight segment compatibility	Segment compatibility > 0.451

phones also cause significance energy loose in the electron reconstruction. The construction of Electron is discussed in section 4.2.1.1 and more details can be found in [69].

Electron ID is constructed to separate Prompt isolated electron(signal) from the background processes. The background can be the electrons from photon conversion, from quark semi-leptonic decay and from misidentification of other particles or jets. The variables used in the identification are related to tracker and ECAL. There are mainly three types of variables. The variable related only to calorimeters. For example, the cluster shape of real electron in ECAL is usually narrower than the shape from hardonic showers and electrons leave most of the energy in ECAL and the energy ratio between ECAL and HCAL is large. The variables related to the matching of measured energy and geometry between tracker and ECAL. The variables related to tracker fitting to explore the difference between electrons and hadrons. These related variables can be used to construct cut-based selection sequence to selection electron. To achieve better performance, MVA based ID with boosted decision tree is also trained. Compared with the cut-based selection, more variables in the three categories mentioned above are used in the training [69]. An example of the BDT electron ID is shown in Figure. 4.2

Electron isolation is used to reject background events in addition to the ID variables, also inverting the requirement in the isolation can be used to setup enriched background control regions. Large numbers of background events that can possible enter the signal selection are misidentified jets or the jets in which there are real electrons, for example the jets from b quark semi-leptonic decay. For these background events, one key different character with respect to signals is more energy flowing around the electron(or misidentified electron) trajectory.

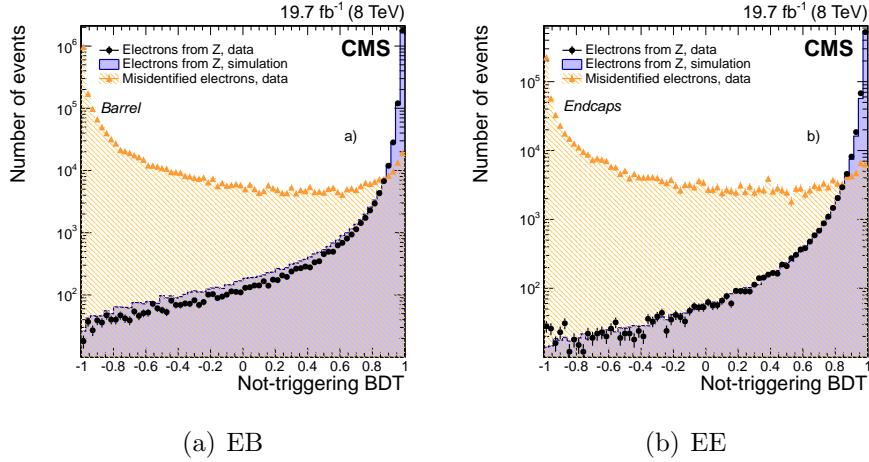


Figure 4.2. Electron BDT-based ID shows good discriminating power against background in both EB and EE [69]

The isolation requirement used in HLT level is summing over the energy depositions either in ECAL or HCAL in the certain core, for example  $\Delta R = 0.3, 0.4$ ,  $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ . The contribution from the particle candidate is removed. In the offline algorithm, particles can better identified with the PF algorithm. Similar in concept to the isolation using the energy flow, the PF isolation summing over the  $p_T$  of the particles in the direction of the reconstructed candidate trajectory momentum. PF algorithm utilizes the whole detector information and the isolation is defined as

$$\text{Iso}_{\text{PF}} = \sum p_T^{\text{charged}} + \max \left[ 0, \sum p_T^{\text{neutral had}} + \sum p_T^\gamma - \frac{1}{2} \sum p_T^{\text{charged,PU}} \right] \quad (4.1)$$

The isolation variable sums over the contribution from charged PF candidates, neutral particles and photons in the selected certain  $\Delta R$  region around the signal.

The contribution from pileup is considered with the  $\Delta\beta$  correction in the last term of Equation. 4.1. The factor 0.5 comes from the naive measurement of neutral to charged particles in jets [25]. The energy of PF objects are better calibrated and the double counting problem which shows up the energy flow isolation described above is solved. PF isolation variable performs better than the isolation variable used in HLT. The performance comparison is shown in Figure. 4.3

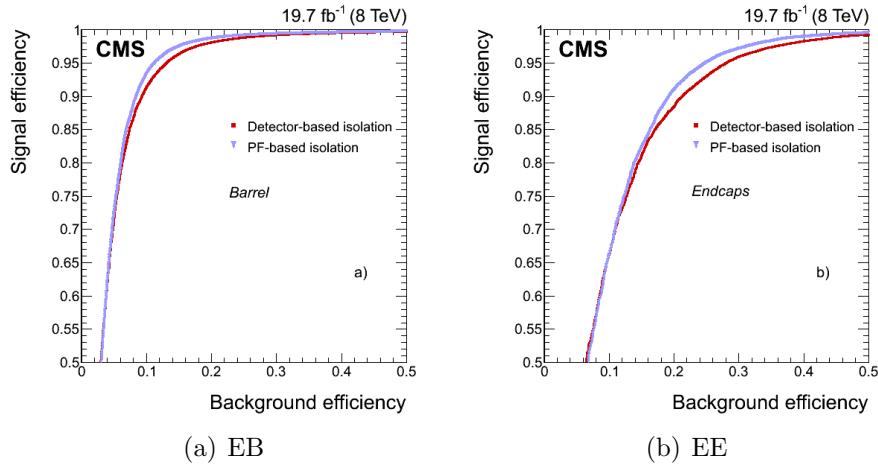


Figure 4.3. PF electron isolation shows better performance in both EB and EE with respect to detector based isolation variable [69]

#### 4.2.4 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 recon-

structed 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 identify different  $\tau$  decay modes and calculate  $\tau$  four-momentum and other quantities [38].

Photon conversions and the bremsstrahlung of electron/positron when traveling inside the CMS detector are well treated by the HPS algorithm. These phenomena 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  $P_T^{strip} > 1 GeV$ . 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. 4.7.

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. 4.2.

$$I_\tau = \sum p_T^{\text{charged}}(d_z < 0.2 \text{ cm}) + \max(0, \sum p_T^\gamma - \Delta\beta \sum p_T^{\text{charged}}(d_z > 0.2 \text{ cm})) \quad (4.2)$$

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:taucutis01 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 [40]. 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  $P_T$

TABLE 4.7

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(%)
$\tau^- \rightarrow h^- v_\tau$			11.6%
$\tau^- \rightarrow h^- \pi^0 v_\tau$	$\rho^-$	770	26.0%
$\tau^- \rightarrow h^- \pi^0 \pi^0 v_\tau$	$\alpha_1^-$	1200	9.5%
$\tau^- \rightarrow h^- h^+ h^- v_\tau$	$\alpha_1^-$	1200	9.8%
$\tau^- \rightarrow h^- h^+ h^- \pi^0 v_\tau$			4.8%

secondary particles. Photons from the neutral pion decay can also go through 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.

$$\begin{aligned}\delta\eta &= f(P_T^\gamma) + f(P_T^{strip}) & f(P_T) &= 0.2 \cdot P_T^{-0.66} \\ \delta\phi &= g(P_T^\gamma) + g(P_T^{strip}) & g(P_T) &= 0.35 \cdot P_T^{-0.71}\end{aligned}$$

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.

$$\begin{aligned}\eta_{strip} &= \frac{1}{P_T^{strip}} \cdot \sum P_T^\gamma \cdot \eta_\gamma \\ \phi_{strip} &= \frac{1}{P_T^{strip}} \cdot \sum P_T^\gamma \cdot \phi_\gamma\end{aligned}$$

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 [41]. The conditions of different hadronic decay mode mass window are listed in the Table. 4.8. With respect to RunI, the difference in mass window is  $\delta m$ , which originates from dynamic clustering.  $\delta m$  is calculated as:

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

with:

$$\frac{\partial m_\tau}{\partial \eta_{strip}} = \frac{P_z^{strip} \cdot E_\tau - E_{strip} \cdot P_z^\tau}{m_\tau}$$

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

TABLE 4.8

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

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

In current algorithm,  $\tau^- \rightarrow h^- h^+ h^- v_\tau$  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 [40, 41]. BDT uses these variables to distinguishing  $\tau_h$  dacay( $H \rightarrow \tau\tau$ ) 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 [40, 41].

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.

#### 4.2.5 Jet reconstruction

Jets are produced in the hadronic processes involving quarks or gluons. In p-p collision, jets involving in most of the processes. Understanding and constructing jets is crucial and difficult.

Anti- $k_t$  jet clustering algorithm [24] is used in CMS for the jet reconstruction. The algorithm starts by introducing the distance  $d_{ij}$  and  $d_{iB}$  as following,

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2}$$

$$d_{iB} = k_{ti}^{2p}$$

$d_{ij}$  is the distance between entities, the particle and the pseudojet.  $\Delta_{ij}$  is the difference of rapidity and azimuth between entry i and entry j.  $\Delta_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ . R is a radius parameter.  $k_{ti}$  and  $k_{tj}$  stands for the momentum of the entries respectively. p is a parameter used to specify jet construction algorithms. For Anti- $k_t$ , p=-1.  $d_{iB}$  is the distance between entry i and the beam. If the parameter p is set for other values, for example p=1 or 0, then the algorithm is the  $k_t$  [57] or Cambridge/Aachen jet reconstruction algorithm [87]. With the Anti- $k_t$ , assuming in one event, there are a couple hard particle with high momentum,  $k_{t1}, k_{t2}$  and so on, also large numbers of soft ones around. Starting from  $k_{t1}$  as an example, if  $d_{1j}$  is smaller, then entry j combined with entry 1, if  $d_{1B}$  is smaller, then entry 1 is set as a jet and removed from the list. The soft entries tend to ground around the high momentum entries to the range  $\Delta_{ij} = R$ , if there is no other hard entry around. If two hard entries are within the distance R, then the two entries are clustered into a single entry. In this case, if  $k_1 \gg k_2$ , then the center is more closed to jet 1. If  $k_1 \sim k_2$ , the boundary between two jets is defined by  $\Delta R_{1b}/k_{t1} = \Delta_{2b}/k_{t2}$ . Jet clustering continues until all of the jets are clustered in the events.

Jet energy is corrected to have the correct energy scale. The correction goes through a couple of steps [71]. The major corrections are derived from simulated samples and the residual corrections from the different response between MC and data are from data-driven methods.

Pileup events can increase the measured jet energy, especially in LHC Run II, the number of pileup per event doubled. Two types of pileup affect the performance the most, in-time pileup(IT PU) and out-of time pileup(OOT PU). The IT PU refers to the additional events produced by the proton-proton collision within the same bunch-crossing as the primary hard collision. The OOT PU refers to the events that produced in previous bunch crossing or subsequent one that affect current bunch. OOT PU can be mitigated by explore the timing window, pulse shape of the calorimeter. The IT PU mitigation is mainly discussed. Charged hadrons from IT PU in CMS are removed with Charged-hadron subtraction(CHS) algorithm in CMS. Tracks with the vertexes that are identified from PU with charged particles inside are removed. CHS removes around 50% of the IT PU within the tracker coverage region. There are also soft jets from pileup interaction. These jets are usually in low energy range and affects the JES by overlapping with the hard jets. Multivariate analysis(MVA) with inputs from jet shape and jet constitution information can more than 90% of pileup jets. This MVA ID is referred as PUJetID [4]. After rejecting the charged particle jets and soft jets from pileup, a jet area method [25] is used to further eliminate the effects from PU. In this method, an estimated energy density bringing in by the PU and effective area of jet is used to calculate the offset energy. After the PU correction, simulated response corrections are performed with simulated QCD multijet sample. A matching between particle-level jets and reconstructed jets is performed. The correction is performed with the anti- $k_T$  jets with distance parameter R=0.5 in  $p_T$  and  $\eta$  distribution and checked with jets corresponding to R in range 0.3 to 1.0. Following the simulated response corrections, the remaining residual correction is small derived with data-based methods. A comparison between data and

MC samples and corrects data only.

B jets is a main component in  $t\bar{t}$  process, so the identification of b jets is important to the analysis involves  $t\bar{t}$ . In CMS, combined secondary vertex version 2(CMSv2) currently is used, which uses multivariable techniques and the character of b jets, to identify b jets [84]. B jets is from the hadronization of b quark. B hadron is relatively heavy, around 5-6 GeV, with high decay multiplicity, relatively high momentum decay products and relatively wide decay core. B hardon has long lifetime( $ct \approx 450\text{um}$ ) which give a traveling distant  $\approx 5$  mm at the energy of 70 GeV. Exploring the property of secondary vertex is used in the identification algorithm. In each jet, at least two tracks presenting in an angular distance  $\Delta R < 0.3$  is required. The jets with combined mass of two tracks compatible with  $k_s^0$  is rejected to reduce the contamination. A number of variables [84] are used as input into the MVA training in different secondary vertices presenting situation. Combining the input variables into a discriminant variable to identify b jets.

#### 4.2.6 Missing transverse momentum

Missing transverse momentum(MET) is defined as the sum  $p_T$  of un-detected process in an event. In CMS, particle flow MET is widely used and the MET can be expressed as  $\vec{\cancel{E}}_T = -\sum \vec{p}_T$ , in which the summing is over the observed PF particles. The magnitude of MET can be affected by various resources, for example the in efficiency of tracker, the thresholds on calorimeter and the asymmetry in detector response. The propagation of JEC to MET can reduce the bias. The corrected MET can be expressed as

$$\vec{E}_T^{corr} = \vec{E}_T - \sum_{jets} (\vec{p}_{T,jet}^{corr} - \vec{p}_{T,jet})$$

The sum is summing over the jets with electromagnetic fraction below 0.9 and corrected  $p_T > 10$  GeV. Further correction aiming for estimation is the effects of PU on MET is derived with charged particles associated with PU vertexes,  $\vec{v} = \sum_{charged} \vec{p}_T$ . The correction is  $f(\vec{v})\vec{v}$ , in which factor  $f(\vec{v}) = c_1(1.0 + \text{erf}(-c_2|\vec{v}|^{c_3}))$ . the coefficients in  $f(\vec{v})$  are estimated from fitting the MC minimum bias events. The correction on MET for the effect of PU is

$$\vec{E}_T^{corr} = \vec{E}_T - \sum_{PU} f(\vec{v})\vec{v}$$

The observed  $\vec{E}_T$  asymmetry in  $\phi$  shows approximately linear relationship with the number of reconstructed vertices. The correction is performed with  $N_{vtx}$  separately in x and y axis.

$M_{col}$  is the most important variable in LFV Higgs decay analysis, which is closely related to MET. Higgs is more massive than the leptons( $e, \mu, \tau$ ), so the decay products are boosted. In  $H \rightarrow \mu\tau_h$  and  $H \rightarrow e\tau_h$  analysis,  $\tau$  decays hadronically. The neutrinos in  $\tau$  decay products are highly Lorentz boosted and are considered collinear with  $\tau$ . The direction of  $\tau$  and the projection of MET amplitude on  $\vec{\tau}$  direction are used to approximate the direction and energy of the neutrinos. With this approximation and the visible mass which is the mass formed with  $e-\tau$  or  $\mu-\tau$  without the approximation,  $M_{col}$  can be derived as

$$M_{col} = M_{vis} / \sqrt{x_\tau^{vis}}$$

$$x_\tau^{vis} = p_T^{\tilde{\tau}^{vis}} / (p_T^{\tilde{\tau}^{vis}} + p_T^{\nu,est})$$

### 4.3 Event simulation

Monte Carlo simulation is used widely from the detector related processes, event reconstruction to physics analysis etc. The generation of MC sample in the generator involving the following steps.

In LHC, p-p collision, QCD hard subprocesses involving large momentum transferring, multiply final states and complex surrounding environment. For a particular process, MC generator starts by calculating the related QCD cross section. This involving calculating the matrix-element, choosing parton distribution functions(PDFs) and possibly performing high order correction. Followed by the parton showers. In a hard process, parton showers uses the basic building block like  $q \rightarrow qg$ ,  $q \rightarrow gq$ ,  $g \rightarrow gg$  and  $g \rightarrow q\bar{q}$ , iteratively showering from high energy to low energy, typically, for quark and gluon in the order of 1 GeV, when they can not be treated as free particles anymore. In addition to the hard subprocesses and the associated parton showers, a real event consists more additional contributions. Typically more than one pair of partons interacts in one collision. These multiple interactions are dealt with different models and affects the total scattering energy, the number of the particle in the hadronization stage, etc. Because of the color confinement, quarks and gluons forms hadrons after completing the parton shower, this is referred as hadronization. Similarly to the QCD processes, electromagnetic radiation is also taken into account with the evolution according

to the electric charge instead of color. With the generated events through the MC generate, the events are put into GEANT4 [45]

## CHAPTER 5

### LFV event selection

In both 8 TeV  $H \rightarrow e\tau_h$  and 13 TeV  $H \rightarrow \mu\tau_h$  analyses, events are selected in several steps. A loose selection selects on the different IDs, energy, geometry parameters of the analysis related objects. In the  $M_{col}$  fit analysis this is 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  $H \rightarrow e\tau_h$  and  $H \rightarrow \mu\tau_h$  analysis. This selection sequence is referred as  $M_{col}$  fit analysis. In  $H \rightarrow \mu\tau_h$ , there is an alternate selection sequence that follows the loose selection. A multivariate analysis with a Boosted decision tree (BDT) is defined and fits on the BDT discriminator. This is referred as the BDT fit analysis and provides more sensitive results in  $H \rightarrow \mu\tau_h$  analysis.

## 5.1 $H \rightarrow \mu\tau_h$

### 5.1.1 Loose selection

Tau leptons from signal events decay hadronically. A SM Higgs boson is much heavier than its decay products  $\mu$  and  $\tau$ . The  $\mu$  and  $\tau$  leptons are expected to have high  $P_T$ . As the decay products from signal events are boosted, therefore a cut on  $\Delta R = \sqrt{(\Delta phi)^2 + (\Delta eta)^2}$ ,  $\Delta R > 0.3$  is applied. The  $\mu$  and  $\tau$  candidates are required to have opposite sign of charge as the Higgs boson is neutral. Further,

events with additional  $\mu$  and  $\tau$  that pass the loose selection and events with jets that are identified by the combined secondary vertex(CSVv2) b-tagging algorithm [42] as a b quark jets are vetoed. The trigger HLT\_IsoMu24 or HLT\_IsoTkMu24 used in the analysis selects isolated muons that have energy higher than 24 GeV at HLT level. A further  $P_T$  cut on the 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  $I_{rel}^\mu < 0.15$ .

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 will also cause the problem that signal peaks under backgrounds. As leptons from signal process tend to have higher  $P_T$  values, cutting tighter on the lepton  $P_T$ , removes more background events. 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  $H \rightarrow e\tau_h$  analysis, the effect of cutting tighter on lepton  $P_T$  will be shown. In  $H \rightarrow \mu\tau_h$  search, lepton  $P_T$  variables are kept at loose values and optimized on other variables to achieve tighter expected limits.

The optimization procedure uses only Monte Carlo samples so as not to double use data. In  $H \rightarrow \mu\tau_h$  channel, the variables tuned are  $M_{jj}$  and  $M_T(\tau_h)$ . The selection criteria has been optimized to have the most stringent expected limits with the Asimov dataset(background only). The name Asimov dataset is inspired by the short story Franchise, by Isaac Asimov. 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  $H \rightarrow \mu\tau_h$  is shown in Table. 5.1

### 5.1.3 Multivariate analysis

A boosted decision trees(BDT) method is used as the multivariate analysis method in the  $H \rightarrow \mu\tau_h$  search which is more sensitive than the  $M_{col}$  fit analysis. The BDT algorithm used in this search is implemented in the TMVA package [66]. 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 the ggH and VBF Higgs production mode are mixed by weighting with respect to their production cross section. The background sample used in the training is the 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  $E_T^{miss}$ ,  $M_T(\tau_h)$ .
- Missing transverse energy,  $E_T^{miss}$ .
- 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  $E_T^{miss}$ ,  $\Delta\phi(\tau_h, E_T^{miss})$ .

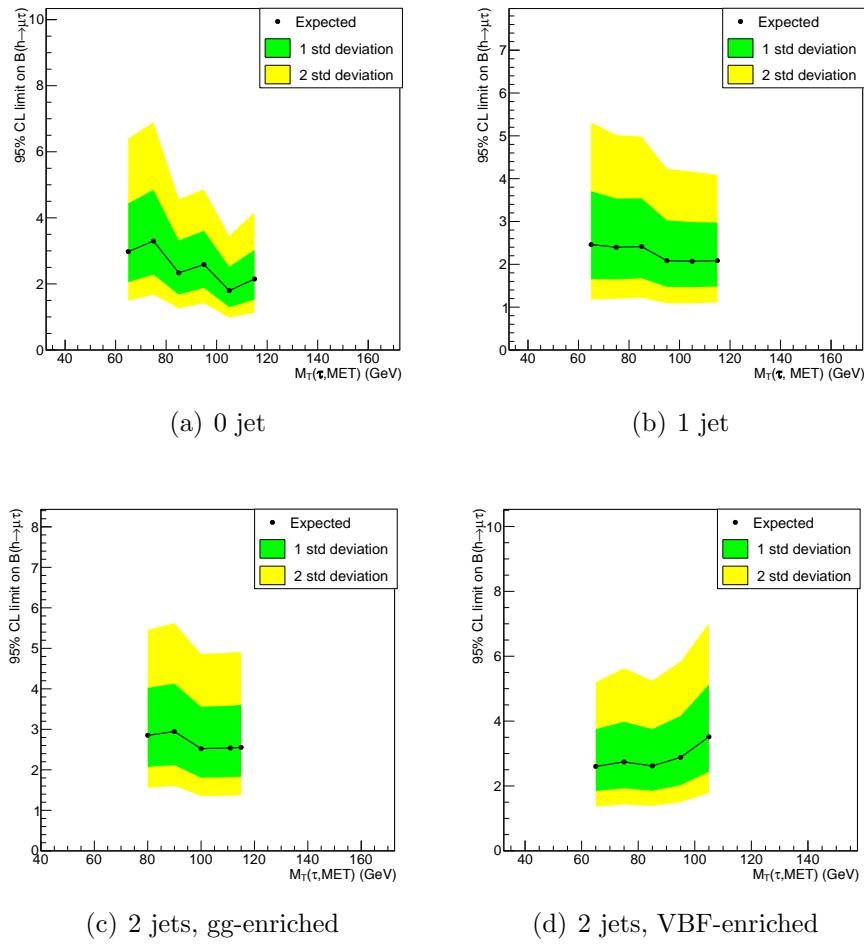


Figure 5.1. Expected limits based on an Asimov dataset as a function of  $M_T(\tau, \text{MET})$  for the different categories.

TABLE 5.1

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

<b>0-jet category</b>
<ul style="list-style-type: none"> <li>• <math>p_T^\mu &gt; 26</math> GeV, <math>p_T^\tau &gt; 30</math> GeV</li> <li>• <math>M_T(\tau) &lt; 105</math> GeV</li> <li>• No jets with <math>p_T^{jet} &gt; 30</math> GeV, <math> \eta  &lt; 4.7</math>, LooseID</li> </ul>
<b>1-jet category</b>
<ul style="list-style-type: none"> <li>• <math>p_T^\mu &gt; 26</math> GeV, <math>p_T^\tau &gt; 30</math> GeV</li> <li>• <math>M_T(\tau) &lt; 105</math> GeV</li> <li>• One jet with <math>p_T^{jet} &gt; 30</math> GeV, <math> \eta  &lt; 4.7</math>, LooseID</li> </ul>
<b>2-jet, gg-enriched category</b>
<ul style="list-style-type: none"> <li>• <math>p_T^\mu &gt; 26</math> GeV, <math>p_T^\tau &gt; 30</math> GeV</li> <li>• <math>M_T(\tau) &lt; 105</math> GeV</li> <li>• <math>p_T^{jet1} &gt; 30</math> GeV, <math>p_T^{jet2} &gt; 30</math> GeV <math> \eta_{jet1}  &lt; 4.7,  \eta_{jet2}  &lt; 4.7</math>, LooseID</li> <li>• <math>M_{jj} &lt; 550</math> GeV</li> <li>• Two jets with <math>p_T^{jet} &gt; 30</math> GeV, <math> \eta  &lt; 4.7</math>, LooseID</li> </ul>
<b>2-jet, VBF-enriched category</b>
<ul style="list-style-type: none"> <li>• <math>p_T^\mu &gt; 26</math> GeV, <math>p_T^\tau &gt; 30</math> GeV</li> <li>• <math>M_T(\tau) &lt; 85</math> GeV</li> <li>• <math>p_T^{jet1} &gt; 30</math> GeV, <math>p_T^{jet2} &gt; 30</math> GeV <math> \eta_{jet1}  &lt; 4.7,  \eta_{jet2}  &lt; 4.7</math>, LooseID</li> <li>• <math>M_{jj} &gt; 550</math> GeV</li> <li>• Two jets with <math>p_T^{jet} &gt; 30</math> GeV, <math> \eta  &lt; 4.7</math>, LooseID</li> </ul>

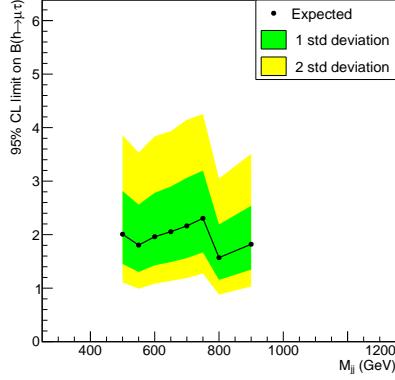


Figure 5.2. Expected limits based on an Asimov dataset as a function of  $M_{jj}$  for the 2 jet categories.

- 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. In the training process, overtraining needs to be carefully treated. The overtraining problem refers to the case that the classifier between signal and background samples are specific to the particular training sample used. The training recognizes the specific features that only occurs in the training samples. Most of the time, these features are caused by the limited number of training events or with the same number of events, too many variables with weak distinguishing power are used. 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

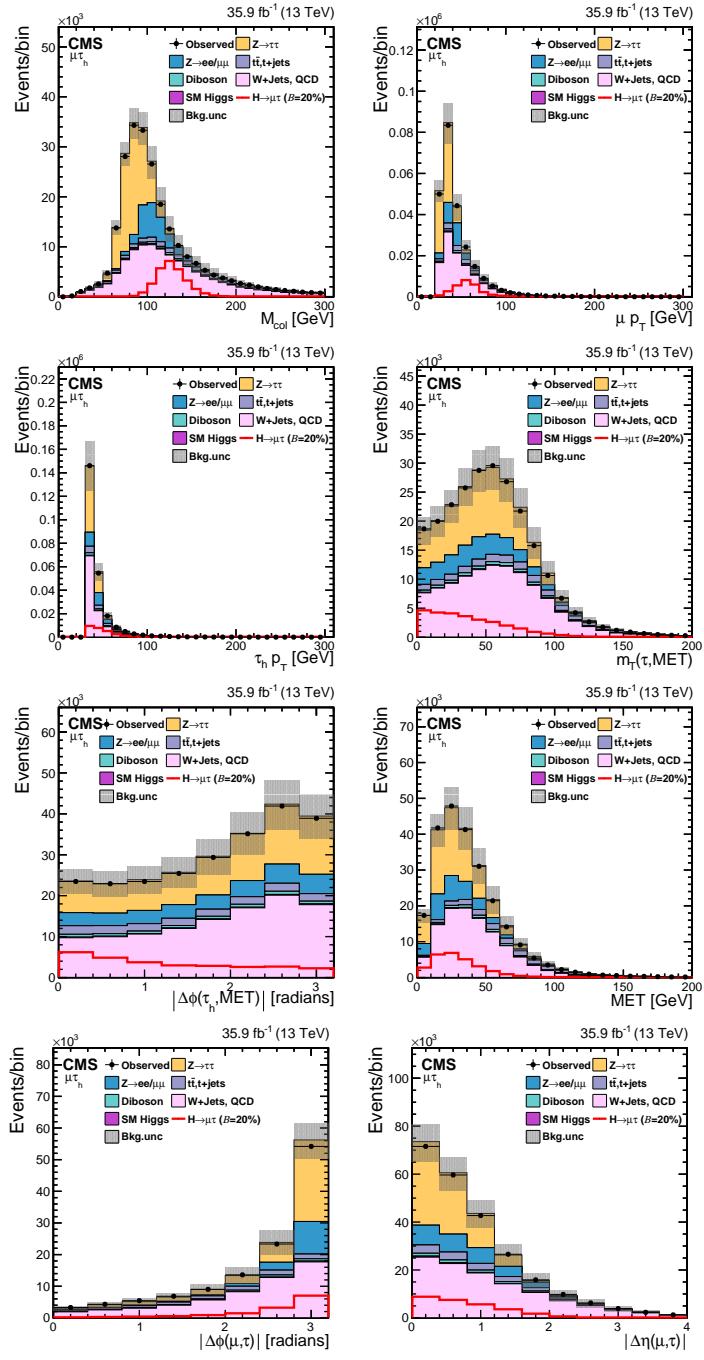
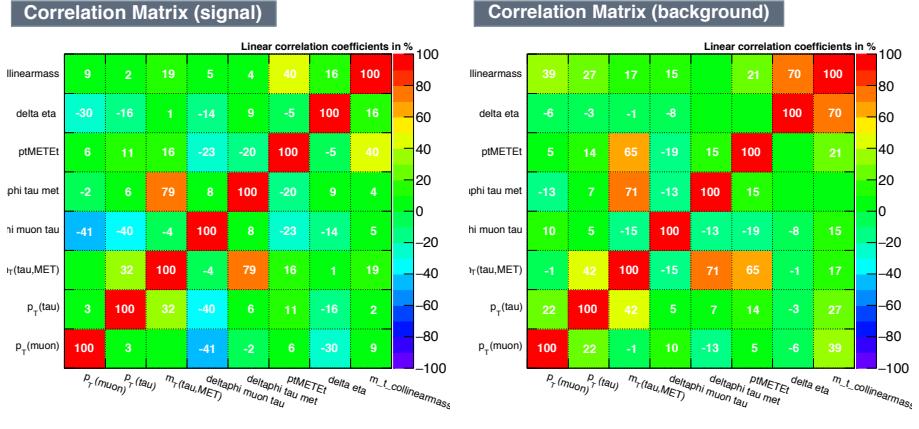


Figure 5.3. Distributions of the input variables to the BDT for the  $H \rightarrow \mu\tau_h$  channel.



(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.

in Figure. 5.5.

## 5.2 $H \rightarrow e\tau_h$

### 5.2.1 Loose selection

In  $H \rightarrow e\tau_h$  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  $D_z < 0.2cm$ .  $D_z$  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  $I_{rel}^e < 0.1$ . 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

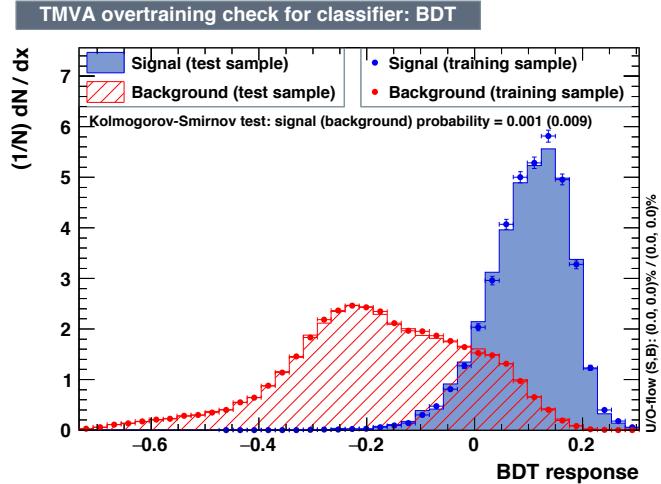


Figure 5.5. Overtraining checking for the BDT training in the TMVA package.

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

### 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 play 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 the 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.

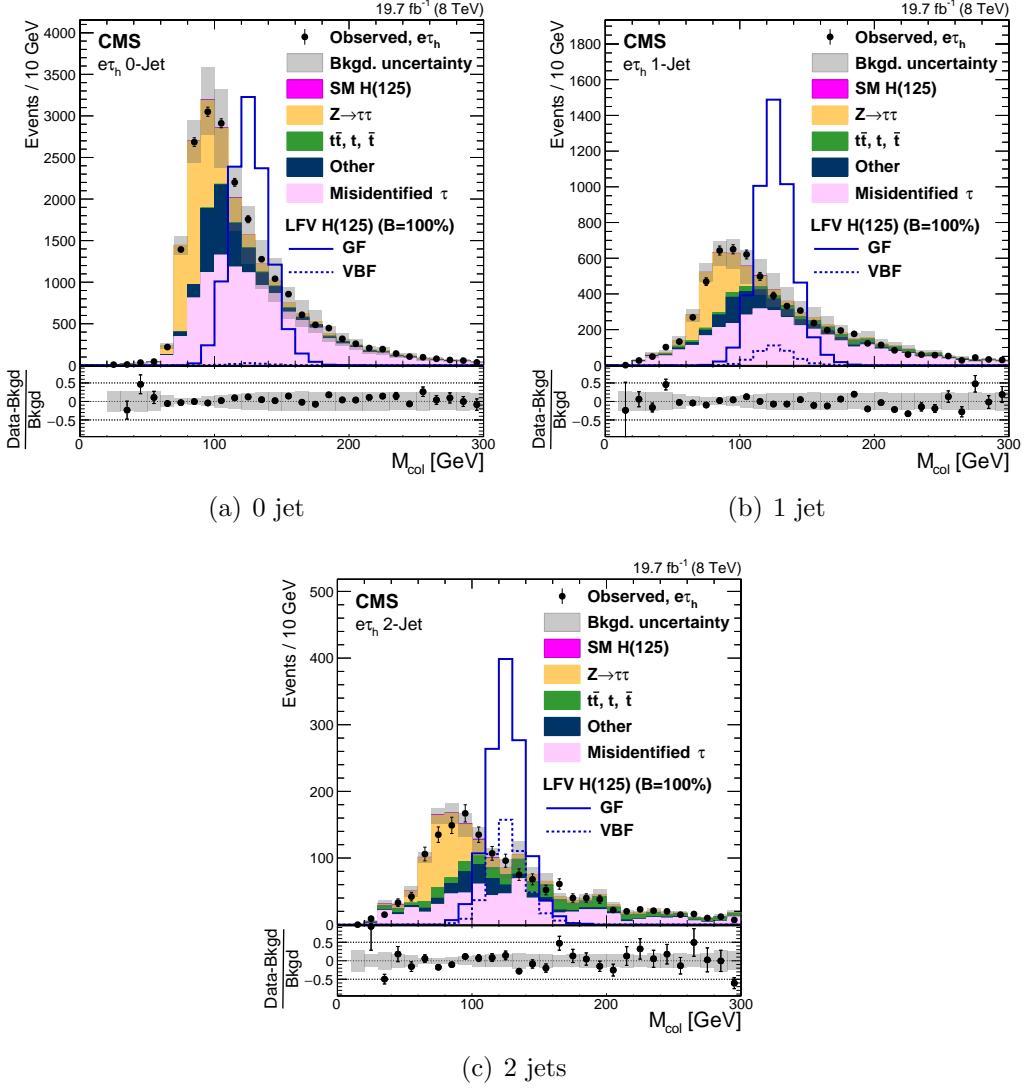


Figure 5.6. With loose selection conditions, the comparison of the observed collinear mass distributions with background from prediction. The shaded grey bands indicate the total background uncertainty. The open histograms correspond to the expected signal distributions for  $\mathcal{B}(H \rightarrow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

TABLE 5.2

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

---

<b>0-jet category</b>
<ul style="list-style-type: none"> <li>• <math>p_T^e &gt; 45</math> GeV, <math>p_T^\tau &gt; 30</math> GeV, <math>\Delta\phi_{e\tau} &gt; 2.3</math></li> <li>• <math>M_T(\tau) &lt; 70</math> GeV</li> <li>• No jets with <math>p_T^{jet} &gt; 30</math> GeV, <math> \eta  &lt; 4.7</math>, LooseID</li> </ul>
<b>1-jet category</b>
<ul style="list-style-type: none"> <li>• <math>p_T^e &gt; 35</math> GeV, <math>p_T^\tau &gt; 40</math> GeV</li> <li>• One jet with <math>p_T^{jet} &gt; 30</math> GeV, <math> \eta  &lt; 4.7</math>, LooseID</li> </ul>
<b>2-jet category</b>
<ul style="list-style-type: none"> <li>• <math>p_T^e &gt; 35</math> GeV, <math>p_T^\tau &gt; 30</math> GeV</li> <li>• <math>M_T(\tau) &lt; 50</math> GeV</li> <li>• <math>p_T^{jet1} &gt; 30</math> GeV, <math>p_T^{jet2} &gt; 30</math> GeV, <math> \eta_{jet1}  &lt; 4.7,  \eta_{jet2}  &lt; 4.7</math>, LooseID</li> <li>• <math>\Delta\eta(jet1, jet2) &gt; 2.3</math></li> <li>• <math>M_{jj} &gt; 400</math> GeV</li> <li>• Two jets with <math>p_T^{jet} &gt; 30</math> GeV, <math> \eta  &lt; 4.7</math>, LooseID</li> </ul>

---

## CHAPTER 6

### LFV analysis background estimation

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

##### 6.1.1 Misidentified Leptons

The misidentified lepton backgrounds are estimated with a fully data driven method using collision data. In  $H \rightarrow \mu\tau_h$  channel, the misidentification  $\tau$  and  $\mu$  lepton rate is obtained from an independent  $Z + \text{jets}$  data sets and then applied to a control region that is 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 except that the  $\mu$  and  $\tau$  lepton pairs have the same electric charge. Similarly 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  $Z + \text{jets}$  collision dataset to estimate both misidentified  $\tau$  and  $\mu$  lepton rates.  $Z$  bosons are selected in the invariant mass window  $70 < M_{ll} < 110$  GeV and the trigger that require isolated muons have  $p_T > 24$  GeV

TABLE 6.1

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

<b>Region I</b>	<b>Region II</b>
$\ell_1^\pm$ (isolated)	$\ell_1^\pm$ (isolated)
$\ell_2^\mp$ (isolated)	$\ell_2^\pm$ (isolated)
<b>Region III</b>	<b>Region IV</b>
$\ell_1^\pm$ (isolated)	$\ell_1^\pm$ (isolated)
$\ell_2^\mp$ (not-isolated )	$\ell_2^\pm$ (not-isolated)

(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,  $|\eta| < 2.4$ , 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. The ratio  $f_\tau$  of tight isolated  $\tau$  to very loose isolated  $\tau$  is extracted first. Then the misidentified  $\tau$  leptons rate ratio  $\tau(f_\tau)$  is calculated as in Equation . 6.2, in which  $\ell$  stands for e,  $\mu$  or  $\tau$  and  $N_\ell$  stands for the number of leptons. The misidentified leptons in the signal region are estimated with  $\tau(f_\tau)$  applied to the events in Region III which is defined in Table. 6.1, specifically, 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.

$$f_\ell = \frac{N_\ell(Z + jets \ell \text{ tight Iso})}{N_\tau(Z + jets \ell \text{ loose Iso})} \quad (6.1)$$

$$\ell(f_\ell) = \frac{f_\ell}{1 - f_\ell} \quad (6.2)$$

$$(6.3)$$

The expected contribution from WW, WZ, ZZ samples are subtracted with simulation, in which the third lepton is genuine lepton. The misidentified  $\tau$  lepton rate  $\tau(f_\tau)$  shows dependence on tau  $p_T$ , tau decay mode and spacial geometry  $\eta$ . 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  $f_\tau$  on  $\eta$  is also shown, which show less dependency.

The estimation of the  $\mu$  misidentified rate the is same as the  $\tau$  misidentified rate, but checks if the jets in Z + jets samples pass the muon isolation. As in Equation. 6.2,  $f_\mu$  is defined as the ratio between tight cut based muon isolation( $I_{rel}^\mu < 0.15$ ) and loose cut based muon isolation( $I_{rel}^\mu < 0.25$ ). The misidentified  $\mu$  lepton rate  $\mu(f_\mu)$  is calculated by Equation. 6.1.  $\mu(f_\mu)$  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 is double-counting between the estimated misidentified muon and tau lepton when both of the leptons are fake ones. The double-counting events are estimated as in Equation. 6.5. 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

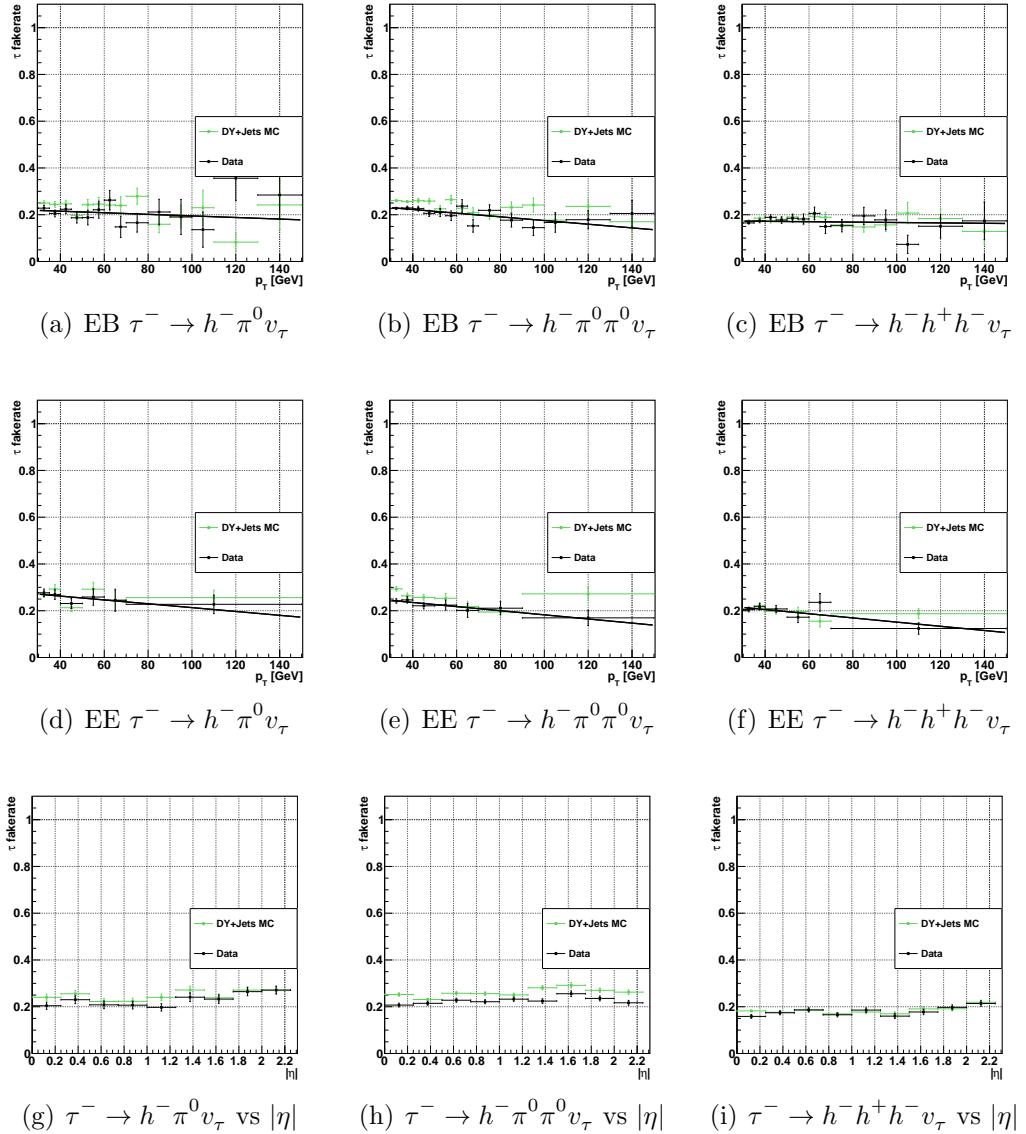


Figure 6.1.  $f_\tau$  ratio for the  $\tau$  fake rate calculation shows in term of tau decay modes,  $p_T$ , ECAL barrel, ECAL endcap, and  $|\eta|$

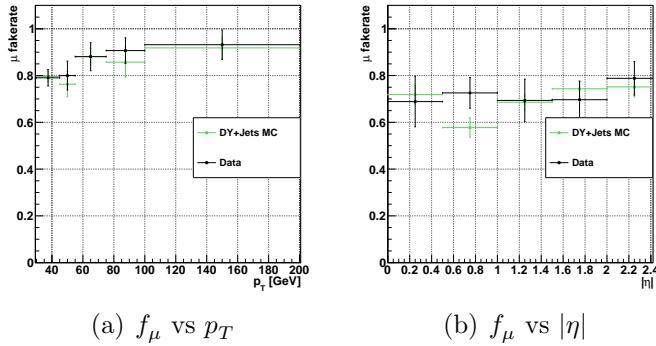


Figure 6.2.  $f_\tau$  ratio for the  $\mu$  fake rate calculation shows in term of muon  $p_T$  and  $|\eta|$ .

background obtained by adding the misidentified muon and tau background and deducts the double-counting events between these two.

$$N(\text{double-counting}) = \frac{f_\tau \cdot f_\mu}{(1 - f_\tau) \cdot (1 - f_\mu)} \cdot N(\text{Region III}) \quad (6.4)$$

In  $H \rightarrow \mu\tau_h$  channel, the validation of the misidentified lepton background with a fully data driven method is performed in a like-sign sample and also a 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. The misidentified background is enriched by 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, the misidentified background is the dominant one, as shown in Fig. 6.3.

Applying the  $\tau(f_\tau)$  as weight, the like-sign and W+jets control regions show good data and MC agreement with the fully data driven method, .

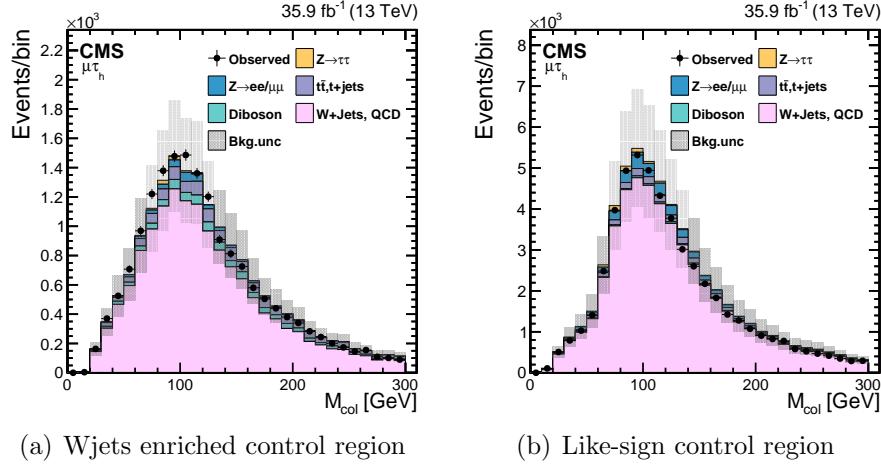


Figure 6.3. The validation of the fully data driven method for the misidentified lepton background. The uncertainties in the distributions include both statistics and systematics uncertainties.

### 6.1.2 Other backgrounds

The decays of  $Z \rightarrow \tau\tau$  and  $Z \rightarrow \mu\mu$  are also major backgrounds in this channel. In  $Z \rightarrow \tau\tau$ , the muons arise from a tau lepton decay, while in  $Z \rightarrow \mu\mu$ , one of the muon 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( $H \rightarrow \tau\tau, WW, t\bar{t}$ ), single top quark, di-boson(WW,WZ,ZZ). A full list of background used is in the

Table. ???. These backgrounds are all estimated with Monte Carlo simulation and normalized to the cross-section of collision dataset.

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 Monte Carlo 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

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

### 6.2.1 Misidentified Leptons

The misidentified lepton background is significant in the  $H \rightarrow e\tau_h$  channel. This background arises mainly from W+jets and QCD multijets. The events may

have one or both leptons are misidentified. The  $H \rightarrow e\tau_h$  channel in the analysis performed on the 8 TeV dataset, both misidentified electron and tau leptons are estimated.

Similar to  $H \rightarrow \mu\tau_h$  channel, the misidentified background in  $H \rightarrow e\tau_h$  is estimated with a fully data driven method, with independent collision dataset of Z+jets, in which the Z boson decays into two electrons. Data sets are divided into four control regions for this estimation as in Table. 6.1. Region II is used to extract the 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, a trigger that selects electron with  $p_T > 27$  GeV(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( $I_{rel}^e < 0.1$ ). The invariant mass formed by the two electrons are required to be in the Z boson mass window( $70$  GeV  $< M_{ee} < 110$  GeV). 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 the tau decay mode finder and discriminator against electrons and muons. Electrons pass the same conditions as the elections used in forming the Z boson mass peak, but with different isolation requirements. Tau leptons use the cut based tight or loose isolation, while loose or tight MVA based electron isolation have the value  $I_{rel}^e < 0.1$  and  $I_{rel}^e < 0.2$  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. 6.1. 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. 6.2. The fake ratio for tau leptons and electrons are shown in Figure. 6.4.

The tau and electron fake ratio is applied as a function of  $\eta$ . The background of three prompt leptons from WW, WZ, ZZ processes are deducted with Monte Carlo simulations. The tau and electron misidentified backgrounds are estimated separately, the double-counting events in which both leptons are misidentified objects are estimated with Equation. 6.5. In the equation,  $N(\text{Region III})$  in  $H \rightarrow e\tau_h$  channel satisfies the condition that both leptons are loose isolated but not tight isolated. The validation of the fully 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. The observed data agrees well with the events from MC and misidentified background from data driven method.

### 6.2.2 $Z \rightarrow \tau\tau$

$Z \rightarrow \tau\tau$  is another significant background in the  $H \rightarrow e\tau_h$  channel. In the case one of the tau leptons decay to a electron and the other hadronically decay. The background is estimated with an embedding technique[37], 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 modelled from data. Compared with the MC prediction, the visible mass peak from the modeling of the embedding sample shifts 2%. This is due to the difference in the final-state radiation of photons from the muon and tau. Identification and isolation correction for this shift to embedded sample are obtained by comparing with Monte Carlo

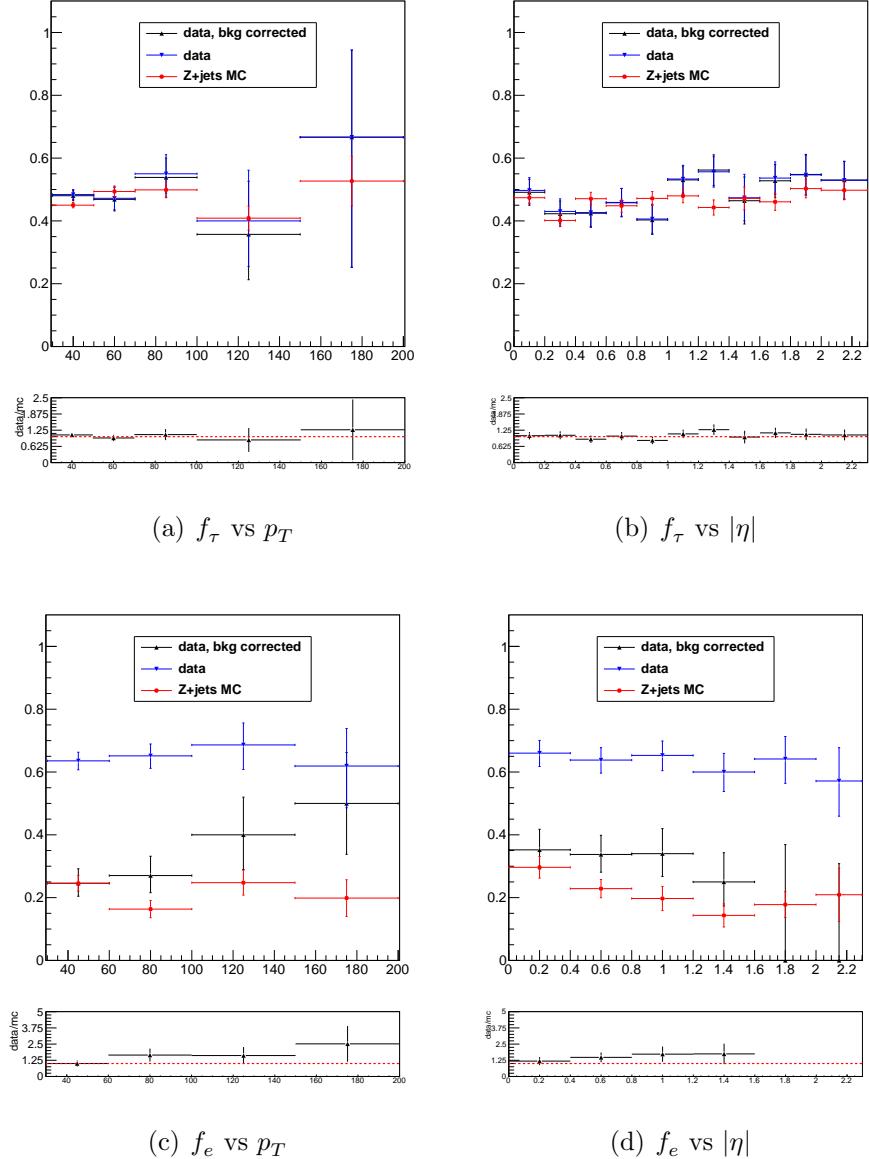


Figure 6.4.  $f_\tau$  and  $f_e$  ratio for the mididentification calculation shows in term of  $p_T$  and  $|\eta|$

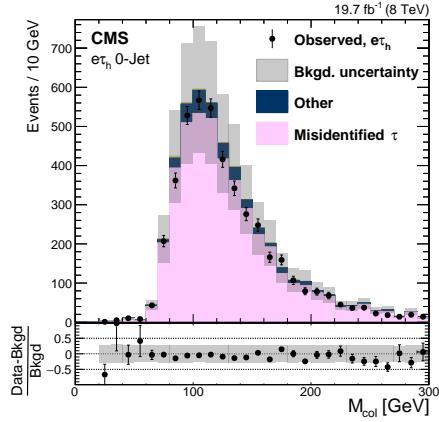


Figure 6.5. Data driven method validation in Region II with  $e$  and  $\tau$  leptons in same sign of charge

simulations.

### 6.2.3 Other backgrounds

The production of  $t\bar{t}$  pairs is also a background. The leptonic decay W bosons can produce opposite sign lepton pair and missing energy. The shape of this background is obtained from simulation samples and normalized to  $t\bar{t}$  control region sample. The control region is defined as the 2-jet selections in this channel but with additional requirement that at least one of the jets is b-tagged. The events 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  $W\gamma^{(*)}$ . These backgrounds are all estimated from Monte Carlo simulation.

## CHAPTER 7

### Signal extraction and systematics

#### 7.1 Boosted decision trees method

Boosted decision trees(BDT) [66, 80] 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.

$$P = \frac{\sum_s W_s}{\sum_s W_s + \sum_b W_b} \quad (7.1)$$

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

$$Criterion = Gini_{n-1} - Gini_{n \ left} - Gini_{n \ right} \quad (7.3)$$

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

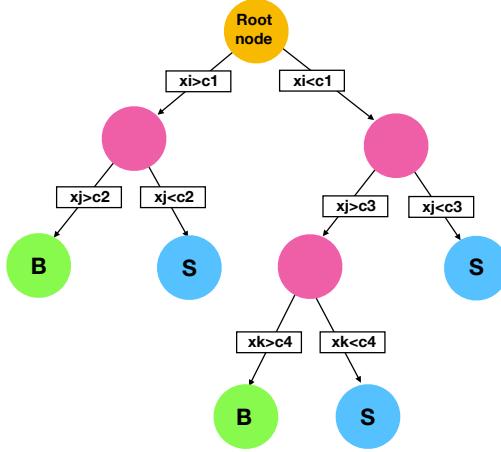


Figure 7.1. Tree structure example in BDT

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.

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

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

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

$$w_i \rightarrow w_i / \sum_{i=1}^N w_i \quad (7.7)$$

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

## 7.2 Statistical methods

The physics results are extracted with statistical methods. The limit of branching ratio with confident level(CL) method, significance and best fit branching fraction are mentioned in the analysis.

Probability density function(PDF) and likelihood function are the base functions used. PDF of a continuous variable  $x$  can be interpreted as the likelihood of  $x$  at its different values. Probability density function holds the property that it is normalized to unity.

$$\int f(x) dx = 1$$

Usually PDF of a variable interested accompanied with a set of parameters,

so as PDF is usually written in the form  $f(data|\alpha)$ , which means the probability density function of variables in data given the parameter set  $\alpha$ . These parameters can come from various sources, for example, from theory estimation, detector response and Monte Carlo simulation. All of the parameters in a model or a system, besides the parameters of interest, the other parameters that have an impact on the results are referred as nuisance parameters.

Likelihood function of the same model  $L(data|\alpha)$  is a function of parameter set  $\alpha$  given data. In particle physics, Likelihood function more often in the form of  $L(data|\mu, \theta)$ , in which  $\mu$  stands for signal strength modifier and  $\theta$  stands for a full suite of nuisance parameters. Data in Likelihood function can be experimental data or pseudo-data which is produced to generate sample distribution. Taking the counting experiment as an example, PDFs and likelihood function in a binned sample can be estimated with Poisson distribution as the PDF can be expressed as

$$f(data|u, \theta) = \prod_i \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-\mu s_i - b_i}$$

The  $s_i$  and  $b_i$  stand for the expected number of signal and background events respectively and both of them are a function of nuisance parameter  $\theta$  as  $s_i(\theta)$  and  $b_i(\theta)$ .  $n_i$  represents the number of events observed in bin  $i$ . The likelihood function in numerical form is similar to PDFs but implement the systematic error pdfs  $p(\theta|\tilde{\theta})$  if there are systematics considered in the model [3]. The form of likelihood function can be expressed as

$$\mathcal{L} = \prod_i \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-\mu s_i - b_i} \cdot p(\tilde{\theta}|\theta)$$

Both PDF and likelihood function rely on the knowledge of nuisance parameters. Auxiliary measurements or control regions are often used for the scale estimation of systematic uncertainties. With observations in different auxiliary measurements, a PDF  $p(\tilde{\theta}|\theta)$  which is referred as the posteriors can be measured. Together with  $\pi(\theta)$  which is referred as a prior, the systematic error pdfs can be constructed under the Bayes' theorem as

$$\rho(\theta|\tilde{\theta}) \sim p(\tilde{\theta}|\theta) \cdot \pi_\theta(\theta)$$

The systematic error pdfs can be improved and less affected by the choice of prior  $\pi(\theta)$  [73] by the auxiliary measurements.

In the search of lepton flavour violaton higgs decay, the discovery will be finding the higgs events either decay into  $\mu\tau$  pair in 13 TeV search or  $e\tau$  pair in 8 TeV search. LFV is forbidden in Standard Model(SM), thus, if take SM as a background only model, roughly speaking, a discovery can be claimed if the observation is not compatible with the background only model [46]. To test the compatibility of the observed data with respect to the background only model and further calculate limits on the signal strength modifier of the signal+background hypothesis, the quantity test statistics and the PDFs of the two model under tested are needed.

Test statistics is a function which can map a set of data into a number. According to the Neyman-Pearson lemma, likelihood ratio of two model under test gives the most powerful test, which is expressed as

$$\tilde{q}_\mu = -2\ln \frac{\mathcal{L}(data|\mu, \tilde{\theta}_\mu)}{\mathcal{L}(data|\hat{\mu}, \hat{\theta})}$$

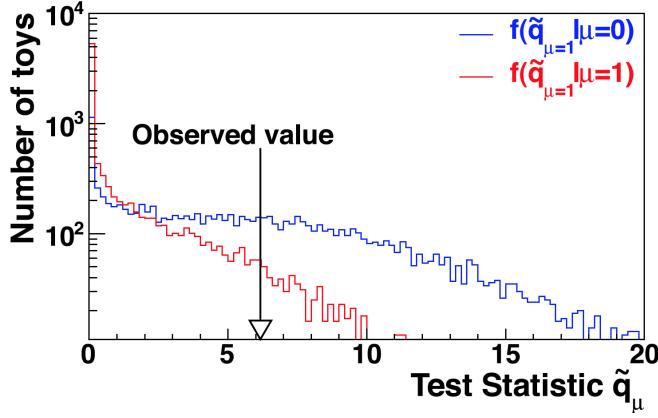


Figure 7.2. Test statistics distribution of signal+background and background only PDFs and the observed valued shown in arrow.

$\mu$  is the given signal strength modifier and  $0 < \hat{\mu} < \mu$ . In the formula of  $\tilde{q}_\mu$ ,  $\hat{\theta}$  is referred as conditional maximum likelihood estimator, given the signal strength modifier  $\mu$ , while in the denominator,  $\hat{\mu}$  and  $\hat{\theta}$  are allowed to flow freely to maximize the likelihood. The PDFs of signal+background model and background only model can be calculated in the following way. The observed  $\theta_\mu^{obs}$  and  $\theta_0^{obs}$  for signal+background and background only model are obtained by maximizing the likelihood function respectively, also the observed  $\tilde{q}_\mu^{obs}$  are calculated give  $\mu$ . The PDFs of signal+background model  $f(\tilde{q}_\mu|\mu, \theta_\mu^{obs})$  and background only model  $f(\tilde{q}_\mu|0, \theta_0^{obs})$  are constructed with the MC pseudo-data as shown in Figure. 7.2.

Confident Level(CLs) are defined with two values  $p_\mu$  and  $1-p_b$ . In signal+background model,  $p_\mu$  is defined as

$$p_\mu = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs} | \text{signal+background}) = \int_{\tilde{q}_\mu^{obs}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{obs}) d\tilde{q}_\mu$$

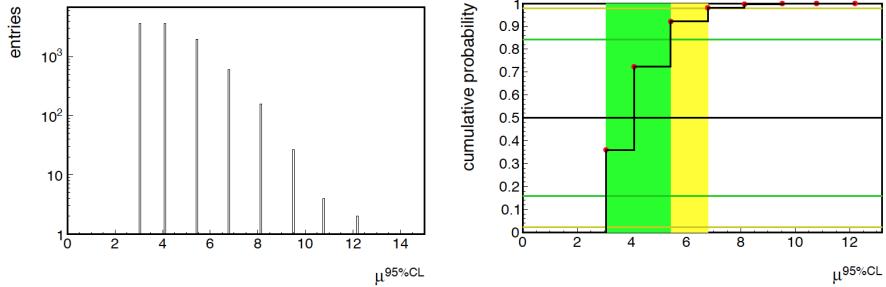


Figure 7.3. Signal strength modifier  $\mu$  at 95%  $CL_s$  distribution from the MC pseudo-data. The right plot is the cumulative distribution of  $\mu^{95\%}$  with  $\pm 1\sigma$  and  $\pm 2\sigma$  bands.

In background only model  $1-p_b$  is defined as

$$1 - p_b = P(\tilde{q} \geq \tilde{q}_\mu^{obs} | \text{background only}) = \int_{\tilde{q}_0^{obs}}^{\infty} f(\tilde{q}_\mu | 0, \hat{\theta}_0^{obs}) d\tilde{q}_\mu$$

The  $CL_s$  then is expressed as

$$CL_s = \frac{p_\mu}{1 - p_b}$$

With a given value of signal strength modifier  $\mu$ ,  $CL_s \leq \alpha$ , then the signal+background is said excluded with  $(1 - \alpha)$   $CL_s$ . Usually  $\alpha$  is picked as 95%.

The expected limit shown in lepton flavour violation Higgs decay is the median 95%  $CL_s$  upper limit with the  $\pm 1\sigma$  and  $\pm 2\sigma$  bands in the background only model. By generating a large number of background only pseudo-data, the CLs and  $\mu^{95\%}$  of each toy are calculated. An example of  $\mu_{95\%}$  distributions is shown in Figure. 7.3. As shown in the cumulative distribution of  $\mu^{95\%}$ , the median corresponds to 50%, while  $\pm 1\sigma$ ,  $\pm 2\sigma$  correspond to range 16% to 84% and 2.5% to 97.5% respectively.

The estimator p value and significance are used to check if the data is compatible with background only model. Test statistics in the checking of background only model is expressed as

$$\tilde{q}_\mu = -2\ln \frac{\mathcal{L}(data|0, \tilde{\theta}_\mu)}{\mathcal{L}(data|\hat{\mu}, \hat{\theta})}$$

Similar procedure as mentioned above to get the PDF of background ground only model  $f(q_0|0, \hat{\theta}_0^{obs})$  and  $q_0^{obs}$ . p-value corresponding to the experimental observable is calculated as

$$p_0 = P(q_0 \geq q_0^{obs}) = \int_{q_0^{obs}}^{\infty} f(q_0|0, \hat{\theta}_0^{obs}) dq_0$$

Significance Z of the observable with respect to the background only model test can be derived from p value [3].

$$p = \int_Z^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx$$

The term used in the discovery as  $5\sigma$  correspond to  $Z=5$  and  $p=2.8 \times 10^{-7}$ . The distribution of test statistics tends to a chi-squared distribution as according to the Wilk's theorem, which gives a quick way of estimating of pdfs in test statistics [46]. In the case when the expected number of events is large, expected limit with this asymptotic approximation gives a fairly well performance and save the computing time for toy samples. Asymptotic approximation can also used in the estimation

of p value as

$$p^{estimate} = \frac{1}{2} \left[ 1 - erf \left( \sqrt{q_0^{obs}/2} \right) \right]$$

### 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 [33, 39, 67, 68, 70]. 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  $Z \rightarrow \tau\tau$ , 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  $Z \rightarrow \mu\mu$  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.  $Z \rightarrow \mu\mu$  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 [49].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 [7] amounts to 2.5%.

TABLE 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

Systematic uncertainty	$H \rightarrow \mu\tau_h$
Muon trigger/identification/isolation	2%
Hadronic tau lepton efficiency	5%
b tagging veto	2.0–4.5%
$Z \rightarrow \tau\tau + \text{jets}$ background	10% $\oplus$ 5%
WW, ZZ background	5% $\oplus$ 5%
t <bar>t background</bar>	10% $\oplus$ 5%
Single top quark background	5% $\oplus$ 5%
$\mu \rightarrow \tau_h$ background	25%
Jet $\rightarrow \tau_h, \mu$ 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  $H \rightarrow \mu\tau_h$  analysis

Systematic uncertainty	$H \rightarrow \mu\tau_h$
$\mu \rightarrow \tau_h$ energy scale	1.5%
$\mu$ energy scale	0.2%
Unclustered energy scale	$\pm 1\sigma$
Renorm./fact. scales (g gH) [49]	3.9%
Renorm./fact. scales (VBF and VH) [49]	0.4%
PDF + $\alpha_s$ (g gH) [49]	3.2%
PDF + $\alpha_s$ (VBF and VH) [49]	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%

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

The systematic uncertainties affect the  $H \rightarrow e\tau_h$  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 [32, 70]. The uncertainties from the  $Z \rightarrow \tau\tau$  is from the uncertainty of the cross section measurement [36] and the  $\tau$  identification in the embedded technique. The normalization uncertainty on  $Z \rightarrow \mu\mu$  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  $Z \rightarrow ee$  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 ratio one standard deviation. The uncertainty from the pileup process are estimated by varying the total inelastic cross section by  $\pm 5$  percentage [35]. 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  $\gamma/Z+jets$  and dijet data [31]. 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  $Z \rightarrow \tau\tau$  distruption between data and MC sample and applied as a shape uncertainty. The theoretical uncertainties considered in the  $H \rightarrow e\tau_h$  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 [75], MSTW [74], NNPDF [19] from the recommendation in PDF4LHC [22]. 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  $H \rightarrow e\tau_h$  analysis. The uncertainties are correlated between categories besides the ones after  $\oplus$ . These uncertainties are uncorrelated between categories.

Systematic uncertainty	$H \rightarrow e\tau_h$		
	0-jet	1-jet	2-jet
Electron trigger/ID/isolation	1%	1%	2%
Efficiency of $\tau_h$	6.7%	6.7%	6.7%
$Z \rightarrow \tau\tau$ background	3% $\oplus$ 5%	3% $\oplus$ 5%	3% $\oplus$ 10%
$Z \rightarrow ee$ 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	$H \rightarrow e\tau_h$
$Z \rightarrow ee$ 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 <sup>-</sup> %	4%	1.5%	2%
Underlying event/parton shower	4%	5 <sup>-</sup> %	10 <sup>-</sup> %	10%	<1%	1 <sup>-</sup> %

## CHAPTER 8

### LFV Higgs decay searching results

The 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 fraction with  $CL_s$  criterion are set with asymptotic formula.

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

In the 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 the  $M_{col}$  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, the BDT discriminator is used as the variable to distinguish signal from background. After fitting to all of the distributions, the observed data versus 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  $H \rightarrow \mu\tau$  with CMS 13 TeV,  $36 \text{ fb}^{-1}$  data is motivated by the previous CMS search on the 8 TeV dataset with  $19.7 \text{ fb}^{-1}$   $H \rightarrow \mu\tau$ , in which a significance of 2.4 standard deviation excess of data with respect to the SM background-only hypothesis was observed. This thesis describes the  $H \rightarrow \mu\tau_h$  search, which combined together with another search for  $H \rightarrow \mu\tau_e$  [43], where the tau lepton decays electronically to an electron, instead of tau hadronically decay, provides the most sensitive search for  $H \rightarrow \mu\tau$ . The combined result is shown with the 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 previous 8 TeV result, the new upper limits on the signal branching fraction improved a factor of 5. The BDT fit analysis is more sensitive than the the  $M_{col}$  fit analysis. The new and tighter limit result comes from the BDT fit analysis.

The constrains on flavour violating Yukawa couplings can be derived from the limits on branching fraction of LFV Higgs decay. The exact relationship between Yukawa couplings and branching fraction of LFV Higgs decays is shown in section. 2.2.1. The 125 GeV Standard Model Higgs decay width is taken as  $\Gamma_{\text{SM}} = 4.1 \text{ MeV}$  [51]. 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.

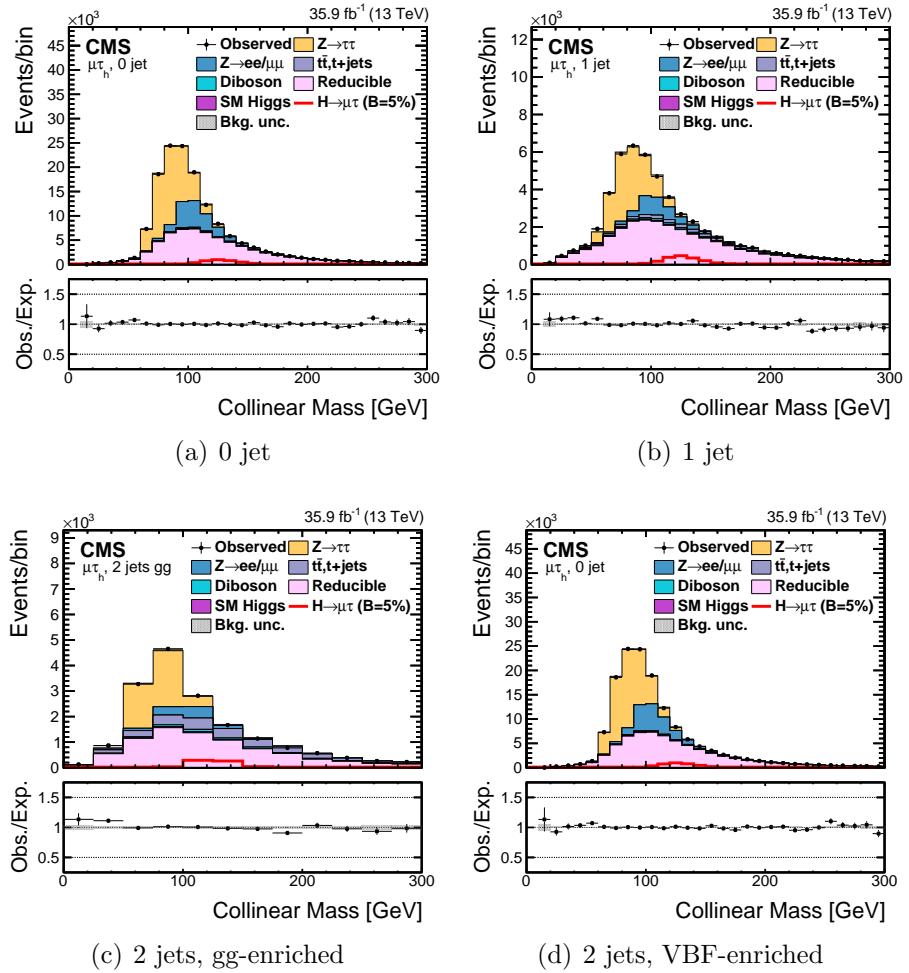


Figure 8.1.  $M_{col}$  distribution after the selection. The signal and backgrounds in the plots have been normalized to the best fit values. The gray bands shows the total uncertainties in each bin and the signal is plotted with the branching ratio of 5% of the Higgs decay branching ratio for visualization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

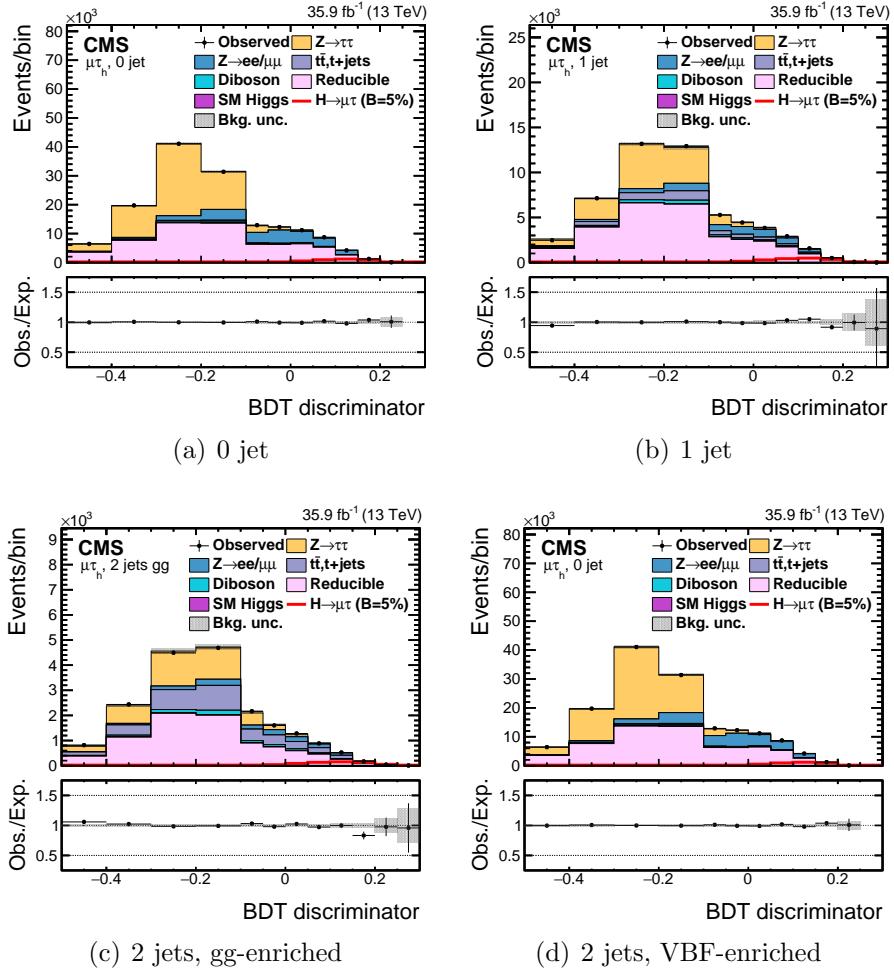


Figure 8.2. The BDT discriminator distribution after the selection. The signal and backgrounds in the plots has been normalized to the best fit values. The gray bands shows the total uncertainties in each bin and the signal is plotted with the branching ratio of 5% of the Higgs decay branching ratio 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 $\pm$ 0.45	0.68 $\pm$ 0.56	-1.23 $\pm$ 1.04	-0.23 $\pm$ 0.66	-0.08 $\pm$ 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
$\mu\tau_h$	$0.12 \pm 0.20$	$-0.05 \pm 0.25$	$-0.72 \pm 0.43$	$-0.22 \pm 0.31$	$-0.04 \pm 0.14$

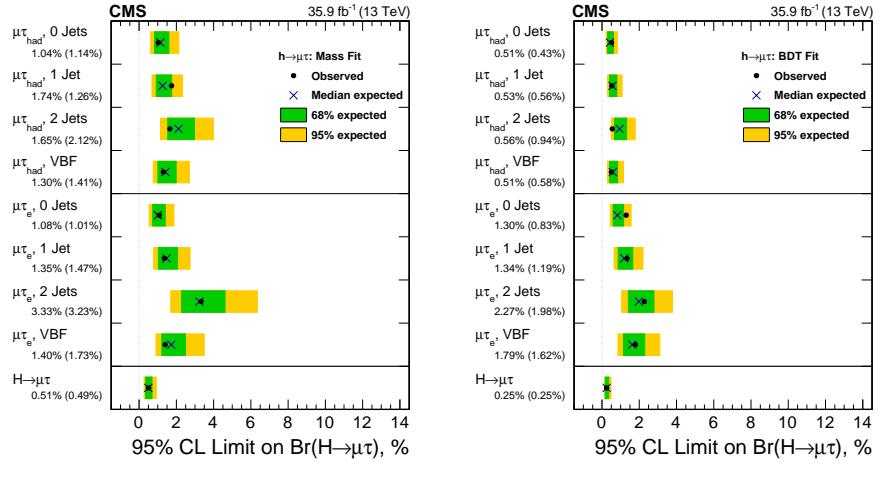


Figure 8.3. The expected and observed 95% CL limits on branching ratio of total  $H \rightarrow \mu\tau$  which shows the results from both  $H \rightarrow \mu\tau_h$  and  $H \rightarrow \mu\tau_e$  in the  $M_{col}$  fit analysis and BDT fit analysis.

### 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  $H \rightarrow e\tau_\mu$  [6]. 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  $\mathcal{B}(H \rightarrow e\tau)$

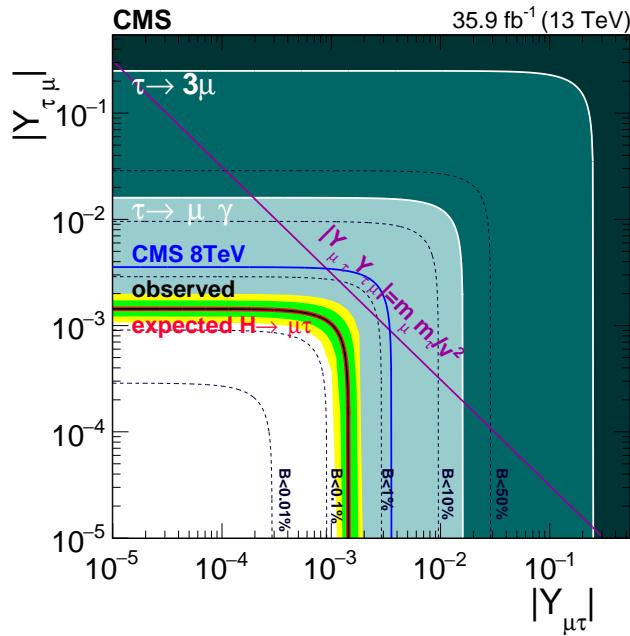


Figure 8.4. Upper limit of flavour violating Yukawa coupling  $|Y_{\mu\tau}|, |Y_{\tau\mu}|$  from the BDT fit analysis. The red solid line represents the expected limit and the black solid line is the observed limit. 68% and 95% range of containing the observed limit is shown in the green and yellow band. The result from  $\tau \rightarrow 3\mu$  is shown in dark green [64, 65, 78] and the one from  $\tau \rightarrow \mu\gamma$  is shown in lighter green [64, 78]. The theoretical naturalness limit  $Y_{ij}Y_{ji} \leq m_i m_j / v^2$  is shown by purple diagonal line [64].

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. The Yukawa coupling between electron and tau leptons is related to decay width  $\Gamma$  and branching fraction  $\mathcal{B}$  as shown in Equation. 2.18 and 2.19. 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.

TABLE 8.3: Events yields of the signal and backgrounds are shown in the mass range  $100 \text{ GeV} < M_{col} < 150 \text{ GeV}$ . 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  $19.7 \text{ fb}^{-1}$ .

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

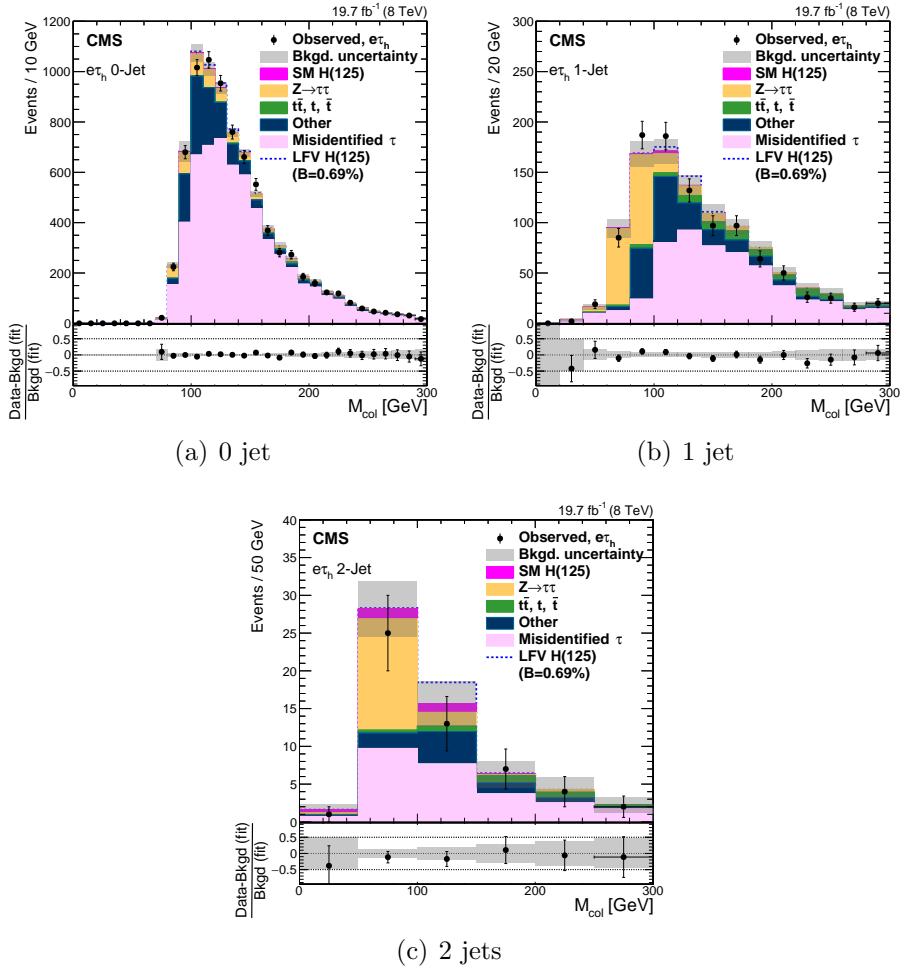


Figure 8.5.  $M_{col}$  distribution after the selection. The signal and backgrounds in the plots have been normalized to the best fit values. The gray bands shows the total uncertainties in each bin and the signal is plotted with the branching ratio of 0.69% of the Higgs decay branching ratio for visualization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

TABLE 8.4

$M_{col}$  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  $H \rightarrow e\tau_\mu$  analyses. 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 (%)			
$e\tau_\mu$	$<1.63^{(+0.66)}_{(-0.44)}$	$<1.54^{(+0.71)}_{(-0.47)}$	$<1.59^{(+0.93)}_{(-0.55)}$
$e\tau_h$	$<2.71^{(+1.05)}_{(-0.75)}$	$<2.76^{(+1.07)}_{(-0.77)}$	$<3.55^{(+1.38)}_{(-0.99)}$
Observed limits at 95% CL (%)			
$e\tau_\mu$	$<1.83$	$<0.94$	$<1.49$
$e\tau_h$	$<3.92$	$<3.00$	$<2.88$
$e\tau$	$<0.69$		

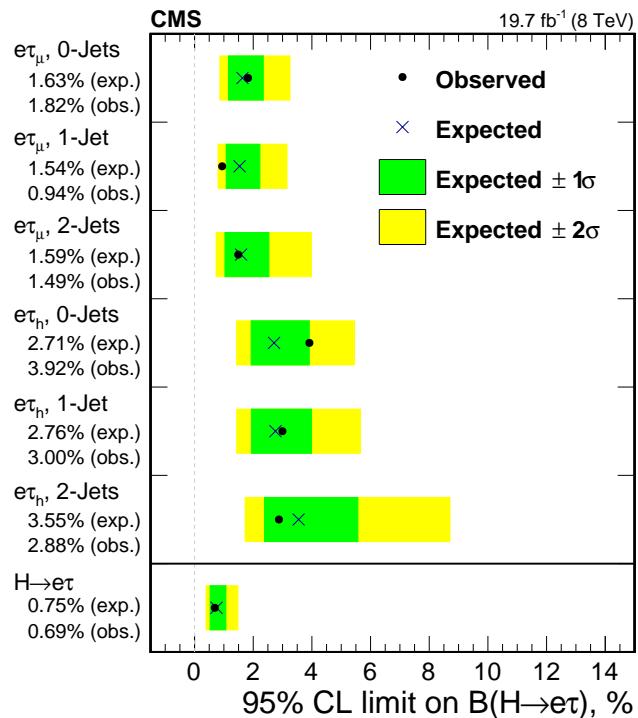


Figure 8.6. 95% CL upper limit on branching ratio on  $H \rightarrow e\tau$  with both the results from  $H \rightarrow e\tau_h$  and  $H \rightarrow e\tau_\mu$  and the combined.

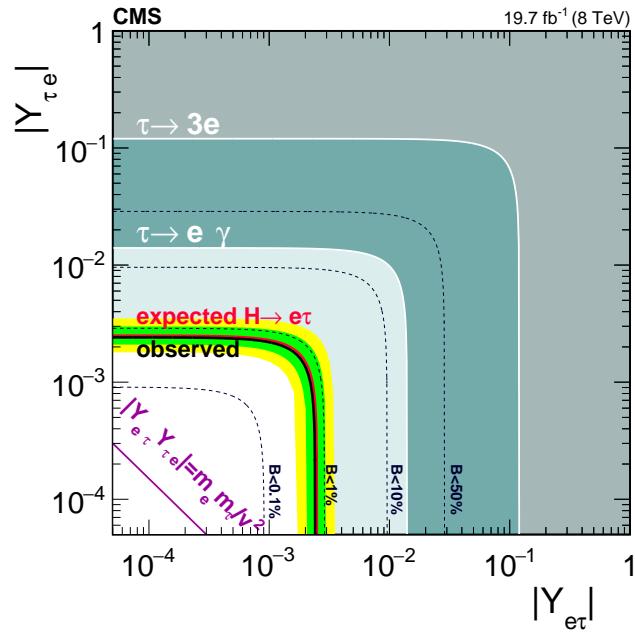


Figure 8.7. Upper limit of flavour violating Yukawa coupling  $|Y_{e\tau}|, |Y_{\tau e}|$  from the combined  $H \rightarrow e\tau$  result. The red solid line represents the expected limit and the black solid line is the observed limit. 68% and 95% range of containing the observed limit is shown in the green and yellow band. The result from  $\tau \rightarrow 3e$  is shown in gray, the one from  $\tau \rightarrow e\gamma$  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 [64]

## CHAPTER 9

### Conclusion

The operation of LHC opens a new era for experimental particle physics search with colliders. The great performance of general purpose detector CMS provides the possibilities to fully explore new physics. Through out this dissertation, the theory background for the search, an introduction of LHC and CMS, the detailed information about the data sets used, the reconstruction methods for the events in CMS and the search in two channels of LFV Higgs decays are presented. The focus is the physics search.

The search for standard model Higgs lepton flavour violation decay on Large hadron collider with CMS detector. One search on  $H \rightarrow e\tau_h$  with  $19.7 \text{ fb}^{-1}$  8 TeV data and a continue search on  $H \rightarrow \mu\tau_h$  with  $35.9 \text{ fb}^{-1}$  13 TeV. The results from both of the searches are presented, also the final results which combined with search in which the tau leptons decay electronically. The  $H \rightarrow e\tau$  search gave the upper limit from the direct measurement for the first time. The combined limit on branching fraction in  $H \rightarrow e\tau_h$  came as  $B < 0.69$  at 95% CL and on Yukawa coupling as  $Y_{e\tau} < 2.4 \times 10^{-3}$ . The continue search on  $H \rightarrow \mu\tau_h$  with  $35.9 \text{ fb}^{-1}$  13 TeV is also presented. Even though the excess observed in 8 TeV search in  $H \rightarrow \mu\tau$  channel is gone, the combined result from 13 TeV search is the tightest limit so far. In  $H \rightarrow \mu\tau$ , a combined limit on is  $B < 0.25\%$  and the upper limit on Yukawa coupling is  $Y_{\mu\tau} < 1.43 \times 10^{-3}$ .

The latest search for 125 GeV Higgs LFV decay  $H \rightarrow \mu\tau$  is presented in the dissertation. While for the  $H \rightarrow e\tau$  search, the 13 TeV results published together with  $H \rightarrow \mu\tau$  search here [44]. Though no significant excesses are observed in both  $\mu\tau$  and  $e\tau$  channel, compared to the  $H \rightarrow e\mu$  channel, there is still huge room to explore in this topic. Whether or not LFV is observed, this is an important topic which has big influences on new physics searches.

## BIBLIOGRAPHY

1. LHC Higgs cross section working group. URL <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG>.
2. Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV. Technical Report CMS-PAS-PFT-10-003, CERN, Geneva, 2010. URL <http://cds.cern.ch/record/1279347>.
3. Procedure for the LHC Higgs boson search combination in Summer 2011. Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug 2011. URL <https://cds.cern.ch/record/1379837>.
4. Pileup Jet Identification. Technical Report CMS-PAS-JME-13-005, CERN, Geneva, 2013. URL <http://cds.cern.ch/record/1581583>.
5. Search for lepton-flavour-violating decays of the higgs boson. *Physics Letters B*, 749:337 – 362, 2015. ISSN 0370-2693. doi: <https://doi.org/10.1016/j.physletb.2015.07.053>. URL <http://www.sciencedirect.com/science/article/pii/S0370269315005638>.
6. 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.

7. 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>.
8. 06 2018. URL [https://en.wikipedia.org/wiki/Standard\\_Model](https://en.wikipedia.org/wiki/Standard_Model).
9. G. Aad et al. Search for lepton-flavour-violating  $H^-$  decays of the Higgs boson with the ATLAS detector. *JHEP*, 11:211, 2015. doi: 10.1007/JHEP11(2015)211.
10. D. Acosta. Cms trigger improvements towards run ii. *Nuclear and Particle Physics Proceedings*, 273-275:1008 – 1013, 2016. ISSN 2405-6014. doi: <https://doi.org/10.1016/j.nuclphysbps.2015.09.158>. URL <http://www.sciencedirect.com/science/article/pii/S2405601415006471>. 37th International Conference on High Energy Physics (ICHEP).
11. W. Adam, R. Frühwirth, A. Strandlie, and T. Todor. Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC. 2005.
12. W. Adam, B. Mangano, T. Speer, and T. Todorov. Track Reconstruction in the CMS tracker. Technical Report CMS-NOTE-2006-041, CERN, Geneva, Dec 2006. URL <https://cds.cern.ch/record/934067>.
13. P. Adzic et al. Energy resolution of the barrel of the CMS electromagnetic calorimeter. *JINST*, 2:P04004, 2007. doi: 10.1088/1748-0221/2/04/P04004.
14. J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their

- matching to parton shower simulations. *JHEP*, 07:079, 2014. doi: 10.1007/JHEP07(2014)079.
15. C. Anastasiou, R. Boughezal, and F. Petriello. Mixed QCD-electroweak corrections to Higgs boson production in gluon fusion. *JHEP*, 04:003, 2009. doi: 10.1088/1126-6708/2009/04/003.
  16. D. Aristizabal Sierra and A. Vicente. Explaining the cms higgs flavor-violating decay excess. *Phys. Rev. D*, 90:115004, Dec 2014. doi: 10.1103/PhysRevD.90.115004. URL <https://link.aps.org/doi/10.1103/PhysRevD.90.115004>.
  17. M. Baber, M. Blake, J. Brooke, M. C. Hermida, S. Dasu, T. Durkin, S. Fayer, E. K. Friis, T. Gorski, G. Hall, K. Harder, G. Iles, S. Ives, J. Jones, P. R. Klabbers, A. G. Levine, C. Lucas, R. Lucas, D. Newbold, J. Marrouche, S. Paramesvaran, T. M. Perry, A. Rose, D. Sankey, W. Smith, A. Tapper, A. Thea, and T. Williams. Development and testing of an upgrade to the cms level-1 calorimeter trigger. *Journal of Instrumentation*, 9(01):C01006, 2014. URL <http://stacks.iop.org/1748-0221/9/i=01/a=C01006>.
  18. A. M. Baldini et al. Search for the lepton flavour violating decay  $\mu^+ \rightarrow e^+ \gamma$  with the full dataset of the MEG experiment. *Eur. Phys. J. C*, 76:434, 2016. doi: 10.1140/epjc/s10052-016-4271-x.
  19. 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.
  20. G. Blankenburg, J. Ellis, and G. Isidori. Flavour-changing decays of a 125

- GeV Higgs-like particle. *Phys. Lett. B*, 712:386, 2012. doi: 10.1016/j.physletb. 2012.05.007.
21. P. Bolzoni, F. Maltoni, S.-O. Moch, and M. Zaro. Higgs production via vector-boson fusion at NNLO in QCD. *Phys. Rev. Lett.*, 105:011801, 2010. doi: 10.1103/PhysRevLett.105.011801.
  22. 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.
  23. G. Branco, P. Ferreira, L. Lavoura, M. Rebelo, M. Sher, and J. P. Silva. Theory and phenomenology of two-higgs-doublet models. *Physics Reports*, 516(1):1 – 102, 2012. ISSN 0370-1573. doi: <https://doi.org/10.1016/j.physrep.2012.02.002>. URL <http://www.sciencedirect.com/science/article/pii/S0370157312000695>. Theory and phenomenology of two-Higgs-doublet models.
  24. M. Cacciari, G. P. Salam, and G. Soyez. The anti- $k_t$  jet clustering algorithm. *JHEP*, 04:063, 2008. doi: 10.1088/1126-6708/2008/04/063.
  25. M. Cacciari, G. P. Salam, and G. Soyez. FastJet user manual. *Eur. Phys. J. C*, 72:1896, 2012. doi: 10.1140/epjc/s10052-012-1896-2.
  26. R. N. Cahn, S. D. Ellis, R. Kleiss, and W. J. Stirling. Transverse-momentum signatures for heavy Higgs bosons. *Phys. Rev. D*, 35:1626, 1987. doi: 10.1103/PhysRevD.35.1626.
  27. S. Catani, D. de Florian, M. Grazzini, and P. Nason. Soft gluon resummation

- for Higgs boson production at hadron colliders. *JHEP*, 07:028, 2003. doi: 10.1088/1126-6708/2003/07/028.
28. A. Celis, V. Cirigliano, and E. Passemar. Lepton flavor violation in the Higgs sector and the role of hadronic tau-lepton decays. *Phys. Rev. D*, 89:013008, 2014. doi: 10.1103/PhysRevD.89.013008.
29. S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008. doi: 10.1088/1748-0221/3/08/S08004.
30. S. Chatrchyan et al. Aligning the CMS Muon Chambers with the Muon Alignment System during an Extended Cosmic Ray Run. *JINST*, 5:T03019, 2010. doi: 10.1088/1748-0221/5/03/T03019.
31. 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.
32. 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.
33. 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.
34. 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.

35. S. Chatrchyan et al. Measurement of the inelastic proton-proton cross section at  $\sqrt{s} = 7\text{TeV}$ . *Phys. Lett. B*, 722:5, 2013. doi: 10.1016/j.physletb.2013.03.024.
36. 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.
37. S. Chatrchyan et al. Evidence for the 125 GeV Higgs boson decaying to a pair of  $\tau$  leptons. 2014.
38. 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>.
39. 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>.
40. C. Collaboration. Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2. 2016.
41. 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.
42. 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>.

43. C. Collaboration. Search for lepton flavour violating decays of the Higgs boson to  $\tau$  and  $e$  in proton-proton collisions at  $\sqrt{s}= 13$  TeV. *JHEP*, 2017.
44. C. collaboration. Search for lepton flavour violating decays of the higgs boson to  $\mu\tau$  and  $e\tau$  in proton-proton collisions at  $\sqrt{s}=13$  tev. *Journal of High Energy Physics*, 2018(6):1, Jun 2018. ISSN 1029-8479. doi: 10.1007/JHEP06(2018)001. URL [https://doi.org/10.1007/JHEP06\(2018\)001](https://doi.org/10.1007/JHEP06(2018)001).
45. G. Collaboration. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303, 2003. ISSN 0168-9002. doi: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8). URL <http://www.sciencedirect.com/science/article/pii/S0168900203013688>.
46. K. Cranmer. Practical Statistics for the LHC. *CERN-2014-003*, pp. 267 - 308, 2015. doi: 10.5170/CERN-2015-004.301.
47. S. Dasu et al. CMS. The TriDAS project. Technical design report, vol. 1: The trigger systems. 2000.
48. D. de Florian, C. Grojean, F. Maltoni, C. Mariotti, A. Nikitenko, M. Pieri, P. Savard, M. Schumacher, R. Tanaka, R. Aggleton, M. Ahmad, B. Allanach, C. Anastasiou, W. Astill, S. Badger, M. Badziak, J. Baglio, E. Bagnaschi, A. Ballestrero, A. Banfi, D. Barducci, M. Beckingham, C. Becot, G. Blanger, J. Bellm, N. Belyaev, F. Bernlochner, C. Beskidt, A. Biektter, F. Bishara, W. Bizon, N. Bomark, M. Bonvini, S. Borowka, V. Bortolotto, S. Boselli, F. Botella, R. Boughezal, G. Branco, J. Brehmer, L. Brenner, S. Bressler, I. Brivio, A. Broggio, H. Brun, G. Buchalla, C. Burgard,

A. Calandri, L. Caminada, R. Caminal Armadans, F. Campanario, J. Campbell, F. Caola, C. Carloni Calame, S. Carrazza, A. Carvalho, M. Casolino, O. Cata, A. Celis, F. Cerutti, N. Chanon, M. Chen, X. Chen, B. Chokouf Nejad, N. Christensen, M. Ciuchini, R. Contino, T. Corbett, R. Costa, D. Curtin, M. Dall'Osso, A. David, S. Dawson, J. de Blas, W. de Boer, P. de Castro Manzano, C. Degrande, R. Delgado, F. Demartin, A. Denner, B. Di Micco, R. Di Nardo, S. Dittmaier, A. Dobado, T. Dorigo, F. Dreyer, M. Dhrssen, C. Duhr, F. Dulat, K. Ecker, K. Ellis, U. Ellwanger, C. Englert, D. Espriu, A. Falkowski, L. Fayard, R. Feger, G. Ferrera, A. Ferroglio, N. Fidanza, T. Figy, M. Flechl, D. Fontes, S. Forte, P. Francavilla, E. Franco, R. Frederix, A. Freitas, F. Freitas, F. Frensch, S. Frixione, B. Fuks, E. Furlan, S. Gadatsch, J. Gao, Y. Gao, M. Garzelli, T. Gehrmann, R. Gerosa, M. Ghezzi, D. Ghosh, S. Gieseke, D. Gillberg, G. Giudice, E. Glover, F. Goertz, D. Goncalves, J. Gonzalez-Fraile, M. Gorbahn, S. Gori, C. Gottardo, M. Gouzevitch, P. Govoni, D. Gray, M. Grazzini, N. Greiner, A. Greljo, J. Grigo, A. Gritsan, R. Grber, S. Guindon, H. Haber, C. Han, T. Han, R. Harlander, M. Harrendorf, H. Hartanto, C. Hays, S. Heinemeyer, G. Heinrich, M. Herrero, F. Herzog, B. Hespel, V. Hirschi, S. Hoeche, S. Honeywell, S. Huber, C. Hugonie, J. Huston, A. Ilnicka, G. Isidori, B. Jger, M. Jaquier, S. Jones, A. Juste, S. Kallweit, A. Kaluza, A. Kardos, A. Karlberg, Z. Kassabov, N. Kauer, D. Kazakov, M. Kerner, W. Kilian, F. Kling, K. Kneke, R. Kogler, R. Konoplich, S. Kortner, S. Kräml, C. Krause, F. Krauss, M. Krawczyk, A. Kulesza, S. Kuttmalai, R. Lane, A. Lazopoulos, G. Lee, P. Lenzi, I. Lewis, Y. Li, S. Liebler, J. Lindert, X. Liu, Z. Liu, F. Llanes-Estrada, H. Logan, D. Lopez-Val, I. Low, G. Luisoni, P. Maierhfer, E. Maina,

B. Mansouli, H. Mantler, M. Mantoani, A. Marini, V. Martinez Outschoorn, S. Marzani, D. Marzocca, A. Massironi, K. Mawatari, J. Mazzitelli, A. McCarn, B. Mellado, K. Melnikov, S. Menari, L. Merlo, C. Meyer, P. Milenovic, K. Mimasu, S. Mishima, B. Mistlberger, S.-O. Moch, A. Mohammadi, P. Monni, G. Montagna, M. Moreno Llcer, N. Moretti, S. Moretti, L. Motyka, A. Mck, M. Mhlleitner, S. Munir, P. Musella, P. Nadolsky, D. Napoletano, M. Nebot, C. Neu, M. Neubert, R. Nevzorov, O. Nicrosini, J. Nielsen, K. Nikolopoulos, J. No, C. O'Brien, T. Ohl, C. Oleari, T. Orimoto, D. Pagani, C. Pandini, A. Papaefstathiou, A. Papanastasiou, G. Passarino, B. Pecjak, M. Pelliccioni, G. Perez, L. Perrozzi, F. Petriello, G. Petrucciani, E. Pianori, F. Piccinini, M. Pierini, A. Pilkington, S. Pltzer, T. Plehn, R. Podskubka, C. Potter, S. Pozzorini, K. Prokofiev, A. Pukhov, I. Puljak, M. Queitsch-Maitland, J. Quevillon, D. Rathlev, M. Rauch, E. Re, M. Rebelo, D. Rebuzzi, L. Reina, C. Reuschle, J. Reuter, M. Riembau, F. Riva, A. Rizzi, T. Robens, R. Rntsch, J. Rojo, J. Romo, N. Rompotis, J. Roskes, R. Roth, G. Salam, R. Salerno, M. Sampaio, R. Santos, V. Sanz, J. Sanz-Cillero, H. Sargsyan, U. Sarica, P. Schichtel, J. Schlenk, T. Schmidt, C. Schmitt, M. Schnherr, U. Schubert, M. Schulze, S. Sekula, M. Sekulla, E. Shabalina, H. Shao, J. Shelton, C. Shepherd-Themistocleous, S. Shim, F. Siegert, A. Signer, J. Silva, L. Silvestrini, M. Sjodahl, P. Slavich, M. Slawinska, L. Soffi, M. Spannowsky, C. Speckner, D. Sperka, M. Spira, O. Stl, F. Staub, T. Stebel, T. Stefaniak, M. Steinhauser, I. Stewart, M. Strassler, J. Streicher, D. Strom, S. Su, X. Sun, F. Tackmann, K. Tackmann, A. Teixeira, R. Teixeira de Lima, V. Theeuwes, R. Thorne, D. Tommasini, P. Torrielli, M. Tosi, F. Tramontano, Z. Trcsnyi, M. Trott, I. Tsinkos, M. Ubiali, P. Vanlaer, W. Verkerke, A. Vicini, L. Viliani,

- E. Vryonidou, D. Wackerlo, C. Wagner, J. Wang, S. Wayand, G. Weiglein, C. Weiss, M. Wiesemann, C. Williams, J. Winter, D. Winterbottom, R. Wolf, M. Xiao, L. Yang, R. Yohay, S. Yuen, G. Zanderighi, M. Zaro, D. Zeppenfeld, R. Ziegler, T. Zirke, and J. Zupan. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. Technical Report FERMILAB-FN-1025-T, Oct 2016. URL <https://cds.cern.ch/record/2227475>. 869 pages, 295 figures, 248 tables and 1645 citations. Working Group web page: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG>.
49. 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.
  50. L. de Lima, C. S. Machado, R. D. Matheus, and L. A. F. do Prado. Higgs flavor violation as a signal to discriminate models. *Journal of High Energy Physics*, 2015(11):74, Nov 2015.
  51. 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.
  52. J. L. Diaz-Cruz and J. J. Toscano. Lepton flavor violating decays of higgs bosons beyond the standard model. *Phys. Rev. D*, 62:116005, Nov 2000. doi: 10.1103/PhysRevD.62.116005. URL <https://link.aps.org/doi/10.1103/PhysRevD.62.116005>.
  53. S. Dittmaier, C. Mariotti, G. Passarino, R. Tanaka, J. Baglio, P. Bolzoni, R. Boughezal, O. Brein, C. Collins-Tooth, S. Dawson, S. Dean, A. Denner, S. Farrington, M. Felcini, M. Flechl, D. de Florian, S. Forte, M. Grazzini,

- C. Hackstein, T. Hahn, R. Harlander, T. Hartonen, S. Heinemeyer, J. Huston, A. Kalinowski, M. Krmer, F. Krauss, J. S. Lee, S. Lehti, F. Maltoni, K. Mazumdar, S. O. Moch, A. Mck, M. Mhlleitner, P. Nason, C. Neu, C. Oleari, J. Olsen, S. Palmer, F. Petriello, G. Piacquadio, A. Pilaftsis, C. T. Potter, I. Puljak, J. Qian, D. Rebuzzi, L. Reina, H. Rzechak, M. Schumacher, P. Slavich, M. Spira, F. Stckli, R. S. Thorne, M. V. Acosta, T. Vickey, A. Vicini, D. Wackeroth, M. Warsinsky, M. Weber, G. Weiglein, C. Weydert, J. Yu, M. Zaro, and T. Zirke. *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*. CERN Yellow Reports: Monographs. CERN, Geneva, 2011. URL <https://cds.cern.ch/record/1318996>. Comments: 153 pages, 43 figures, to be submitted to CERN Report. Working Group web page: <https://twiki.cern.ch/twiki/bin/view/LHCPysics/CrossSections>.
54. S. Dittmaier, C. Mariotti, G. Passarino, R. Tanaka, S. Alekhin, J. Alwall, E. A. Bagnaschi, A. Banfi, J. Blumlein, S. Bolognesi, N. Chanon, T. Cheng, L. Cieri, A. M. Cooper-Sarkar, M. Cutajar, S. Dawson, G. Davies, N. De Filippis, G. Degrassi, A. Denner, D. D'Enterria, S. Diglio, B. Di Micco, R. Di Nardo, R. K. Ellis, A. Farilla, S. Farrington, M. Felcini, G. Ferrera, M. Flechl, D. de Florian, S. Forte, S. Ganjour, M. V. Garzelli, S. Gascon-Shotkin, S. Glazov, S. Goria, M. Grazzini, J. P. Guillet, C. Hackstein, K. Hamilton, R. Harlander, M. Hauru, S. Heinemeyer, S. Hoche, J. Huston, C. Jackson, P. Jimenez-Delgado, M. D. Jorgensen, M. Kado, S. Kallweit, A. Kardos, N. Kauer, H. Kim, M. Kovac, M. Kramer, F. Krauss, C. M. Kuo, S. Lehti, Q. Li, N. Lorenzo, F. Maltoni, B. Mellado, S. O. Moch, A. Muck, M. Muhlleitner, P. Nadolsky, P. Nason, C. Neu, A. Nikitenko, C. Oleari, J. Olsen, S. Palmer, S. Paganis, C. G. Papadopoulos, T. C.

- Petersen, F. Petriello, F. Petrucci, G. Piacquadio, E. Pilon, C. T. Potter, J. Price, I. Puljak, W. Quayle, V. Radescu, D. Rebuzzi, L. Reina, J. Rojo, D. Rosco, G. P. Salam, A. Sapronov, J. Schaarschmidt, M. Schonherr, M. Schumacher, F. Siegert, P. Slavich, M. Spira, I. W. Stewart, W. J. Stirling, F. Stockli, C. Sturm, F. J. Tackmann, R. S. Thorne, D. Tommasini, P. Torrielli, F. Tramontano, Z. Trocsanyi, M. Ubiali, S. Uccirati, M. V. Acosta, T. Vickey, A. Vicini, W. J. Waalewijn, D. Wackeroth, M. Warsinsky, M. Weber, M. Wiesemann, G. Weiglein, J. Yu, and G. Zanderighi. Handbook of LHC Higgs Cross Sections: 2. Differential Distributions. Technical Report arXiv:1201.3084. CERN-2012-002, Geneva, 2012. URL <https://cds.cern.ch/record/1416519>. Comments: 275 pages, 136 figures, to be submitted to CERN Report. Working Group web page: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections>.
55. A. Djouadi. The anatomy of electroweak symmetry breaking: Tome i: The higgs boson in the standard model. *Physics Reports*, 457(1):1 – 216, 2008. ISSN 0370-1573. doi: <https://doi.org/10.1016/j.physrep.2007.10.004>. URL <http://www.sciencedirect.com/science/article/pii/S0370157307004334>.
56. A. Djouadi, M. Spira, and P. M. Zerwas. Production of Higgs bosons in proton colliders: QCD corrections. *Phys. Lett.*, B264:440–446, 1991. doi: 10.1016/0370-2693(91)90375-Z.
57. S. D. Ellis and D. E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev.*, D48:3160–3166, 1993. doi: 10.1103/PhysRevD.48.3160.

58. L. Evans and P. Bryant. Lhc machine. *Journal of Instrumentation*, 3(08):S08001, 2008. URL <http://stacks.iop.org/1748-0221/3/i=08/a=S08001>.
59. R. Fruhwirth. Application of Kalman filtering to track and vertex fitting. *Nucl. Instrum. Meth.*, A262:444–450, 1987. doi: 10.1016/0168-9002(87)90887-4.
60. H. M. Georgi, S. L. Glashow, M. E. Machacek, and D. V. Nanopoulos. Higgs bosons from two gluon annihilation in proton proton collisions. *Phys. Rev. Lett.*, 40:692, 1978. doi: 10.1103/PhysRevLett.40.692.
61. V. Gori. The CMS High Level Trigger. *Int. J. Mod. Phys. Conf. Ser.*, 31:1460297, 2014. doi: 10.1142/S201019451460297X.
62. D. J. Griffiths. *Introduction to elementary particles; 2nd rev. version*. Physics textbook. Wiley, New York, NY, 2008. URL <https://cds.cern.ch/record/111880>.
63. R. Harnik, J. Kopp, and J. Zupan. Flavor violating higgs decays. *JHEP*, 03:26, 2013. doi: 10.1007/JHEP03(2013)026.
64. R. Harnik, J. Kopp, and J. Zupan. Flavor violating higgs decays. *JHEP*, 03:26, 2013. doi: 10.1007/JHEP03(2013)026.
65. 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.
66. 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.
67. 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.
  68. 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.
  69. 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.
  70. 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.
  71. V. Khachatryan et al. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *JINST*, 12:P02014, 2017. doi: 10.1088/1748-0221/12/02/P02014.
  72. C. Lefèvre. The CERN accelerator complex. Complexe des accélérateurs du CERN. Dec 2008. URL <https://cds.cern.ch/record/1260465>.
  73. L. Lista. Practical Statistics for Particle Physicists. 2016.

74. 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.
75. 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.
76. G. N. Davidson et al. Universal interface of tauola technical and physics documentation. doi: 10.1016/j.cpc.2011.12.009.
77. C. Olari. The POWHEG-BOX. *Nucl.Phys.Proc.Suppl.205-206:36-41,2010*, 2010. doi: 10.1016/j.nuclphysbps.2010.08.016.
78. 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.
79. M. E. Peskin and D. V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995. ISBN 9780201503975, 0201503972. URL <http://www.slac.stanford.edu/~mpeskin/QFT.html>.
80. B. P. Roe et al. Boosted Decision Trees as an Alternative to Artificial Neural Networks for Particle Identification. 2005.
81. D. Rabady, J. Ero, G. Flouris, J. Fulcher, N. Loukas, E. Paradas, T. Reis, H. Sakulin, and C.-E. Wulz. Upgrade of the CMS muon trigger system in the barrel region. *Nucl. Instrum. Meth.*, A845:616–620, 2017. doi: 10.1016/j.nima.2016.05.071.

82. A. M. Sirunyan et al. Particle-flow reconstruction and global event description with the cms detector. *JINST*, 12:P10003, 2017. doi: 10.1088/1748-0221/12/10/P10003.
83. A. M. Sirunyan et al. Particle-flow reconstruction and global event description with the cms detector. *JINST*, 12:P10003, 2017. doi: 10.1088/1748-0221/12/10/P10003.
84. A. M. Sirunyan et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST*, 13:P05011, 2018. doi: 10.1088/1748-0221/13/05/P05011.
85. T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159, 2015. doi: 10.1016/j.cpc.2015.01.024.
86. E. Yazgan and the CMS ECAL/HCAL Collaborations. The cms barrel calorimeter response to particle beams from 2 to 350 gev/c. *Journal of Physics: Conference Series*, 160(1):012056, 2009. URL <http://stacks.iop.org/1742-6596/160/i=1/a=012056>.
87. S. Yu.L.Dokshitzer, G.D.Leder and B.R.Webber. Better Jet Clustering Algorithms. *JHEP* 9708:001, 1997, 1997. doi: 10.1088/1126-6708/1997/08/001.
88. A. Zabi, F. Beaudette, L. Cadamuro, L. Mastrolorenzo, T. Romanteau, J. Sauvan, T. Strebler, J. Marrouche, N. Wardle, R. Aggleton, F. Ball, J. Brooke, D. Newbold, S. Paramesvaran, D. Smith, M. Baber, A. Bundred, M. Citron, A. Elwood, G. Hall, G. Iles, C. Laner, B. Penning, A. Rose, A. Tap-

per, T. Durkin, K. Harder, S. Harper, C. Shepherd-Themistocleous, A. Thea, and T. Williams. Triggering on electrons, jets and tau leptons with the cms upgraded calorimeter trigger for the lhc run ii. *Journal of Instrumentation*, 11(02):C02008, 2016. URL <http://stacks.iop.org/1748-0221/11/i=02/a=C02008>.

*This document was prepared & typeset with pdfL<sup>A</sup>T<sub>E</sub>X, and formatted with NDDiss2<sub>E</sub> classfile (v3.0[2005/07/27]) provided by Sameer Vijay.*