Measuring the $\mathcal{CP}$ nature of the Yukawa coupling between the Higgs boson and tau leptons

Mohammad Hassan Hassanshahi

Imperial College London
Department of Physics

A thesis submitted to Imperial College London
for the degree of Doctor of Philosophy

Abstract

The first direct measurement of the $\mathcal{CP}$ nature of the Yukawa coupling between the Higgs boson and tau leptons is presented. The data used in this analysis were collected in proton-proton collisions recorded by the CMS experiment at the LHC and correspond to a total integrated luminosity of 137 fb^{-1} and a centre-of-mass energy of $\sqrt{s}$=13 TeV. $\mathcal{CP}$-sensitive observables are reconstructed using the angular correlation between the decay planes of tau leptons in the $H \to \tau^+\tau^-$ decay. The observed (expected) value of the effective mixing angle between the $\mathcal{CP}$-even and $\mathcal{CP}$-odd couplings is found to be $-1\pm19^\circ$ ($0\pm21^\circ$) at 68% confidence level. The result is in agreement with the Standard Model predictions within uncertainties and disfavours a pure $\mathcal{CP}$-odd coupling at 3.0 σ confidence level.

To my family

Declaration

I declare that the work in this thesis is my own. It was produced as an extension of the existing work from several individuals and the Compact Muon Solenoid (CMS) collaboration. Where relevant, the work of others has been cited appropriately. Chapter 1-4 are not my work, but are written in my words to provide an overview of the theory and motivation, the CMS experiment, and the techniques used to reconstruct physics objects. Chapter 5 is my own work and describes the model I developed for identifying hadronic tau decay modes. A summary of this work was published as a CMS detector performance summary in Ref. [1]. Chapter 6 is partially my own work and it describes the analysis of the $\mathcal{CP}$ structure of the Yukawa coupling between the Higgs boson and tau leptons. The result of this analysis was published in Ref. [2]. The "MVA decay mode" subsection in section 6.4.2 shows the improvement in the $\mathcal{CP}$ sensitivity brought by the model I developed for identifying hadronic tau decay modes. The simulation corrections described in section 6.6 are my own work unless otherwise stated. The rest of chapter 6 describes various aspects of the analysis along with the result, in my own words. The first 5 sections of chapter 7 describe, in my own words, the high granularity calorimeter (HGCal) detector, the role of the HGCal in the CMS Level-1 trigger, and other preliminary information needed for the rest of the chapter. Sections 7.6 and 7.7 are my own work on developing an optimization algorithm for the Level-1 trigger of the CMS during the high-luminosity LHC (HL-LHC) and studying the effect of this algorithm on the position and energy resolutions of hadronic jets. Chapter 8 are the conclusions and an outlook for this thesis work, in my own words.

Mohammad Hassan Hassanshahi

Acknowledgements

I would like to thank the Imperial College HEP group and the Imperial College President's Scholarship scheme for giving me the opportunity to carry out this research. Thanks to my supervisors, David Colling and Paul Dauncey, for their invaluable support during my PhD and for teaching me how to think as a physicist. David's enthusiasm and energy for physics have deeply inspired me and made working on the $H \to \tau^+\tau^-$ analysis a great experience. I have been truly inspired by Paul's physics intuition, thoroughness, and organization during my work on the HGCal detector. Additionally, Paul removed bureaucratic obstacles from my path on several occasions, allowing me to make the most of my PhD.

I am grateful to Danny and Albert for their guidance during the first months of the $H \to \tau^+\tau^-$ analysis and for answering my questions. I would like to extend my gratitude to Sasha for providing useful advice. I thank Samuel for helping me start the HGCal project and for always being available to help. I also thank Vito, Alex, and Greg for guiding me to study the performance of the HGCal electronic boards, a project which is not included in this thesis.

Thanks to all my ICL friends, especially those at B512, with whom I had a lot of fun chatting and laughing. I would like to thank my friends at home; although I could rarely see them during my PhD, their messages made my days so much brighter. I really enjoyed spending time with my friends in London and I am very grateful to them for that.

I would like to express my deepest gratitude to all members of my family for their support, love, and patience. I owe a great deal to my family and this thesis is therefore dedicated to them.

Contents

1 Introduction **1**

2 Theory and motivation **3**

2.1 The Standard Model of particle physics 3

2.1.1 Fundamental particles 3

2.1.2 The Higgs mechanism 5

2.2 Higgs boson at the Large Hadron Collider 8

2.3 $\mathcal{CP}$ in the Higgs sector 11

2.3.1 Introduction 11

2.3.2 Phenomenology 13

3 The LHC and the CMS detector **21**

3.1 The LHC 21

3.2 The CMS detector 22

3.3 Silicon tracker 25

3.4 Electromagnetic calorimeter 26

3.5 Hadronic calorimeter 27

3.6 Solenoid magnet 29

3.7 Muon system 29

3.8 Triggering 30

3.8.1 Level-1 trigger 31

3.8.2 High-Level Trigger 33

4 Physics object reconstruction **35**

4.1 Introduction 35

4.2 Tracks 35

4.2.1 General method 35

4.2.2 Electron tracking 36

4.2.3 Muon tracking 37

4.3 Vertices . . . 37
4.4 Calorimeter clusters . . . 39
4.5 Particle Flow . . . 39
4.5.1 Muons . . . 41
4.5.2 Electrons and isolated photons . . . 41
4.5.3 Hadrons and non-isolated photons . . . 42
4.6 Jets . . . 42
4.7 Missing transverse momentum . . . 43
4.8 Hadronic taus . . . 45

5 MVA-based hadronic tau decay mode identification 53
5.1 Motivation . . . 53
5.2 Boosted Decision Tree classifiers . . . 55
5.3 MC samples and event selection . . . 56
5.4 Training . . . 56
5.4.1 1-charged-prong decay . . . 56
5.4.2 3-charged-prong decay . . . 61
5.5 Comparing MVA and HPS . . . 64

6 Measuring the $\mathcal{CP}$ properties of the Higgs-tau Yukawa coupling 71
6.1 Introduction . . . 71
6.2 Simulated samples . . . 71
6.3 Physics object and event selection . . . 72
6.4 $\phi_{\mathcal{CP}}$: strategy and optimization . . . 74
6.4.1 Strategy . . . 74
6.4.2 Optimization . . . 75
6.5 Background estimation . . . 76
6.5.1 Fake factor method . . . 76
6.5.2 τ-embedding method . . . 77
6.5.3 Other backgrounds . . . 77
6.6 Simulation correction on efficiencies . . . 78
6.6.1 Electron and muon efficiency correction . . . 78
6.6.2 Hadronic taus efficiency correction . . . 82
6.7 Event categorization . . . 91
6.8 $\phi_{\mathcal{CP}}$ distribution in MVA score bins . . . 91
6.9 Systematic uncertainties . . . 93
6.9.1 Normalization uncertainties . . . 93
6.9.2 Shape uncertainties . . . 95

6.10 Results . . . 95
6.10.1 Measured value of $\alpha^{H\tau\tau}$ . . . 96

7 The high granularity calorimeter 103
7.1 The high-luminosity LHC . . . 104
7.2 Phase-2 CMS upgrade . . . 104
7.3 The high granularity calorimeter . . . 106
7.3.1 Essential features . . . 106
7.3.2 Design . . . 108
7.4 Phase-2 L1T upgrade . . . 110
7.5 The HGCal in the calorimetry triggering path . . . 112
7.5.1 The HGCal trigger primitive generator . . . 112
7.5.2 The two-stage system of the HGCal TPG . . . 113
7.6 Assigning module sums to trigger towers . . . 114
7.6.1 Introduction . . . 114
7.6.2 Silicon modules . . . 116
7.6.3 Scintillator modules . . . 117
7.7 Jet reconstruction performance . . . 121

8 Conclusions and outlook 125

A MVA decay mode feature-importance 127
A.1 1-charged-prong decay . . . 127
A.2 3-charged-prong decay . . . 128

B Electron and muon scale factors 133

C Hadronic tau scale factors 141

List of Figures

2.1 The Standard Model particles. 4
2.2 Measured coupling strengths of the Higgs boson to fermions and to vector bosons. 9
2.3 Feynman diagrams of the dominant Higgs boson production processes. . 10
2.4 Illustration of $\phi_{\mathcal{CP}}$ angle. 14
2.5 $\phi_{\mathcal{CP}}$ distribution for different $\mathcal{CP}$ scenarios. 15
2.6 Illustration of decay plane methods. 17
2.7 Feynman diagram of the $a_1^- \to 2\pi^-\pi^+$ decay. 18

3.1 Schematic of the LHC complex. 22
3.2 Cross sections of different physics processes in proton-(anti)proton collisions as a function of the centre-of-mass energy. 23
3.3 Perspective view of the CMS detector. 24
3.4 Cross sectional view of the tracker system. 26
3.5 Schematic view of the ECAL. 27
3.6 Cross sectional view of the HCAL. 28
3.7 Cross sectional view of the muon system. 30
3.8 Dataflow in the Level-1 trigger. 32

4.1 PUPPI-MET alpha distributions 45
4.2 Schematic of 1-charged-prong τ_h decays. 47
4.3 Schematic of 3-charged-prong τ_h decays. 47
4.4 Distance between τ_h and the γ/e from its decay. 49
4.5 Architecture of DeepTau. 51
4.6 DeepTau: Grids for extracting low-level information. 51

5.1 HPS decay mode: confusion matrices 55
5.2 MVA decay mode: MVA score distributions for 1-prong decays 60
5.3 MVA decay mode: ROC curves for 1-prong decays 61

5.4 MVA decay mode: MVA score distributions for 3-prong decays 63
5.5 MVA decay mode: ROC curves for 3-prong decays 63
5.6 MVA decay mode: purity and efficiency 64
5.7 MVA decay mode: confusion matrices (purity) 65
5.8 MVA decay mode: confusion matrices (efficiency) 66
5.9 MVA decay mode: Data/MC agreement 68

6.1 Improvement of $\mathcal{CP}$ sensitivity by using MVA decay mode. 75
6.2 Examples of muon efficiency measurement fits. 81
6.3 Muon scale factors as a function of p_{T} in the central detector region. . . 83
6.4 Electron scale factors as a function of p_{T} in the central detector region. . 84
6.5 Examples of τ_h identification scale factors. 87
6.6 Examples of τ_h identification postfit plots. 88
6.7 Examples of τ_h trigger efficiencies and scale factors. 90
6.8 MVA score distribution of Genuine and Mis-ID backgrounds. 92
6.9 $\phi_{\mathcal{CP}}$ distribution in bins of MVA score. 94
6.10 Negative log-likelihood scan of $\alpha^{H\tau\tau}$. 97
6.11 Negative 2D log-likelihood scan of the κ_τ and $\tilde{\kappa}_\tau$ couplings. 98
6.12 Negative 2D log-likelihood scan of $\alpha^{H\tau\tau}$ and μ. 99
6.13 Weighted $\phi_{\mathcal{CP}}$ distributions of the background-subtracted data and the predictions of the $\mathcal{CP}$-even ($\alpha^{H\tau\tau}$=0°) and $\mathcal{CP}$-odd ($\alpha^{H\tau\tau}$=90°) scenarios. 100

7.1 CMS integrated luminosity during 2010-18. 105
7.2 LHC and HL-LHC plan. 105
7.3 Absorbed dose of the HGCal after 3000 fb^{-1}. 107
7.4 A cross sectional view of the HGCal. 109
7.5 Trigger cells in silicon modules. 110
7.6 Dataflow of the Phase-2 CMS Level-1 trigger. 111
7.7 The two-stage system of the HGCal TPG. 114
7.8 Number of towers overlapped with modules. 115
7.9 Illustration of module sum splitting in silicon modules. 117
7.10 An example of splitting a module sum over the overlapping towers. . . . 118
7.11 Illustration of the position of scintillator modules with respect to towers. 119
7.12 The η range of u0 and u1 scintillator modules and the module sum share each overlapping tower receives. 119
7.13 A view of the position of two towers in a layer. 120
7.14 Comparing the E_{T} distribution between the splitting and non-splitting cases for an event. 122

7.15 Comparing the transverse energy resolution of the leading jet between the splitting and non-splitting cases 123
7.16 Comparing the splitting and non-splitting cases based on the η and ϕ resolutions of the jets. 123

A.1 Appendix: MVA decay mode 1-prong feature-importance list 129
A.2 Appendix: MVA decay mode 3-prong feature-importance list 132

B.1 Electron identification scale factor as a function of p_{T} for different regions of the detector. 134
B.2 Electron isolation scale factor as a function of p_{T} for different regions of the detector. 135
B.3 Single-electron trigger scale factor as a function of p_{T} for different regions of the detector. 136
B.4 Muon identification scale factor as a function of p_{T} for different regions of the detector. 137
B.5 Muon isolation scale factor as a function of p_{T} for different regions of the detector. 138
B.6 Single-muon trigger scale factor as a function of p_{T} for different regions of the detector. 139
B.7 Muon+tau cross trigger scale factor as a function of p_{T} for different regions of the detector. 140

C.1 τ_h identification scale factor for the embedded sample, valid for the $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels. 141
C.2 τ_h identification scale factor for the MC sample, valid for the $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels. 142
C.3 τ_h identification scale factor for the embedded sample, valid for the $\tau_e\tau_h$ channel. 142
C.4 τ_h identification scale factor for the MC sample, valid for the $\tau_e\tau_h$ channel. 142
C.5 τ_h identification postfit plots for the embedded samples of 2016 data-taking period. The τ_h candidates are required to have $p_{\mathrm{T}}^{\tau_h} > 40$ GeV. . . 143
C.6 τ_h identification postfit plots for the MC samples of 2016 data-taking period. The τ_h candidates are required to have $p_{\mathrm{T}}^{\tau_h} > 40$ GeV. 144
C.7 τ_h identification postfit plots for the embedded samples of 2016 data-taking period. The τ_h candidates are required to have $20 < p_{\mathrm{T}}^{\tau_h} < 40$ GeV. 145

C.8 τ_h identification postfit plots for the MC samples of 2016 data-taking period. The τ_h candidates are required to have $20 < p_{\mathrm{T}}^{\tau_h} < 40$ GeV. . . . 146
C.9 τ_h identification postfit plots for the embedded samples of 2017 data-taking period. The τ_h candidates are required to have $p_{\mathrm{T}}^{\tau_h} > 40$ GeV. . . 147
C.10 τ_h identification postfit plots for the MC samples of 2017 data-taking period. The τ_h candidates are required to have $p_{\mathrm{T}}^{\tau_h} > 40$ GeV. 148
C.11 τ_h identification postfit plots for the embedded samples of 2017 data-taking period. The τ_h candidates are required to have $20 < p_{\mathrm{T}}^{\tau_h} < 40$ GeV. 149
C.12 τ_h identification postfit plots for the MC samples of 2017 data-taking period. The τ_h candidates are required to have $20 < p_{\mathrm{T}}^{\tau_h} < 40$ GeV. . . . 150
C.13 τ_h identification postfit plots for the embedded samples of 2018 data-taking period. The τ_h candidates are required to have $p_{\mathrm{T}}^{\tau_h} > 40$ GeV. . . 151
C.14 τ_h identification postfit plots for the MC samples of 2018 data-taking period. The τ_h candidates are required to have $p_{\mathrm{T}}^{\tau_h} > 40$ GeV. 152
C.15 τ_h identification postfit plots for the embedded samples of 2018 data-taking period. The τ_h candidates are required to have $20 < p_{\mathrm{T}}^{\tau_h} < 40$ GeV. 153
C.16 τ_h identification postfit plots for the MC samples of 2018 data-taking period. The τ_h candidates are required to have $20 < p_{\mathrm{T}}^{\tau_h} < 40$ GeV. . . . 154
C.17 τ_h trigger efficiencies and scale factors for $\tau_\mu\tau_h$ channel using data and embedded samples based on the 2016 experimental conditions. 155
C.18 τ_h trigger efficiencies and scale factors for $\tau_\mu\tau_h$ channel using data and MC samples based on the 2016 experimental conditions. 156
C.19 τ_h trigger efficiencies and scale factors for $\tau_h\tau_h$ channel using data and embedded samples based on the 2016 experimental conditions. 157
C.20 τ_h trigger efficiencies and scale factors for $\tau_h\tau_h$ channel using data and MC samples based on the 2016 experimental conditions. 158
C.21 τ_h trigger efficiencies and scale factors for $\tau_e\tau_h$ channel using data and embedded samples based on the 2017 experimental conditions. 159
C.22 τ_h trigger efficiencies and scale factors for $\tau_e\tau_h$ channel using data and MC samples based on the 2017 experimental conditions. 160
C.23 τ_h trigger efficiencies and scale factors for $\tau_\mu\tau_h$ channel using data and embedded samples based on the 2017 experimental conditions. 161
C.24 τ_h trigger efficiencies and scale factors for $\tau_\mu\tau_h$ channel using data and MC samples based on the 2017 experimental conditions. 162

C.25 τ_h trigger efficiencies and scale factors for $\tau_h\tau_h$ channel using data and embedded samples based on the 2017 experimental conditions. 163
C.26 τ_h trigger efficiencies and scale factors for $\tau_h\tau_h$ channel using data and MC samples based on the 2017 experimental conditions. 164
C.27 τ_h trigger efficiencies and scale factors for $\tau_e\tau_h$ channel using data and embedded samples based on the 2018 experimental conditions. 165
C.28 τ_h trigger efficiencies and scale factors for $\tau_e\tau_h$ channel using data and MC samples based on the 2018 experimental conditions. 166
C.29 τ_h trigger efficiencies and scale factors for $\tau_\mu\tau_h$ channel using data and embedded samples based on the 2018 experimental conditions. 167
C.30 τ_h trigger efficiencies and scale factors for $\tau_\mu\tau_h$ channel using data and MC samples based on the 2018 experimental conditions. 168
C.31 τ_h trigger efficiencies and scale factors for $\tau_h\tau_h$ channel using data and embedded samples based on the 2018 experimental conditions. 169
C.32 τ_h trigger efficiencies and scale factors for $\tau_h\tau_h$ channel using data and MC samples based on the 2018 experimental conditions. 170

List of Tables

2.1 Cross section of the dominant Higgs boson production processes. 10
2.2 Tau decay modes. 13

4.1 The branching ratios of tau decay modes. 46
4.2 Mass conditions on tau decay modes. 50

6.1 Online triggers and offline p_{T} requirements. 73
6.2 Methods for modeling backgrounds in the $\tau_h\tau_h$ channel. 78
6.3 Methods for modeling backgrounds in the $\tau_l\tau_h$ channel. 78

Chapter 1

Introduction

The Standard Model (SM) [3] of particle physics is a theory which describes fundamental particles and their interactions. The theory explains the strong interaction and also beautifully combines the electromagnetic and weak interactions into the electroweak (EW) theory. The EW theory exhibits a certain symmetry which is broken by the Higgs mechanism. The Higgs mechanism not only describes the symmetry breaking mechanism, but also explains how the SM particles obtain mass. A natural consequence of the broken symmetry is the existence of a massive scalar particle, widely known as the Higgs boson.

The SM theory has been verified with increasingly impressive precision. However, this theory cannot explain certain observations, such as the existence of dark matter and the origin of neutrino mass. In addition, the SM is not capable of explaining the level of the matter-antimatter asymmetry in the Universe. There are theories beyond the Standard Model (BSM) which can provide a description for all or some of the SM shortcomings.

One way to probe BSM physics is to measure the properties of the SM to a high precision and compare them with the predicted value. In particular, the charge conjugate-parity ($\mathcal{CP}$) is an important property of the Higgs boson and is predicted by the SM, so a measurement can shed light on BSM physics. The SM predicts the $\mathcal{CP}$ of the Higgs boson to be *even*; therefore, any deviation from a pure $\mathcal{CP}$-even scenario is a clear indication of BSM physics.

This thesis describes the measurement of the $\mathcal{CP}$ structure of the Yukawa coupling between the Higgs boson and tau leptons using the decay of the Higgs boson to two tau leptons. The data used in this measurement was recorded by the CMS experiment at the CERN's Large Hadron Collider (LHC) during 2016 to 2018. The results of this measurement were published in Ref. [2]. The structure of this thesis is as follows.

Chapter 2 begins by describing the SM theory of particle physics and in particular

the Higgs mechanism. This is followed by explaining the production processes of the Higgs boson at the LHC and presenting the latest measurements of the properties of this particle. In the last section of this chapter, the Higgs couplings for which $\mathcal{CP}$ structure can be measured at the LHC are compared and the methods used at the LHC for reconstructing observables sensitive to the $\mathcal{CP}$ structure of the Higgs-tau coupling are illustrated.

Chapter 3 describes the LHC and the CMS detector. The CMS detector is comprised of several subdetectors designed for identifying different particles along with measuring their properties and kinematic quantities. At the end of this chapter, the CMS triggering system is explained.

Chapter 4 describes the methods employed in CMS to identify and reconstruct physics objects (such as electrons and taus) using particle tracks and energy deposits as ingredients. The methods are optimized such that objects are identified with a high efficiency and a low misidentification rate.

Chapter 5 describes the multivariate analysis (MVA) I developed for identifying different hadronic decay modes of taus. In this analysis, tau decay modes differ in their $\mathcal{CP}$ sensitivity, and it is therefore essential to identify them to a high degree of accuracy. A summary of this chapter is published as a CMS detector performance summary (DPS) in Ref. [1].

Chapter 6, which is the core of the thesis, describes the analysis of the $\mathcal{CP}$ structure of the Yukawa coupling between the Higgs boson and tau leptons. Different steps of the analysis are described in detail and the result of this measurement is presented at the end of the chapter. My contributions to this analysis are elaborated in more detail. The analysis was published in Ref. [2].

Chapter 7 firstly describes the upgrade plan of the CMS with a particular focus on the high granularity calorimeter (HGCal) detector and its role in the L1 trigger upgrade. Then, a description of the optimization algorithm I developed to determine how data from individual HGCal channels should be processed is presented.

Finally, chapter 8 discusses the conclusions of the analysis and the HGCal work. In addition, a perspective on future work is provided.

Chapter 2

Theory and motivation

This chapter describes the theoretical framework behind and the motivation for the Higgs $\mathcal{CP}$ measurement along with the phenomenology of this measurement at the Large Hadron Collider (LHC). The established theory of particle physics, the Standard Model (SM), is briefly explained in section 2.1, the Higgs boson production channels at the LHC along with its decay channels are described in section 2.2, and finally in section 2.3, the phenomenology of measuring the Higgs $\mathcal{CP}$ is illustrated.

2.1 The Standard Model of particle physics

2.1.1 Fundamental particles

The SM is a renormalizable quantum field theory which describes three fundamental forces in our Universe: the electromagnetic (EM), weak, and strong forces. This model represents matter as spin-1/2 fermions and forces as spin-1 vector bosons. There is only one spin-0 scalar boson in the SM, the Higgs boson, which plays a crucial and unique role in this model and is responsible for the mass of massive fermions and vector bosons through the spontaneous symmetry breaking mechanism described later in this chapter. The SM Lagrangian is invariant under certain local gauge transformations, which is an essential property of this model. The symmetry group corresponding to the gauge invariance is

$$G_{\mathrm{SM}} = SU(3)_{\mathrm{C}} \times SU(2)_{\mathrm{L}} \times U(1)_{\mathrm{Y}} \tag{2.1}$$

which will be discussed shortly.

The fundamental particles of the SM are shown in Fig. 2.1. For each fermion, there is an antiparticle with opposite quantum numbers. Fermions can be grouped into three generations with similar quantum numbers but different masses. Fermions are made

up of twelve *flavours*, six of which interact with the strong interaction, called quarks, and six of which do not, called leptons. Quarks comes in three *colour* charges, where a colour is a quantum number for particles interacting with the strong force.

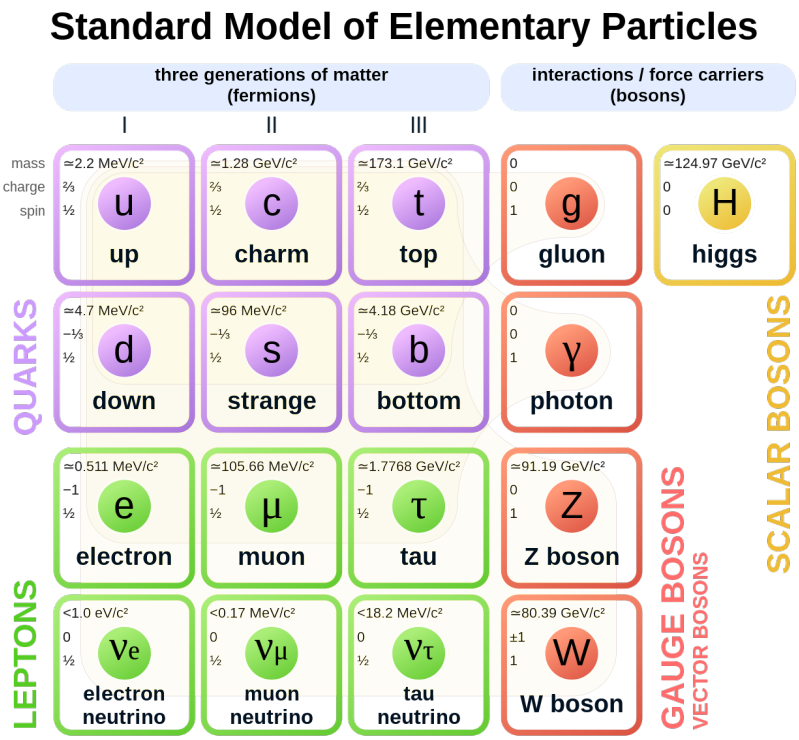

Figure 2.1: The fundamental particles of the Standard Model [4].

The strong force is described by the theory of *quantum chromodynamics* (QCD). The gauge symmetry of this force is $SU(3)$ (see Eq. 2.1), which governs the interaction of quarks and gluons. There are a total of eight gluons carrying the strong force, which correspond to the eight generators of the $SU(3)$ group. Gluons are massless and, since they have colour charge, they can self-interact; accordingly, the strong force possesses two unique properties:

- **Asymptotic freedom:** The coupling strength of the strong interaction decreases with energy.
- **Confinement:** The binding energy between quarks increases linearly with distance. Therefore, no free quarks are observed in nature as they quickly join quark-antiquark pairs, produced from vacuum, to form colourless hadrons.

The electromagnetic and weak forces were combined into a theory called *electroweak* (EW) in 1960s by the work of Glashow [5], Weinberg [6], and Salam [7]. The theory

possesses a $SU(2)_{\rm L}\times U(1)_{\rm Y}$ gauge symmetry, where "L" shows that the $SU(2)$ symmetry applies to the left-chiral fermions and "Y" is a symbol for the hypercharge symmetry. The generators of the $SU(2)$ symmetry are associated to the weak isospin gauge bosons, shown with W^i (i=1,2,3), while the generator of the $U(1)$ symmetry corresponds to the hypercharge boson, shown with B. The electric charge of fermions are related to the hypercharge (Y) and the third component of the weak isospin ($I_w^{(3)}$) as

$$Q = I_w^{(3)} + \frac{Y}{2} \tag{2.2}$$

The weak isospin and hypercharge bosons, with coupling strengths to fermions respectively shown by g and g', are mixed to form the physical gauge bosons of the weak and electromagnetic interactions

$$\begin{aligned} Z_\mu &= \cos\theta_w W_\mu^3 - \sin\theta_w B_\mu \\ A_\mu &= \sin\theta_w W_\mu^3 + \cos\theta_w B_\mu \\ W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \mp i W_\mu^2) \end{aligned} \tag{2.3}$$

where μ is the four-vector index, $Z/W^\pm$ are the weak interaction gauge bosons, A is the EM gauge boson (photon), and $\theta_w = \arctan(g'/g)$ is called the *weak mixing angle*. The EW theory was confirmed by the discovery of the W [8, 9] and Z [10, 11] bosons in 1983.

The EW theory beautifully unifies the EM and weak interactions. However, it cannot provide a mass term in the Lagrangian for W or Z bosons, which are known to be massive. Adding a $m_W^2 W_\mu^+ W_\mu^-$ or $\frac{1}{2}m_Z^2 Z_\mu Z^\mu$ term by hand explicitly breaks the EW gauge symmetry and makes it non-renormalizable. Likewise, one cannot provide a mass term for fermions since a term of the form $m\bar{\psi}\psi$ is not gauge invariant, where ψ is the spinor of the fermion. As we will see, the issue is solved by introducing the Higgs mechanism.

2.1.2 The Higgs mechanism

In order for the gauge bosons of the weak interaction to acquire mass while keeping the Lagrangian gauge invariant, the EW symmetry must be broken through spontaneous (rather than explicit) symmetry breaking. In this type of symmetry breaking, the vacuum state of the system does not respect the Lagrangian symmetry.

The idea of spontaneous symmetry breaking had been used in condensed matter physics before getting into particle physics. When Salam and Weinberg tried to use this idea, they ran into an obstacle, which is known as *Goldstone theorem*: spontaneous

symmetry breaking leads to a massless spin-0 particle, called the *Goldstone boson*. The problem was that such particles had not been observed although, if they had a reasonable interaction strength, they should have been easily seen [12].

In 1964, however, three research groups, firstly Brout and Englert [13], then Higgs [14, 15], and finally Guralnik, Hagen and Kibble [16], independently discovered that Goldstone bosons can be avoided in a gauge theory and that they manifest themselves as the longitudinal polarization of the gauge bosons when they acquire mass from spontaneous symmetry breaking. In the case of the EW theory, spontaneous symmetry breaking is achieved through the Higgs mechanism, in which a new spin-0 particle is introduced: the Higgs boson. The minimal way to introduce such a particle which can spontaneously break the symmetry while retaining the gauge invariance of the Lagrangian consists of a $SU(2)_L$ doublet of the form:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{2.4}$$

where both elements of the doublet are complex scalar fields. The Lagrangian describing the ϕ field is

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger (D^\mu \phi) - \underbrace{\left(\mu^2 \phi\phi^\dagger + \lambda(\phi\phi^\dagger)^2\right)}_{V(\phi)}, \tag{2.5}$$

where $V(\phi)$ is the Higgs potential and D_μ is the covariant derivative defined with the help of the Pauli matrices, σ^i, to ensure that the Lagrangian is gauge invariant,

$$D_\mu = \partial_\mu + i\frac{g}{2} W_\mu^i \sigma^i + ig'\frac{Y}{2} B_\mu. \tag{2.6}$$

Assuming a negative value for μ^2 in the Higgs potential, the absolute minimum of the potential becomes

$$\phi^\dagger \phi = -\frac{\mu^2}{2\lambda} \tag{2.7}$$

indicating that there are infinitely many ground state solutions. Since at the end of the calculation, we want to end up with a massless **neutral** vector boson (i.e. photon), we take the vacuum state to be one of the components of ϕ^0 (rather than ϕ^+),

$$\langle 0|\phi|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \sqrt{\frac{-\mu^2}{\lambda}} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \tag{2.8}$$

where $v \sim 246$ GeV is the vacuum expectation value (VEV). The Higgs field, H, is

then defined by expanding ϕ about its vacuum

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \ + \ H \end{pmatrix}. \tag{2.9}$$

Substituting Eq. 2.9 in Eq. 2.5, keeping the terms up to the second order in the fields, and using physical vector bosons introduced in Eq. 2.3 gives

$$\mathcal{L}_{\text{Higgs}} = \frac{1}{2}\partial_\mu H \partial^\mu H + \frac{g^2 v^2}{4} W_\mu^- W^{+\mu} + \frac{\left(g^2 + g'^2\right) v^2}{8} Z_\mu Z^\mu + \lambda v^2 H^2 + O(3) \tag{2.10}$$

which leads to m_A=0, m_W=$\frac{1}{2}vg$, m_Z=$\frac{1}{2}v\sqrt{g^2+g'^2}$, and m_H=$\sqrt{2\lambda}v$. The higher-order terms include the Higgs boson self-interaction as well as its interaction with the vector bosons.

The Higgs mechanism can also provide gauge invariant mass terms for fermions. Defining $\tilde{\phi} \equiv i\sigma_2\phi^*$, the gauge invariant interaction between the ϕ and fermions can be written as

$$\mathcal{L}_{\text{Yukawa}} = -Y^e_{ij}(\overline{L_{Li}} \cdot \phi)E_{Rj} - Y^d_{ij}(\overline{Q_{Li}} \cdot \phi)D_{Rj} - Y^u_{ij}(\overline{Q_{Li}} \cdot \tilde{\phi})U_{Rj} + \text{h.c.}, \tag{2.11}$$

where Y_{ij} are the Yukawa matrices which mix different fermion generations while the spinors are defined as

$$\begin{aligned} \overline{L_{L1}} &= \begin{pmatrix} \overline{\nu_e}_L \\ \overline{e}_L \end{pmatrix}, \\ \overline{Q_{L1}} &= \begin{pmatrix} \overline{u}_L \\ \overline{d}_L \end{pmatrix} \end{aligned} \tag{2.12}$$

and

$$\begin{aligned} E_{R1} &= e_R, \\ D_{R1} &= d_R, \\ U_{R1} &= u_R \end{aligned} \tag{2.13}$$

for the first generation and similarly for the rest. Since neutrinos are massless in the SM, the Yukawa matrix for mixing lepton generations, Y^e_{ij}, can be simultaneously diagonalized in the mass and interaction (with W bosons) eigenstates. The matrix therefore has 3 physical degrees of freedom which are the charged lepton masses. However, the mass and interaction bases can be (and are) different for the quark Yukawa matrices, Y^d_{ij} and Y^u_{ij}. These two matrices have 10 degrees of freedom in total, of which 6 correspond to the quark masses while 4 are encapsulated in a matrix known as Cabibbo–Kobayashi–Maskawa (CKM) which transforms the mass basis to the interac-

tion basis. From the 4 degrees of freedom in the CKM matrix, 3 are the matrix mixing angles and 1 is the $\mathcal{CP}$-violating phase, which is the only source of $\mathcal{CP}$ violation within the SM theory.

After the Higgs boson acquires a non-zero VEV, the Yukawa interaction (Eq. 2.11) for the particular case of tau leptons can be written as

$$\mathcal{L}^{\tau}_{\text{Yukawa}} = -m_{\tau}\overline{\tau}\tau - \frac{m_{\tau}}{v}H\overline{\tau}\tau \tag{2.14}$$

where the first term of the RHS is the mass term while the second term shows the interaction of the Higgs boson and tau leptons. The interaction $H\overline{\tau}\tau$ is *even* under $\mathcal{CP}$ transformation whereas the $H\overline{\tau}i\gamma_5\tau$ term, which is allowed in some BSM theories, is *odd* under $\mathcal{CP}$ transformation. We will come back to the $\mathcal{CP}$ of this interaction later in this chapter.

2.2 Higgs boson at the Large Hadron Collider

In 2012, the ATLAS and CMS collaborations jointly announced [17, 18] the discovery of a new particle with mass near 125 GeV. Since then, there has been a great deal of effort put into measuring the properties of this particle and all have shown compatibility with the SM Higgs boson properties within the uncertainties. The observed particle is found to be spin-0 and incompatible with a pure $\mathcal{CP}$-odd Higgs boson [19, 20]. The observed particle, therefore, can couple to other particles either in a pure $\mathcal{CP}$-even state, as predicted by the SM, or in a mixed state of $\mathcal{CP}$-even and $\mathcal{CP}$-odd, which would be a clear indication of physics beyond the SM.

The observed decay channels of this particle are $\gamma\gamma$ [21, 22], ZZ [23, 24], WW [25, 26], $\tau\bar{\tau}$ [27, 28], and $b\bar{b}$ [29, 30]. Additionally, a direct measurement of the coupling to the top quark has been performed in the Higgs production associated with $t\bar{t}$ pairs [31, 32]. Further, the CMS Collaboration has recently announced the observation of evidence for the decay of the Higgs boson to $\mu\bar{\mu}$ [33], which is the first evidence for the Higgs boson coupling to the second generation of fermions. The measured coupling strengths of the Higgs boson to fermions and vector bosons are shown in Fig. 2.2. All coupling strengths have shown a good compatibility with the SM predictions.

At the LHC, the Higgs boson is produced via four main processes (see Fig. 2.3): gluon-gluon fusion (ggH), vector boson fusion (VBF), W/Z-associated production (VH), and $t\bar{t}$-associated production (ttH). Table 2.1 shows the cross section of these processes. The dominant production cross section belongs to the ggH process although it occurs through a loop of fermions, mainly top quark. The VBF process with approximately one order of magnitude smaller cross section is the sub-dominant process. However,

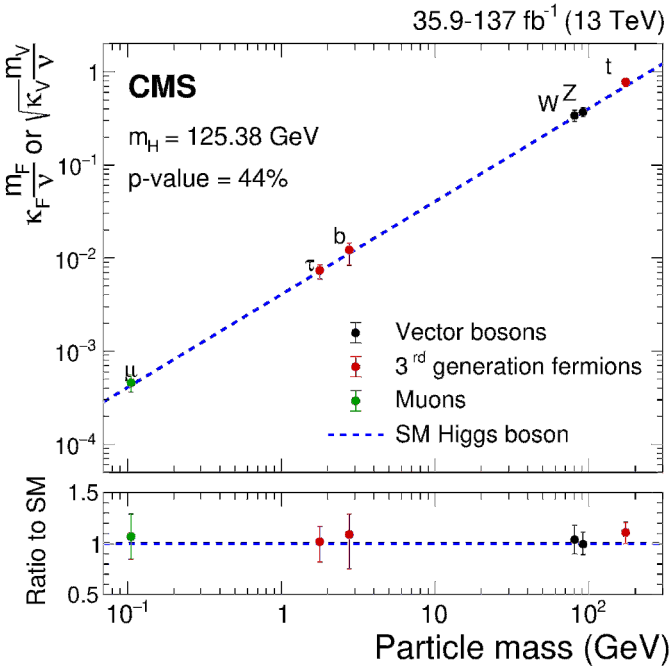

Figure 2.2: The reduced Higgs boson coupling to fermions and vector bosons as a function of their mass, measured by the CMS experiment. The bottom panel displays the ratio of the measurement to the SM perdictions. The error bars represent 68% confidence intervals [33].

Production mode	ggH	VBF	WH	ZH	ttH
Cross section (pb)	48.71	3.78	1.37	0.88	0.51

Table 2.1: Cross section of the dominant Higgs boson production processes in the SM derived for $m_H = 125$ GeV and proton-proton interactions at the centre-of-mass energy of $\sqrt{13}$ TeV [34].

the VBF process is important from the experimental point of view for the distinctive topology of its associated jets: large spatial separation and large invariant mass. Due to these distinctive features, the Higgs boson produced in this process is easier to distinguish from the background than in other processes.

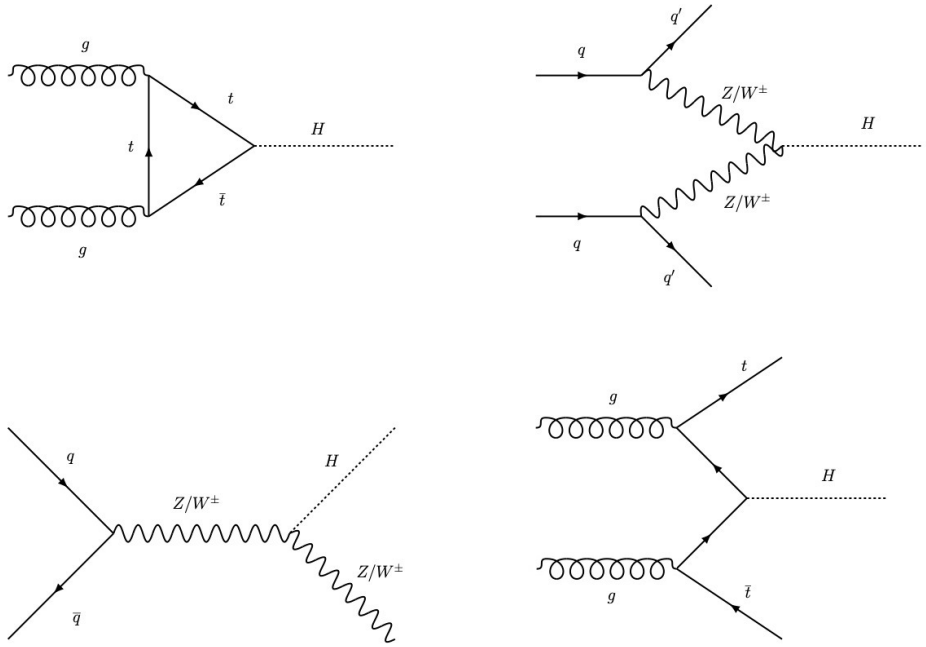

Figure 2.3: Feynman diagrams showing the dominant Higgs boson production processes. The processes are known as gluon-gluon fusion (top left), vector boson fusion (top right), W/Z-associated production (bottom left), and $t\bar{t}$-associated production (bottom right).

2.3 $\mathcal{CP}$ in the Higgs sector

2.3.1 Introduction

The SM has been verified with increasingly impressive accuracy. However, there are concrete observations which cannot be explained by this theory such as the existence of dark matter, dark energy, and the mass of neutrinos. In addition, there are some fundamental questions which are not answered in the SM, such as the naturalness problem

and the reason for the observed amount of the matter-antimatter asymmetry in the Universe. Since the discovery of the Higgs boson, some of the particles predicted by BSM theories have been searched for, because the existence of such particles, e.g. those from supersymmetric (SUSY) models, could explain the observations and be an answer to the fundamental questions. Nonetheless, no new fundamental particles have been discovered, nor any deviations from the SM have been seen. This fact led to an increasing interest among particle physicists to perform indirect BSM searches, in particular by precisely measuring the SM parameters and look for a possible deviation. The growing amount of data coming from the LHC makes this strategy even more sensible.

Among the SM parameters, the Higgs boson properties are of great interest because of its unique role in the SM. A $\mathcal{CP}$-odd component for the Higgs boson would lead to $\mathcal{CP}$ violation in the Higgs sector which could then explain the matter-antimatter asymmetry in the Universe and would be a proof of the existence of new physics. Such a component is predicted by some BSM models such as SUSY and multi-Higgs-doublet models.

At the LHC, $\mathcal{CP}$ studies in the Higgs sector consist of two types of measurements: measuring the $\mathcal{CP}$ nature of the coupling of the Higgs boson to (i) vector bosons and to (ii) fermions. The couplings of which $\mathcal{CP}$ nature can be studied with the current available data of the LHC are the Higgs-to-gluon effective coupling (Hgg) and the Higgs-to-W/Z couplings (HVV) from type (i), and the Higgs-to-top (Htt) and Higgs-to-tau ($H\tau\tau$) couplings from type (ii). In the following list, the couplings are introduced and compared to the $H\tau\tau$, which is the one measured in the analysis discussed in this thesis. The latest result for the measurement of these couplings are also reported.

- **Hgg:** This coupling can be measured in the ggH production mode of the Higgs boson. Gluons are massless and hence couple to the Higgs boson through loops of fermions (and possibly BSM particles) which makes this measurement model-dependent, as opposed to the $H\tau\tau$ which is measured model-independently at the LHC. Recent measurements of the $\mathcal{CP}$ of this coupling have shown compatibility with the SM predictions [35–38]

- **HVV:** The Higgs-to-W/Z couplings can be measured on the decay side in the $H \to VV$ decay, or on the production side in the VBF or VH production modes. Contrary to the $H\tau\tau$ coupling, no renormalizable $\mathcal{CP}$-odd term exists for the HVV interaction and hence such contribution would be suppressed by powers of $(1/\Lambda)^2$, in which Λ is the energy scale of new physics. The $\mathcal{CP}$ measurements on this coupling have been consistent with the SM [35, 36, 38–41].

- **Htt:** This coupling is measured mainly through the ttH production mode. It is complementary to the $H\tau\tau$ coupling measurement in the sense that in some BSM models (e.g. complex two-Higgs-doublet model, or C2HDM for short [42]) constraining the $\mathcal{CP}$ nature of one of the couplings does not constrain that of the other. The ATLAS and CMS Collaborations have recently measured this coupling in the $H \to \gamma\gamma$ final state and excluded a pure $\mathcal{CP}$-odd component by more than 3 standard deviations [43, 44]. Similarly, the result from the four-lepton final state $H \to 4\ell$ agrees with the SM predictions within the uncertainties [35, 36].

- ***Hττ*:** This coupling is measured in the $H \to \tau^+\tau^-$ decay. The only direct measurement of this coupling has been performed [2] by the CMS Collaboration and is discussed in this thesis.

The SM Lagrangian describing the interaction of the Higgs boson and tau leptons can be generalized to include a $\mathcal{CP}$-odd term (c.f. second term in the RHS of Eq. 2.14)

$$\mathcal{L}^{H\tau\tau}_{\text{Yukawa}} = -\frac{m_\tau}{v} H(\kappa_\tau \overline{\tau}\tau + \tilde{\kappa}_\tau \overline{\tau} i\gamma_5 \tau) \tag{2.15}$$

where κ_τ and $\tilde{\kappa}_\tau$ are coupling strength modifiers for the $\mathcal{CP}$-even and $\mathcal{CP}$-odd components, respectively. An effective mixing angle $\alpha^{H\tau\tau}$ can then be defined in terms of these coupling strength modifiers as

$$\tan(\alpha^{H\tau\tau}) = \frac{\tilde{\kappa}_\tau}{\kappa_\tau}. \tag{2.16}$$

A pure $\mathcal{CP}$-even ($\mathcal{CP}$-odd) coupling corresponds to a mixing angle of $\alpha^{H\tau\tau} = 0(90)^\circ$ while any mixing angle with $0^\circ < |\alpha^{H\tau\tau}| < 90^\circ$ is a $\mathcal{CP}$-violating scenario in which the Higgs boson couples to taus with a combination of the $\mathcal{CP}$-even and $\mathcal{CP}$-odd components. The maximum mixing of the components occur in $\alpha^{H\tau\tau} = 45^\circ$.

A non-zero value of $\alpha^{H\tau\tau}$ would be a clear evidence of new physics and has implications for BSM theories. For example, a sizable mixing angle would disfavour the minimal supersymmetric model (MSSM), while in the next-to-minimal supersymmetric model (NMSSM), a mixing angle of up to 27° is allowed [45].

2.3.2 Phenomenology

The $\phi_{\mathcal{CP}}$ angle

In this section, the methods for measuring the mixing angle $\alpha^{H\tau\tau}$ are described. For the sake of brevity, tau decay modes are shown with the following symbols: $\tau^\pm \to e^\pm$ with e, $\tau^\pm \to \mu^\pm$ with μ, $\tau^\pm \to \pi^\pm$ with π, $\tau^\pm \to \pi^\pm\pi^0$ with ρ, $\tau^\pm \to \pi^\pm 2\pi^0$ with $a_1^{1\text{pr}}$,

Decay mode	Dominant resonance	BR(%)	Symbol
$e^-\bar{\nu}_e\nu_\tau$		17.8	e
$\mu^-\bar{\nu}_\mu\nu_\tau$		17.4	μ
$h^\pm\nu_\tau$		11.5	π
$h^\pm\pi^0\nu_\tau$	$\rho^\pm$ (770)	25.9	ρ
$h^\pm 2\pi^0\nu_\tau$	$a_1^\pm(1260)$	9.5	$a_1^{1\text{pr}}$
$3h^\pm\nu_\tau$	$a_1^\pm(1260)$	9.8	$a_1^{3\text{pr}}$
$3h^\pm\pi^0\nu_\tau$	excited $\rho^\pm$	4.8	-
Other		3.3	-

Table 2.2: The intermediate resonances and branching ratios of the dominant tau decays [46]. The $h^\pm$ mesons are predominantly $\pi^\pm$. The "symbol" column shows how the decay modes are referred to throughout this thesis. The "pr" index in $a_1^{1\text{pr}}$ and $a_1^{3\text{pr}}$ indicates the number of charged particles (**pr**ongs) in the final state.

and $\tau^\pm \to 3\pi^\pm$ with $a_1^{3\text{pr}}$, where neutrinos are not shown in the decay products for simplicity as they are not detected in the CMS detector. These symbols represent one of the final states or the intermediate resonances of the decays (see Table 2.2). Other decay modes are not used for measuring the $\mathcal{CP}$ nature of the $H\tau\tau$ coupling and hence not symbolized. The decay channels of $H \to \tau^+\tau^-$ are shown by pairs of the symbols; for example, $\rho\, a_1^{1\text{pr}}$ is the decay channel when one of the taus decays to ρ and the other to $a_1^{1\text{pr}}$.

In the $H \to \tau^+\tau^-$ decay, the spins of the taus are correlated based on the value of the mixing angle. Besides, the decay products in a tau decay, and hence the decay planes reconstructed with the products, are correlated with the tau spin. Consequently, the mixing angle determines the correlation between the decay planes produced by the taus from a Higgs boson decay. Fig. 2.4, for instance, shows these planes in the $\pi\,\pi$ channel depicted in the Higgs rest frame. One can define an angle between the two planes, called $\phi_{\mathcal{CP}}$, which ranges from 0° to 360° (as opposed to 180°) when the relative direction of each tau and its $\pi^\pm$ meson is taken into account. The $H \to \tau^+\tau^-$ differential decay width with respect to the $\phi_{\mathcal{CP}}$ at the leading order (LO) can be written as [47]

$$\frac{\mathrm{d}\Gamma}{\mathrm{d}\phi_{\mathcal{CP}}}(H \to \tau^+\tau^-) \approx 1 - b(E_+)\, b(E_-)\, \frac{\pi^2}{16}\, \cos(\phi_{\mathcal{CP}} - 2\alpha^{H\tau\tau}), \tag{2.17}$$

where for the $\tau^\pm$, $E_\pm$ is the energy of the outgoing charged particle, while $b(E_\pm)$ is its τ-spin analyzing power, encapsulating the correlation between the tau spin and the momentum of the outgoing charged particle.

Fig. 2.5 shows the $\phi_{\mathcal{CP}}$ distribution for the $\mathcal{CP}$-even, $\mathcal{CP}$-odd, and maximum mixing

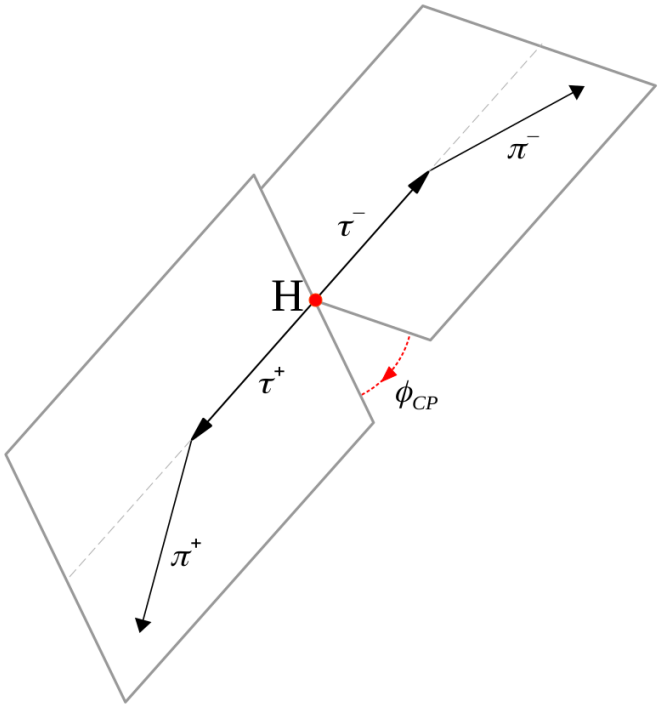

Figure 2.4: An illustration of the decay planes produced by the taus in the $\pi\,\pi$ decay channel. The angle between the planes is called $\phi_{\mathcal{CP}}$. The angle is defined in the Higgs rest frame [2].

Higgs scenarios. The figure also includes the $\phi_{\mathcal{CP}}$ distribution for a Z boson decay, one of the main backgrounds in this analysis. Each distribution is derived by fitting the normalized histogram produced by simulated events. The distributions are produced in the rest frame of the bosons at the generator-level, the level where no detector effects are included. The decay channel is $\pi\,\pi$ and taus with visible transverse momenta (p_{T}) less than 33 GeV are excluded. The simulated event samples used for producing these distributions are explained in chapter 6.

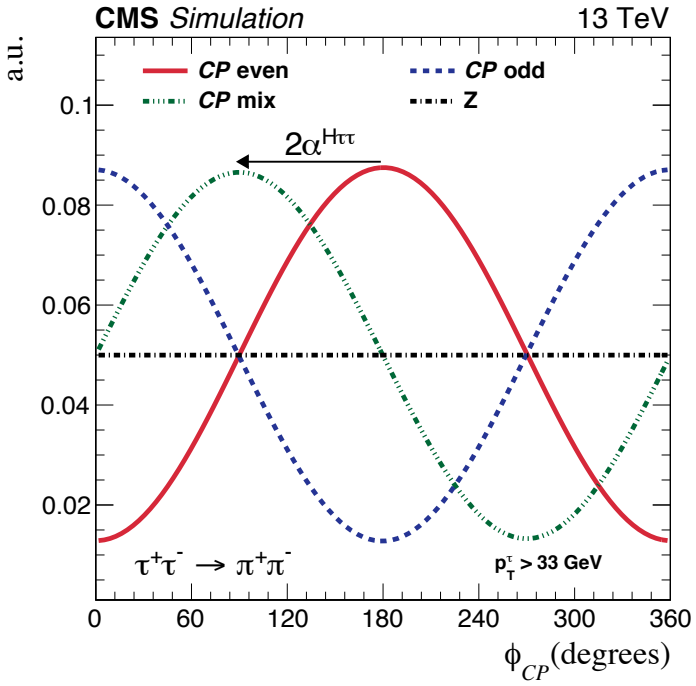

Figure 2.5: $\phi_{\mathcal{CP}}$ distribution at the generator-level for the $\mathcal{CP}$-even (red), $\mathcal{CP}$-odd (dash blue), and maximum mixing (dash-dot-dot green) Higgs boson scenarios along with the distribution for the Z boson (black dash-dot). To derive the distributions, a fit is performed to the normalized histogram of each process. The bosons decay to $\pi\,\pi$ final state and the visible tau decay products are required to have a minimum transverse momentum of 33 GeV [2].

A non-zero mixing angle would shift the sinusoidal $\phi_{\mathcal{CP}}$ distribution of the $\mathcal{CP}$-even scenario by $2 \times \alpha^{H\tau\tau}$ while retaining the amplitude of the distribution. The Z bosons, however, show a flat distribution and therefore this background does not distort the sinusoidal shape of the Higgs boson distribution at the generator-level.

It is, however, experimentally difficult to measure the $\phi_{\mathcal{CP}}$ angle (except for the $a_1^{\mathrm{3pr}}\,a_1^{\mathrm{3pr}}$ channel) since the momentum of the neutrinos in the tau decays cannot be well constrained and hence the Higgs boson rest frame cannot be well reconstructed. Fortunately, alternative, though similar, methods are proposed by phenomenologists

(as described and cited appropriately below) to reconstruct a decay plane for each of the tau leptons in an experimentally accessible way. The angle between the planes in the alternative methods is sensitive to the $\mathcal{CP}$ nature of the coupling in the same way as discussed although the rest frame is not the bosons' rest frame in most of the methods. In this thesis, the notation $\phi_{\mathcal{CP}}$ is used for the angle between the decay planes regardless of the method. The reference frame in which the $\phi_{\mathcal{CP}}$ angle is defined is the Higgs rest frame in the $a_1^{\text{3pr}}\, a_1^{\text{3pr}}$ channel whereas in all the other channels, the reference frame is the zero-momentum-frame (ZMF) of two charged particles chosen from the decay products of taus, as will be explained later in this section. In the kinematic phase space used in this analysis, the τ-spin analyzing power (see Eq. 2.17) in the e and μ decay modes has the opposite sign of that of the other decay modes and therefore the $\phi_{\mathcal{CP}}$ distribution is shifted by 180° in these two decay modes [48].

The alternative methods are the *impact parameter* method [49, 50], *neutral-pion* method [50, 51], *combined* method [50], and *polarimetric vector* method [52], which are explained below. The applicability and $\mathcal{CP}$ sensitivity of these methods vary in different $H \to \tau^+\tau^-$ decay channels. As will be discussed in chapter 6, for each of the $H \to \tau^+\tau^-$ decays, the method which yields the maximum $\mathcal{CP}$ sensitivity is used.

Impact parameter method

This method can be applied to all decay channels but, as explained in section 6.4.1, the optimal CP sensitivity is achieved when the method is used when each tau decays to one of the following decay modes: e, μ, or π. The idea of this method is to exploit the finite lifetime of tau leptons to reconstruct their decay plane. The plane is reconstructed using the momentum of the output charged particle and its impact parameter. The ZMF is defined using the two output charged particles.

The impact parameter vector $\vec{j}^{\pm}$ for an output charged particle, with $\pm$ showing the charge of the particle, is defined as the vector connecting the primary vertex (PV) to the closest point on the particle's track, where the PV is the reconstructed 3D position of the Higgs boson decay point. The four-component vector $\lambda^{\pm} = (0, \vec{j}^{\pm})$ is then constructed, as proposed in [49], and boosted similarly to a four-vector to the ZMF and is denoted by $\lambda^{\text{ZMF}\pm}$. The output charged particles are also boosted to this frame and are denoted by $q^{\text{ZMF}\pm}$. The $\lambda^{\text{ZMF}\pm}$ and $q^{\text{ZMF}\pm}$ are then considered as 3-component vectors by discarding their 0^{th} component. After that, the normalized vector of the transverse component of the $\lambda^{\text{ZMF}\pm}$ with respect to the $q^{\text{ZMF}\pm}$ is computed and shown with $\hat{\lambda}_{\perp}^{\text{ZMF}\pm}$. Using $\hat{\lambda}_{\perp}^{\text{ZMF}\pm}$ and the normalized vectors of the output charged particles

$\hat{q}^{\mathrm{ZMF}\pm}$, the following variables are defined

$$\begin{aligned} \phi^{\mathrm{ZMF}} &= \arccos(\hat{\lambda}_{\perp}^{\mathrm{ZMF}+} \cdot \hat{\lambda}_{\perp}^{\mathrm{ZMF}-}), \text{ and} \\ O^{\mathrm{ZMF}} &= \hat{q}^{\mathrm{ZMF}-} \cdot (\hat{\lambda}_{\perp}^{\mathrm{ZMF}+} \times \hat{\lambda}_{\perp}^{\mathrm{ZMF}-}) \end{aligned} \tag{2.18}$$

from which $\phi_{\mathcal{CP}}$ is computed as

$$\phi_{\mathcal{CP}} = \begin{cases} \phi^{\mathrm{ZMF}} & \text{if } O^{\mathrm{ZMF}} \geq 0 \\ 360^\circ - \phi^{\mathrm{ZMF}} & \text{if } O^{\mathrm{ZMF}} < 0 \end{cases}. \tag{2.19}$$

Fig. 2.6 (left) illustrates this method.

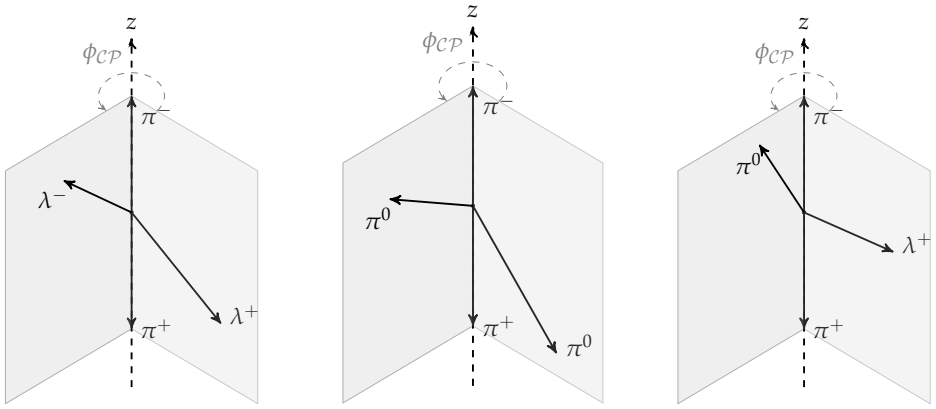

Figure 2.6: Illustration of reconstructing the $\phi_{\mathcal{CP}}$ angle in the impact parameter method (left), neutral-pion method (middle), and combined method (right). The planes are reconstructed in the frame in which the momenta of the charged particles are summed to zero. The decay planes reconstructed in the left, middle, and right figures correspond to the $\pi\,\pi$, $\rho\,\rho$, and $\pi\,\rho$ decay channels, respectively [2].

Neutral-pion method

The neutral-pion method is applicable to the decay channels in which both taus decay to more than one hadron, i.e. when each tau decay mode is among ρ, $a_1^{1\mathrm{pr}}$, or $a_1^{3\mathrm{pr}}$. The plane reconstruction method for each of these decay modes is described below.

In the ρ decay mode where we have a $\pi^\pm$ and a π^0 meson in the final state, a similar approach to the impact parameter method is taken to define the $\phi_{\mathcal{CP}}$ angle. The $\lambda^\pm$ defined in the impact parameter method is replaced with the four-momentum of the π^0. To estimate the four-momentum of the π^0, its energy is taken from the four-momentum sum of all photons/electrons[1] in the *strip* reconstructed by the HPS

[1]In CMS, the word "electron" is in general used for both electrons and positrons unless the two

algorithm described in section 4.8. The direction of the π^0 is, however, taken from the leading photon/electron in the reconstructed strip. Finally, the known mass of the π^0 meson is set as the mass of the reconstructed π^0. After replacing $\lambda^{\pm}$ with the π^0 four-momentum, Eq. 2.18 and Eq. 2.19 are used to define the $\phi_{\mathcal{CP}}$ angle.

To define a decay plane for the $a_1^{1\text{pr}}$ decay mode, which includes a pair of π^0 mesons in the final state that are experimentally difficult to resolve, the same approach as for the ρ decay mode is taken except that the four-momentum sum of the pair plays the role of the π^0 for reconstructing the decay plane.

To reconstruct the decay plane in the $a_1^{3\text{pr}}$ decay mode, from the two pairs of oppositely charged pions, the pair with an invariant mass closer to the ρ^0 meson mass is considered as coming from a ρ^0 meson decay (see Fig. 2.7). In this pair, the charged pion with the same charge as the tau is used to define the ZMF while the oppositely charged pion is treated as if it was a π^0 and therefore the decay plane is reconstructed with this pair using the neutral-pion method described for the ρ decay mode.

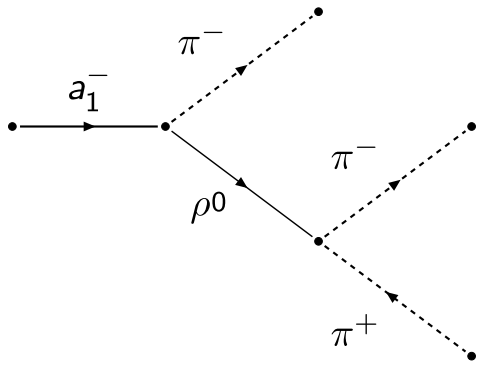

Figure 2.7: Feynman diagram showing the decay of a_1^- meson to three charged pions [2].

When reconstructing the $\phi_{\mathcal{CP}}$ angle in the neutral-pion method, in order to avoid destructive interference due to differently polarized intermediate mesons, the following parameters are defined in the lab frame [50]

$$\begin{aligned} y^{\tau^{\pm}} &= \frac{E_{\pi^{\pm}} - E_{\pi^0}}{E_{\pi^{\pm}} + E_{\pi^0}}, \text{ and} \\ y^{\tau} &= y^{\tau^-} y^{\tau^+}. \end{aligned} \tag{2.20}$$

When y^{τ} is negative, $\phi_{\mathcal{CP}}$ is set to $360^\circ - \phi_{\mathcal{CP}}$.

particles are contrasted in a sentence.

To summarize, the neutral-pion method is similar to the impact parameter method but the $\lambda^{\pm}$ is replaced with the "π^0" four-momentum, where "π^0" is the π^0 meson in the ρ decay mode, the four-vector sum of the two π^0 mesons in the a_1^{1pr} decay mode, and the $\pi^{\pm}$ with an opposite charge to the tau in the a_1^{3pr} decay mode. The ZMF is defined using the $\pi^{\pm}$ meson with the same charge as the tau, and in the a_1^{3pr} decay mode in which there are two choices, the one which is more likely to come from the ρ^0 meson decay is selected. Finally, the destructive interference is avoided by applying a shift to the $\phi_{\mathcal{CP}}$ according to Eq. 2.20.

The neutral-pion method is illustrated in Fig. 2.6 (middle).

Combined method

The combined method is a combination of the impact parameter and neutral-pion methods and is applied to the decay channels where exactly one of the taus decays to multiple hadrons such as $\pi\, a_1^{\text{3pr}}$ and $e\, a_1^{\text{1pr}}$. For reconstructing the decay planes in these decay channels, the neutral-pion and impact parameter methods are applied to the tau decaying to multiple hadrons and the other tau, respectively. To avoid destructive interference, if the $y^{\tau^{\pm}}$ (see Eq. 2.20) for the decay plane reconstructed with the neutral-pion method is negative, then the shift $360^\circ - \phi_{\mathcal{CP}}$ is applied to obtain the $\phi_{\mathcal{CP}}$ angle.

An illustration of the combined method is displayed in Fig. 2.6 (right).

Polarimetric vector method

The last method used in this analysis is the polarimetric vector method, which is only applicable to the $a_1^{\text{3pr}}\, a_1^{\text{3pr}}$ decay channel. In this channel, the Higgs rest frame can be reconstructed thanks to the reconstructable secondary vertex (SV) in the a_1^{3pr} decay. The momentum direction of each tau is set to the PV-SV direction and, considering the two-body decay of $\tau^{\pm} \to a_1^{\pm}\nu_\tau$ and assuming a massless neutrino, the magnitude of the tau momentum can be derived to be

$$|\vec{p}_\tau| = \frac{(m_{a_1}^2 + m_\tau^2)|\vec{p}_{a_1}|\cos\theta_{\text{GJ}} \pm \sqrt{\left(m_{a_1}^2 + |\vec{p}_{a_1}|^2\right)\left((m_{a_1}^2 - m_\tau^2)^2 - 4m_\tau^2|\vec{p}_{a_1}|^2\sin^2\theta_{\text{GJ}}\right)}}{2(m_{a_1}^2 + |\vec{p}_{a_1}|^2\sin^2\theta_{\text{GJ}})} \tag{2.21}$$

where θ_{GJ} is the Gottfried-Jackson angle defined as the angle between the τ lepton and a_1 directions in the lab frame [53]. If, in a tau decay, θ_{GJ} is measured to be more than

$$\theta_{\text{GJ}}^{\text{max}} = \arcsin\left(\frac{m_\tau^2 - m_{a_1}^2}{2m_\tau|\vec{p}_{a_1}|}\right). \tag{2.22}$$

θ_{GJ} is set to $\theta_{\mathrm{GJ}}^{\mathrm{max}}$. Values above $\theta_{\mathrm{GJ}}^{\mathrm{max}}$ are unphysical but can still be measured because of the finite detector resolution in measuring the tau and a_1 directions. The square root term in Eq. 2.21 shows that there are up to two answers for the magnitude of the tau momentum. When the square root is zero, ending up with a unique answer for the tau momentum, the a_1 direction in the tau rest frame is perpendicular to the tau direction in the lab frame. On the other hand, a non-zero square root indicates two possible directions for the a_1 in the tau rest frame: the same as or the opposite of the tau momentum in the lab frame. Therefore, there are up to four pairs of solutions for the momenta of the tau pair in the $a_1^{\mathrm{3pr}}\, a_1^{\mathrm{3pr}}$ channel. The pair of solutions with the invariant mass closest to the Higgs boson mass is selected.

Now that the momentum of the tau and a_1 are known, the polarimetric method can be applied. The polarimetric vector $\vec{h}$ can be considered as an estimate of the most likely direction of the tau spin $\vec{s}$ in the tau rest frame [52]. This vector can be computed from the momentum of the tau and a_1 as implemented in the TAUOLA [54–56] program. The decay plane can be reconstructed using the tau direction and the polarimetric vector. The normal vector $\vec{k}$ to the decay plane is reconstructed with

$$\vec{k}_{1,2} = \frac{\vec{h}_{1,2} \times \vec{n}_{1,2}}{|\vec{h}_{1,2} \times \vec{n}_{1,2}|} \tag{2.23}$$

where $\vec{n}$ is the unit vector pointing along the direction of each tau and the indices run over the two taus. In the next step, the Higgs rest frame is used to define ϕ^* and O^* as (c.f. Eq. 2.18)

$$\begin{aligned} \phi^* &= \arccos(\vec{k}_1 \cdot \vec{k}_2), \text{ and} \\ O^* &= -(\vec{h}_1 \times \vec{h}_2) \cdot \vec{n}_1 \end{aligned} \tag{2.24}$$

from which $\phi_{\mathcal{CP}}$ angle can be reconstructed using Eq. 2.19 while replacing ϕ^{ZMF} and O^{ZMF} with ϕ^* and O^*, respectively.

The $\phi_{\mathcal{CP}}$ angle can also be reconstructed in the $a_1^{\mathrm{3pr}}\, a_1^{\mathrm{3pr}}$ channel using the neutral-pion method. However, using the polarimetric vector method enhances the resolving power between the $\mathcal{CP}$-even and $\mathcal{CP}$-odd scenarios by approximately a factor of two. Therefore, the polarimetric vector method is used in this channel.

Chapter 3

The LHC and the CMS detector

The data used in the Higgs-tau $\mathcal{CP}$ analysis presented in this thesis was recorded by the Compact Muon Solenoid (CMS) detector [57] located at CERN, near Geneva. CMS is placed about 100 meters underground and is designed to detect particles produced from proton-proton (pp) collisions. These protons are accelerated and collided in the Large Hadron Collider (LHC) [58]. This chapter describes the LHC along with the CMS detector.

3.1 The LHC

The LHC is a circular hadron accelerator and collider with approximately a 27 km circumference, built at CERN in an underground tunnel previously used by the Large Electron-Positron (LEP) collider [59] during 1989-2000. Fig. 3.1 shows a schematic view of the LHC complex. The process of obtaining protons leading up to their collisions is the following. Protons are obtained by ionizing hydrogen gas with an electric field before being accelerated up to an energy of 50 MeV in a linear accelerator (LINAC2). After that, their energy is increased to 1.4, 25, and 450 GeV using the *Proton Synchroton* (PS) Booster, the PS itself, and the *Super Proton Synchroton* (SPS) circular accelerators, respectively. Finally, they are injected into the LHC beam pipes in counter-rotating directions, with each beam subsequently reaching an energy of 6.5 TeV. Each proton beam is split into bunches of $O(10^{11})$ protons separated by 25 ns, which are accelerated using radio-frequency (RF) cavities. In order for the protons to stay in the circular path of the LHC beam pipes, several 8.3 T superconducting dipole magnets are installed at the LHC operating at a temperature of 1.9 K. Proton bunches from the two beams are collided in four points, surrounded by the ALICE [60], ATLAS [61], CMS [57], and LHCb [62] detectors.

The expected number of events containing a specific physics process can be calcu-

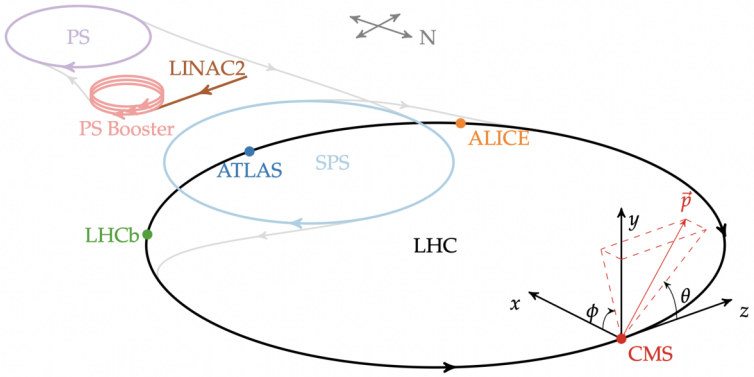

Figure 3.1: A not-to-scale schematic of the LHC complex. The coordinate system depicted on "CMS" shows the orientation of the Cartesian coordinate used in CMS. In addition, the azimuthal (ϕ) and polar (θ) angles for a particle with momentum $\vec{p}$ are displayed [63].

lated using

$$N = \mathcal{L} \times \sigma, \tag{3.1}$$

where $\mathcal{L}$ is the integrated luminosity and σ is the cross section of the process. Fig. 3.2 shows the cross sections of certain physics processes as a function of the centre-of-mass energy. The largest Higgs production cross section is for the gluon-gluon fusion (ggH) process, which is about nine orders of magnitude smaller than the total pp cross section. Therefore, the study of the Higgs boson properties requires a large number of pp collisions, and hence a large integrated luminosity.

The LHC has had several running periods. The first three periods known as Run-1, Run-2 and Run-3, were during 2010-12, 2015-18, and will be 2022-25, respectively. Following Run-3, the LHC will upgrade to the *high-luminosity LHC* (HL-LHC) with two running periods known as Run-4 and Run-5, operating during 2029-32 and 2035-38, respectively [65].

The data used for the Higgs-tau $\mathcal{CP}$ analysis described in this thesis were collected during Run-2 in 2016, 2017, and 2018, which correspond to integrated luminosities of 35.9, 41.5, 59.7 fb^{-1}, respectively, and a centre-of-mass energy of $\sqrt{s} = 13$ TeV.

3.2 The CMS detector

CMS is a general-purpose detector which was designed to discover and subsequently study the properties of the Higgs boson. CMS is cylindrical in shape with a diameter of 14.6 m, a length of 21.6 m, and a weight of 12500 tonnes. The distinctive feature of this detector is its superconducting solenoid magnet which provides a magnetic field of about

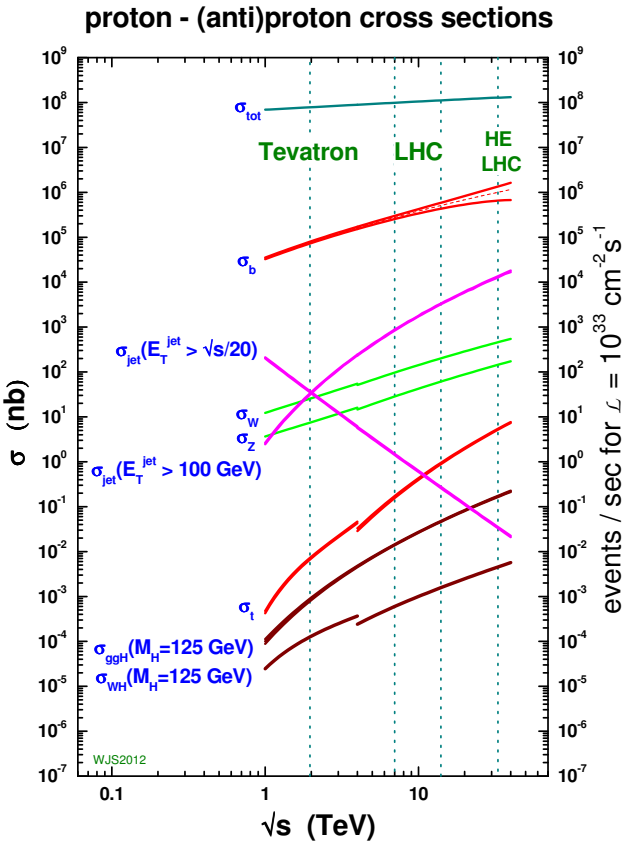

Figure 3.2: The total cross section of pp collisions (σ_{tot}), along with the cross sections of certain physics processes which can occur in such collisions, all plotted as functions of the centre-of-mass energy [64]. Interesting physics processes, such as the Higgs boson production, are usually several orders of magnitude smaller than the total cross section.

4 and 2 T inside and outside the solenoid, respectively. The detector consists of several subdetectors arranged in an onion-like structure and centred at the pp collision region (see Fig. 3.3). From the collision point outwards, the first subdetector is the silicon tracker which is designed to measure the position and, thanks to the magnetic field, the momentum of charged particles. The next subdetectors are the electromagnetic calorimeter (ECAL) and hadron calorimeter (HCAL) which measure the energies of particles. These three subdetectors are placed within the solenoid magnet, whereas the muon system and the iron return yoke are outside.

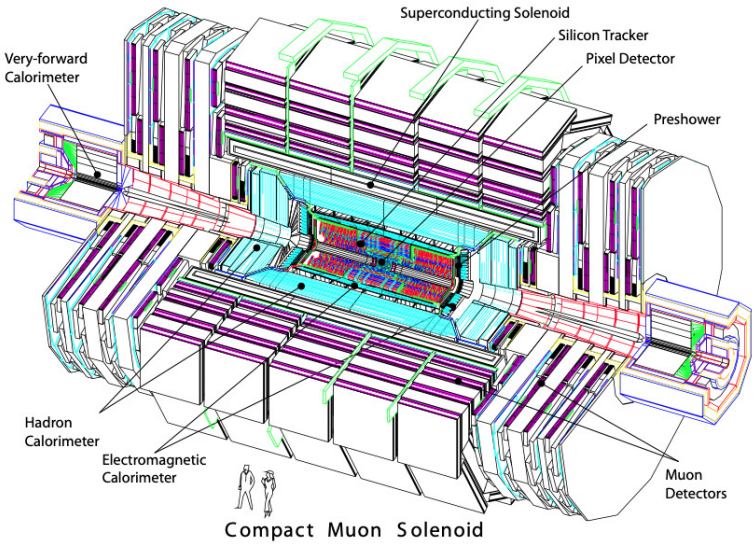

Figure 3.3: A perspective view of the CMS detector [57].

As illustrated in Fig. 3.1, the coordinate system used in CMS is defined such that the x axis points towards the centre of the LHC ring, the y axis points upward, and the z axis is derived with the right-hand rule and points along one of the beam directions. A more common coordinate system used in CMS is two-dimensional $\eta-\phi$ coordinates, where η is the pseudorapidity defined using the polar angle θ

$$\eta = -\ln\left(\tan\left(\theta/2\right)\right) \tag{3.2}$$

and ϕ is defined in the $x-y$ plane as the azimuthal angle with respect to the x axis. Distance in these coordinates is defined as

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}. \tag{3.3}$$

The transverse component of an object (such as its momentum, denoted by p_{T}) is defined in the $x-y$ plane. In CMS, the low pseudorapidity region is called the *barrel*,

whereas the two ends of the detector with high pseudorapidity are called the *endcaps*.

3.3 Silicon tracker

The first subdetector encountered by particles produced from pp collisions is the tracker [66], which consists of a silicon pixel detector surrounded by a silicon strip tracker. The tracker is used to measure the position of charged particles by collecting the produced ionization deposits (known as hits). A collection of hits is then used to reconstruct the curved trajectory of charged particles and subsequently measure their p_{T} through their radius of curvature. The trajectories are also used to reconstruct other quantities of interest such as the position of pp collisions as well as the decay position (and subsequently the impact parameter) of finite-lifetime particles. The tracker is made from silicon in order to precisely reconstruct these quantities, have a fast response, and withstand extreme radiation conditions given the large ($\approx$1000 [67]) number of charged particles traversing the tracker at each bunch crossing.

Fig. 3.4 displays a schematic view of the tracker in the $r-z$ plane. The tracker acceptance range is $|\eta| < 2.5$. The pixel detector was originally composed of three layers in the barrel and two layers in each endcap with a total of 66 million pixels. This design was optimized for reconstructing the 3D position of particles and handling the large number of particles produced at the LHC design luminosity 1×10^{34} cm^{-2} s^{-1}. In winter 2017, the pixel detector was upgraded to cope with a factor of two larger luminosity such that it is comprised of one more layer in the barrel and in each of the endcaps and the total number of pixels increased to more than 100 million. The upgraded pixel detector has a pixel size of 100×150 μm^2 and is closer to the pp collision region. It is made with less material budget which reduces photon conversions and multiple scattering [68].

Surrounding the silicon pixel is the strip tracker which is comprised of four subsystems, namely, the tracker inner barrel (TIB), tracker inner disk (TID), tracker outer barrel (TOB) and tracker endcap (TEC), which differ by the position and size of the silicon strips. The TIB and TOB are installed in the barrel and consist of four and six silicon layers, respectively. The TID (TEC) is made of three (nine) disks placed in the inner (outer) part of each endcap.

The performance of the CMS tracker is summarized in Ref. [67]. Charged particles produced as far as 60 cm from the beam line, and those with p_{T} as low as 0.1 GeV are reconstructable. The track-reconstruction efficiency for charged particles with $p_{\mathrm{T}} > 0.9$ GeV is measured to be 94% and 85% in the barrel and endcap regions, respectively, using simulated $t\bar{t}$ samples and assuming the typical LHC pile-up conditions of 2011

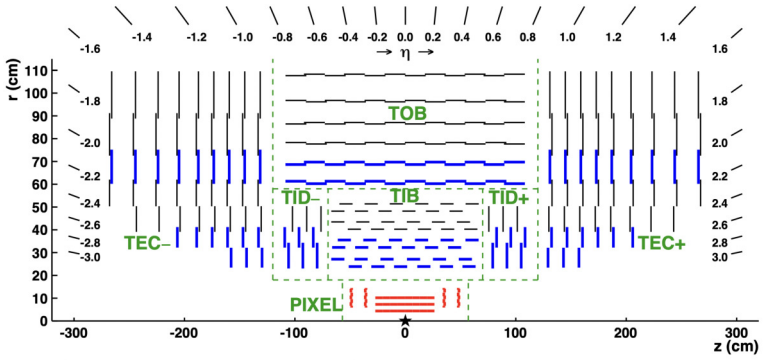

Figure 3.4: A cross sectional view of the tracker system. Only the upper half of the system is shown, while the lower half is symmetric with respect to the z axis. The pp collision region is marked with a star and the green lines separate the tracker subsystems. The pixel modules, shown in red lines, provide 3D hits. The silicon modules shown in blue and black provide 3D and 2D hits, respectively [67].

running period. The dominant reason for the inefficiency is the nuclear interactions of hadrons in the tracker. The ratio of tracks which are falsely reconstructed is at the few percent level. The p_{T} resolution of tracks with $1 < p_{\mathrm{T}} < 10$ GeV in the central region is about 1.5%. The transverse (longitudinal) impact parameter resolution of these tracks changes from 90 µm (150 µm) at p_{T}=1 GeV to 25 µm (45 µm) at p_{T}=10 GeV.

3.4 Electromagnetic calorimeter

Around the tracker lies the electromagnetic calorimeter (ECAL) [66] which is hermetic and homogeneous with 61200 lead-tungstate ($PbWO_4$) scintillating crystals in the barrel (EB) and 7324 in each endcap (EE). In front of each endcap calorimeter, a preshower detector (ES) is placed to improve the identification of photons mainly from $\pi^0 \to \gamma\gamma$ decay. Avalanche photodiodes (APDs) and vacuum phototriodes (VPTs) are used as photodetectors in the EB and EE, respectively. The ECAL is constructed in a compact and granular design, thanks to the short radiation length ($X_0 = 0.89$ cm), small Moliere radius ($R_M = 2.2$ cm), and high density ($\rho = 8.28$ g/cm^3) of the $PbWO_4$ crystals.

Fig. 3.5 displays a schematic of the ECAL. The ECAL crystals face (almost) towards the pp collision region with a slight tilt to reduce the probability of particle passing through the inter-crystal gaps. Crystals in the EB (EE) have an approximate area of 2.2 cm × 2.2 cm (2.86 cm × 2.86 cm) and a length of 23 cm (22 cm). The EB coverage is up to $|\eta| = 1.48$ and the EE extends the coverage to $|\eta| = 3.0$. The fiducial area of the ES covers $1.65 < |\eta| < 2.6$. Discriminating photons and electrons is possible up to $|\eta| = 2.5$, which is the acceptance region of the tracker system.

The energy resolution of electrons in the EB as a function of their energy E is

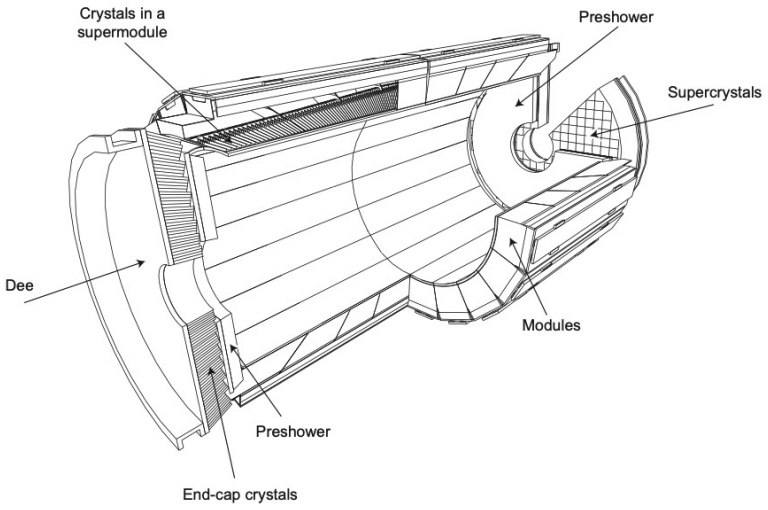

Figure 3.5: A schematic view of the ECAL [69].

measured to be:

$$\frac{\sigma_E}{E} = \frac{0.028}{\sqrt{E}} \oplus \frac{0.12}{E} \oplus 0.003 \tag{3.4}$$

where E is expressed in GeV, and the three terms on the right hand side correspond to the stochastic, noise, and constant term contributions. The stochastic term includes fluctuations from gain process, electromagnetic shower containment, and the number of produced photoelectrons. The noise term, which has energy-independent absolute value, corresponds to the noise in the readout system. The constant term is the dominant term in high energy region and accounts for energy leakage from the back of the crystal as well as non-uniformity in the single-channel response and longitudinal light collection [69].

3.5 Hadronic calorimeter

The hadronic calorimeter (HCAL) [66] is located behind the ECAL to identify strongly interacting particles as well as measure their energy. This subdetector plays an important role in the reconstruction of quark or gluon initiated jets as well as missing transverse momentum (both defined in chapter 4). As illustrated in Fig. 3.6, the HCAL is comprised of four subsystems: the barrel hadronic (HB), endcap hadronic (HE), outer hadronic (HO), and forward hadronic (HF) calorimeters. The HB ($|\eta| < 1.3$) and HE ($1.3 < |\eta| < 3.0$) consist of sampling calorimeters with brass as absorber and plastic scintillator tiles as active material. The sampling calorimeters are segmented in the $\eta-\phi$ plane, where each segment is called a *tower*. The size of towers in $\Delta\eta \times \Delta\phi$ is 0.087×0.087 for $|\eta| < 1.6$ and 0.17×0.17 for higher pseudorapidity regions. The scintil-

lation light is absorbed and re-emitted by wavelength-shifting (WLS) fibers embedded in the scintillator tiles and then transferred to photodetectors.

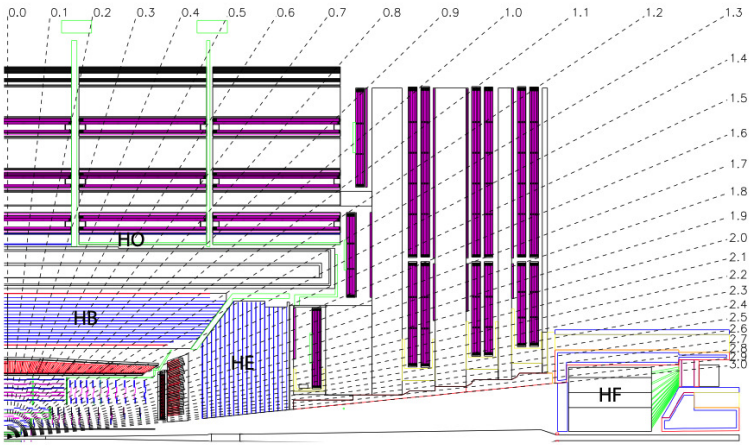

Figure 3.6: A cross sectional view of the HCAL and its subsystems overlaid with constant-pseudorapidity lines (dashed black) [57].

The HB is placed in the gap between the ECAL and solenoid magnet, which constrains the amount of absorber material fittable in the gap. Due to this constraint, the absorber material in the HB ranges from 5.82 to 10.6 interaction length (λ_I) when η increases from 0 to 1.3, while larger values of λ_I are needed to minimize hadronic shower leakage. The total λ_I is increased to 11 by placing the HO behind the HCAL and the solenoid magnet when looking from the centre of the CMS detector. The ECAL crystals add a further 1.1 λ_I of material. The HF is placed in the forward region to extend the HCAL coverage to $|\eta| = 5$. It is a Cherenkov detector with iron as the hadron absorber and quartz-fibre as the active material, inside which particles emit Cherenkov light that is subsequently detected by photomultiplier tubes (PMT) [57].

The energy resolution of the ECAL and HCAL calorimeters, combined, as a function of energy is measured [70, 71] for a charged pion to be

$$\frac{\sigma_E}{E} = \frac{1.1}{\sqrt{E}} \oplus \frac{0.09}{E} \tag{3.5}$$

where E is in units of GeV.

3.6 Solenoid magnet

The superconducting solenoid magnet [72], as a key component of the CMS detector, provides approximately 4 T magnetic field inside and 2 T outside the solenoid to identify the charge and measure the momentum (and hence the energy) of charged particles.

This high magnetic field is essential to exploit the physics potential of the LHC. With a high magnetic field, the ECAL occupancy decreases through trapping low-p_{T} charged particles in the tracker. In addition, the muon trigger efficiency increases while the trigger rate remains within the allowed limit. (See section 3.8 for more information on triggering.) Another advantage is that the momentum of muons can be measured precisely, which is important for studying several physics processes [72]. The solenoid, with a diameter of 6 m and a length of 12.5 m, can store 2.6 GJ of energy at full current. Helium liquid cooling is used to provide a 4.5 K temperature for the superconducting magnet to operate. In order to return the flux, a 10000 tonne iron yoke is used which consists of 6 endcap discs and 5 barrel wheels [57, 72].

3.7 Muon system

The detection of muons with a high resolution is of high importance in CMS (as is implied by the middle name of the experiment) since it is a powerful tool for BSM searches as well as precise SM measurements. For example, the role of muons was important in the search for the $H \to ZZ \to 4\ell$ channel because of the clean signal they leave in the detector.

The outermost subdetector of the CMS is the muon system [66] responsible for identifying muons, measuring their momentum, and triggering on them. The muon system, shown in Fig. 3.7, is comprised of three gas detectors: (i) drift tube chambers (DT), (ii) cathode strip chambers (CSC), and (iii) resistive plate chambers (RPC). These detectors are interspersed in the iron return yoke of the solenoid magnet. The DT chambers, covering $|\eta| < 1.2$, are made of tube-shaped drift cells in which electrons, ionized as a result of traversing muons, drift towards anode wires located in the cell. The DT chambers include four stations, with each station containing several chambers. The first three stations measure muon coordinates in $r{-}\phi$ as well as in z, while the fourth station only measures them in $r{-}\phi$. The position and number of chambers and drift cells are carefully chosen to optimize angular resolution, eliminate dead spots in the efficiency, measure muon time with an excellent time resolution, and achieve good efficiency in linking muon hits for reconstructing muon tracks. In the endcaps, where muon rate is high and the magnetic field is strong and non-uniform, the CSCs are used. The CSCs, covering $0.9 < |\eta| < 2.4$, are finely segmented, radiation-hard, and fast in response time. They consist of four stations of which chambers are positioned perpendicular to the beam line. The cathode strips provide $r{-}\phi$ position measurements while the anode wires measure the η coordinate as well as the muon beam-crossing time. The DT and CSC subsystems can each efficiently trigger on the p_{T} of a muon

independent of the rest of the detector. The RPCs are added in both the barrel and endcaps in order to provide further dedicated triggers on muons. Operated in avalanche mode, they can operate at high rates and provide independent, highly-segmented, and fast trigger over $|\eta| < 1.6$ [57].

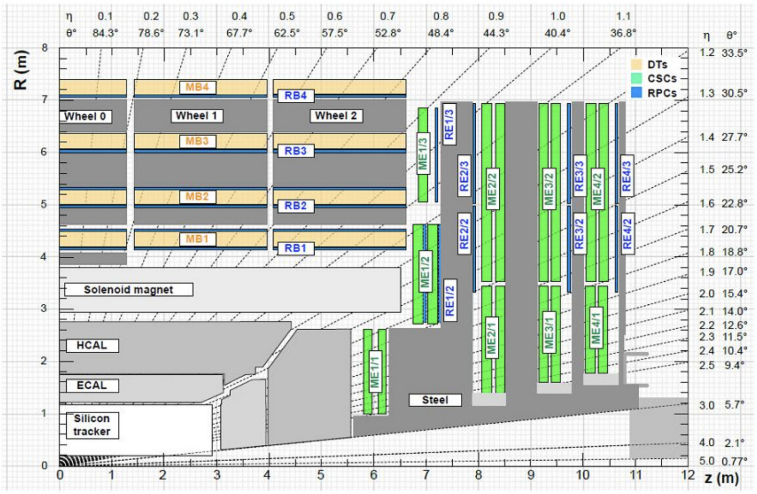

Figure 3.7: A cross sectional view of the muon system and its subsystems: DT (yellow), CSC (green), and RPC (blue) [73].

The performance of the muon system was studied in Ref. [74] using a proton-proton data sample collected by CMS in 2010 corresponding to an integrated luminosity of 40 fb^{-1}. The spatial resolutions in the DT, CSC, and RPC chambers were measured to be 80-120 µm, 40-150 µm, and 0.8-1.3 cm, respectively. A time resolution of 3 ns per chamber was achievable. The efficiency of reconstructing muon tracks traversing the muon chambers was 95-98%.

3.8 Triggering

The LHC bunch crossing occurs at 40 MHz frequency while, due to the limited available resources, only a tiny fraction of the events (O(1 kHz)) can be processed and stored in memory. Furthermore, most pp collisions result in events in which new physics is not expected to emerge, such as those including low energy multijets. Therefore, an *online* triggering system is designed for making a judicial decision of accepting the events likely to contain new physics and rejecting the rest. The word "online" is contrasted to "offline" and refers to the real-time processing tasks with strict timing constraints (typically a few µs to less than a second)

The online triggering system in CMS consists of two tiers: the Level-1 trigger (L1T) [75, 76] and High-Level Trigger (HLT) [77]. Data is firstly processed by the

L1T, which is based on custom-programmable processors and then those accepted by the L1T are processed by the HLT, which is performed on commercial CPUs. Events accepted by the HLT are stored on disk for offline use.

3.8.1 Level-1 trigger

Fig. 3.8 shows a diagram of the dataflow in the CMS Level-1 trigger. In this trigger level, the calorimeters and the muon system independently trigger on objects, whereas the tracker is not used in the triggering process. The first objects to be processed are trigger primitives (TPs) which are basic particle signatures in the detector, such as track segments in the muon system and energy deposits in the calorimeter. This processing step includes simple tasks such as clustering and calibration, as well as assigning $\eta-\phi$ position, energy, and transverse momentum to clusters and muon tracks. In the next processing step, physics objects such as taus, muons, and hadronic jets are identified. These objects are sorted by $p_{\mathrm{T}}/E_{\mathrm{T}}$ ratio and the leading twelve objects for each object type are then sent to the global trigger. The global trigger accepts or rejects events by combining the information from the calorimeters trigger and the muon trigger. The accepted events are passed to the HLT. The L1T has a latency of 3.8 μs and an output rate of 100 kHz, so 1 out of 400 collisions is selected at this level.

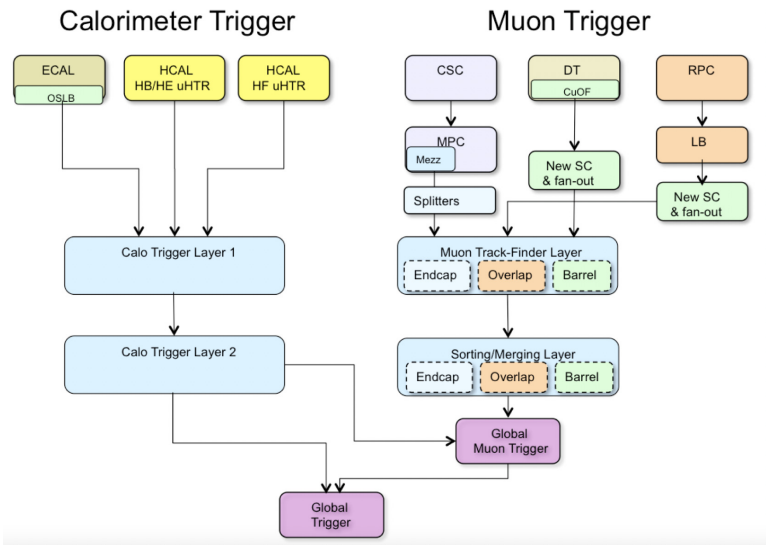

Figure 3.8: Dataflow in the Level-1 trigger [76].

The calorimeters trigger makes use of a time-multiplexed system. In the layer 1 of the trigger (see Fig. 3.8), each node processes only a slice of the detector but for several bunch crossings. The processed data are then passed to the layer 2 in which these slices join and the information from a bunch crossing is processed together on a

single node.

To reconstruct physics objects in the L1T, the calorimeter trigger makes use of *trigger towers* (TTs). A TT shows the amount of energy stored in ECAL and HCAL in a certain $\eta-\phi$ region. TT regions have a $\Delta\phi \times \Delta\eta$ size of 0.087×0.087 in the barrel region and larger in the endcaps. In the barrel, TTs map exactly to 5×5 ECAL crystals and to 1×1 HCAL towers. In the endcaps, the sizes of TTs are larger and irregular but are as close as possible to the HCAL towers.

The objects in the L1T are reconstructed in the following methods [78]:

- **Photons and electrons:** In the L1T, photons and electrons, collectively shown in this documents as γ/e, are reconstructed in the same way since they cannot be distinguished given the absence of the tracker information in this step. The algorithm finds a TT for which the energy deposit is locally maximum and then adds nearby TTs to construct a tower cluster. The shape of the tower cluster as well as the fraction of energy stored in the ECAL part of it are used for discriminating against hadronic jets. In addition, the algorithm determines if the object is isolated and also assigns an energy-weighted position to it.

- **Hadronic tau decays (τ_h):** The basic idea of reconstructing τ_h and γ/e candidates is the same. However, since τ_h candidates can decay to multiple hadrons and therefore are generally more spread in the $\eta-\phi$ plane than electrons, nearby tower clusters are merged to reconstruct τ_h candidates. In addition, a dedicated algorithm is developed for discriminating τ_h against jets.

- **Hadronic jets and missing transverse energy:** After finding a tower which is a local maximum, 9×9 TTs centring at the local maximum tower are constructed and considered as coming from a jet. The energy of the reconstructed jet is the sum of its TTs after deducting an estimated pile-up contribution. In addition, the imbalance in the energy deposited in the calorimeters (called missing transverse energy) and the scalar sum of the energy of all jets are other reconstructed L1T objects which are important for identifying certain physics processes.

- **Muons:** Muon identification and reconstruction algorithms, which are performed in the muon trigger, are different depending on the η of the TPs. In the $|\eta| < 0.83$ and $0.83 < |\eta| < 2.40$ regions, the track finding algorithm utilizes extrapolation and pattern matching methods, respectively. The transverse momentum assigned to each muon track is estimated using the difference in ϕ coordinate (in $|\eta| < 0.83$), pattern matching (in $0.83 < |\eta| < 1.20$), and boosted decision tree (BDT) regression (in $1.2 < |\eta| < 2.4$) methods.

3.8.2 High-Level Trigger

The HLT processes events which are accepted by the L1T with algorithms as close as possible to those used offline. The HLT algorithms, however, are optimized to run two orders of magnitude faster than the offline algorithms in order to fit into the limited available resources. The algorithms are designed in several modules between which filters are applied to reject events which do not meet certain criteria. This allows rejection of uninteresting events without being fully reconstructed. After an event is fully reconstructed, hundreds of signatures for interesting physics are searched for, after which the HLT accepts the event if at least one of the signatures is found. The bandwidth required for the HLT output is 2 GB/s and the latency of the HLT is ≈300 ms [78].

Chapter 4

Physics object reconstruction

4.1 Introduction

This chapter reviews the methods used in CMS for reconstructing physics objects such as taus and muons from low-level signatures in the detector such as hits in the tracker and energy deposits in the calorimeters. The core algorithm for reconstructing all individual particles in CMS is the *Particle Flow* (PF) [71] algorithm and the particles reconstructed by this algorithm are called *PF candidates*. The main idea behind this algorithm is to utilize the signatures left by particles in different subdetectors for identifying and reconstructing the particles along with measuring their kinematic quantities. Higher-level physics objects, such as hadronic jets and hadronic decays of taus (τ_h), are reconstructed by grouping PF candidates with dedicated algorithms.

4.2 Tracks

4.2.1 General method

A track is reconstructed by fitting a curve to the hits in the silicon or muon tracker for the purpose of measuring the position and, due to the presence of the CMS magnetic field, the momentum and electric charge of charged particles. In order to reconstruct tracks, CMS employs the *Combinatorial Track Finder* (CTF) algorithm [67] in which *Kalman filter* (KF) [79] is used iteratively. In the initial iterations, the algorithm searches for tracks which are easier to find, such as those with high-p_{T} originating from or close to the pp interaction region. In contrast, more difficult tracks, such as low-p_{T} ones and those with a largely displaced origin, are reconstructed in later iterations. After each iteration, hits associated to the already-reconstructed tracks are removed to simplify track-finding in the subsequent iterations.

The strategy for finding tracks in each iteration is the following. Firstly, in order to generate track seeds, two or three pixel hits are selected and used to estimate the parameters and the associated uncertainties of a particle track. Then, making use of the KF algorithm, the track is extrapolated in the expected direction of flight where a hit which is compatible with the track is selected. With the new hit involved, the track parameters and the uncertainties are updated using the best fit to the hits. This extrapolation continues and each time the track parameters are updated using the new hit. After the track is reconstructed, it is required to pass a set of quality conditions to suppress fake tracks which are those not originating from a real particle.

4.2.2 Electron tracking

The energy loss of electrons[1] through bremsstrahlung radiation in the tracker is usually significant, causing non-negligible changes to their trajectory. For instance, on average, approximately 35% of electrons lose more than 70% of their initial energy before reaching the ECAL. Therefore, the track reconstruction algorithm needs to be modified to take into account this effect by combining information from both the tracker and ECAL.

Two methods are employed to find the seeds of electron tracks. The first method [80], which is the traditional method in CMS (called the *ECAL-based* method hereafter), is performed by searching for ECAL *clusters* with similar η over a relatively wide ϕ range, combining them as a *supercluster*, and associating the supercluster to a seed in the tracker. (See section 4.4 for information about clustering in the ECAL.) Choosing a narrow (wide) range of η (ϕ) is based on the fact that the direction of the CMS magnetic field bends the trajectory of charged particles in ϕ rather than η and hence the radiation is (locally) spread only over ϕ. Finally, a track seed compatible with the supercluster and originating near the centre of the *beam spot* is selected as an electron seed, where beam spot is a 3D profile of the luminous region computed from an average over several bunch crossings . Electrons which are missed in the ECAL-based reconstruction method are targeted by the second method [81] which is developed in the context of the PF algorithm (hereafter called the *track-based* method). In this method, the track seeds compatible with electron seeds are selected by the general method described in section 4.2.1. The compatibility for an electron with low-energy radiation means finding a cluster in the ECAL close to the extrapolation curve of the track. In contrast, for an electron with high-energy radiation, it is defined as tracks with poor goodness-of-fit or few associated hits.

[1] In CMS, the word "electron" is in general used for both electrons and positrons unless the two particles are contrasted in a sentence.

The seeds collected by the two methods are merged and considered as electron seeds. In order to get the best track parameters, a modified version of the KF algorithm, called the *Gaussian Sum Filter* (GSF) [82], is used to correct for the non-Gaussian distribution of the Bethe-Heitler electron energy-loss formula. The GSF basically decomposes the non-Gaussian distribution into a set of Gaussian ones. Finally, a requirement on the score of a boosted decision tree (BDT) is applied for a track to be selected. The BDT features include track parameters, hits properties, goodness-of-fit, energy-loss along the track and the distance between the extrapolated track and the ECAL cluster.

4.2.3 Muon tracking

Muons can pass through the tracker and calorimeters and reach the muon system. Therefore, there are several possible ways to reconstruct a muon track depending on which subdetector(s) is used. CMS reconstructs the following types of muon tracks:

- **Standalone muon:** Only hits in the muon system are used to seed and reconstruct a muon.
- **Global muon:** A track reconstructed in the tracker is matched to a standalone muon track and a global fit is performed to form a global muon track.
- **Tracker muon:** A track reconstructed only using the tracker is matched to at least one muon segment. A minimum transverse (total) momentum of 0.5 (2) GeV is required for the track.

Among the above-mentioned muon tracks, standalone muons usually have the worst momentum resolution and are contaminated by cosmic muons. In contrast, global muons are more efficient than the other two when a muon penetrates at least two muon detector planes.

More information about track-finding in CMS can be found in [71] and [67] .

4.3 Vertices

This section describes how the 3D position and the associated uncertainty of pp collision points, known as vertices, are computed using the information of tracks. This process consists of three steps: selecting tracks, clustering those appearing to originate from the same interaction point, and fitting each track cluster to find the position of the corresponding vertex.

In the first step, tracks which are compatible with prompt production in the primary interaction region are selected. The compatibility condition is imposed by vetoing

tracks with a small associated number of pixel and strip hits, those with large χ^2 from the track fits, and those with a transverse impact parameter with very large significance compared to the centre of the beam spot.

In the next step, the selected tracks are clustered based on the z-coordinate of their point of closest approach with respect to the centre of the beam spot, in order to determine the number and (a rough estimate of) the position of vertices. These values are determined by performing the *deterministic annealing* (DA) [83] algorithm which finds the global minimum for a problem with many degrees of freedom. The problem here is analogous to minimizing free energy in statistical mechanics, with the vertex χ^2 playing the role of energy. Since the DA algorithm per se is not robust to outlier tracks, such as those from secondary vertices, the algorithm is slightly modified to achieve the required robustness.

Having clustered tracks and identified the corresponding vertices, the last step is to perform a fit to each cluster of tracks (with at least two tracks) using the *adaptive vertex fitter* (AVF) [84] to determine the best estimate of the 3D position of the vertex along with the associated covariance matrix. The AVF algorithm can also provide indicators for the success of the fit. One of the indicators is a weight between 0 and 1 for each track, showing its consistency with the vertex associated to it, where weights close to 1 represent high levels of consistency whereas tracks far from the vertex are assigned smaller weights. Another indicator is the number of degrees of freedom assigned to each vertex, which is defined as:

$$n_{\text{dof}} \coloneqq -3 + 2 \sum_{i=1}^{\#\text{tracks}} w_i \tag{4.1}$$

where w_i is the weight of i^{th} track which is associated to the vertex. n_{dof} can be helpful in identifying true pp collisions as tracks originating from such collisions tend to be consistent with the vertex candidate.

The primary vertex (PV) is then defined as the vertex with the largest Σp_{T}^2 of the associated physics objects, where physics objects are anti-k_t jets, as defined in section 4.6, along with $\vec{p}_{\text{T}}^{\,\text{miss}}$ which is the negative vector sum of the jets associated to the vertex.

More information regarding vertex reconstruction in CMS can be found in [67].

4.4 Calorimeter clusters

A calorimeter clustering algorithm has been developed in CMS for various reasons such as identifying and measuring the energy of photons and neutral hadrons. The

clustering algorithm is used separately in the following subdetectors: ECAL barrel, ECAL endcaps, HCAL barrel, HCAL endcaps, and the two preshower detectors. In all cases, firstly *cluster seeds* are identified by finding a calorimeter cell with a deposited energy above a subdetector-dependent seed threshold. The seed energy must be larger than its neighbours, where the neighbours are either the 4 cells sharing one side or the 8 cells sharing at least one corner with the seed candidate, depending on the subdetector. Then, initializing a cluster with the seed cell, *topological clusters* are defined by repeatedly extending the cluster to cells with at least one common corner with the cell(s) already in the cluster, provided that the energy of the new cells are at least twice the noise level. Topological clusters can end up encompassing more than one seed.

Clusters are reconstructed with an expectation-maximization algorithm using a model which is comprised of a finite number of Gaussian functions. In this model, the distribution of the energy deposited in a topological cluster with N seeds is approximated with a sum of N Gaussian functions. The free parameters of the fit for each of the functions are the mean of the Gaussian in the $\eta-\phi$ plane and an amplitude for scaling the Gaussian function, while the widths of the Gaussian functions are fixed for each subdetector. The expectation-maximization algorithm is performed iteratively with two steps per iteration. In the first step, the fit parameters are kept constant and the sum of the N Gaussian functions is used as the initial estimate of the energy deposited in the topological cluster cells. The relative contribution of each Gaussian function to the energy estimated for each cell is measured in this step. In the second step, the fit parameters are updated using the maximum likelihood method. The energy and position of the seeds are used for initializing the Gaussian functions in the first step and the iteration is repeated until convergence. The position and energy derived from the Gaussian functions after convergence are considered as the cluster reconstructed values.

The clustering algorithm is described in more detail in [71].

4.5 Particle Flow

In CMS, the PF algorithm is employed to reconstruct individual particles, namely muons, electrons, photons, neutral and charged hadrons. The algorithm basically connects *particle elements*, i.e. the objects reconstructed solely in each subdetector such as tracks and calorimeter clusters, using a *link algorithm*. The probability that a PF candidate consists of elements from only one particle is limited by the granularity of the subdetectors and the number of to-be-resolved particles per unit solid angle. The

probability that all elements of a particle are reconstructed and included in a PF candidate is limited by the amount of material upstream of the calorimeters and muon system because a large amount of material can lead to particle trajectory kinks as well as secondary particle production. When two elements are linked, a distance between them is defined by the algorithm in order to be used as a measure to quantify the quality of the link.

A link between a track and a calorimeter cluster is established if the extrapolation of the track lies within the cluster area which is defined as the union of all its cells. This area is extended by up to 1 cell in each direction to account for various effects such as multiple scattering in low energy charged particles. The link distance is defined as the distance between the cluster position and the extrapolation of the track in the $\eta-\phi$ plane. When more than one track (cluster) is linked to a cluster (track), the link with the shortest distance is kept. Tracks and ECAL clusters are also matched to account for electrons' bremsstrahlung radiation if tangents to the GSF tracks extrapolated to the ECAL pass through the envelop of an ECAL cluster. Additionally, a dedicated conversion finder algorithm has been developed to find photons converted to e^+e^- pairs. If the tracks from e^+e^- are compatible with a photon conversion, the photon and e^+e^- are linked. Two clusters in the HCAL and ECAL or in the ECAL and preshower can be linked if the cluster in the more granular subdetector is located within the boundaries of the other cluster. Nuclear interactions of charged particles in the tracker are reconstructed by finding a secondary vertex with at least three associated tracks, of which at most one is an incoming track. Finally, a track in the tracker system and one from the muon system can also be linked to reconstruct global muons, as described in section 4.2.3.

The order of identification and reconstruction in the PF algorithm is as follows. Firstly, PF muons are reconstructed and their associated PF elements are removed from the detector, as summarized in section 4.5.1. Electrons and isolated photons are reconstructed in the next step and their associated PF elements are removed, as described in section 4.5.2. In the final step, which is described in section 4.5.3, charged and neutral hadrons along with non-isolated photons are reconstructed. After all identification and reconstruction processes are completed, the event is revisited for possible misidentification or misreconstruction.

4.5.1 Muons

In the muon identification process, a set of conditions are applied to global and tracker muons to reduce misidentification. To suppress hadrons misidentified as global muons, an isolation criterion is applied to these muons: the sum of the p_{T} of the tracks and

E_{T} of the clusters within a cone size of $\Delta R < 0.3$ with respect to the muon direction in the $\eta-\phi$ plane must be less than 10% of the muon p_{T}. More stringent criteria are required for muons inside jets. Non-isolated global muons are also collected without any isolation condition but tighter requirements are applied to reduce backgrounds mainly from high-p_{T} charged hadrons which reach the muon system or from the case where a track and a standalone muon are accidentally associated.

The muon momentum is measured using only its track in the tracker when the p_{T} of the muon is below 200 GeV. For higher p_{T} muons, different track fits (e.g. tracker only, global, tracker and first muon detector plane, etc.) are compared and the fit with the lowest χ^2 is chosen for the momentum calculation. Some charged hadron candidates may later be re-identified as muon candidates if they do not satisfy specific conditions.

4.5.2 Electrons and isolated photons

Electrons and isolated photons manifest themselves in the detector with similar properties as they can be produced from each other through bremsstrahlung and pair production. Therefore, they are reconstructed in the same step of the PF algorithm by combining information from the tracker and calorimeter. Track-based electrons are seeded by a GSF track associated to an ECAL cluster with fewer than three additional linked GSF tracks. Photons, on the other hand, are seeded by an ECAL supercluster (a group of clusters, broad in ϕ and narrow in η) with $E_{\mathrm{T}} > 10$ GeV and no link to a GSF track. In the case of photon candidates as well as ECAL-based electron candidates, the energy deposited in the HCAL close to the supercluster must be less than 10% of the supercluster energy. The energy of ECAL clusters are corrected to account for the energy missed in the association process.

Afterwards, the energy and direction of superclusters are considered as those of photons. The direction of GSF tracks is assumed to correspond to that of electrons, while an optimized mixture of the momentum of GSF tracks and the ECAL energy is used to represent the energy of electrons. Further requirements are applied to electrons and photons to improve their identification. The properties of GSF tracks as well as ECAL and HCAL energy deposits are incorporated into a set of boosted decision trees (BDT) trained for enhancing electron identification. In the case of photons, the HCAL to ECAL energy deposit ratio must be compatible with a photon shower. Only isolated photons are collected in this step.

4.5.3 Hadrons and non-isolated photons

The final step of particle reconstruction in the PF algorithm, after removing PF elements of previously-reconstructed particles, is to reconstruct hadrons and non-isolated

photons. These particles include charged hadrons (e.g. protons and charged mesons such as $K^{\pm}$ and $\pi^{\pm}$), neutral hadrons (e.g. K_L^0 mesons and neutrons), non-isolated photons (e.g. from π^0 decay), and, less frequently, muons from an early decay of hadrons. The strategy for seeking these particles is different for the tracker acceptance region ($|\eta| < 2.5$) and for beyond it because of the ability to find the track of charged hadrons in the acceptance region. A hadronic jet deposits on average 25%, 3%, and 22% of its energy as photons, neutral and charged hadrons, respectively, in the ECAL. Therefore, if no track is associated to an ECAL (HCAL) cluster within the tracker acceptance region, all of the deposited energy is assigned to a photon (neutral hadron), neglecting the relatively small deposition of neutral hadrons in the ECAL. Beyond the tracker acceptance region, the ECAL clusters linked to a given HCAL cluster are assumed to originate from an unknown-charge hadron. In this region, ECAL clusters with no links to an HCAL cluster are labeled as photon candidates. Finally, any remaining HCAL clusters in the detector are linked to one or more of the tracks not already linked to any other HCAL cluster. These tracks may be subsequently linked to some of the remaining ECAL clusters. The energy of ECAL and HCAL clusters associated to a neutral or charged hadron are corrected depending on the clusters' energy and pseudorapidity. The energy of the hadron is then computed by combining the corrected ECAL and HCAL clusters.

4.6 Jets

A cluster of collimated hadrons produced from hadronization of quarks and gluons is called a jet. Reconstructing jets is needed for measuring the kinematics of the initial quarks or gluons. Jets are reconstructed in CMS using the *anti-k_t* [85] algorithm which is implemented in the FastJet [86] C++ package. This algorithm defines a distance between PF objects i and j as

$$d_{ij} \coloneqq min\left(\frac{1}{{p_T}_i^2}, \frac{1}{{p_T}_j^2}\right) \cdot \frac{\Delta R_{ij}^2}{R^2} \tag{4.2}$$

and the distance between PF object i and the beam line is defined as

$$d_{iB} \coloneqq \frac{1}{{p_T}_i^2} \tag{4.3}$$

where $\Delta R_{ij}^2 = (\phi_i - \phi_j)^2 + (\eta_i - \eta_j)^2$ and $R = 0.4$ is the distance parameter, loosely referred to as the jet cone size. The distances are sorted and added into a list. Then:

- if the shortest distance is one of the d_{ij}: objects i and j are combined into a new

object by summing their four-momenta and the list is updated.

- if the shortest distance is one of the d_{iB}: object i is labeled as a jet and removed from the list.

This procedure is repeated until no objects remain.

b-jets

In the Higgs $\mathcal{CP}$ analysis, identifying and vetoing jets originating from b-quarks (b-jets) can remove a large fraction of $t\bar{t}$ background. In CMS, b-jets are identified by the deepCSV algorithm [87, 88] in which several features from these objects are fed into a deep neural network (DNN). The DNN combines low-level and high-level features such as the secondary vertex properties for exploiting the relatively long lifetime of b-quarks, the properties of the tracks reconstructed by the anti-k_t algorithm, the kinematics of the jet, and so forth.

4.7 Missing transverse momentum

Momentum conservation in the plane transverse to the LHC proton beam requires the vector sum of the p_{T} of all outgoing particles from a pp collision to cancel out. The presence of neutrinos or potential BSM particles not interacting with the CMS detector results in an imbalance in the p_{T} sum, leading to a "missing" transverse momentum known as $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$. The magnitude of the missing transverse momentum is sometimes known as missing transverse energy (MET). Measuring $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ with a sufficient resolution is crucial in the Higgs $\mathcal{CP}$ analysis as this parameter helps suppress W+jet background in the $\tau_\mu\tau_h$ final state as well as differentiate the signal ($H \to \tau^+\tau^-$) from background ($Z \to \tau^+\tau^-$).

An estimation of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is calculated in the PF algorithm and is known as PF-MET [89]:

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = -\sum_{i\in\mathrm{particles}} \vec{p}_{\mathrm{T}i} \tag{4.4}$$

where the sum runs over all PF particles reconstructed in the detector. A more precise approach for estimating $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is to consider only the particles coming from the PV. This is performed in an algorithm called *pile-up per particle identification* (PUPPI) [90] in which a weight, w_i, is assigned to each PF particle, indicating the likelihood of originating from the PV. In this case, the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ (PUPPI-MET) is defined as

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = -\sum_{i\in\mathrm{particles}} w_i\,\vec{p}_{\mathrm{T}i} \tag{4.5}$$

where w_i can get values from 0 to 1 with particles coming from the PV (pile-up) being assigned $w_i = 1$ ($w_i = 0$) in the ideal case. These weights are easier to measure for charged particles in the **central** region of the detector covered by the tracker ($|\eta| < 2.5$) using the method described in section 4.3. The challenge is to measure these weights for charged particles in the **forward** region as well as neutral particles throughout the detector.

In order to measure these weights, the following shape-variables are defined:

$$\begin{aligned} \alpha_i^F &\coloneqq \log \sum_{j \in \text{particles}} \frac{p_{\text{T}j}}{\Delta R_{ij}}, \quad \text{with } 0.02 < \Delta R_{ij} < 0.3 \\ \alpha_i^C &\coloneqq \log \sum_{j \in \text{Ch,PV}} \frac{p_{\text{T}j}}{\Delta R_{ij}}, \quad \text{with } 0.02 < \Delta R_{ij} < 0.3 \end{aligned} \tag{4.6}$$

where α_i^F (α_i^C) will be used for identifying the vertex type, PV or pile-up, of particle i in the forward (central) region of the detector. In α_i^F, j is summed over all particles in the event while in α_i^C, it is summed over charged particles in the central region originating from the PV. If particle i originates from pile-up, it generally obtains smaller values of α_i^F and α_i^C (see Fig. 4.1) because in such cases, the activity around the direction of particle i is usually uncorrelated with it. α_i^C is more powerful but not applicable to the forward region, whereas α_i^F can be applied to any direction of the detector.

The strategy for finding the weights is based on two facts:

- MC simulations have shown that charged and neutral particles have similar distributions of α_i^F and α_i^C (see Fig. 4.1).

- The PV/pile-up origin of charged particles in the central region of the detector is known.

The decision on the origin of neutral particles in the central region of the detector can therefore be made using the more powerful discriminant α_i^C, whereas for the (neutral and charged) particles in the forward region, only α_i^F can be used. The likelihood of a particle coming from the PV or pile-up is calculated by measuring its compatibility with the distribution of $\alpha_i^{C/F}$ corresponding to pile-up. For each event, $\alpha_i^{C/F}$ values are computed for all charged particles originating from pile-up in the central region and subsequently their median and RMS are calculated. For both neutral and charged particles, the median and RMS of α_i^C (α_i^F), computed for an event, are considered as estimators of the median and RMS of the true pile-up distribution of α_i^C (α_i^F). α_i^F is extrapolated to the forward region by correcting for the pile-up pseudorapidity dependence.

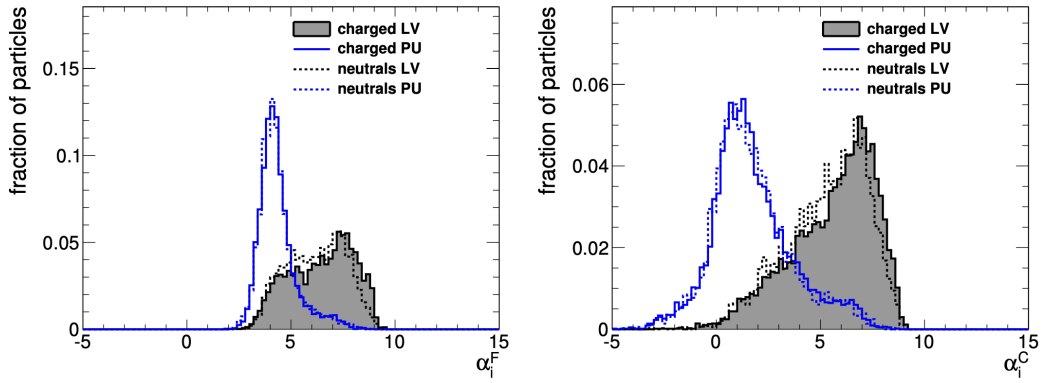

Figure 4.1: The distributions of α_i^F (left) and α_i^C (right) discriminants, over many events, for particle i from the PV (gray) and pile-up (blue). Here, LV stands for "Leading vertex" and PU stands for "pile-up". Neutral (dotted line) and charged (solid line) particles have similar distributions in all cases. Whether a particle is charged or neutral, α_i^C is used in the central region ($|\eta| < 2.5$) while α_i^F is used in the forward region in order to find if the particle originates from the PV or pile-up [90].

To measure the compatibility of a particle vertex type with pile-up, the following quantity is introduced for particle i:

$$\chi_i^2 = \Theta(\alpha_i - \bar{\alpha}_{\text{pile-up}}) \cdot \frac{(\alpha_i - \bar{\alpha}_{\text{pile-up}})^2}{\sigma^2_{\text{pile-up}}} \tag{4.7}$$

where Θ is the Heaviside step function, while $\bar{\alpha}_{\text{pile-up}}$ and $\sigma_{\text{pile-up}}$ are respectively the median and RMS of the pile-up distribution of α_i^C for the central region and α_i^F for the forward region, as described earlier. The weight used in Eq. 4.5 is then defined as

$$w_i = F_{\chi^2, NDF=1}(\chi_i^2) \tag{4.8}$$

where F_{χ^2} is the cumulative distribution function of the χ^2 distribution.

4.8 Hadronic taus

Tau leptons, with a mass of 1.78 GeV, are heavy enough to decay to hadrons and leptons, as shown in table 4.1. They have a lifetime of 2.9×10^{-13} seconds, so a 50 GeV tau lepton, which has the typical energy of taus produced at the LHC, propagates only a few millimeters before it decays and almost never reaches even the innermost subdetector. This small decay length, along with the fact that neutrinos in leptonic decays of taus (τ_l) escape the detector, makes it practically impossible to distinguish between a τ_l candidate and a light lepton (e/μ) from a prompt decay. Hence, the reconstruction

Decay mode	Dominant resonance	BR(%)
Leptonic decays (τ_l)		35.2
$e^-\bar{\nu}_e\nu_\tau$		17.8
$\mu^-\bar{\nu}_\mu\nu_\tau$		17.4
Hadronic decays (τ_h)		64.8
$h^\pm\nu_\tau$		11.5
$h^\pm\pi^0\nu_\tau$	$\rho^\pm$ (770)	25.9
$h^\pm 2\pi^0\nu_\tau$	$a_1^\pm(1260)$ ($\rightarrow\pi^0\rho^\pm$)	9.5
$3h^\pm\nu_\tau$	$a_1^\pm(1260)$ ($\rightarrow h^\pm\rho^0$)	9.8
$3h^\pm\pi^0\nu_\tau$	excited $\rho^\pm$ mesons	4.8
Other		3.3

Table 4.1: The intermediate resonances and branching ratios of the dominant tau decays [46].

procedure for τ_l candidates are the same as those introduced in sections 4.5.1 and 4.5.2. Although individual τ_l candidates cannot be discriminated from prompt leptons, the discrimination is feasible to a good extent in the τ_l candidates from a $H \rightarrow \tau^+\tau^-$ decay using the properties of both taus, such as di-tau mass, as will be discussed later in chapter 6. In this section, the method used in CMS for reconstructing hadronic decays of taus (τ_h) is discussed.

The Hadron-Plus-Strip algorithm

The reconstruction of τ_h candidates in CMS is performed using the Hadron-Plus-Strip (HPS) [91] algorithm which targets dominant τ_h decays, listed in Table 4.1 and shown in Fig. 4.2 and 4.3. The algorithm starts with an anti-k_t jet as a seed and all PF candidates within an $\eta-\phi$ cone size of $\Delta R = 0.5$ around the jet axis are considered for the next steps. The PF charged hadrons in the cone are considered as true $h^\pm$ particles. The PF γ/e candidates in the cone are grouped into what is called a "strip" and the strip is considered as coming from a true π^0 meson decay. π^0 mesons almost always decay to two photons, after which, given the large amount of material in the tracker, the photons may convert to e^+e^- which in turn bend in the CMS magnetic field and could produce further photons through bremsstrahlung radiation. A strip is expected to collect these γ/e and group them into a single object which represents a π^0 meson.

Jets are required to have $p_\mathrm{T} > 14$ GeV and $|\eta| < 2.5$. To ensure sufficient track quality and a minimal number of hits in the tracker, all charged particles in the jets must have a minimum p_T of 0.5 GeV. Considering the finite flight length of tau leptons, the transverse impact parameter requirement with respect to the PV is relaxed to $d_{xy} < 0.1$ cm.

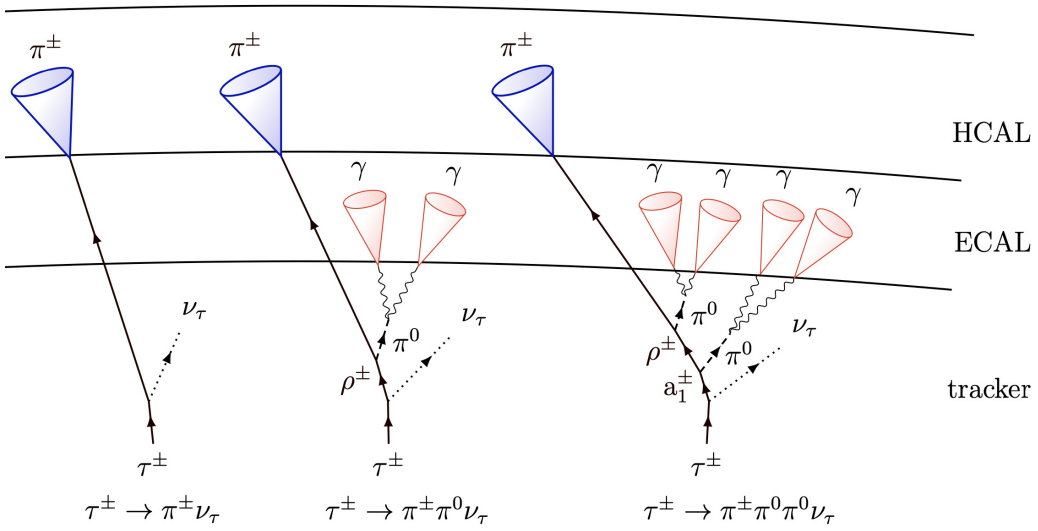

Figure 4.2: Schematic of the main τ_h decays with one charged pion in the final state. Charged hadrons in τ_h decays are predominantly charged pions. The original LaTeX script for producing such figures is taken from [92].

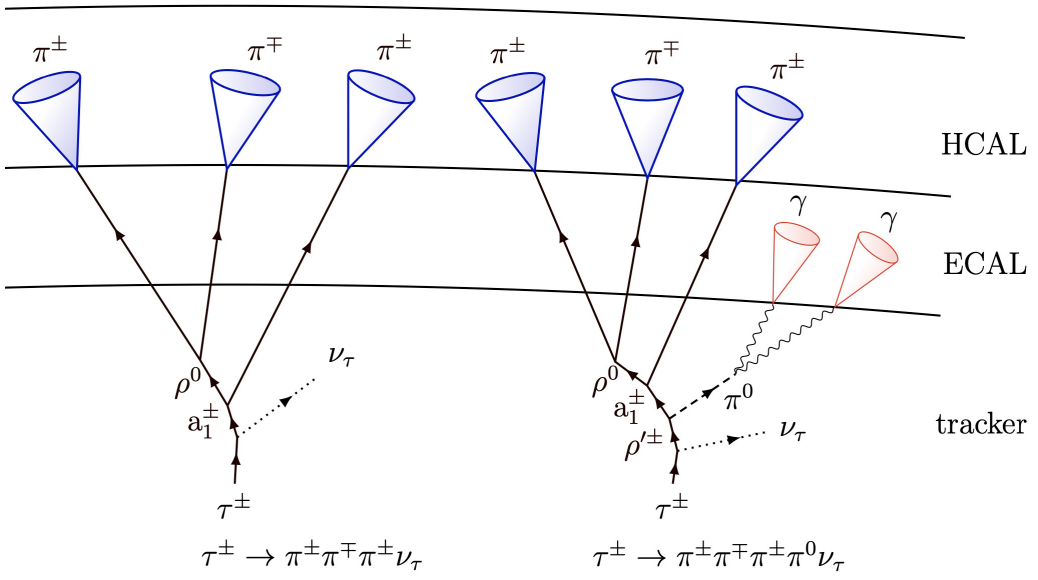

Figure 4.3: Schematic of the main τ_h decays with three charged pions in the final state. Charged hadrons in τ_h decays are predominantly charged pions. The original LaTeX script for producing such figures is taken from [92].

Strips are reconstructed in the following procedure:

1. The highest-p_{T} photon or electron (γ/e) not yet included in any strips is taken as a seed for a strip. The $\eta-\phi$ position and p_{T} of the strip is defined as those of the γ/e.

2. The second highest-p_{T} γ/e is selected and included in the strip if it is within

$$\begin{aligned} \Delta\eta &= f(p_{\mathrm{T}}^{\gamma/e}) + f(p_{\mathrm{T}}^{strip}) \quad \text{and} \\ \Delta\phi &= g(p_{\mathrm{T}}^{\gamma/e}) + g(p_{\mathrm{T}}^{strip}) \end{aligned} \tag{4.9}$$

   with respect to the strip position, where

$$\begin{aligned} f(p_{\mathrm{T}}) &= 0.20\, p_{\mathrm{T}}^{-0.66} \quad \text{and} \\ g(p_{\mathrm{T}}) &= 0.35\, p_{\mathrm{T}}^{-0.71} \end{aligned} \tag{4.10}$$

   functions, with p_{T} in units of GeV, are derived by fitting to the simulated events such that on average 95% of the γ/e from π^0 decays are included in the strip (see Fig. 4.4). $\Delta\eta$ and $\Delta\phi$ are bounded in the range (0.05-0.15) and (0.05-0.3), respectively.

3. The strip position is updated to the p_{T}-weighted sum of the position of all γ/e candidates in the strip. The p_{T} of the strip is set to the vector sum of the p_{T} of the γ/e candidates.

4. The strip reconstruction ends when no remaining γ/e candidates are close enough to the strip. The procedure continues by reconstructing a new strip.

Strips with $p_{\mathrm{T}}^{strip} < 2.5$ GeV are not considered as π^0 mesons. The τ_h candidate is not considered as a true τ_h decay if the direction of any π^0 or $h^{\pm}$ candidates lie outside a signal cone (R_{sig}) around the τ_h direction. For 30 GeV $< p_{\mathrm{T}}^{\tau_h} <$ 60 GeV, $R_{sig} = 3.0/p_{\mathrm{T}}^{\tau_h}$ in which $p_{\mathrm{T}}^{\tau_h}$ is in units of GeV. For $p_{\mathrm{T}}^{\tau_h}$ outside this range, the boundary values are assigned to R_{sig}. Although the direction of each strip, as a π^0 candidate, must be within the signal cone, part of the strip can lie outside the cone.

Afterwards, the HPS algorithm assigns decay modes to τ_h candidates by comparing the number of reconstructed $h^{\pm}$ and π^0 candidates with those expected from the τ_h decay modes shown in table 4.1 and by also requiring conditions on the invariant mass and charge of τ_h candidates to ensure that the reconstructed objects are compatible with τ_h decays. Table 4.2 summarizes the reconstructed decay modes and the associated τ_h mass conditions. Since neutrinos from taus are not practically detectable in CMS, τ_h mass is defined as the invariant mass of hadronic tau decay products excluding

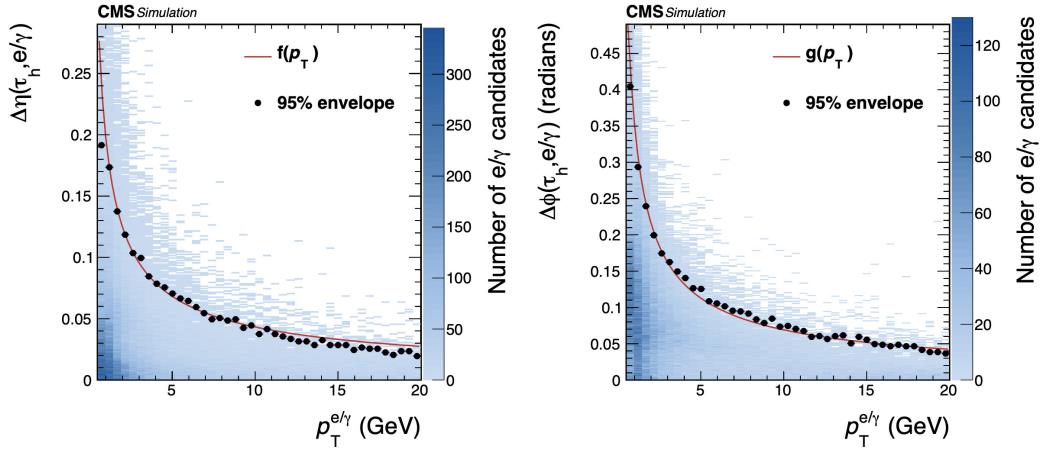

Figure 4.4: $\Delta\eta$ (left) and $\Delta\phi$ (right) between the τ_h direction and the produced γ/e as a function of $p_{\mathrm{T}}^{\gamma/e}$ plotted for a large number of τ_h decays. The $f(p_{\mathrm{T}})$ and $g(p_{\mathrm{T}})$ functions are fits to the 95% envelope of points using a function with the analytical form of $a/(p_{\mathrm{T}})^b$ [91].

neutrinos, a.k.a. "visible mass". The mass condition of each τ_h decay mode covers the mass of the intermediate resonance through which the decay occurs. A small fraction of the reconstructed decay modes include two $h^{\pm}$ candidates, which usually means that τ_h decays with three $h^{\pm}$ in the final state are not fully reconstructed and one of the three $h^{\pm}$ is "missed", for instance, because of misidentification.

The DeepTau algorithm

After reconstructing τ_h with the HPS algorithm, a deep convolutional neutral network (CNN) algorithm, called *DeepTau* [93], is used to suppress the objects which are misidentified as τ_h. These objects could be jets, muons, or electrons; therefore the output of DeepTau consists of four scores showing the probability that an object is a genuine τ_h decay or one of the three contaminating objects.

The architecture of DeepTau is depicted in Fig. 4.5. DeepTau exploits high-level and low-level features from τ_h candidates by incorporating both classes of features into several DNNs. High-level features include general event properties such as an estimation of the pile-up density in the event as well as the τ_h candidate properties such as its four-momentum, its compatibility with the PV, and the number of neutral and charged particles used in τ_h reconstruction. These features are pre-processed in a DNN, changing the number of features from 47 to 57. In order to construct low-level features, an inner 11×11 grid with a cell size of 0.02×0.02 along with an outer 21×21 grid with a cell size of 0.05×0.05, both centred on the τ_h axis, are defined to cover the τ_h signal and isolation cone, respectively (see Fig. 4.6). In each cell, a total of

No. of $h^{\pm}$ and π^0	m_{τ_h} mass window (lower, upper) [GeV]	Upper limit constraint [GeV]
$1h^{\pm}$ plus 0 π^0	$(0.0, 1.0)$	
$1h^{\pm}$ plus 1 π^0	$(0.3 - \Delta m_{\tau_h}, 1.3\sqrt{p_{\mathrm{T}}^{\tau_h}/100} + \Delta m_{\tau_h})$	min=1.3 max=4.2
$1h^{\pm}$ plus 2 π^0	$(0.4 - \Delta m_{\tau_h}, 1.2\sqrt{p_{\mathrm{T}}^{\tau_h}/100} + \Delta m_{\tau_h})$	min=1.2 max=4.0
$2h^{\pm}$ plus 0 π^0	$(0.0, 1.2)$	
$2h^{\pm}$ plus 1 π^0	$(0.0, 1.2\sqrt{p_{\mathrm{T}}^{\tau_h}/100} + \Delta m_{\tau_h})$	min=1.2 max=4.0
$3h^{\pm}$ plus 0 π^0	$(0.8, 1.4)$	
$3h^{\pm}$ plus 1 π^0	$(0.9 - \Delta m_{\tau_h}, 1.6 + \Delta m_{\tau_h})$	

Table 4.2: The mass conditions required for reconstructing a τ_h decay. The Δm_{τ_h}, defined in [91], is an estimation of the change in the τ_h mass brought by including the strip constituents.

188 features, such as the HCAL/ECAL energy deposit ratio, the quality of tracks, the number of hits in tracks and the probability of originating from pile-up, are defined for each particle type found in the cell. If more than one particle of the same type are found in a cell, only the one with the highest-p_{T} is considered. The features in the inner and outer cells undergo similar operations. The number of features per cell are reduced to 64 by being pre-processed using four DNNs. The cells corresponding to the same feature are then transformed into a single cell using a deep CNN, resulting in a total of 64 numbers, each representing a feature. High- and low-level features are subsequently fed into a final DNN with four outputs showing the score for an object to be an electron, muon, τ_h, or jet.

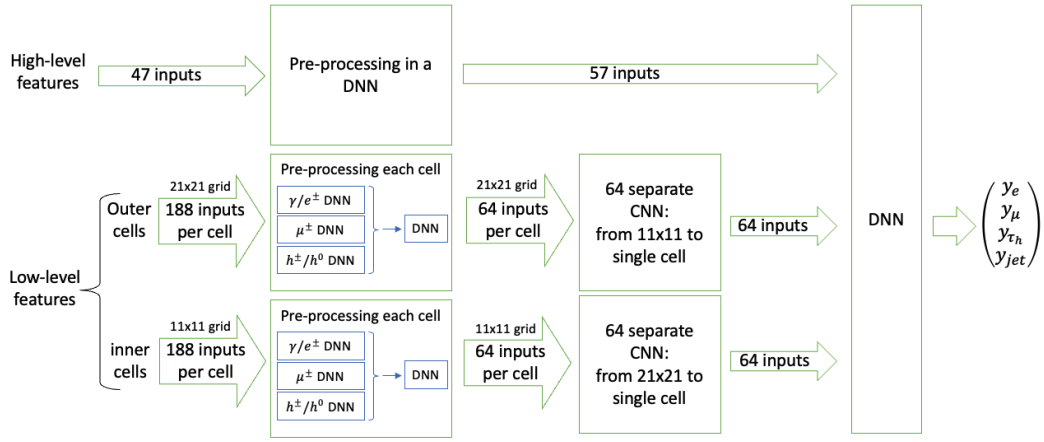

Figure 4.5: The DeepTau architecture.

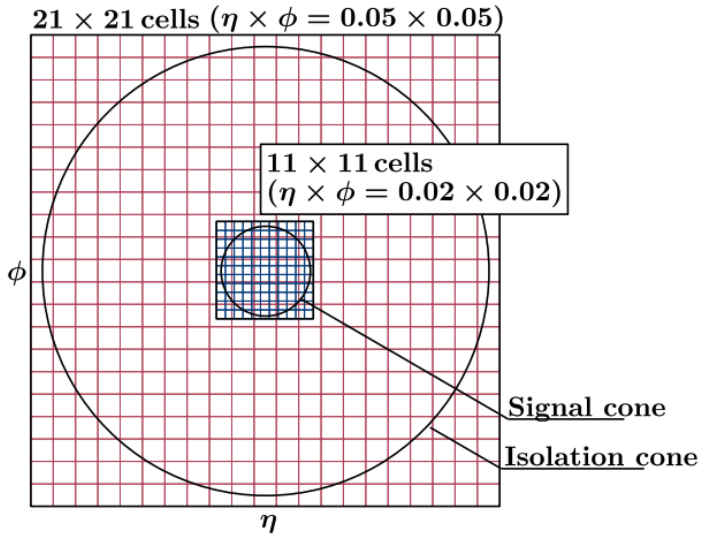

Figure 4.6: Inner (blue) and outer (red) grids defined in $\eta-\phi$ coordinates covering signal and isolation cone, respectively, to extract low-level features from a τ_h decay [93].

The performance of DeepTau has been evaluated and compared with the previous algorithm for discriminating τ_h candidates against background. The previous algorithm is explained, after introducing the HPS algorithm, in [91]. The comparison is performed by measuring the yield of genuine τ_h candidates and background in the invariant mass distribution of the four-vector sum of τ_μ and τ_h candidates from a $Z \to \tau^+\tau^-$ decay. The comparison shows that using DeepTau increases the yield of genuine τ_h by about 20% and, at the same time, decreases the background yield by about 23%. The event selection and background modeling is explained in [93].

Chapter 5

MVA-based hadronic tau decay mode identification

5.1 Motivation

In order to achieve optimal sensitivity in measuring the $\mathcal{CP}$ properties of the Higgs-tau Yukawa coupling using tau decay planes, one needs to correctly identify τ_h decay modes to a great extent, where τ_h symbol represents hadronic decays of tau leptons. This is because the sensitivity of this measurement varies for different τ_h decay modes for theoretical and experimental reasons.

From the theoretical perspective, for instance, the τ-spin analyzing power of $\pi^{\pm}$ is constant and maximal in $\tau^{\pm} \to \pi^{\pm}$ decay[1] while it is energy-dependent in $\tau^{\pm} \to \pi^{\pm}\pi^0$ decay. Therefore, the even-odd $\mathcal{CP}$ separation is different in these decays even at generator-level (GEN-level) [47, 48, 50, 94].

Experimental issues can also affect $\mathcal{CP}$ sensitivity. As will be discussed in chapter 6, different decay-plane methods are used for different decay modes to achieve maximum sensitivity. (See chapter 2 for more information on decay-plane methods.) Each method has its own sources of uncertainty associated with decay-plane reconstruction, for example impact parameter resolution in the impact parameter method and π^0 direction resolution in the neutral-pion method. Since the sources are different, the $\mathcal{CP}$ sensitivities are not necessarily the same because one source can decrease sensitivity more than another.

Additionally, as a decay-plane method is not generally applicable to all decay modes, a decay mode misidentification (due to detector effects such as detector malfunction, reconstruction failures, etc.) dilutes the sensitivity. For instance, applying the neutral-

[1] ν_τ is omitted for the sake of brevity.

pion method to a $\tau^{\pm} \to \pi^{\pm}$ decay misidentified as $\tau^{\pm} \to \pi^{\pm}\pi^0$ constructs a plane using the $\pi^{\pm}$ and a fake π^0. Therefore, it is important to correctly identify decay modes to avoid diluting $\mathcal{CP}$ sensitivity.

The main τ_h decay modes, their branching ratio (BR), and intermediate resonances was previously shown in table 4.1. τ_h decays include hadrons and a tau neutrino. Approximately all of the hadrons are charged pions while they also include a small fraction charged kaons. Since the CMS detector cannot distinguish between them, all $h^{\pm}$ are assumed to be $\pi^{\pm}$ in this thesis. When discussing decay modes, ν_{τ} is usually omitted for the sake of brevity.

As described in section 4.8, CMS employs the HPS algorithm to identify, reconstruct and assign a decay mode to τ_h. However, decay mode identification is not the primary task of this algorithm and hence the HPS algorithm is not optimized for this purpose. The algorithm is highly effective in identifying the number of charged hadrons in a τ_h decay. Nonetheless, since the p_T-dependent strip size is intentionally enlarged in the last released version of the algorithm [91] to ensure photons and electrons from π^0 decay are well encompassed, multiple π^0 mesons could be reconstructed in a single strip. For instance, $\tau^{\pm} \to \pi^{\pm}2\pi^0$ decays are (almost) always reconstructed as $\tau^{\pm} \to \pi^{\pm}\pi^0$ or $\tau^{\pm} \to \pi^{\pm}$.

The confusion matrices shown in Fig. 5.1 illustrate the performance of the HPS algorithm in τ_h decay mode finding. In general, whenever the number of π^0 in the final state is not correctly estimated by the HPS, it is under- (rather than over-) estimated. This can be seen from the confusion matrix normalized by true label, in which major off-diagonal values are **under** the main diagonal (ignoring the "other" category which is in an arbitrary position). The HPS algorithm is, however, very effective in identifying $\pi^{\pm}$ and hence the confusion matrix is almost block-diagonal, with the two blocks being decays with $1\pi^{\pm}$ and $3\pi^{\pm}$ in the final state. As a result, the main task of any new algorithm for improving decay mode identification is to concentrate on π^0 identification efficiency.

To improve τ_h decay mode identification, two independent boosted decision tree (BDT) algorithms were developed using the XGBoost library [95] for identifying, respectively, *1-* and *3-charged-prong* decays[2]. One could develop a single BDT for all decay modes; however, with two BDTs, each can have dedicated features for exploiting π^0 identification. One of the BDTs is trained to identify $\tau^{\pm} \to \pi^{\pm}$, $\tau^{\pm} \to \pi^{\pm}\pi^0$ and $\tau^{\pm} \to \pi^{\pm}2\pi^0$ decays, whereas the other searches for $\tau^{\pm} \to 3\pi^{\pm}$ and $\tau^{\pm} \to 3\pi^{\pm}\pi^0$ decays.

This chapter is dedicated to a detailed description of these BDTs. A summary of

[2] "Charged-prong" refers to charged particle candidates.

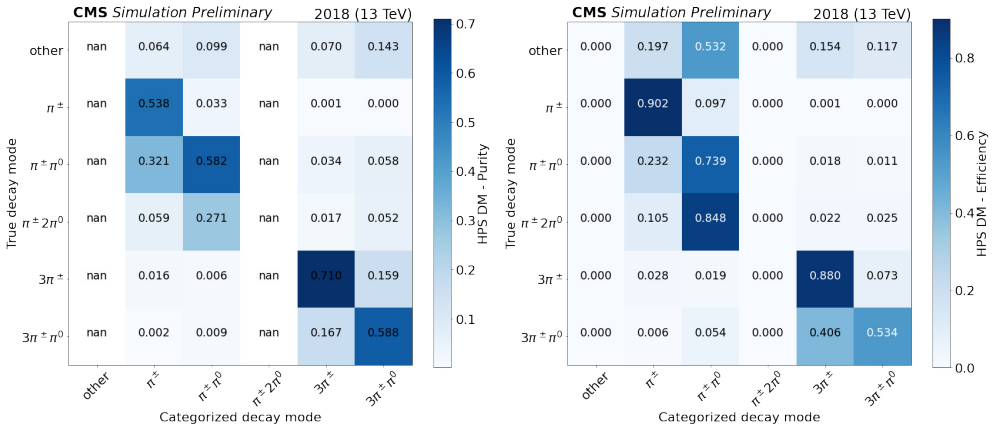

Figure 5.1: The confusion matrix of τ_h decays reconstructed by the HPS algorithm, normalized to the categorized (left) and true (right) decay modes. The former (latter) is sometimes known as the purity (efficiency) matrix. τ_h candidates are collected from $H \to \tau^+\tau^-$ decays in $\tau_\mu\tau_h$ final state where the Higgs boson is produced via ggH and VBF processes. The simulated events are based on experimental conditions of the 2018 data-taking period.

this chapter is published in [1].

5.2 Boosted Decision Tree classifiers

A decision tree classifier is a learning algorithm with a tree-like structure which is utilized for classification tasks. This algorithm splits input data recursively using input features and subsequently assigns a class to each data point. A BDT classifier is an iterative algorithm which combines several (weak) decision trees into a strong classifier. In the first iteration, the input data is classified using a decision tree. In the second iteration, a new decision tree is used to classify the data points but the data points which were misclassified in the first iteration are given larger weights. With this reweighting, the misclassified data points are more likely to be assigned their correct classes. In other words, the second decision tree attempts to correct the mistakes of the first one. In the next iteration, a third decision tree is used with the goal of correcting the mistakes of the second decision tree. This process continues until a user-defined stopping condition is reached. Each data point is assigned a class based on a weighted sum of the classes assigned by each decision tree. In this sum, decision trees with smaller errors (i.e. fewer misclassified data points) are given larger weights.

5.3 MC samples and event selection

MC samples for this study are produced with POWHEG and are based on the experimental conditions of the 2017 data-taking period. The samples include the two dominant Standard Model Higgs boson production processes: gluon-gluon and vector boson fusion. The produced Higgs bosons subsequently decay to two tau leptons. Among decay channels, $\tau_\mu\tau_h$ and $\tau_h\tau_h$ are selected after which τ_h candidates, which are reconstructed by the HPS algorithm, are extracted and merged independent of their production process. The candidates are then matched to GEN-level taus, and an extra matching is performed to ensure that all GEN-level taus undergo hadronic decay.

Further conditions are applied on τ_h candidates to be in line with the selections in the Higgs $\mathcal{CP}$ analysis. The conditions include some pre-selections to remove misidentified objects: a *VLoose* (very loose) "MVA isolation" cut against jets and electrons in both decay channels and *Loose* (*Tight*) cut against muons in $\tau_h\tau_h$ ($\tau_\mu\tau_h$). MVA isolation is a multivariate analysis method for improving τ_h identification against background and it is currently superseded by DeepTau which was introduced in section 4.8.

The τ_h candidates in $\tau_\mu\tau_h$ and $\tau_h\tau_h$ decay channels are required to have p_T above 20 and 40 GeV, respectively. After kinematic selections, the average p_T of τ_h candidates is 52 GeV. It should be noted that although no $\tau_e\tau_h$ sample is used in the training, the p_T range of τ_h in this decay mode is already covered in the other channels.

5.4 Training

In this section, the details of training the BDTs are described. Some terminology will be introduced and used hereafter. Each BDT is simply called an *MVA* (for multivariate analysis)[3]. The decay mode classified by the MVA is called the *MVA decay mode*, as opposed to the *HPS decay mode* which is that classified by the HPS algorithm. The scores associated to the output of the MVAs are called the *MVA scores*. The MVA decay mode, as we will see, is constructed by labeling each reconstructed τ_h candidate with the decay mode having the highest MVA score.

5.4.1 1-charged-prong decay

After the kinematic selections mentioned in section 5.3, τ_h candidates categorized as the following decay modes by the HPS algorithm are collected: $\pi^\pm$, $\pi^\pm\pi^0$ and $\pi^\pm 2\pi^0$ (the last one is empty as mentioned before, but included just for completeness). The training

[3]Not to be confused with "MVA isolation" mentioned in section 5.3 which is developed for τ_h identification against jets, electrons and muons.

is performed with a Multi-class BDT with four classes in the output to represent the three above-mentioned 1-charged-prong decays as well as one class, named "other", to collect a small fraction of events which are not any of the mentioned decays (e.g. a misidentified 3- or 5-charged-prong decay).

The XGBoost parameters in the training are the following:

- "objective": "multi:softprob"
- "max_depth": 5
- "subsample": 0.9
- "learning_rate": 0.05
- "n_estimators": 2000

The *multi:softprob* objective provides a probability for each of the outputs, which is equivalent to the MVA score defined above. The *max_depth* shows the maximum depth of the decision trees while the *subsample* determines the fraction of the training data set which are randomly selected and used in the decision trees. The values of these two parameters are chosen slightly less than the default values to prevent the MVA to be overtrained. Similarly, a value smaller than the default is selected for the *learning_rate* as higher values deteriorated the performance. The training stops if it iterates *n_estimators*=2000 times or the "mlogloss" score does not improve in 100 sequential iterations.

Half of the sample (every 2^{nd}) are selected as a validation set. In the training, the events are weighted by the Standard Model Higgs $\mathcal{CP}$ which is extracted using the TauSpinner package [96]. This, however, does not bias the training towards hadronic taus originating from the Standard Model Higgs $\mathcal{CP}$ because the $\mathcal{CP}$ structure of the Higgs affects the correlation between the spins of the tau pair while the MVA is trained only on single taus.

The main differences between the decays are the number of π^0 mesons in the final state and the intermediate resonances. Both differences are exploited to optimize the classification performance, as explained below.

A good strategy to identify the number of π^0 mesons in the final state is to exploit the properties of the HPS strips as they are designed for π^0 reconstruction. A procedure is introduced for selecting one of the strips in each decay in a consistent way and incorporating its properties as features into the MVA. It should be noted that it often happens that although a strip is reconstructed, it is not considered as a π^0 decay by the algorithm because it has failed to meet at least one of the algorithm requirements, e.g. p_{T} threshold.

To select one strip per τ_h decay, a strip-like object $\boldsymbol{\mathcal{S}}$ is defined in the following way:

- Select the leading strip in the signal cone (if it exists), otherwise the leading strip in the isolation cone (if it exists). If no strips exist, values outside physical range are assigned to all of its related features.

- All photons or electrons (γ/e) with $p_{\mathrm{T}} < 1$ GeV are removed from $\mathcal{S}$. This choice is made to remove γ/e with poor position resolution or those likely to originate from pile-up.

  In both of the MVAs introduced in this chapter, photons and electrons are treated as the same object.

- The four-vector (p_{T}, η, ϕ, E) of $\mathcal{S}$ is defined as follows: E is the sum of the energy of all γ/e candidates in $\mathcal{S}$, η and ϕ are the position coordinates of the leading γ/e, and the transverse momentum (p_{T}) is derived using E, η, and the π^0 mass. Other studies [2] show that using the leading γ/e candidate leads to a more precise π^0 position resolution while using all γ/e candidates in a strip can estimate the π^0 energy with a more precise resolution.

The decays also differ by their intermediate resonances; $\tau^{\pm} \to \pi^{\pm}$ has no intermediate resonance whereas the other two decay modes mainly occur as $\tau^{\pm} \to \rho^{\pm} \to \pi^{\pm}\pi^0$ and $\tau^{\pm} \to a_1^{\pm} \to \rho^{\pm}\pi^0 \to \pi^{\pm}2\pi^0$. To exploit these differences, several mass variables are introduced which are defined using the charged hadron, the $\mathcal{S}$, and the two leading γ/e candidates in the $\mathcal{S}$.

The features used in this MVA are listed below. The symbol *m(obj1, obj2,...)* is used to show the invariant mass of the four-vector sum of objects in parentheses.

- **Mass variables:** $m(\pi^{\pm}, \mathcal{S})$, $m(\pi^{\pm}, \gamma/e^{lead})$, $m(\pi^{\pm}, \gamma/e^{sublead})$, $m(\pi^{\pm}, \gamma/e^{lead}, \gamma/e^{sublead})$, $m(\mathcal{S})$, $m(\gamma/e^{lead}, \gamma/e^{sublead})$

- $\boldsymbol{\tau_h}$ **properties:** η and p_{T}.

- **Relative energies:** $E_{\gamma/e^{lead}}$, $E_{\gamma/e^{sublead}}$, and $E_{\pi^{\pm}}$, each divided by E_{τ_h}.

- $\boldsymbol{\mathcal{S}}$ **properties:** p_{T} and energy.

- $\boldsymbol{\pi^{\pm}}$ **properties:** Energy.

- **HPS decay mode**

- **$\langle \Delta R^2 \rangle$-related variables:** $\langle \Delta R^2 \rangle$ and $\langle \Delta R^2 \rangle \times {E_{\tau_h}}^2$, in which $\langle \Delta R^2 \rangle = \frac{\Sigma_i \Delta R_i^2 p_{T_i}^2}{\Sigma_i p_{T_i}^2}$ and i sums over $\pi^\pm$ as well as all γ/e in $\mathcal{S}$. ΔR is measured with respect to the τ_h direction.
- **Angular quantities between $\pi^\pm$ and $\mathcal{S}$:** $\Delta\eta$, $\Delta\phi$, $\Delta\eta \times E_{\tau_h}$, and $\Delta\phi \times E_{\tau_h}$.
- **Angular quantities between γ/e^{lead} and $\gamma/e^{sublead}$:** $\Delta\eta$, $\Delta\eta \times E_{\tau_h}$, and $\Delta R \times E_{\tau_h}$.

After the training is performed, the importance of the features is computed using the "plot_importance" function from the XGBoost library. Two typical feature-importance metrics for BDTs are *Gain* and *Weight*. The former is defined as the average gain achieved when a feature is used for splitting while the latter shows the number of times a features is used for splitting. The HPS decay mode, the p_T of $\mathcal{S}$, and $E_{\pi^\pm}/E_{\tau_h}$ are the features with the highest *Gain*, whereas the features with the highest *Weight* are the η of τ_h, m($\pi^\pm$, $\mathcal{S}$), and $E_{\gamma/e^{lead}}/E_{\tau_h}$. The importance of all features is shown in the Appendix A.

Figure 5.2 displays the normalized-by-integral distributions of 1-charged-prong MVA scores, each split by generator-level decay information. (Being negligible in size, the "other" category is not shown in some of the figures in this chapter) As expected, the MVA has pushed the signal in each MVA decay mode to high MVA score values. In the $\pi^\pm$ MVA score, the signal is almost perfectly separated from the other two backgrounds (note the log-scale). In the $\pi^\pm\pi^0$ MVA score, there is a strong differentiation power between true $\pi^\pm\pi^0$ and $\pi^\pm$ while $\pi^\pm\pi^0$ and $\pi^\pm 2\pi^0$ are scored more similarly by the MVA. In the $\pi^\pm 2\pi^0$ MVA score, similarly, $\pi^\pm$ is well identified as background while less separation is seen between the other two decay modes. Most of the signal is below 0.8 with a flat distribution meaning that the MVA is less certain about the signal in this MVA decay mode.

As mentioned earlier in this section, MVA decay modes are constructed based on the decay mode with the highest MVA score. However, different categorizations can be performed based on needs. For example, one can collect a purer $\pi^\pm 2\pi^0$ sample by requiring the corresponding MVA score to be above 0.6.

The MVA distinguishing power can also be seen in receiver operating characteristic (ROC) curves. A ROC curve shows the fraction of signal and background one gets by changing minimum threshold on MVA scores. In fact, one can compute these fractions using the distributions in Fig. 5.2 before normalization. The final choice of a threshold is a trade-off between efficiency and purity and is therefore analysis-dependent.

Fig. 5.3 displays the ROC curves for 1-charged-prong MVA scores. The curves show that the MVA can strongly differentiate $\pi^\pm$ final state from background while the

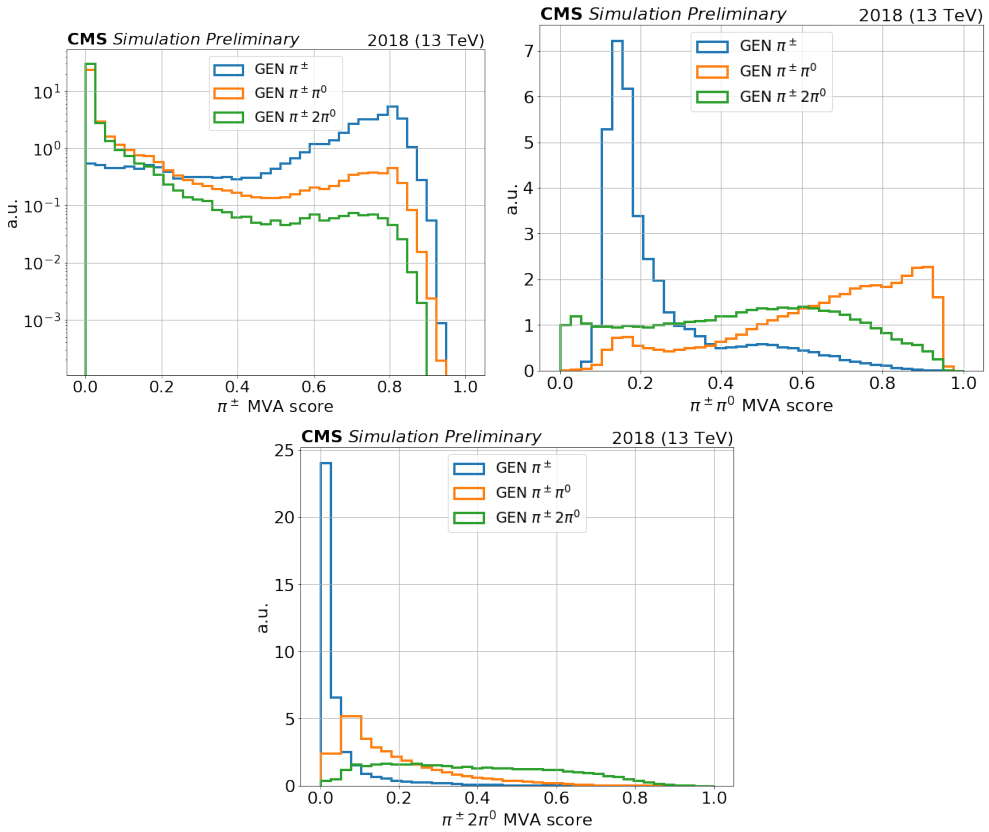

Figure 5.2: The MVA score distribution for $\pi^\pm$ (top left), $\pi^\pm\pi^0$ (top right), and $\pi^\pm 2\pi^0$ (bottom), split by generator-level decay information [1].

differentiation power is relatively less, though still high, for the other decay modes.

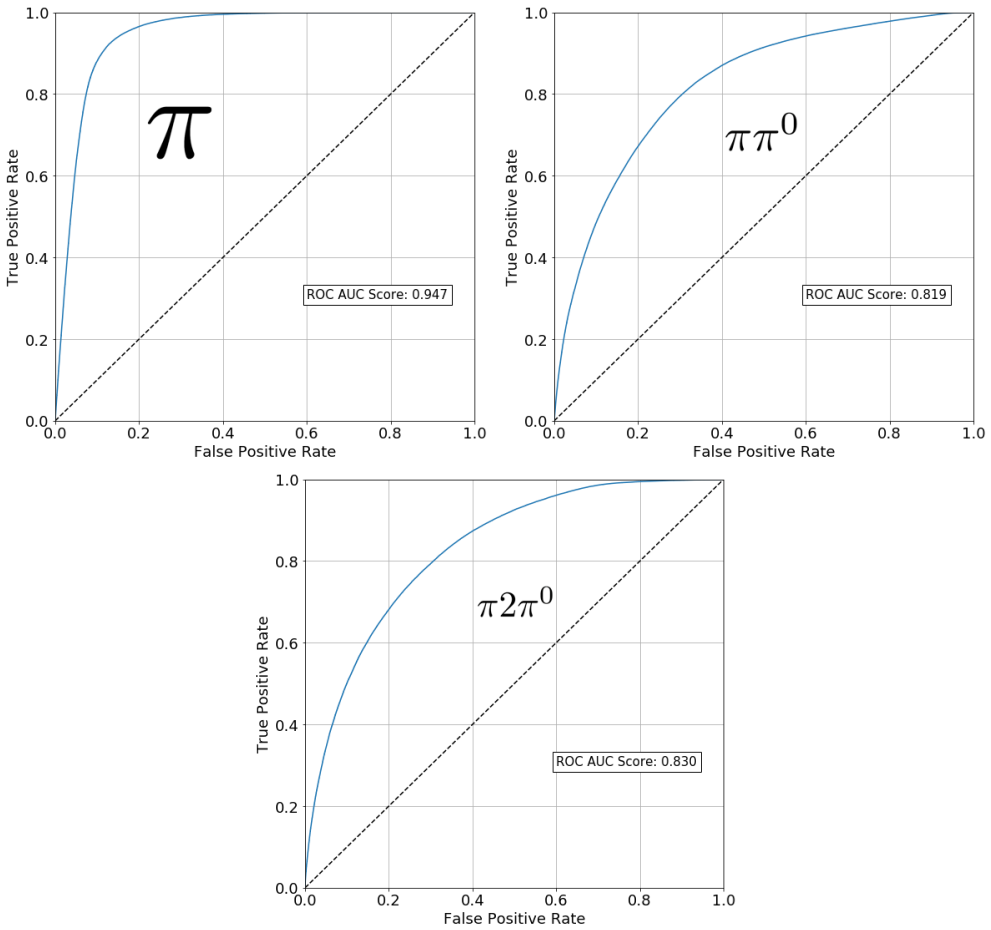

Figure 5.3: ROC curves for $\pi^{\pm}$ (top left), $\pi^{\pm}\pi^0$ (top right) and $\pi^{\pm}2\pi^0$ (bottom) reconstructed by the MVA trained on 1-charged-prong decays. ROC AUC score shows the area under the curve.

5.4.2 3-charged-prong decay

A similar MVA is developed for 3-charged-prong decays. As input to the MVA, after the kinematic selections described in section 5.3, τ_h candidates which are classified as $3\pi^{\pm}$ and $3\pi^{\pm}\pi^0$ by the HPS algorithm are collected. The MVA has three outputs corresponding to the score of the two mentioned decay modes and the last one, "other", for collecting other decays misidentified as 3-charged-prong decays.

The strategy is similar to the one used for 1-charged-prong decays (see section 5.4.1). The decay modes differ by the number of π^0 mesons in the final state and the type of intermediate resonances. $3\pi^{\pm}$ decays occur through production of a_1 mesons, i.e. $\tau^{\pm} \to a_1^{\pm} \to \rho^0\pi^{\pm} \to 3\pi^{\pm}$. In contrast, $3\pi^{\pm}\pi^0$ decays occur through multiple resonances [97,

98]. Therefore, again, adding features associated to strips as well as mass variables can help distinguish the decay modes.

As there are 3 charged pions in the above decay modes (two same-sign and one opposite-sign), one needs a consistent way to define mass variables. Since the decay $\tau^{\pm} \to a_1^{\pm} \to \rho^0\pi^{\pm} \to 3\pi^{\pm}$ includes $\rho^0 \to \pi^+\pi^-$, in an effort to reconstruct the ρ^0 mass, the invariant mass of the single opposite-sign pion summed with each of the same-sign pions are computed and considered as two different features based on the closeness of the mass to the ρ^0 mass, i.e. 0.7755 GeV. In the rest of this chapter, the opposite-sign pion is referred to as π_1 and the same-sign pion which results in $m(\pi^+, \pi^-)$ closer to (further from) the ρ^0 mass is referred to as π_2 (π_3).

The features used in the training are listed below. The $\mathcal{S}$ is defined in exactly the same way as the one defined in section 5.4.1.

- **Mass variables:** $m(\pi_1, \pi_2)$, $m(\pi_1, \pi_3)$, $m(\pi_1, \pi_2, \pi_3)$, $m(\mathcal{S})$, and $m(\gamma/e^{lead}, \gamma/e^{sublead})$.
- **HPS decay mode**
- **Absolute energy of $\boldsymbol{\pi^{\pm}}$ and $\boldsymbol{\gamma/e}$:** E_{π_1}, E_{π_2}, E_{π_3}, $E_{\gamma/e^{lead}}$, and $E_{\gamma/e^{sublead}}$.
- **Relative energies:** E_{π_1}, E_{π_2}, and E_{π_3}, each divided by E_{τ_h}. Also E_{π_1} and E_{π_2}, each divided by $E_{\pi_1} + E_{\pi_2} + E_{\pi_3}$.
- **$\boldsymbol{\mathcal{S}}$ properties:** p_{T} and energy.
- **$\boldsymbol{\tau_h}$ properties:** p_{T}, energy, and η.
- **Angular quantities between charged pions:** $\Delta\phi(\pi_i, \pi_j)$, $\Delta\eta(\pi_i, \pi_j)$, $(E_{\pi_i} + E_{\pi_j}) \times \Delta\phi(\pi_i, \pi_j)$, and $(E_{\pi_i} + E_{\pi_j}) \times \Delta\eta(\pi_i, \pi_j)$ with $(i, j) \in \{(1,2), (2,3), (1,3)\}$.
- **Angular quantities between $\boldsymbol{\tau_h}$ and $\boldsymbol{\mathcal{S}}$:** $\Delta\phi$, $\Delta\eta$, $E_{\tau_h} \times \Delta\eta$, and $E_{\tau_h} \times \Delta\phi$.
- **Angular quantities between $\boldsymbol{\gamma/e^{lead}}$ and $\boldsymbol{\gamma/e^{sublead}}$:** $\Delta\eta$, $E_{\tau_h} \times \Delta\eta$, and $E_{\tau_h} \times \Delta R$.

Among the features, HPS decay mode, the p_T of $\mathcal{S}$, and m(π_1, π_2) are ranked highest in the *Gain* metric while in the *Weight* metric, m(π_1, π_3), m(π_1, π_2), and m(π_1, π_2, π_3) are the most important features. The full feature-importance lists can be found in Appendix A.

The MVA score distributions for 3-charged-prong decays are shown in Fig. 5.4. The signal and background in each MVA score are very well separated. The associated ROC curves are displayed in Fig. 5.5. The MVA provides a strong differentiation power between $3\pi^{\pm}$ and $3\pi^{\pm}\pi^0$ final states.

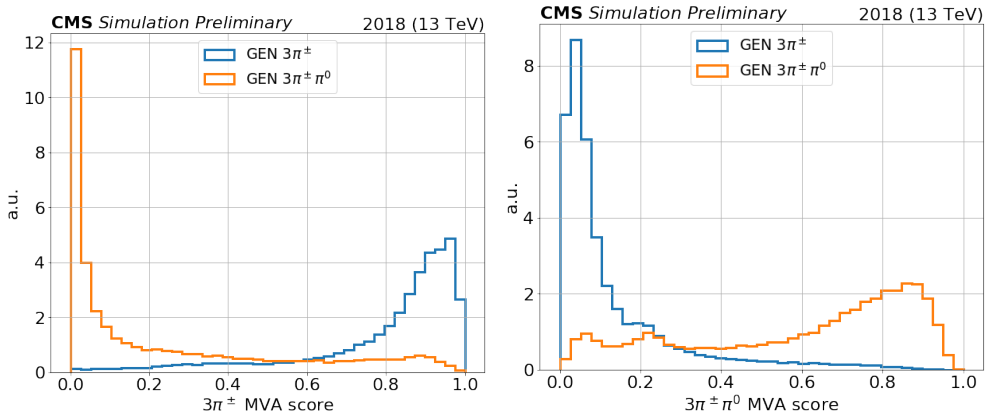

Figure 5.4: The MVA score distribution for $3\pi^{\pm}$ (left) and $3\pi^{\pm}\pi^{0}$ (right), split by generator-level decay information [1].

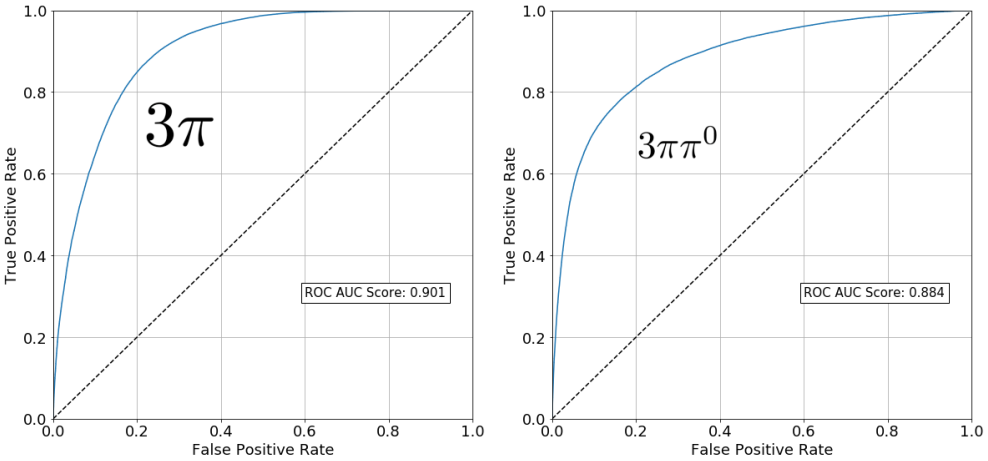

Figure 5.5: ROC curves for $3\pi^{\pm}$ (left) and $3\pi^{\pm}\pi^{0}$ (right) reconstructed by the MVA trained on 3-charged-prong decays. ROC AUC score shows the area under the curve.

5.5 Comparing MVA and HPS

In order to compare MVA and HPS decay modes, the purity and efficiency of decay mode $\tau_h \to X$ is defined as:

$$\text{purity} = \frac{\sum_{events} \text{event is categorized as and is a true } \tau_h \to X \text{ decay}}{\sum_{events} \text{event is categorized as a } \tau_h \to X \text{ decay}} \tag{5.1}$$

$$\text{efficiency} = \frac{\sum_{events} \text{event is categorized as and is a true } \tau_h \to X \text{ decay}}{\sum_{events} \text{event is a true } \tau_h \to X \text{ decay}} \tag{5.2}$$

In the above definitions, only true τ_h decays are considered, which means that no contamination from other physics processes misidentified as τ_h decays is included. Fig. 5.6 compares the performance of MVA and HPS decay categorization in terms of purity and efficiency. In 1-charged-prong decays, the purity of $\pi^\pm$ decay mode has increased by 16 percentage points but has 7 percentage points less efficiency. In $\pi^\pm\pi^0$ final state, purity and efficiency improved by 10 and 5 percentage points, respectively. While no sample for $\pi^\pm 2\pi^0$ final state is provided by the HPS algorithm, a sample with 55 percentage points purity and 39 percentage points efficiency is now available in the corresponding MVA decay mode.

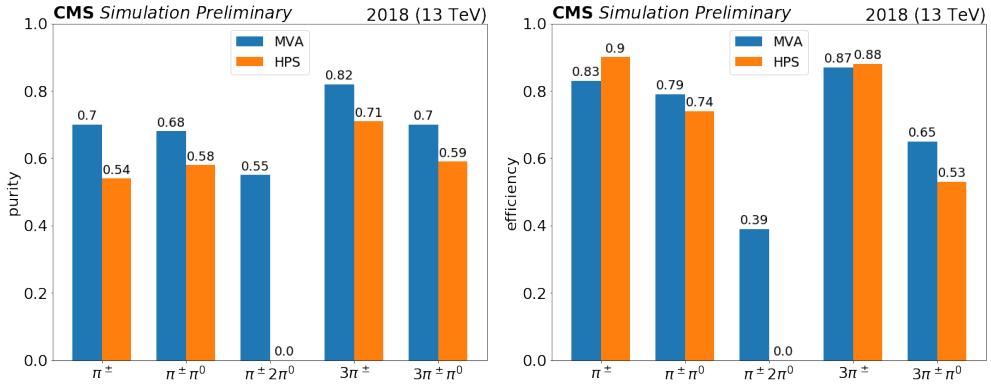

Figure 5.6: The purity (left) and efficiency (right) of different τ_h decay modes, compared between the HPS (orange) and MVA (blue) algorithms. τ_h candidates are collected from $H \to \tau^+\tau^-$ decays in $\tau_\mu\tau_h$ final state where the Higgs boson is produced via ggH and VBF processes [1].

The performance in the identification of 3-charged-prong decays is also improved. The sample provided for $3\pi^\pm$ final state is now 11 percentage points purer with almost the same efficiency. The purity and efficiency in the $3\pi^\pm\pi^0$ final state are increased by

11 and 12 percentage points, respectively.

As mentioned in section 5.3, the training is performed based on 2017 experimental conditions. The performance of the MVAs with the experimental conditions of other data-taking periods needs to be evaluated. Fig. 5.7 and 5.8 are MVA decay mode confusion matrices normalized to the categorized and true decay modes, respectively. In each figure, a comparison of the performance of the MVA in the 2016, 2017, and 2018 data-taking periods is demonstrated. The confusion matrices are very similar, which means that there is no need to train separate MVAs for each year as the 2017 one is applicable to all years.

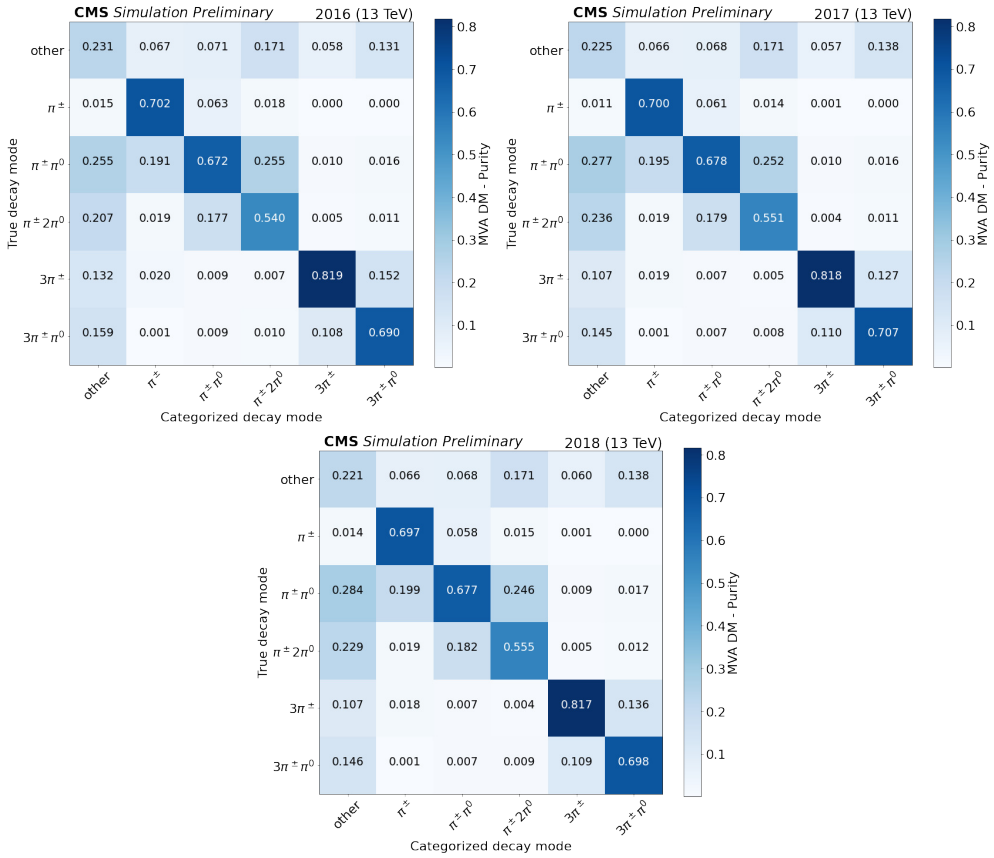

Figure 5.7: The confusion matrix of MVA decay modes, normalized to the categorized decay mode (a.k.a **purity** matrices), applied to simulated events with experimental conditions of 2016 (top left), 2017 (top right), and 2018 (bottom) data-taking period. τ_h candidates are collected from $H \to \tau^+\tau^-$ decays in $\tau_\mu\tau_h$ final state where the Higgs boson is produced via ggH and VBF processes. The performance of the MVA in all years are similar. The bottom plot can be found in [1].

The confusion matrices can provide valuable information regarding the composition of each MVA decay mode. Consider the efficiency and purity matrices for one of the

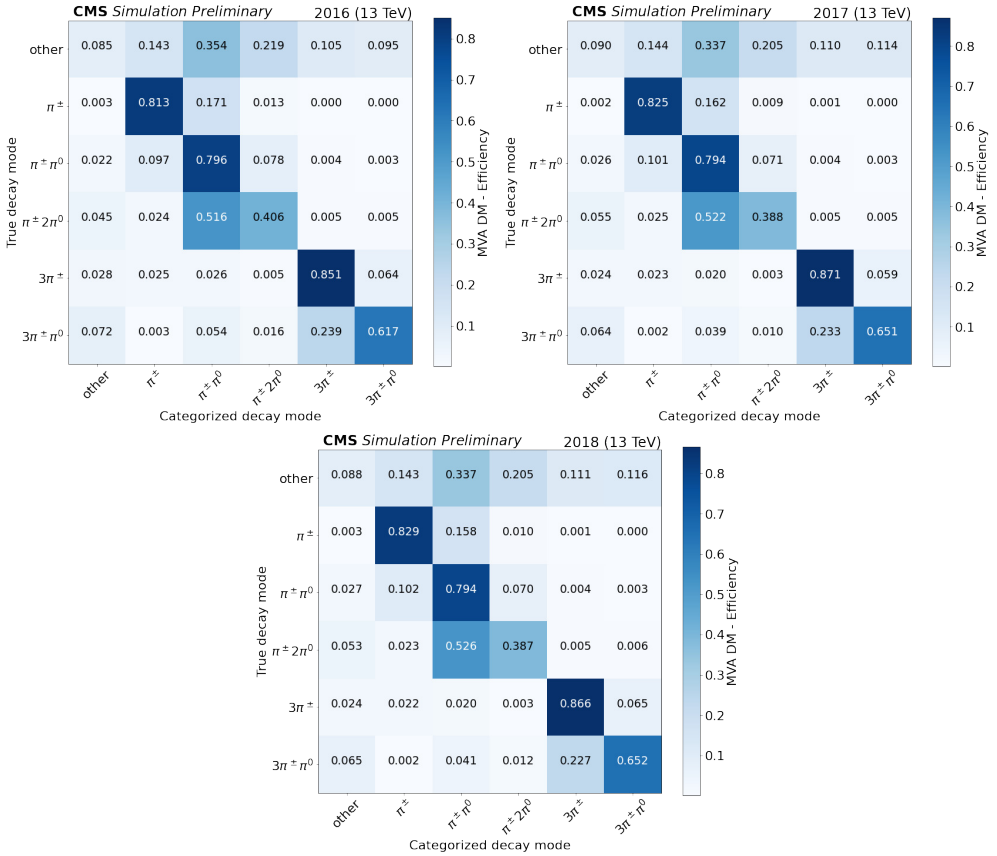

Figure 5.8: The confusion matrix of MVA decay modes, normalized to the true decay mode (a.k.a **efficiency** matrices), applied to simulated events with experimental conditions of 2016 (top left), 2017 (top right), and 2018 (bottom) data-taking period. τ_h candidates are collected from $H \to \tau^+\tau^-$ decays in $\tau_\mu\tau_h$ final state where the Higgs boson is produced via ggH and VBF processes. The performance of the MVA in all years are similar. The bottom plot can be found in [1].

years, say 2018. Comparing these matrices with the corresponding ones for the HPS, as shown in Fig. 5.1 and discussed in section 5.1, it is clear that the mixing between 1- and 3-charged-prong decays remains negligible, which is expected because separate MVAs are developed for them.

The following conclusions can be made based on the purity and efficiency matrices:

- 1-charged-prong decays:
  - **$\pi^{\pm}$:** Approximately 83% of true $\pi^{\pm}$ decays are correctly identified and virtually all the rest are misidentified as $\pi^{\pm}\pi^0$.
  - **$\pi^{\pm}\pi^0$:** True $\pi^{\pm}\pi^0$ decays are correctly identified about 79% of the time with the rest being misidentified as $\pi^{\pm}$ or $\pi^{\pm}2\pi^0$ with comparable ($\approx 10\%$) probabilities. Since the $\pi^{\pm}\pi^0$ decay mode has about 2.5 times larger branching fraction than the other two (as was shown in table 4.1), this small contamination causes a relatively large impurity (about 20 to 25%) in the decay modes reconstructed for the other two decays.
  - **$\pi^{\pm}2\pi^0$:** The MVA identifies nearly 40% of true $\pi^{\pm}2\pi^0$ decays while the rest are misidentified mainly as $\pi^{\pm}\pi^0$. The reconstructed $\pi^{\pm}2\pi^0$ sample has about 25% contamination from $\pi^{\pm}\pi^0$ and the rest of the contamination (about 17%) originates from the "other" category.
- 3-charged-prong decays:
  - **$3\pi^{\pm}$:** This decay mode has the highest purity and efficiency of all. More than 85% of true $3\pi^{\pm}$ decays are correctly identified and the corresponding reconstructed decay mode is more than 80% pure while the rest are misidentified chiefly as $3\pi^{\pm}\pi^0$.
  - **$3\pi^{\pm}\pi^0$:** About 65% of true $3\pi^{\pm}\pi^0$ decays are identified by the MVA with approximately 23% misidentified as $3\pi^{\pm}$. The corresponding reconstructed decay mode is about 70% pure with equal contamination from $3\pi^{\pm}$ and the "other" category.

It is crucial to validate simulated MVA decay modes with data. To this end, data and simulated samples from a region enriched in $Z \to \tau^+\tau^-$ with $\tau_\mu\tau_h$ final state are collected and plotted in bins of MVA decay mode, see Fig. 5.9. The following conditions are applied on the events. Muons are required to pass *single-muon* trigger and *Medium* muon ID while having an isolation parameter of less than 0.15. They are also required to have $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.1$. In addition, τ_h candidates must satisfy $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.3$. In order to suppress background processes faking τ_h,

these candidates must pass *Medium*, *VVLoose* (very very loose), and *Tight* DeepTau working points against jets, electrons, and muons. Events with more than two leptons are vetoed.

A transverse mass condition, $m_T(\mu, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}) < 40$, is applied to reduce most of W+jets (Electroweak) background, where the transverse mass of a light lepton (e/μ) and missing momentum is defined as

$$m_T(l, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}) \equiv \sqrt{2p_{\mathrm{T}}^{l}|\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}|\,(1-\cos(\Delta\phi))} \tag{5.3}$$

where $\Delta\phi$ is the azimuthal angle between the lepton and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$.

The main backgrounds are W+jets and QCD. The W+jets background is modeled with simulation while its normalization factor is measured from the high-m_T region which is enriched with W+jets. The QCD background is estimated from a control region using the data-driven *fake factor* method. These methods will be discussed in chapter 6. All other background processes are produced with simulated events.

Incorporating the systematic uncertainties of the simulated events shows an approximately 7% uncertainty in the total yield, which is included as a constant (flat) prefit uncertainty. This uncertainty is subsequently added in quadrature to the (relatively small) statistical uncertainty of the simulated events.

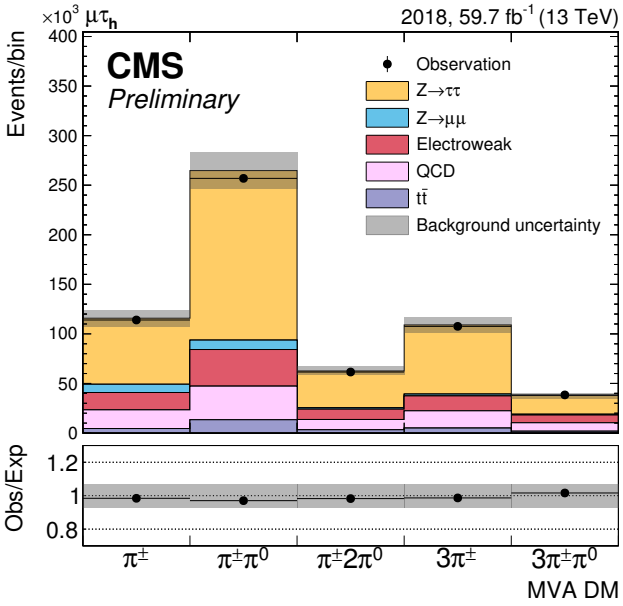

Figure 5.9: MVA decay mode distribution for data and simulated $Z \to \tau^+\tau^-$ enhanced events. There is a good agreement between data and simlation. See text for more information [1].

Data and simulated samples well agree within uncertainties. We therefore used MVA decay mode in the Higgs $\mathcal{CP}$ analysis instead of HPS decay mode, as will be described in chapter 6. The improvement of the $\mathcal{CP}$ sensitivity brought by the MVA decay mode will also be shown in that chapter. As the corresponding code is already implemented in the CMS software (CMSSW), MVA decay mode can also be used by any future CMS analyses which are sensitive to τ_h decay modes.

Chapter 6

Measuring the $\mathcal{CP}$ properties of the Higgs-tau Yukawa coupling

6.1 Introduction

In this chapter, the procedure for measuring the $\mathcal{CP}$ of the $H\tau\tau$ coupling is explained. This chapter follows closely the paper [2] published on this analysis, while my personal contribution is elaborated in greater depth. The data collected for this analysis is taken from the CMS proton-proton (pp) collisions at $\sqrt{s} = 13$ TeV during 2016, 2017 and 2018 with integrated luminosities of 35.9, 41.5, and 59.7 fb^{-1}, respectively. The most sensitive decay channels are targeted, namely $\tau_h\tau_h$, $\tau_\mu\tau_h$, and $\tau_e\tau_h$, where τ_h represents the hadronic decays of taus while τ_μ and τ_e (collectively shown as τ_l) are the decays of taus to a muon and an electron, respectively. In total, 70% of all possible $H \to \tau^+\tau^-$ decays are considered in this analysis.

6.2 Simulated samples

This section describes how Monte Carlo (MC) simulated events are produced. Higgs bosons (signal) produced through ggH, VBF, and VH processes are simulated with the POWHEG 2.0 [99–105] event generator. In order to ensure that the measurement of $\alpha^{H\tau\tau} = \arctan(\tilde{\kappa}/\kappa)$ (defined in section 2.3.1) is not sensitive to the assumptions about the $\mathcal{CP}$ in the Higgs boson production interaction, variables that are sensitive to the $\mathcal{CP}$ of the production, such as $\Delta\phi$ between the two leading jets [106], are not used as inputs to signal discriminants or for event selection. The decay of the Higgs boson to a tau pair is simulated using the PYTHIA generator version 8.2 [107] while keeping the spin of the tau pair uncorrelated. The TAUSPINNER package [96] is then deployed to correlate

the spin of the tau pair based on the value of the mixing angle, $\alpha^{H\tau\tau}$. This correlation is included as three weights for each event which correspond to $\alpha^{H\tau\tau} = \{0°, 90°, 45°\}$. The distribution for any $\alpha^{H\tau\tau}$ can be derived by combining these weights.

The simulation of single top quark and $t\bar{t}$ processes are performed using the POWHEG generator while the MADGRAPH5_aMC@NLO generator is used for diboson production as well as processes including a W or Z boson associated with up to four hadronic jets, which are denoted as W+jets and Z+jets, respectively.

In order to model parton shower and fragmentation as well as the decay of tau leptons, the generators are interfaced with PYTHIA 8.2. Simulated events are then passed through GEANT4 to include smearing effects from the CMS detector. The methods applied to MC events for reconstructing physics objects are the same as those applied to data. Simultaneously with a collision of interest, a set of inelastic pp collisions occur which are known as *pile-up*. Pile-up is modeled with PYTHIA 8.2. The pile-up distribution is weighted to match the data.

6.3 Physics object and event selection

In Chapter 4, the methods for reconstructing physics objects were described. In this section, the selections applied on the objects and events used for this analysis along with some of the corresponding corrections are explained. The selections closely follow those used in the measurement of the Higgs-tau coupling [108]. The main differences are the use of the DeepTau discriminator (as it has been recently developed), the requirement on the visible mass for suppressing Drell-Yan background, and the requirement on the impact parameter (described in Section 6.7).

Physics object selection

Muons from the $\tau_\mu\tau_h$ channel need to be well discriminated against other particles and non-prompt muons. To this end, an isolation variable is defined as the following. The scalar sum of the p_{T} of particles originating from the PV and within a cone of radius $\Delta R = 0.4$ around the muon direction is estimated and then divided by the muon p_{T}. This ratio is used as the isolation variable and is required to be less than 15%. Likewise, an isolation variable is defined for electrons in the $\tau_e\tau_h$ channel with an analogous cone of size $\Delta R = 0.4$. The isolation variable must be less than 15%.

The momenta of hadronic jets are defined as the vectorial sum of all their constituents. Corrections are applied to account for the energy scale of jets and to remove pile-up contribution. In the 2017 data-taking period, a large amount of noise affected part of the endcap which led to data/MC disagreement. Therefore, jets with

Channel	Year	Online trigger	Offline p_{T} (GeV)
$\tau_h\tau_h$	All years	$\tau_h(35)$ & $\tau_h(35)$	$p_{\mathrm{T}}^{\tau_h} > 40$
$\tau_\mu\tau_h$	2016	$\mu(22)$, $\mu(19)$ & $\tau_h(20)$	$p_{\mathrm{T}}^{\mu} > 20$ & $p_{\mathrm{T}}^{\tau_h} > 25$
	2017, 2018	$\mu(24)$, $\mu(20)$ & $\tau_h(27)$	$p_{\mathrm{T}}^{\mu} > 21$ & $p_{\mathrm{T}}^{\tau_h} > 32$
$\tau_e\tau_h$	2016	e(25)	$p_{\mathrm{T}}^{e} > 26$
	2017	e(27), e(24) & $\tau_h(30)$	$p_{\mathrm{T}}^{e} > 25$ & $p_{\mathrm{T}}^{\tau_h} > 35$
	2018	e(32), e(24) & $\tau_h(30)$	$p_{\mathrm{T}}^{e} > 25$ & $p_{\mathrm{T}}^{\tau_h} > 35$

Table 6.1: The online triggers and offline p_{T} requirements applied to the $\tau_h\tau_h$, $\tau_\mu\tau_h$, and $\tau_e\tau_h$ channels for different data-taking periods. The trigger p_{T} requirement (in GeV) on each object is shown in parenthesis after the object.

$2.65 < |\eta| < 3.10$ and $p_{\mathrm{T}} < 50$ GeV which are reconstructed in this data-taking period are discarded. Jets initiated from b-quarks (b-jets) are reconstructed with the deepCSV algorithm, as explained in section 4.6. The working point used in this analysis provides around 70% efficiency for b-jet identification with a misidentification rate for gluon or light quark jets of around 1%.

The working point used for τ_h identification (ID) corresponds to 60% τ_h identification efficiency with a misidentification rate of 5×10^{-3}. The invariant mass of the tau pair from a Higgs decay, denoted as $m_{\tau\tau}$, is estimated using the SVFIT algorithm [109]. This algorithm combines $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ and its uncertainty matrix along with the four-vectors of the two tau candidates (excluding neutrinos) to estimate $m_{\tau\tau}$ based on an event-by-event likelihood optimization.

Event selection

From the events collected by the online trigger system, the di-tau trigger events are used for the $\tau_h\tau_h$ channel while the events which pass either of the single-lepton or lepton+τ_h trigger are used for the $\tau_l\tau_h$ channels. Several offline selection criteria are applied to the events. The two reconstructed tau candidates must be separated by $\Delta R > 0.5$ and each candidate must match the object reconstructed by the trigger system within $\Delta R < 0.5$. The online triggers and offline p_{T} requirements are summarized in table 6.1. The offline p_{T} threshold on τ_l (τ_h) is set 1 (5) GeV above the online trigger because of the turn-on shape of the trigger efficiency. In addition, if the $\tau_l\tau_h$ channel is triggered by a single-lepton trigger, an offline p_{T} threshold of 20 GeV is required on the τ_h leg.

In the $\tau_l\tau_h$ channel, W+jets background is suppressed by requiring $m_T(l, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}) < 50$, where $m_T(l, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$ is defined in Eq. 5.3. Events with $m_{vis} < 40$ GeV are vetoed to slightly decrease the Drell-Yan background. The longitudinal and transverse impact parameter of τ_l candidates, respectively shown as d_z and d_{xy}, must satisfy $|d_z| < 0.2$ cm and $|d_{xy}| < 0.045$ cm. These impact parameters are computed using minimization

in the transverse plane only, while in calculating the $\phi_{\mathcal{CP}}$ observable in the impact parameter method, a 3D minimization is performed (see section 6.4.2). For the leading τ_h in the $\tau_h\tau_h$ channel, only $|d_z| < 0.2$ cm is applied to prevent large efficiency loss. In the $\tau_l\tau_h$ decay channel, events containing b-jets are vetoed. Whenever multiple tau pairs are reconstructed, they are ranked based on the DeepTau score of τ_h and lepton isolation score of τ_l, after which the pair with the highest rank is chosen.

6.4 $\phi_{\mathcal{CP}}$: strategy and optimization

In Section 2.3.2, different methods for reconstructing tau decay planes and subsequently the $\phi_{\mathcal{CP}}$ observables were described. In this section, firstly the strategy of selecting decay plane methods is explained and then the techniques used for optimizing $\phi_{\mathcal{CP}}$ sensitivity are introduced.

6.4.1 Strategy

In the impact parameter method, reconstructing the tau impact parameter with sufficiently high resolution is essential. The tau impact parameter is not large compared to its resolution, despite the excellent CMS tracker resolution. As a result, the $\phi_{\mathcal{CP}}$ observable cannot be precisely reconstructed with this method. However, in the neutral-pion method, the π^0 four-momentum needs to be reconstructed instead of the tau impact parameter. Thanks to the small size of ECAL modules and the relatively large distance of the ECAL to the tau decay point, a precise reconstruction of the π^0 four-momentum, and hence the $\phi_{\mathcal{CP}}$ observable, is possible. We compared the impact parameter and neutral-pion methods in a phase space where the impact parameter has an optimal performance. The two methods showed the same sensitivity normalized to the number of events. However, when the required selections (explained in section 6.7) are applied to the impact parameter of taus, a large fraction of events are discarded, which decreases the sample size by a factor of two. Therefore, although the impact parameter method can be used in all decay channels, we used it only for the $e\pi$, $\mu\pi$, and $\pi\pi$ channels, in which the neutral-pion method is not applicable.

The neutral-pion method is applied to the $\rho\rho$, ρa_1^{1pr}, ρa_1^{3pr}, $a_1^{\text{1pr}} a_1^{\text{1pr}}$, and $a_1^{\text{1pr}} a_1^{\text{3pr}}$ channels and the polarimetric vector method is used only for the $a_1^{\text{3pr}} a_1^{\text{3pr}}$ channel, where the Higgs rest frame is reconstructable. In the decay channels where one of the taus decays to $\{e, \mu, \pi\}$ and the other decays to $\{\rho, a_1^{\text{1pr}}, a_1^{\text{3pr}}\}$, the combined method is employed.

6.4.2 Optimization

MVA decay mode

Correctly identifying tau decay modes is essential to exploit the $\mathcal{CP}$ sensitivity in this analysis. As described in detail in chapter 5, a multivariate analysis (MVA) was developed to improve the identification of τ_h decay modes. The decay mode identification performed by the MVA is called the MVA decay mode, as opposed to the HPS decay mode which is the one performed by the HPS algorithm. The efficiency and purity of the MVA and HPS decay modes were compared in Fig. 5.6.

Fig. 6.1 is the negative log-likelihood scan of $\alpha^{H\tau\tau}$ showing that using the MVA decay mode improves the expected sensitivity of distinguishing pure $\mathcal{CP}$-even from pure $\mathcal{CP}$-odd couplings by about 25% in the $\rho\rho$ decay channel, one of the most $\mathcal{CP}$ sensitive channels in our analysis. The simulated events for this comparison are produced based on the experimental conditions of the 2016 data-taking period. Only statistical uncertainties are included in the fit and the background modeling is described in section 6.5.

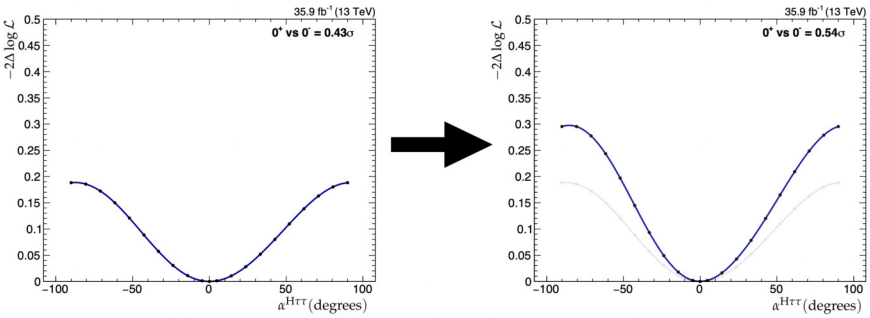

Figure 6.1: The negative log-likelihood scan of $\alpha^{H\tau\tau}$ in the $\rho\rho$ channel when using the HPS decay mode (left) compared to when using the MVA decay mode (right). The sensitivity improved by $\approx 25\%$. The simulated events are based on the 2016 data-taking period conditions [110].

Primary vertex refitting

Due to the finite lifetime of tau leptons, the decay position of taus (having enough p_{T} to pass the analysis minimum requirements) in the lab frame is displaced from the primary vertex (PV), which is essentially where they are produced (up to the experiment resolution). Therefore, when estimating the PV by fitting to the PV tracks, the displaced tracks of the tau pair deteriorate the PV position resolution. Therefore, in this analysis, the tracks of the decay products of the two taus are removed and the new position of the PV is estimated by refitting to the remaining tracks using the AVF

algorithm described in section 4.3.

The beam spot position, defined in section 4.2.2, has a relatively low uncertainty. This uncertainty was added to the AVF algorithm to constrain the PV position, which led to a factor of 3 (4) more precise resolution in the xy plane for the Higgs (Drell-Yan) sample while the resolution in the z axis did not change significantly.

In the impact parameter and polarimetric vector method, where the reconstruction of the PV is needed, the refitted PV is used.

Impact parameter significance

In order to measure the impact parameter of a tau, a function is fitted to the helical shape of the track of the outgoing charged particle. The point on the fitted curve which has the least 3D distance to the PV is used for defining the impact parameter. The minimization in three dimensions has two benefits over the transverse-plane-only minimization: (i) higher resolution in the z coordinate of the impact parameter in high η region and (ii) the impact parameter and its uncertainty can be measured and propagated into S_{IP}, which is defined as the ratio of the impact parameter to its uncertainty. Small values of S_{IP} show no $\mathcal{CP}$ sensitivity and are therefore discarded (see section 6.7).

6.5 Background estimation

The backgrounds considered for this analysis are from the Drell-Yan process, QCD processes, Z/W+jets, diboson production, $t\bar{t}$, and single top quarks. Particles produced via QCD processes, such as jets, light leptons, and τ_l decays can be misidentified as a τ_h decay. QCD products can also fake τ_l by jets and prompt leptons. Most of the backgrounds involving jets misidentified as τ_h are modeled with a data-driven method called the *fake factor* (F_F) method. (The fake factor method is described in Ref. [111].) All backgrounds involving two genuine taus are modeled with the τ-embedding data-driven method [112]. The remaining minor backgrounds are modeled with MC simulation.

6.5.1 Fake factor method

This method is applied in the $\tau_l\tau_h$ ($\tau_h\tau_h$) channel to estimate the contribution of jets misidentified as the τ_h (the leading τ_h). Jets which are misidentified as τ_h are denoted as jet $\rightarrow \tau_h$. This method can estimate both the shape and normalization of this contribution. A determination region is defined orthogonal to the signal region and enriched in a background process including jet $\rightarrow \tau_h$. A τ_h *nominal ID* is defined as

the DeepTau DNN identification requirement on τ_h mentioned in section 6.3 and a τ_h *relaxed ID* is defined by requiring τ_h candidates to fail the nominal ID but pass a looser DeepTau DNN requirement. The ratio of the nominal ID rate to the relaxed ID rate in the determination region is called fake factor. An application region is then defined with exactly the same event selection criteria as in the signal region but by requiring a τ_h candidate (which is the leading τ_h in $\tau_h\tau_h$) to pass the relaxed ID instead of the nominal ID. In the application region, the contribution of processes other than jet $\rightarrow \tau_h$ are estimated using simulated events and subtracted from the total number of events. In the next step, the jet $\rightarrow \tau_h$ yield in the application region is scaled by F_F and considered as the estimated jet $\rightarrow \tau_h$ contribution in the signal region. To account for data/MC residual difference, corrections are applied to kinematic variables such as the lepton p_{T} and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ spectra.

6.5.2 τ-embedding method

Backgrounds including two genuine τ_h decays, which are mainly from $Z \rightarrow \tau^+\tau^-$ events, are estimated using the τ-embedding method which relies on the lepton universality. A dedicated di-μ trigger is used in all data-taking periods for collecting oppositely-charged muon pairs. Then in each of the collected di-μ events, the hits associated to the muon pair are removed from the detector. In parallel, a $Z \rightarrow \tau^+\tau^-$ decay is simulated in an empty detector and is further required to decay to $\tau_l\tau_h$ or $\tau_h\tau_h$ channels with the tau pair having the same kinematics as the muon pair. After adding detector effects to the decays of tau pair, they are merged with the data event. The corrections which are applied to account for mis-modeling include corrections on the efficiency of identification and tracking, isolation, and trigger of muons and of electrons. Another set of corrections are applied to the identification as well as trigger efficiency of τ_h candidates. The *tag-and-probe* method [113] is used to measure these corrections as will be elaborated in section 6.6. Besides, further corrections are applied on the τ_h energy scale and the impact parameter of electrons, muons, and charged pions.

6.5.3 Other backgrounds

The remaining minor backgrounds (≈10%) are estimated using MC simulation. Tables 6.2 and 6.3 summarize the method used for estimating the backgrounds in the $\tau_h\tau_h$ and $\tau_l\tau_h$ channels, respectively.

Leading τ_h	Subleading τ_h		
	Genuine τ_h	jet $\to \tau_h$	(Prompt lepton/τ_l) $\to \tau_h$
Genuine τ_h	τ-Embedding	Simulation	Simulation
jet $\to \tau_h$	F_F	F_F	F_F
(Prompt lepton/τ_l) $\to \tau_h$	Simulation	Simulation	Simulation

Table 6.2: Methods used to model different sources of backgrounds in the $\tau_h\tau_h$ channel.

τ_l	τ_h		
	Genuine τ_h	jet $\to \tau_h$	(Prompt lepton/τ_l) $\to \tau_h$
Genuine τ_h	τ-Embedding	F_F	Simulation
jet $\to \tau_h$	Simulation	F_F	Simulation
(Prompt lepton/τ_l) $\to \tau_h$	Simulation	F_F	Simulation

Table 6.3: Methods used to model different sources of backgrounds in the $\tau_l\tau_h$ channel.

6.6 Simulation correction on efficiencies

In this section, the efficiency of identification, isolation, and trigger requirements applied to electron, muon, and τ_h candidates are measured. The efficiencies are measured for simulated events and data, after which the corresponding *scale factors* defined as the ratio of the two efficiencies are calculated. Scale factors are applied to the simulated events to correct the difference between data and simulated events stemming from mis-modeling in simulation. The simulated events can be generated solely from MC or using the τ-embedding technique (called embedded samples).

The corrections in this section are measured using the tag-and-probe method [113]. In this method, to measure the efficiency of object $\mathcal{P}$ (e.g. a τ_h) after a requirement is applied on it, firstly a process containing $\mathcal{P}$ and a tagging object $\mathcal{T}$ (e.g. a well-isolated μ) in the final state is considered. The process is tagged with $\mathcal{T}$ after which the requirement is applied on the probe leg $\mathcal{P}$ on which the corresponding efficiency is measured.

The procedure for extracting the electron and muon efficiencies is explained in section 6.6.1. The efficiencies for the 2016 data-taking year were extracted by me while my colleagues extracted those of the 2017 and 2018 data-taking years. The procedure for extracting the hadronic tau efficiencies, which was performed solely by me (for all data-taking periods), is described in section 6.6.2.

6.6.1 Electron and muon efficiency correction

The efficiency of identification, isolation, and trigger requirements on electrons and muons are measured for both data and simulated samples, separately for MC and

embedded samples, using the tag-and-probe method. The measurements for electrons and muons are performed on phase spaces enriched in $Z \to ee$ and $Z \to \mu\mu$ events, respectively.

In order to measure the efficiencies for MC, Drell-Yan Z+jets events are used which are modelled at the leading-order (LO) with MadGraph5 [114]. On the other hand, the embedded events are produced using the same method described in section 6.5.2, except that the muon pair removed from the detector in a data event is replaced with an electron/muon pair (rather than taus) coming from a Z boson decay simulated with MC.

The electrons (muons) selected for measuring the efficiencies using the tag-and-probe method are required to have $|\eta| < 2.5$ (2.4) and $p_{\mathrm{T}} > 10$ GeV. The impact parameter of the leptons must satisfy $d_{xy} < 0.045$ cm and $d_z < 0.2$ cm. An event is discarded if it contains more than one muon/electron pair satisfying the above conditions. The lepton pair must be separated by $\Delta R > 0.5$ and have invariant mass $50 < m_{ll} < 150$ GeV to increase the Z boson yield.

In each event, from the two selected leptons, one is considered as a tag and the other as a probe. To increase statistics, the tag and probe method is applied twice for each event, with the role of tag and probe being swapped. The tags must pass identification and isolation requirements to suppress fake leptons. Muon tags are required to pass a single-muon trigger with an online p_{T} threshold of 24 (27) GeV in the 2016 (2017 and 2018) data-taking periods. They are further required to pass an offline threshold of 25 (28) GeV. Electron tags must pass a single-electron trigger with a 25 (35) GeV online p_{T} threshold in 2016 (2017 and 2018). They must additionally have a minimum offline p_{T} of 26 (36) GeV.

The identification, isolation, and trigger requirements are applied to the probe lepton sequentially, meaning that the identification scale factor is measured first, then it is applied before the isolation scale factor is measured, and finally both the identification and isolation scale factors are applied before the trigger scale factor is measured. The lepton trigger scale factor is derived separately for single-lepton trigger and lepton+tau cross trigger[1]. However, the electron+tau cross trigger is only applied to the 2017 and 2018 data-taking years (which are not derived by me) and therefore the results for this trigger is not elaborated in this section.

The efficiencies are measured in bins of p_{T} and η of the probe lepton. For each selection requirement (ID/isolation/trigger), the events in each bin are split into "pass" and "fail" categories depending on whether or not the probe lepton in the event meets the selection requirement. Fake leptons produced by background processes can pass the

[1]The word "cross trigger" means that two objects are used in the trigger.

selection requirements as well so these events must be taken into account by modeling the signal and backgrounds. The efficiency is derived by a simultaneous fit to the pass and fail categories and is defined as

$$\epsilon = \frac{N_{\text{pass}}}{N_{\text{pass}} + N_{\text{fail}}}, \tag{6.1}$$

where N_{pass} and N_{fail} are the number of events in the pass and fail categories, respectively, extracted from the fitted **signal** functions to the di-lepton mass (m_{ll}) distribution.

The background for the ID requirement is modeled using an error function times an exponentially decaying function (called the CMS-shape function [115]). Such backgrounds can normally be modeled solely with an exponentially decaying function. However, because of the minimum p_{T} threshold on the lepton pair, the background in the low mass region is reduced compared to the exponential function and hence the error function is added to include this reduction. The ID requirement removes a large fraction of backgrounds; therefore, the backgrounds for the isolation and trigger requirements are modeled with a decaying exponential only.

Two functions are used to model the signal. The most commonly used function is a Double-Voigtian function (DV) which is a Breit Wigner (BW) function with a width equal to the Z boson width (2.495 GeV) convoluted with the superposition of two Gaussian functions. The other function is the superposition of two Crystal Ball functions convoluted with a BW with the Z boson width, where the Crystal Ball function is defined in Eq. F-1 of [116]. The latter is used for fitting isolation efficiencies in the 2016 data-taking period while the former is used for all other signals. Fig. 6.2 shows an example of the fit for each of the ID (top), isolation (middle), and trigger (bottom) efficiency measurements.

In the isolation scale factor measurement, some of the m_{ll} distributions in the fail category show a two-peak structure, one around the Z boson mass and the other at a lower mass ($\approx$75 GeV). Studying this effect with MC revealed that the lower mass peak is the result of final state radiation (FSR), where one or both of leptons radiates a photon carrying a share of the lepton energy. Considering these events as signal has two problems. Firstly, these scale factors are measured with prompt leptons while in the actual analysis, the scale factors are applied to non-prompt leptons which have different FSR rates and kinematics. Secondly, fitting the two-peak structure needs a function with a large number of degrees of freedom which sometimes results in fitting part of the background as signal or vice versa. Therefore, an effort is made to veto these events in the following procedure. The highest p_{T} photon with $\Delta R < 0.4$ with respect to the

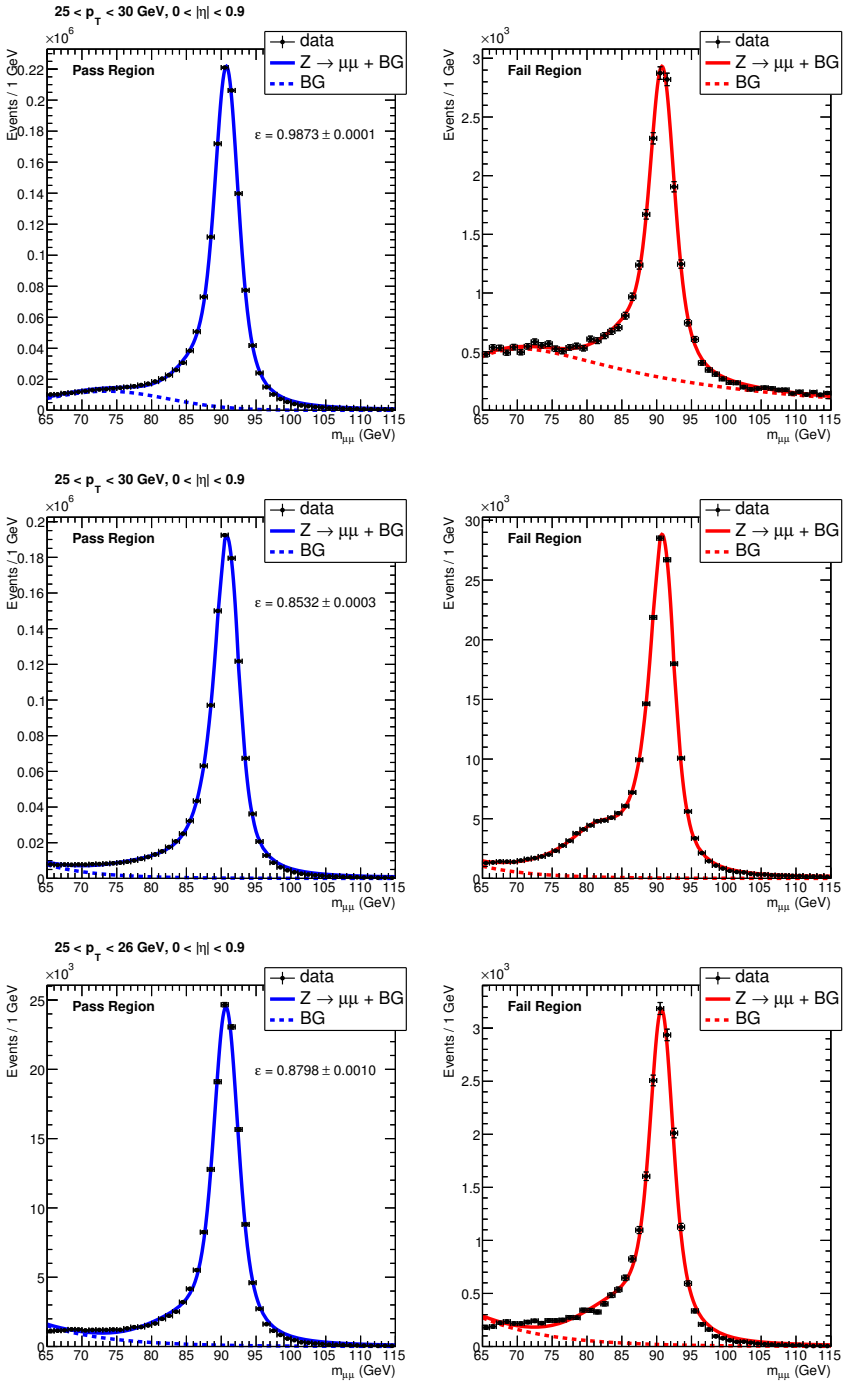

Figure 6.2: Examples of muon efficiency measurement fits. The top, middle and bottom pair of plots show the fits to the pass (left) and fail (right) regions of the identification, isolation, and single-muon trigger requirement, respectively. The data points are derived from real CMS data, the fits to the background are shown with dashed lines, and the fits to the sums of background and Drell-Yan events are shown with solid lines.

leading lepton and the one with respect to the subleading lepton are collected. This collection can include two, one, or no photons depending on whether such photons exist. Any photons with $p_{\mathrm{T}} < 10$ GeV are discarded from the collection. If no photons are left in the collection, the event is used for the scale factor measurement and is considered as a no-FSR event. Otherwise, in order to distinguish genuine FSR events from those by chance having photons around the leptons, the invariant mass of the lepton-photon system is calculated ($m_{ll\gamma}$ or $m_{ll\gamma\gamma}$) and required to be between 80 and 100 GeV, close to the Z boson peak mass. Fig. 6.3 and Fig. 6.4 display respectively the muon and electron scale factors for the central region of the detector derived for the 2016 data-taking year. The efficiencies and scale factors approach a constant value in the high-p_{T} region as expected. The lower panel of each sub-figure shows the data/MC and data/embedded ratios. This ratios are applied in the analysis to correct the simulated events. The scale factor plots for other regions of the detector can be found in Appendix B.

6.6.2 Hadronic taus efficiency correction

τ_h identification efficiency

Corrections are applied to taus to account for the difference between data and simulation efficiency after hadronic taus pass the reconstruction and identification algorithms. The corrections derived for previous CMS analyses were in bins of p_{T} for $p_{\mathrm{T}}^{\tau_h} < 40$ GeV and in bins of HPS decay mode for $p_{\mathrm{T}}^{\tau_h} > 40$ GeV. However, since we use MVA decay mode in this analysis, we need these corrections in bins of MVA decay mode. Therefore, the corrections are derived in bins of MVA decay mode, separately for $p_{\mathrm{T}}^{\tau_h} < 40$ and $p_{\mathrm{T}}^{\tau_h} > 40$. Separate corrections are derived for embedded and MC samples.

In order to measure the scale factors, a sample enriched in $Z/\gamma^* \to \tau_\mu \tau_h$ is provided by applying the following selections for the 2016 (2017/2018) data-taking period(s):

- The muon must pass a single-muon trigger and a medium identification criterion. It also needs to be well isolated and have $p_{\mathrm{T}} > 23$ (28) GeV.
- The τ_h candidate is required to pass the HPS algorithm and have $p_{\mathrm{T}} > 20$ GeV. It must also pass a medium working point against hadronic jets, a tight working point against muons and a very loose working point against electrons based on the DeepTau DNN scores.
- The muon and τ_h must be oppositely charged and separated by $\Delta R > 0.5$. If more than one pair is found, the most isolated pair is kept.
- The transverse mass $m_T(\mu, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}})$ must be less than 40 GeV to suppress W+jets background.

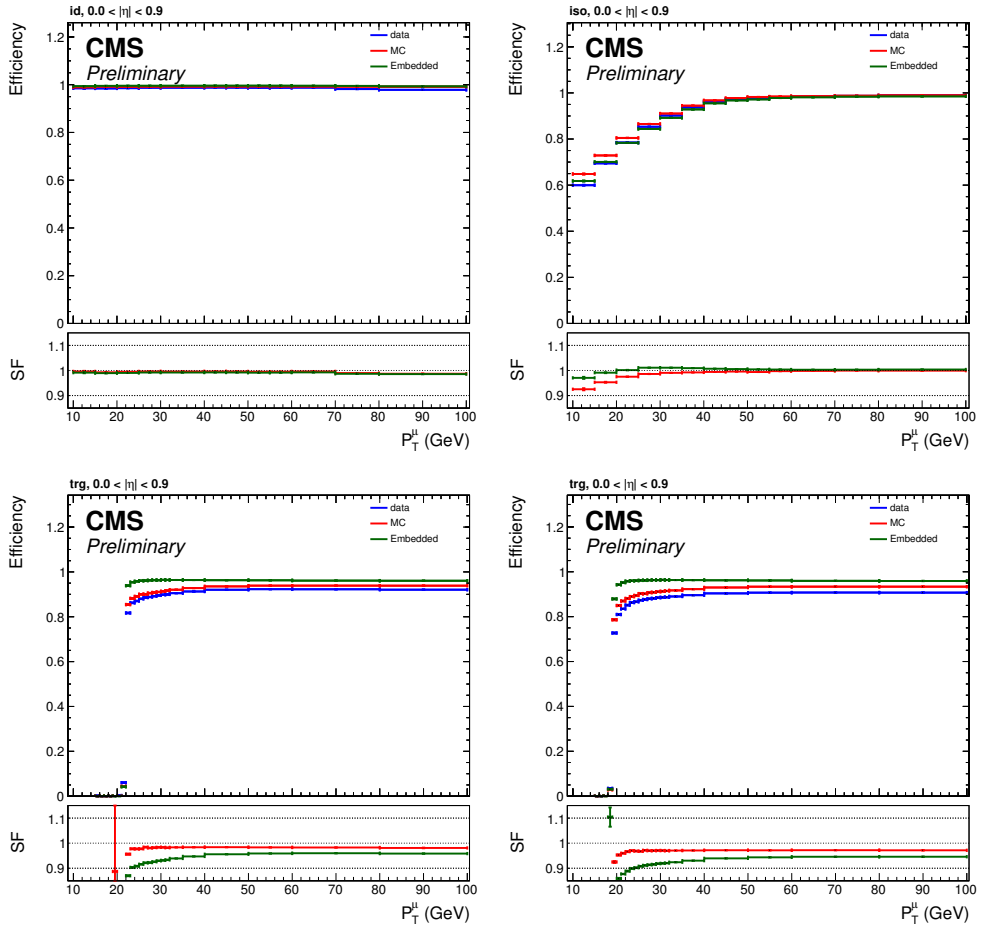

Figure 6.3: The efficiencies and scale factors for muon identification (top-left), muon isolation (top-right), single-muon trigger (bottom-left), and muon+tau cross trigger (bottom-right) requirements are displayed. The efficiency of data (blue), MC (red), and embedded (green) samples are shown in the upper panel of each sub-figure while the ratio of data to the MC and embedded samples are shown in the lower panel.

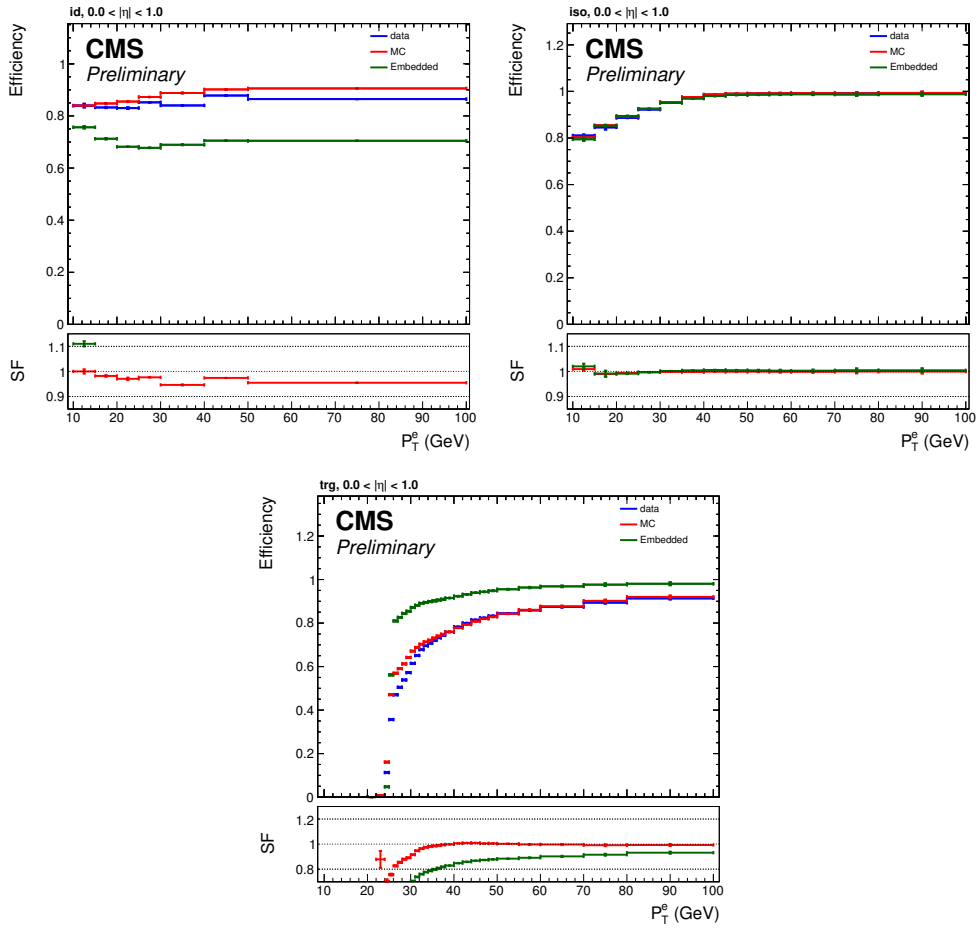

Figure 6.4: The efficiencies and scale factors for electron identification (top-left), electron isolation (top-right), and single-electron trigger (bottom) requirements are displayed. The efficiency of data (blue), MC (red), and embedded (green) samples are shown in the upper panel of each sub-figure while the ratio of data to the MC and embedded samples are shown in the lower panel.

- Events with additional electrons, muons, or b-jets are vetoed.

In addition, a sample enriched in $Z/\gamma^* \to \mu\mu$ is collected and included in the fit to ensure that the fit is able to differentiate between shifts in $Z/\gamma^* \to \tau_\mu\tau_h$ due to the τ_h identification (τ_h ID) scale factor and other sources (such as cross section) as the τ_h ID scale factor modifies only the $Z/\gamma^* \to \tau_\mu\tau_h$ yield while sources like cross section modify both simultaneously. The di-muon selection criteria are therefore chosen similar to the $\tau_\mu\tau_h$ case. The selections for the 2016 (2017/2018) data-taking period(s) are the following:

- Events must contain two muons passing a medium identification criterion, have $p_{\mathrm{T}} > 10$ GeV, and be well isolated.
- The leading muon must additionally pass the single-muon trigger and have $p_{\mathrm{T}} >$ 23 (28) GeV.
- The two muons are required to be separated by $\Delta R > 0.5$ and be oppositely charged. If more than one muon pair is found, the most isolated pair is retained.
- The invariant mass of the muon pair must be between 70 and 120 GeV.
- Events with additional electrons, muons, or b-jets are vetoed.

There are therefore two categories: $\tau_\mu\tau_h$ and $\mu\mu$. In the $\tau_\mu\tau_h$ category, the invariant mass of τ_μ and τ_h is used as a variable to directly measure the scale factor using a fit. This is in contrast to measuring the identification scale factor of muons and electrons described in section 6.6.1 during which the efficiencies are measured and the scale factors are subsequently derived using their ratio. In the $\mu\mu$ category, a single bin is used in the fit to constrain the Drell-Yan yield. A rate parameter is included in the fit to shift the Drell-Yan yields in both categories simultaneously. Another rate parameter is added to shift the τ_h ID scale factor, which is the parameter of interest. A simultaneous maximum-likelihood fit is performed to both categories to extract the τ_h ID scale factor.

The background processes considered in this measurement are QCD, W+jets, single-top, diboson, and $t\bar{t}$. All backgrounds except QCD are estimated with MC while QCD is estimated using the fake-factor method described in section 6.5.1. The yield of W+jets is estimated using the data in a high-m_T side-band region ($m_T > 70$ GeV) which is enriched in W+jets events.

The fits are performed separately for MC and embedded samples. In the embedded samples, the $\mu\mu$ category is modeled using the same procedure described in section 6.5.2

for modeling $Z \to \tau\tau$ events while here, after removing the muon pair from data, a $Z \to \mu\mu$ is simulated.

The systematic uncertainties included in the fit are divided into normalization and shape uncertainties, where the former only affects the yield of a distribution while the latter can change the shape of it. The following systematic uncertainties are added:
Normalization uncertainties:

- **Trigger efficiency:** A 2% uncertainty on the single-muon trigger efficiency.
- **b-jet efficiency:** A 5% uncertainty applied to diboson and $t\bar{t}$ events to account for the b-tagging efficiency because of vetoing b-jets.
- **Luminosity:** A 2.5% uncertainty applied to all processes for which the yield is estimated from MC [117–119].
- **Muon identification/isolation efficiency:** A 1% uncertainty per muon is applied.
- **jet $\to \tau_h$ fake-rate**: A 20% uncertainty applied to all processes containing a jet $\to \tau_h$ fake if the yield is estimated from MC.
- **$\mu \to \tau_h$ fake-rate:** A 20% uncertainty applied to $Z \to \mu\mu$ process in the $\tau_\mu\tau_h$ category in which one of the muons fakes a τ_h.
- **Background normalization uncertainty:**
  - A 10% uncertainty on the W+jets yield for extrapolating from the high-m_T to low-m_T regions when estimating W yields.
  - A 10% uncertainty on the QCD yield for extrapolating the determination region to the application and signal regions.
  - A 4% uncertainty on the embedded $Z \to \mu\mu$ and $Z \to \tau\tau$ yields.
  - A 2% uncertainty on the MC $Z \to \mu\mu$ and $Z \to \tau\tau$ yields for the Drell-Yan cross section.
  - A 5% uncertainty on the diboson and single-top yields for their production cross sections [120, 121].
  - A 4.2% uncertainty on the $t\bar{t}$ yield for its production cross section.

Shape uncertainties:

- **μ energy scale:** A $0.4 - 2.7\%$ uncertainty depending on η.

- **Bin-by-bin:** An uncertainty to account for the statistical fluctuation of the signal and background samples, estimated using the "Barlow-Beeston" method [122, 123].
- **Top and Z p_T reweighting:** The top quark and Z boson p_T distributions derived from simulation are corrected by comparing them with data in phase-spaces dominated by $t\bar{t}$ and $Z \to \mu^+\mu^-$, respectively. An uncertainty is included for each of the two corrections.
- **τ_h energy scale:** A $1.0 - 1.5\%$ uncertainty depending on τ_h decay mode.
- **Uncertainties on the $\vec{p}_T^{\,miss}$ scale and resolution:** These uncertainties account for correcting the contribution of hadronic jets to $\vec{p}_T^{\,miss}$.

Fig. 6.5 shows the maximum likelihood fit results for the τ_h identification scale factor in the $\tau_\mu\tau_h$ channel when embedded samples are used. The scale factors for different data-taking years are not necessarily the same as the experimental conditions are different. Fig. 6.6 shows the postfit plots of the $\tau_\mu\tau_h$ visible mass distribution after applying τ_h ID scale factors (for 2016 period and $p_T > 40$ GeV). The data and simulated events perfectly agree within the uncertainties, which shows the effectiveness of the measured scale factors. These plots are shown as examples; all the τ_h ID and postfit plots can be found in Appendix C.

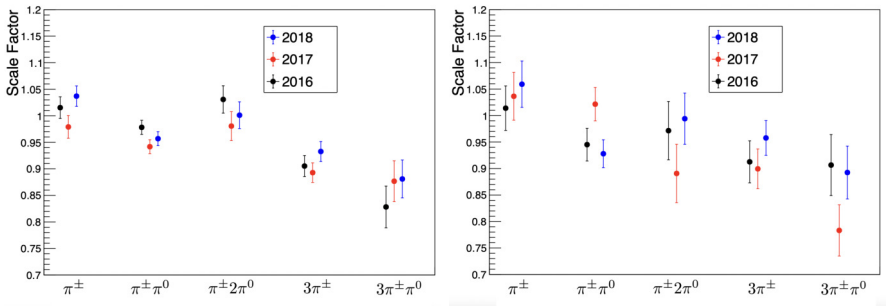

Figure 6.5: The τ_h identification scale factors for different MVA decay modes and data-taking periods. The left (right) plot shows the scale factors for $20 < p_T^{\tau_h} < 40$ GeV ($p_T^{\tau_h} > 40$ GeV). These scale factors are derived for the embedded sample and are valid for $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels.

τ_h trigger efficiency

In this section, the efficiencies of data, MC and embedded samples are extracted for the case when the trigger requirement is applied on τ_h and the corresponding scale factors are calculated. The existing efficiencies are calculated in bins of HPS decay mode

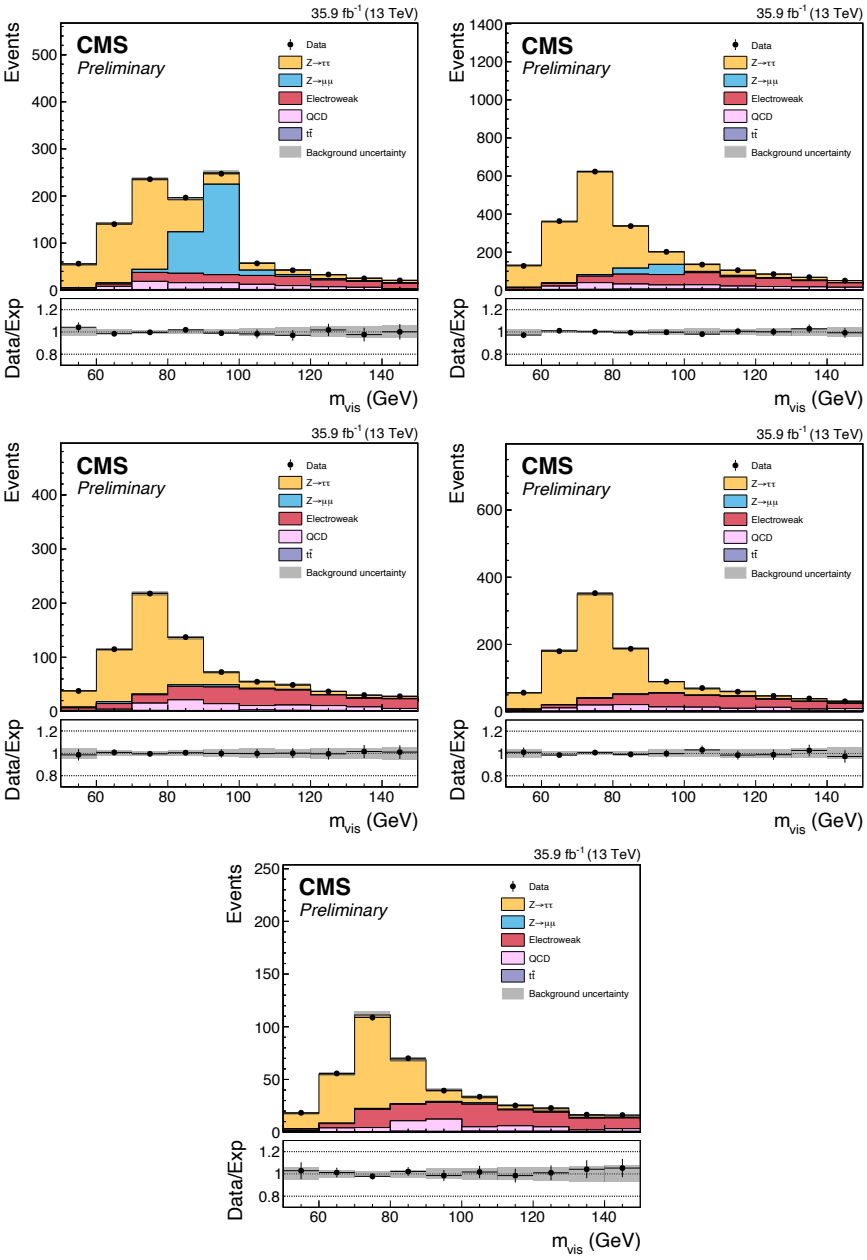

Figure 6.6: Postfit plots of m_{vis} distributions in the $\tau_\mu\tau_h$ channel after applying the τ_h identification scale factors for the $\pi^\pm$ (top-left), $\pi^\pm\pi^0$ (top-right), $\pi^\pm 2\pi^0$ (middle-left), $3\pi^\pm$ (middle-right), and $3\pi^\pm\pi^0$ (bottom) MVA decay modes. The plots are for the embedded samples of 2016 data-taking period and the τ_h candidates are required to have $p_{\mathrm{T}}^{\tau_h} > 40$ GeV.

which may not be accurate enough for our analysis. We measured these efficiencies as a function of MVA decay mode for $\tau_e\tau_h$, $\tau_\mu\tau_h$, and $\tau_h\tau_h$ channels in all data-taking periods. In this measurement, the tag-and-probe method is used on the $\tau_\mu\tau_h$ events by tagging the muon leg and probing the τ_h leg.

The background is estimated similar to the method described earlier in this section for τ_h ID scale factor. The following conditions are required for the 2016 (2017/2018) data-taking periods:

- The muon must pass the single-muon trigger and be well isolated. It must also have a minimum p_{T} of 25 (29) GeV.
- Events with additional leptons or with b-jets are vetoed.
- A medium, very loose, and tight DeepTau DNN working points are used against jets, electrons and muons, respectively.
- To suppress W+jets background, $m_T(\mu, \vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}) < 40$ GeV is required.
- The visible mass of μ and τ_h is required to be within $50 < m_{vis} < 90$ GeV.

The efficiency for each $p_{\mathrm{T}}^{\tau_h}$ bin is initially calculated as the ratio of the events passing the trigger in that bin. The final efficiency is extracted using a fit to the graph of efficiencies. Attempts have been made to fit the graph with a parametric function, such as the error function. However, the turn-on part of the graph does not always fit well as the graph sometimes exhibits a double-step shape, which is a result of relaxing the isolation condition on τ_h candidates above a certain p_{T}. Therefore, Gaussian Process (GP) regression implemented in Scikit-learn python package [124] is used to fit the graph with a non-parametric function. This fitting method can also provide the uncertainty band of the fit, where the uncertainties between the graph points are estimated using interpolation. In this method, kernels available in the python package are used to determine the general shape of the fitted curves. In the high-p_{T} region (plateau), a constant kernel is used, in contrast to the low-p_{T} region (turn-on) in which a constant kernel is multiplied by a Matern kernel (with parameter $\nu = 1$) to allow fitting the turn-on. The bin uncertainties are included in the fit through a White Noise kernel for both p_{T} regions. Finally a tanh() function is used to smooth the intersection of the high and low p_{T} regions.

The scale factors are calculated through dividing the function fitted to data by the one fitted to MC/embedded sample and the uncertainties of the scale factors are derived using standard error propagation. Fig. 6.7 shows the fits for the $\tau_\mu\tau_h$ channel in the 2016 data-taking year when using embedded samples. (See Appendix C for all

fits.) A large amount of data is used to produce embedded samples, leading to a small uncertainty compared to the MC samples. The vertical red dashed line shows the p_{T} threshold below which the events are excluded from the analysis. The scale factor functions have a smooth shape and are approximately flat in the p_{T} range used in the analysis.

6.7 Event categorization

In order to improve the $\mathcal{CP}$ sensitivity, classifiers are trained to identify the signal (Higgs) from background. A deep neural network (DNN) is used for the $\tau_l\tau_h$ channel while a boosted decision tree (BDT) from the XGBoost library is employed for the $\tau_h\tau_h$ channel. The classifiers (called *MVA* hereafter) are trained separately for different data-taking years. The features used in the MVA include the transverse momentum of taus, jets and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ as well as the mass of di-tau and of the two leading jets. The output of the MVA is a score for each of the following categories:

- **"Higgs":** This is the signal category. This category includes the ggH, VBF, and VH production modes which are weighted by their cross sections before being merged and used in the training. The $\mathcal{CP}$ of the $H\tau\tau$ coupling is derived using this category.

- **"Genuine":** This category consists of all background events containing two genuine taus in the final state. The dominant background in this category is the Drell-Yan process.

- **"Mis-ID":** This category contains all background processes which include at least one misidentified tau.

The categories are mutually exclusive. Since low values of S_{IP} in the impact parameter method show no $\mathcal{CP}$ sensitivity, a $S_{\mathrm{IP}} > 1.5$ condition is applied to the electrons, muons, and single charged pions in the Higgs category to prevent diluting the sensitivity. In the background categories, the S_{IP} cutoff is only applied to single charged pions.

Fig. 6.8 shows the post-fit MVA score distribution of the Genuine and Mis-ID categories in the $\tau_h\tau_h$ decay channel. Most of the genuine di-tau and jet $\rightarrow \tau_h$ background events are given high MVA scores in the Genuine and Mis-ID categories, respectively, showing the effective performance of the MVA. The data and simulated events agree within uncertainties. The analogous figures for the $\tau_l\tau_h$ channels can be found in [2].

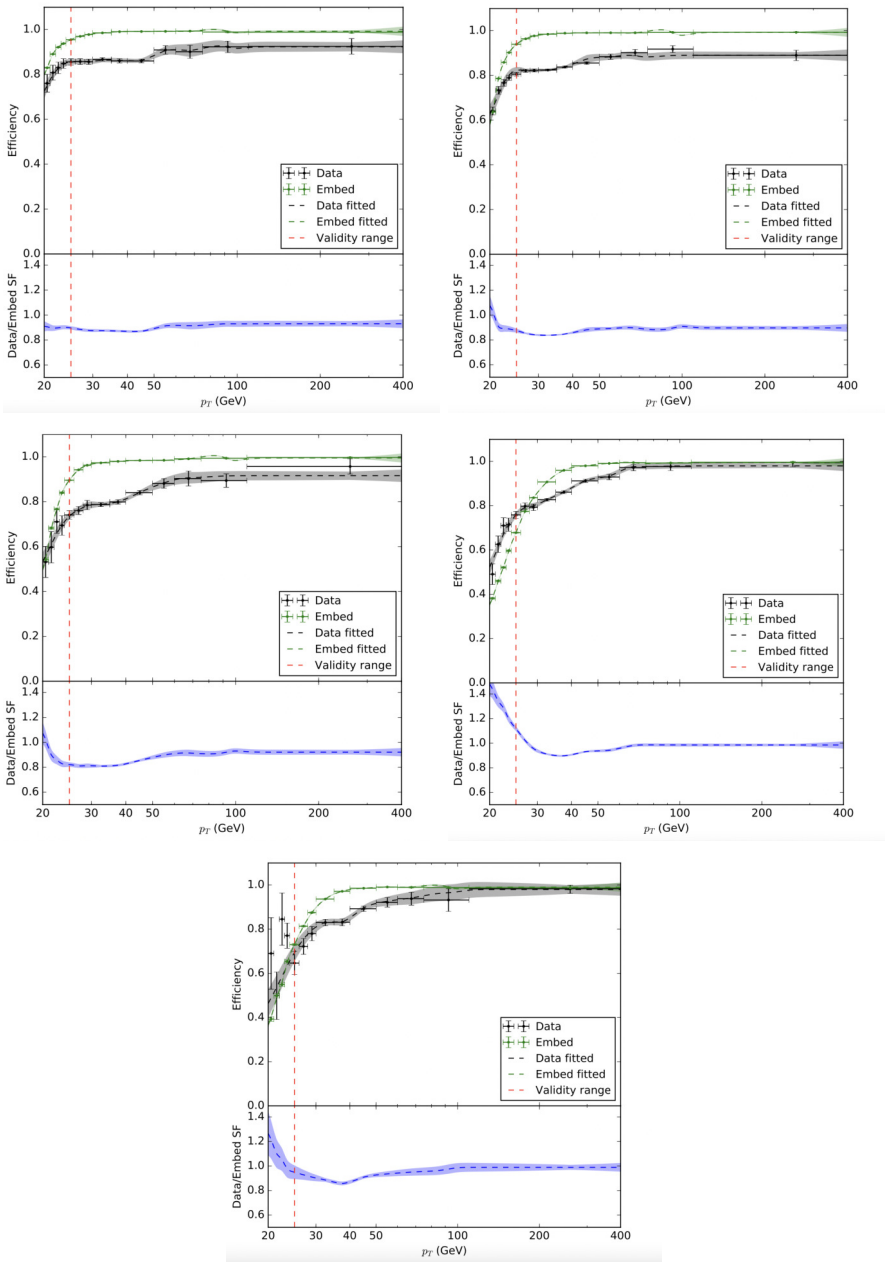

Figure 6.7: The efficiencies and scale factors of the τ_h trigger in the $\tau_\mu\tau_h$ channel as a function of $p_{\mathrm{T}}^{\tau_h}$ using 2016 data and embedded samples for the $\pi^{\pm}$ (top-left), $\pi^{\pm}\pi^0$ (top-right), $\pi^{\pm}2\pi^0$ (middle-left), $3\pi^{\pm}$ (middle-right), and $3\pi^{\pm}\pi^0$ (bottom) MVA decay modes. The efficiency graphs along with their fits and fit uncertainties are shown in the upper panel of each sub-figure while the scale factors are shown in the lower panel. The red dashed line shows the offline threshold below which the events are excluded from the analysis.

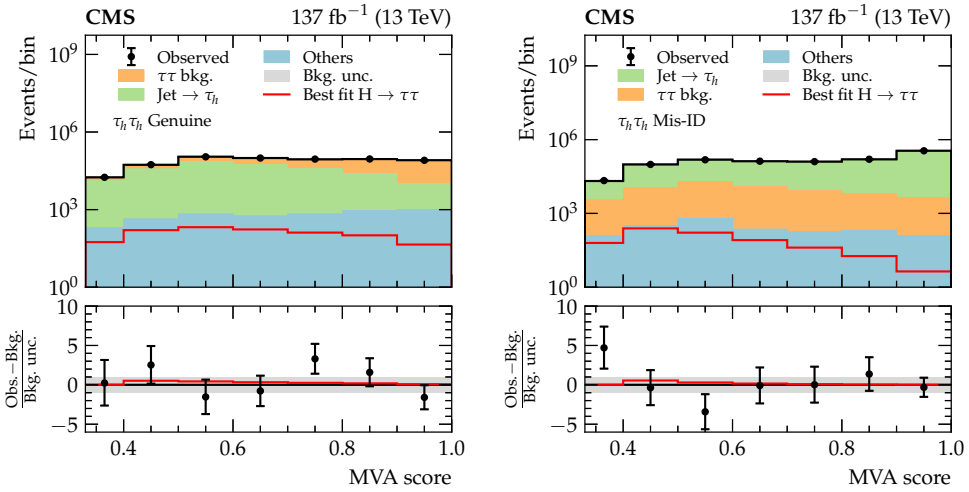

Figure 6.8: The postfit MVA score distribution of the Genuine and Mis-ID categories in the $\tau_h\tau_h$ decay channel. The plots are inclusive in τ_h decay modes. The distributions of the best fit signal are overlaid. The lower panel shows data minus background over background uncertainty. The red curve in the lower panel is derived by dividing the signal by the background uncertainty [2].

6.8 $\phi_{\mathcal{CP}}$ distribution in MVA score bins

To exploit the $\mathcal{CP}$ sensitivity, the $\phi_{\mathcal{CP}}$ distribution in each decay channel is analyzed in bins of the MVA score, because higher MVA scores exhibit larger signal-to-background ratios. These $\phi_{\mathcal{CP}}$ distributions are used to fit to data.

The background distribution estimates, known as background templates, have large fluctuations which can reduce the $\mathcal{CP}$ sensitivity. These fluctuations are reduced with the following techniques. Background templates including two genuine taus are known to have flat $\phi_{\mathcal{CP}}$ distribution at the generator-level (see for example the Drell-Yan distribution in Fig. 2.5). In the neutral-pion and combined methods, the detector effects do not change the shape of these distributions. Therefore, when these two methods are used, we flattened these templates by merging all $\phi_{\mathcal{CP}}$ bins. The jet $\to \tau_h$ background templates are non-flat in all decay channels for kinematic reasons. However, since these templates are symmetric around $\phi_{\mathcal{CP}} = 180°$, we "symmetrized" theses templates by averaging bins with equal distance to $\phi_{\mathcal{CP}} = 180°$. Other backgrounds, such as $\mu \to \tau_h$, are also flat and hence flattened. In the channels where the impact parameter and polarimetric methods are used, the detector smearing effects on the PV are correlated with the $\phi_{\mathcal{CP}}$ variable, leading to a deficit of events around $\phi_{\mathcal{CP}} = 180°$ and an excess about $\phi_{\mathcal{CP}} = 0°$ or $360°$ [47]. These templates, however, remain symmetric around $\phi_{\mathcal{CP}} = 180°$ and are therefore symmetrized. The signal templates ($\mathcal{CP}$-even, $\mathcal{CP}$-

odd, and maximum mixing) are also symmetrized around the peak of their sinusoidal distribution.

Fig. 6.9 displays the $\phi_{\mathcal{CP}}$ distribution of the data and the postfit background templates in bins of MVA score for the $\rho\rho$ and $\mu\pi$ channels. The signal distribution for a pure $\mathcal{CP}$-odd coupling (shown with "PS", meaning *pseudoscalar*) and for the best fit are overlaid. The effect of flattening (symmetrizing) the di-tau background in the $\rho\rho$ ($\mu\pi$) channel is visible in the figure. In addition, a 180° phase shift can be seen in the signal $\phi_{\mathcal{CP}}$ distribution when comparing the $\mu\pi$ and $\rho\rho$ channels. The reasons for this phase shift is the sign flip of the τ-spin analyzing power in the decay channels containing e or μ, as mentioned in section 2.3.2. The figure also demonstrates the effectiveness of the MVA in assigning large MVA scores to the signal events and low scores to the backgrounds. The $\phi_{\mathcal{CP}}$ distributions for other decay channels can be found in the published paper [2].

6.9 Systematic uncertainties

The systematic uncertainties included in this analysis are divided into two types:

- **normalization uncertainties:** affect only the normalization of a distribution.
- **shape uncertainties:** can change the shape of a distribution via correlated changes across bins.

The uncertainties are incorporated into the final fit as nuisance parameters, which are variables that affect the measured value but are not, per se, of interest. A nuisance parameter is allowed to float in the fit and is usually constrained by a function added to the likelihood which penalizes deviations from its nominal expectation.

The systematic uncertainties included in this analysis are briefly explained below while a more detailed explanation can be found in [2].

6.9.1 Normalization uncertainties

The uncertainty in the muon and electron reconstruction efficiencies are 1% and 2%, respectively. A 2% uncertainty is added separately for muons and electrons to account for their trigger efficiency. The uncertainty in the integrated luminosity varies between 2.3 to 2.5% depending on the data-taking year [117–119]. The extracted uncertainty for the cross section of $t\bar{t}$, W+jets, and Z+jets are respectively 4%, 4%, and 2%. A combined uncertainty of 5% is estimated for the single-top quark and diboson production processes using CMS measurements [120, 121].

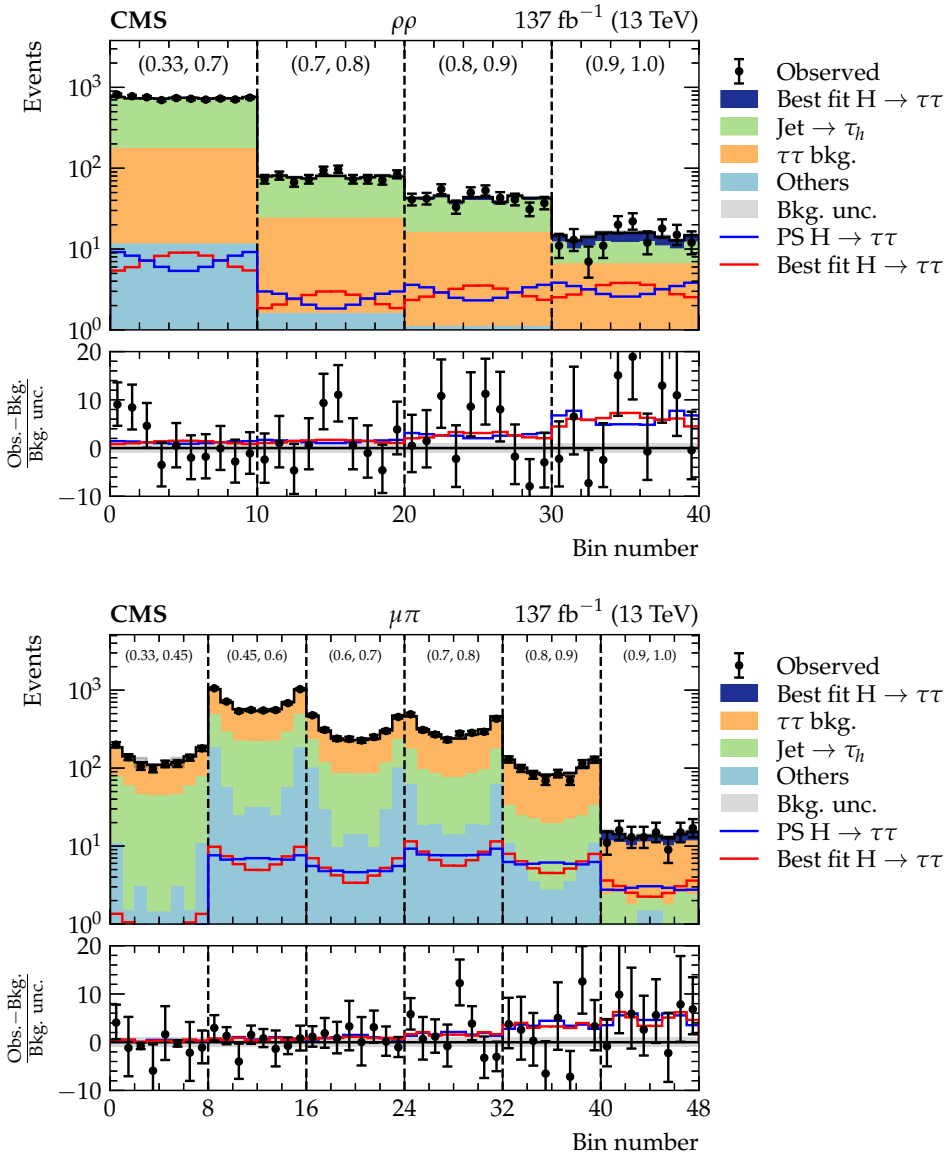

Figure 6.9: The $\phi_{\mathcal{CP}}$ distribution in the $\rho\rho$ (top) and $\mu\pi$ (bottom) channels in bins of MVA score. The $\mathcal{CP}$-odd (shown with "PS", standing for pseudoscalar) and best fit signal distributions are overlaid. The lower panel shows data minus background over background uncertainty along with the pseudoscalar and best fit signals over the background uncertainty [2].

Using the values recommended in Ref. [34], the Higgs boson production cross section and its di-tau decay branching fraction are added to the normalization uncertainties. Separate uncertainties are included to account for the following misidentification rates: $e \to \tau_h$ in $\tau_h\tau_h$, $e \to \tau_h$ in $\tau_e\tau_h$, and $\mu \to \tau_h$ in $\tau_\mu\tau_h$. In the $\tau_l\tau_h$ channel where b-jets are vetoed, 1-9% uncertainty is considered for the $t\bar{t}$ and diboson event yields due to b-jet tagging scale factor. The uncertainty for τ_h identification efficiency against muons and electrons using DNN is set to 3%. The uncertainty of secondary vertex (SV) reconstruction efficiency in the $a_1^{\text{3pr}}a_1^{\text{3pr}}$ channel is 2%. An uncertainty is added for the correction of the S_{IP} in the e, μ, and π decays.

6.9.2 Shape uncertainties

The muon energy scale uncertainty ranges between 0.4 and 2.7% depending on the muon η, while the electron energy scale uncertainty is less than 1%. The (statistically dominated) uncertainty of the τ_h identification and reconstruction efficiency as well as its trigger efficiency are measured as a function of the p_{T} and MVA decay mode, as explained in section 6.6. The $e \to \tau_h$ and $\mu \to \tau_h$ misidentification energy scale uncertainties are included. The uncertainties on jet energy scale, jet energy resolution, and τ_h energy scale are also incorporated. Fake factors, F_F, are measured as a function of different parameters such as p_{T}. A function is then fitted to the F_F values, where the uncertainty of the fit is included as a shape uncertainty (the procedure is similar to the method described in Ref. [28]). In the 2016-2017 data-taking periods, the forward endcap region ($2.5 < |\eta| < 3.0$) showed an inefficiency due to a timing issue in the Level-1 trigger; a specific uncertainty is added to cover this issue. The shape uncertainties also include parton showering as well as renormalization and factorization scales uncertainties for the signal samples. Statistical fluctuations of the templates are included using the "Barlow-Beeston" method [122, 123] which provides one nuisance parameter per bin. The bin-by-bin uncertainties of the flattened bins are all fully correlated while in the symmetrized bins, only bin pairs that are used for symmetrization are fully correlated. The uncertainties of hadronic recoil resolution and response, jet energy resolution and scale, as well as unclustered energy are all propagated to $\vec{p}_{\text{T}}^{\text{miss}}$ and included as shape uncertainties.

6.10 Results

The mixing angle $\alpha^{H\tau\tau}$ is derived using a simultaneous fit to the data with a likelihood of the form

$$L(\mathcal{L}, \vec{\mu}, \alpha^{H\tau\tau}, \vec{\theta}) = \prod_{j}^{N_{\text{categories}}} \prod_{i}^{N_{\text{bin}}} P\left(n_{i,j} \mid S_{i,j}(\mathcal{L}, \vec{\mu}, \alpha^{H\tau\tau}, \vec{\theta}) + B_{i,j}(\vec{\theta})\right) \times \prod_{m}^{N_{\text{nuisance}}} C_m(\vec{\theta}). \tag{6.2}$$

The signal $S_{i,j}(\mathcal{L}, \vec{\mu}, \alpha^{H\tau\tau}, \vec{\theta}) = \mathcal{L}\ \vec{\mu}\ \vec{A}_{i,j}(\vec{\theta}, \alpha^{H\tau\tau})$ is derived by multiplying the integrated luminosity $\mathcal{L}$ by the signal strength modifiers $\vec{\mu} = (\mu_{ggH}, \mu_{qqH})$ and the signal acceptance $\vec{A}_{i,j}$ in bin i category j. The Poisson distribution P calculates the probability of observing $n_{i,j}$ given the sum of the signal $S_{i,j}$ and background $B_{i,j}$ expectations. The signal strength modifiers $\vec{\mu} = (\mu_{ggH}, \mu_{qqH})$, defined as the multiplication of the Higgs cross section and $H \to \tau^+\tau^-$ branching fraction with respect to the SM prediction, modify the ggH and qqH processes as free parameters in the fit, where qqH combines VBF and VH processes. The qqH and ggH processes can be distinguished in the fit because the VBF distribution is more pushed towards the higher MVA scores in the signal category (due to its distinctive topology) while the ggH process has a broader MVA score distribution.

The systematic uncertainties are incorporated as nuisance parameters $\vec{\theta}$ which are constrained via C_m functions. A log-normal (Gaussian) distribution is assumed and included in the constraint term C_m for the normalization (shape) uncertainties. The inputs to the likelihood for the signal category are the $\phi_{\mathcal{CP}}$ distributions in bins of MVA score while for the background categories, the inputs are the MVA score distribution (compare Figs. 6.9 and 6.8). Adding background categories to the likelihood helps to further constrain the background contribution and the associated systematic uncertainties. To compute confidence intervals, a negative log-likelihood is defined as

$$-2\Delta \ln L(\alpha^{H\tau\tau}) = -2\left(\ln\left(L(\alpha^{H\tau\tau})\right) - \ln\left(L(\alpha^{H\tau\tau}_{\text{best fit}})\right)\right), \tag{6.3}$$

where the values $-2\Delta \ln L = 1.00$, 4.02, and 8.81 correspond to 68.3, 95.5, and 99.7% confidence levels (CL).

6.10.1 Measured value of $\alpha^{H\tau\tau}$

Fig. 6.10 shows the observed and expected negative log-likelihood scan of the $\alpha^{H\tau\tau}$ for the combination of the $\tau_e\tau_h$, $\tau_\mu\tau_h$, and $\tau_h\tau_h$ channels. The observed (expected) value for the $\alpha^{H\tau\tau}$ is found to be $-1 \pm 19°$ ($0 \pm 21°$) at 1σ CL which excludes a pure $\mathcal{CP}$-odd $H\tau\tau$ coupling at 3.0σ (2.6σ) CL. The uncertainty of the observed value of $\alpha^{H\tau\tau}$ can

be decomposed into statistical, experimental systematic, bin-by-bin, and theoretical uncertainties:

$$\alpha^{H\tau\tau} = (-1 \pm 19\,\text{stat} \pm 1\,\text{syst} \pm 2\,(\text{bin-by-bin}) \pm 1\,(\text{theory}))^{\circ}. \tag{6.4}$$

The result is fully statistically dominated and compatible with the SM prediction within experimental uncertainties. The subdominant uncertainty is the bin-by-bin uncertainty corresponding to the fluctuation of the background templates. Since most of the templates are data-driven, the limiting factor for the bin-by-bin uncertainty is the size of the data used for producing the templates. The next dominant uncertainties are the τ_h trigger efficiency, theory uncertainties, and τ_h energy scale.

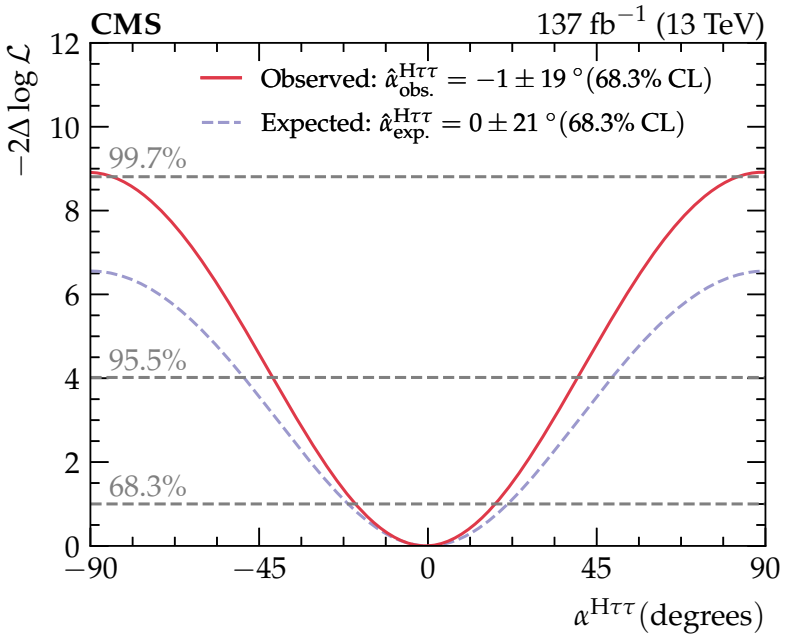

Figure 6.10: The negative log-likelihood scan of the mixing angle $\alpha^{H\tau\tau}$ [2]. See the text for more information.

The best fit values for the signal strength modifiers are $\mu_{ggH} = 0.59^{+0.28}_{-0.32}$ and $\mu_{qqH} = 1.39^{+0.56}_{-0.47}$ which have a $\rho = -0.76$ correlation coefficient. The reason for the large anti-correlation is that the analysis is not optimized to distinguish these production processes.

The sensitivities of the $\tau_h\tau_h$, $\tau_\mu\tau_h$, and $\tau_e\tau_h$ channels are 1.8σ, 1.5σ and 1.0σ, respectively. The most sensitive modes are $\mu\rho$, $\rho\rho$ and $\pi\rho$ with 1.2σ, 1.1σ, and 1.0σ, respectively.

In the κ_τ and $\tilde{\kappa}_\tau$ parameterization, a 2D NLL scan is performed while keeping other

Yukawa couplings equal to their SM value (see Fig. 6.11). In the 2D case, the 68.3, 95.5, and 99.7% CLs correspond to $-2\Delta \ln L_{2D} = 2.30$, 6.20, and 11.62, respectively. This analysis is sensitive only to the ratio of the signs of the κ_τ and $\tilde{\kappa}_\tau$ couplings (rather than individual signs) and hence there are two global minima for the NLL.

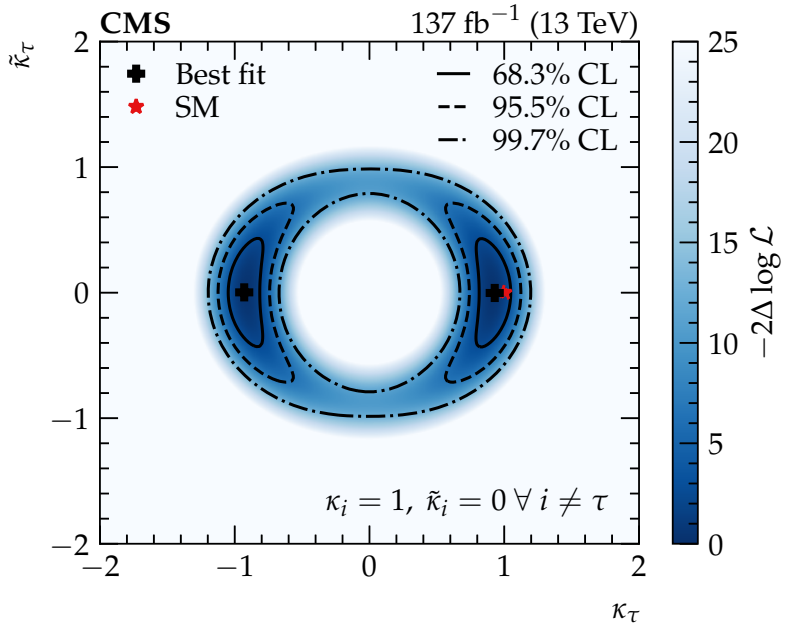

Figure 6.11: The negative 2D log-likelihood scan of the κ_τ and $\tilde{\kappa}_\tau$ Yukawa couplings, while other kappas are set to the SM value [2].

A separate fit is performed to the data by assuming $\mu_{ggH} = \mu_{qqH} = \mu$, in which μ is defined as the total signal strength modifier which scales the multiplication of the total Higgs production cross section and the $H \to \tau^+\tau^-$ branching fraction divided by the SM value. The 2D NLL scan of $\alpha^{H\tau\tau}$ and μ is shown in Fig. 6.12. No strong correlation is seen between the two parameters.

The compatibility of the data with a $\mathcal{CP}$-even coupling can be seen more clearly in Fig. 6.13, which displays the *weighted* $\phi_{\mathcal{CP}}$ distributions of the four most $\mathcal{CP}$-sensitive decay channels, namely $\rho\rho$, $\pi\rho$, $\mu\rho$, and $e\rho$. The $\phi_{\mathcal{CP}}$ distributions of the MVA score bins belonging to these decay channels are combined, with each distribution weighted as explained below. (See Fig. 6.9 for examples of $\phi_{\mathcal{CP}}$ distributions in the MVA score bins.) The data is weighted after the background is subtracted and the uncertainty on the subtracted background is calculated and displayed in Fig. 6.13. The weights are calculated separately for each MVA score bin and used in order to provide a clear visual contrast between $\mathcal{CP}$ scenarios. The definition of the weight for the MVA score

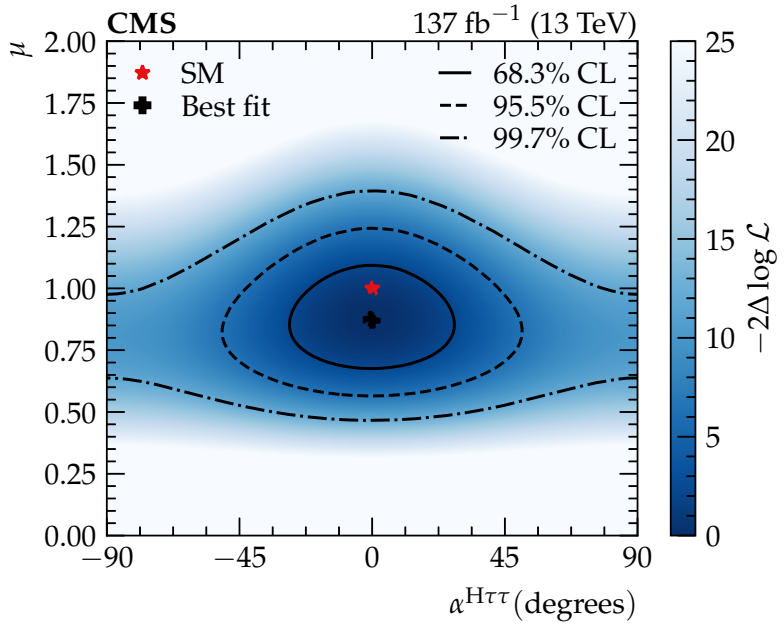

Figure 6.12: The negative 2D log-likelihood scan of the mixing angle $\alpha^{H\tau\tau}$ and signal strength modifier μ. No strong correlation is seen [2].

bin j is

$$W_j = A_j \times \frac{S_j}{S_j + B_j}, \tag{6.5}$$

where S_j and B_j are the number of signal and background events in the MVA score bin, respectively, while A_j shows an average asymmetry between the $\mathcal{CP}$-even and $\mathcal{CP}$-odd scenarios. The A_j is defined as

$$A_j = \frac{1}{N_j^{\text{bins}}} \sum_i \frac{|\mathcal{CP}_i^{\text{even},j} - \mathcal{CP}_i^{\text{odd},j}|}{\mathcal{CP}_i^{\text{even},j} + \mathcal{CP}_i^{\text{odd},j}}, \tag{6.6}$$

where $\mathcal{CP}_i^{\text{even},j}$ and $\mathcal{CP}_i^{\text{odd},j}$ represent the expected number of $\mathcal{CP}$-even and $\mathcal{CP}$-odd events in the $\phi_{\mathcal{CP}}$ bin i of the MVA score bin j, respectively, while N_j^{bins} is the number of $\phi_{\mathcal{CP}}$ bins in the MVA score bin j.

Adding A_j to the standard $S_j/(S_j + B_j)$ weight when combining the distributions is important to maximize the separation between the $\mathcal{CP}$-even and $\mathcal{CP}$-odd distributions in Fig. 6.13. As can be seen from Eq. 2.17, the signal (S_j) includes a constant term, insensitive to the $\mathcal{CP}$ nature of the Higgs-tau coupling, and a cosine term with an amplitude which determines the separation between the $\mathcal{CP}$-even and $\mathcal{CP}$-odd distributions. This amplitude varies with the $H \to \tau^+\tau^-$ decay channels and therefore, a larger S_j in a decay channel does not necessarily imply a larger separation in that

channel. In contrast, A_j directly reflects the separation for each decay channel and each MVA score bin.

As an example of calculating W_j, in the $\rho\rho$ decay channel shown in Fig. 6.9 (top), the MVA score bin j=(0.9, 1.0) includes 10 $\phi_{\mathcal{CP}}$ bins, i.e. $N_j^{\text{bins}} = 10$. The $\mathcal{CP}$-odd distribution is overlaid on the histogram and shown with a blue curve; $\mathcal{CP}_i^{\text{odd},j}$ is the value at bin i of this distribution. The values of $\mathcal{CP}_i^{\text{even},j}$ are extracted similarly. The $\mathcal{CP}$-even distribution is not shown in the figure but it is very similar to the "best fit" distribution shown in a red curve, given that the result of this measurement was found to be SM-like. The S_j and B_j can be extracted by integrating over the signal (dark blue) and background (other colours) contributions to the histogram, respectively.

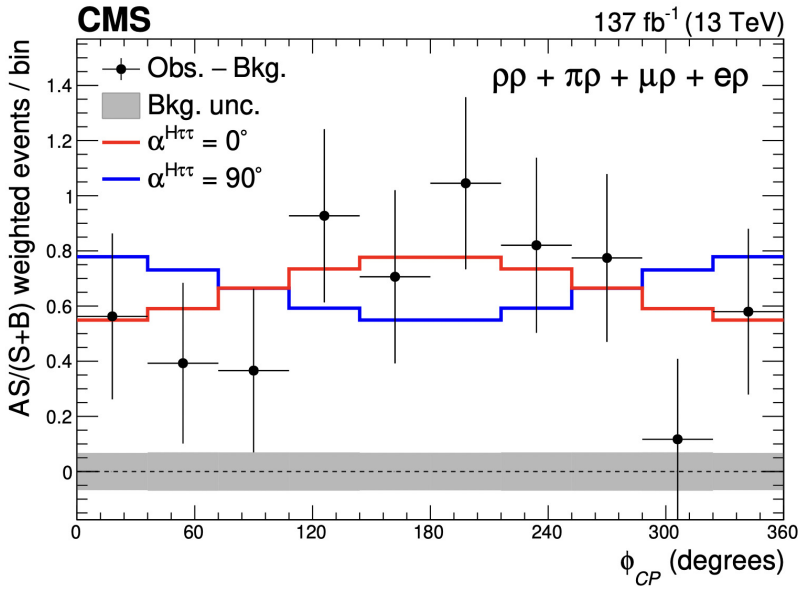

Figure 6.13: Weighted $\phi_{\mathcal{CP}}$ distributions of the background-subtracted data and the predictions of the $\mathcal{CP}$-even (red) and $\mathcal{CP}$-odd (blue) distributions displayed for the most $\mathcal{CP}$-sensitive decay channels, i.e. $\rho\rho$, $\pi\rho$, $\mu\rho$, and $e\rho$. The uncertainty of the background is shown in grey. The best fit value of the signal strength is used for calculating the yield of the predicted $\mathcal{CP}$ distributions. For the reason described in section 2.3.2, when combining the decay channels, a 180° phase shift is applied to the decay channels with e or μ in the final state. The data favours the $\mathcal{CP}$-even scenario [2]. More information on how the figure is produced can be found in text.

Due to statistical fluctuations, the amplitude and phase of the sinusoidal $\phi_{\mathcal{CP}}$ distribution of the data may not perfectly match those of the simulated events. In general, a larger (smaller) amplitude shows more (less) sensitivity in discriminating $\mathcal{CP}$ scenarios and less (more) uncertainty for $\alpha^{H\tau\tau}$. The final uncertainty of $\alpha^{H\tau\tau}$ is derived from the likelihood function which receives contributions from the $\phi_{\mathcal{CP}}$ distributions of all MVA

score bins. As mentioned earlier, the observed and expected uncertainties of $\alpha^{H\tau\tau}$ were found to be 19° and 21°, respectively, meaning that a greater contribution was received from the MVA score bins in which the amplitude is larger in data than in simulated events. The difference in the amplitudes of the data and simulated events might be noticed[2] in Fig. 6.13.

[2]Since the observed and expected uncertainties are comparable, the amplitude difference is hardly noticable by eye. The reader can find a larger amplitude difference in a similar plot in Ref. [125], the preliminary version of this analysis (in which the $\tau_e\tau_h$ decay channel was not included). The observed (expected) uncertainty in that analysis was found to be 17° (23°).

Chapter 7

The high granularity calorimeter

The LHC will increase luminosity in the next upgrade known as *the high-luminosity LHC* (HL-LHC) [126], which leads to a higher amount of radiation in the detector, especially in the forward region. Therefore, the current endcap ECAL and HCAL calorimeters need to be replaced with a radiation-tolerant detector. The proposed detector is called *the high granularity calorimeter* (HGCal) [127], which not only can sustain the harsh radiation environment, but also is highly granular to be able to distinguish the large number of particles produced in the forward region. In addition, the Level-1 Trigger (L1T) will be upgraded in order to accommodate dedicated devices and algorithms for handling the increased number of particles in the detector and identifying interesting physics processes.

Using the HGCal detector can enhance several physics analyses, in particular the Higgs boson analyses. The $H \to \gamma\gamma$ decay channel, thanks to its clean signal in the detector, was one of the most important channels in the Higgs boson discovery and has been one of the best channels for measuring the properties of this particle. With the HGCal detector in the forward region, the efficiency of this analysis increases by 12% as the photons in the $2.5 < |\eta| < 3.0$ region will be included. Additionally, since the high granularity of this detector allows a precise measurement of jet properties, the VBF and ggH processes can be well discriminated. Another interesting channel is $H \to \tau^+\tau^-$ as it provides the best direct sensitivity to the Yukawa couplings between the Higgs boson and leptons. A leading variable for distinguishing this process from $Z \to \tau^+\tau^-$ background is the visible mass of di-tau, of which resolution depends on a precise reconstruction of jets. Due to the HGCal high precision in reconstructing jets, the resolution of this mass variable was shown to be similar in Run-2 and the HL-LHC despite the high pile-up conditions of the HL-LHC. In addition, a study of the di-Higgs production in the bb$\tau\tau$ decay channel showed that the larger acceptance region provided by the HGCal improves the VBF and ggH processes in the signal region

by 10-15% and 4-8%, respectively [127].

In this chapter, after describing the HL-LHC, the upgrade plan of CMS for the HL-LHC era is briefly discussed while the HGCal detector is more elaborated on. After that, the architecture of the proposed L1T upgrade is explained with a focus on the role of the HGCal *trigger primitives generator* (TPG). This is followed by presenting the algorithm I developed to improve jet identification and reconstruction with trigger towers. Finally, the position and energy resolutions of jets with the new algorithm are compared with the previous algorithm.

7.1 The high-luminosity LHC

The LHC running periods were briefly described in section 3.1. The integrated luminosity of the data collected by CMS so far is plotted as a function of time in Fig. 7.1. The Run-1 of the LHC (2010-12), which lead to the discovery of the Higgs boson, provided $\approx$6 fb^{-1} integrated luminosity during 2010-11 at a centre-of-mass energy of $\sqrt{s} = 7$ TeV and $\approx$23 fb^{-1} in 2012 at $\sqrt{s} = 8$ TeV. The Run-2 of the LHC (2015-18) operated at $\sqrt{s} = 13$ TeV and collected a further $\approx$164 fb^{-1} integrated luminosity[1]. Thus, the total integrated luminosity collected by the LHC to date is $\approx$193 fb^{-1}. The instantaneous luminosity during Run-2 reached a maximum of 1.7×10^{34} cm^{-2} s^{-1} which is 1.7 times the LHC design luminosity.

By the end of the Run-3, which is planned to operate during 2022-25, the total integrated luminosity of the LHC is expected to reach $\approx$450 fb^{-1}. Afterwards, the LHC will be upgraded to the HL-LHC, during which the instantaneous luminosity will be 5 to 7×10^{34} cm^{-2} s^{-1}. The total integrated luminosity collected by the LHC machine will reach 3000 to 4000 fb^{-1} by the end of the HL-LHC (late-2030s). Fig. 7.2 shows the schedule for the LHC and HL-LHC programs, a.k.a. the Phase-1 and Phase-2 of the LHC, respectively.

7.2 Phase-2 CMS upgrade

In the HL-LHC, the average number of simultaneous pp collisions per bunch crossing, known as pile-up, will be 140 while it can be as high as 200 when operating at the maximum capacity [65]. Several changes to the CMS detector are considered for the Phase-2 upgrade [130] in order to maintain (or even improve) the physics performance of the detector in the high pile-up condition. These changes are summarized below:

[1]The collected data is not 100% efficient for physics analysis. That's why only 137 fb^{-1} integrated luminosity from Run-2 was available for the Higgs $\mathcal{CP}$ analysis explained in chapter 6.

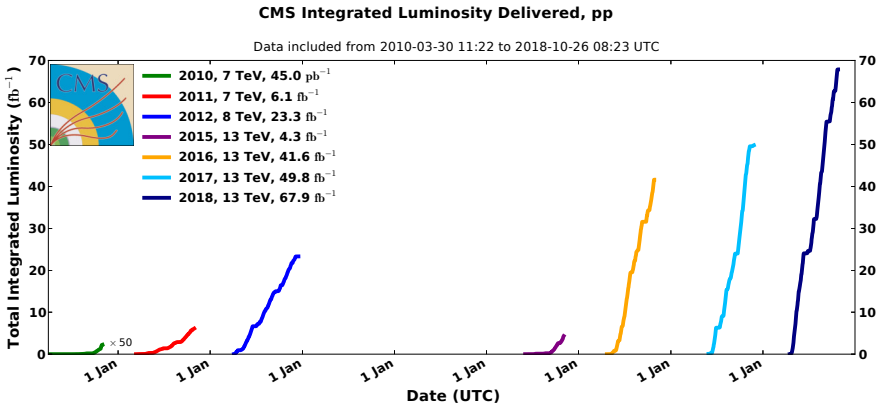

Figure 7.1: The integrated luminosity collected by CMS as a function of time [128].

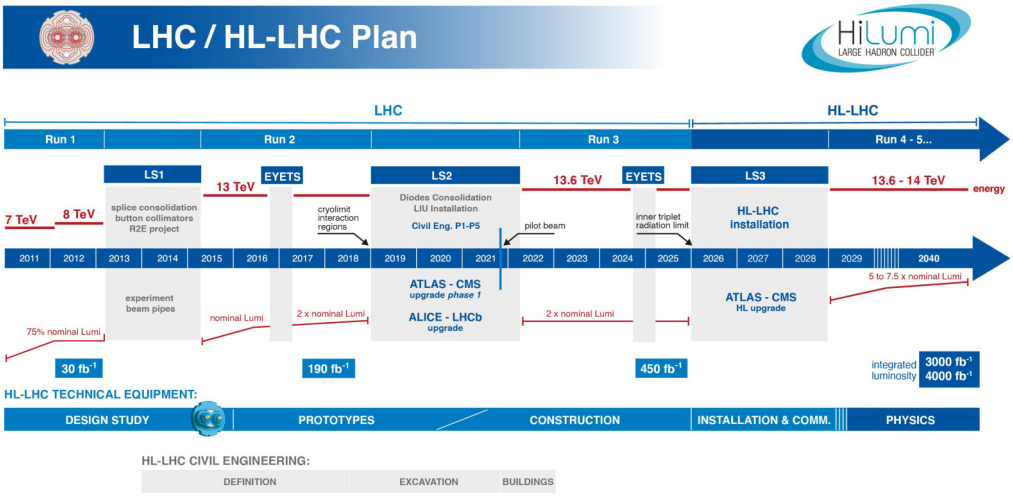

Figure 7.2: The plan for the LHC and HL-LHC [129].

- **Tracker:** The tracker will be fully replaced by a new tracker which is more radiation-tolerant and more granular. It also covers a larger acceptance range and has less material budget. Moreover, the new tracker is designed to allow track information in the L1T.

- **Calorimeters:** The electronics of the EB need to be fully replaced to meet the L1T requirement. The ECAL crystals in the EB do not need to be replaced as the low-η region suffers less from radiation. For instance, the expected radiation damage to the ECAL crystals located in the highest-η region of the EB after 3000 fb^{-1} is equal to parts of ECAL crystals in the EE after 30 fb^{-1}. In the HB, a subset of scintillator tiles will be replaced with more radiation-tolerant ones. In contrast, both ECAL and HCAL endcap calorimeters will be replaced with the HGCal, which allows for precise position, energy, and time measurement of showers thanks to its high transverse and longitudinal granularity.

- **Muon system:** The forward region will be equipped with additional muon subsystems, namely Gas Electron Multipliers (GEM) and improved RPCs (iRPC), to increase redundancy, improve muon reconstruction, and enhance trigger performance. In order to take advantage of the extended coverage of the tracker, the GEM subsystem extends the pseudorapidity coverage of the muon system to $\eta \approx 3$.

- **MIP timing detector:** A new detector, called *MIP timing detector* (MTD) [131], will be installed between the tracker and calorimeters to precisely measure the timing information of minimum ionizing particles (MIPs) in the barrel and endcaps. The MTD can significantly impact the CMS physics program in the HL-LHC by allowing several improvements such as in vertex identification and $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ reconstruction through pile-up rejection.

- **Trigger and data acquisition:** The latency of the L1T at Phase-2 will increase to 12.5 µs and the maximum event selection rate will increase to 750 kHz. Therefore, the trigger and the data acquisition will be upgraded to accommodate more complex algorithms and handle the higher data rate.

7.3 The high granularity calorimeter

7.3.1 Essential features

The endcap calorimeters will be replaced in the HL-LHC with a radiation-tolerant detector, namely the HGCal, due to the ten-fold increase in the luminosity. The existing

calorimeters in the endcaps (i.e. $PbWO_4$ crystals in the EE and sampling calorimeters based on plastic scintillators in the HE, as described in chapter 3) are designed for a maximum integrated luminosity of 500 fb^{-1}, beyond which the physics performance degrades significantly.

Fig. 7.3 shows the level of ionizing radiation accumulated in the HGCal after 3000 fb^{-1} integrated luminosity. The dose will be as high as 2 MGy and the fluence will reach 10^{16} n_{eq}/cm^2, where n_{eq}/cm^2 indicates the number of 1 MeV equivalent neutrons per square centimeter. R&D studies have shown [127] that silicon sensors are capable of retaining sufficient charge collection after a fluence of 1.5×10^{16} n_{eq}/cm^2, 50% higher than the maximum fluence after 3000 fb^{-1}. Therefore, silicon sensors are used as active material in the front section of the HGCal, which is more exposed to radiation. In order to minimize the level of electronic noise after such irradiation, the sensors are kept at a low temperature of $-30°C$. In the rear section of the HGCal, which is less exposed to radiation, plastic scintillator tiles are used with silicon photomultipliers (SiPM) as readout. Likewise, the SiPMs operate optimally at a low temperature and hence the temperature of the whole HGCal is kept at $-30°C$.

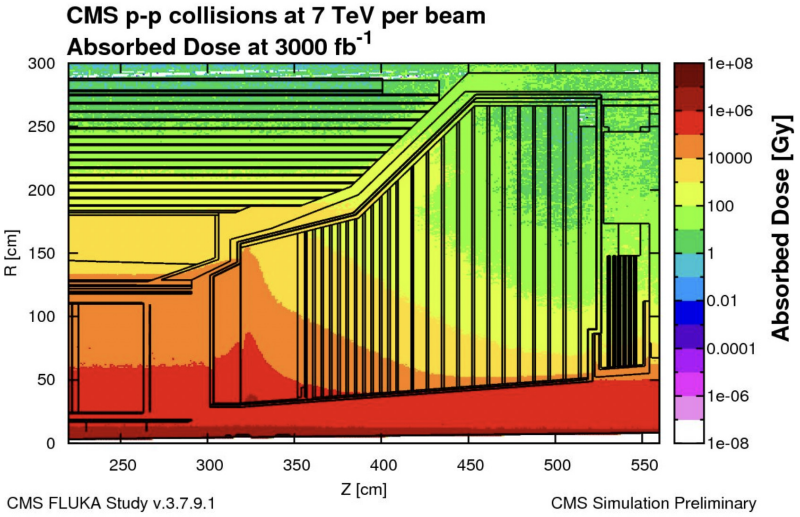

Figure 7.3: The absorbed dose of the HGCal after 3000 fb^{-1} integrated luminosity for pp collisions at a centre-of-mass energy of $\sqrt{s} = 14$ TeV [127].

The proposed HGCal design has several features which are essential to achieve the physics goals of the HL-LHC. Due to the dense structure of the HGCal, the lateral spread of particle showers is small in this detector. The central feature of the HGCal is its high granularity (as is implied by its name), which is important for performing particle-flow algorithms and extracting features from showers. The longitudinal granularity enables a great power of pattern recognition, pile-up discrimination, and

electromagnetic energy measurement. The lateral granularity helps separate two nearby showers, identify narrow jets, and minimize pile-up inclusion in energy measurement. In addition, the HGCal allows for a precise timing measurement in the forward region, which is crucial for pile-up mitigation. The detector can maintain energy resolution even after full lifetime of the HL-LHC because of its radiation tolerance. Another feature of the HGCal design is the ability to contribute to the L1T.

7.3.2 Design

A cross sectional view of the upper-half of the HGCal is shown in Fig. 7.4. This detector covers the $1.52 < |\eta| < 3$ region and includes an electromagnetic calorimeter and a hadronic calorimeter, which are known as the CE-E and CE-H, respectively.

The CE-E is comprised of 28 silicon layers with a total thickness of 34 cm, a radiation length of $26X_0$, and an interaction length of 1.3 λ. The information from every other layer of the CE-E (including the first layer) is used for triggering. The silicon layers consist of smaller units, known as *silicon modules*, which are hexagonal 163 mm wide silicon sensors fabricated on 8 inch (8") wafers. The absorber for each module consists of a base plate made of copper-tungsten alloy and a cooling plate made of copper. Modules are chosen to be hexagonal to most efficiently use the wafer areas which are initially produced in large circular shapes.

The CE-H includes 22 layers[2] with stainless steel absorber of thickness 35 mm in the first 12 layers and of thickness 68 mm in the remaining 10 layers. The active material in the inner (outer) region of the CE-H is silicon modules (plastic scintillator tiles). Unlike the silicon modules, the plastic scintillators are projective to $\eta-\phi$ plane. The CE-H adds a further 8.5 interaction lengths (λ) to the HGCal [132].

The silicon sensors are made in thicknesses of 300, 200 and 120 µm, with thinner sensors used in regions with higher fluence. The 120 µm silicon modules are made up of hexagonal *sensor cells* (SC) with 0.52 cm^2 area while SCs in the rest of the modules have an area of 1.18 cm^2. The former (latter) modules are called *high density* (*low density*) modules. The cell size is driven by physics performance and the fact that the cell capacitance needs to be within a manageable range. SCs are the most basic unit of the HGCal silicon section.

As illustrated in Fig. 7.5, groups of nine (four) nearby SCs in the high (low) density modules form *trigger cells* (TCs). TCs are ingredients for making trigger primitives (defined in section 3.8.1) in the L1T. Both high and low density silicon modules have

[2]The design of the HGCal is still not finalized. Therefore, some of the numbers may not be the same as in the HGCal technical desing report (TDR) [127] or in the final system. The numbers used in this thesis correspond to the simulation geometry used for the results in this chapter.

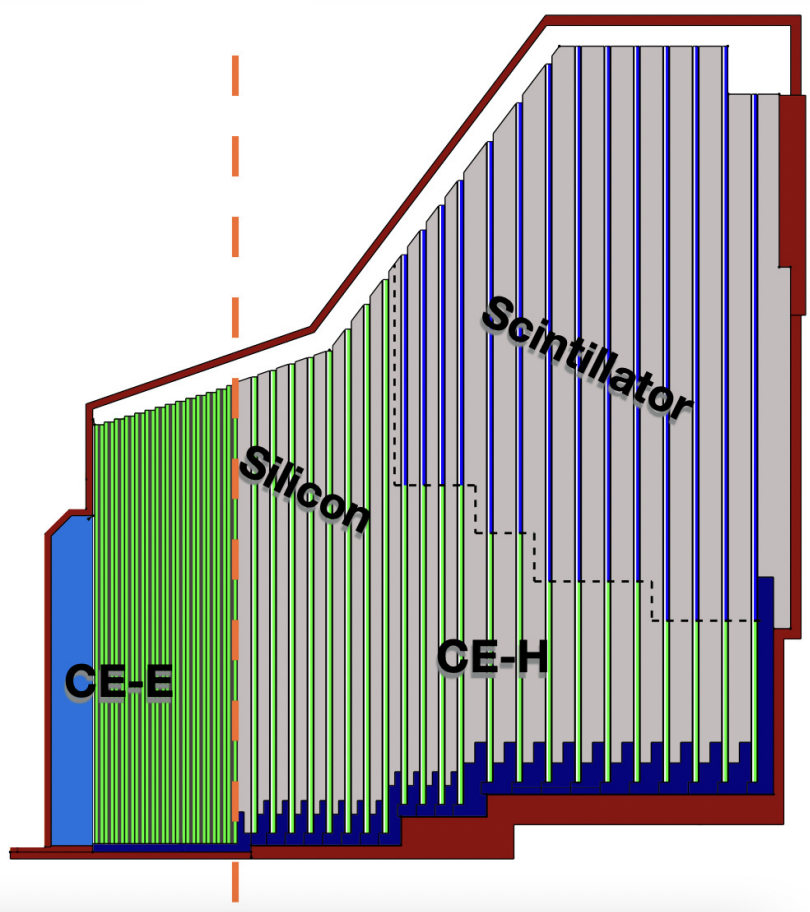

Figure 7.4: A cross sectional view of the upper-half of the HGCal in one of the endcaps. The CE-E and CE-H sections are separated by the orange dotted line. The CE-E and CE-H consist of 28 and 22 layers, respectively, interleaved with absorber layers. The active material in the CE-E is silicon sensors (green). In the CE-H, silicon sensors (green) are only used in the regions with higher level of radiation while in the low-radiation regions plastic scintillators (dark blue) are used as the active material. The neutron moderator layer (light blue) is located closer to the pp interaction point compared to the CE-E and CE-H sections [132].

48 TCs, apart from non-hexagonal *partial* modules found in space-constrained regions such as the edges of the HGCal. Likewise, TCs in the scintillator section are defined as groups of scintillating tiles, as described in section 7.5.

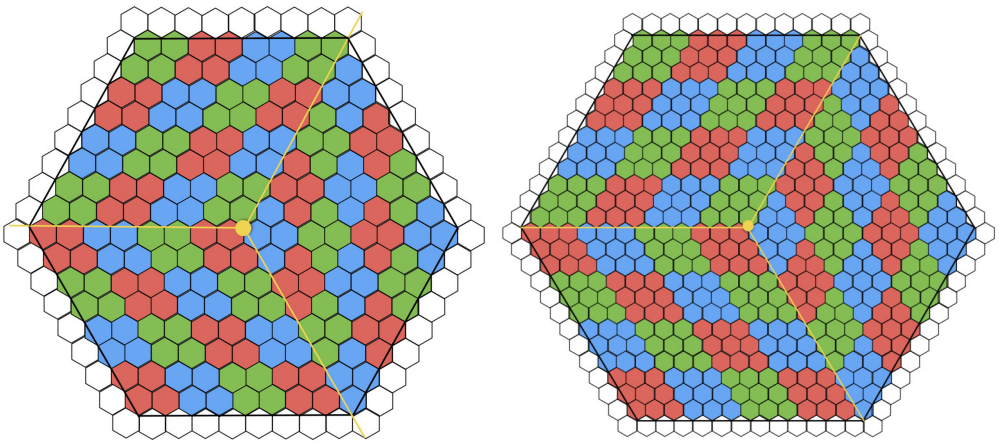

Figure 7.5: A low (left) and a high (right) density silicon module. The size of both types of modules are the same (163 mm wide) and both have 48 trigger cells. Trigger cells are made of four (nine) sensor cells in the low (high) density modules and are shown with different colours [127].

7.4 Phase-2 L1T upgrade

The unprecedented high instantaneous luminosity and hence pile-up in the HL-LHC calls for a major upgrade in the L1T architecture in order to retain signal selection efficiency and enhance new physics signature identification. In the CMS Phase-2, the maximum rate of the L1T will increase from 100 to 750 kHz. The L1T latency will rise from 3.8 to 12.5 µs in order to handle the rate increase and to allow the inclusion of the tracker and HGCal information. Furthermore, a longer latency enables more complex algorithms, such as particle-flow reconstruction techniques, to be performed.

To achieve these goals, the design of the Phase-2 L1T utilizes cutting-edge technology. State-of-the-art field-programmable gate array (FPGA) devices are extensively used for processing (identifying, reconstructing, etc.) trigger objects with the high granularity information of the HGCal taken as input. Additionally, high-speed optical links enable aggregating data from the entire detector into the same processing board, which allows for reconstructing global quantities such as $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ and selecting certain topologies such as two jets with a large pseudorapidity difference in the VBF Higgs production mode. Furthermore, a flexible and modular architecture is implemented in the proposed L1T design to allow later reconfiguration depending on the HL-LHC running conditions.

Fig. 7.6 displays a diagram of the dataflow and architecture of the CMS Phase-2 L1T. There are two key changes compared to the Phase-1 L1T: (i) using the tracker information for the first time in the L1T and (ii) introducing the *Correlator Trigger*, which employs particle-flow techniques by including the calorimeter, muon and tracker information.

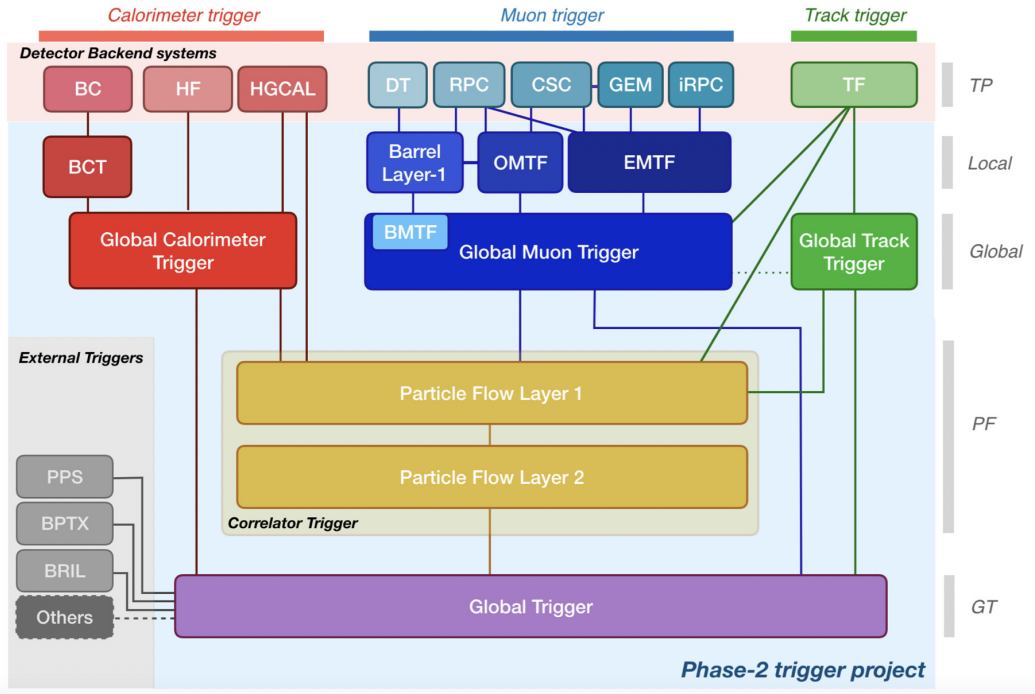

Figure 7.6: The dataflow of the Phase-2 CMS Level-1 trigger. The HGCal is near the top left corner [133].

There are four triggering paths in the architecture: calorimetry, muon systems, tracking, and particle-flow techniques, where the first three use the information of the detectors individually while the last one takes advantage of combining them. With this strategy, complementary types of trigger objects can be generated to achieve optimal flexibility in identifying various physics signatures in the detector as well as allowing efficiency determination from orthogonal triggers. The final decision is made in the *Global Trigger* (GT), which receives inputs from the *Global Calorimeter Trigger* (GCT), *Global Muon Trigger* (GMT), *Global Track Trigger* (GTT), and *Correlator Trigger* (CT), corresponding to the four triggering paths, respectively. These sections of the L1T, which utilize trigger primitives (TPs) for making a trigger decision, are sometimes referred to as the *central L1T*. This is contrasted with the TP generators (TPG) which generate TPs for the use of the central L1T. (See section 3.8 for more information on TPs.) Further information on the Phase-2 L1T can be found in its dedicated TDR [133].

In the remainder of this chapter, the focus will be on the role of the HGCal in the calorimetry triggering path.

7.5 The HGCal in the calorimetry triggering path

This section describes how the HGCal contributes to the L1T based on the calorimetry triggering path.

7.5.1 The HGCal trigger primitive generator

On-detector (a.k.a *Front-end*, or FE for short) electronic systems process signals from particles hitting the HGCal and prepare them for the L1T. The prepared data are then sent to off-detector (a.k.a *Back-end*, or BE for short) electronic systems for further processing. In the BE, the data is processed in the HGCal *trigger primitive generator* (HGCal TPG), which generates TPs for the central L1T, as the name suggests. The TPs generated by the HGCal TPG are duplicated and a copy is sent to each of the GCT and CT to perform L1T algorithms. The TPs processed in the CT are very briefly discussed while those processed in the GCT are the main focus of this chapter.

The basic ingredients for the HGCal TPG are TCs, which are defined differently in the silicon and scintillator sections. In the silicon section, a TC is a group of hexagonal SCs (as shown in Fig. 7.5) with a combined granularity of approximately 4 cm^2. In contrast, a TC in the scintillator section is a group of scintillating tiles with a combined azimuthal (ϕ) angle of 2.5 degrees corresponding to a side of about 4 to 10 cm in ϕ and with a similar size in r, where r is the distance to the beam axis.

Due to bandwidth limitation, not all TCs can be sent to the HGCal TPG and, instead, only those above a certain p_{T} threshold are sent. To have a more realistic estimate of the total deposited energy, the sum of the energy of all TCs in each module, called a *module sum*, is also sent to the HGCal TPG. The number of TCs in a silicon module is 48 (apart from partial modules) and a comparable number of TCs exist in a scintillator module.

Data from the FE to the BE are sent using the lpGBT [134] (low power GigaBit Transceiver) link protocols which are connected to the HGCal modules. The connections are arranged such that the data load is (on average) spread evenly among lpGBTs. Therefore, the data of modules which tend to see more particles, such as those closer to the pp interaction region, may be sent via multiple lpGBTs while the data of several modules far from the interaction region could be sent through a single lpGBT.

There are two sets of data sent from the HGCal TPG to the central L1T:

- **TC cluster information:** The energy of TCs from all modules across the HGCal detector are projected to make a 2D histogram in the $r/z-\phi$ plane. The histogram is then smoothed to reduce fluctuations, after which local maxima above certain threshold are found and considered as *seeds* for clustering. TCs are attached to the nearest seed if they are within a layer-dependent distance to the seed in the $x/z-y/z$ plane. After that, a cluster is reconstructed from its attached TCs. The properties of the clusters, such as their position, energy, shape, and substructure, are sent to the central L1T.

- **Trigger tower information:** The HGCal TPG sends the information of a fixed number of trigger towers (TTs) to the central L1T. (TTs are defined in section 3.8.1.) To construct TTs, the energy deposited in each module (module sum) is assigned to the TT located at the same $\eta-\phi$ position as the module. Since most of the modules overlap with several TTs, assigning a module sum to only one of the overlapping TTs may not accurately reflect the true distribution of energy. Therefore, each module sum can be split over multiple overlapping towers according to the method described in section 7.6.

  The dimension of TTs in the $\eta-\phi$ plane is 0.087×0.087 both for the barrel and endcap regions, which means that compared to the L1T in the Phase-1 CMS, a finer η dimension is used for the TTs in the endcaps in order to take advantage of the fine granularity of the HGCal. The HGCal TPG sends the total energy deposited in the TTs to the central L1T. In addition, the fraction of energy deposited in the CE-E for each TT is sent to the central L1T in order to help discriminate between γ/e and hadrons.

7.5.2 The two-stage system of the HGCal TPG

Data from each endcap of the HGCal are processed separately in the HGCal TPG, which is designed based on a two-stage system (as illustrated in Fig. 7.7). The first stage, known as the Stage-1, includes processing the data received from the FE, such as sorting, calibrating and repacking. The Stage-1 includes three groups of 14 electronic boards with each group processing data sent from one third of an endcap, i.e. a $\Delta\phi = 120^\circ$ sector. Each board contains one FPGA which processes about 2.4% ($\approx 1/(3\times14)$) of one endcap. The data processed in the Stage-1 are subsequently sent to the Stage-2 boards in a time multiplexing fashion.

In the Stage-2, there are 18 boards per 120° region, each containing one FPGA. For each bunch crossing, the data of a 120° sector, which are spread among the 14 Stage-1 FPGAs, are subsequently merged in one of the 18 Stage-2 FPGAs allowing for a full

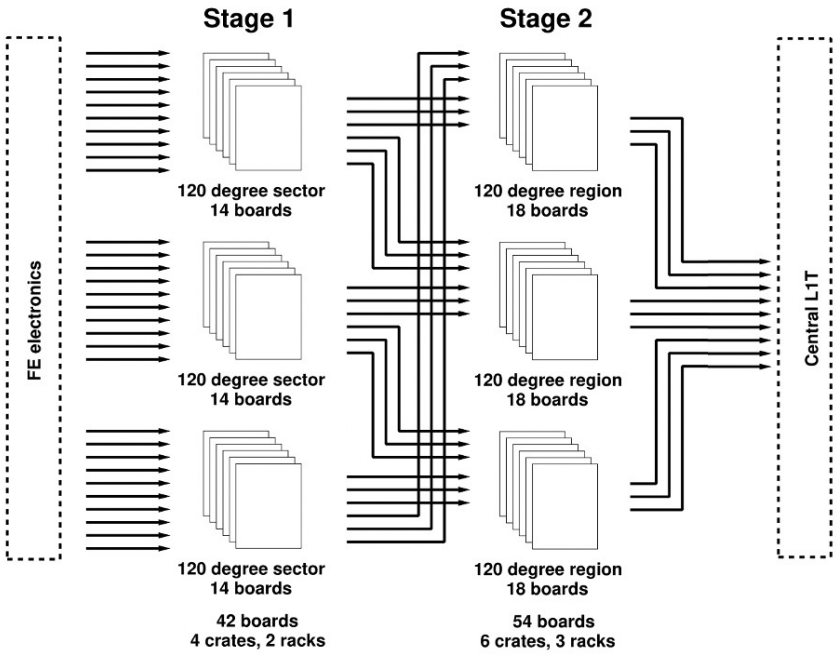

Figure 7.7: The two-stage system of the HGCal TPG. The information of the TCs close to the border of a 120° sector is copied and sent to two Stage-2 FPGAs, the one corresponding to the nominal region and the adjacent one.

process of a 120° region. Therefore, each Stage-2 FPGA processes one bunch crossing in a time interval of 18 bunch crossings (18×25 ns $= 450$ ns) while the other Stage-2 FPGAs are processing other bunch crossings in parallel. As a result, 18 consecutive bunch crossings are being processed at the same time in the Stage-2 FPGAs.

Particles passing close to the border of a 120° sector can have their energy partially transferred to the adjacent sector. To ensure that the clusters of such particles can be fully reconstructed in the Stage-2, the information of TCs close to the border is duplicated and sent to the FPGA handling the adjacent sector. Since no clustering is performed with the TTs in the HGCal TPG, the information of TTs are not duplicated. Compared to the Stage-1, the Stage-2 performs higher-level algorithms such as clustering TCs, as explained in section 7.5.1.

7.6 Assigning module sums to trigger towers

7.6.1 Introduction

As mentioned earlier, TTs in the endcaps are constructed using the energy deposited in modules, known as module sums. The initial plan was to assign each module sum to one trigger tower with the same position as the module, where the module position

could be defined as the average of the module's sensor cell (SC) positions. One could, instead, split the energy of module sums over multiple towers overlapping with the module for the following reasons:

- Modules can overlap with many towers, as demonstrated in Fig. 7.8. Summing all of the energies deposited anywhere in a module and assigning it to the centre of the module would distort energy distributions. Splitting the energy over multiple towers mitigates the distortion effect.
- Assigning to a single tower means more than 20 towers in a 120° region would receive no module sums from any modules in any layers. This means that these towers are not at the centre of any modules. As a result, regardless of the direction and energy of particles, these towers have always zero deposited energy. After dividing module sums (as discussed below) and splitting them over their overlapping towers, all towers receive energy from at least one module.

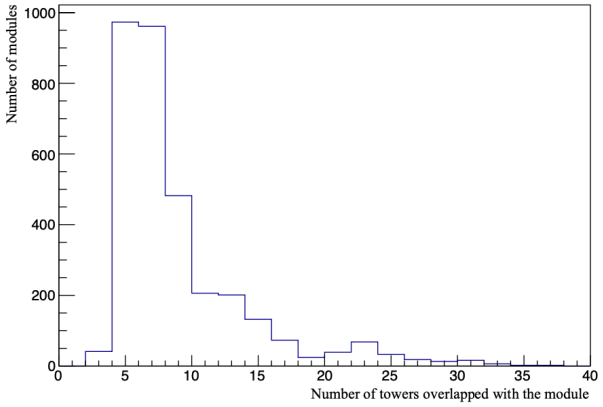

Figure 7.8: A histogram showing the number of towers a module overlaps with. Here, a module is said to overlap with a tower, if the tower covers at least one sensor cell of the module.

Therefore, it was decided to split the module sums. A tower which overlaps with a module should receive a share of the module sum proportional to the overlapping area in the $x-y$ plane rather than the $\eta-\phi$ plane. The reason is that module sums are made from summing the energy deposited in their SCs and the SCs are evenly distributed in the $x-y$ plane. In addition, since FPGAs efficiently perform number operations in binary integers, dividing module sums by 2^n, where n is a positive integer, is extremely fast and hence preferable. The n=0 case corresponds to the previous plan where a module sum is only assigned to one tower while larger values of n allow finer splitting, but require more resources as more towers receive energy shares. After investigating

the resource limitations, n=3 and n=4 were chosen and fixed for silicon and scintillator modules, respectively. Therefore, a silicon (scintillator) module sum is firstly divided by 8 (16) and then multiples of the fraction are assigned to the overlapping towers proportional to their area. In other words, a tower i overlapping with a module receives a share

$$w_i = \frac{k_i}{2^n} \tag{7.1}$$

of the module sum such that

$$\begin{aligned} &\Sigma_i w_i = 1, \text{ and} \\ &k_i \in \mathbb{Z}{\geq}0 \end{aligned} \tag{7.2}$$

To find the optimized w_i values for a module, the ratio of the overlapping area of towers and the module to the full area of the module is needed. Sections 7.6.2 and 7.6.3 describe how to estimate the relative overlapping area and how to assign a share (w_i) of the module sum to the overlapping tower.

7.6.2 Silicon modules

Fig. 7.9 depicts how a silicon module sum is split over trigger towers proportional to their overlapping area. Due to the hexagonal shape of silicon modules and the distorted shape of towers in the $x-y$ plane, estimating the relative overlapping area is challenging. Several methods were considered for estimating the relative area, such as using an analytical calculation, Monte Carlo integration with a pseudorandom number generator, and counting the relative number of sensor cells. The method of counting the relative number of sensor cells can estimate the relative area to a sufficiently accurate extent and is reasonably fast and less prone to mistakes. Therefore, this method was eventually chosen for silicon modules.

After the relative areas of the overlapping towers are estimated for a module, the areas are modified (or, in other words, smoothed in the space of tower areas) by adding a weighted average of the areas of the 24 nearby towers using a 5×5 kernel. This modification breaks the possible degeneracy of the areas and therefore allows the optimization algorithm to find a unique solution for splitting module sums. The resulting areas a_i are then normalized such that

$$\Sigma_i a_i = 1. \tag{7.3}$$

After normalizing, the areas are sorted by their size $a_1, a_2, ..., a_m$, where a_1 is the largest.

The areas smaller than a_8 do not receive any share as the smallest share a tower

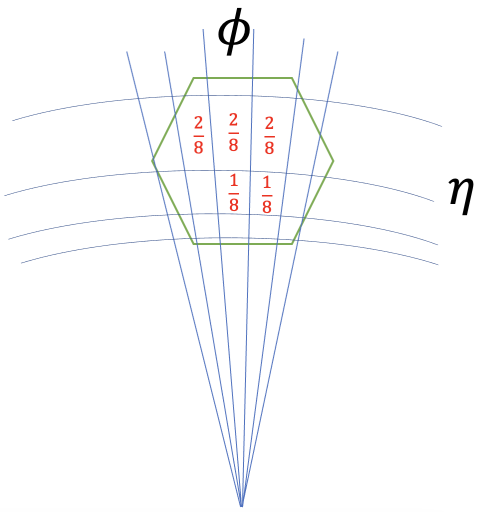

Figure 7.9: An illustration of how one unit of energy which is deposited in a module is split over the overlapping towers proportional to the overlapping area. The green hexagon represents the module and the blue lines show tower borders. The fraction of module sum each tower receives is shown in red. The figure is just for illustration and does not show a real module-tower overlap.

can get is 1/8; therefore towers with smaller relative areas are discarded[3]. Then the optimized w_i values are found by minimizing a loss function defined as

$$L = \Sigma_i (w_i - a_i)^2. \tag{7.4}$$

The optimized w_i values are then used for splitting module sums over the modules' overlapping towers. Fig. 7.10 shows, as an example, the result of this optimization for a silicon module.

7.6.3 Scintillator modules

Since scintillator modules are projective in the $\eta-\phi$ plane, finding the optimized w_i values is more straightforward. Each scintillator module covers $\Delta\phi$=10° which is twice the size of towers in ϕ and since there is no offset in ϕ between towers and scintillator modules, one of the two ϕ borders of a tower is aligned with a module border and the other divides the module in half (see Fig. 7.11). However, module borders in η are not necessarily aligned with tower borders and they can cover different numbers of towers. Within $\Delta\phi$=10° module borders, exactly two modules exist, which are called u0 and u1. The η range of these modules for different layers is shown in Fig. 7.12. The η range

[3]Note that m can be smaller than or equal to eight, in which case no towers are discarded in this step.

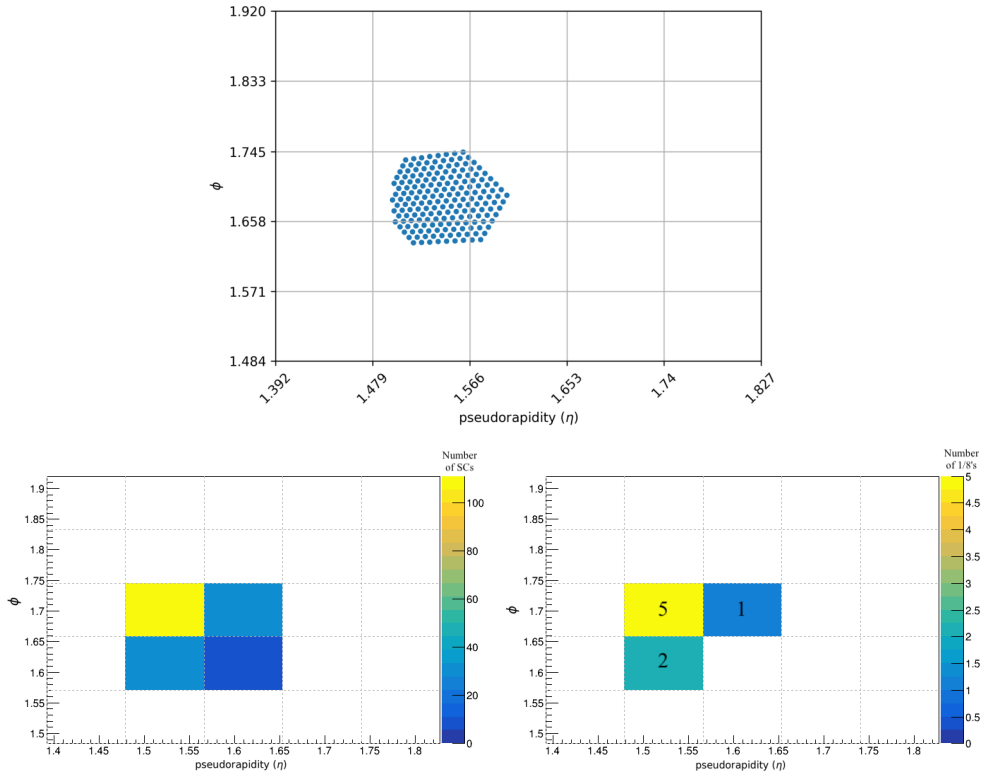

Figure 7.10: An example of splitting a module sum over the overlapping towers. The top figure shows the position of the module SCs in the $\eta-\phi$ plane with tower borders overlapped. The SCs of this module overlap with four towers. The bottom-left figure is a histogram of the number of SCs in each tower. The number of SCs can be used to estimate the ratio of the overlapping area to the module area. The 5×5 kernel for smoothing the areas (mentioned in the text) is applied on the histogram in the bottom-left figure. The bottom-right figure shows the share of the module sum given to each of the towers to minimize the loss function defined in Eq. 7.4. One of the four overlapping towers does not receive any share.

of u0 and u1 modules in each layer is estimated by assuming that their trigger cells are evenly spread across the modules.

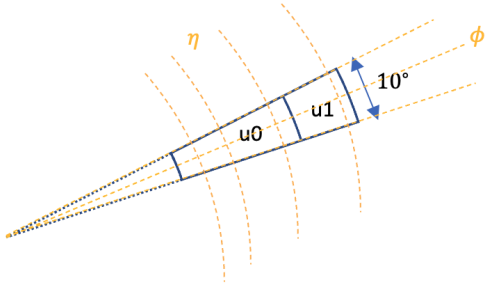

Figure 7.11: An illustration of the position of scintillator modules with respect to towers. Tower borders are shown in orange while the borders of the modules are shown in blue. There are exactly two modules, namely u0 and u1, in a $\Delta\phi$=10° sector. Tower borders in ϕ are either aligned with the module borders or divide the modules in half. The module borders in η are irregular.

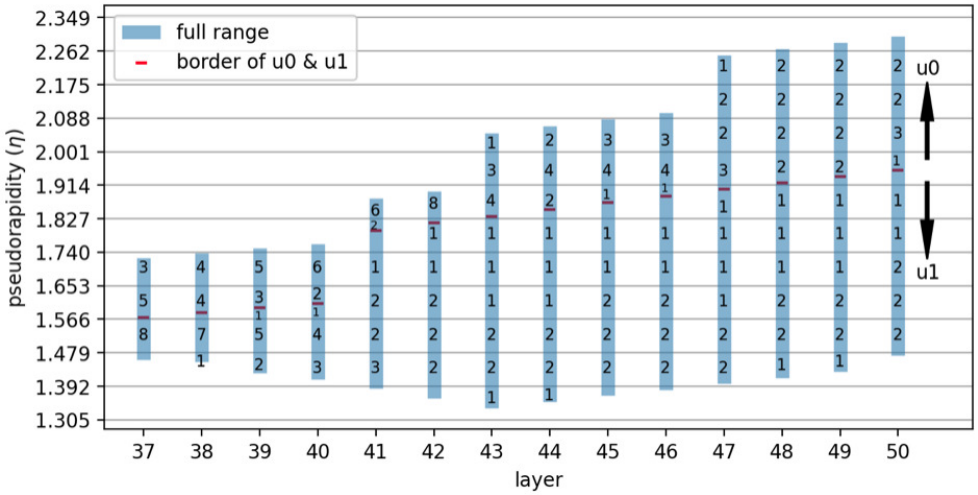

Figure 7.12: The η range of u0 and u1 scintillator modules in different layers and the module sum share each overlapping tower receives. The shares are $8 \times w_i$ for towers in a $\Delta\phi$=5° interval. The horizontal lines show tower borders in η.

Since the pattern of modules in a $\Delta\phi$=10° interval is repeated across ϕ, the optimized w_i values are derived for only one $\Delta\phi$=10° interval, separately for the u0 and u1 modules and for all layers. As mentioned earlier, scintillator module sums are divided by 16. Since both u0 and u1 modules are divided in half by tower borders in ϕ, the optimization problem reduces to firstly divide module sums in half and then divide each half by 8 and split it over the towers in a $\Delta\phi$=5° interval.

For each module, the relative area of each overlapping tower is calculated analyti-

cally. Fig. 7.13 shows a schematic of two towers, A and B, for which the relative area is to be calculated. Both towers cover $\Delta\phi$=5° (the same as every tower) while tower A (B) is bounded between R_3 and R_2 (R_2 and R_1), where R is the distance to the beam axis. Assuming the layer is located at a distance Z with respect to the pp interaction point, η_i corresponding to R_i is derived using the relation between R, Z and η

$$\frac{R}{Z} = \text{cosech}(\eta), \tag{7.5}$$

where cosech is hyperbolic cosecant, i.e. 1/sinh. This equation holds for both endcaps but for simplicity the following calculation is derived for the endcap with $\eta > 0$ and the result of these calculation is applicable to the other endcap because of the symmetry between the endcaps. The areas of A and B are derived as

$$\begin{aligned} A &= \Delta\phi\ \frac{R_3^2}{2} - \Delta\phi\ \frac{R_2^2}{2}, \text{ and} \\ B &= \Delta\phi\ \frac{R_2^2}{2} - \Delta\phi\ \frac{R_1^2}{2} \end{aligned} \tag{7.6}$$

The relative areas are therefore

$$\frac{A}{B} = \frac{R_3^2 - R_2^2}{R_2^2 - R_1^2} = \frac{\text{cosech}^2(\eta_3) - \text{cosech}^2(\eta_2)}{\text{cosech}^2(\eta_2) - \text{cosech}^2(\eta_1)}, \tag{7.7}$$

where Z is canceled out as it is equal for all towers in a layer.

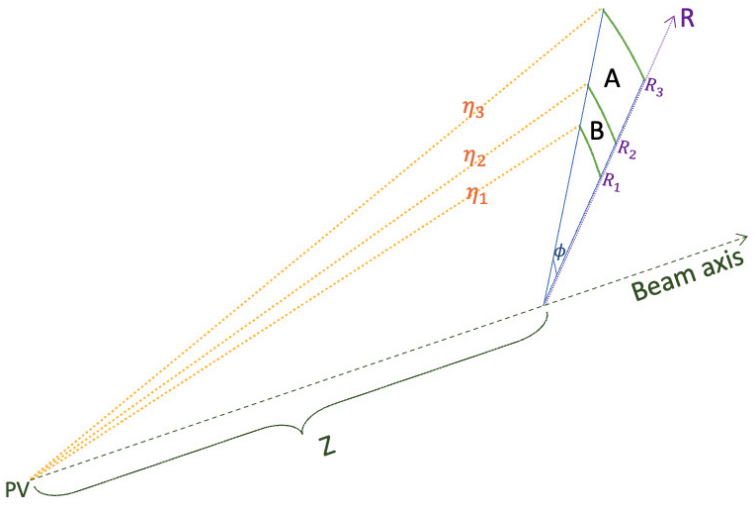

Figure 7.13: A 3D view of the position of two towers, A and B, in a layer with distance Z to the PV. The area of A relative to B is estimated analytically as described in the text.

Using Eq. 7.7, the relative areas for all pairs of towers are derivable and, consequently, the relative areas $a_1, a_2, ..., a_m$ defined in Eq. 7.3 can be calculated. The optimized w_i values are subsequently derived by minimizing the loss function in Eq. 7.4. The optimized w_i values (multiplied by 8) are shown in Fig. 7.12 for towers overlapping u0 and u1 modules.[4] Towers which fully lie within a module have the same area in the $\eta-\phi$ plane; however, in the $x-y$ plane, those located in lower η have larger areas and therefore receive larger or equal shares compared to those in higher η. The η range that scintillators cover varies with layer as can also be seen in Fig. 7.4.

7.7 Jet reconstruction performance

In this section the effect of splitting module sums on the position and energy resolutions of hadronic jets is studied. Hereafter, the case where module sums are split over multiple towers is referred to as the "splitting" case, as opposed to the "non-splitting" case which describes the situation where a module sum is assigned to a single tower. A Higgs boson sample with an average of 200 pile-up interactions is used for this study, where the Higgs boson is produced via the VBF production mode and it subsequently decays to hypothetical particles leaving no trace in the detector. The production and decay of the Higgs boson are simulated using POWHEG and PYTHIA, respectively. Among the hadronic jets produced in each event, the leading jet is used for the resolution study.

In the Phase-2 CMS, reconstructing jets with towers is performed with a 7×7 window of towers, which approximately corresponds to a cone size of 0.4 in the anti-k_T algorithm. So in this study, the tower towards which the *true* direction of the leading jet is pointing is taken as the centre of the 7×7 window, where "true" means using generator-level information. Events are discarded if the window partially or fully lies outside the range of the HGCal towers.

Fig. 7.14 compares the E_{T} distribution between the splitting and non-splitting cases for an event in which the leading jet has a true E_{T}=335 GeV transverse energy. In the splitting case, the E_{T} distribution of the leading jet is more spread, which is an expected consequence of splitting module sums. The shape of the energy deposition is more perceptible in the splitting case while in the non-splitting case, most of the energy is concentrated in a single tower and a its structure is difficult to discern. The L1T analyzes the shape of energy depositions for object identification; therefore, splitting module sums should bring improvement to this area (to be confirmed by future L1T studies).

[4]Note that these are w_i values for a $\Delta\phi$=5° sector and hence sum up to 8/16 for each u0 or u1 module. These modules receive the same w_i values for the other $\Delta\phi$=5° sector they cover, so the values sum up to 16/16.

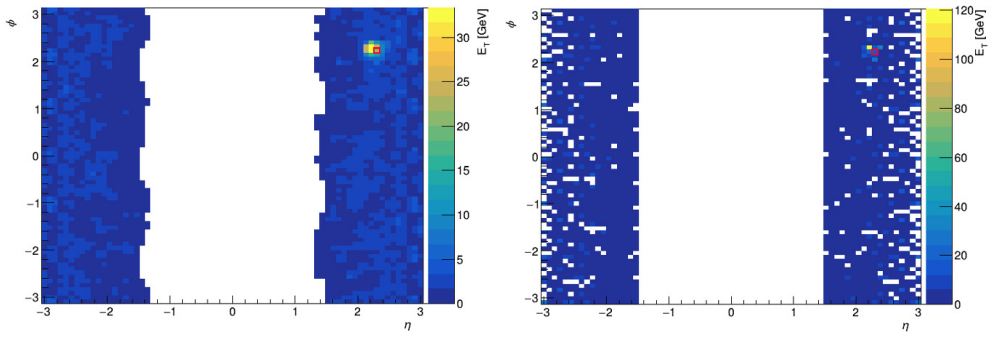

Figure 7.14: Comparing the E_{T} distribution between the splitting (left) and non-splitting (right) cases for an event in which the true transverse energy of the leading jet is E_{T}=335 GeV. A red box is drawn around the tower towards which the true direction of the leading jet is pointing; this direction is the same in both figures as they display the same event.

The E_{T} of the jets is reconstructed by summing the E_{T} of the towers within the 7×7 window. In Fig. 7.15, the scatter plot of the reconstructed E_{T} as a function of the generated E_{T} is displayed for the splitting and non-splitting cases. The mean and uncertainty of the reconstructed E_{T} are subsequently calculated in bins of the generated E_{T} and overlaid in the plot. Due to the large number of pile-up interactions, there is a positive shift in the reconstructed E_{T} in all energy regimes. The plot demonstrates that the energy resolution has not significantly changed after splitting module sums. The reason is that the tower window is wide enough to include (almost) all of the E_{T} of the jets regardless of whether or not module sums are split.

The position of the jets in the $\eta-\phi$ plane is reconstructed using an energy-weighted average of the positions of the towers inside the 7×7 window. Fig. 7.16 shows the η and ϕ resolutions of the jets for the splitting and non-splitting cases. The η and ϕ resolutions improves by about 7 and 19%, respectively, after splitting module sums.

To conclude, splitting module sums over multiple towers improves the position resolution of jets while it does not change their energy resolution. It should be noted that the jets considered in this study tend to be well isolated as they come from the VBF process. Therefore, this result may not be applicable to non-isolated cases, such as the b-jets from boosted $H \to b\bar{b}$ decay. Moreover, this study does not cover the performance of splitting module sums in object identification with towers (such as identifying taus from jets). By utilizing the shape of objects, object identification power could be improved in the splitting case, where the shapes are more perceptible. Future L1T studies should provide more information on the effect of module sum splitting on object identification power and non-isolated jet resolution.

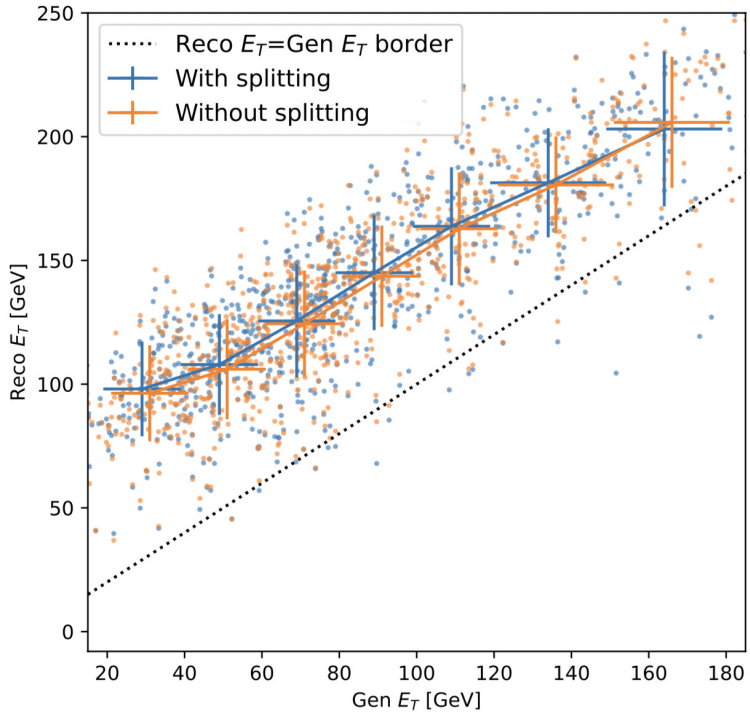

Figure 7.15: Comparing the transverse energy resolution of the leading jet between the splitting and non-splitting cases. The splitting (non-splitting) case is shown in blue (orange) dots. The mean and uncertainty of the reconstructed E_{T} in both cases are derived in bins of the generated E_{T} and are overlaid in the plot.

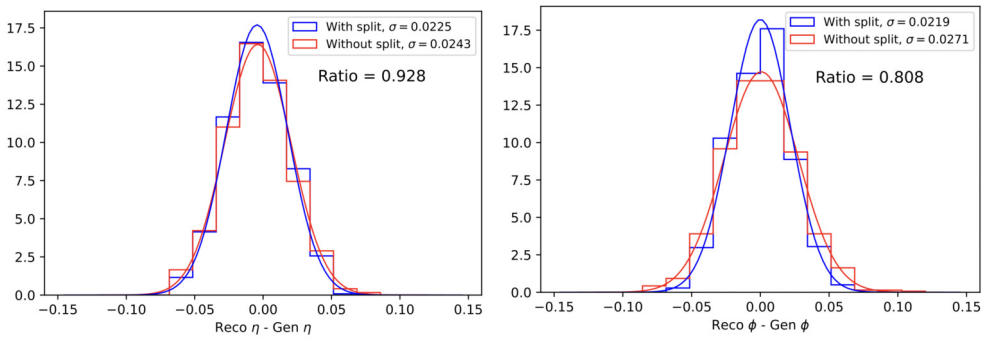

Figure 7.16: Comparing the splitting and non-splitting cases based on the η (left) and ϕ (right) resolutions of the jets.

Chapter 8

Conclusions and outlook

The thesis presented the results of measuring the $\mathcal{CP}$ nature of the Yukawa coupling between the Higgs boson and tau leptons using proton-proton collision data recorded by the CMS experiment, corresponding to an integrated luminosity of 137 fb^{-1} at $\sqrt{s}$=13 TeV. The data included $H \to \tau^+\tau^-$ events in the final states where at least one tau decays hadronically. Several optimization techniques were employed to improve the $\mathcal{CP}$ sensitivity. In particular, an MVA was developed for identifying different hadronic decay modes of taus. The MVA improved the $\mathcal{CP}$ sensitivity by about 20% depending on the exact final state. The effective mixing angle, $\alpha^{H\tau\tau}$, was found to have an observed (expected) value of $-1 \pm 19^\circ$ ($0 \pm 21^\circ$) at 68% CL. This result is compatible with the Standard model predictions and excludes a pure $\mathcal{CP}$-odd coupling at 3.0 σ CL. The statistical uncertainty of the data was by far the largest source of uncertainty for this result; therefore future data, which will be collected during the Run-3 of the LHC and the HL-LHC, is needed to significantly reduce the uncertainty.

The projection of the expected sensitivity for $\alpha^{H\tau\tau}$ is found to be $0 \pm 13^\circ$ and $0 \pm 5^\circ$ after the Run-3 of the LHC and the HL-LHC, respectively. However, several improvements are planned in different areas which could result in an expected sensitivity of $0 \pm 10^\circ$ after Run-3. Among the improvements are: (i) developing a more complex algorithm to increase the efficiency of selecting τ_h candidates at trigger level, (ii) utilizing low- and high-level features of τ_h decays in a more advanced ML algorithm for identifying τ_h decay modes, (iii) using a new method [53] which can reconstruct $\phi_{\mathcal{CP}}$ in the Higgs boson rest frame for certain decay channels, and (iv) estimating the direction of PF candidates (such as π^0) using ML techniques. A prediction for the improvement in the expected sensitivity after the HL-LHC is not yet available as the systematic uncertainties begin to compete with or dominate the statistical uncertainties and therefore a detailed study is needed.

In addition to the Higgs $\mathcal{CP}$ analysis, the thesis presented an optimization algorithm

which was developed to determine how data should be processed in the CMS Level-1 triggering system during the HL-LHC. The total energy deposited in a module of the HGCal detector needs to be split over multiple trigger towers overlapping with the module. The trigger towers should receive a share proportional to the overlapping area. With this condition, and after imposing the bandwidth limitation constraints, the algorithm found the share each tower should receive. The thesis demonstrated that splitting module energies over multiple towers improves the η and ϕ resolutions of hadronic jets by 7% and 19%, respectively, compared with the case where the total energy of modules is assigned to only one tower overlapping with the module. However, the energy resolution did not change significantly because the "window" in the $\eta-\phi$ plane inside which trigger towers are used for jet energy measurements is sufficiently large to contain almost all of the jet energy regardless of the splitting choice.

Future Level-1 trigger studies should reveal more information about the effects of splitting module energies. The sample used in this study was prepared by collecting the leading jet in the events where a Higgs boson is produced through the VBF production mode. Since these jets tend to be isolated, the conclusion on the resolution of the jets is not necessarily applicable to non-isolated jets. Moreover, this study did not cover the effect of splitting module energies on physics object identification. These effects should be studied separately.

Appendix A

MVA decay mode feature-importance

A.1 1-charged-prong decay

The features used in the 1-charged-prong MVA are listed below and the corresponding feature-importance ranking is shown in Fig. A.1.

- `eta`: η of τ_h
- `pt`: p_T of τ_h
- `Mrho`: m($\pi^{\pm}$, $\mathcal{S}$)
- `Mrho_OneHighGammas`: m($\pi^{\pm}$, γ/e^{lead})
- `Mrho_subleadingGamma`: m($\pi^{\pm}$, $\gamma/e^{sublead}$)
- `Mrho_TwoHighGammas`: m($\pi^{\pm}$, γ/e^{lead}, $\gamma/e^{sublead}$)
- `Egamma1_tau`: $\frac{E_{\gamma/e^{lead}}}{E_{\tau_h}}$
- `Egamma2_tau`: $\frac{E_{\gamma/e^{sublead}}}{E_{\tau_h}}$
- `Mpi0`: m($\mathcal{S}$)
- `Mpi0_TwoHighGammas`: m(γ/e^{lead}, $\gamma/e^{sublead}$)
- `strip_pt`: p_T of $\mathcal{S}$
- `Epi0`: energy of $\mathcal{S}$

- `Epi_tau`: $\frac{E_{\pi^\pm}}{E_{\tau_h}}$
- `Epi`: $E_{\pi^\pm}$
- `tau_decay_mode`: HPS decay mode
- `DeltaR2WRTtau`: $\langle \Delta R^2 \rangle = \frac{\Sigma_i \Delta R_i^2 p_{T_i}^2}{\Sigma_i p_{T_i}^2}$, in which i sums over $\pi^\pm$ and all γ/e in $\mathcal{S}$. ΔR is with respect to τ_h direction)
- `DeltaR2WRTtau_tau`: $\langle \Delta R^2 \rangle \times (E_{\tau_h})^2$
- `rho_dEta`: $\Delta\eta(\pi^\pm, \mathcal{S})$
- `rho_dEta_tau`: $E_{\tau_h} \times \Delta\eta(\pi^\pm, \mathcal{S})$
- `rho_dphi`: $\Delta\phi(\pi^\pm, \mathcal{S})$
- `rho_dphi_tau`: $E_{\tau_h} \times \Delta\phi(\pi^\pm, \mathcal{S})$
- `gammas_dEta`: $\Delta\eta(\gamma/e^{lead}, \gamma/e^{sublead})$
- `gammas_dEta_tau`: $E_{\tau_h} \times \Delta\eta(\gamma/e^{lead}, \gamma/e^{sublead})$
- `gammas_dR_tau`: $E_{\tau_h} \times \Delta R(\gamma/e^{lead}, \gamma/e^{sublead})$

A.2 3-charged-prong decay

The features used in the 3-charged-prong MVA are listed below and the corresponding feature-importance ranking is shown in Fig. A.2.

- `E1_overEa1`: $\frac{E_{\pi_1}}{E_{\pi_1}+E_{\pi_2}+E_{\pi_3}}$
- `E2_overEa1`: $\frac{E_{\pi_2}}{E_{\pi_1}+E_{\pi_2}+E_{\pi_3}}$
- `E1_overEtau`: $\frac{E_{\pi_1}}{E_{\tau_h}}$
- `E2_overEtau`: $\frac{E_{\pi_2}}{E_{\tau_h}}$
- `E3_overEtau`: $\frac{E_{\pi_3}}{E_{\tau_h}}$
- `tau_decay_mode`: HPS decay mode
- `mass0`: m(π_1, π_2, π_3)
- `mass1`: m(π_1, π_2)

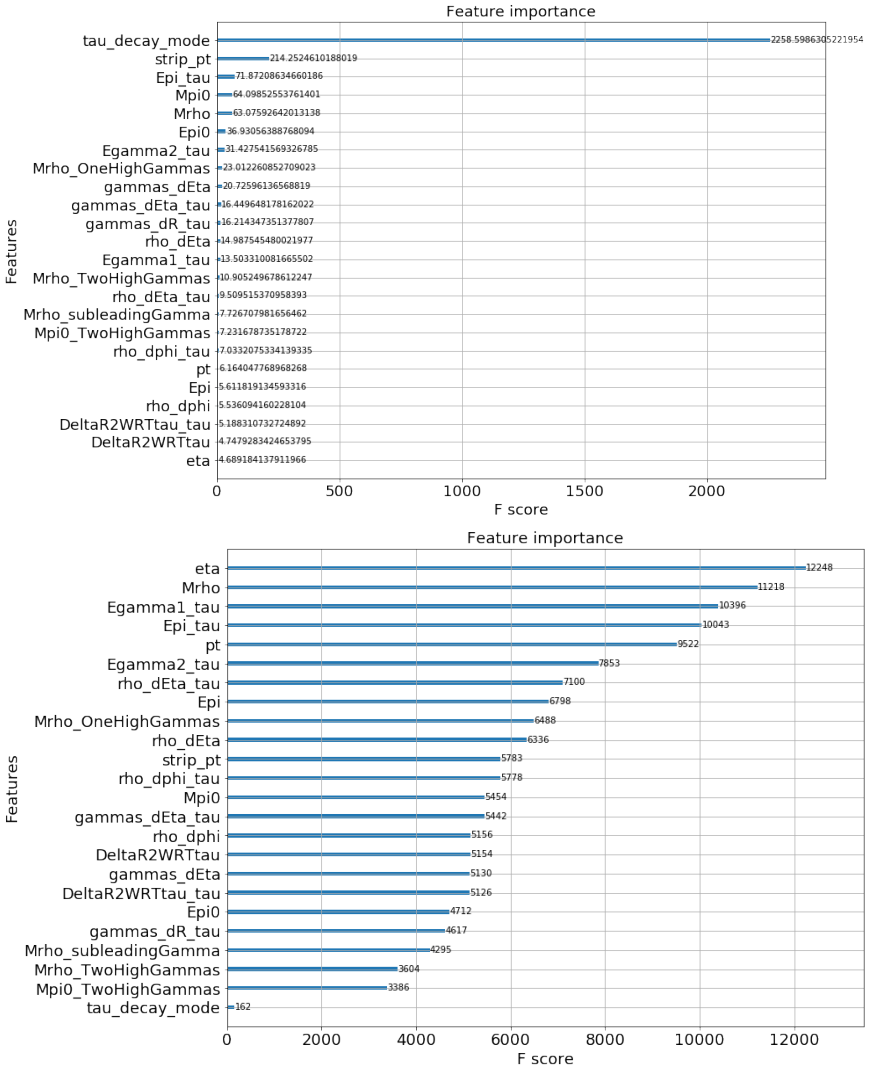

Figure A.1: Feature-importance list for 1-charged-prong decay MVA with *Gain* (top) and *Weight* (bottom) metrics. These metrics are defined in section 5.4.1.

- `mass2`: m(π_1, π_3)
- `E1`: E_{π_1}
- `E2`: E_{π_2}
- `E3`: E_{π_3}
- `strip_E`: energy of $\mathcal{S}$
- `strip_pt`: p_T of $\mathcal{S}$
- `pt`: p_T of τ_h
- `eta`: η of τ_h
- `E`: energy of τ_h
- `Egamma1`: $E_{\gamma/e^{lead}}$
- `Egamma2`: $E_{\gamma/e^{sublead}}$
- `Mpi0`: m($\mathcal{S}$)
- `Mpi0_TwoHighGammas`: m(γ/e^{lead}, $\gamma/e^{sublead}$)
- `h1_h2_dphi`: $\Delta\phi(\pi_1, \pi_2)$
- `h1_h3_dphi`: $\Delta\phi(\pi_1, \pi_3)$
- `h2_h3_dphi`: $\Delta\phi(\pi_2, \pi_3)$
- `h1_h2_dEta`: $\Delta\eta(\pi_1, \pi_2)$
- `h1_h3_dEta`: $\Delta\eta(\pi_1, \pi_3)$
- `h2_h3_dEta`: $\Delta\eta(\pi_2, \pi_3)$
- `h1_h2_dphi_timesE12`: $(E_{\pi_1} + E_{\pi_2}) \times \Delta\phi(\pi_1, \pi_2)$
- `h1_h3_dphi_timesE13`: $(E_{\pi_1} + E_{\pi_3}) \times \Delta\phi(\pi_1, \pi_3)$
- `h2_h3_dphi_timesE23`: $(E_{\pi_2} + E_{\pi_3}) \times \Delta\phi(\pi_2, \pi_3)$
- `h1_h2_dEta_timesE12`: $(E_{\pi_1} + E_{\pi_2}) \times \Delta\eta(\pi_1, \pi_2)$
- `h1_h3_dEta_timesE13`: $(E_{\pi_1} + E_{\pi_3}) \times \Delta\eta(\pi_1, \pi_3)$
- `h2_h3_dEta_timesE23`: $(E_{\pi_2} + E_{\pi_3}) \times \Delta\eta(\pi_2, \pi_3)$

- `a1_pi0_dEta`: $\Delta\eta(\tau_h, \mathcal{S})$
- `a1_pi0_dphi`: $\Delta\phi(\tau_h, \mathcal{S})$
- `a1_pi0_dEta_timesEtau`: $E_{\tau_h} \times \Delta\eta(\tau_h, \mathcal{S})$
- `a1_pi0_dphi_timesEtau`: $E_{\tau_h} \times \Delta\phi(\tau_h, \mathcal{S})$
- `gammas_dEta`: $\Delta\eta(\gamma/e^{lead}, \gamma/e^{sublead})$
- `gammas_dEta_timesEtau`: $E_{\tau_h} \times \Delta\eta(\gamma/e^{lead}, \gamma/e^{sublead})$
- `gammas_dR_timesEtau`: $E_{\tau_h} \times \Delta R(\gamma/e^{lead}, \gamma/e^{sublead})$

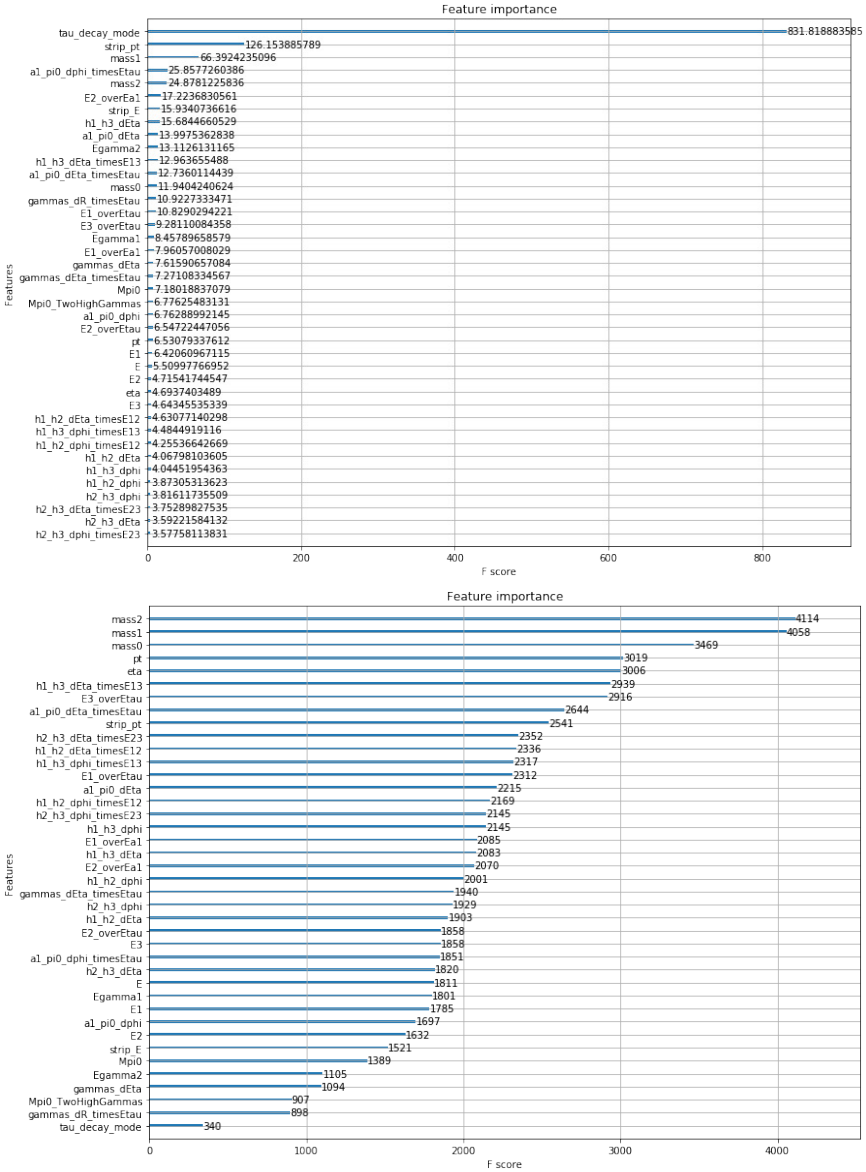

Figure A.2: Feature-importance list for 3-charged-prong decay MVA with *Gain* (top) and *Weight* (bottom) metrics. These metrics are defined in section 5.4.1.

Appendix B

Electron and muon scale factors

In this section, the efficiencies and scale factors for identification, isolation, and trigger requirements which are derived separately for electrons and muons in different regions of the detector are plotted as a function of the transverse momentum of the leptons.

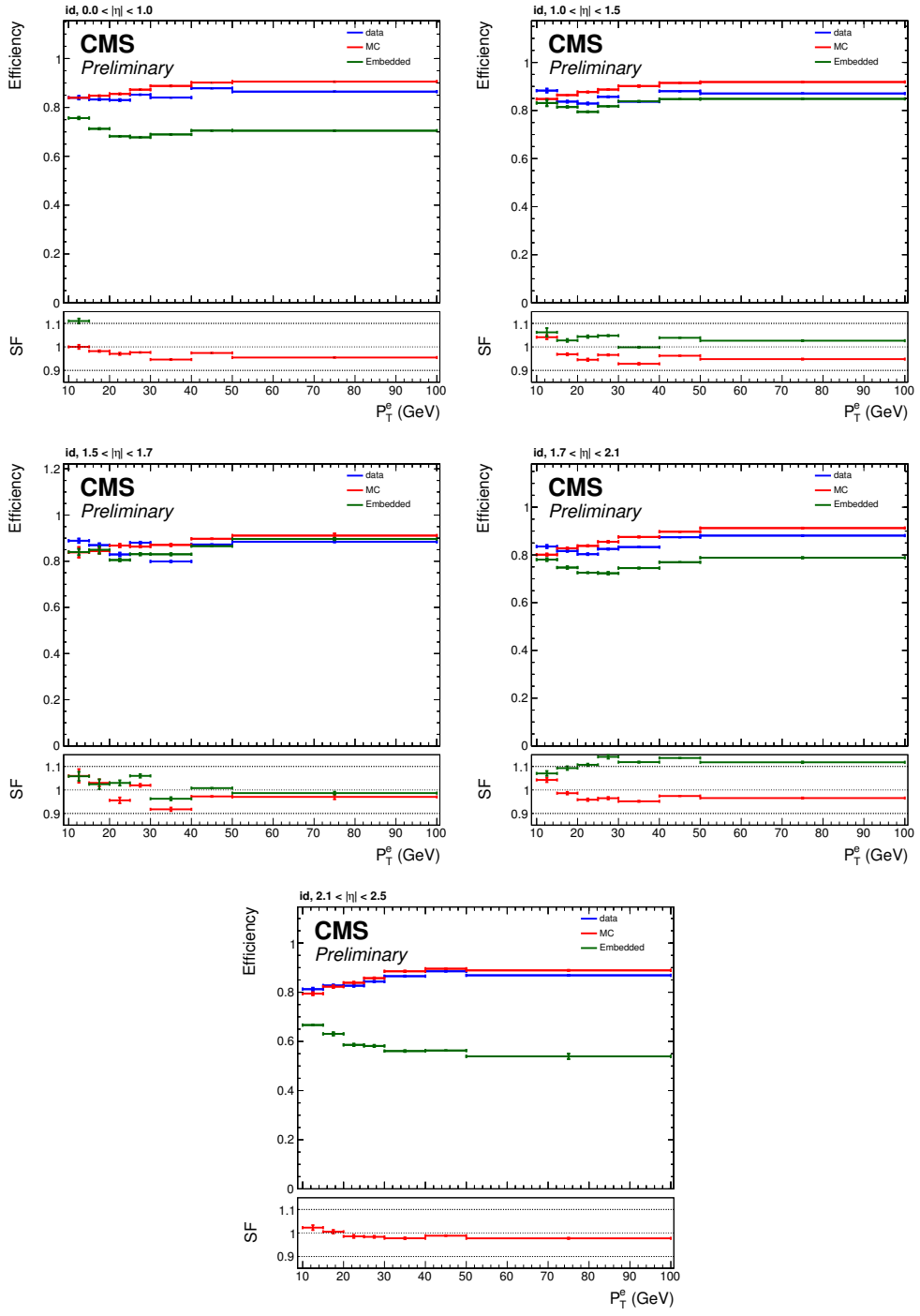

Figure B.1: The efficiencies and scale factors for electron identification are displayed for the $0.0 < |\eta| < 1.0$, $1.0 < |\eta| < 1.5$, $1.5 < |\eta| < 1.7$, $1.7 < |\eta| < 2.1$, and $2.1 < |\eta| < 2.5$ regions. The efficiency of data (blue), MC (red), and embedded (green) samples are shown in the upper panel of each sub-figure while the ratio of data to the MC and embedded samples are shown in the lower panel.

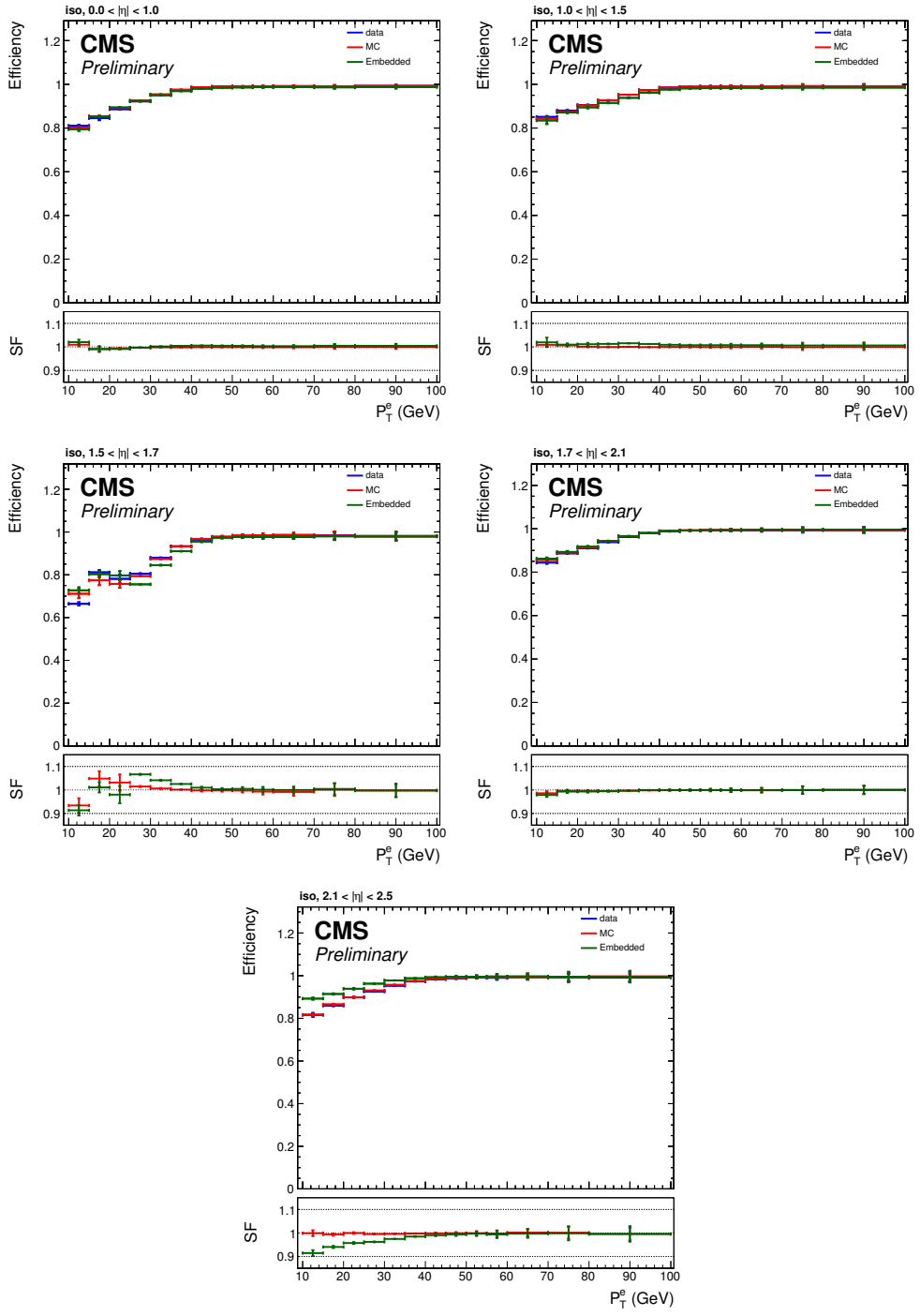

Figure B.2: The efficiencies and scale factors for electron isolation are displayed for the $0.0 < |\eta| < 1.0$, $1.0 < |\eta| < 1.5$, $1.5 < |\eta| < 1.7$, $1.7 < |\eta| < 2.1$, and $2.1 < |\eta| < 2.5$ regions. The efficiency of data (blue), MC (red), and embedded (green) samples are shown in the upper panel of each sub-figure while the ratio of data to the MC and embedded samples are shown in the lower panel.

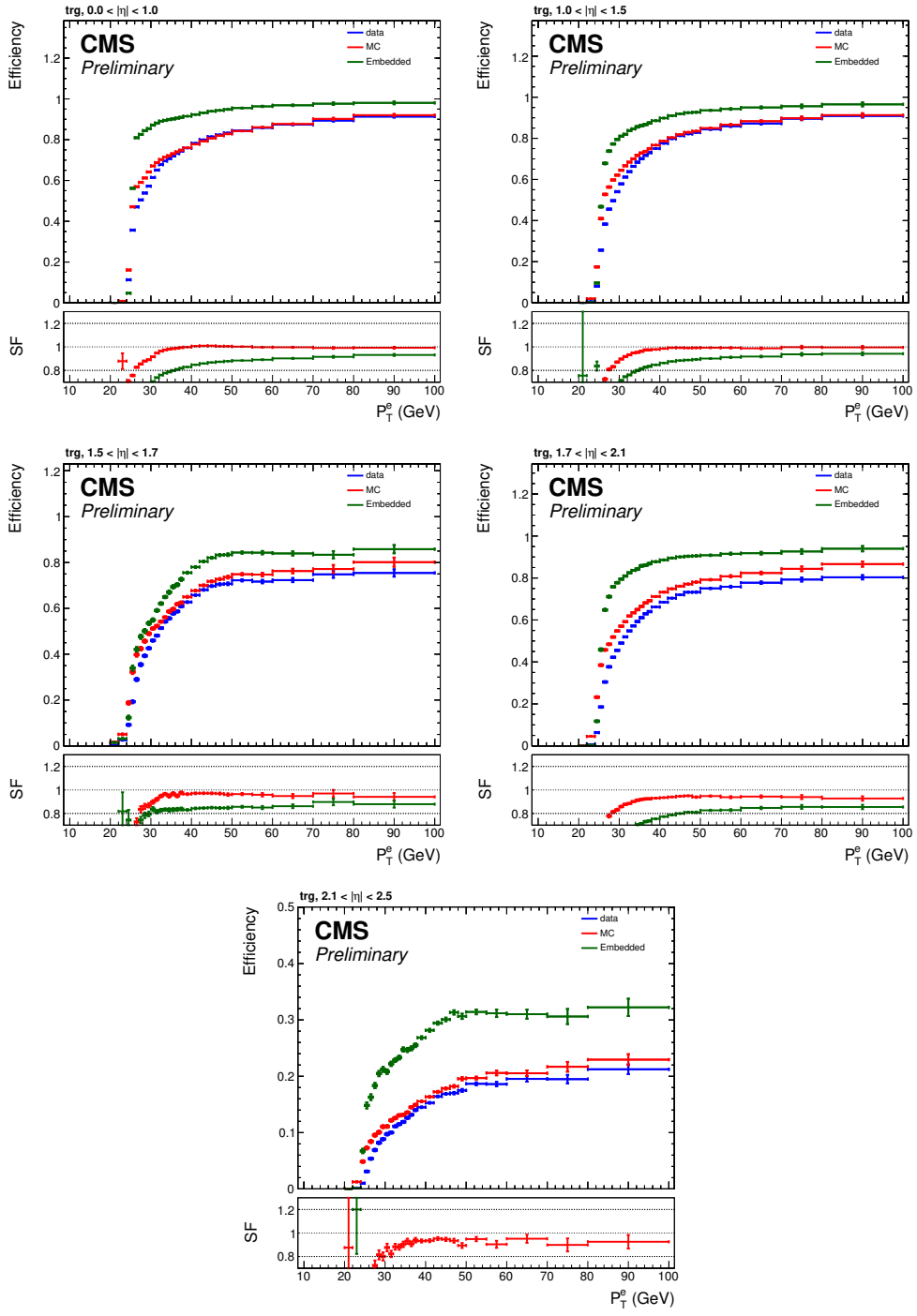

Figure B.3: The efficiencies and scale factors for single-electron trigger are displayed for the $0.0 < |\eta| < 1.0$, $1.0 < |\eta| < 1.5$, $1.5 < |\eta| < 1.7$, $1.7 < |\eta| < 2.1$, and $2.1 < |\eta| < 2.5$ regions. The efficiency of data (blue), MC (red), and embedded (green) samples are shown in the upper panel of each sub-figure while the ratio of data to the MC and embedded samples are shown in the lower panel.

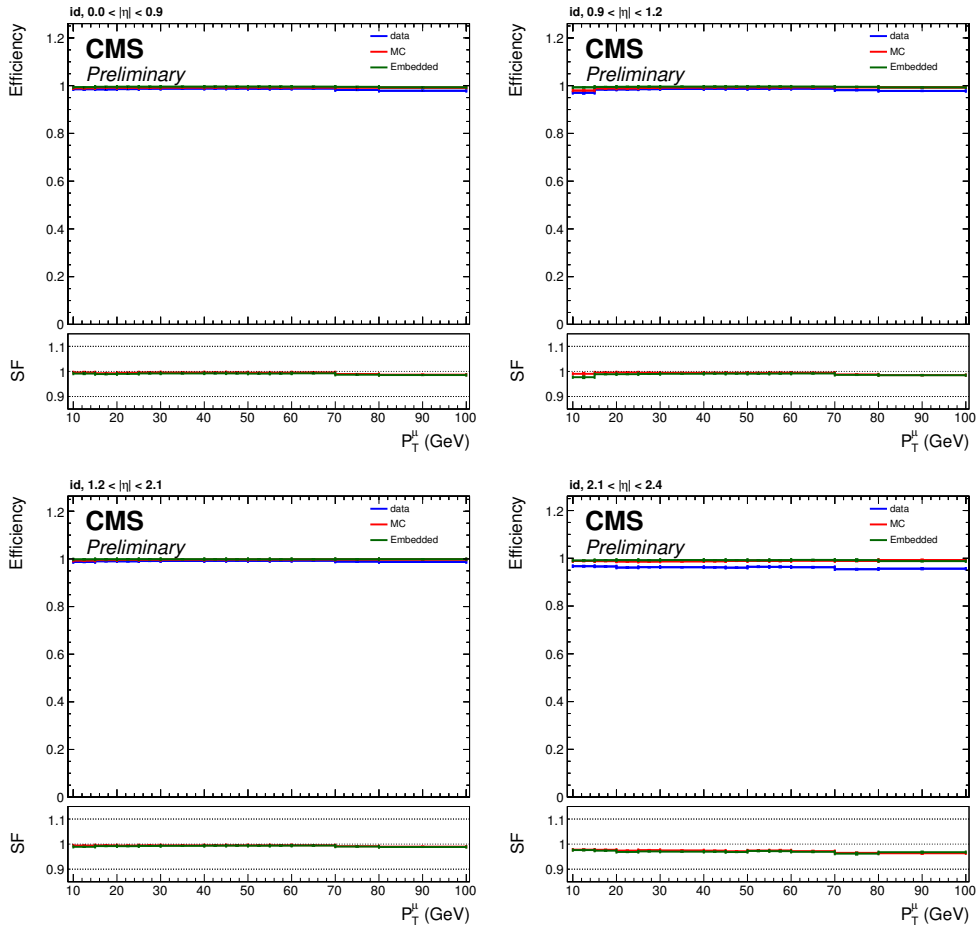

Figure B.4: The efficiencies and scale factors for muon identification are displayed for the $0.0 < |\eta| < 0.9$, $0.9 < |\eta| < 1.2$, $1.2 < |\eta| < 2.1$, and $2.1 < |\eta| < 2.4$ regions. The efficiency of data (blue), MC (red), and embedded (green) samples are shown in the upper panel of each sub-figure while the ratio of data to the MC and embedded samples are shown in the lower panel.

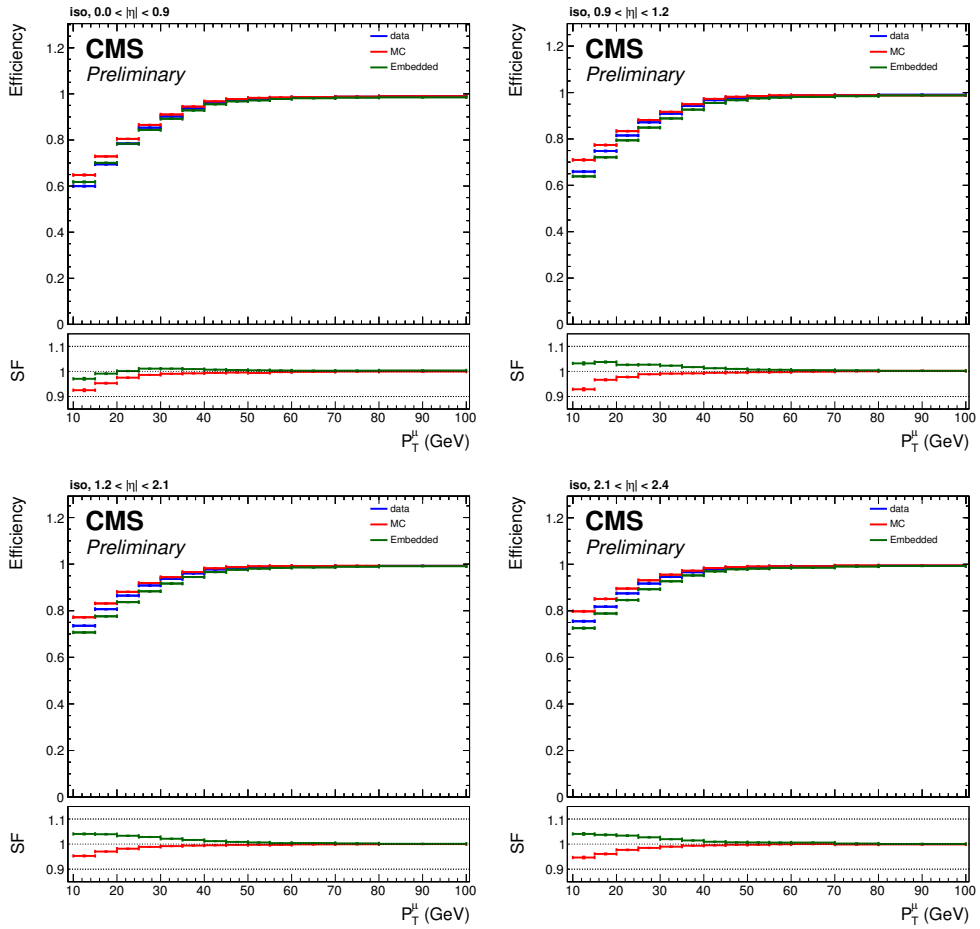

Figure B.5: The efficiencies and scale factors for muon isolation are displayed for the $0.0 < |\eta| < 0.9$, $0.9 < |\eta| < 1.2$, $1.2 < |\eta| < 2.1$, and $2.1 < |\eta| < 2.4$ regions. The efficiency of data (blue), MC (red), and embedded (green) samples are shown in the upper panel of each sub-figure while the ratio of data to the MC and embedded samples are shown in the lower panel.

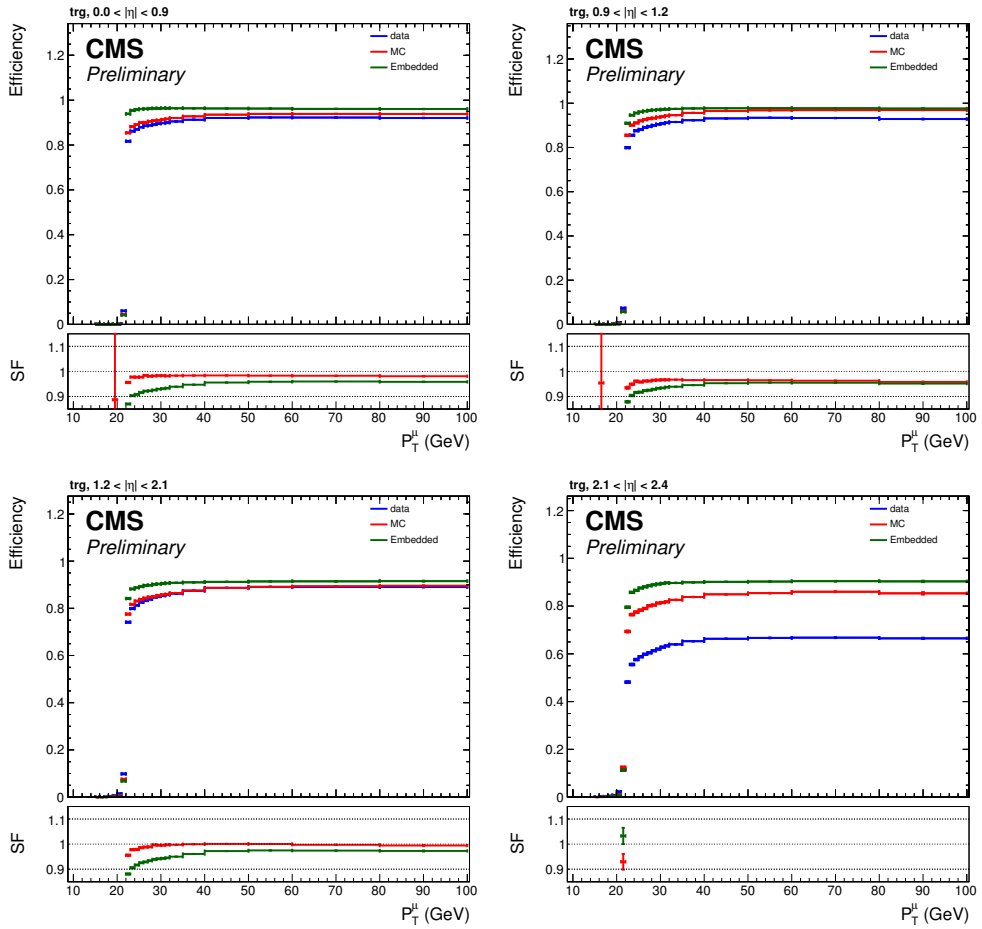

Figure B.6: The efficiencies and scale factors for single-muon trigger are displayed for the $0.0 < |\eta| < 0.9$, $0.9 < |\eta| < 1.2$, $1.2 < |\eta| < 2.1$, and $2.1 < |\eta| < 2.4$ regions. The efficiency of data (blue), MC (red), and embedded (green) samples are shown in the upper panel of each sub-figure while the ratio of data to the MC and embedded samples are shown in the lower panel.

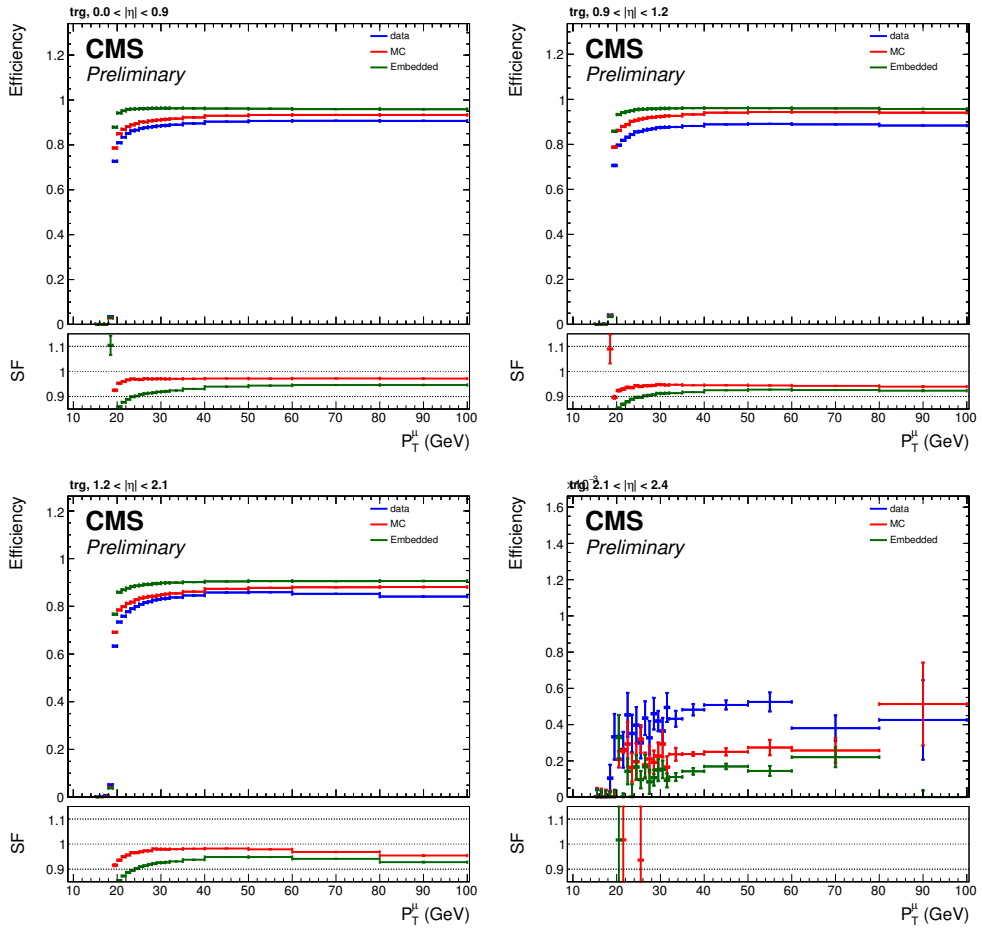

Figure B.7: The efficiencies and scale factors for muon+tau cross trigger are displayed for the $0.0 < |\eta| < 0.9$, $0.9 < |\eta| < 1.2$, $1.2 < |\eta| < 2.1$, and $2.1 < |\eta| < 2.4$ regions. The efficiency of data (blue), MC (red), and embedded (green) samples are shown in the upper panel of each sub-figure while the ratio of data to the MC and embedded samples are shown in the lower panel.

Appendix C

Hadronic tau scale factors

The first two sets of plots in this section show the scale factors derived for τ_h identification requirement along with the postfit plots illustrating the agreement between data and simulated events after applying the scale factors. The third set of plots displays the efficiencies and scale factors derived for τ_h trigger requirement.

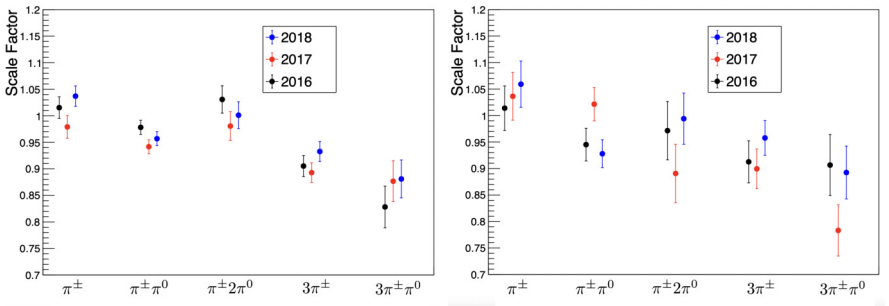

Figure C.1: The τ_h identification scale factors for different MVA decay modes and data-taking periods. The left (right) plot shows the scale factors for $20 < p_{\mathrm{T}}^{\tau_h} < 40$ GeV ($p_{\mathrm{T}}^{\tau_h} > 40$ GeV). These scale factors are derived for the embedded sample and are valid for the $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels.

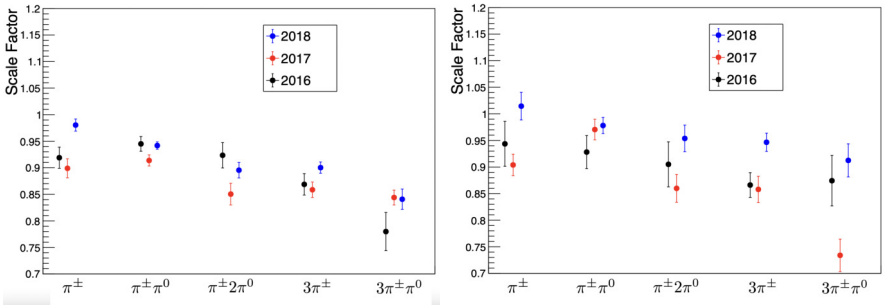

Figure C.2: The τ_h identification scale factors for different MVA decay modes and data-taking periods. The left (right) plot shows the scale factors for $20 < p_{\mathrm{T}}^{\tau_h} < 40$ GeV ($p_{\mathrm{T}}^{\tau_h} > 40$ GeV). These scale factors are derived for the MC sample and are valid for the $\tau_\mu\tau_h$ and $\tau_h\tau_h$ channels.

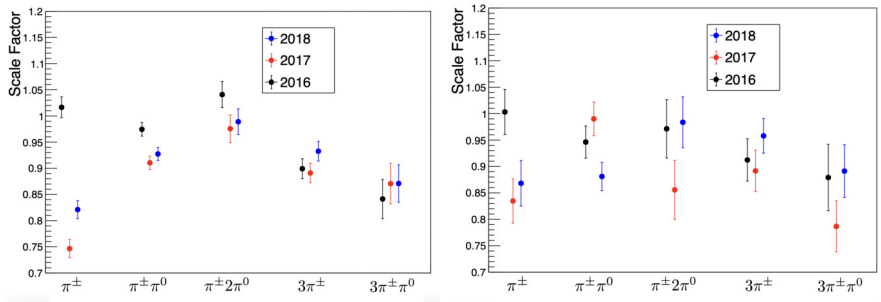

Figure C.3: The τ_h identification scale factors for different MVA decay modes and data-taking periods. The left (right) plot shows the scale factors for $20 < p_{\mathrm{T}}^{\tau_h} < 40$ GeV ($p_{\mathrm{T}}^{\tau_h} > 40$ GeV). These scale factors are derived for the embedded sample and are valid for the $\tau_e\tau_h$ channel.

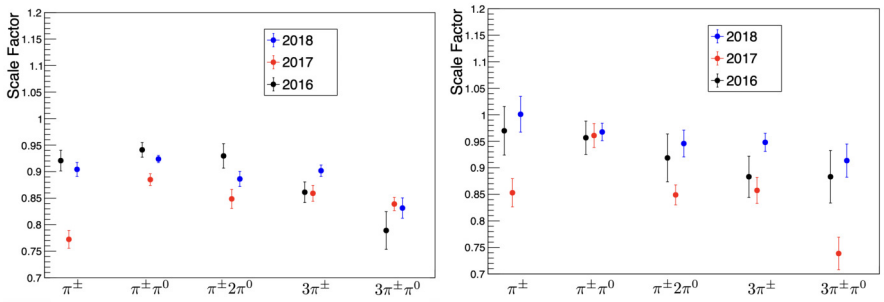

Figure C.4: The τ_h identification scale factors for different MVA decay modes and data-taking periods. The left (right) plot shows the scale factors for $20 < p_{\mathrm{T}}^{\tau_h} < 40$ GeV ($p_{\mathrm{T}}^{\tau_h} > 40$ GeV). These scale factors are derived for the MC sample and are valid for the $\tau_e\tau_h$ channel.

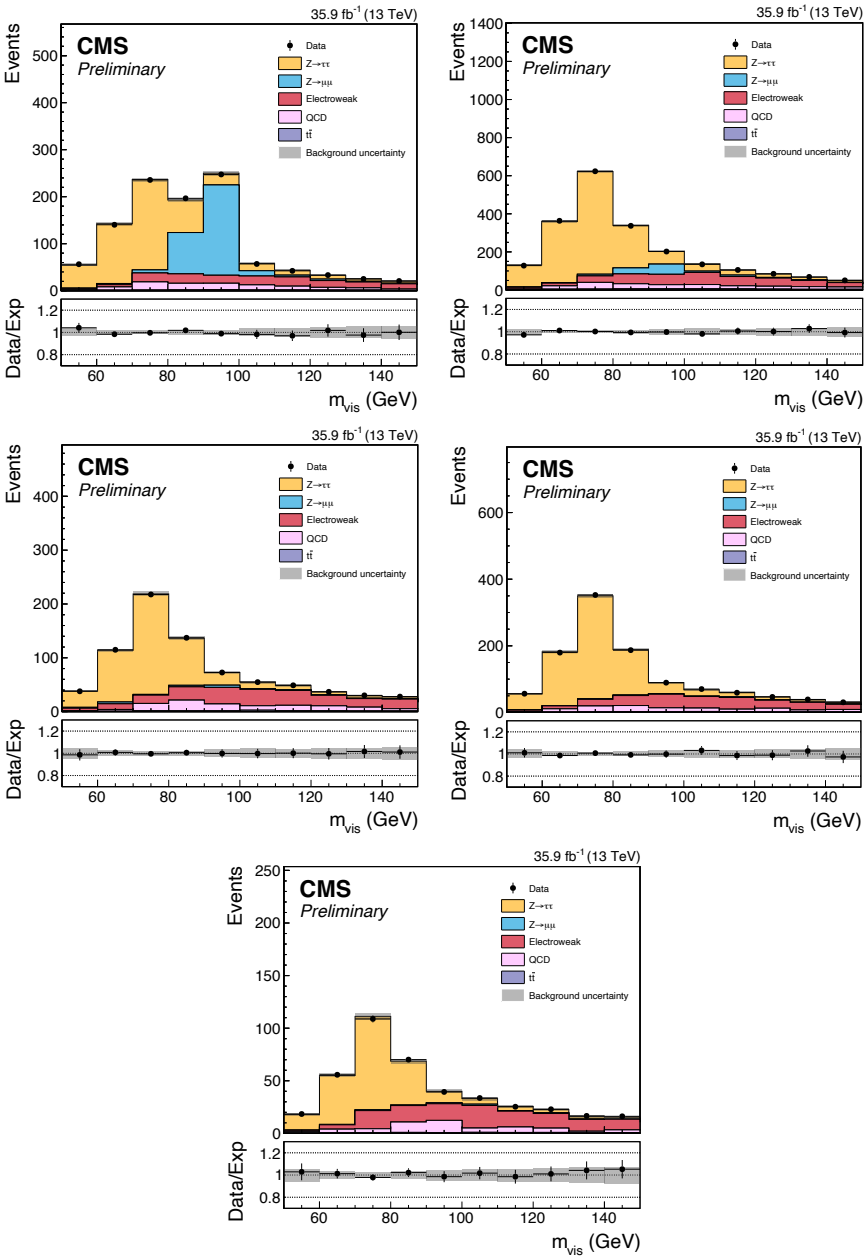

Figure C.5: Postfit plots of m_{vis} distributions in the $\tau_\mu\tau_h$ channel after applying the τ_h identification scale factors for the $\pi^\pm$ (top-left), $\pi^\pm\pi^0$ (top-right), $\pi^\pm 2\pi^0$ (middle-left), $3\pi^\pm$ (middle-right), and $3\pi^\pm\pi^0$ (bottom) MVA decay modes. The plots are for the embedded samples of 2016 data-taking period and the τ_h candidates are required to have $p_{\mathrm{T}}^{\tau_h} > 40$ GeV.

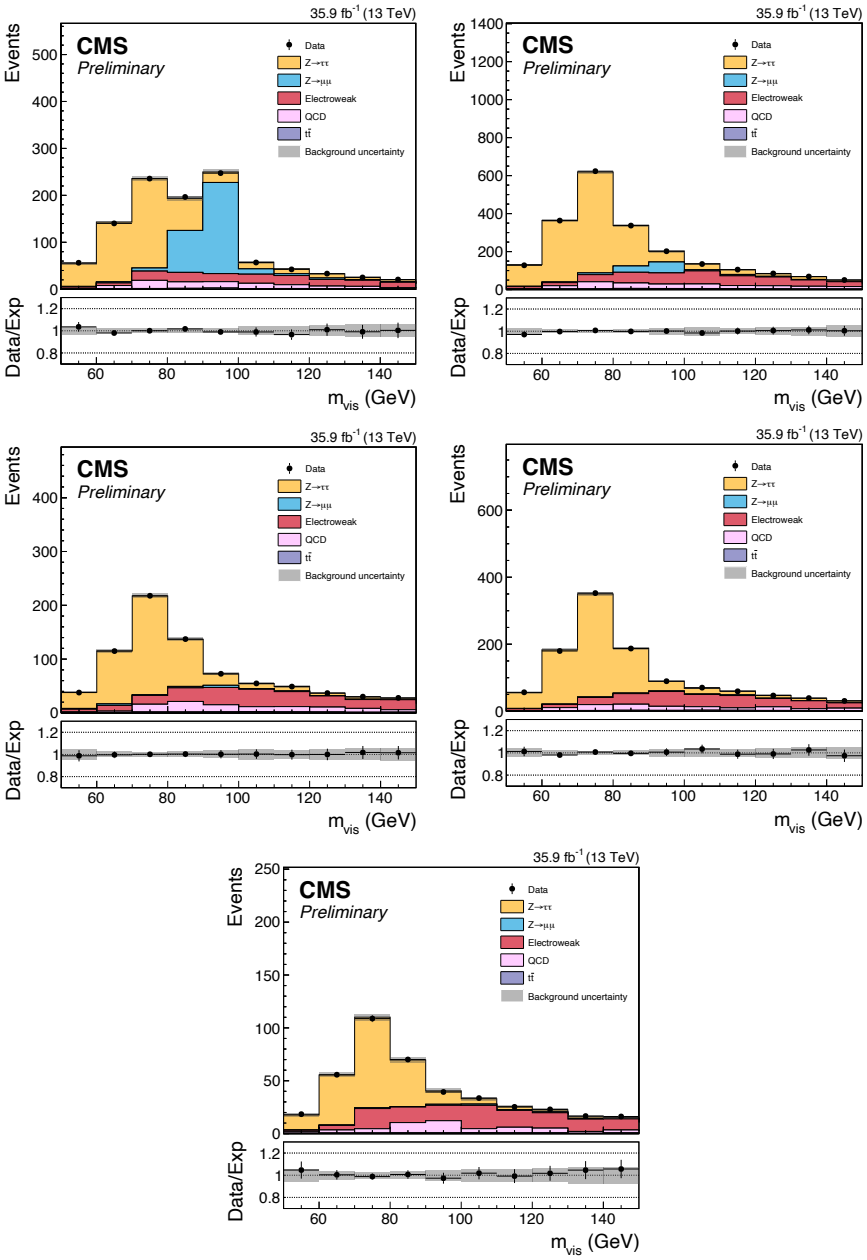

Figure C.6: Postfit plots of m_{vis} distributions in the $\tau_\mu\tau_h$ channel after applying the τ_h identification scale factors for the $\pi^\pm$ (top-left), $\pi^\pm\pi^0$ (top-right), $\pi^\pm 2\pi^0$ (middle-left), $3\pi^\pm$ (middle-right), and $3\pi^\pm\pi^0$ (bottom) MVA decay modes. The plots are for the MC samples of 2016 data-taking period and the τ_h candidates are required to have $p_{\mathrm{T}}^{\tau_h} > 40$ GeV.

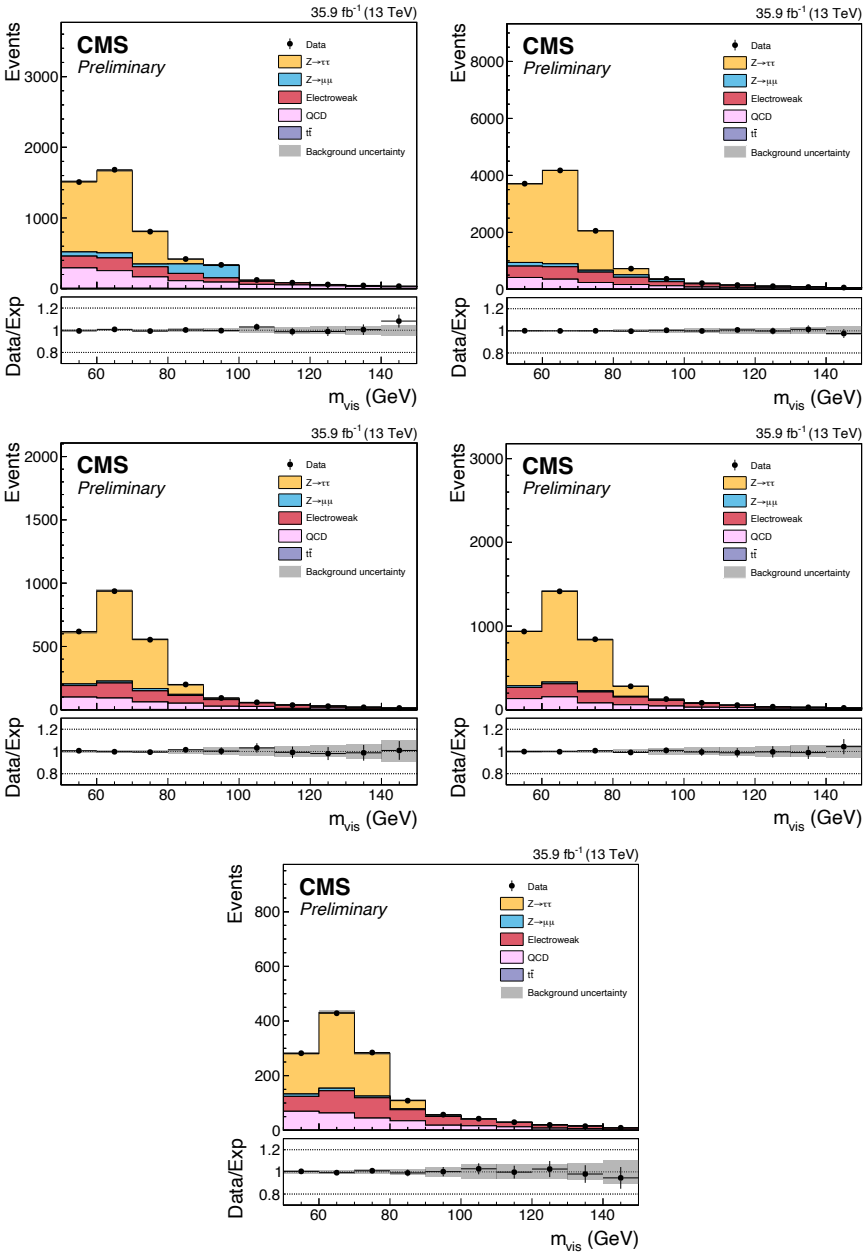

Figure C.7: Postfit plots of m_{vis} distributions in the $\tau_\mu\tau_h$ channel after applying the τ_h identification scale factors for the $\pi^\pm$ (top-left), $\pi^\pm\pi^0$ (top-right), $\pi^\pm 2\pi^0$ (middle-left), $3\pi^\pm$ (middle-right), and $3\pi^\pm\pi^0$ (bottom) MVA decay modes. The plots are for the embedded samples of 2016 data-taking period and the τ_h candidates are required to have $20 < p_{\mathrm{T}}^{\tau_h} < 40$ GeV.

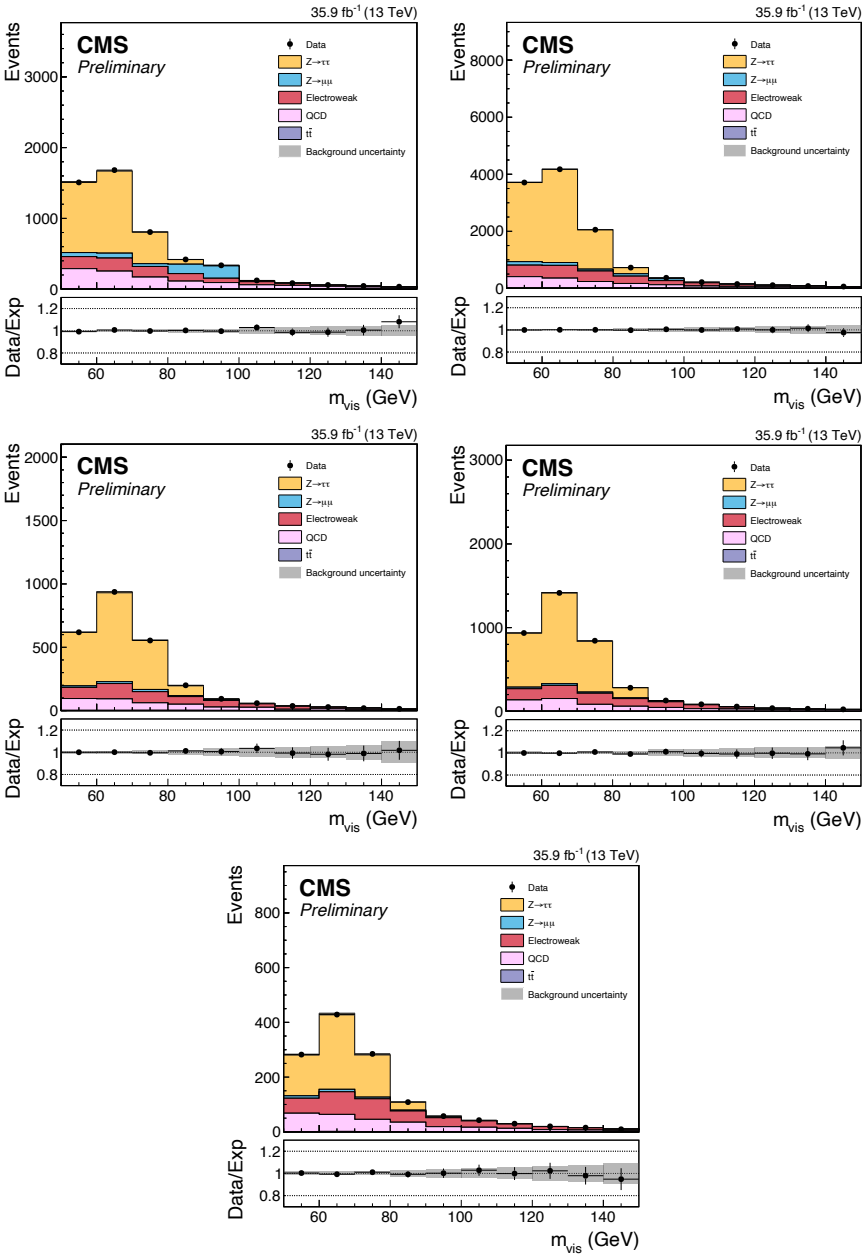

Figure C.8: Postfit plots of m_{vis} distributions in the $\tau_\mu\tau_h$ channel after applying the τ_h identification scale factors for the $\pi^\pm$ (top-left), $\pi^\pm\pi^0$ (top-right), $\pi^\pm 2\pi^0$ (middle-left), $3\pi^\pm$ (middle-right), and $3\pi^\pm\pi^0$ (bottom) MVA decay modes. The plots are for the MC samples of 2016 data-taking period and the τ_h candidates are required to have $20 < p_{\mathrm{T}}^{\tau_h} < 40$ GeV.

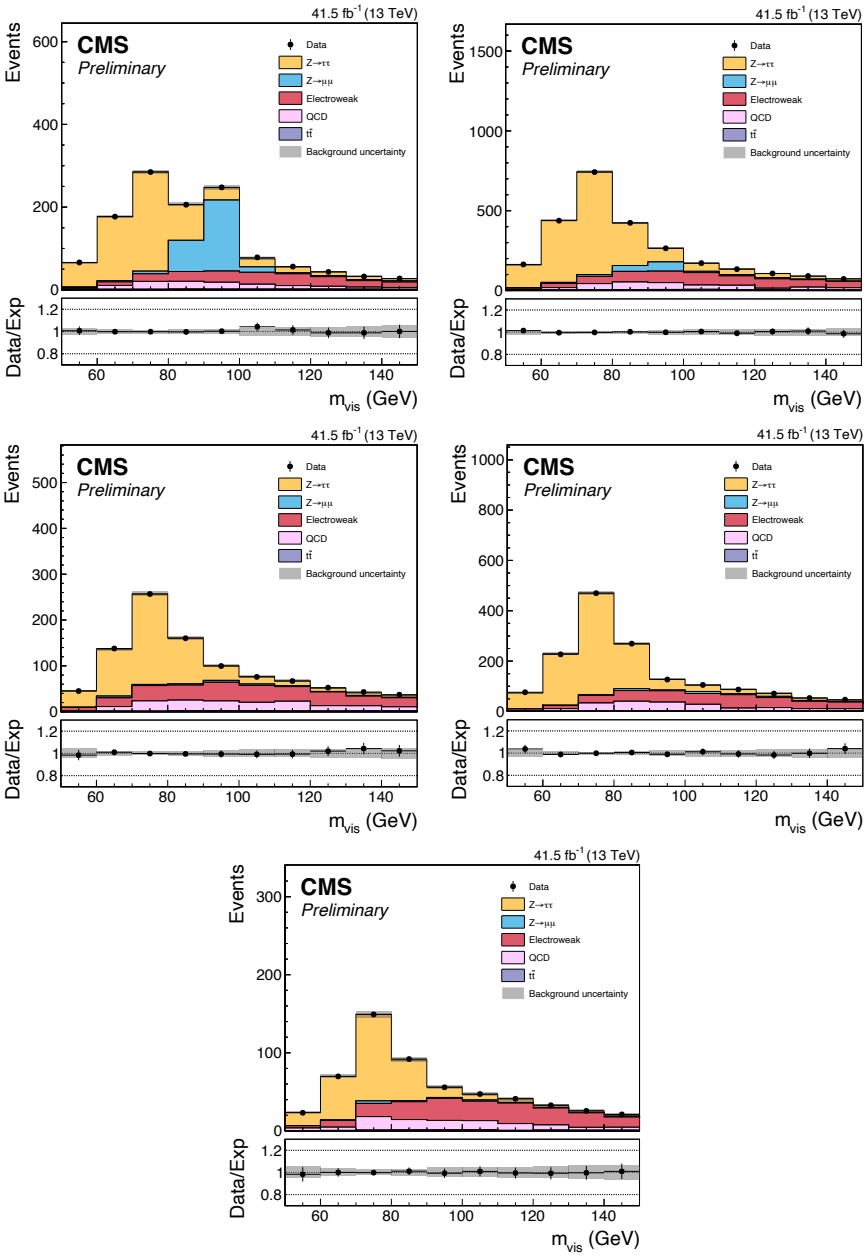

Figure C.9: Postfit plots of m_{vis} distributions in the $\tau_\mu\tau_h$ channel after applying the τ_h identification scale factors for the $\pi^\pm$ (top-left), $\pi^\pm\pi^0$ (top-right), $\pi^\pm 2\pi^0$ (middle-left), $3\pi^\pm$ (middle-right), and $3\pi^\pm\pi^0$ (bottom) MVA decay modes. The plots are for the embedded samples of 2017 data-taking period and the τ_h candidates are required to have $p_{\mathrm{T}}^{\tau_h} > 40$ GeV.

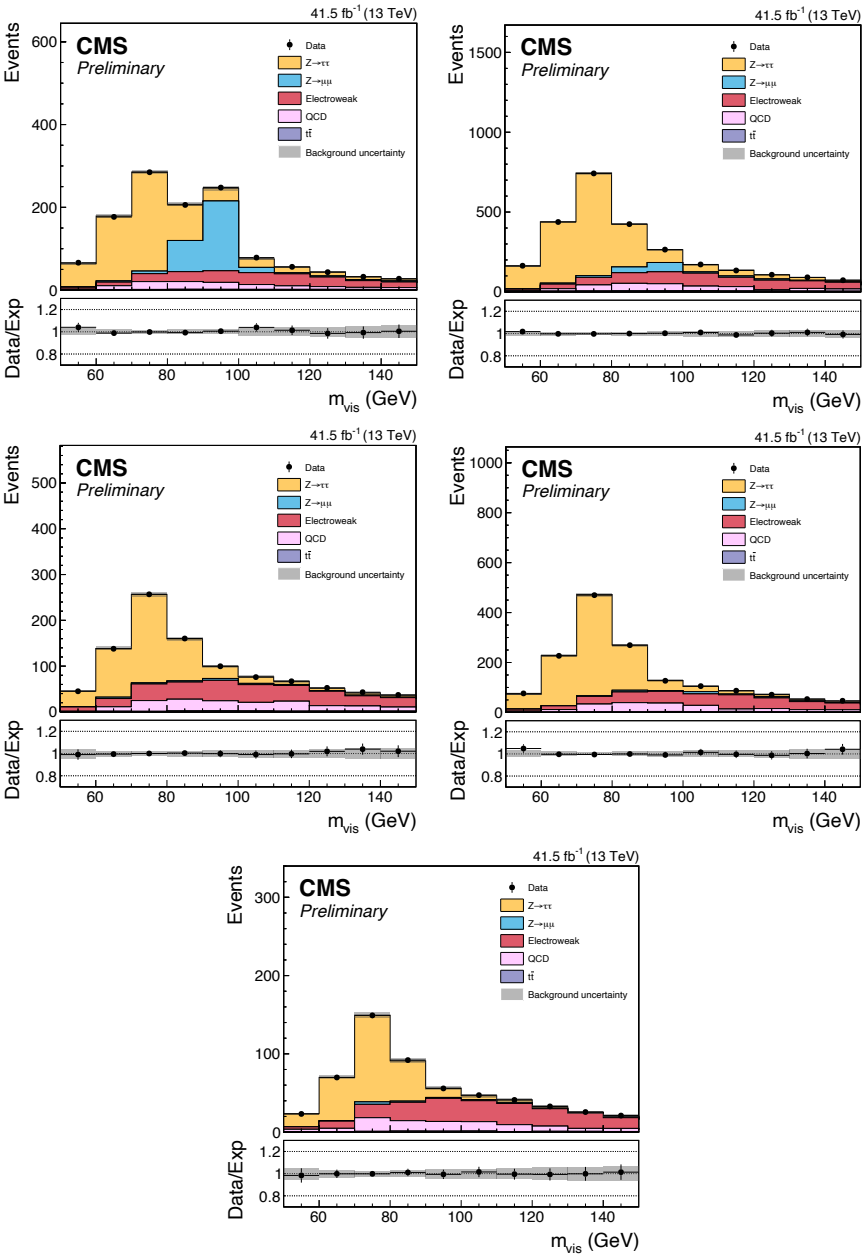

Figure C.10: Postfit plots of m_{vis} distributions in the $\tau_\mu\tau_h$ channel after applying the τ_h identification scale factors for the $\pi^\pm$ (top-left), $\pi^\pm\pi^0$ (top-right), $\pi^\pm 2\pi^0$ (middle-left), $3\pi^\pm$ (middle-right), and $3\pi^\pm\pi^0$ (bottom) MVA decay modes. The plots are for the MC samples of 2017 data-taking period and the τ_h candidates are required to have $p_{\mathrm{T}}^{\tau_h} > 40$ GeV.

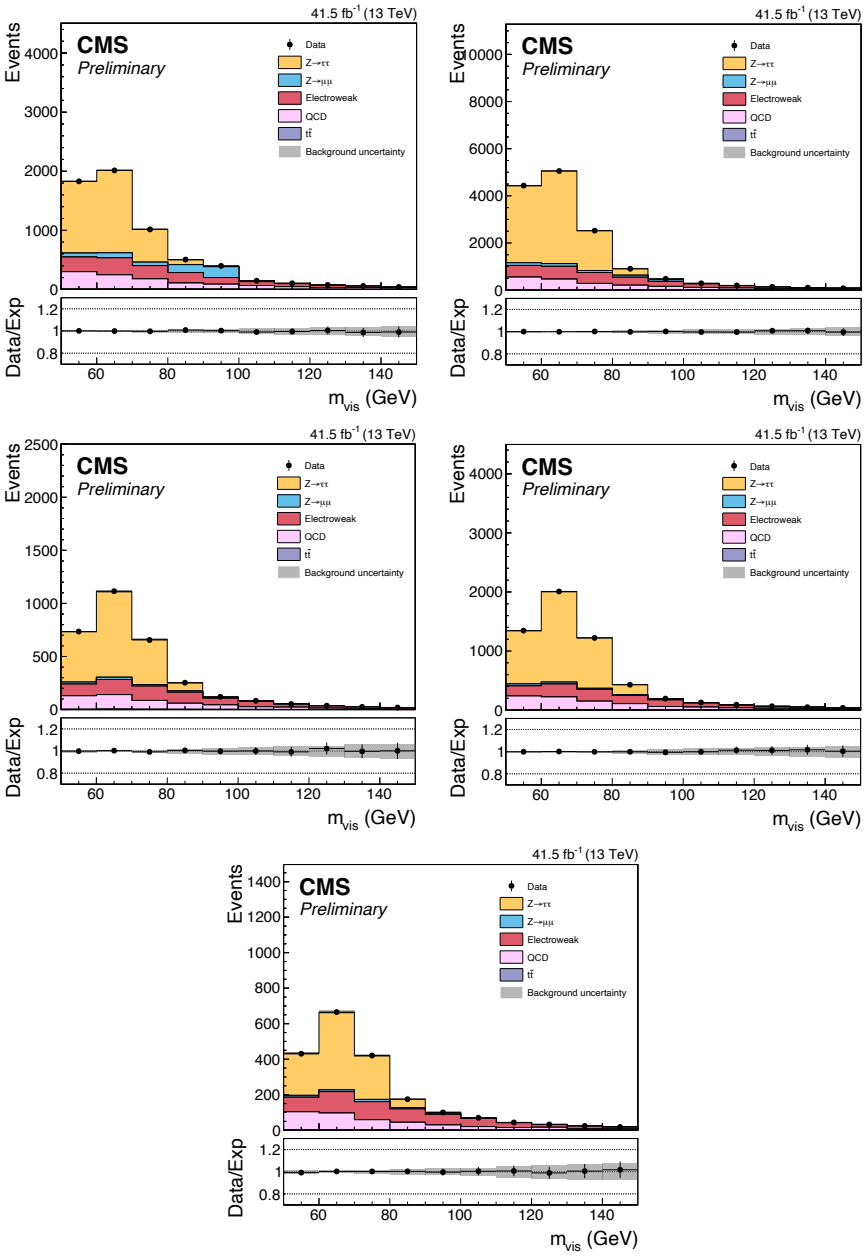

Figure C.11: Postfit plots of m_{vis} distributions in the $\tau_\mu\tau_h$ channel after applying the τ_h identification scale factors for the $\pi^\pm$ (top-left), $\pi^\pm\pi^0$ (top-right), $\pi^\pm 2\pi^0$ (middle-left), $3\pi^\pm$ (middle-right), and $3\pi^\pm\pi^0$ (bottom) MVA decay modes. The plots are for the embedded samples of 2017 data-taking period and the τ_h candidates are required to have $20 < p_{\mathrm{T}}^{\tau_h} < 40$ GeV.

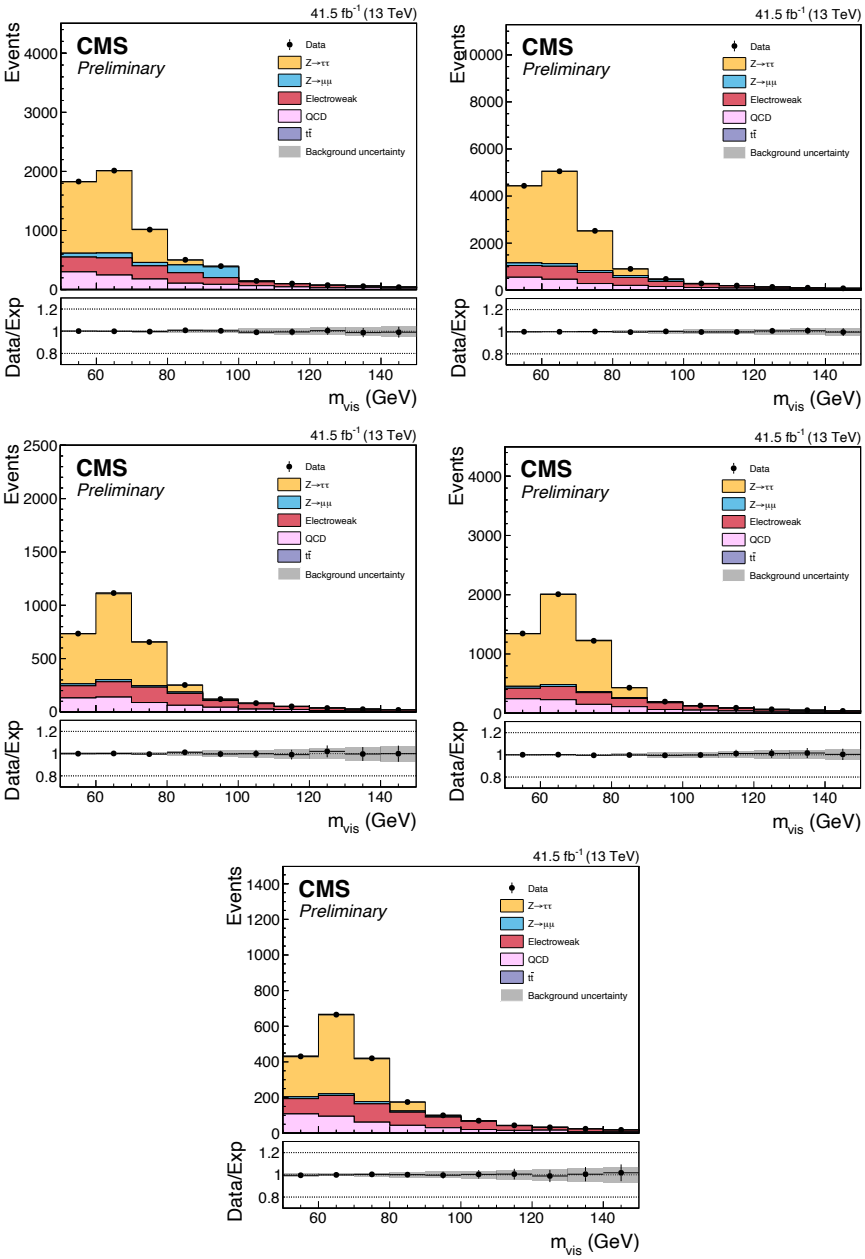

Figure C.12: Postfit plots of m_{vis} distributions in the $\tau_\mu\tau_h$ channel after applying the τ_h identification scale factors for the $\pi^\pm$ (top-left), $\pi^\pm\pi^0$ (top-right), $\pi^\pm 2\pi^0$ (middle-left), $3\pi^\pm$ (middle-right), and $3\pi^\pm\pi^0$ (bottom) MVA decay modes. The plots are for the MC samples of 2017 data-taking period and the τ_h candidates are required to have $20 < p_{\mathrm{T}}^{\tau_h} < 40$ GeV.

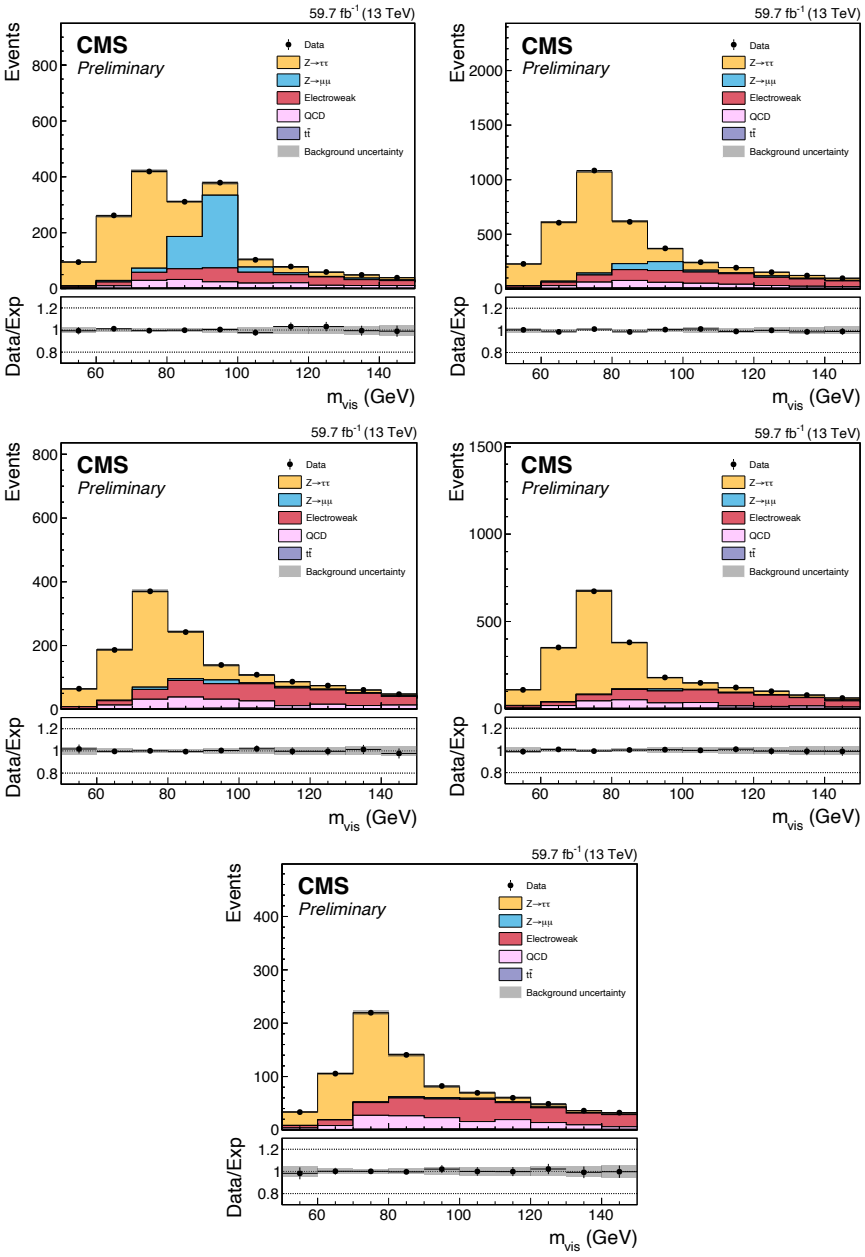

Figure C.13: Postfit plots of m_{vis} distributions in the $\tau_\mu\tau_h$ channel after applying the τ_h identification scale factors for the $\pi^\pm$ (top-left), $\pi^\pm\pi^0$ (top-right), $\pi^\pm 2\pi^0$ (middle-left), $3\pi^\pm$ (middle-right), and $3\pi^\pm\pi^0$ (bottom) MVA decay modes. The plots are for the embedded samples of 2018 data-taking period and the τ_h candidates are required to have $p_{\mathrm{T}}^{\tau_h} > 40$ GeV.

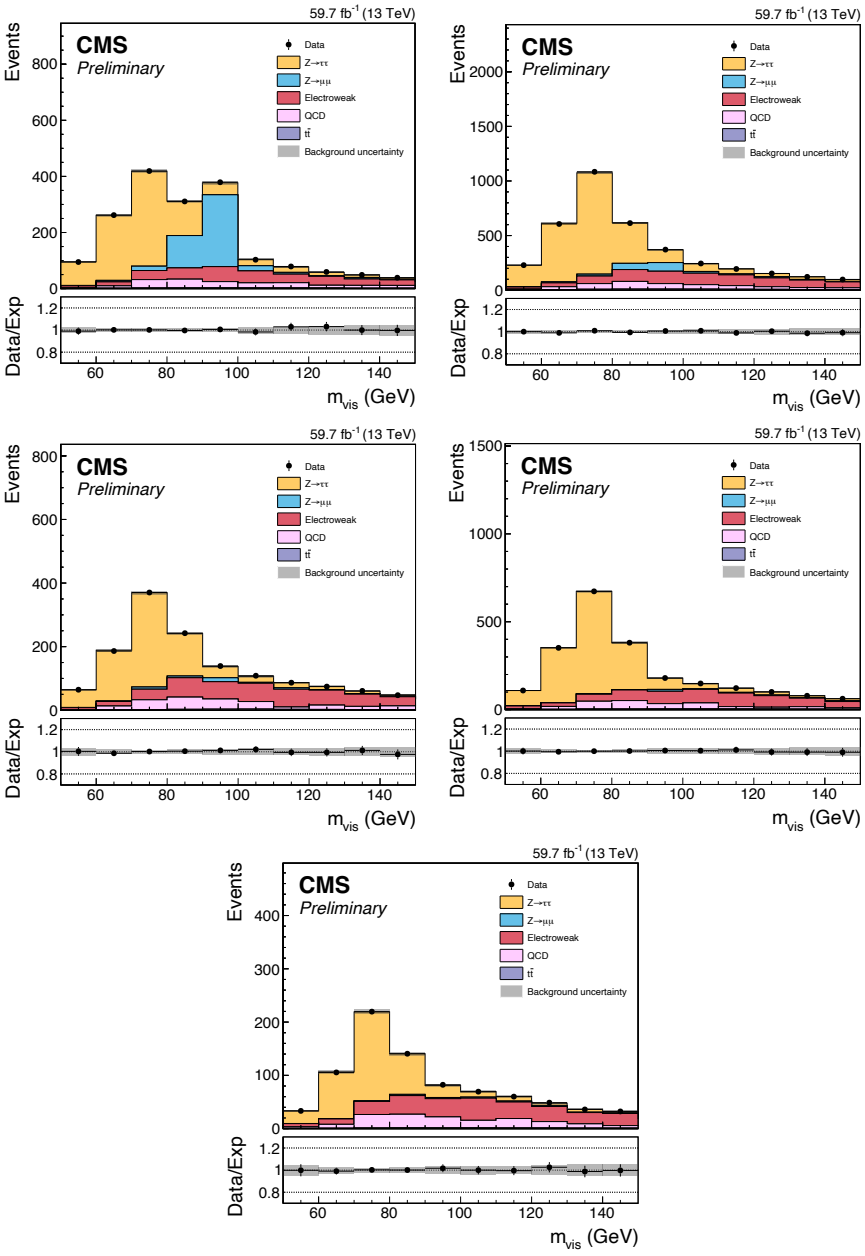

Figure C.14: Postfit plots of m_{vis} distributions in the $\tau_\mu\tau_h$ channel after applying the τ_h identification scale factors for the $\pi^\pm$ (top-left), $\pi^\pm\pi^0$ (top-right), $\pi^\pm 2\pi^0$ (middle-left), $3\pi^\pm$ (middle-right), and $3\pi^\pm\pi^0$ (bottom) MVA decay modes. The plots are for the MC samples of 2018 data-taking period and the τ_h candidates are required to have $p_{\mathrm{T}}^{\tau_h} > 40$ GeV.

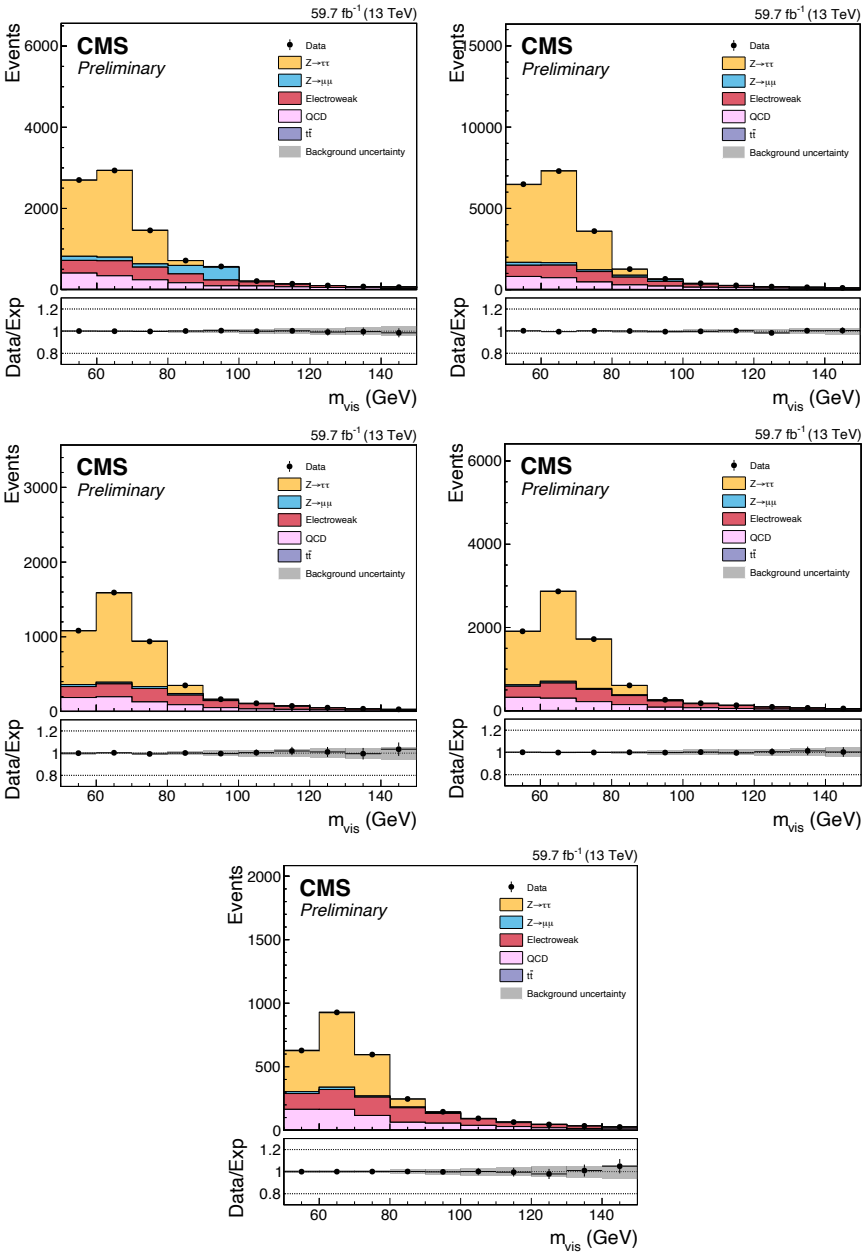

Figure C.15: Postfit plots of m_{vis} distributions in the $\tau_\mu\tau_h$ channel after applying the τ_h identification scale factors for the $\pi^\pm$ (top-left), $\pi^\pm\pi^0$ (top-right), $\pi^\pm 2\pi^0$ (middle-left), $3\pi^\pm$ (middle-right), and $3\pi^\pm\pi^0$ (bottom) MVA decay modes. The plots are for the embedded samples of 2018 data-taking period and the τ_h candidates are required to have $20 < p_{\mathrm{T}}^{\tau_h} < 40$ GeV.

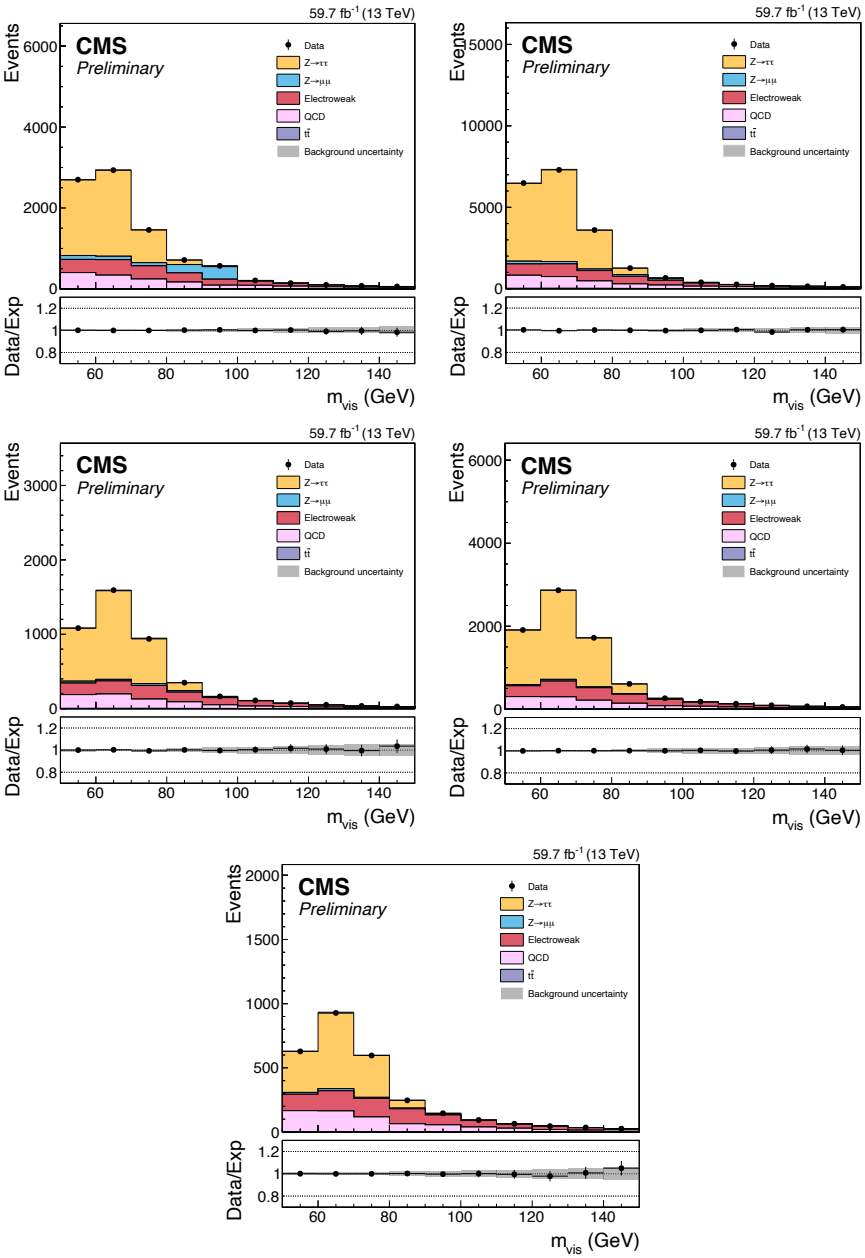

Figure C.16: Postfit plots of m_{vis} distributions in the $\tau_\mu\tau_h$ channel after applying the τ_h identification scale factors for the $\pi^\pm$ (top-left), $\pi^\pm\pi^0$ (top-right), $\pi^\pm 2\pi^0$ (middle-left), $3\pi^\pm$ (middle-right), and $3\pi^\pm\pi^0$ (bottom) MVA decay modes. The plots are for the MC samples of 2018 data-taking period and the τ_h candidates are required to have $20 < p_{\mathrm{T}}^{\tau_h} < 40$ GeV.

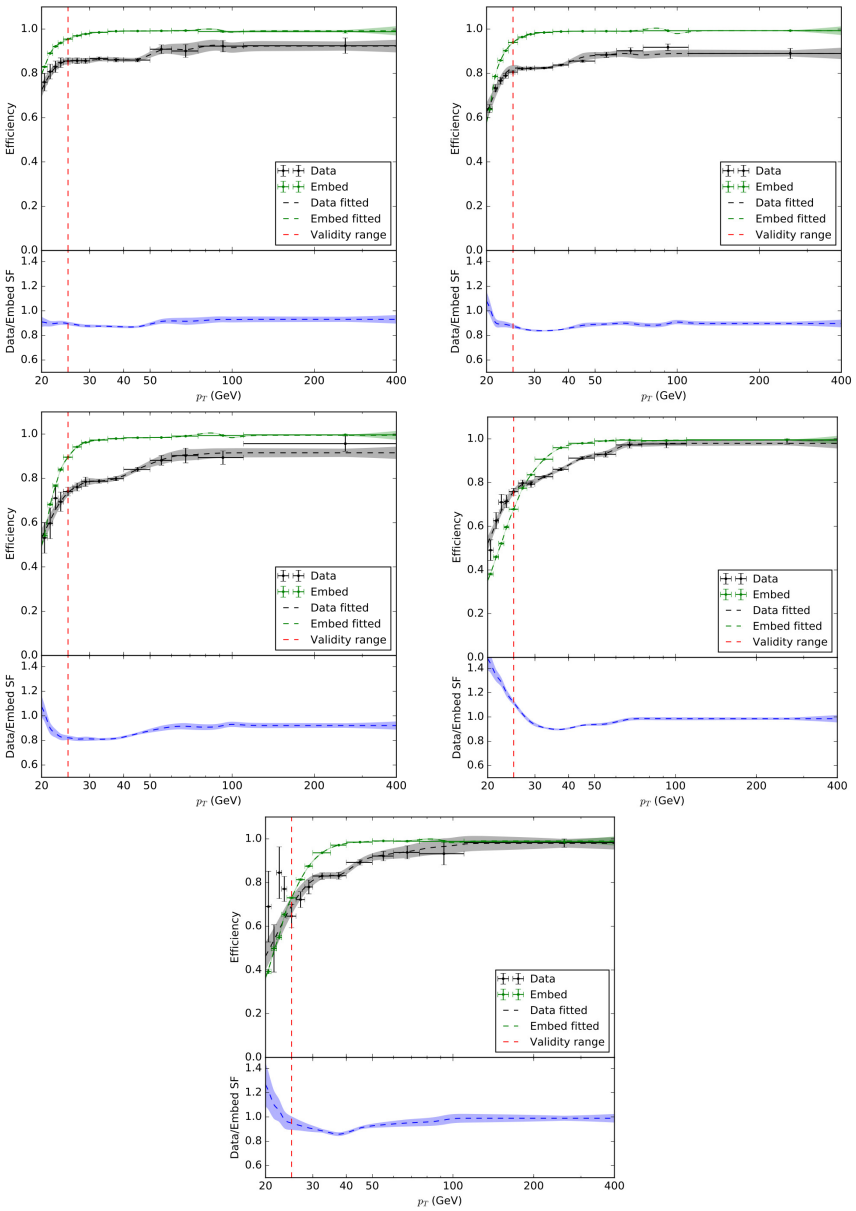

Figure C.17: The efficiencies and scale factors of the τ_h trigger in the $\tau_\mu\tau_h$ channel as a function of $p_{\mathrm{T}}^{\tau_h}$ using 2016 data and embedded samples for the $\pi^\pm$ (top-left), $\pi^\pm\pi^0$ (top-right), $\pi^\pm 2\pi^0$ (middle-left), $3\pi^\pm$ (middle-right), and $3\pi^\pm\pi^0$ (bottom) MVA decay modes. The efficiency graphs along with their fits and fit uncertainties are shown in the upper panel of each sub-figure while the scale factors are shown in the lower panel. The red dashed line shows the offline threshold below which the events are excluded from the analysis.

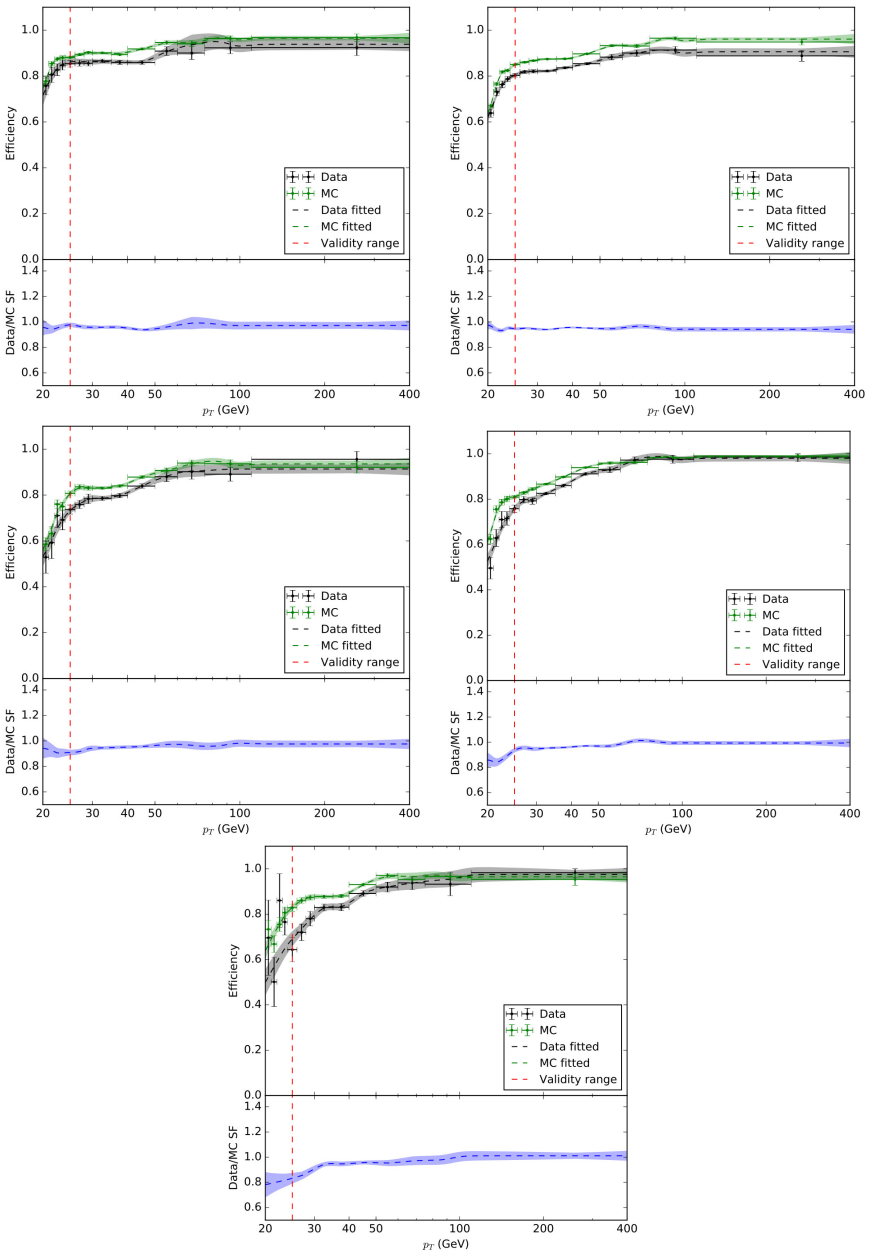

Figure C.18: The efficiencies and scale factors of the τ_h trigger in the $\tau_\mu\tau_h$ channel as a function of $p_{\mathrm{T}}^{\tau_h}$ using 2016 data and MC samples for the $\pi^\pm$ (top-left), $\pi^\pm\pi^0$ (top-right), $\pi^\pm 2\pi^0$ (middle-left), $3\pi^\pm$ (middle-right), and $3\pi^\pm\pi^0$ (bottom) MVA decay modes. The efficiency graphs along with their fits and fit uncertainties are shown in the upper panel of each sub-figure while the scale factors are shown in the lower panel. The red dashed line shows the offline threshold below which the events are excluded from the analysis.

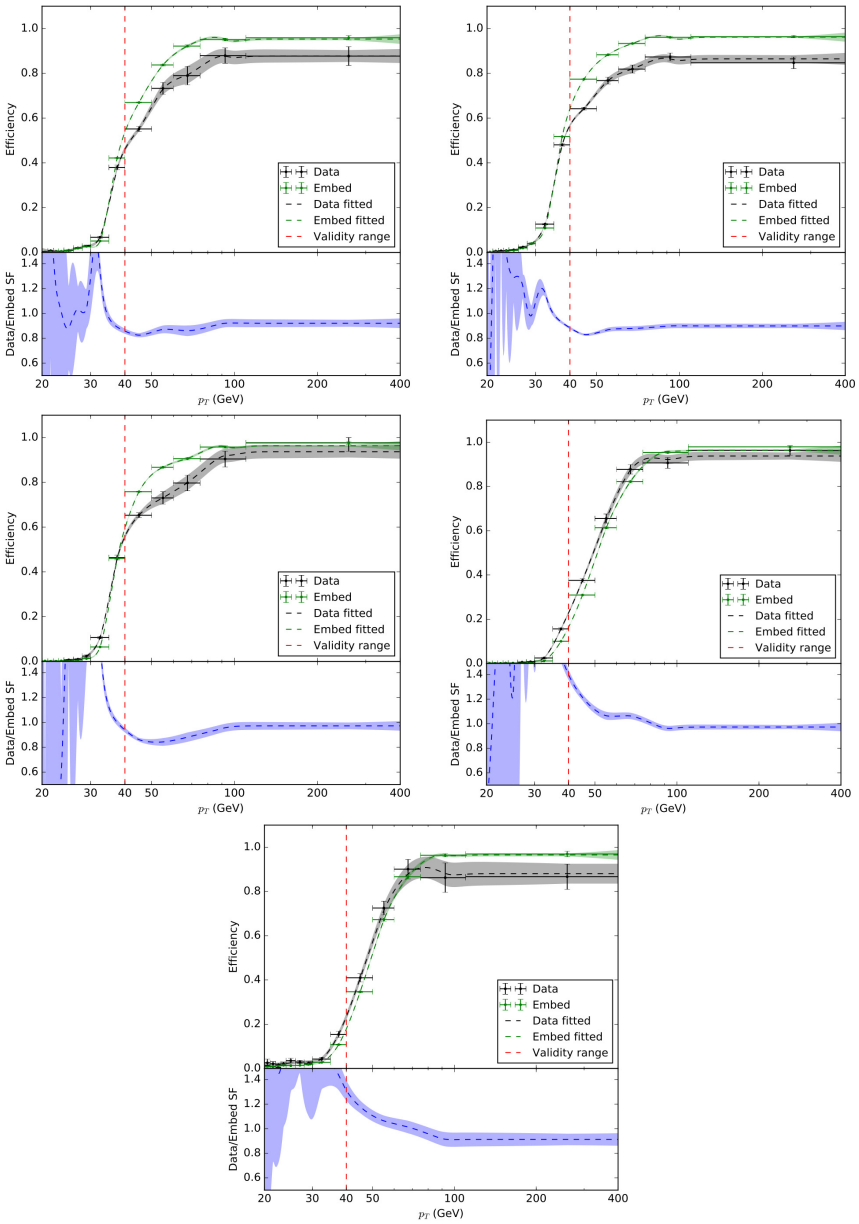

Figure C.19: The efficiencies and scale factors of the τ_h trigger in the $\tau_h\tau_h$ channel as a function of $p_{\mathrm{T}}^{\tau_h}$ using 2016 data and embedded samples for the $\pi^{\pm}$ (top-left), $\pi^{\pm}\pi^0$ (top-right), $\pi^{\pm}2\pi^0$ (middle-left), $3\pi^{\pm}$ (middle-right), and $3\pi^{\pm}\pi^0$ (bottom) MVA decay modes. The efficiency graphs along with their fits and fit uncertainties are shown in the upper panel of each sub-figure while the scale factors are shown in the lower panel. The red dashed line shows the offline threshold below which the events are excluded from the analysis.

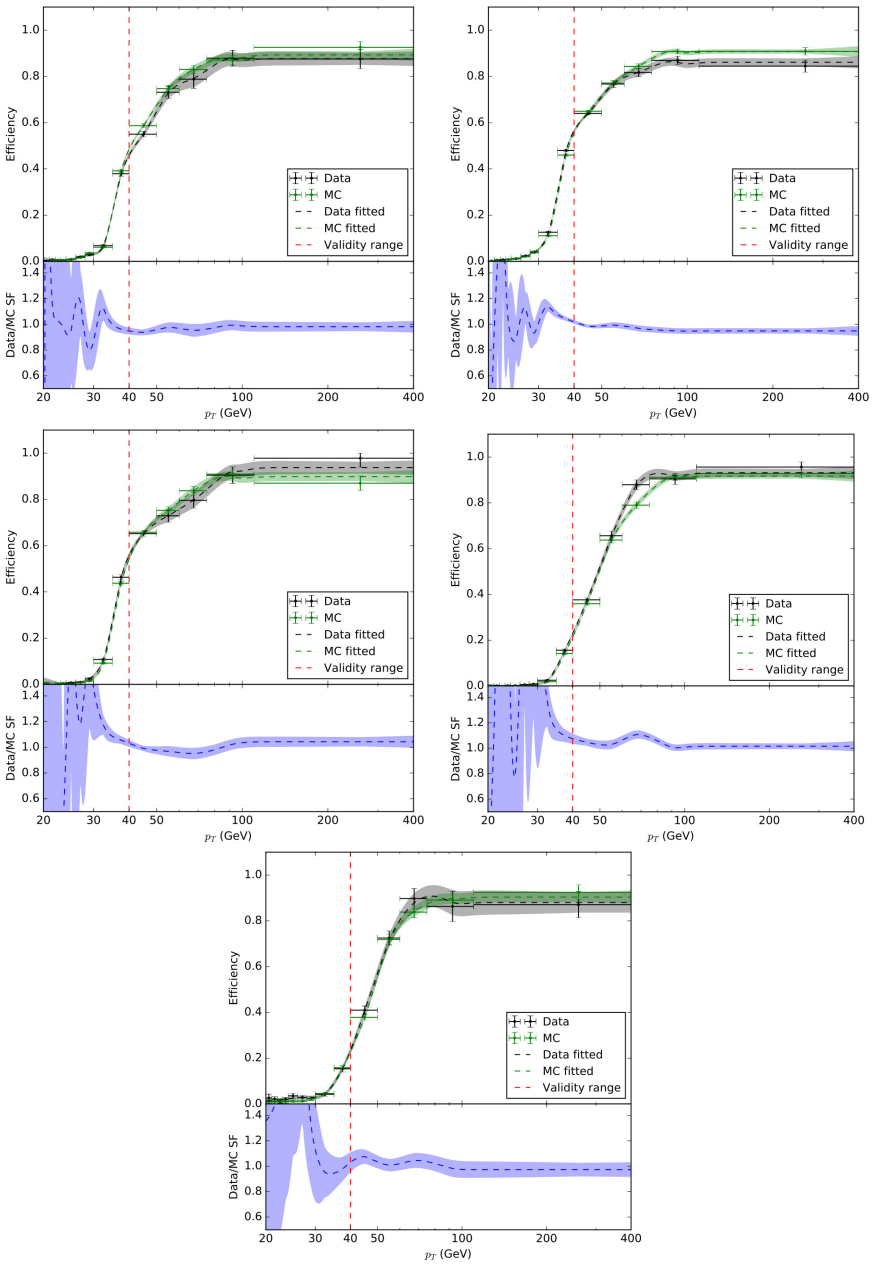

Figure C.20: The efficiencies and scale factors of the τ_h trigger in the $\tau_h\tau_h$ channel as a function of $p_{\mathrm{T}}^{\tau_h}$ using 2016 data and MC samples for the $\pi^{\pm}$ (top-left), $\pi^{\pm}\pi^0$ (top-right), $\pi^{\pm}2\pi^0$ (middle-left), $3\pi^{\pm}$ (middle-right), and $3\pi^{\pm}\pi^0$ (bottom) MVA decay modes. The efficiency graphs along with their fits and fit uncertainties are shown in the upper panel of each sub-figure while the scale factors are shown in the lower panel. The red dashed line shows the offline threshold below which the events are excluded from the analysis.

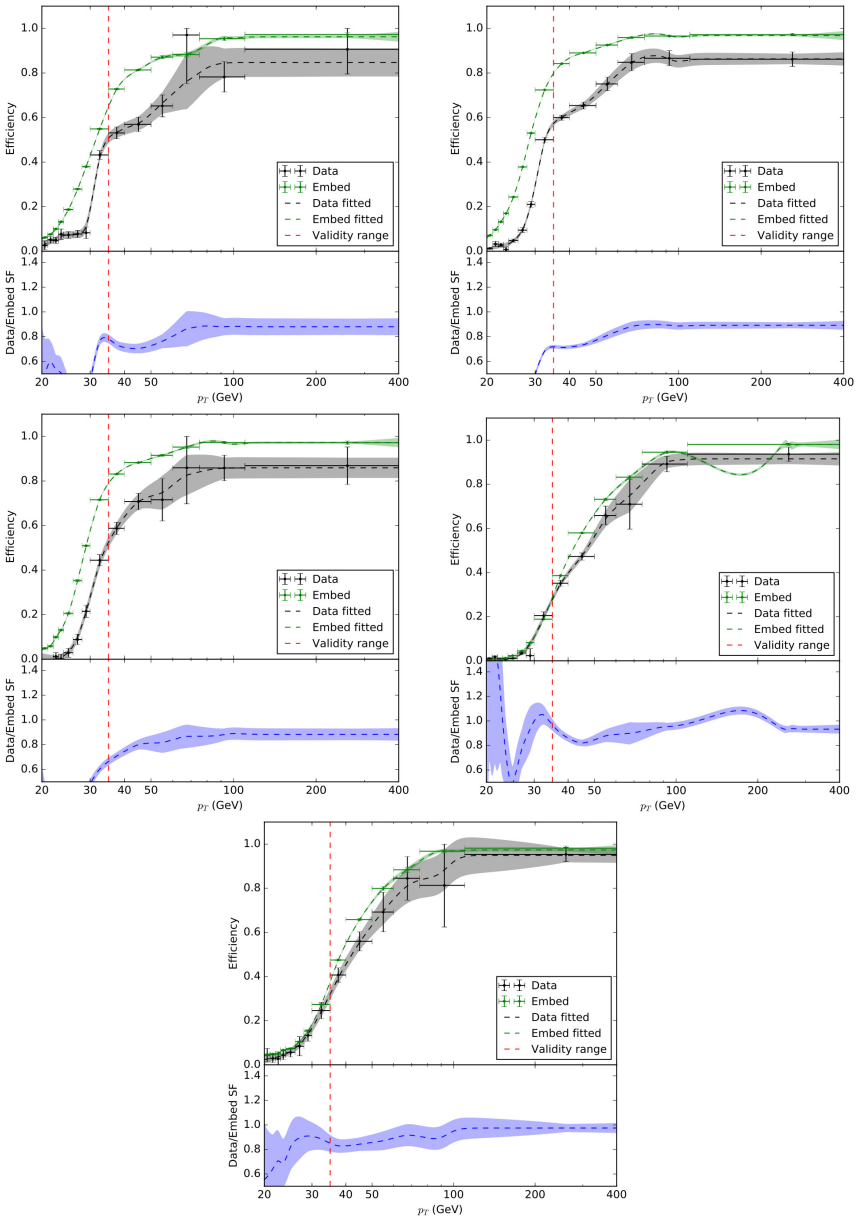

Figure C.21: The efficiencies and scale factors of the τ_h trigger in the $\tau_e\tau_h$ channel as a function of $p_{\mathrm{T}}^{\tau_h}$ using 2017 data and embedded samples for the $\pi^{\pm}$ (top-left), $\pi^{\pm}\pi^0$ (top-right), $\pi^{\pm}2\pi^0$ (middle-left), $3\pi^{\pm}$ (middle-right), and $3\pi^{\pm}\pi^0$ (bottom) MVA decay modes. The efficiency graphs along with their fits and fit uncertainties are shown in the upper panel of each sub-figure while the scale factors are shown in the lower panel. The red dashed line shows the offline threshold below which the events are excluded from the analysis.

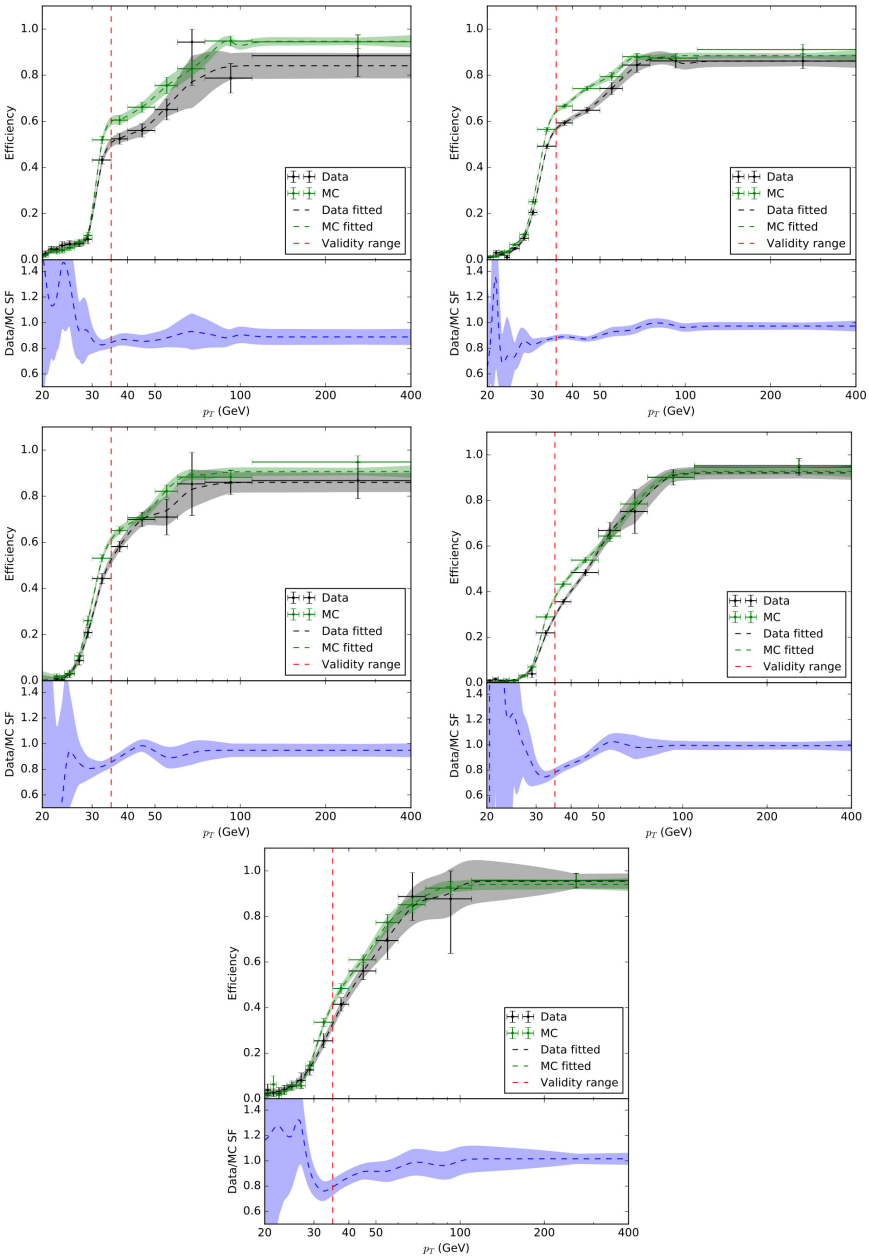

Figure C.22: The efficiencies and scale factors of the τ_h trigger in the $\tau_e\tau_h$ channel as a function of $p_{\mathrm{T}}^{\tau_h}$ using 2017 data and MC samples for the $\pi^{\pm}$ (top-left), $\pi^{\pm}\pi^0$ (top-right), $\pi^{\pm}2\pi^0$ (middle-left), $3\pi^{\pm}$ (middle-right), and $3\pi^{\pm}\pi^0$ (bottom) MVA decay modes. The efficiency graphs along with their fits and fit uncertainties are shown in the upper panel of each sub-figure while the scale factors are shown in the lower panel. The red dashed line shows the offline threshold below which the events are excluded from the analysis.

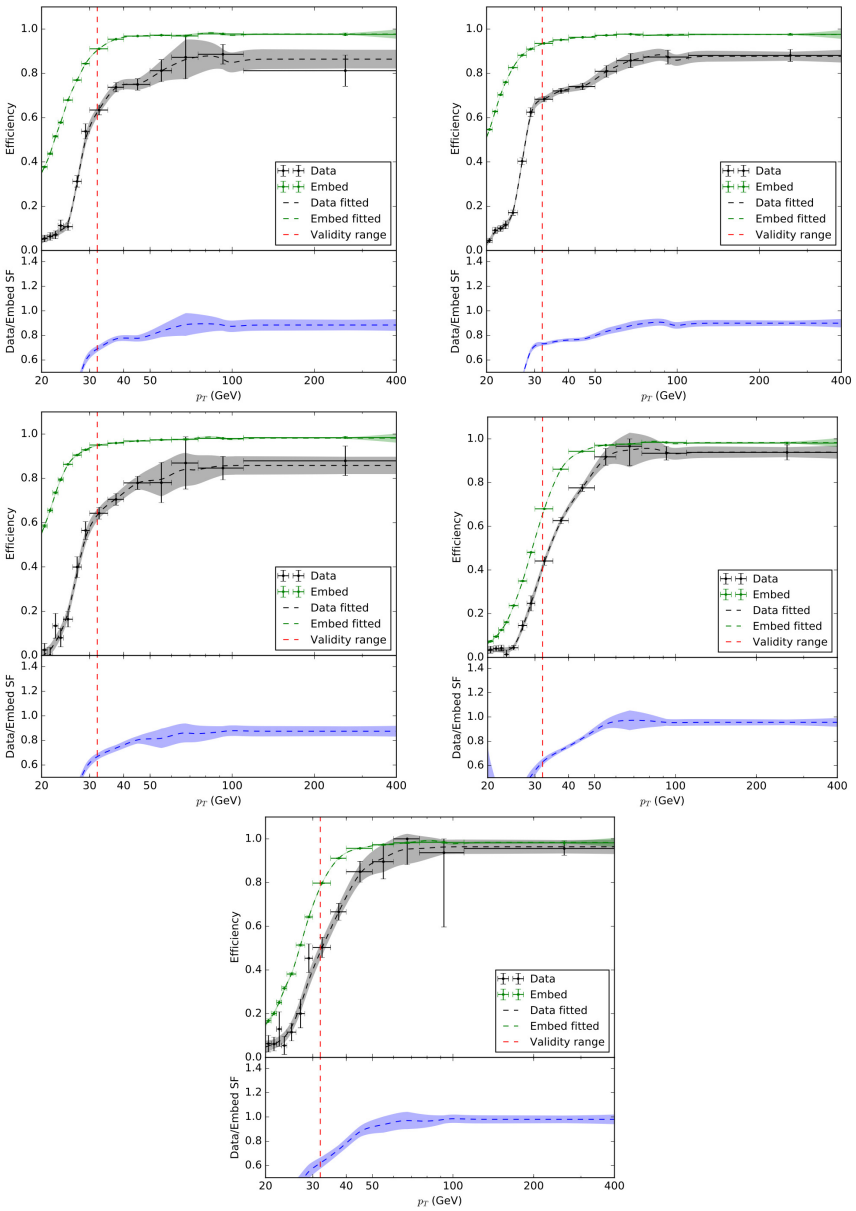

Figure C.23: The efficiencies and scale factors of the τ_h trigger in the $\tau_\mu\tau_h$ channel as a function of $p_{\mathrm{T}}^{\tau_h}$ using 2017 data and embedded samples for the $\pi^\pm$ (top-left), $\pi^\pm\pi^0$ (top-right), $\pi^\pm 2\pi^0$ (middle-left), $3\pi^\pm$ (middle-right), and $3\pi^\pm\pi^0$ (bottom) MVA decay modes. The efficiency graphs along with their fits and fit uncertainties are shown in the upper panel of each sub-figure while the scale factors are shown in the lower panel. The red dashed line shows the offline threshold below which the events are excluded from the analysis.

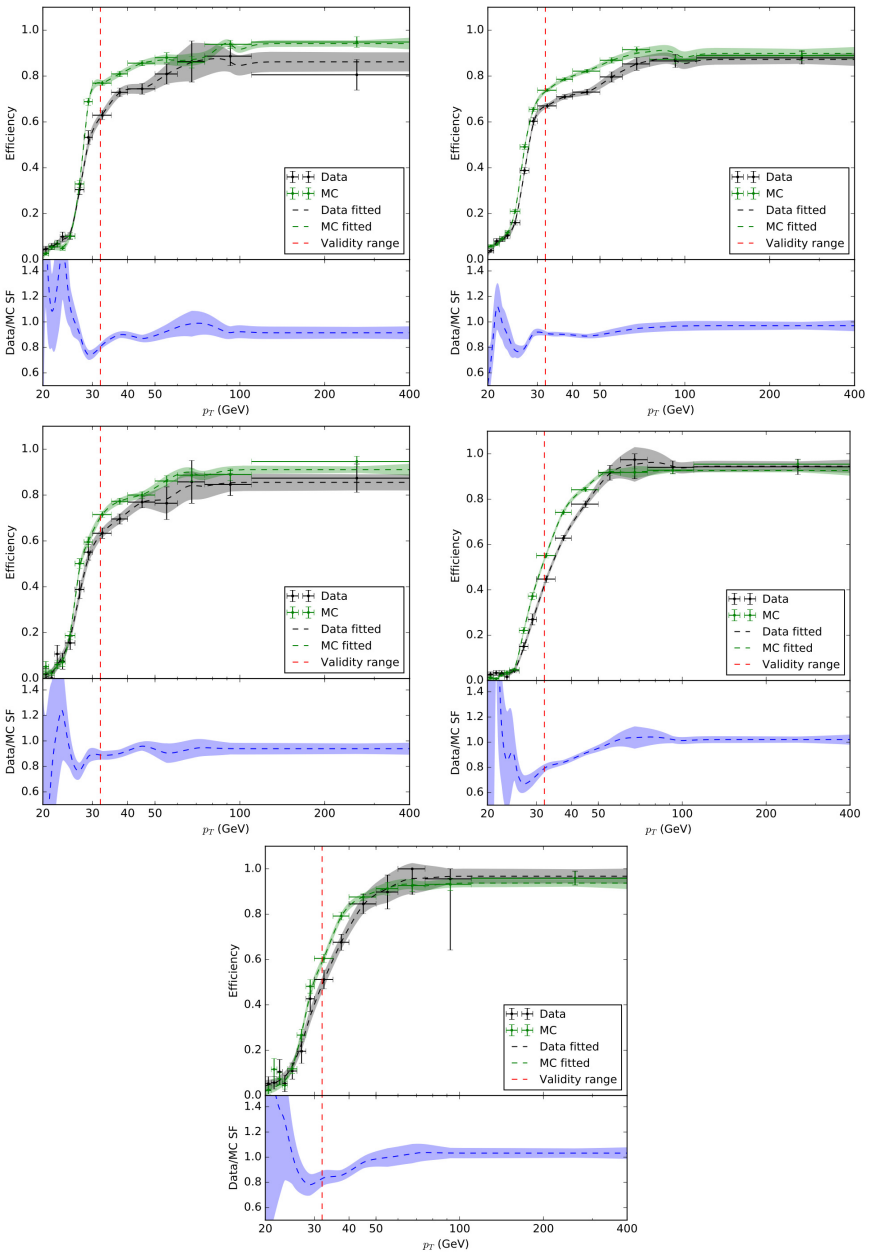

Figure C.24: The efficiencies and scale factors of the τ_h trigger in the $\tau_\mu\tau_h$ channel as a function of $p_{\mathrm{T}}^{\tau_h}$ using 2017 data and MC samples for the $\pi^\pm$ (top-left), $\pi^\pm\pi^0$ (top-right), $\pi^\pm 2\pi^0$ (middle-left), $3\pi^\pm$ (middle-right), and $3\pi^\pm\pi^0$ (bottom) MVA decay modes. The efficiency graphs along with their fits and fit uncertainties are shown in the upper panel of each sub-figure while the scale factors are shown in the lower panel. The red dashed line shows the offline threshold below which the events are excluded from the analysis.

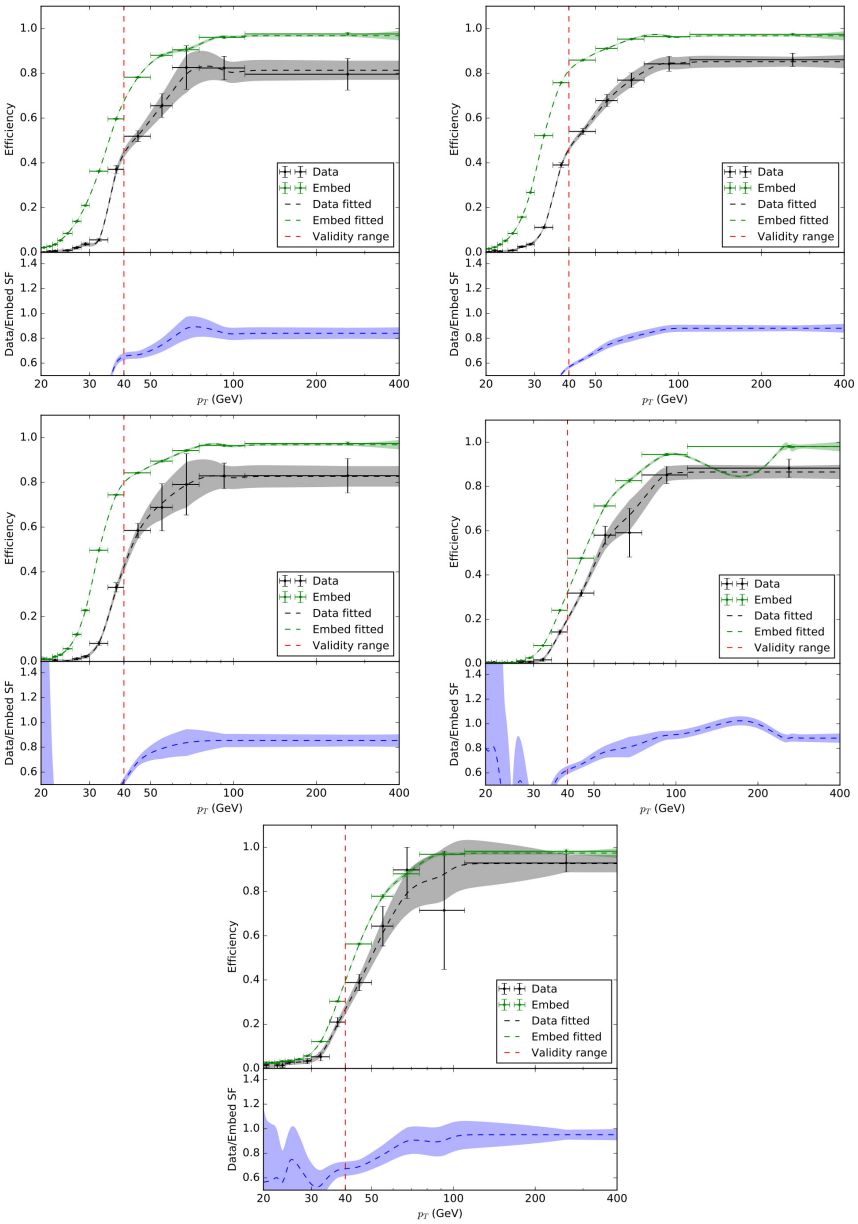

Figure C.25: The efficiencies and scale factors of the τ_h trigger in the $\tau_h\tau_h$ channel as a function of $p_{\mathrm{T}}^{\tau_h}$ using 2017 data and embedded samples for the $\pi^{\pm}$ (top-left), $\pi^{\pm}\pi^0$ (top-right), $\pi^{\pm}2\pi^0$ (middle-left), $3\pi^{\pm}$ (middle-right), and $3\pi^{\pm}\pi^0$ (bottom) MVA decay modes. The efficiency graphs along with their fits and fit uncertainties are shown in the upper panel of each sub-figure while the scale factors are shown in the lower panel. The red dashed line shows the offline threshold below which the events are excluded from the analysis.

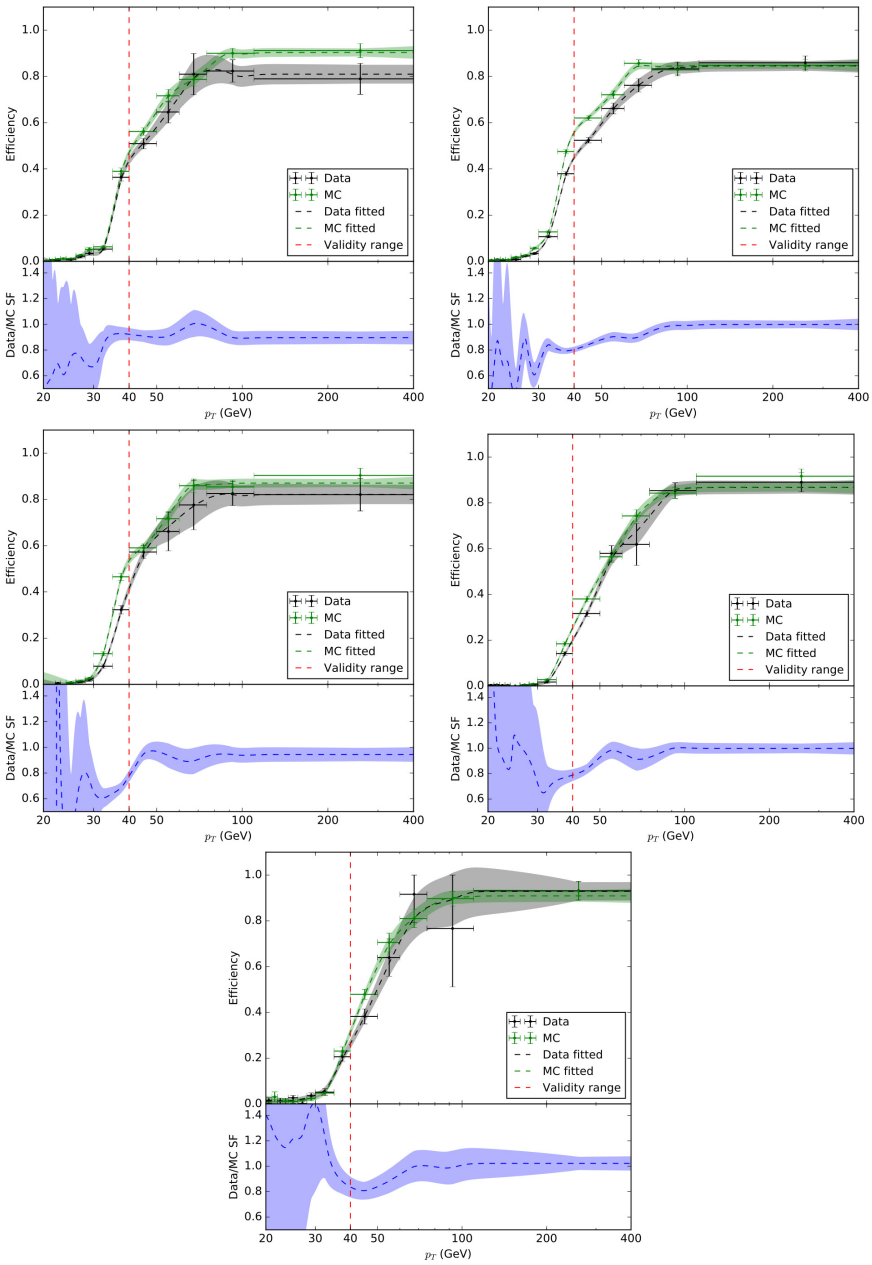

Figure C.26: The efficiencies and scale factors of the τ_h trigger in the $\tau_h\tau_h$ channel as a function of $p_{\mathrm{T}}^{\tau_h}$ using 2017 data and MC samples for the $\pi^{\pm}$ (top-left), $\pi^{\pm}\pi^0$ (top-right), $\pi^{\pm}2\pi^0$ (middle-left), $3\pi^{\pm}$ (middle-right), and $3\pi^{\pm}\pi^0$ (bottom) MVA decay modes. The efficiency graphs along with their fits and fit uncertainties are shown in the upper panel of each sub-figure while the scale factors are shown in the lower panel. The red dashed line shows the offline threshold below which the events are excluded from the analysis.

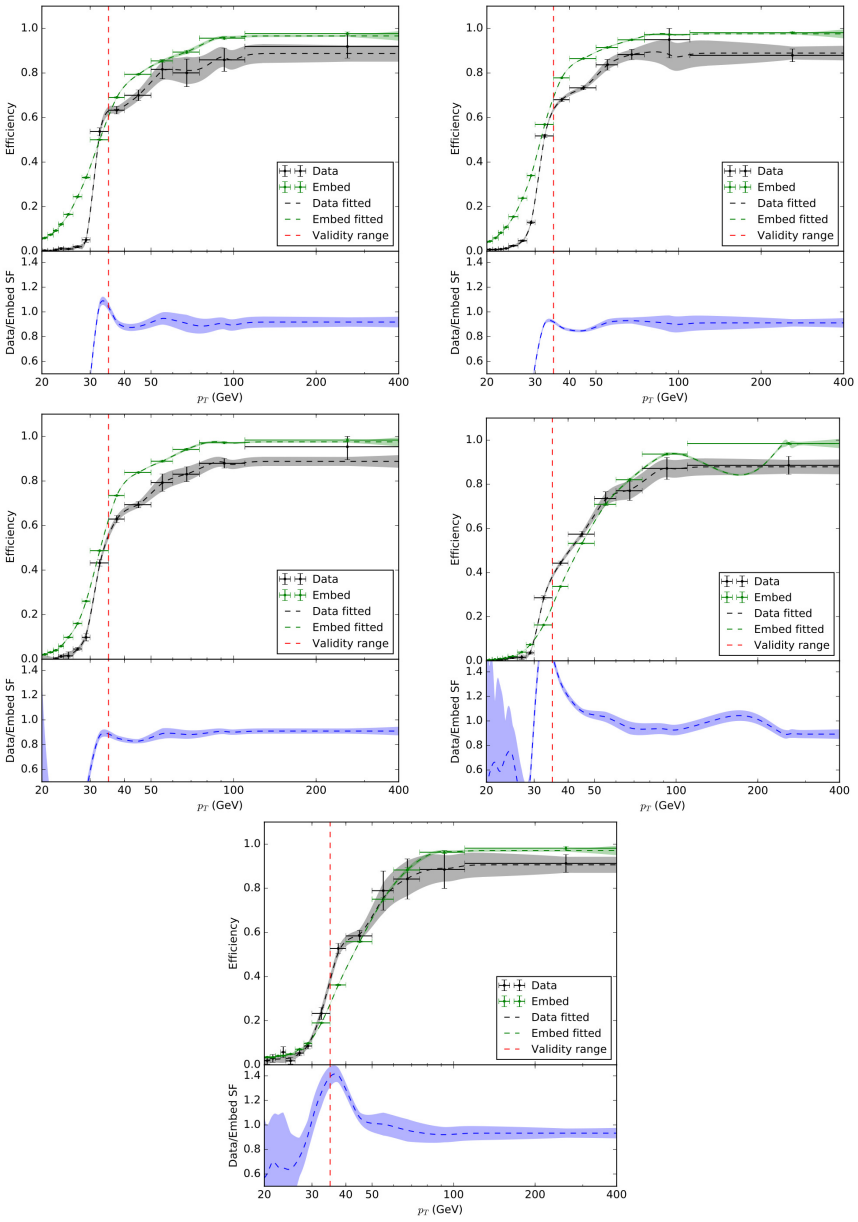

Figure C.27: The efficiencies and scale factors of the τ_h trigger in the $\tau_e\tau_h$ channel as a function of $p_{\mathrm{T}}^{\tau_h}$ using 2018 data and embedded samples for the $\pi^{\pm}$ (top-left), $\pi^{\pm}\pi^0$ (top-right), $\pi^{\pm}2\pi^0$ (middle-left), $3\pi^{\pm}$ (middle-right), and $3\pi^{\pm}\pi^0$ (bottom) MVA decay modes. The efficiency graphs along with their fits and fit uncertainties are shown in the upper panel of each sub-figure while the scale factors are shown in the lower panel. The red dashed line shows the offline threshold below which the events are excluded from the analysis.

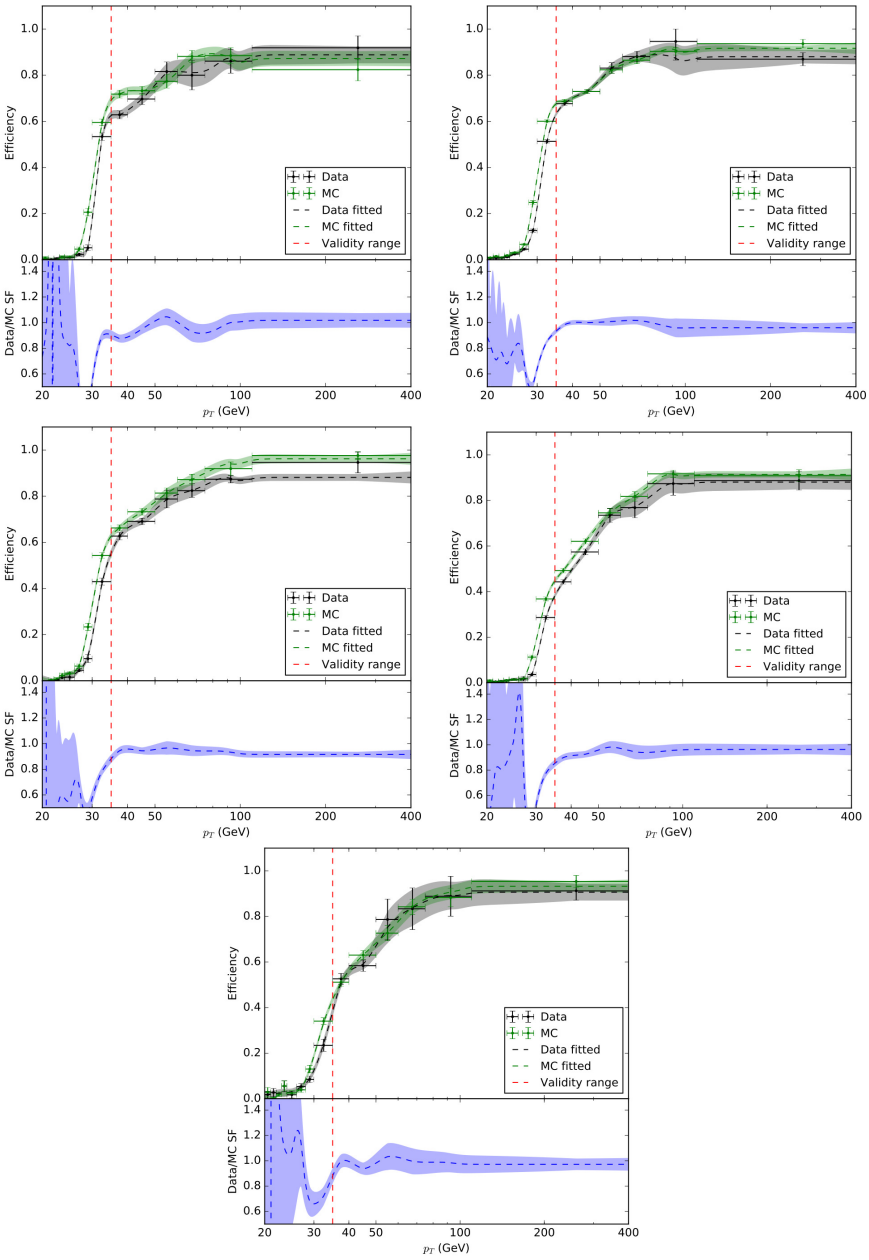

Figure C.28: The efficiencies and scale factors of the τ_h trigger in the $\tau_e\tau_h$ channel as a function of $p_{\mathrm{T}}^{\tau_h}$ using 2018 data and MC samples for the $\pi^{\pm}$ (top-left), $\pi^{\pm}\pi^0$ (top-right), $\pi^{\pm}2\pi^0$ (middle-left), $3\pi^{\pm}$ (middle-right), and $3\pi^{\pm}\pi^0$ (bottom) MVA decay modes. The efficiency graphs along with their fits and fit uncertainties are shown in the upper panel of each sub-figure while the scale factors are shown in the lower panel. The red dashed line shows the offline threshold below which the events are excluded from the analysis.

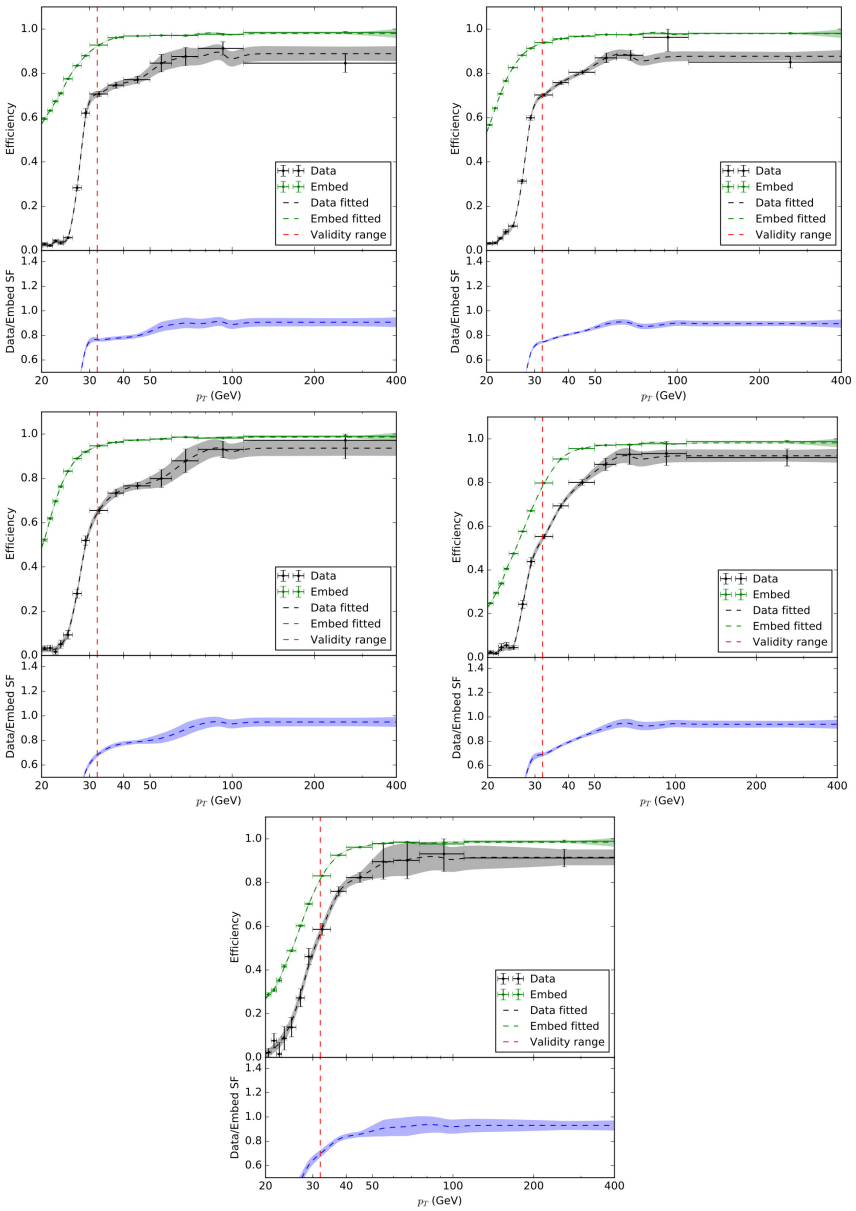

Figure C.29: The efficiencies and scale factors of the τ_h trigger in the $\tau_\mu\tau_h$ channel as a function of $p_{\mathrm{T}}^{\tau_h}$ using 2018 data and embedded samples for the $\pi^\pm$ (top-left), $\pi^\pm\pi^0$ (top-right), $\pi^\pm 2\pi^0$ (middle-left), $3\pi^\pm$ (middle-right), and $3\pi^\pm\pi^0$ (bottom) MVA decay modes. The efficiency graphs along with their fits and fit uncertainties are shown in the upper panel of each sub-figure while the scale factors are shown in the lower panel. The red dashed line shows the offline threshold below which the events are excluded from the analysis.

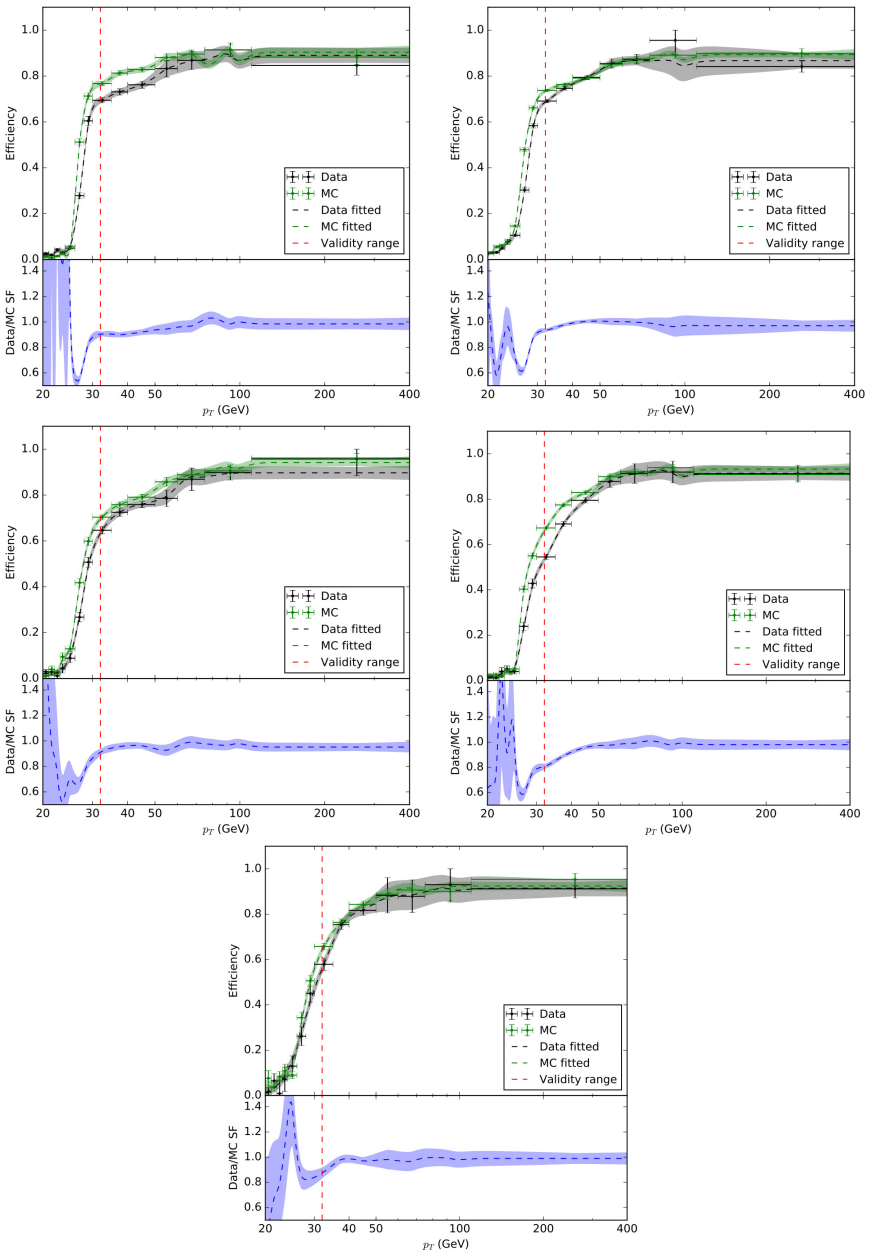

Figure C.30: The efficiencies and scale factors of the τ_h trigger in the $\tau_\mu\tau_h$ channel as a function of $p_{\mathrm{T}}^{\tau_h}$ using 2018 data and MC samples for the $\pi^\pm$ (top-left), $\pi^\pm\pi^0$ (top-right), $\pi^\pm 2\pi^0$ (middle-left), $3\pi^\pm$ (middle-right), and $3\pi^\pm\pi^0$ (bottom) MVA decay modes. The efficiency graphs along with their fits and fit uncertainties are shown in the upper panel of each sub-figure while the scale factors are shown in the lower panel. The red dashed line shows the offline threshold below which the events are excluded from the analysis.

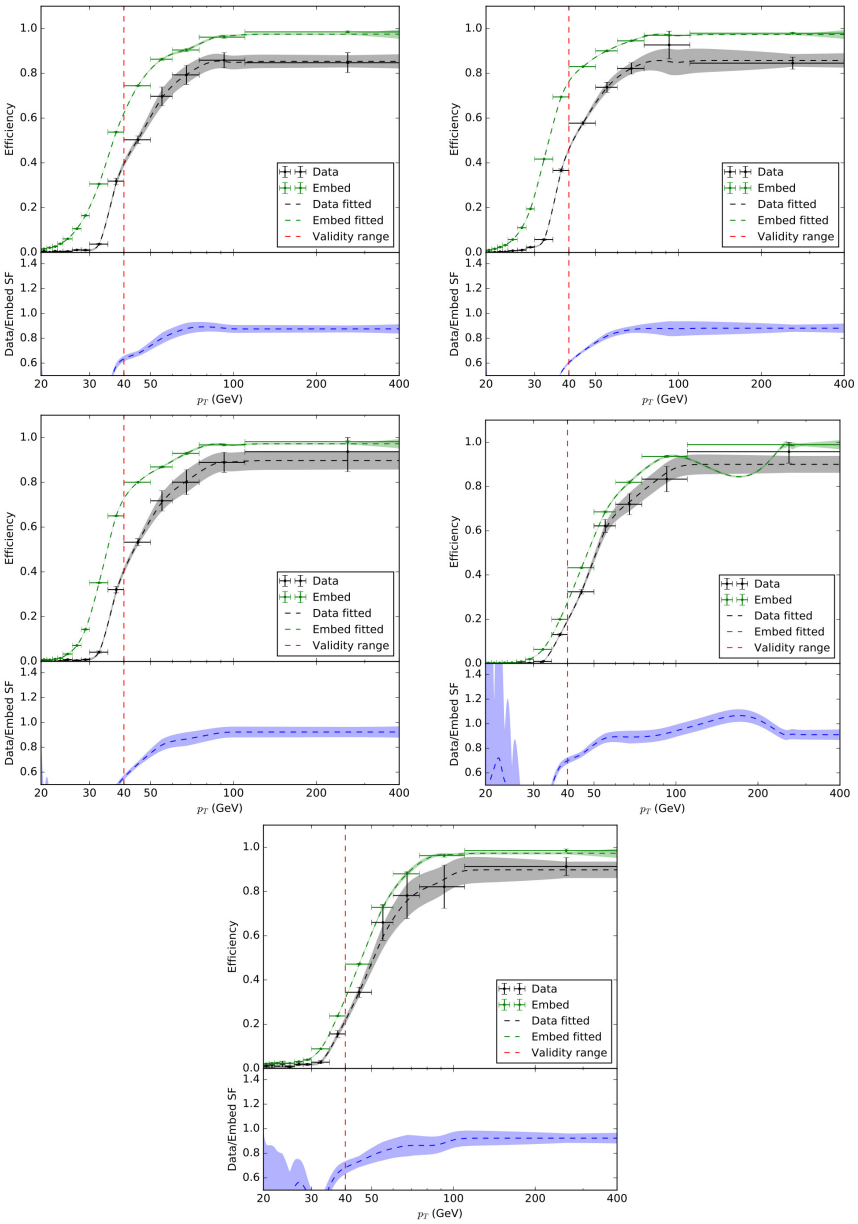

Figure C.31: The efficiencies and scale factors of the τ_h trigger in the $\tau_h\tau_h$ channel as a function of $p_{\mathrm{T}}^{\tau_h}$ using 2018 data and embedded samples for the $\pi^{\pm}$ (top-left), $\pi^{\pm}\pi^0$ (top-right), $\pi^{\pm}2\pi^0$ (middle-left), $3\pi^{\pm}$ (middle-right), and $3\pi^{\pm}\pi^0$ (bottom) MVA decay modes. The efficiency graphs along with their fits and fit uncertainties are shown in the upper panel of each sub-figure while the scale factors are shown in the lower panel. The red dashed line shows the offline threshold below which the events are excluded from the analysis.

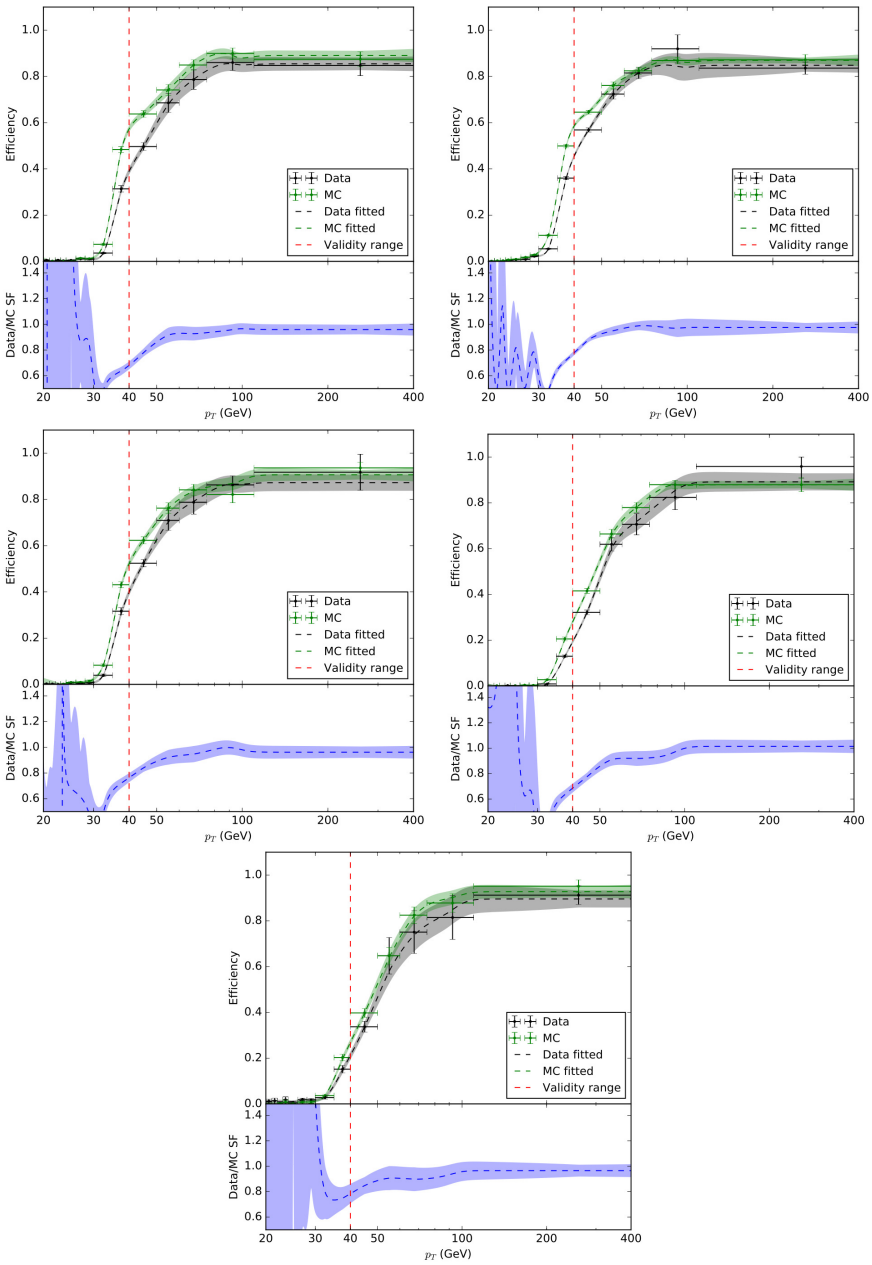

Figure C.32: The efficiencies and scale factors of the τ_h trigger in the $\tau_h\tau_h$ channel as a function of $p_{\mathrm{T}}^{\tau_h}$ using 2018 data and MC samples for the $\pi^{\pm}$ (top-left), $\pi^{\pm}\pi^0$ (top-right), $\pi^{\pm}2\pi^0$ (middle-left), $3\pi^{\pm}$ (middle-right), and $3\pi^{\pm}\pi^0$ (bottom) MVA decay modes. The efficiency graphs along with their fits and fit uncertainties are shown in the upper panel of each sub-figure while the scale factors are shown in the lower panel. The red dashed line shows the offline threshold below which the events are excluded from the analysis.

Bibliography

[1] CMS Collaboration. *Identification of hadronic tau decay channels using multivariate analysis (MVA decay mode)*. Tech. rep. CMS-DP-2020-041. 2020. URL: http://cds.cern.ch/record/2727092.

[2] CMS Collaboration. "Analysis of the $\mathcal{CP}$ structure of the Yukawa coupling between the Higgs boson and τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 06 (2022), p. 012. DOI: 10.1007/JHEP06(2022)012. arXiv: 2110.04836 [hep-ex].

[3] C. P. Burgess and G. D. Moore. *The standard model: A primer*. Cambridge University Press, Dec. 2006. ISBN: 978-0-511-25485-7, 978-1-107-40426-7, 978-0-521-86036-9.

[4] *Standard Model of Elementary Particles*. [Online; accessed June 2022]. URL: https://en.wikipedia.org/wiki/Standard_Model.

[5] S. L. Glashow. "Partial Symmetries of Weak Interactions". In: *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.

[6] S. Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[7] A. Salam. "Weak and Electromagnetic Interactions". In: *Conf. Proc. C* 680519 (1968), pp. 367–377. DOI: 10.1142/9789812795915_0034.

[8] UA1 Collaboration. "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV". In: *Phys. Lett. B* 122 (1983), pp. 103–116. DOI: 10.1016/0370-2693(83)91177-2.

[9] UA2 Collaboration. "Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider". In: *Phys. Lett. B* 122 (1983), pp. 476–485. DOI: 10.1016/0370-2693(83)91605-2.

[10] UA1 Collaboration. "Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/c^2 at the CERN SPS Collider". In: *Phys. Lett. B* 126 (1983), pp. 398–410. DOI: 10.1016/0370-2693(83)90188-0.

[11] UA2 Collaboration. "Evidence for $Z^0 \to e^+e^-$ at the CERN $\bar{p}p$ Collider". In: *Phys. Lett. B* 129 (1983), pp. 130–140. DOI: 10.1016/0370-2693(83)90744-X.

[12] T. W. B. Kibble. "History of electroweak symmetry breaking". In: *J. Phys. Conf. Ser.* 626.1 (2015). Ed. by H. T. Elze et al., p. 012001. DOI: 10.1088/1742-6596/626/1/012001. arXiv: 1502.06276 [physics.hist-ph].

[13] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[14] P. W. Higgs. "Broken symmetries, massless particles and gauge fields". In: *Phys. Lett.* 12 (1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9.

[15] P. W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[16] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 585–587. DOI: 10.1103/PhysRevLett.13.585.

[17] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[18] CMS Collaboration. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[19] CMS Collaboration. "Study of the Mass and Spin-Parity of the Higgs Boson Candidate Via Its Decays to Z Boson Pairs". In: *Phys. Rev. Lett.* 110.8 (2013), p. 081803. DOI: 10.1103/PhysRevLett.110.081803. arXiv: 1212.6639 [hep-ex].

[20] ATLAS Collaboration. "Evidence for the spin-0 nature of the Higgs boson using ATLAS data". In: *Phys. Lett. B* 726 (2013), pp. 120–144. DOI: 10.1016/j.physletb.2013.08.026. arXiv: 1307.1432 [hep-ex].

[21] ATLAS Collaboration. "Measurement of Higgs boson production in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector". In: *Phys. Rev. D* 90.11 (2014), p. 112015. DOI: 10.1103/PhysRevD.90.112015. arXiv: 1408.7084 [hep-ex].

[22] CMS Collaboration. "Observation of the Diphoton Decay of the Higgs Boson and Measurement of Its Properties". In: *Eur. Phys. J. C* 74.10 (2014), p. 3076. DOI: 10.1140/epjc/s10052-014-3076-z. arXiv: 1407.0558 [hep-ex].

[23] ATLAS Collaboration. "Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector". In: *Phys. Rev. D* 91.1 (2015), p. 012006. DOI: 10.1103/PhysRevD.91.012006. arXiv: 1408.5191 [hep-ex].

[24] CMS Collaboration. "Measurement of the Properties of a Higgs Boson in the Four-Lepton Final State". In: *Phys. Rev. D* 89.9 (2014), p. 092007. DOI: 10.1103/PhysRevD.89.092007. arXiv: 1312.5353 [hep-ex].

[25] CMS Collaboration. "Measurements of properties of the Higgs boson decaying to a W boson pair in pp collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Lett. B* 791 (2019), p. 96. DOI: 10.1016/j.physletb.2018.12.073. arXiv: 1806.05246 [hep-ex].

[26] ATLAS Collaboration. "Observation and measurement of Higgs boson decays to WW* with the ATLAS detector". In: *Phys. Rev. D* 92.1 (2015), p. 012006. DOI: 10.1103/PhysRevD.92.012006. arXiv: 1412.2641 [hep-ex].

[27] ATLAS Collaboration. "Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Rev. D* 99 (2019), p. 072001. DOI: 10.1103/PhysRevD.99.072001. arXiv: 1811.08856 [hep-ex].

[28] CMS Collaboration. "Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector". In: *Phys. Lett. B* 779 (2018), pp. 283–316. DOI: 10.1016/j.physletb.2018.02.004. arXiv: 1708.00373 [hep-ex].

[29] ATLAS Collaboration. "Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector". In: *Phys. Lett. B* 786 (2018), pp. 59–86. DOI: 10.1016/j.physletb.2018.09.013. arXiv: 1808.08238 [hep-ex].

[30] CMS Collaboration. "Observation of Higgs boson decay to bottom quarks". In: *Phys. Rev. Lett.* 121.12 (2018), p. 121801. DOI: 10.1103/PhysRevLett.121.121801. arXiv: 1808.08242 [hep-ex].

[31] ATLAS Collaboration. "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector". In: *Phys. Lett. B* 784 (2018), pp. 173–191. DOI: 10.1016/j.physletb.2018.07.035. arXiv: 1806.00425 [hep-ex].

[32] CMS Collaboration. "Observation of $t\bar{t}H$ production". In: *Phys. Rev. Lett.* 120.23 (2018), p. 231801. DOI: 10.1103/PhysRevLett.120.231801. arXiv: 1804.02610 [hep-ex].

[33] CMS Collaboration. "Evidence for Higgs boson decay to a pair of muons". In: *JHEP* 01 (2021), p. 148. DOI: 10.1007/JHEP01(2021)148. arXiv: 2009.04363 [hep-ex].

[34] D. de Florian et al. "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector". In: 2/2017 (Oct. 2016). DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph].

[35] CMS Collaboration. "Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state". In: *Phys. Rev. D* 104.5 (2021), p. 052004. DOI: 10.1103/PhysRevD.104.052004. arXiv: 2104.12152 [hep-ex].

[36] ATLAS Collaboration. "Higgs boson production cross-section measurements and their EFT interpretation in the 4ℓ decay channel at $\sqrt{s}$ =13 TeV with the ATLAS detector". In: *Eur. Phys. J. C* 80.10 (2020). [Erratum: Eur.Phys.J.C 81, 29 (2021), Erratum: Eur.Phys.J.C 81, 398 (2021)], p. 957. DOI: 10.1140/epjc/s10052-020-8227-9. arXiv: 2004.03447 [hep-ex].

[37] ATLAS Collaboration. *Constraints on Higgs boson properties using $WW^*(\to e\nu\mu\nu)jj$ production in 36.1 fb^{-1} of $\sqrt{s}$=13 TeV pp collisions with the ATLAS detector*. Tech. rep. CERN-EP-2021-096. Sept. 2021. arXiv: 2109.13808 [hep-ex].

[38] CMS Collaboration. *Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production with associated particles using the $H \to \tau\tau$ final state*. Tech. rep. CMS-PAS-HIG-20-007. 2021. URL: https://cds.cern.ch/record/2784578/.

[39] CMS Collaboration. *Evidence for off-shell Higgs boson production and first measurement of its width*. Tech. rep. CMS-PAS-HIG-21-013. 2021. URL: https://cds.cern.ch/record/2784590/.

[40] CMS Collaboration. "Constraints on anomalous HVV couplings from the production of Higgs bosons decaying to τ lepton pairs". In: *Phys. Rev. D* 100.11 (2019), p. 112002. DOI: 10.1103/PhysRevD.100.112002. arXiv: 1903.06973 [hep-ex].

[41] ATLAS Collaboration. "Test of $\mathcal{CP}$ invariance in vector-boson fusion production of the Higgs boson in the H$\to \tau\tau$ channel in proton-proton collisions at $\sqrt{s}$=13 TeV with the ATLAS detector". In: *Phys. Lett. B* 805 (2020), p. 135426. DOI: 10.1016/j.physletb.2020.135426. arXiv: 2002.05315 [hep-ex].

[42] D. Fontes et al. "The C2HDM revisited". In: *JHEP* 02 (2018), p. 073. DOI: 10.1007/JHEP02(2018)073. arXiv: 1711.09419 [hep-ph].

[43] ATLAS Collaboration. "$\mathcal{CP}$ Properties of Higgs Boson Interactions with Top Quarks in the $t\bar{t}H$ and tH Processes Using $H \to \gamma\gamma$ with the ATLAS Detector". In: *Phys. Rev. Lett.* 125.6 (2020), p. 061802. DOI: 10.1103/PhysRevLett.125.061802. arXiv: 2004.04545 [hep-ex].

[44] CMS Collaboration. "Measurements of t$\bar{\text{t}}H$ Production and the $\mathcal{CP}$ Structure of the Yukawa Interaction between the Higgs Boson and Top Quark in the Diphoton Decay Channel". In: *Phys. Rev. Lett.* 125.6 (2020), p. 061801. DOI: 10.1103/PhysRevLett.125.061801. arXiv: 2003.10866 [hep-ex].

[45] S. F. King et al. "Exploring the $\mathcal{CP}$-violating NMSSM: EDM Constraints and Phenomenology". In: *Nucl. Phys. B* 901 (2015), pp. 526–555. DOI: 10.1016/j.nuclphysb.2015.11.003. arXiv: 1508.03255 [hep-ph].

[46] M. Tanabashi et al. "Review of Particle Physics". In: *Phys. Rev. D* 98.3 (2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001.

[47] S. Berge, W. Bernreuther, and S. Kirchner. "Determination of the Higgs $\mathcal{CP}$-mixing angle in the tau decay channels". In: *Nucl. Part. Phys. Proc.* 273-275 (2016), pp. 841–845. DOI: 10.1016/j.nuclphysbps.2015.09.129. arXiv: 1410.6362 [hep-ph].

[48] S. Berge et al. "How to pin down the $\mathcal{CP}$ quantum numbers of a Higgs boson in its tau decays at the LHC". In: *Phys. Rev. D* 84 (2011), p. 116003. DOI: 10.1103/PhysRevD.84.116003. arXiv: 1108.0670 [hep-ph].

[49] S. Berge and W. Bernreuther. "Determining the $\mathcal{CP}$ parity of Higgs bosons at the LHC in the tau to 1-prong decay channels". In: *Phys. Lett. B* 671 (2009), pp. 470–476. DOI: 10.1016/j.physletb.2008.12.065. arXiv: 0812.1910 [hep-ph].

[50] S. Berge, W. Bernreuther, and S. Kirchner. "Prospects of constraining the Higgs boson's $\mathcal{CP}$ nature in the tau decay channel at the LHC". In: *Phys. Rev. D* 92 (2015), p. 096012. DOI: 10.1103/PhysRevD.92.096012. arXiv: 1510.03850 [hep-ph].

[51] G. R. Bower et al. "Measuring the Higgs boson's parity using $\tau \to \rho\nu$". In: *Phys. Lett. B* 543 (2002), pp. 227–234. DOI: 10.1016/S0370-2693(02)02445-0. arXiv: hep-ph/0204292.

[52] V. Cherepanov, E. Richter-Was, and Z. Was. "Monte Carlo, fitting and Machine Learning for Tau leptons". In: *SciPost Phys. Proc.* 1 (2019), p. 018. DOI: 10.21468/SciPostPhysProc.1.018. arXiv: 1811.03969 [hep-ph].

[53] V. Cherepanov and A. Zotz. *Kinematic reconstruction of $Z/H \to \tau\tau$ decay in proton-proton collisions*. May 2018. arXiv: 1805.06988 [hep-ph].

[54] S. Jadach, J. H. Kuhn, and Z. Was. "TAUOLA: A Library of Monte Carlo programs to simulate decays of polarized tau leptons". In: *Comput. Phys. Commun.* 64 (1990), pp. 275–299. DOI: 10.1016/0010-4655(91)90038-M.

[55] M. Jezabek et al. "The tau decay library TAUOLA, update with exact O(alpha) QED corrections in $\tau \to \mu(e)\nu\bar{\nu}$ decay modes". In: *Comput. Phys. Commun.* 70 (1992), pp. 69–76. DOI: 10.1016/0010-4655(92)90092-D.

[56] S. Jadach et al. "The tau decay library TAUOLA: Version 2.4". In: *Comput. Phys. Commun.* 76 (1993), pp. 361–380. DOI: 10.1016/0010-4655(93)90061-G.

[57] CMS Collaboration. "The CMS Experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[58] *LHC Design Report Vol.1: The LHC Main Ring*. Tech. rep. CERN-2004-003-V1, CERN-2004-003, CERN-2004-003-V-1. June 2004. DOI: 10.5170/CERN-2004-003-V-1.

[59] *LEP Design Report: Vol.2. The LEP Main Ring*. Tech. rep. CERN-LEP-84-01. June 1984.

[60] ALICE Collaboration. "The ALICE experiment at the CERN LHC". In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002.

[61] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[62] LHCb Collaboration. "The LHCb Detector at the LHC". In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[63] S. D. Breeze. "Precision measurement of the Z invisible width with the CMS experiment". PhD thesis. Imperial Coll., London, 2019. DOI: 10.25560/80981.

[64] W. J. Stirling. URL: https://twiki.cern.ch/twiki/pub/LHCPhysics/HiggsEuropeanStrategy2012/Stirling_crosssections2012HE.pdf.

[65] *LHC commissioning with beam*. [Online; accessed May 2022]. URL: https://lhc-commissioning.web.cern.ch/schedule/LHC-long-term.htm.

[66] CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Tech. rep. CERN-LHCC-2006-001, CMS-TDR-8-1. 2006.

[67] CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *JINST* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[68] CMS Collaboration. *CMS Technical Design Report for the Pixel Detector Upgrade*. Tech. rep. CERN-LHCC-2012-016, CMS-TDR-011. Sept. 2012. DOI: 10.2172/1151650.

[69] CMS Collaboration. "Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in *pp* Collisions at $\sqrt{s} = 7$ TeV". In: *JINST* 8 (2013), P09009. DOI: 10.1088/1748-0221/8/09/P09009. arXiv: 1306.2016 [hep-ex].

[70] USCMS and ECAL/HCAL Collaboration. "The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c". In: *Eur. Phys. J. C* 60 (2009). [Erratum: Eur.Phys.J.C 61, 353–356 (2009)], pp. 359–373. DOI: 10.1140/epjc/s10052-009-0959-5.

[71] CMS Collaboration. "Particle-flow reconstruction and global event description with the CMS detector". In: *JINST* 12.10 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[72] CMS Collaboration. *CMS, the magnet project: Technical design report*. Tech. rep. CERN-LHCC-97-10. May 1997.

[73] M. Shopova et al. "Performance of Resistive Plate Chambers installed during the first long shutdown of the CMS experiment". In: *13th Workshop on Resistive Plate Chambers and Related Detectors*. May 2016. arXiv: 1605.06798 [physics.ins-det].

[74] CMS Collaboration. "The Performance of the CMS Muon Detector in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV at the LHC". In: *JINST* 8 (2013), P11002. DOI: 10.1088/1748-0221/8/11/P11002. arXiv: 1306.6905 [physics.ins-det].

[75] CMS Collaboration. *CMS TriDAS project : Technical Design Report, Volume 1: The Trigger Systems*. Tech. rep. CERN-LHCC-2000-038. Dec. 2000.

[76] *CMS Technical Design Report for the Level-1 Trigger Upgrade*. Tech. rep. CERN-LHCC-2013-011, CMS-TDR-12, CMS-TDR-012. June 2013.

[77] CMS Collaboration. *CMS: The TriDAS project. Technical design report, Vol. 2: Data acquisition and high-level trigger*. Tech. rep. CERN-LHCC-2002-026. Dec. 2002.

[78] P. Das. "An overview of the trigger system at the CMS experiment". In: *Phys. Scripta* 97.5 (2022), p. 054008. DOI: 10.1088/1402-4896/ac6302.

[79] W. Adam et al. *Track reconstruction in the CMS tracker*. Tech. rep. CERN-CMS-NOTE-2006-041, CMS-NOTE-2006-041. Dec. 2005. URL: https://cds.cern.ch/record/934067.

[80] S. Baffioni et al. "Electron reconstruction in CMS". In: *Eur. Phys. J. C* 49 (2007), pp. 1099–1116. DOI: 10.1140/epjc/s10052-006-0175-5.

[81] CMS Collaboration. "Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV". In: CMS-PAS-PFT-10-003, CMS-PAS-PFT-10-003 (2010). URL: https://cds.cern.ch/record/1279347.

[82] W. Adam et al. *Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC*. Tech. rep. CMS-NOTE-2005-001, CERN-CMS-NOTE-2005-001. Jan. 2005. URL: https://cds.cern.ch/record/815410.

[83] K. Rose. "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems". In: *IEEE Proc.* 86.11 (1998), pp. 2210–2239. DOI: 10.1109/5.726788.

[84] R. Fruhwirth, W. Waltenberger, and P. Vanlaer. "Adaptive vertex fitting". In: *J. Phys. G* 34 (2007), N343. DOI: 10.1088/0954-3899/34/12/N01.

[85] M. Cacciari, G. P. Salam, and G. Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[86] M. Cacciari, G. P. Salam, and G. Soyez. "FastJet User Manual". In: *Eur. Phys. J. C* 72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[87] D. Guest et al. "Jet Flavor Classification in High-Energy Physics with Deep Neural Networks". In: *Phys. Rev. D* 94.11 (2016), p. 112002. DOI: 10.1103/PhysRevD.94.112002. arXiv: 1607.08633 [hep-ex].

[88] CMS Collaboration. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *JINST* 13.05 (2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158 [physics.ins-det].

[89] CMS Collaboration. "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector". In: *JINST* 14.07 (2019), P07004. DOI: 10.1088/1748-0221/14/07/P07004. arXiv: 1903.06078 [hep-ex].

[90] D. Bertolini et al. "Pileup Per Particle Identification". In: *JHEP* 10 (2014), p. 059. DOI: 10.1007/JHEP10(2014)059. arXiv: 1407.6013 [hep-ph].

[91] CMS Collaboration. "Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV". In: *JINST* 13.10 (2018), P10005. DOI: 10.1088/1748-0221/13/10/P10005. arXiv: 1809.02816 [hep-ex].

[92] I. Neutelings. *Graphics with TikZ in LaTeX*. https://tikz.net.

[93] A. M. Sirunyan et al. *Identification of hadronic tau lepton decays using a deep neural network*. Submitted to *JINST*. 2022. arXiv: 2201.08458 [hep-ex].

[94] S. Berge, W. Bernreuther, and H. Spiesberger. "Higgs $\mathcal{CP}$ properties using the τ decay modes at the ILC". In: *Phys. Lett. B* 727 (2013), pp. 488–495. DOI: 10.1016/j.physletb.2013.11.006. arXiv: 1308.2674 [hep-ph].

[95] T. Chen and C. Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: Association for Computing Machinery, 2016, pp. 785–794. ISBN: 9781450342322. DOI: 10.1145/2939672.2939785.

[96] T. Przedzinski, E. Richter-Was, and Z. Was. "Documentation of *TauSpinner* algorithms: program for simulating spin effects in τ -lepton production at LHC". In: *Eur. Phys. J. C* 79.2 (2019), p. 91. DOI: 10.1140/epjc/s10052-018-6527-0. arXiv: 1802.05459 [hep-ph].

[97] A. E. Bondar et al. "Novosibirsk hadronic currents for $\tau \to 4\pi$ channels of τ decay library TAUOLA". In: *Comput. Phys. Commun.* 146 (2002), pp. 139–153. DOI: 10.1016/S0010-4655(02)00262-X. arXiv: hep-ph/0201149.

[98] K. W. Edwards et al. "Resonant structure of $\tau \to 3\pi\pi^0\nu_\tau$ and $\tau \to \omega\pi\nu_\tau$ decays". In: *Phys. Rev. D* 61 (2000), p. 072003. DOI: 10.1103/PhysRevD.61.072003. arXiv: hep-ex/9908024.

[99] P. Nason. "A New method for combining NLO QCD with shower Monte Carlo algorithms". In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146.

[100] S. Frixione, P. Nason, and C. Oleari. "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method". In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[101] S. Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[102] E. Bagnaschi et al. "Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM". In: *JHEP* 02 (2012), p. 088. DOI: 10.1007/JHEP02(2012)088. arXiv: 1111.2854 [hep-ph].

[103] P. Nason and C. Oleari. "NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG". In: *JHEP* 02 (2010), p. 037. DOI: 10.1007/JHEP02(2010)037. arXiv: 0911.5299 [hep-ph].

[104] T. Ježo and P. Nason. "On the Treatment of Resonances in Next-to-Leading Order Calculations Matched to a Parton Shower". In: *JHEP* 12 (2015), p. 065. DOI: 10.1007/JHEP12(2015)065. arXiv: 1509.09071 [hep-ph].

[105] F. Granata et al. "NLO QCD+EW predictions for HV and HV +jet production including parton-shower effects". In: *JHEP* 09 (2017), p. 012. DOI: 10.1007/JHEP09(2017)012. arXiv: 1706.03522 [hep-ph].

[106] G. Klamke and D. Zeppenfeld. "Higgs plus two jet production via gluon fusion as a signal at the CERN LHC". In: *JHEP* 04 (2007), p. 052. DOI: 10.1088/1126-6708/2007/04/052. arXiv: hep-ph/0703202.

[107] T. Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[108] CMS Collaboration. *Measurement of Higgs boson production in the decay channel with a pair of τ leptons*. Tech. rep. Geneva: CERN, 2020. URL: https://cds.cern.ch/record/2725590.

[109] L. Bianchini et al. "Reconstruction of the Higgs mass in $H \to \tau\tau$ Events by Dynamical Likelihood techniques". In: *J. Phys. Conf. Ser.* 513 (2014). Ed. by D. L. Groep and D. Bonacorsi, p. 022035. DOI: 10.1088/1742-6596/513/2/022035.

[110] Private communication with Albert Dow.

[111] CMS Collaboration. "Measurement of the $Z\gamma^* \to \tau\tau$ cross section in pp collisions at $\sqrt{s} = 13$ TeV and validation of τ lepton analysis techniques". In: *Eur. Phys. J. C* 78.9 (2018), p. 708. DOI: 10.1140/epjc/s10052-018-6146-9. arXiv: 1801.03535 [hep-ex].

[112] CMS Collaboration. "An embedding technique to determine $\tau\tau$ backgrounds in proton-proton collision data". In: *JINST* 14.06 (2019), P06032. DOI: 10.1088/1748-0221/14/06/P06032. arXiv: 1903.01216 [hep-ex].

[113] CMS Collaboration. "Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV". In: *JHEP* 01 (2011), p. 080. DOI: 10.1007/JHEP01(2011)080. arXiv: 1012.2466 [hep-ex].

[114] J. Alwall et al. "MadGraph 5 : Going Beyond". In: *JHEP* 06 (2011), p. 128. DOI: 10.1007/JHEP06(2011)128. arXiv: 1106.0522 [hep-ph].

[115] N. Adam, A. Hunt, and K. Mishra. URL: https://github.com/cms-sw/cmssw/blob/master/PhysicsTools/TagAndProbe/src/RooCMSShape.cc.

[116] J. E. Gaiser. "Charmonium Spectroscopy From Radiative Decays of the J/ψ and ψ'". M.Sc. thesis. Aug. 1982.

[117] CMS Collaboration. *CMS Luminosity Measurements for the 2016 Data Taking Period*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2257069.

[118] CMS Collaboration. *CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV*. Tech. rep. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2621960.

[119] CMS Collaboration. *CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV*. Tech. rep. Geneva: CERN, 2019. URL: https://cds.cern.ch/record/2676164.

[120] CMS Collaboration. "Cross section measurement of t-channel single top quark production in pp collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Lett. B* 772 (2017), pp. 752–776. DOI: 10.1016/j.physletb.2017.07.047. arXiv: 1610.00678 [hep-ex].

[121] CMS Collaboration. "Measurement of the WZ production cross section in pp collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Lett. B* 766 (2017), pp. 268–290. DOI: 10.1016/j.physletb.2017.01.011. arXiv: 1607.06943 [hep-ex].

[122] R. J. Barlow and C. Beeston. "Fitting using finite Monte Carlo samples". In: *Comput. Phys. Commun.* 77 (1993), pp. 219–228. DOI: 10.1016/0010-4655(93)90005-W.

[123] J. S. Conway. "Incorporating Nuisance Parameters in Likelihoods for Multi-source Spectra". In: *PHYSTAT 2011*. 2011, pp. 115–120. DOI: 10.5170/CERN-2011-006.115. arXiv: 1103.0354 [physics.data-an].

[124] F. Pedregosa et al. "Scikit-learn: Machine learning in Python". In: *the Journal of machine Learning research* 12 (2011), pp. 2825–2830.

[125] CMS Collaboration. *Analysis of the CP structure of the Yukawa coupling between the Higgs boson and τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV*. Tech. rep. Geneva: CERN, 2020. URL: https://cds.cern.ch/record/2725571.

[126] G. Apollinari et al. "High Luminosity Large Hadron Collider HL-LHC". In: *CERN Yellow Rep.* 5 (2015). Ed. by G. Apollinari et al., pp. 1–19. DOI: 10.5170/CERN-2015-005.1. arXiv: 1705.08830 [physics.acc-ph].

[127] CMS Collaboration. *The Phase-2 Upgrade of the CMS Endcap Calorimeter*. Tech. rep. CERN-LHCC-2017-023, CMS-TDR-019. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2293646.

[128] CMS Collaboration. *CMS Luminosity - Public Results*. [Online; accessed May 2022]. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResult.

[129] *The HL-LHC project*. [Online; accessed May 2022]. URL: https://hilumilhc.web.cern.ch/content/hl-lhc-project.

[130] CMS Collaboration. *Technical Proposal for the Phase-II Upgrade of the CMS Detector*. Tech. rep. CERN-LHCC-2015-010, LHCC-P-008, CMS-TDR-15-02. Geneva, June 2015.

[131] CMS Collaboration. *Technical proposal for a MIP timing detector in the CMS experiment Phase 2 upgrade*. Tech. rep. CERN-LHCC-2017-027. LHCC-P-009. Geneva: CERN, Dec. 2017. DOI: 10.17181/CERN.2RSJ.UE8W. URL: https://cds.cern.ch/record/2296612.

[132] M. Bonanomi. "Beam-tests of CMS High Granularity Calorimeter prototypes at CERN". In: *JINST* 15.04 (2020). Ed. by M. Nessi, p. C04001. DOI: 10.1088/1748-0221/15/04/C04001.

[133] A. Zabi et al. *The Phase-2 Upgrade of the CMS Level-1 Trigger*. Tech. rep. CERN-LHCC-2020-004, CMS-TDR-021. 2020.

[134] CERN. *LpGBT specification document*. URL: https://espace.cern.ch/GBT-Project/LpGBT/Specifications/LpGbtxSpecifications.pdf..