



---

# Searches for New Physics With Tau Leptons at the CMS Experiment

---

George Uttley

Imperial College London  
Department of Physics

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

## Abstract

The Standard Model of Particle Physics is currently the best model of the fundamental particles and their interactions. However, there are still significant theoretical issues and recently seen experimental tensions with the model. The theoretical issues include the hierarchy problem which forecasts the breakdown of the Standard Model when looking at the size of corrections needed to calculate the mass of the newest found member of the theory, the Higgs boson particle. The current experimental tensions include the B-anomalies and the g-2 measurement. These results, although they do not yet sit at the required  $5\sigma$  deviation for a discovery, offer the most prominent leads into where new physics may be hiding. Looking for signatures of theoretical explanations of these anomalies offers excellent search options for new fundamental particles. This thesis describes the search for new physics that can explain both the theoretical problems and experimental tensions. This is done using tau leptons seen during Run-2 of the Large Hadron Collider (LHC) at the Compact Muon Solenoid (CMS) experiment. The Beyond Standard Model theories searched for range from Supersymmetry, leptoquarks, to type-X two Higgs doublet models. Each theory is separately studied and an analysis is tailored to find its most sensitive signature. In the process of optimisation, data-driven background modelling is improved to aid the reliability of the results. The results are currently blinded however the expected limits offer some of the largest constraints that are placed on these prominent Beyond Standard Model Theories.

## Declaration

I did this work I promise

## Acknowledgements

Emmeline

# Contents

<b>1 Theory and Motivation</b>	<b>17</b>
1.1 The Standard Model of Particle Physics . . . . .	18
1.1.1 Fundamental Particles and Interactions . . . . .	18
1.1.2 Higgs Sector . . . . .	20
1.2 Extended Higgs Sector . . . . .	22
1.3 Theoretical Problems and Potential Solutions . . . . .	24
1.3.1 Hierarchy Problem . . . . .	24
1.4 Experimental Tensions and Potential Solutions . . . . .	26
1.4.1 B Anomalies . . . . .	26
1.4.2 g-2 Anomaly . . . . .	26
<b>2 The LHC and CMS Experiment</b>	<b>29</b>
2.1 The LHC . . . . .	29
2.2 The CMS Detector . . . . .	32
2.2.1 Tracker . . . . .	32
2.2.2 Electromagnetic calorimeter . . . . .	34
2.2.3 Hadron calorimeter . . . . .	35
2.2.4 Muon system . . . . .	35
2.2.5 Triggering . . . . .	35
<b>3 Object Reconstruction</b>	<b>37</b>
3.1 Tracks and vertices . . . . .	37
3.2 Particle flow . . . . .	37
3.3 Muons . . . . .	38
3.4 Electrons . . . . .	40
3.5 Jets . . . . .	41
3.6 b jets . . . . .	42
3.7 Missing transverse energy . . . . .	42

---

3.8 Taus . . . . .	42
<b>4 Searches for New Physics in <math>\tau^+\tau^-</math> Final States</b>	<b>46</b>
4.1 Signal Modelling . . . . .	47
4.1.1 Additional Higgs Bosons . . . . .	47
4.1.2 Vector Leptoquarks . . . . .	50
4.2 Event Selection . . . . .	52
4.2.1 Trigger Requirements . . . . .	53
4.2.2 Offline Requirements . . . . .	54
4.3 Search Optimisation . . . . .	55
4.4 Background Modelling Overview . . . . .	59
4.5 QCD Estimation in the $e\mu$ Channel . . . . .	60
4.6 Embedding Method . . . . .	61
4.7 Fake Factor Method . . . . .	63
4.7.1 Determination Regions . . . . .	64
4.7.2 Parametrisation . . . . .	66
4.7.3 Corrections . . . . .	70
4.7.4 Applying Fake Factors . . . . .	72
4.8 MC Corrections . . . . .	74
4.9 Uncertainty Model . . . . .	75
4.10 Signal Extraction . . . . .	79
4.11 Postfit Plots . . . . .	82
4.12 Model Independent Results . . . . .	83
4.12.1 Limits . . . . .	83
4.12.2 Significance and Compatibility . . . . .	88
4.12.3 2D Likelihood Scans . . . . .	90
4.13 Model Dependent Limits . . . . .	90
<b>5 Search for New Physics in <math>\tau^+\tau^-\tau^+\tau^-</math> Final States</b>	<b>95</b>
5.1 Signal Modelling . . . . .	96
5.2 Event Selection . . . . .	98
5.2.1 Trigger Requirements . . . . .	100
5.2.2 Offline Requirements . . . . .	101
5.3 Search Optimisation . . . . .	102
5.4 Background Modelling Overview . . . . .	103
5.5 ZZ Modelling . . . . .	104

5.6	Machine Learning Fake Factor Method . . . . .	104
5.6.1	BDT Reweighting . . . . .	106
5.6.2	Fitting Regions . . . . .	106
5.6.3	Variables Used . . . . .	108
5.6.4	Machine Learning Subtraction Method . . . . .	108
5.6.5	Fitting . . . . .	110
5.6.6	Applying Fake Factors . . . . .	114
5.7	Uncertainty Model . . . . .	115
5.8	Signal Extraction . . . . .	117
5.8.1	Postfit Plots . . . . .	117
5.9	Model Independent Results . . . . .	117
5.10	Model Dependent Limits . . . . .	117
<b>6</b>	<b>Conclusion</b>	<b>125</b>
6.1	Global Interpretations of Results . . . . .	125
6.2	Outlook . . . . .	125

# List of Figures

1.1	Diagram of the fundamental particles the constitute the SM. Also displayed are the fermion generation shown in Roman numerals, the particles measured mass, charge, spin and colours available for the strong interaction. This is taken and adjusted from Ref. [1] . . . . .	19
1.2	One-loop corrections to the Higgs mass by a fermion $f$ (a) and a scalar $S$ (b). . . . .	24
1.3	Examples of one-loop Bar-Zee feynman diagrams for the contribution from $\phi$ and $A$ in (a) and $H^\pm$ in (b). . . . .	27
1.4	Examples of two-loop Bar-Zee feynman diagrams for the contribution from $\phi$ and $A$ in (a) and $H^\pm$ in (b). Note that other fundamental particles can also contribute to the loop. . . . .	27
2.1	A schematic diagram of the CERN [1]. . . . .	30
2.2	The total integrated luminosity for proton-proton collisions collected by the CMS experiment between 2010 and May 2023 [2]. . . . .	31
2.3	A perspective view of the CMS detector [2]. . . . .	33
2.4	Schematic of the CMS tracker (pre-pixel upgrade) in the $r$ - $z$ plane, showing the position of the pixel detector as well as the tracker inner barrel (TIB), tracker inner disks (TID), tracker outer barrel (TOB), and tracker endcaps (TEC) strip detectors. The lines represent detector modules and the double lines represent back-to-back modules [2].	33
2.5	A schematic of the CMS ECAL detector [2]. . . . .	36
2.6	. . . . .	36
3.1	. . . . .	45

4.1	Diagram (a) shows the production of neutral Higgs bosons from gluon fusion. The dominant loop contributions to this diagrams are from top-only, bottom-only and top-bottom interference. Diagrams (b) and (c) show production in association with b quarks. . . . .	47
4.2	$p_T$ density distributions of the A (top) and H (bottom) boson, with contributions to the gluon fusion loop displayed individually and summed. These are shown for $\tan\beta$ values of 10 (left) and 30 (right) where $m_A = 500$ GeV in the MSSM $M_h^{125}$ scenario. . . . .	49
4.3	Feynman diagrams showing the contribution from $U_1$ vector leptoquarks to the final state with a pair of oppositely charged $\tau$ leptons. Diagram (a) shows t-channel, (b) pair and (c) single production of a vector leptoquark. . . . .	50
4.4	The generator level $m_{\tau\tau}$ density distributions of the t-channel vector leptoquark signal and the interference with Drell-Yan. This is shown in the VLQ BM 1 scenario for a leptoquark of mass 1 TeV for coupling strengths of $g_U = 3$ (a) and $g_U = 5$ (b). . . . .	52
4.5	Diagram of inputs to the $D_\zeta$ variable. . . . .	55
4.6	Overview of the categories used for the extraction of the signal in the high mass optimisation procedure. . . . .	58
4.7	Overview of the categories used for the extraction of the signal in the low mass optimisation procedure. . . . .	58
4.8	Schematic of the embedding method to model genuine di- $\tau$ backgrounds from di-muon events in data. . . . .	62
4.9	Closure plot showing the di-muon mass on the dedicated embedding validation samples. . . . .	63
4.10	Schematic of the regions used for fake factor derivation. . . . .	65
4.11	A 2D heat map of the fake factors determined from $t\bar{t}$ MC for the full run-2 dataset in the combined $e\tau_h$ and $\mu\tau_h$ channels. This is shown with respect to the $\tau_h p_T$ and the $p_T$ of the jet matched to the $\tau_h$ . The ratio of jet to $\tau_h p_T$ categorisation used is shown split by the dashed lines. . . . .	67
4.12	Flow chart of the algorithm used to determine where binned values are taken instead of the fit. The blue box represents the input, the green diamonds represent the decisions and the red boxes represent the outputs. . . . .	68

4.13	Fake factor fits in $\tau_h\tau_h$ channel for the QCD $N_{\text{pre b jets}} = 0$ category with 2018 data. The three jet $p_T$ to $\tau_h$ $p_T$ categories are shown. . . . .	69
4.14	Fake factor fits in $\mu\tau_h$ channel for the QCD and W + Jets $N_{\text{pre b jets}} = 0$ category with 2018 data. The three jet $p_T$ to $\tau_h$ $p_T$ categories are shown for each process. . . . .	69
4.15	Diagram showing how fake $E_T^{\text{miss}}$ arises from mismodelling jet constituents in $\tau_h$ identification. . . . .	70
4.16	Determination region closure correction fits with 2018 data. (a) is the correction parametrised by $\Delta R$ in events with $N_{\text{b jets}} = 0$ in the $\tau_h\tau_h$ channel. (b) and (c) show the correction for the $\mu\tau_h$ channel parametrised by the specific correction variables defined in Equation 4.6 and 4.7 for QCD and W + jets processes respectively. . . . .	71
4.17	Determination region to application region closure correction fits with 2018 data. (a) is the correction moving from same sign to opposite sign $\tau$ leptons the parameterised by leading $\tau_h$ $p_T$ in events with $N_{\text{b-jets}} = 0$ in the $\tau_h\tau_h$ channel. (b) and (c) show the correction for the $\mu\tau_h$ channel moving from same sign to opposite sign $\tau$ leptons and high $m_T$ to low $m_T$ both parameterised by the the muon $p_T$ for QCD and W + jets processes respectively. . . . .	72
4.18	The expected application region fractions of the processes in the $\mu\tau_h$ channel. (a) and (b) show the no b tag Tight- $m_T$ and Loose- $m_T$ categories and (c) and (d) show the b tag Tight- $m_T$ and Loose- $m_T$ categories respectively. . . . .	73
4.19	Distributions of $m_{\tau\tau}$ in the no b tag second highest (left) and highest (right) $p_T$ category for the $\tau_h\tau_h$ (top), the combined $e\tau_h$ and $\mu\tau_h$ (middle) and the $e\mu$ (bottom) channels. The solid histograms show the stacked background predictions after a signal plus background fit to the data. The best fit gluon fusion signal for $m_\phi = 100$ GeV is shown by the red line. Also shown by a blue dashed line on the bottom pad is the ratio of the background predictions for the background only fit to the signal plus background fit . . . . .	84

4.20 Distributions of $m_T^{\text{tot}}$ in the $\tau_h\tau_h$ no b tag (a) and b tag (b) categories, the combined $e\tau_h$ and $\mu\tau_h$ no b tag (c) and b tag (d) Tight- $m_T$ categories and the $e\mu$ no b tag (e) and b tag (f) Medium- $D_\zeta$ categories. The solid histograms show the stacked background predictions after a background only fit to the data. The best fit gluon fusion signal for $m_\phi = 1.2$ TeV is shown by the red line, b associated production and $U_1$ signals are also shown for illustrative purposes. . . . .	85
4.21 Expected (dashed line) and observed (solid line and dots) 95% CL upper limits on the product of the cross sections and branching fraction for the decay into $\tau$ leptons for (a) $gg\phi$ and (b) $bb\phi$ production in a mass range of $60 \leq m_\phi \leq 3500$ GeV. The dark green and bright yellow bands indicate the central 68% and 95% intervals for the expected exclusion limit. . . . .	86
4.22 Comparison of the expected 95% CL upper limits on the product of the cross sections and branching fraction for the decay into $\tau$ leptons for (a) $gg\phi$ and (b) $bb\phi$ production, split by the $\tau\tau$ decay products fit individually. . . . .	87
4.23 Comparison of the expected 95% CL upper limits on the product of the cross sections and branching fraction for the decay into $\tau$ leptons for (a) $gg\phi$ and (b) $bb\phi$ production, split by the CMS result detailed in this thesis and the ATLAS result from Ref. [3]. . . . .	87
4.24 Local $p$ -value and significance of a $gg\phi$ (left) and $bb\phi$ (right) signal as a function of $m_\phi$ . . . . .	88
4.25 Compatibility plots of the 100 GeV excess split into analysis channels (a) and categories (b). In each case the fitted signal strength is decoupled in the bin shown on the plot. . . . .	89
4.26 Compatibility plots of the 1.2 TeV excess split by analysis channels. In each case the fitted signal strength is decoupled in each channel. . . . .	89
4.27 Maximum likelihood scans, including 68% and 95% CL contours obtained from the signal likelihood for the model-independent search. The scans are shown for selected values of $m_\phi$ between 60 GeV and 3.5 TeV. . . . .	91

4.28	Expected and observed 95% CL exclusion contours in the MSSM $M_h^{125}$ (a) and $M_{h,EFT}^{125}$ (b) scenarios. The exclusion limit only on background expectation is shown as a dashed black line, the dark and bright grey bands show the 68% and 95% intervals of the expected exclusion and the observed exclusion contour is shown by the blue area. The parameter space where $m_h$ deviates by more than $\pm 3$ GeV from the observed SM Higgs boson mass is shown by a red hatched area. . . . .	93
4.29	Expected and observed 95% CL upper limits on $g_U$ in the VLQ BM 1 (a) and 2 (b) scenarios, in a mass range of $1 < m_U < 5$ TeV. The exclusion limit only on background expectation is shown as a dashed black line, the dark and bright grey bands show the 68% and 95% intervals of the expected exclusion and the observed exclusion contour is shown by the blue area. The 95% confidence interval for the preferred region from the global fit presented in Ref. [4] is also shown by the green shaded area. . . . .	94
5.1	Diagram of production of two additional neutral Higgs bosons from an off-shell Z boson and their decays to tau leptons. . . . .	96
5.2	Generator level distributions of the visible mass densities for A (a) and $\phi$ (b) for the signal process. . . . .	97
5.3	Calculated production cross sections for the $Z^* \rightarrow \phi A$ process, varying the masses of $\phi$ and A. . . . .	98
5.4	Calculated branching fractions of A (a) and $\phi$ (b) decaying to a pair of $\tau$ leptons for various mass scenarios. . . . .	99
5.5	Calculated branching fractions in the $\cos(\beta - \alpha)$ - $\tan \beta$ phase space for $\phi$ of mass 200 GeV decaying to a pair of $\tau$ leptons, in the scenario where $m_A = 160$ GeV. . . . .	99
5.6	Overview of the categories used for the extraction of the signal in the $Z^* \rightarrow \phi A \rightarrow 4\tau$ . . . . .	102
5.7	Distributions in the $\mu\mu\mu\mu$ channel are shown for the variables total transverse mass, $m_T^{\text{tot}}$ , (a) and the total mass, $m_{\mu\mu\mu\mu}$ (b). The solid histograms show the stacked background predictions. . . . .	105

---

5.8 Comparison of histograms produced via the BDT subtraction method to histogram subtraction. Also shown is the histogram produced when no subtraction is performed. The uncertainty bands contain statistical uncertainties and uncertainties derived on the non-closure of the method. The is shown for four of the fitted variable: $\tau_h\text{-}p_T$ , the ratio of $\tau_h\text{-}p_T$ to jet- $p_T$ , the $\tau_h$ HPS decay mode and the $\tau_h\text{-}\eta$ . . . . .	111
5.9 Average fake factors (reweights) calculated by the BDT reweighting method shown individually in regions A, B and D NEED TO CHANGE THE REGIONS as defined in Figure 4.10. The is shown for four of the fitted variable: $\tau_h\text{-}p_T$ , the ratio of $\tau_h\text{-}p_T$ to jet- $p_T$ , the $\tau_h$ HPS decay mode and the $\tau_h\text{-}\eta$ . . . . .	112
5.10 Comparison of histograms produced using the fake factors applied to the fitted fail region compared to fitted pass region. The uncertainty bands contain statistical uncertainties and uncertainties derived on the non-closure of the method. The is shown for three of the fitted variable: $\tau_h\text{-}p_T$ , the ratio of $\tau_h\text{-}p_T$ to jet- $p_T$ , the $\tau_h$ HPS decay mode and the $\tau_h\text{-}\eta$ . The $\chi^2$ divided by number of degrees of freedom between the two histograms is also shown. . . . .	113
5.11 Distributions of $m_T^{\text{tot}}$ in the $ee\tau_h\tau_h$ OS Leptons (a) and $\mu\mu\tau_h\tau_h$ OS Leptons (b), $e\mu\tau_h\tau_h$ OS Leptons (c), $e\mu\tau_h\tau_h$ SS Leptons (d) and $\tau_h\tau_h\tau_h$ channel and categories. The solid histograms show the stacked background predictions after a background only fit to the data. The $m_\phi = 100$ GeV and $m_A = 100$ GeV signal is also shown for illustrative purposes. . . . .	118
5.12 Distributions of events with respect to the $\log_{10}(S/(S + B))$ of all analysis category bins fit for two signal scenarios: $m_\phi = 100$ GeV and $m_A = 100$ GeV and $m_\phi = 160$ GeV and $m_A = 160$ GeV. The solid histograms show the stacked background predictions after a background only fit to the data. The signal model considered is displayed with a solid line. . . . .	120

5.13 Expected (dashed line) and observed (solid line and dots) 95% CL upper limits on the product of the cross sections and branching fractions for the decay of both additional Higgs bosons into $\tau$ leptons. Different $m_A$ hypotheses are scaled by different orders of magnitudes (written on plot) in order to make mass points distinguishable. The dark green and bright yellow bands indicate the central 68% and 95% intervals for the expected exclusion limit. . . . .	121
5.14 Comparison of the expected 95% CL upper limits on the product of the cross sections and branching fractions for the decay into $\tau$ leptons, split by the $\tau\tau\tau\tau$ decay products fit individually. . . . .	122
5.15 Compatibility plots for the $m_A = 100$ GeV and $m_\phi = 100$ GeV mass scenario in analysis decay channels. In each case the fitted signal strength is decoupled in the bin shown on the plot. . . . .	122
5.16 Expected and observed 95% CL exclusion contours on the $m_A$ - $\tan\beta$ phase space in the type X 2HDM alignment scenario for $m_\phi$ scenarios of 100 GeV (a) and 200 GeV (b). The exclusion limit only on background expectation is shown as a dashed black line, the dark and bright grey bands show the 68% and 95% intervals of the expected exclusion and the observed exclusion contour is shown by the blue area. . . . .	123
5.17 Expected and observed 95% CL exclusion contours on the $\cos(\beta - \alpha)$ - $\tan\beta$ phase space in the type X 2HDM alignment scenario with $m_\phi$ equal to 200 GeV and $m_A$ scenarios of 100 GeV (a) and 160 GeV (b). The exclusion limit only on background expectation is shown as a dashed black line, the dark and bright grey bands show the 68% and 95% intervals of the expected exclusion and the observed exclusion contour is shown by the blue area. . . . .	124
6.1 . . . . .	126

# List of Tables

1.1	Table showing which Higgs doublet different fermion groups different types of 2HDMs couple to. By convention the $u$ quark is chosen to couple to $\Phi_2$ .	23
1.2	Table showing the couplings of fermion groups to additional neutral Higgs bosons in different types of 2HDMs. These are dependent on the mixing angles $\alpha$ and $\beta$ . $t_x$ , $s_x$ and $c_x$ represent $\tan x$ , $\sin x$ and $\cos x$ respectively.	23
1.3	Regions of interest for g-2 anomaly with respect to the type X 2HDM in the normal and inverted alignment scenarios as suggested in Ref. [5].	28
3.1	Electron effective areas used for the $\rho$ -corrected isolation computation.	41
3.2	Measured branching fractions, that are greater than 2%, for the tau lepton. $h$ represents a charged hadron either a pion or a kaon.	43
4.1	Branching fractions of the decays of two $\tau$ leptons.	53
4.2	Lower trigger light lepton thresholds $p_T$ in GeV for the $e\tau_h$ and $\mu\tau_h$ channels.	53
4.3	pdfs used for nuisance parameters.	79
5.1	Branching fractions of four $\tau$ leptons, where $e$ and $\mu$ represent the leptonic decay of the $\tau$ and $\tau_h$ represent the hadronic decay of the $\tau$ .	100
5.2	Vetoos set on the number of leptons for each channel.	101
5.3	Regions modelled by the fake factor method when using the leading $\tau_h$ ( $\tau_h^1$ ), the sub-leading $\tau_h$ ( $\tau_h^2$ ) and both, to model the jet $\rightarrow \tau_h$ background and whether that specific object is a genuine $\tau$ lepton ( $\tau$ ) or a jet ( $j$ ). Also shown is a combination of three regions to fully model all possible combinations.	114

5.4 Table of background, data and signal integrals, for the $m_\phi = 100$ GeV and $m_A = 100$ GeV scenario, in all analysis categories after a background only fit to the data. The background uncertainties are separated into their statistical and systematic and shown along side the background integral. The number of bins used in the fit are also shown. . . . .	119
--	-----

# Chapter 1

## Theory and Motivation

The Standard Model (SM) of particle physics is our current best theory to describe the fundamental particles and their interactions. The SM describes the strong force, as well as unifying the weak and electromagnetic forces. The latter is partly done through the Brout-Englert-Higgs (BEH) mechanism for spontaneous symmetry breaking, that allows many fundamental particles to obtain masses. It also predicts a new scalar boson, named the Higgs boson. In 2012, the ATLAS [6] and CMS collaboration [7] discovered a Higgs boson-like particle when colliding protons at high energy at the Large Hadron Collider (LHC) and further measurements of this particle's properties have been consistent with such a particle. This discovery experimentally completed the SM particle constituents.

However, the SM is not without theoretical problems and experimental tensions. Firstly, the hierarchy problem describes the issue of lightness of the observed Higgs boson mass and the "unnatural" balancing of inputs needed to explain the theorised loop corrections to the predicted mass. These are orders of magnitude larger than the observed mass. A solution to this problem is Supersymmetry, but no experimental evidence for this theory has yet been found that separates it from the SM. Secondly, results from the LHCb experiment [8, 9, 10, 11, 12, 13, 14, 15] and the g-2 experiment [16] have shown deviations from the SM predictions. Although not the statistical significance for a discovery, they offer intriguing hints at potential Beyond SM (BSM) physics. BSM particles, produced from extended Higgs sectors or otherwise, have been theorised to explain these deviations . This chapter will explain the SM and the Higgs sector theory, as well as detailing the BSM extensions that can help resolve the theoretical tensions and experimental tensions.

## 1.1 The Standard Model of Particle Physics

### 1.1.1 Fundamental Particles and Interactions

The SM is a set of fundamental particles, as shown in Figure 1.1, and rules that govern the interactions between particles. The interactions between these particles are able to model the strong, weak and electromagnetic force, unifying the later two into one electroweak interaction. The SM consists of 6 quarks, 3 charged leptons and 3 neutrinos, which are grouped in fermions because of their shared half integer spin. Each of these particles contain an anti-partner with opposite quantum numbers but the same mass. The SM also consists of a number of bosons (integer spin) that describe the fundamental forces of nature: the strong, weak, and electromagnetic forces. The gluon is the mediator of the strong force, the W and Z bosons mediate the weak force, and the photon mediates the electromagnetic force.

The SM is a renormalisable quantum field theory that is built on the principle of local gauge invariance. The  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$  is the gauge symmetry group of the SM. This means that the Lagrangian, that governs the particles interactions, is invariant under such a transformation.  $SU(2)_L \otimes U(1)_Y$  is the symmetry of the electroweak unification and  $SU(3)_C$  is the symmetry of the theory for strong force, name Quantum Chromodynamic (QCD).

SOME STUFF ABOUT QCD HERE

ALSO MAYBE SOME STUFF ABOUT WHY  $SU(2)_L$  AND NOT JUST  $SU(2)$

Electroweak unification was initially proposed by Glashow [], Weinberg [] and Salam [] to combine the theories for the weak and electromagnetic forces into one. It is built on the premise of the Dirac equation. The Dirac Lagrangian for a massless field,  $\psi$ , is shown below.

$$\mathcal{L}_{\text{Dirac}} = i\bar{\psi}\gamma^\mu\partial_\mu\psi \quad (1.1)$$

The  $SU(2)_L$  transformation operates on the weak isospin,  $I$ , only for the left handed spinors and the  $U(1)_Y$  transformation operates on the weak hypercharge,  $Y = 2(Q - I_3)$ , where  $Q$  is the charge of the fundamental particles and  $I_3$  is the third component of the weak isospin. By invoking gauge invariance of the Dirac Lagrangian under such a transformation, four associated gauge fields are present,  $\mathbf{W}_\mu = (W_\mu^1, W_\mu^2, W_\mu^3)$

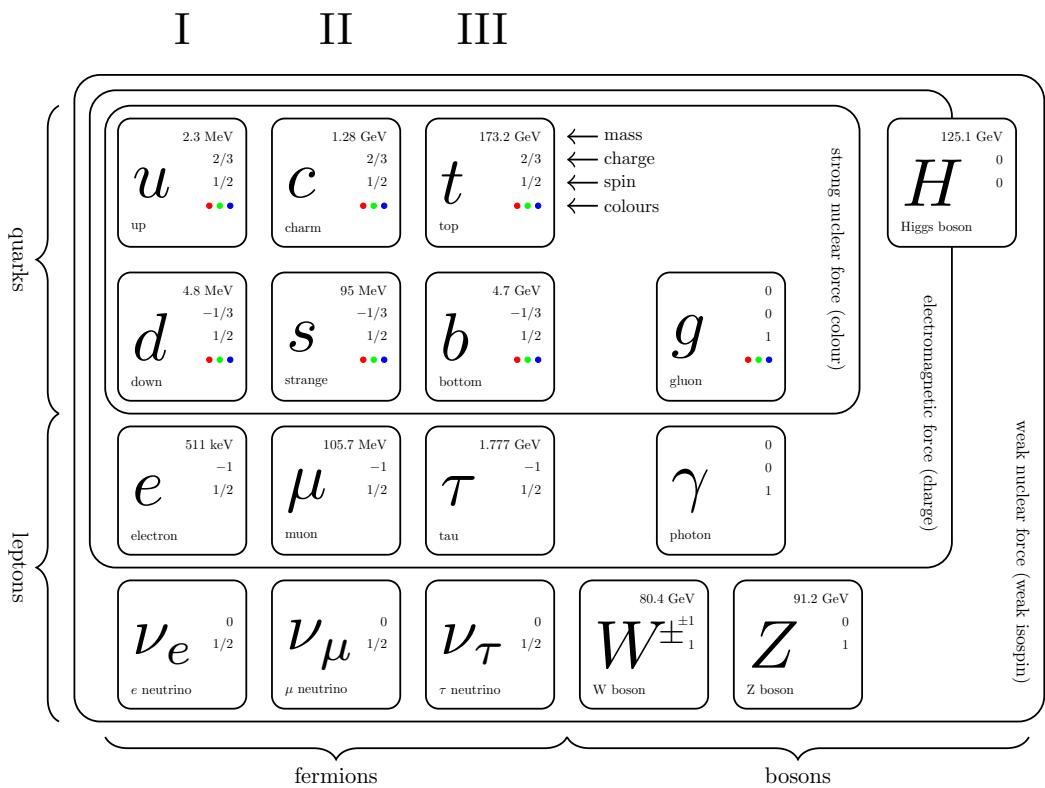


Figure 1.1: Diagram of the fundamental particles the constitute the SM. Also displayed are the fermion generation shown in Roman numerals, the particles measured mass, charge, spin and colours available for the strong interaction. This is taken and adjusted from Ref. [1]

and  $B_\mu$ .

$$\begin{aligned}\mathcal{L}_{\text{Electroweak}} = & \bar{\psi}_L \left( \partial_\mu + \frac{i}{2} g \mathbf{W}_\mu \cdot \boldsymbol{\sigma} + g' Y B_\mu \right) \psi_L \\ & + \bar{\psi}_R \left( \partial_\mu + g' Y B_\mu \right) \psi_R - \frac{1}{4} \mathbf{W}_{\mu\nu} \cdot \mathbf{W}^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu},\end{aligned}\quad (1.2)$$

where the partial derivatives have been replaced with the covariant derivative and the added field tensors  $\hat{\mathbf{W}}_{\mu\nu}$  and  $\hat{B}_{\mu\nu}$  that are defined as,

$$\mathbf{W}_{\mu\nu} = \partial_\mu \mathbf{W}_\nu - \partial_\nu \mathbf{W}_\mu - ig[\mathbf{W}_\mu, \mathbf{W}_\nu] \quad (1.3)$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \quad (1.4)$$

These fields can be rotated in physical fields for the photon, Z and W boson with the following transformations,

$$\begin{aligned}W_\mu^\pm &= \frac{1}{\sqrt{2}} (W_\mu^1 \mp W_\mu^2), \\ Z_\mu &= W_\mu^3 \cos \theta_w - B_\mu \sin \theta_w, \\ A_\mu &= W_\mu^3 \sin \theta_w - B_\mu \cos \theta_w,\end{aligned}\quad (1.5)$$

where  $\theta_w$  represents the weak mixing angle and is defined such that

$$\sin \theta_w = \frac{g'}{\sqrt{g^2 + g'^2}}, \quad \cos \theta_w = \frac{g}{\sqrt{g^2 + g'^2}}. \quad (1.6)$$

The  $W^\pm$  and Z bosons were discovered by the UA1 and UA2 collaborations in 1983, which confirmed the predictions made by electroweak unification. However, the  $W^\pm$  and Z bosons were measured to have a non-zero mass and Lagrangian mass terms of the form  $\frac{1}{2} m_Z^2 Z_\mu Z^\mu$  or  $m_W^2 W_\mu^- W^{+\mu}$  cannot be included as they are not invariant under the  $SU(2)_L \otimes U(1)_Y$  gauge symmetry. The same issue also exist in massive fermions, as the fermion mass terms left- and right-handed chiral states,  $m\bar{\psi}\psi = m(\bar{\psi}_R \psi_L + \bar{\psi}_L \psi_R)$ , transform differently and therefore does not remain invariant under the gauge transformation. To resolve this, the Brout-Englert-Higgs (BEH) mechanism for spontaneous symmetry breaking was theorised.

### 1.1.2 Higgs Sector

The BEH mechanism was proposed in the 1960s by Englert and Brout [17], Higgs [18, 19, 20] and Guralnik, Hagen and Kibble [21, 22]. It works on the principles of

spontaneous symmetry breaking of the  $SU(2)_L \otimes U(1)_Y$  gauge symmetry. It does this by introducing a gauge invariant field that has a non-zero vacuum expectation value (VEV). The formalism of this is shown below, starting from a complex scalar doublet,  $\Phi$ .

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \quad (1.7)$$

where  $\phi^+$  and  $\phi^0$  are complex functions. The Lagrangian and the potential, chosen to fulfil the conditions stated above, then takes the form,

$$\mathcal{L} = (\partial_\mu \Phi)^\dagger (\partial^\mu \Phi) - V(\Phi) \quad (1.8)$$

with,

$$V(\phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2, \quad (1.9)$$

where  $\mu^2$  and  $\lambda$  are two real parameters,  $\lambda$  is required to be positive for the vacuum to be stable. If  $\mu^2$  is negative,  $\Phi$  will have a non-zero VEV and the field will be able to spontaneously breaking the gauge symmetry. In the vacuum state, the field must satisfy the criteria,

$$\Phi^\dagger \Phi = -\frac{\mu^2}{2\lambda}. \quad (1.10)$$

This potential fulfils the criteria of a non-zero VEV as  $\Phi$  must be non-zero. Choosing a gauge, to remove the massless scalar bosons predicted by Goldstone's theorem [23], by taking  $\phi^+$  to be zero and  $\phi$  to be real. The ground state of  $\Phi$  is found as,

$$\langle 0 | \Phi | 0 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu \end{pmatrix}, \quad (1.11)$$

where  $\nu^2 = \mu^2/\lambda$ , where  $\nu$  is the non-zero VEV. The complex doublet can then be written as an expansion around the minimum of the potential,

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + h(x) \end{pmatrix} \quad (1.12)$$

Applying the covariant derivative for the electroweak gauge symmetry to produce a kinematic term for the field,  $\Phi$ , yields

$$(D_\mu \Phi)^\dagger (D^\mu \Phi) = \frac{1}{2}(\partial_\mu h)(\partial^\mu h) + \frac{1}{8}g^2(W_\mu^1 + iW_\mu^2)(W^{1\mu} - iW^{2\mu})(\nu + h)^2 \\ + \frac{1}{8}(gW_\mu^3 - g'B_\mu)(gW^{3\mu} - g'B^\mu)(\nu + h)^2. \quad (1.13)$$

The Lagrangian incorporating this term, gives mass terms for  $W_\mu^1$  and  $W_\mu^2$  and hence the mass of the W boson is  $m_W = \frac{1}{2}g\nu$ . Once again the remaining states need to be rotated to give the photon and the Z boson fields. Setting up the non-diagonal mass matrix for the fields  $W_\mu^3$  and  $B_\mu$  and calculating the eigenvalues to find the diagonal basis, masses of two fields are found, one equal to 0 and the other equal to  $\frac{1}{2}\nu\sqrt{g^2 + g'^2}$  which represent the photon and Z boson mass respectively. The transformation to the photon and Z fields as parametrised in Eq. 1.5 utilises,

$$\tan \theta_W = \frac{g'}{g}. \quad (1.14)$$

A mass term for the Higgs boson also arises from the potential of the BEH mechanism with  $m_h = \sqrt{-2\mu^2}$ . The same mechanism can also be used to add mass terms for quarks and charged leptons. Gauge invariant couplings to the Higgs bosons and fermion mass terms can be generated from the BEH mechanism applied to a Lagrangian term of the form,

$$\bar{\psi}_L \Phi \psi_R + \bar{\psi}_R \Phi \psi_L. \quad (1.15)$$

## 1.2 Extended Higgs Sector

There is no theoretical limitation to only have one Higgs doublet in the theory. Therefore, a natural extension to the SM Higgs sector is the two Higgs doublet model (2HDM). 2HDMs predict 5 Higgs bosons; 1 lighter and 1 heavier scalar ( $h$  and  $H$ ), 1 pseudoscalar ( $A$ ) and 2 charged particles ( $H^\pm$ ). The Lagrangian for such a theory is shown below.

$$\mathcal{L}_{\text{yukawa}}^{\text{2HDM}} = - \sum_{f=u,d,l} \left( \frac{m_f}{\nu} g_h^f \bar{f} f h + \frac{m_f}{\nu} g_H^f \bar{f} f H - i \frac{m_f}{\nu} g_A^f \bar{f} \gamma_5 f A \right) \\ - \left[ \frac{\sqrt{2}V_{ud}}{\nu} \bar{u}(m_u g_A^u P_L + m_d g_A^d P_R) d H^+ + \frac{\sqrt{2}m_l g_A^d}{\nu} \bar{\nu}_L l_R H^+ + h.c. \right], \quad (1.16)$$

where  $u$ ,  $d$  and  $l$  represent up-like quarks, down-like quarks and charged leptons.  $m_f$  are the fermion masses,  $\nu$  is the vacuum expectation value of the SM Higgs doublet,

$g$  are the couplings (relative to the SM couplings) of fermion fields,  $f$ , to Higgs fields,  $h$ ,  $H$ ,  $A$  and  $H^+$ . NEED TO WRITE ABOUT THE OTHER PARAMETERS I DON'T CARE ABOUT ALSO. There are four main types of 2HDM that are dependent on which Higgs doublet couples to which group of fermions, named type I, II, X (lepton-specific) and Y (flipped). The couplings of the fermion groups to the Higgs doublets are shown in Tab. 1.1, by convention  $\Phi_2$  is chosen to couple to up-like quarks.

	Type I	Type II	Type X	Type Y
$u$	$\Phi_2$	$\Phi_2$	$\Phi_2$	$\Phi_2$
$d$	$\Phi_2$	$\Phi_1$	$\Phi_2$	$\Phi_1$
$l$	$\Phi_2$	$\Phi_1$	$\Phi_1$	$\Phi_2$

Table 1.1: Table showing which Higgs doublet different fermion groups different types of 2HDMs couple to. By convention the  $u$  quark is chosen to couple to  $\Phi_2$ .

The type of 2HDM determines the formulae for the couplings,  $g$ , that are functions on two parameters: the CP-even ( $\alpha$ ) and CP-odd mixing angles ( $\beta$ ). These relative couplings are shown in Tab. 1.2.

	Type I	Type II	Type X	Type Y
$g_h^u$	$c_\alpha/s_\beta$	$c_\alpha/s_\beta$	$c_\alpha/s_\beta$	$c_\alpha/s_\beta$
$g_h^d$	$c_\alpha/s_\beta$	$-s_\alpha/c_\beta$	$c_\alpha/s_\beta$	$-s_\alpha/c_\beta$
$g_h^l$	$c_\alpha/s_\beta$	$-s_\alpha/c_\beta$	$-s_\alpha/c_\beta$	$c_\alpha/s_\beta$
$g_H^u$	$s_\alpha/s_\beta$	$s_\alpha/s_\beta$	$s_\alpha/s_\beta$	$s_\alpha/s_\beta$
$g_H^d$	$s_\alpha/s_\beta$	$c_\alpha/c_\beta$	$s_\alpha/s_\beta$	$c_\alpha/c_\beta$
$g_H^l$	$s_\alpha/s_\beta$	$c_\alpha/c_\beta$	$c_\alpha/c_\beta$	$s_\alpha/s_\beta$
$g_A^u$	$1/t_\beta$	$1/t_\beta$	$1/t_\beta$	$1/t_\beta$
$g_A^d$	$1/t_\beta$	$t_\beta$	$-1/t_\beta$	$t_\beta$
$g_A^l$	$1/t_\beta$	$t_\beta$	$t_\beta$	$-1/t_\beta$

Table 1.2: Table showing the couplings of fermion groups to additional neutral Higgs bosons in different types of 2HDMs. These are dependent on the mixing angles  $\alpha$  and  $\beta$ .  $t_x$ ,  $s_x$  and  $c_x$  represent  $\tan x$ ,  $\sin x$  and  $\cos x$  respectively.

In the alignment limit, for the normal scenario  $h_{SM} = h$ ,  $\sin(\beta - \alpha) = 1$ ,  $\Rightarrow c_\alpha/s_\beta = 1$ ,  $s_\alpha/c_\beta = -1$ ,  $s_\alpha/s_\beta = -1/t_\beta$ ,  $c_\alpha/c_\beta = t_\beta$ .

In the alignment limit, for the inverted scenario  $h_{SM} = H$ ,  $\cos(\beta - \alpha) = 1 \implies c_\alpha/s_\beta = 1/t_\beta$ ,  $s_\alpha/c_\beta = t_\beta$ ,  $s_\alpha/s_\beta = 1$ ,  $c_\alpha/c_\beta = 1$ .

## 1.3 Theoretical Problems and Potential Solutions

### 1.3.1 Hierarchy Problem

The current best measurement for the SM Higgs boson's mass from the CMS experiment is  $125.38 \pm 0.14$  GeV [24]. The lightness is a concern when considering "naturalness" and physics beyond the SM. The Higgs boson has loop corrections to calculations of its mass. Known heavy fermions already provide significant contributions (orders of magnitude greater than the observed mass) to the Higgs boson's calculated mass. On top of this, treating the SM as an effective field theory, new physics would be expected in the unexplored regions between the weak scale and the Planck scale. But if a new fermion were to be found,  $f$ , or a heavy scalar,  $S$ , in such a range, the Higgs boson would be subject to even greater changes in its predicted mass. The Feynman diagrams for the mass corrections for a fermion and a scalar are shown in Fig. 1.2.

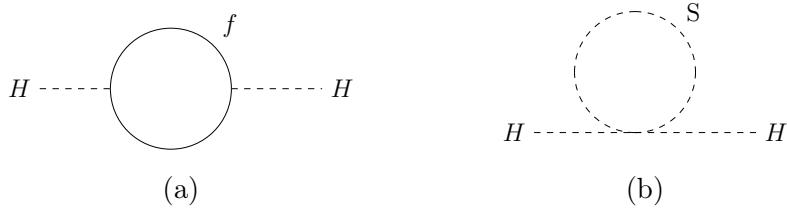


Figure 1.2: One-loop corrections to the Higgs mass by a fermion  $f$  (a) and a scalar  $S$  (b).

Figure 1.2a shows a diagrammatic representation of the fermionic correction to the Higgs boson's mass. The Lagrangian term for the coupling of the Higgs field to fermions is  $-\lambda_f H \bar{f} f$ . Hence, it can be determined that the correction to the mass is

$$\Delta m_H^2 = -\frac{\lambda_f}{16\pi^2} \left[ \Lambda_{UV}^2 - 2m_f^2 \ln \left( \frac{\Lambda_{UV}}{m_f} \right) \right] + \dots, \quad (1.17)$$

where  $\Lambda_{UV}$  is the ultraviolet momentum cut off [25]. Beyond this, our effective field theory would be expected to break down and new physics to be found.

Figure 1.2b illustrates a correction to the Higgs mass by a scalar particle. The coupling of a scalar to the Higgs field is represented by the Lagrangian term  $-\lambda_S |H|^2 |S|^2$ . The Higgs mass correction for such a term is derived to be

$$\Delta m_H^2 = \frac{\lambda_S^2}{16\pi^2} \left[ \Lambda_{UV}^2 - 2m_S^2 \ln \left( \frac{\Lambda_{UV}}{m_S} \right) \right] + \dots \quad (1.18)$$

Eq.1.17 and Eq.1.18 show that if the mass of the scalar is equivalent to that of the fermion and  $\lambda_f = \lambda_S^2$ , then the Higgs mass corrections cancel. This offers a solution to the hierarchy problem introducing a new symmetry that extends the Standard Model. The symmetry relates fermions and bosons and is known as Supersymmetry. It states that fermions and bosons exist in groups called supermultiplets. Each supermultiplet contains fermion and boson states which are superpartners of one another. On-shell each supermultiplet must have an equivalent number of fermionic and bosonic degrees of freedom. In order for this to also hold off-shell, an auxiliary field is added to balance the number of degrees of freedom.

If Supersymmetry is an unbroken theory, then one would expect to have the superpartners at the same mass as the Standard Model particles. This has not been seen experimentally, therefore Supersymmetry must be a broken theory in the vacuum state. Soft supersymmetry breaking can be introduced through the addition of SUSY violating Lagrangian term  $\mathcal{L}_{\text{soft}}$ , where

$$\mathcal{L} = \mathcal{L}_{\text{SUSY}} + \mathcal{L}_{\text{soft}}. \quad (1.19)$$

$\mathcal{L}_{\text{soft}}$  contains only mass terms and coupling parameters. Defining  $m_{\text{soft}}$  as the largest mass scale involved in the soft Lagrangian,  $m_{\text{soft}}$  also then defines the mass splitting between the SM and supersymmetric particles. If the mass splitting becomes significant, the hierarchy problem would be reintroduced as corrections to the Higgs mass would again become large.

The Minimal Supersymmetric Standard Model (MSSM) is simplest implementation of Supersymmetry. It introduces sets of new particles named squarks, sleptons, gauginos and Higgsinos as superpartners to quarks, leptons, gauge bosons and the Higgs bosons. Also added are neutralinos and charginos that are combined of gauginos and Higgsinos. Additional contributions to the particle content comes from the Higgs sector. The Higgs sector of the MSSM is required to be extended to two Higgs doublets in order to maintain the gauge symmetries and to cancel quantum mechanical inconsistencies []. In particular a type II 2HDM is needed to ensure natural Yukawa couplings, minimal flavour violation and tree-level mass relations,

that all cannot be achieved with a different extended Higgs sector.

## 1.4 Experimental Tensions and Potential Solutions

### 1.4.1 B Anomalies

Recent measurements from the LHCb experiment, testing lepton flavour conservation, have found deviations away from the Standard Model. The measurement of  $R_K$ ,  $R_{K^*}$  and  $R_{D(0)}$  have deviations with significances of 3.1, 2.1-2.5 and 3.1  $\sigma$  respectively [?, ?, ?]. These B-anomalies have prompted the idea for a short range lepton flavour violating interaction. This interaction is theorised to be mediated by leptoquarks. In an attempt to fit a model that offers a combined explanation of these results, it was found that a  $U_1$  vector leptoquark was the only leptoquark that could offer a simultaneous explanation of all anomalous results [?]. Such a leptoquark would couple to fermions by the Lagrangian shown below.

$$\mathcal{L}_U = \frac{g_U}{\sqrt{2}} U^\mu [\beta_L^{i\alpha} (\bar{q}_L^i \gamma_\mu l_L^\alpha) + \beta_R^{i\alpha} (\bar{d}_R^i \gamma_\mu e_R^\alpha)] + \text{h.c.} \quad (1.20)$$

where  $g_U$  is the coupling scaling parameter and  $\beta_L$  and  $\beta_R$  are the left and right-handed mixing matrices

$$\beta_L = \begin{pmatrix} 0 & 0 & \beta_L^{d\tau} \\ 0 & \beta_L^{s\mu} & \beta_L^{s\tau} \\ 0 & \beta_L^{b\mu} & 1 \end{pmatrix}, \quad \beta_R = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \beta_R^{b\tau} \end{pmatrix}. \quad (1.21)$$

The fit to B-anomalies done in Ref. [?], found the best fit values for each left-handed mixing matrix parameter based on two scenarios for  $\beta_R^{b\tau}$ , namely  $\beta_R^{b\tau} = 0$  and  $\beta_R^{b\tau} = -1$ . These represent no and maximal right-handed contributions. The most stringent constraints on these models are imposed by high- $p_T$  di-tau tails. At the LHC the most dominant production mode of this final state would be given in Figure 4.3.

### 1.4.2 g-2 Anomaly

The measurement of the muon anomalous magnetic moment from the Fermilab National Accelerator Laboratory (FNAL) Muon g-2 experiment [16], combined with

earlier results from the Brookhaven National Laboratory (BNL) E821 measurement [26], finds the difference of  $a_\mu$  between experiment and SM prediction to be,

$$\Delta a_\mu^{\text{obs}} = a_\mu^{\text{exp}} - a_\mu^{\text{SM}} = (251 \pm 59) \times 10^{-11}, \quad (1.22)$$

where  $a_\mu = (g - 2)_\mu / 2$ . This is a  $4.2 \sigma$  deviation away from the SM expectation. One potential solution to this deviation is a 2HDM. This contributes in two ways to  $\Delta a_\mu$ , one-loop and two-loop Bar-Zee contributions [27, 28], that can be mediated by any of the additional Higgs bosons.

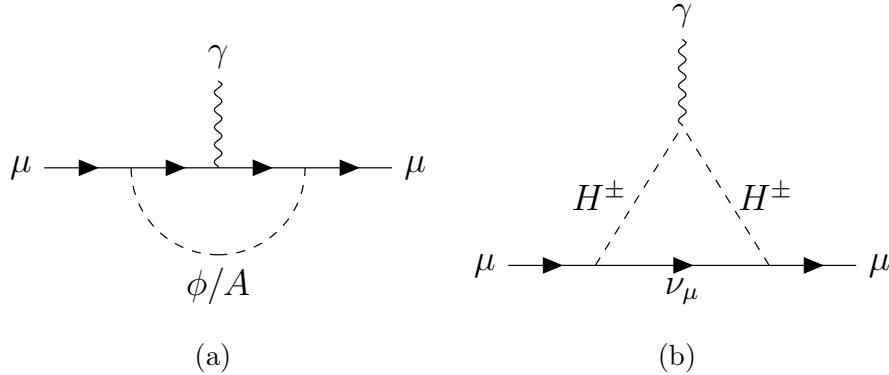


Figure 1.3: Examples of one-loop Bar-Zee Feynman diagrams for the contribution from  $\phi$  and  $A$  in (a) and  $H^\pm$  in (b).

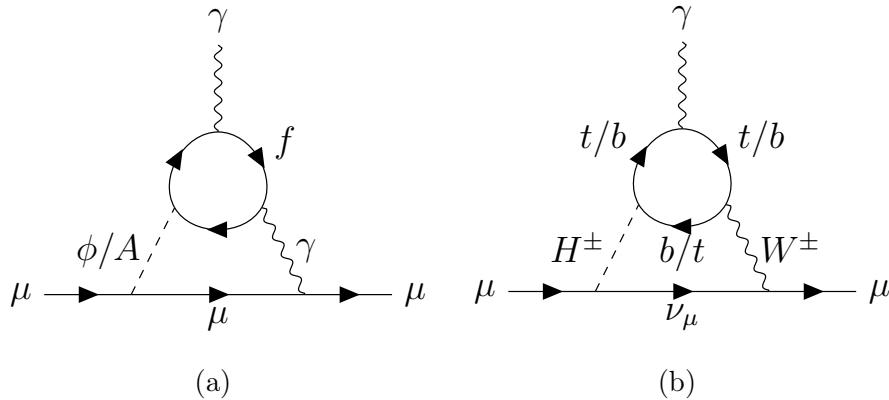


Figure 1.4: Examples of two-loop Bar-Zee Feynman diagrams for the contribution from  $\phi$  and  $A$  in (a) and  $H^\pm$  in (b). Note that other fundamental particles can also contribute to the loop.

The one-loop contribution to  $\Delta a_\mu$  is positive for  $\phi$  and negative from  $A$  and  $H^\pm$ , however the more significant contribution comes from two-loop Bar-Zee diagrams with heavy fermions in the loop which gives a positive shift to  $\Delta a_\mu$ . In order to

fulfil the requirements for the anomaly, an enhanced couplings between additional Higgs bosons to muons is needed. This gives two options: a type II or a type X 2HDM, both at large values of  $\tan \beta$ . The type II 2HDM, that also offers enhanced couplings to down-like quarks, is heavily constrained by LEP, Tevatron and LHC searches and an available region of phase space to explain the anomaly is not easily found. However, the type X 2HDM, with only lepton coupling enhancements at high  $\tan \beta$ , is relatively unconstrained due to suppressed quark initiated production modes of additional Higgs bosons.

Ref. [5] puts an upper limit on the  $m_A \lesssim 200$  GeV, that is also required for an explanation of  $\Delta a_\mu^{\text{obs}}$ . Further constraints are placed on the phase space from theoretical stabilities, electroweak precision measurements and collider bounds. The final available phase space for a type X 2HDM in the alignment scenario to explain the g-2 anomaly is shown in Tab. 1.3.

Scenario	$\tan \beta$	$m_A$ (GeV)	$m_\phi$ (GeV)	$m_{H^\pm}$ (GeV)
Normal	$\geq 90$	[62.5,145]	[130,245]	[95,285]
Inverted	$\geq 120$	[70,105]	[100,120]	[95,185]

Table 1.3: Regions of interest for g-2 anomaly with respect to the type X 2HDM in the normal and inverted alignment scenarios as suggested in Ref. [5].

# Chapter 2

## The LHC and CMS Experiment

### 2.1 The LHC

The Large Hadron Collider (LHC) [29], located at CERN just outside Geneva, is a synchrotron measuring 27 km in circumference, installed in the tunnel previously used by the Large Electron-Positron (LEP) accelerator [30]. Upon design, its primary purpose was to provide collisions between proton beams, generating center-of-mass energies ( $\sqrt{s}$ ) of up to 14 TeV and an instantaneous luminosity of approximately  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ . Fig 2.1 illustrates the arrangement of the LHC and the CERN accelerator complex. Protons are supplied to the LHC through a sequence of accelerators: Linac4 (Linac2 pre-2020), Proton Synchrotron Booster (PSB), Proton Synchrotron, and Super Proton Synchrotron (SPS), that successively raise the energies of the protons to 50 MeV, 1.4 GeV, 25 GeV, and 450 GeV respectively. When the final energy has been achieved, the SPS injects bunched protons into the LHC's beam pipes as two counter-rotating beams. Each bunch containing over  $10^{11}$  protons, with each one separated by 25 ns and each beam consists of 2808 bunches. The protons are accelerated to collision energy using eight 400 MHz radio frequency (RF) cavities and kept in a circular trajectory with 1232 niobium-titanium superconducting dipole magnets. These magnets must be maintained at their operating temperature of 1.9 K to generate magnetic fields up to 8.4 T, requiring the use of superfluid helium. The bunches are collided at four intersections point surrounded by the ALICE [31], ATLAS [32], CMS [2] and LHCb [33] detectors, at a collision rate of 40 MHz.

The rate of events for a process produced in LHC collisions can be expressed as,

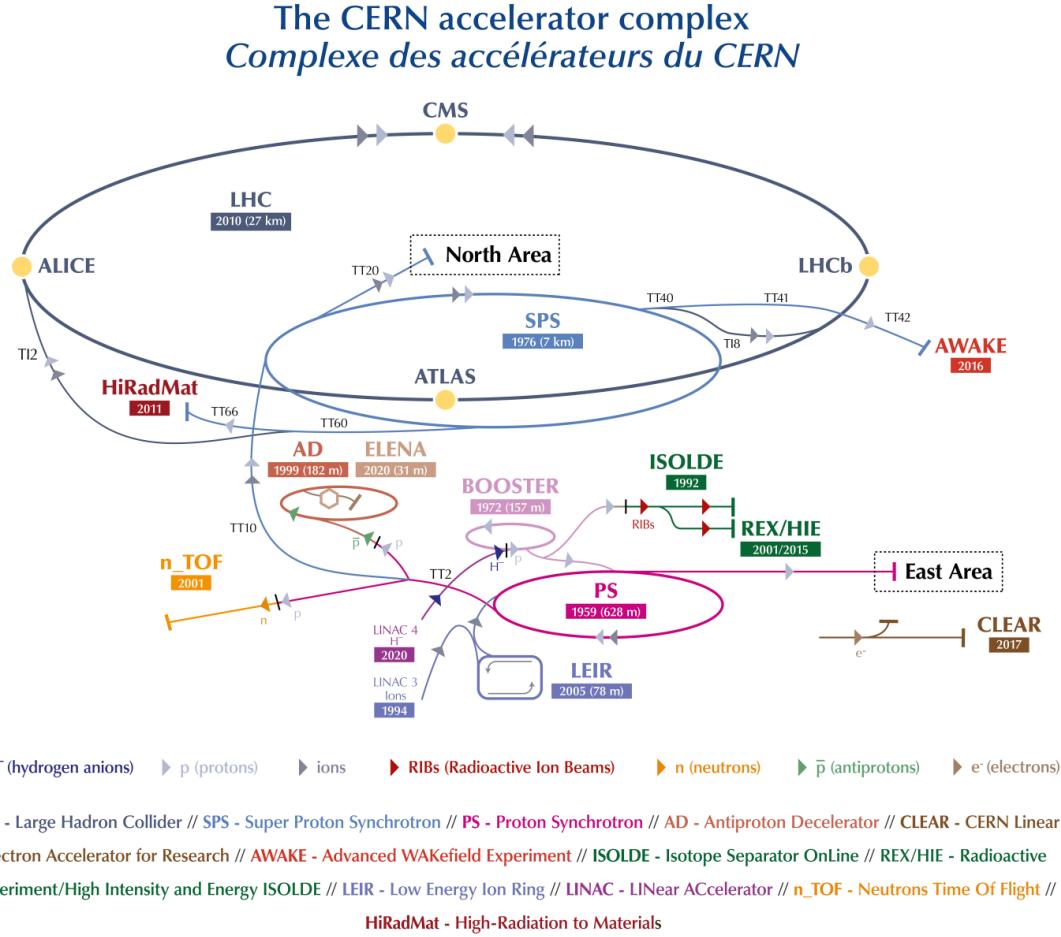


Figure 2.1: A schematic diagram of the CERN [1].

$$R_{\text{event}} = L_{\text{inst}} \sigma(\sqrt{s}), \quad (2.1)$$

where  $\sigma$  represents the cross section of the process and is a function of the center-of-mass energy, and  $L_{\text{inst}}$  denotes the LHC machine instantaneous luminosity, that depends only on the beam parameters and can be calculated by,

$$L_{\text{inst}} = \frac{n_b N_b^2 f_{\text{rev}} \gamma_r}{4\pi \epsilon_n \beta^*} F. \quad (2.2)$$

where  $n_b$  is the number of bunches per beam,  $N_b$  is the number of particles per bunch,  $f_{\text{rev}}$  is the revolution frequency,  $\gamma_r$  is the relativistic gamma factor,  $\epsilon_n$  is the normalized transverse beam emittance,  $\beta^*$  is the beta-function at the collision point, and  $F$  is a reduction factor which accounts for the crossing angle of the beams at the collision point. One disadvantage of an increase the instantaneous luminosity is the increase of pileup (PU), defined as the number of additional inelastic proton-proton

collisions per bunch crossing.

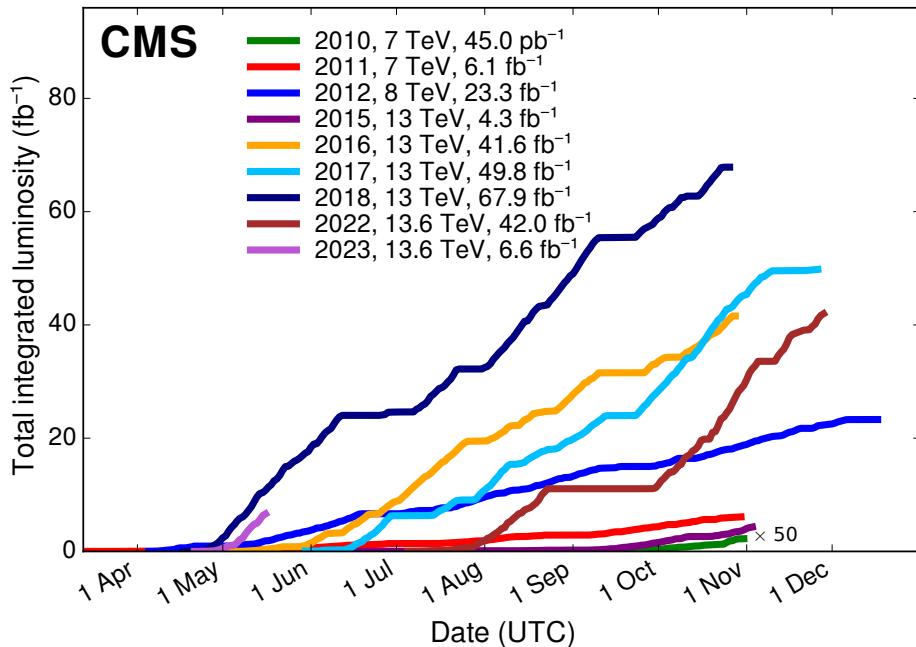


Figure 2.2: The total integrated luminosity for proton-proton collisions collected by the CMS experiment between 2010 and May 2023 [1].

The LHC began its first physics collisions in 2010, initially colliding beams at  $\sqrt{s} = 7$  TeV, which was increased to 8 TeV throughout 2012. Following this data-taking period, known as Run 1, the LHC underwent a two-year Long Shutdown 1 (LS1) to undergo upgrades in preparation for an increase in  $\sqrt{s}$ . In 2015, the LHC was restarted, initiating its Run 2 phase of data collection at  $\sqrt{s} = 13$  TeV, which lasted until the end of 2018. During Run 2, the LHC achieved and surpassed its original design luminosity by obtaining a record peak luminosity of  $2.1 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$  in 2018. Subsequently, the LHC performed a second update period, the Long Shutdown 2 (LS2), lasting for approximately three years, where-after the data collection for Run 3 began at  $\sqrt{s} = 13.6$  TeV. Fig 2.2 illustrates the total integrated luminosity of proton-proton collisions delivered to the CMS detector at the time of writing. The data used for the analyses described in this thesis correspond to the full Run 2 dataset collected during the 2016–2018 data-taking periods at 13 TeV. Only data recorded by CMS, where all sub-detectors were functioning correctly, are certified for use in physics analyses. This equates to  $36.3 \text{ fb}^{-1}$ ,  $41.5 \text{ fb}^{-1}$  and  $59.7 \text{ fb}^{-1}$  of

data collected in 2016, 2017 and 2018 respectively.

## 2.2 The CMS Detector

The Compact Muon Solenoid (CMS) detector was engineered to fulfil the rigorous demands of the LHC physics program. Its primary objective is to achieve sensitivity to the Higgs boson and novel phenomena at the TeV energy scale. Weighing 12,500 tonnes and measuring 21.6 m in length with a diameter of 14.6 m, the CMS detector uses an array of sub-detectors encircling the central beam axis. The layout of the CMS detector is shown in Fig. 2.3. A superconducting solenoid, operating at a magnetic field strength of 3.8 T, which encompasses the inner tracker, electromagnetic calorimeter, and hadronic calorimeter. Situated outside the solenoid within the iron return yoke are gaseous muon detectors, positioned to accurately measure muons. The CMS detector adopts a coordinate system centered at the collision point, with the y-axis oriented vertically, the x-axis directed radially inward toward the LHC center, and the z-axis aligned with the beam direction. Transverse energy ( $E_T$ ) and transverse momentum ( $p_T$ ) are defined in the x-y plane, and the azimuthal angle ( $\phi$ ) and polar angle ( $\theta$ ) are measured relative to the x-axis and z-axis, respectively.  $r$  is used as the radial distance in the x-y plane. Pseudorapidity ( $\eta$ ), defined as  $\eta = -\ln[\tan(\theta/2)]$ , is used due to its gauge invariance, and distances in the  $\phi$ - $\eta$  plane are characterised by the metric  $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ .

### 2.2.1 Tracker

Closest to the interaction point in the CMS experiment is the tracker [2, 34, 35], which is essential for precise measurements of charged particle trajectories and the determination of primary and secondary vertices. To meet the requirements of high granularity and fast response for the large number of particles generated in each bunch crossing, as well as being radiation-hard to deal with the particle flux, a silicon tracking detector is utilised. The tracker consists of a pixel detector and a silicon strip detector as shown in Fig. 2.4.

The pixel detector covers the pseudorapidity range  $|\eta| < 2.5$  and is composed of three cylindrical pixel modules and two disk pixel modules. It contains 66 million silicon pixels, each with dimensions of  $100 \times 150 \mu\text{m}^2$ . This configuration provides a

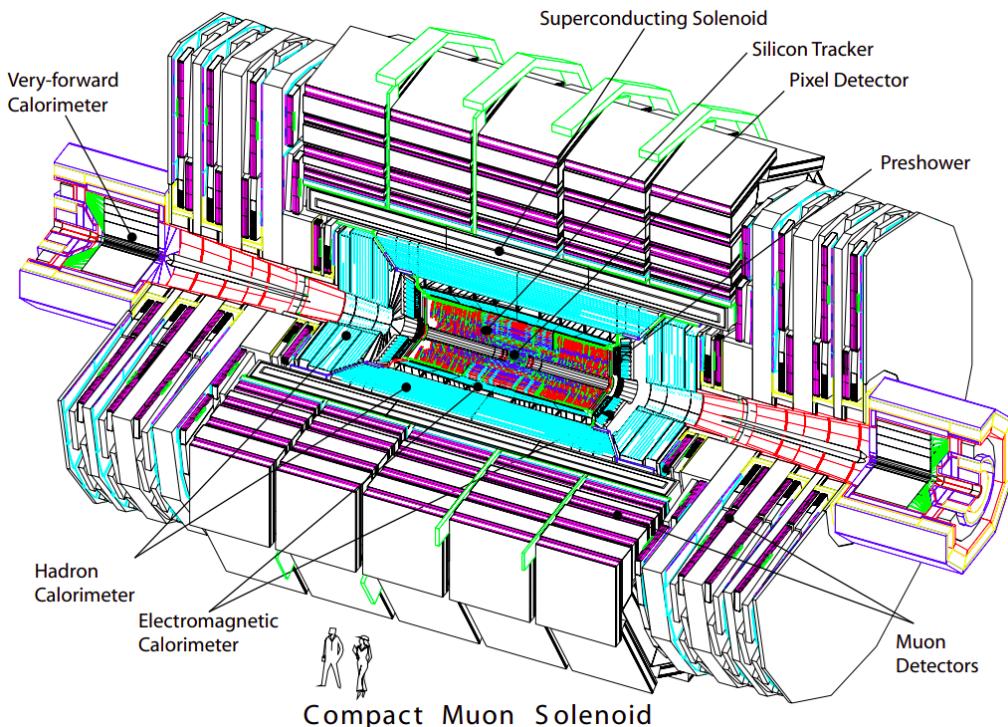


Figure 2.3: A perspective view of the CMS detector [2].

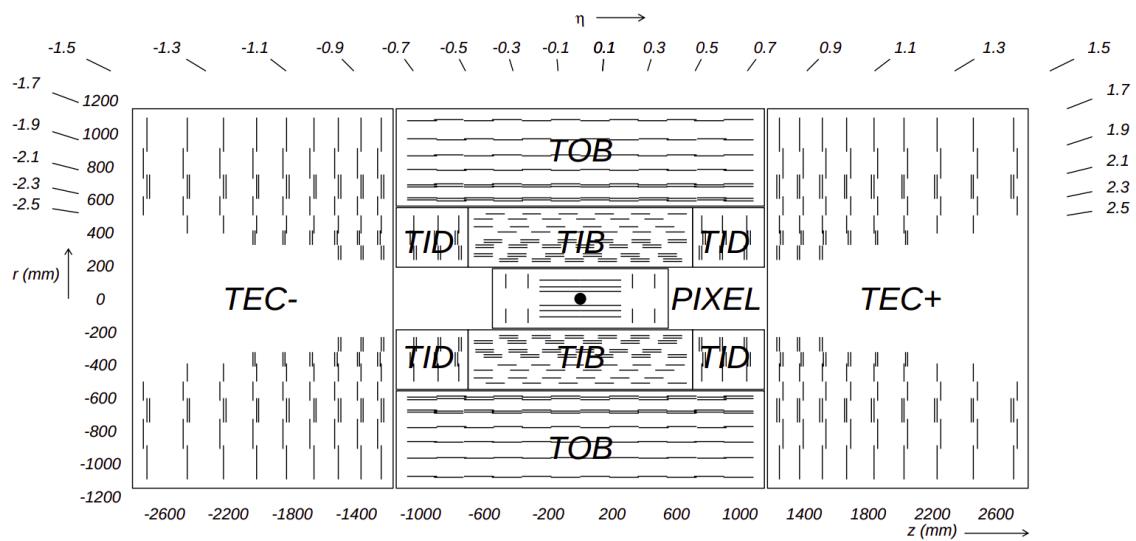


Figure 2.4: Schematic of the CMS tracker (pre-pixel upgrade) in the  $r$ - $z$  plane, showing the position of the pixel detector as well as the tracker inner barrel (TIB), tracker inner disks (TID), tracker outer barrel (TOB), and tracker endcaps (TEC) strip detectors. The lines represent detector modules and the double lines represent back-to-back modules [2].

spatial resolution of  $15 - 20 \mu\text{m}$  in both the  $r$ - $\phi$  plane and the  $z$  direction, enabling 3-dimensional vertex reconstruction. The original tracker was built to handle as

instantaneous luminosity of  $10^{34} \text{ cm}^2\text{s}^{-1}$  and an average PU of 25. To handle higher luminosities and increased event PU, the pixel detector was upgraded in 2016/2017 to handle approximately double the instantaneous luminosity and PU [35]. The upgraded detector consists of four barrel module layers and three endcap disks, resulting in 124 million pixels.

Surrounding the pixel detector is the silicon strip detector, which is divided into four subsystems: the tracker inner barrel (TIB), tracker inner disks (TID), tracker outer barrel (TOB), and tracker endcaps (TEC). The TIB and TID provide four layers of silicon strip detectors in the barrel region and three disks at each end, extending up to a radius of 55 cm. The TOB consists of six barrel layers extending up to an outer radius of 116 cm, while the TEC comprises nine disks covering a range of  $|z|$  from 124 cm to 282 cm. The silicon strips have various thickness and widths, providing multiple measurements of the  $r\phi$  position with resolutions ranging from 23–35  $\mu\text{m}$  in the TIB to 35–53  $\mu\text{m}$  in the TOB.

In addition to the main components, the tracker includes back-to-back mounted micro-strip detector modules to provide additional measurements of the  $z$  coordinate in specific regions. The overall tracker layout ensures the presence of at least three hits in the pixel detector (at least four hits for the upgraded detector) and at least nine hits in the silicon-strip tracker, with a minimum of four 2-dimensional measurements among them.

### 2.2.2 Electromagnetic calorimeter

The CMS electromagnetic calorimeter (ECAL) is a hermetic homogeneous calorimeter designed to detect high-energy electrons and photons [2, 36]. In total, 75,848 lead tungstate ( $\text{PbWO}_4$ ) crystals are sorted in a barrel and endcap configuration.  $\text{PbWO}_4$  was chosen as the crystal material due to its radiation-hardness, high density, short radiation length, and small Molière radius (the average radius containing on average 90% of a shower’s total energy deposit), enabling the construction of a compact and finely granular ECAL. Scintillation light produced by showering electrons and photons in the crystals is converted into an electrical signal by photodetectors such as avalanche photodiodes in the barrel and vacuum phototriodes in the endcaps. The scintillation decay time of the crystals matches the 25 ns LHC bunch crossing time, ensuring that a significant portion of the light is emitted between bunch crossings.

The ECAL is placed outside of the tracker but within the magnet bore and covers the pseudorapidity range  $|\eta| < 3.0$ . It comprises of an ECAL barrel (EB) covering  $|\eta| < 1.479$  and an ECAL endcap (EE) covering  $1.479 < |\eta| < 3.0$ . A diagram of this is shown in Fig. 2.5. The barrel region consists of 61,200 crystals with 360-fold granularity in  $\phi$  and  $(2 \times 85)$ -fold granularity in  $\eta$ . The crystals are tapered with a front face area of  $0.0174 \times 0.0174$  in  $\eta\text{-}\phi$  ( $22 \times 22$  mm $^2$ ) and a length of 230 mm. The endcaps house 7,324 crystals arranged in a rectangular  $x\text{-}y$  grid, each with a front face area of  $28.62 \times 28.62$  mm $^2$  and a length of 220 mm.

The ECAL system also includes Preshower detectors placed in front of each endcap to identify photons from neutral pion decays and improve electron identification and position resolution. These detectors consist of lead radiators to initiate showering and silicon-strip sensors to measure the deposited energy. The ECAL energy resolution is parametrized as a function of the incident particle energy (E), with terms for the stochastic (S), noise (N), and constant (C) contributions.

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right) + C^2 \quad (2.3)$$

The stochastic term captures fluctuations in lateral shower containment and photon yield, the noise term accounts for electronics noise and pileup, and the constant term arises from non-uniformity of the longitudinal response and calibration errors. Measurements using electron beams have determined the values of  $S = 0.028$  GeV $^{1/2}$ ,  $N = 0.12$  GeV, and  $C = 0.003$  for the ECAL energy resolution.

### 2.2.3 Hadron calorimeter

### 2.2.4 Muon system

### 2.2.5 Triggering

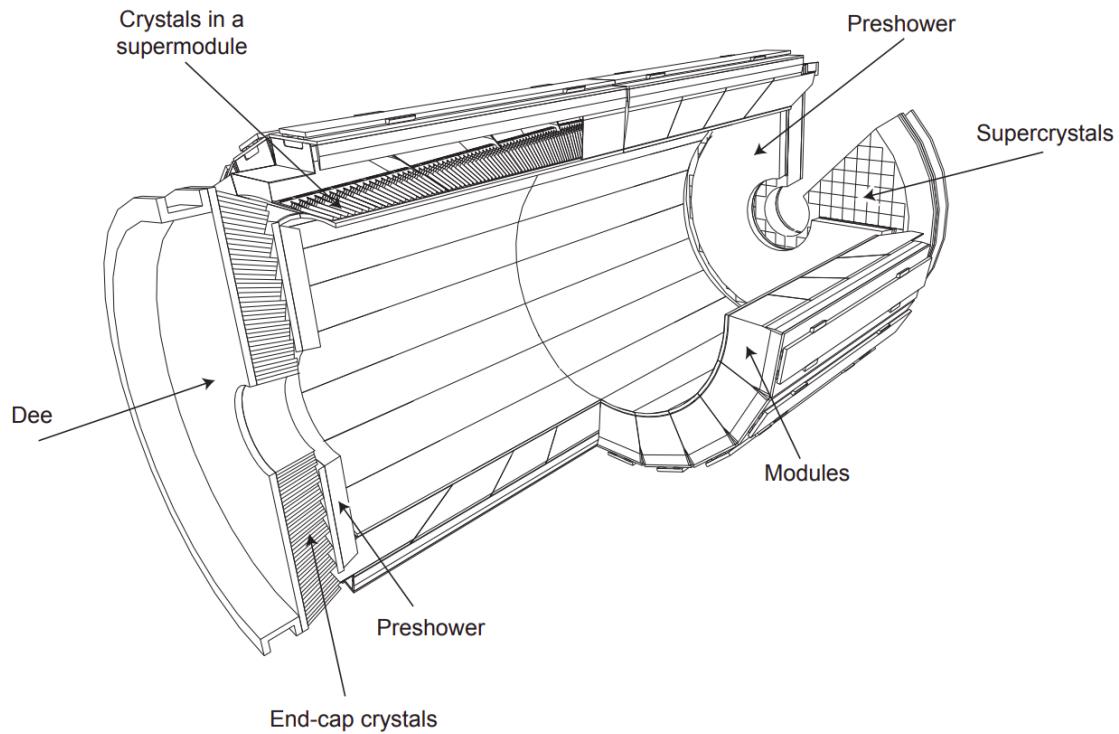


Figure 2.5: A schematic of the CMS ECAL detector [2].

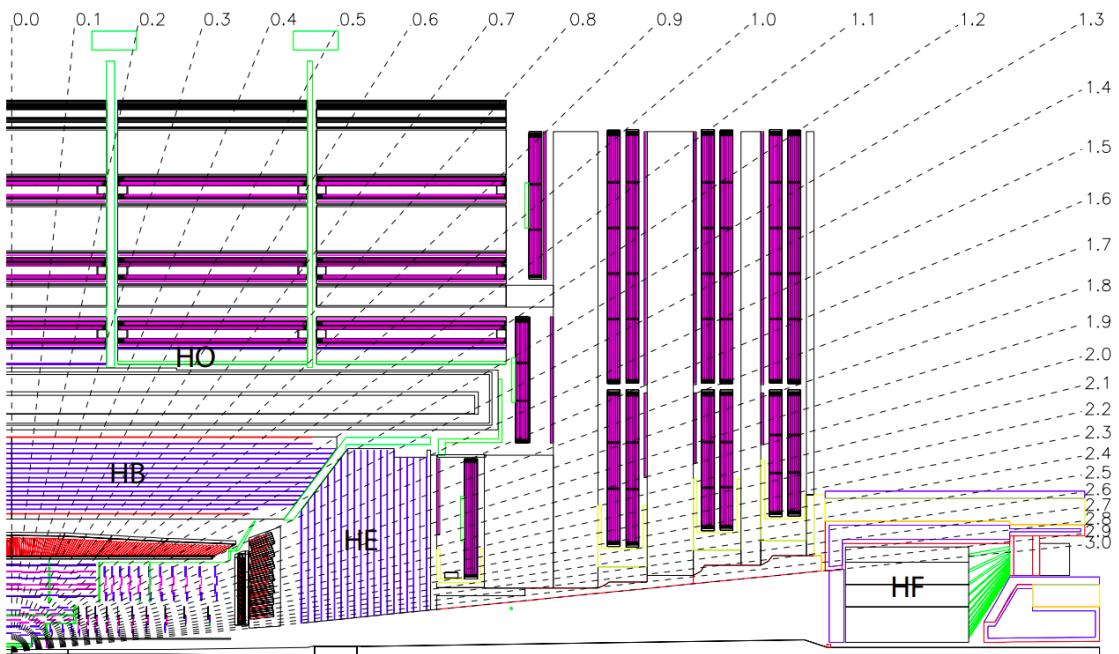


Figure 2.6: .

# Chapter 3

## Object Reconstruction

### 3.1 Tracks and vertices

The reconstructed vertex with the largest value of summed physics-object  $p_T^2$  is taken to be the primary pp interaction vertex. The physics objects for this purpose are the jets, clustered using the jet finding algorithm [37, 38], as described below, with the tracks assigned to the vertex as inputs, and the associated missing transverse momentum, taken as the negative vector sum of the  $p_T$  of those jets. Any other collision vertices in the event are associated with additional soft inelastic pp collisions called *pileup*.

### 3.2 Particle flow

The particle-flow (PF) algorithm reconstructs the products of the LHC pp collisions and is described in full in Ref.[39]. It utilises all the information available from the tracker, ECAL, HCAL and muon detectors combined to produce a list of particle candidates. These candidates are either a photon, electron, muon or a neutral or charged hadrons. It begins with defining an event as the data taken per bunch crossing. The PF algorithm then reconstructs the tracks of the particle candidates in order to find the collision vertices. The primary collision vertex is taken to be the one with the largest value of  $p_T^2$  summed over all physics objects originated from that vertex. Physics objects are not only defined to include particle candidate tracks, but also missing tracks represented by the negative vectorial sum of all particle candidate tracks. Other pp collisions vertices are referred to as pileup.

In reconstructing electron and muons, the energy deposits in the ECAL and the track hits in the muon chamber respectively, working alongside the tracker, provide the basis of electron and muon identification. However, additional requirements are used to drop misidentification rates by ensuring that the electron or muon is isolated from any hadronic activity in the detector, as leptons do not carry colour charge. This is done by defining a relative isolation variable  $I_{\text{rel}}^{e(\mu)}$  in the following way

$$I_{\text{rel}}^{e\mu} = \frac{\sum p_{T,i} + \sum E_{T,i}}{p_T^{e\mu}}. \quad (3.1)$$

The sums are over all particles included in a cone of radius  $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$  excluding the electron or muon itself.  $\Delta\eta$  and  $\Delta\phi$  are the angular distance in  $\eta$  and  $\phi$  around the electron or muon direction from the primary vertex. To remove problems with pileup, only charged particles originating from the primary vertex are included. To remove neutral particles from pileup in the cone, the  $p_T$  for neutral particles is estimated by subtracting half of the sum of the  $p_T$  of charged particle in the cone, due to the approximate ratio of charged to neutral hadron production. The cone size selected for electrons is  $\Delta R < 0.3$  and the isolation variable is  $I_{\text{rel}}^{e(\mu)} < 0.1$ . For muons is cone size is  $\Delta R < 0.4$  and the isolation variable is  $I_{\text{rel}}^{e(\mu)} < 0.15$ .

Jets originating from the hadronisation of b quarks, are identified using the combined secondary vertex b-tagging algorithm. This discriminates between jets originating from b quarks from other jets, utilising track impact parameters and secondary vertex related variables [40]. This plays a key part in the analysis, as b-tagging is used for categorisation purposes described in Section ???. The missing transverse momentum,  $\vec{p}_T^{\text{miss}}$ , is also used in categorisation of events and is calculated as the negative vector sum of all PF reconstructed transverse momenta.

### 3.3 Muons

Muons are required to be reconstructed as muon by the particle flow reconstruction. They additionally must pass *Medium muon* requirements [?], as recommended by the Muon POG. These requirements are:

- The muon is reconstructed by the *tracker* or *global* muon reconstruction algorithm.
- The impact parameters  $d_{xy}$  and  $d_z$  between the muon track ('best track') and

the primary vertex are restricted as  $d_{xy} < 0.045$  cm and  $d_z < 0.2$  cm to ensure the muon is associated with the primary vertex

- At least 80% of the tracker hits have to be valid

In addition, either of the following two sets of criteria must be satisfied:

- The muon is reconstructed by the *global* muon reconstruction algorithm.
- The  $\chi^2/\text{ndof}$  of the global track fit is smaller than 3
- The  $\chi^2$  of the tracker-standalone position match is smaller than 12
- The  $\chi^2$  of the track kink finder is less than 20
- The muon segment compatibility is  $> 0.303$

*or:*

- The muon segment compatibility is  $> 0.451$

## Isolation

To reduce the contamination from muons originating from heavy-flavoured quark decays in jets or decays in flight, selected muons are required to be isolated. The isolation is based on photon and neutral/charged hadron particle flow candidates within a cone size of  $\Delta R < 0.4$  of the selected muon. A relative combined isolation variable is defined as:

$$I_{\text{rel}}^\mu = \frac{\Sigma P_T(\text{charged}) + \max(\Sigma E_T(\text{neutral}) + \Sigma E_T(\text{photon}) - \Delta\beta\Sigma E_T(\text{PU}), 0)}{P_T^\mu}, \quad (3.2)$$

where  $P_T(\text{charged})$  corresponds to the  $P_T$  of all charged hadronic candidates from primary vertex,  $E_T(\text{photon, neutral})$  to the transverse energy of all photon and neutral hadron candidates, and  $\Delta\beta$  is the energy estimate of neutral particles due to pileup, which is taken to be 0.5. Furthermore  $E_T(\text{PU})$  corresponds to the transverse momentum of all charged hadronic candidates from pileup. The muon isolation used is  $I_{\text{rel}} < 0.15$ .

## 3.4 Electrons

In addition, electrons are required to pass an identification variable based on a Boosted Decision Tree (BDT) discriminator which uses track quality, shower shapes and kinematic quantities as input. The following variables are used as input to the BDT: [?]

- Cluster shape variables  $\sigma_{i\eta,i\eta}$  and  $\sigma_{i\phi,i\phi}$ , with  $i\eta$  and  $i\phi$  the integer label of the  $\eta$  and  $\phi$  calorimeter cell. The circularity =  $1 - \frac{E1 \times 5}{E5 \times 5}$ , with  $E1 \times 5$  and  $E5 \times 5$  the energies in a  $1 \times 5$  and a  $5 \times 5$  grid around the super cluster seed, respectively. Shape variable  $R9 = \frac{E3 \times 3}{E_{SC}}$ , with  $E3 \times 3$  the energy in a  $3 \times 3$  grid of cells around the super cluster seed and  $E_{SC}$  the raw energy of the super cluster.
- The number of valid hits in the track fit, the  $\chi^2$  of the track fit and the  $\chi^2$  of the GSFTrack fit
- The number of GSFtrack hits, the number of expected missing inner hits and the result of the conversion vertex fit
- The distance  $\Delta\eta$  and  $\Delta\phi$  between the reconstructed super cluster and the associated track at the position of the primary vertex, and the distance in  $\eta$  between the super cluster and the track at the calorimeter surface.
- H/E, the ratio of the hadronic energy over the electromagnetic energy in the super cluster and E/P, the ratio of the super cluster energy over the momentum of the track associated with the electron
- The ratio of the energy of the electron cluster and the momentum of the associated track, evaluated at the electron cluster, and  $1/E_e - 1/P_e$ , with  $E_e$  the energy of the electron candidate and  $P_e$  its momentum.

The BDT was trained on a  $Z/\gamma^*$  Monte Carlo sample generated with MadGraph 5, in 3  $\eta$  bins for electrons with  $p_T > 10$  GeV. This analysis uses the version of the ID BDT without the isolation included in the training, instead using an additional cut on the electron isolation. This was shown to give similar performance to the version of the ID including the isolation and makes defining side-band regions used for estimation of backgrounds simpler.

The electron are additionally required to pass the 90% efficiency working point. Finally electrons are also subject to the same impact parameter requirements as

the muons: the impact parameters  $d_{xy}$  and  $d_z$  between the electron track and the primary vertex are restricted as  $d_{xy} < 0.045$  cm and  $d_z < 0.2$  cm to ensure the electron is associated with the primary vertex.

The isolation used is based on photon and neutral/charged hadron particle flow candidates within a cone size of  $\Delta R < 0.3$  of the selected electron but used a different method to estimate the neutral particles due to pileup, namely the rho-effective-area method. The pileup in this method is estimated as  $PU = \rho \cdot EA$ , where  $\rho$  is the event-specific average pile-up energy density per unit area in the  $\phi\text{-}\eta$  plane and the EA is the effective area specific to the given type of isolation. The rho-effective-area subtracted relative combined isolation variable is defined as:

$$I_{\text{rel}}^e = \frac{\Sigma P_T(\text{charged}) + \max(\Sigma E_T(\text{neutral}) + \Sigma E_T(\text{photon}) - \rho \cdot EA, 0)}{P_T^e}, \quad (3.3)$$

where the EA is measured in bins of  $\eta$  as listed in Table 3.1. The electron isolation used is  $I_{\text{rel}} < 0.15$ .

$\eta$ range	EA
$0.0 \leq \eta < 1.0$	0.1440
$1.0 \leq \eta < 1.479$	0.1562
$1.479 \leq \eta < 2.0$	0.1032
$2.0 \leq \eta < 2.2$	0.0859
$2.2 \leq \eta < 2.3$	0.1116
$2.3 \leq \eta < 2.4$	0.1321
$2.4 \leq \eta < 5.0$	0.1654

Table 3.1: Electron effective areas used for the  $\rho$ -corrected isolation computation.

## 3.5 Jets

Jets are clustered from all particle flow candidates using the anti  $k_T$  algorithm [38] with a cone size parameter  $R=0.4$ . To reduce the contribution of jets that originate from pile-up vertices, charged hadrons that are not associated with the hard-scattering primary vertex are removed from the particle flow candidates that form the input to the anti- $k_T$ -algorithm (charged hadron subtraction).

Practically, this means we use the "slimmedJets" collection from the miniAOD dataformat, where the jet-energy corrections from global tags

The jets are required to pass the tight working point of the PF jet ID discriminator provided by the jetMET POG [?]. To exclude selected  $e, \mu, \tau_h$  candidates from the

jet collection, the jets are required to be separated from the selected  $\tau$  candidate by  $\Delta R > 0.5$ . Jets with a corrected  $p_T > 30$  GeV and  $|\eta| < 4.7$  are considered.

High endcap ECAL noise in 2017 led to large amounts of noise in reconstructed jets [?]. To mitigate the noise issue, jets reconstructed on 2017 data also used a veto of jets with raw  $p_t < 50$  and  $2.65 < |\eta| < 3.139$ .

### 3.6 b jets

For determining information about the b-tagging status of the jets, the DeepJet algorithm [?] is used. Jets with  $p_T > 20$  and  $|\eta| < 2.4(2.5)$  in 2016 (2017 and 2018) are considered b-tagged if their discriminator value is larger than 0.3093, 0.3033 or 0.2770 in the 2016, 2017 or 2018 analysis, respectively. This corresponds to the medium working point provided by the BTV POG [?, ?, ?].

### 3.7 Missing transverse energy

The missing transverse energy in the event is reconstructed using the PUPPI algorithm [?], which is more robust towards pileup compared to PF MET, which tends to increase for Run 2 analyses. This algorithm is an extension of the PFCHS algorithm and assigns weights to PF candidates based on the probability that they come from pileup or the primary vertex. Type-I corrections are applied to correct the MET estimated in this way for over- or underestimation due to inefficiencies in the detector. These corrections entail the propagation of the JECs to the MET as

$$\vec{E}_T^{\text{miss,corr}} = \vec{E}_T^{\text{miss}} - \sum_{\text{jets}} (\vec{p}_T^{\text{corr}} - \vec{p}_T). \quad (3.4)$$

### 3.8 Taus

Also fundamental to this analysis is the identification of tau particles. The tau lepton is measured to have a mean lifetime of  $2.9 \times 10^{-13}$ s. This short lifetimes means that the tau lepton is not directly observable in the CMS detector. In order to detect these particles, it is important to understand how the tau decays. Due to the heavy nature of the particle, it does not only decay leptonically, but unlike the muon, it can also decay hadronically. A list of prominent decays of the tau lepton are shown in the table below.

Decay Mode	Branching Fraction
<b>Leptonic Decay (<math>e, \mu</math>)</b>	<b>35.2%</b>
$e^- \bar{\nu}_e \nu_\tau$	17.8%
$\mu^- \bar{\nu}_\mu \nu_\tau$	17.4%
<b>Hadronic Decay (<math>\tau_h</math>)</b>	<b>64.8%</b>
$h^- \pi^0 \nu_\tau$	25.9%
$h^- \nu_\tau$	11.5%
$h^- 2\pi^0 \nu_\tau$	9.3%
$\pi^- \pi^- \pi^+ \nu_\tau$	9.0%
$\pi^- \pi^- \pi^+ \pi^0 \nu_\tau$	2.7%
other	6.4%

Table 3.2: Measured branching fractions, that are greater than 2%, for the tau lepton. h represents a charged hadron either a pion or a kaon.

These decays can be split into three groups: the 17.8% of taus that decay to an electron ( $e$ ), the 17.4% that decay into a muon ( $\mu$ ) and hadronic tau decays ( $\tau_h$ ) that make up the final 64.8% of tau decays. The leptonic decays of the tau can be accounted for by the identification of electrons and muons as discussed in the previous subsection. The hadron-plus-strips (HPS) algorithm is used to identify hadronic taus [41, 42]. This algorithm groups electrons, positrons and photons and names this cluster as a "strip". This is defined to represent the decay products of the  $\pi^0$  meson. The strip size is variable depending on the  $p_T$  of its components. In a jet, the number of strips and charged particles are counted. If the numbers are corresponding to the number of  $\pi^0$  mesons and charged hadrons shown in Table 3.2 for hadronic decays, it is concluded that the jet may originate from a tau lepton.

Hadronic tau decays are reconstructed using the `hadrons plus strips` (HPS) algorithm [?, 43], which uses particle flow candidates to reconstruct and identify hadronic tau decays. In the reconstruction step, particle-flow charged and neutral particles are grouped into combinations compatible with specific hadronic tau decay modes, and the 4-momentum of the candidates is computed. In the identification step, discriminators that separate hadronic taus from quark and gluon jets and from electrons and muons are evaluated. Taus in the hadrons plus strips algorithm are seeded by jets clustered with the anti- $k_T$  algorithm with a distance parameter  $\Delta R = 0.4$ . To reconstruct the energy deposits  $\pi^0$  candidates leave in the ECAL, photon and

electron constituents of the jet that seeds the  $\tau$  reconstruction are clustered into strips. The  $e$  or  $\gamma$  with highest  $p_T$  that is not yet included in a strip is used to build a new strip. The  $\eta$  and  $\phi$  of this candidate determine the initial position of the strip, the next highest  $p_T$   $e$  or  $\gamma$  within an  $\eta \times \phi$  window centred on the strip location is added to the strip and the position is recomputed as the energy-weighted average of the electron/photon constituents in the strip. This procedure is repeated until there are no more electrons or photons with  $p_T > 0.5$  GeV within the strip window. The  $\Delta\eta$  and  $\Delta\phi$  of the strip are varied based on the  $p_T$  or  $E_T$  to be added to the strip and on the energy the strip already has, as

$$\begin{aligned}\Delta\eta &= f(p_T^{e/\gamma}) + f(p_T^{\text{strip}}) \\ \Delta\phi &= g(p_T^{e/\gamma}) + g(p_T^{\text{strip}}),\end{aligned}\tag{3.5}$$

where  $p_T^{e/\gamma}$  is the transverse momentum of the candidate to be added to the strip and  $p_T^{\text{strip}}$  is the transverse momentum of the strip before merging a new candidate in. In addition, the strip size is bounded as  $0.05 < \Delta\eta < 0.15$ ,  $0.05\Delta\phi < 0.3$ .

The functions  $f(p_T)$  and  $g(p_T)$  are defined as

$$\begin{aligned}f(p_T) &= 0.2 \cdot p_T^{-0.66} \\ g(p_T) &= 0.35 \cdot p_T^{-0.71}.\end{aligned}\tag{3.6}$$

If the  $\Sigma p_T$  of the strip is at least 2.5 GeV, it is considered as a  $\pi^0$  candidate. To reconstruct hadronic taus, charged particles and strips are combined into different signatures which are compatible with a certain decay mode if the set of requirements listed below is satisfied. If a candidate satisfies more than one of the hypotheses, the one that maximises the  $p_T$  is retained.

The decay modes (with `decayModeFindingNewDMs`) considered for reconstructing taus are:

- **One prong, 0  $\pi^0$**  : One charged particle, no strips.
- **One prong, 1  $\pi^0$**  : One charged particle + one strip with mass  $0.3 < m_\tau < 1.3\sqrt{p_T/100}$  GeV. The mass window upper limit is constrained to lie between 1.3 and 4.2 GeV.
- **One prong, 2  $\pi^0$**  : One charged particle + two strips. The  $\tau_h$  mass should be  $0.4 < m_\tau < 1.2\sqrt{p_T/100}$  GeV. The upper limit on the mass window is constrained to lie between 1.2 and 4.0 GeV.

- **Two prong, 0  $\pi^0$**  : Two charged particle + no strips.
- **Two prong, 1  $\pi^0$**  : Two charged particle + one strips.
- **Two prong, 2  $\pi^0$**  : Two charged particle + two strips.
- **Three prong, 0  $\pi^0$** : Three charged particles with mass  $0.8 < m_\tau < 1.5$  GeV.  
The tracks are required to originate within  $\Delta z < 0.4$  cm of the same vertex.
- **Three prong, 1  $\pi^0$** : Three charged particles + one strip.

Requiring the reconstructed  $\tau_h$  to be isolated reduces the jet $\rightarrow\tau_h$  fake rate. Both a cut-based discriminator and MVA-based ones are available, in this analysis we make use of the deepTau discriminator [44].

The deepTauv2p1 is a DNN-based algorithm that allows to define three quantities for discrimination of tau vs jets, tau vs electrons and tau vs muons [?], where the anti-lepton discriminators provide reduction in  $e \rightarrow \tau_h$  and  $\mu \rightarrow \tau_h$  fake rates.

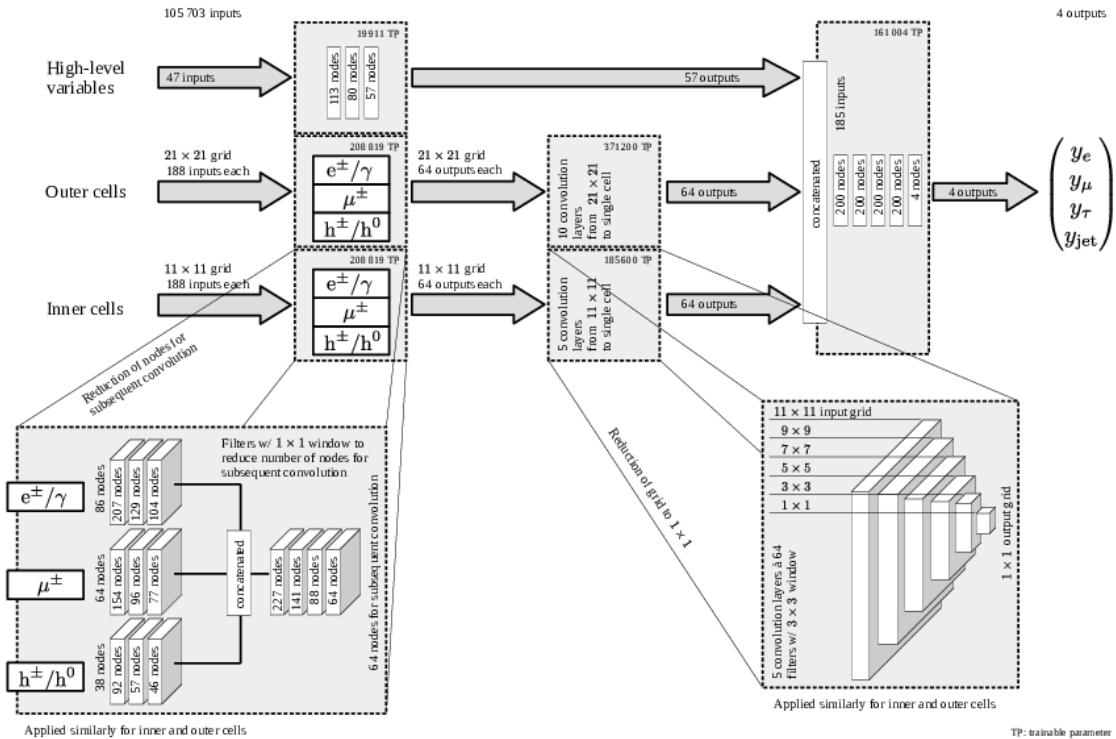


Figure 3.1

# Chapter 4

## Searches for New Physics in $\tau^+\tau^-$ Final States

The  $\tau^+\tau^-$  final states are a powerful tool to search for new physics at collider experiments. As the heaviest lepton,  $\tau$  particles are sensitive to resonant production of new neutral particles where the couplings have mass hierarchy. They are also sensitive to non-resonant effects from new physics mediators. This chapter will detail the searches for two such areas of new physics: additional Higgs bosons and vector leptoquarks. These searches are split up into three sections:

- i) A model independent search for single narrow spin-0 resonance ( $\phi$ ), produced via gluon fusion ( $gg\phi$ ) or in association with a b quark ( $bb\phi$ ). The SM Higgs boson is treated as a background and the Yukawa couplings of the spin-0 resonance that contribute to the gluon fusion loop are set to SM values.
- ii) A search for the MSSM Higgs sector in a number of benchmark scenarios. The benchmark scenarios were proposed in Refs. [45, 46, 47] and summarised in Ref. [48]. The  $M_h^{125}$  and  $M_{h,\text{EFT}}^{125}$  scenarios are shown in this chapter. The production of the observed Higgs boson particle at 125 GeV is also used to constrain the available phase space.
- iii) A search for the t-channel exchange of a  $U_1$  vector leptoquark. Two scenarios are taken, motivated by the best fit to the b anomalies. These scenarios are detailed in Section 4.1.2.

These searches are performed with the full run-2 dataset ( $138 \text{ fb}^{-1}$ ) collected by the CMS experiment. The search for additional Higgs bosons had previously been

performed with data collected in 2016 ( $39 \text{ fb}^{-1}$ ) and results were consistent with the SM background prediction [49].

## 4.1 Signal Modelling

### 4.1.1 Additional Higgs Bosons

Extended Higgs sectors, such as that of the MSSM, can be probed by direct searches for the additional bosons and further precise measurements of the SM Higgs boson. This search for an extended Higgs sector is motivated by Type II 2HDMs, such as the MSSM. In these models  $\tan\beta$  enhances couplings of additional Higgs bosons to down-type quarks and leptons, whilst up-type quark couplings are suppressed. This divides the dominant production modes of the Higgs boson into two categories: Gluon fusion and production in association with a bottom quark. Examples of the Feynman diagrams for such processes are shown in Figure 4.1.

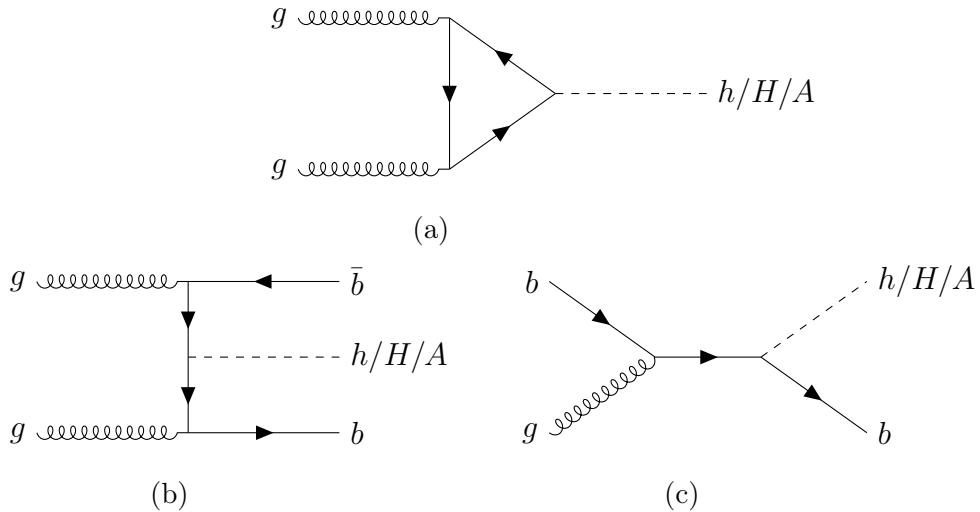


Figure 4.1: Diagram (a) shows the production of neutral Higgs bosons from gluon fusion. The dominant loop contributions to this diagram are from top-only, bottom-only and top-bottom interference. Diagrams (b) and (c) show production in association with  $b$  quarks.

With the  $\tan\beta$  enhancement, the decays of additional Higgs bosons to  $\tau$  leptons and  $b$  quarks are most likely.  $\tau$  leptons are identified with a higher purity than bottom quarks at the CMS detector. It is also easier to separate  $\tau^+\tau^-$  from the large QCD

multiparticle background produced from the high energy proton-proton collisions. This hypothesis was tested with the 2016 dataset and although no deviations were observed, the strongest limits on the MSSM phase space was placed by the  $\tau^+\tau^-$  final states [49, 50].

For this analysis, signal templates for the production of additional Higgs bosons over a mass range of 60 GeV to 3.5 TeV are generated. Gluon fusion is simulated at NLO precision using the 2HDM implementation of PowHEG 2.0 [51, 52, 53, 54]. The kinematic properties are highly dependent on the contributions to the loop, that are dependent on the specific signal model. To account for the different loop contributions at the NLO plus parton shower prediction, weights based off the  $p_T$  spectra are calculated to split the contributions from the t quark only, b quark only, and tb-interference. Once individual templates have been determined for each contribution to the loop, the 2HDM samples can be scaled to the MSSM scenario prediction with the following formula,

$$\frac{d\sigma_{\text{MSSM}}}{dp_T} = \left( \frac{Y_{t,\text{MSSM}}}{Y_{t,2\text{HDM}}} \right)^2 \frac{d\sigma_{2\text{HDM}}^t}{dp_T}(Q_t) + \left( \frac{Y_{b,\text{MSSM}}}{Y_{b,2\text{HDM}}} \right)^2 \frac{d\sigma_{2\text{HDM}}^b}{dp_T}(Q_b) \\ + \left( \frac{Y_{t,\text{MSSM}}}{Y_{t,2\text{HDM}}} \frac{Y_{b,\text{MSSM}}}{Y_{b,2\text{HDM}}} \right) \left\{ \frac{d\sigma_{2\text{HDM}}^{t+b}}{dp_T}(Q_{tb}) - \frac{d\sigma_{2\text{HDM}}^t}{dp_T}(Q_{tb}) - \frac{d\sigma_{2\text{HDM}}^b}{dp_T}(Q_{tb}) \right\}, \quad (4.1)$$

where  $Q_i$  are resummation scales that depend on the mass of the additional Higgs boson. Further contributions from any Supersymmetric partners were checked and account for less than a few percent and so are neglected. The  $p_T$  reweighting is done separately for the scalar and pseudoscalar additional Higgs bosons, as the  $p_T$  distributions can differ. The benchmark scenarios provide the relative Yukawa couplings (to calculate the cross sections) and branching fractions of the MSSM Higgs bosons. An example of the distributions for gluon fusion production, in the MSSM  $M_h^{125}$  scenario with  $m_A = 1600$  GeV and  $\tan\beta$  varying, is shown in Figure 4.2. The distributions peak at a higher  $p_T$  for the top quark loop, therefore at smaller  $\tan\beta$ , where the top quark contribution is dominant, an additional Higgs boson would be more boosted.

Production in association with bottom quarks is simulated at NLO precision using the corresponding PowHEG 2.0 implementation in the four-flavour scheme [51, 52, 53, 54]. All additional Higgs boson signal generation is performed using the parton

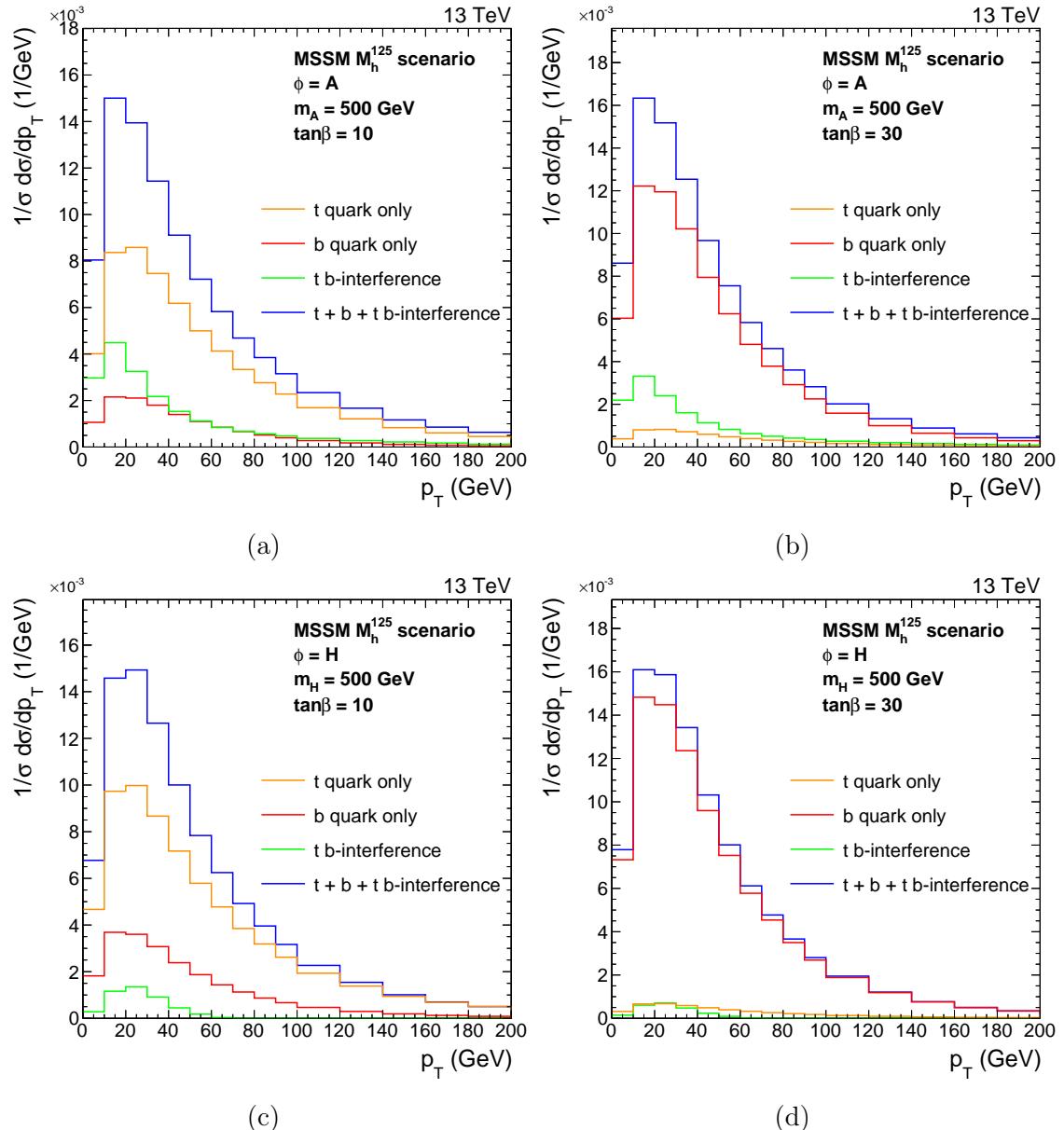


Figure 4.2:  $p_T$  density distributions of the A (top) and H (bottom) boson, with contributions to the gluon fusion loop displayed individually and summed. These are shown for  $\tan\beta$  values of 10 (left) and 30 (right) where  $m_A = 500 \text{ GeV}$  in the MSSM  $M_h^{125}$  scenario.

distribution function (PDF) NNPDF3.1 [55, 56].  $\tau$  lepton decay, parton showering and hadronisation are all modelled with the PYTHIA event generator where the PU profile is matched to data [57, 58]. All events generated are passed through a GEANT4-based [59] simulation of the CMS detector and reconstructed in the same way as data.

The model-dependent search for the MSSM also looks to find differences from the observed SM Higgs boson and the predicted MSSM SM-like Higgs boson. In each MSSM benchmark scenario, an uncertainty of  $\pm 3$  GeV is given on the prediction for the SM Higgs boson mass. This uncertainty is to reflect the contribution from any unknown higher-order corrections. The value of the mass is allowed to vary within this window, however the Yukawa couplings are rescaled the observed mass.

#### 4.1.2 Vector Leptoquarks

The best fit to the b anomalies in the available phase space for vector leptoquarks yielded large b quark and  $\tau$  lepton couplings to the  $U_1$  particle. The possible production modes of a  $\tau^+\tau^-$  final state are shown in Figure 4.3.

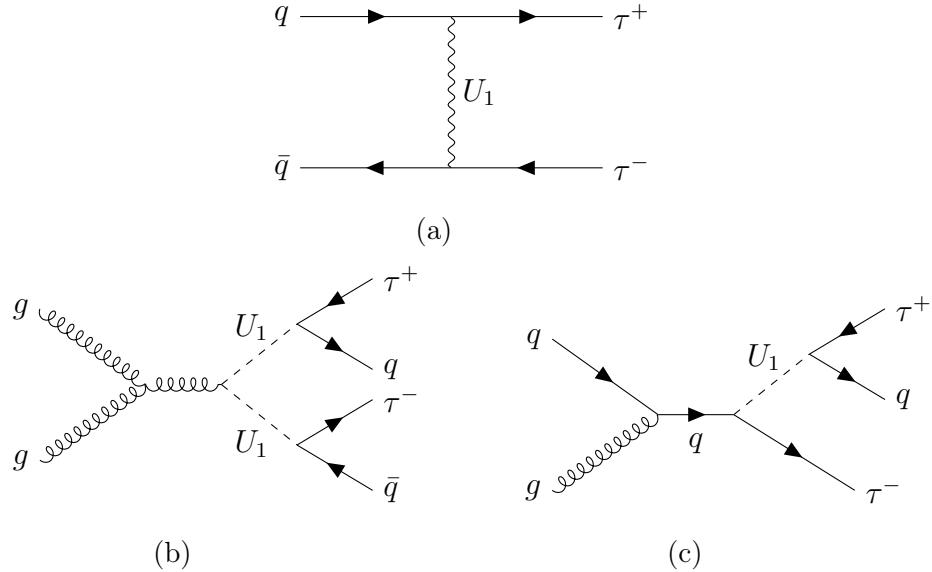


Figure 4.3: Feynman diagrams showing the contribution from  $U_1$  vector leptoquarks to the final state with a pair of oppositely charged  $\tau$  leptons. Diagram (a) shows t-channel, (b) pair and (c) single production of a vector leptoquark.

Pair and single production of a vector leptoquark are dependent on its strong coupling, which is highly model dependent. For large mass,  $m_U$ , the probability of producing an on-shell  $U_1$  singlet or pair is heavily suppressed due to the momentum of the initial partons. These production processes are not discussed further in this search. Further studies have been used to search for single and pair production at the CMS experiment and no statistically significance derivation was observed [60]. Single production places the loosest constraints out of the processes mentioned whilst pair production puts a lower limit on the leptoquark mass, as the process is approximately independent on  $g_U$ , and this limit is heavily dependent on the values taken for the strong coupling.

The t-channel process contain two vertices with a  $U_1$  vector leptoquark, a quark and a  $\tau$  lepton, and hence the cross section will scale with  $g_U^4$ . From the best fit to b anomalies, the vertex is dominated by the b quark and hence the initial state will be mostly from  $b\bar{b}$ , with sub-dominant contributions from  $b\bar{s}$ ,  $s\bar{b}$  and  $s\bar{s}$ . Although there are no additional b quarks in the final state in the LO process, initial state radiation can lead to additional b quarks in the final state. In this search the two scenarios discussed in Section 1.4.1 are considered. The only non negligible freely floating parameter in each fit, for  $\tau^+\tau^-$  final states in the  $m_U$ - $g_U$  phase space is the  $\beta_L^{s\tau}$  parameter. This is set to the best fit value.

The signal process of the  $U_1$  t-channel exchange is simulated in the five-flavour scheme (5FS) at LO precision using the MADGRAPH5\_amc@NLO v2.6.5 event generator [61]. Events are generated with one or fewer outgoing partons from the matrix element and the MLM prescription [62] is then used for matching, with a scale set to 40 GeV. Negligible dependence of the  $U_1$  decay width is observed, for simulation this is chosen to approximately match the value predicted by the b anomaly fit. Samples with a mass between 1 and 5 TeV at  $g_U = 1$  are generated.

The interference between the  $U_1$  signal and  $Z/\gamma^* \rightarrow \tau\tau$  production was checked and a large destructive affect was observed, with magnitude dependent on  $g_U$ . To account for this, separate samples are produced for this interference, generated in the same way as the t-channel exchange. The interference samples are then split into two with a di- $\tau$  mass split in order to have a sufficient number of events in the high di- $\tau$  mass regions. The cross section of these interference samples scale with  $g_U^2$ . Examples of the generator level di- $\tau$  mass distributions are shown in Figure 4.4.

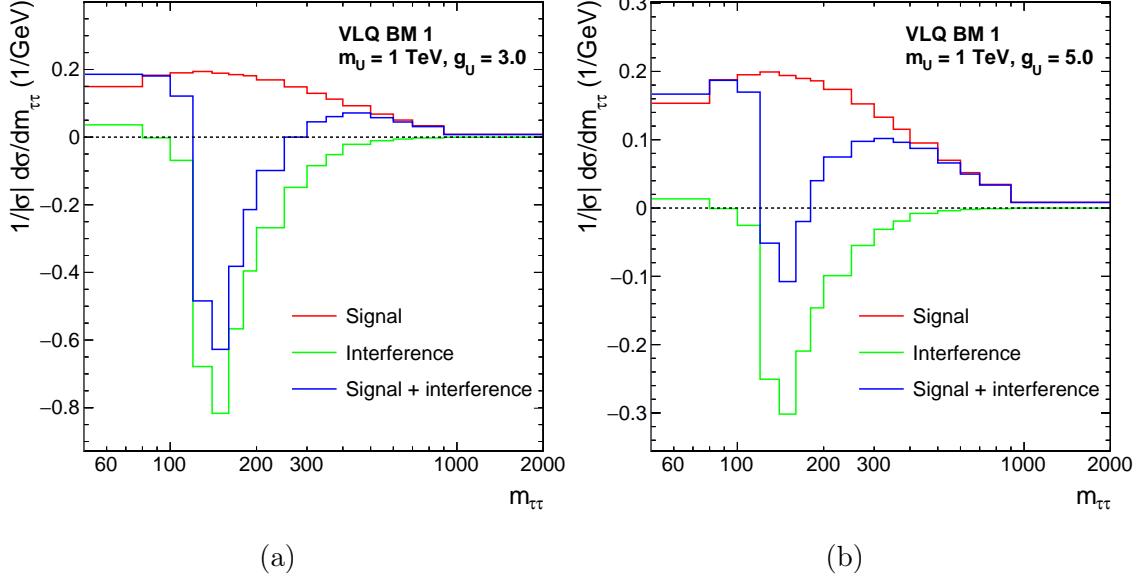


Figure 4.4: The generator level  $m_{\tau\tau}$  density distributions of the t-channel vector leptoquark signal and the interference with Drell-Yan. This is shown in the VLQ BM 1 scenario for a leptoquark of mass 1 TeV for coupling strengths of  $g_U = 3$  (a) and  $g_U = 5$  (b).

The t-channel signal produces a broad distribution in  $m_{\tau\tau}$  due to its non-resonant nature. The interference is mostly a destructive effect (except for at small  $m_{\tau\tau}$ ), with the yield becoming less negative at higher  $m_{\tau\tau}$ . The interference peaks negatively between 100 and 200 GeV and in this region the combined yield can be negative. Due to the difference in scaling of the two effects, at small  $g_U$  the interference is more dominant than the signal and hence the yield of the combined result is reduced.

## 4.2 Event Selection

The possible decays of two  $\tau$  leptons and their branching fractions, where the  $\tau$  decay is grouped into three categories e,  $\mu$  and  $\tau_h$  as defined in Section 3, are shown in Table 4.1. For this search the four largest branching fraction channels used:  $\tau_h\tau_h$ ,  $e\tau_h$ ,  $\mu\tau_h$  and  $e\mu$ . This accounts for approximately 94% of di- $\tau$  events. The two same lepton channels are neglected due to small branching ratio and the dominating  $Z \rightarrow ee$  and  $Z \rightarrow \mu\mu$  backgrounds.

Channel	Branching Fraction
$\tau_h\tau_h$	42.0%
$e\tau_h$	23.1%
$\mu\tau_h$	22.6%
$e\mu$	6.2%
$ee$	3.2%
$\mu\mu$	3.0%

Table 4.1: Branching fractions of the decays of two  $\tau$  leptons.

#### 4.2.1 Trigger Requirements

In the four final state pairs, a number of different online trigger requirements are needed. In the  $\tau_h\tau_h$  channel, two possible triggers are available: the double- $\tau_h$  and single- $\tau_h$  triggers. The single- $\tau_h$  trigger has a high  $p_T$  threshold at 120 (180) GeV for events recorded in 2016 (2017-2018), whilst the double- $\tau_h$  has a  $p_T$  threshold at 40 GeV. Therefore, the double- $\tau_h$  trigger is used individually where the  $\tau_h$  has  $p_T$  below the single- $\tau_h$  threshold and the union of single- $\tau_h$  and double- $\tau_h$  triggered events are taken above the threshold.

In the  $e\tau_h$  and  $\mu\tau_h$  channels, there are three possible triggers available: the single- $e/\mu$ , single- $\tau_h$  and the  $e/\mu-\tau_h$  cross-trigger. The cross-trigger is used for events where the light lepton has  $p_T$  between the thresholds for the cross-trigger and single- $e/\mu$  shown in Table 4.2. The light lepton used in the cross-trigger is required to be in the central barrel of the detector within  $|\eta| < 2.1$ . Above these light lepton  $p_T$  thresholds the single- $e/\mu$  trigger is used, where it is required that the  $\tau_h$  has  $p_T > 30$  GeV. At  $\tau_h$   $p_T$  above the single- $\tau_h$  thresholds, the single- $\tau_h$  trigger is used in combination with the single- $e/\mu$  trigger.

Year/ Trigger	e- $\tau_h$ cross-trigger	single- $e$	$\mu-\tau_h$ cross-trigger	single- $\mu$
2016	23	26	20	23
2017	25	28	20	25
2018	25	33	21	25

Table 4.2: Lower trigger light lepton thresholds  $p_T$  in GeV for the  $e\tau_h$  and  $\mu\tau_h$  channels.

In the  $e\mu$  channel, there are three possible triggers available: the single- $e$ , single- $\mu$  and the  $e\text{-}\mu$  cross-trigger. However, only the cross-trigger is used in this analysis, due to the larger efficiencies of correctly selecting light leptons. The  $e$  and  $\mu$  are required to have  $p_T > 15$  GeV and  $|\eta| < 2.4$ .

### 4.2.2 Offline Requirements

All offline selections stated are in addition the object selection discussed in Section 3. In this analysis,  $\tau_h$  candidates are required to pass the **Medium**  $D_{\text{jet}}^{\text{WP}}$ .  $D_e^{\text{WP}}$  and  $D_\mu^{\text{WP}}$  are dependent on the decay channel. The **VVLoose**, **Tight**, **VVLoose**  $D_e^{\text{WP}}$  and the **VLoose**, **VLoose** and **Tight**  $D_\mu^{\text{WP}}$  are used in the  $\tau_h\tau_h$ ,  $e\tau_h$  and  $\mu\tau_h$  channels respectively. The tighter working point for the same light lepton discrimination as tagged in the event is used to remove light leptons misidentified as  $\tau_h$  from the  $Z \rightarrow ll$  process. The light lepton isolation requirement is  $I_{\text{rel}}^{e/\mu} < 0.15$  except for in the  $e\mu$  channel where the muon is required to have  $I_{\text{rel}}^\mu < 0.2$ .

The selected  $\tau$  lepton decay candidates are required to have opposite charge and to be separated by more than  $\Delta R > 0.5$  in all channels except  $e\mu$  where  $\Delta R > 0.3$ . In events where the numbers of an object in the event is greater than the required number of objects in the  $\tau\tau$  decay channel, the objects are sorted by the maximum  $D_{\text{jet}}^{\text{score}}$  if a  $\tau_h$  candidate or minimum  $I_{\text{rel}}$  if a light lepton candidate and the leading objects are chosen. In order to maintain orthogonality between channels, events with additional light leptons passing looser selections than the nominal requirements, are rejected from the selection. The looser selections help to suppress the  $Z \rightarrow ll$  background process further.

In the  $e\tau_h$  and  $\mu\tau_h$  channels, a cut is placed at 70 GeV on the transverse mass between the light lepton  $\vec{p}_T$  and the missing  $\vec{p}_T$ , where the transverse momentum is defined as,

$$m_T(\vec{p}_T^i, \vec{p}_T^j) = \sqrt{2p_T^i p_T^j (1 - \cos \Delta\phi)}, \quad (4.2)$$

where  $\Delta\phi$  to the azimuthal angle between  $\vec{p}_T^i$  and  $\vec{p}_T^j$ . The variable is used to remove  $W + \text{jets}$  background events, where a jet is misidentified as a  $\tau_h$  and the MET and light lepton from the  $W$  decay are aligned and hence the event has a large  $m_T(\vec{p}_T^{e/\mu}, \vec{p}_T^{\text{MET}})$ .

In the  $e\mu$  channel, an additional cut is placed on the variable  $D_\zeta$ , which is defined

as,

$$D_\zeta = p_\zeta^{\text{miss}} - 0.85 p_\zeta^{\text{vis}}, \quad p_\zeta^{\text{miss}} = \vec{p}_T^{\text{miss}} \cdot \hat{\zeta}; \quad p_\zeta^{\text{vis}} = (\vec{p}_T^e + \vec{p}_T^\mu) \cdot \hat{\zeta} \quad (4.3)$$

where  $\vec{p}_T^{e/\mu}$  corresponds to the transverse momentum vector of the electron or muon and  $\hat{\zeta}$  to the bisectional direction between the electron and the muon in the transverse plane. A diagram of the inputs is shown Figure 4.5.

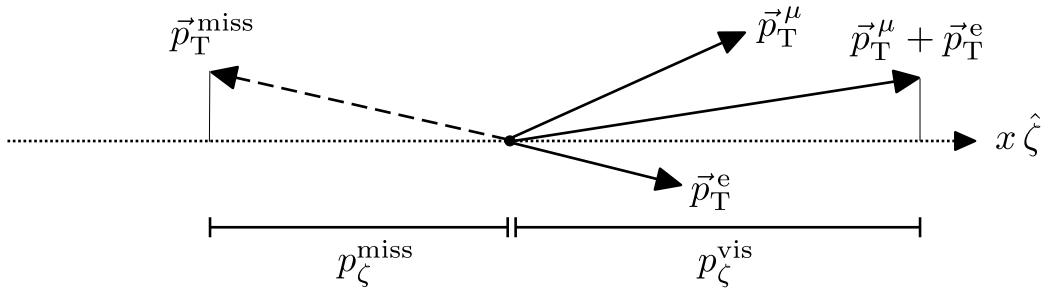


Figure 4.5: Diagram of inputs to the  $D_\zeta$  variable.

The linear combination is optimised for genuine di- $\tau$  events to peak around  $D_\zeta = 0$  GeV. It is motivated by the expectation that in di- $\tau$  decays from a resonance, the visible and missing (from  $\tau$  neutrinos) momenta are roughly aligned and of similar magnitudes. In W + jets and  $t\bar{t}$  events the directions of the visible and missing products are expected to be more randomly distributed and lead to a non-peaking  $D_\zeta$ . Therefore, only events with  $D_\zeta > -35$  GeV are considered for signal events. Events with no b tag with this cut are vetoed and events with a b tag also with this cut are used for a  $t\bar{t}$  control region and discussed further in Section 4.4.

### 4.3 Search Optimisation

The optimisation of the signal extraction depends on which of the three scenarios, set out at beginning of this section, is searched for. The components of the optimisation are named the high-mass, low-mass and SM Higgs optimisation procedures. For the model independent search (i) the high- or low-mass optimisation procedures are used depending on whether the mass of the resonance is greater or less than 250 GeV. The search for the MSSM Higgs sector (ii) uses the high mass or the SM Higgs optimisation procedures depending on whether the reconstructed di- $\tau$  mass is greater or less than 250 GeV. Due to more than one neutrino in the event, it is

not possible to fully reconstruct the di- $\tau$  mass. To resolve this issue a likelihood-based method named SVFit algorithm is used [63]. Finally, the search for vector leptoquarks (iii) uses only the high mass optimisation procedure. The procedures are discussed in detail below.

The high-mass optimisation procedure follows what was done in Ref. [49]. Each event is initially split into categories depending on whether it has 0 or  $\geq 1$  b tagged jets in the event. Firstly, this helps target the additional Higgs boson production modes gluon fusion and b associated production respectively. In particular,  $b\bar{b}\phi$  has final state b quarks that if tagged can significantly aid the sensitivity of this signal. Secondly, the initial state radiation of a t-channel vector leptoquark signal dominated by initial states of b quarks, can lead to additional b jets in the final state. The reduced backgrounds in b tagged events allows for a more sensitive vector leptoquark search in this category.

The  $e\tau_h$  and  $\mu\tau_h$  channels are further subdivided into categories depending on the transverse mass between the light lepton and missing transverse momentum vectors as defined in Equation 4.2. The corresponding categories are defined as:

- **Tight- $m_T$ :**  $m_T(\vec{p}_T^{e/\mu}, \vec{p}_T^{\text{miss}}) < 40$  GeV;
- **Loose- $m_T$ :**  $40 \leq m_T(\vec{p}_T^{e/\mu}, \vec{p}_T^{\text{miss}}) < 70$  GeV.

The majority of the signal events fall within the **Tight- $m_T$**  sub-category. The **Loose- $m_T$**  category is used to improve the signal acceptance for resonant masses of  $m_\phi > 700$  GeV.

In the  $e\mu$  channel, is subdivided into three signal categories based on the variable  $D_\zeta$  as defined in Equation 4.3. The three categories are defined as:

- **Low- $D_\zeta$ :**  $-35 \leq D_\zeta < -10$  GeV;
- **Medium- $D_\zeta$ :**  $-10 \leq D_\zeta[\text{GeV}] < 30$  GeV;
- **High- $D_\zeta$ :**  $D_\zeta[\text{GeV}] \geq 30$  GeV.

By design, the majority of signal events are located in the **Medium- $D_\zeta$**  sub-category. The **Low-** and **High- $D_\zeta$**  categories are used to catch the tail of the signal distributions. A schematic of all high mass optimisation categories are shown in Figure 4.6.

Once all category divisions have been applied, events are drawn in histograms based off a discriminating variable. The discriminating variable used in this analysis is  $m_T^{\text{tot}}$  and is defined below.

$$m_T^{\text{tot}} = \sqrt{m_T(\vec{p}_T^{\tau_1}, \vec{p}_T^{\text{miss}})^2 + m_T(\vec{p}_T^{\tau_2}, \vec{p}_T^{\text{miss}})^2 + m_T(\vec{p}_T^{\tau_1}, \vec{p}_T^{\tau_2})^2}, \quad (4.4)$$

where  $\tau_1$  and  $\tau_2$  refer to the visible products of the two  $\tau$  lepton decays. This variable provides excellent discriminating power between higher mass resonant signals compared to other non-peaking backgrounds, whilst still maintaining some separation between signal masses. It is also excellent at separating the high-mass non-resonant di- $\tau$  signatures where a di- $\tau$  mass is unphysical for the signal. This is due to the use of the transverse momenta and MET in the variable definition. For a t-channel signal where the mediator has high mass, no significant mass separation is expected in any variable.

The low-mass optimisation procedure loosely follows the high-mass procedure with a few key difference. Firstly, categories that are only sensitive to high-mass signals are dropped. This includes the **Low- $D_\zeta$**  and **Loose- $m_T$**  categories. Each no b tag subcategory is further divided into four bins of reconstructed di- $\tau$  visible  $p_T$  with bin edges: 0,50,100,200 and  $\infty$ . This is not done in the b tag subcategories due to the lack of statistics in this region. A schematic of the categories used in the low mass optimisation procedure is shown in Figure 4.7. The final difference with the high-mass optimisation procedure is the discriminator used. In the low-mass optimisation procedure the reconstructed di- $\tau$  mass is used. This helps to separate signal events from the Z boson peak in this region.

Finally, the SM Higgs optimisation procedure is taken from the CMS SM  $H \rightarrow \tau\tau$  analysis and is detailed in Ref. [64]. This was previously used for simplified template cross section measurements. This uses a neural-network-based (NN) categorisation to obtain the most precise estimates from data of the SM Higgs produced via gluon fusion, vector boson fusion or vector boson associated production. The NN based analysis introduces 26 categories, 8 of which are optimised to pull out the Higgs boson signal. Although the NN is trained specifically to target events with an SM-like Higgs boson, signal events with differing masses can also enter the NN categories.

	No b tag			b tag		
e $\mu$	Low- $D_\zeta$	Medium- $D_\zeta$	High- $D_\zeta$	Low- $D_\zeta$	Medium- $D_\zeta$	High- $D_\zeta$
e $\tau_h$	Loose- $m_T$	Tight- $m_T$		Loose- $m_T$	Tight- $m_T$	
$\mu\tau_h$	Loose- $m_T$	Tight- $m_T$		Loose- $m_T$	Tight- $m_T$	
$\tau_h\tau_h$						
$t\bar{t}(e\mu)$						
	Signal region (SR)			$D_\zeta < -35 \text{ GeV}$		
	Control region					

Figure 4.6: Overview of the categories used for the extraction of the signal in the high mass optimisation procedure.

	No b tag			b tag		
e $\mu$	Medium- $D_\zeta$ $p_T^{\tau\tau} < 50 \text{ GeV}$	High- $D_\zeta$ $p_T^{\tau\tau} < 50 \text{ GeV}$		Medium- $D_\zeta$	High- $D_\zeta$	
	50 < $p_T^{\tau\tau} < 100 \text{ GeV}$	50 < $p_T^{\tau\tau} < 100 \text{ GeV}$				
	100 < $p_T^{\tau\tau} < 200 \text{ GeV}$	100 < $p_T^{\tau\tau} < 200 \text{ GeV}$				
	$p_T^{\tau\tau} > 200 \text{ GeV}$	$p_T^{\tau\tau} > 200 \text{ GeV}$				
e $\tau_h$	Tight- $m_T$ $p_T^{\tau\tau} < 50 \text{ GeV}$	Tight- $m_T$ $p_T^{\tau\tau} < 50 \text{ GeV}$		Tight- $m_T$		
	50 < $p_T^{\tau\tau} < 100 \text{ GeV}$	50 < $p_T^{\tau\tau} < 100 \text{ GeV}$				
	100 < $p_T^{\tau\tau} < 200 \text{ GeV}$	100 < $p_T^{\tau\tau} < 200 \text{ GeV}$				
	$p_T^{\tau\tau} > 200 \text{ GeV}$	$p_T^{\tau\tau} > 200 \text{ GeV}$				
$\mu\tau_h$	Tight- $m_T$ $p_T^{\tau\tau} < 50 \text{ GeV}$	Tight- $m_T$ $p_T^{\tau\tau} < 50 \text{ GeV}$		Tight- $m_T$		
	50 < $p_T^{\tau\tau} < 100 \text{ GeV}$	50 < $p_T^{\tau\tau} < 100 \text{ GeV}$				
	100 < $p_T^{\tau\tau} < 200 \text{ GeV}$	100 < $p_T^{\tau\tau} < 200 \text{ GeV}$				
	$p_T^{\tau\tau} > 200 \text{ GeV}$	$p_T^{\tau\tau} > 200 \text{ GeV}$				
$\tau_h\tau_h$						
$t\bar{t}(e\mu)$						
	Signal region (SR)			$D_\zeta < -35 \text{ GeV}$		
	Control region					

Figure 4.7: Overview of the categories used for the extraction of the signal in the low mass optimisation procedure.

## 4.4 Background Modelling Overview

The analysis considers several backgrounds including Drell-Yan,  $t\bar{t}$ , W+jets, QCD, di-boson, single-top, and electroweak W and Z bosons production. These are split into a five categories:

- i) Events containing only genuine  $\tau$  leptons.
- ii) Events with a jet misidentified as a  $\tau_h$  ( $\text{jet} \rightarrow \tau_h$ ) in the  $e\tau_h$ ,  $\mu\tau_h$  or  $\tau_h\tau_h$  channels.
- iii) Events with jets faking both light leptons ( $\text{jet} \rightarrow l$ ) in the  $e\mu$  channel.
- iv) Events from  $t\bar{t}$  with a prompt light lepton ( $e$  or  $\mu$  not from a  $\tau$  decay) and the other object (if there are not two prompt light lepton) is from a genuine  $\tau$  leptons.
- v) Other events. This is a small contribution and hence why it is grouped.
  - Non  $t\bar{t}$  events with a prompt light lepton ( $e$  or  $\mu$  not from a  $\tau$  decay) and the other object (if there are not two prompt light lepton) is from a genuine  $\tau$  leptons.
  - Events with a light lepton misidentified as a  $\tau_h$  and the other object (if there are not two light leptons faking a  $\tau_h$ ) is reconstructed as prompt light lepton or from genuine  $\tau$  leptons.
  - Events with a jet misidentified as a light lepton and the other object is from genuine  $\tau$  leptons in the  $e\tau_h$ ,  $\mu\tau_h$  or  $\tau_h\tau_h$  channels.
  - Events with a muon misidentified as an electron in the  $e\mu$  channel.
  - Events with one jet misidentified as a light lepton and the other object from a prompt light lepton in the  $e\mu$  channel.

Backgrounds from (i) consists of largely  $Z/\gamma^* \rightarrow \tau\tau$  events but there are also smaller contributions from other processes. This background is modelled by a data-simulation hybrid method called the embedding method and this is described in detail in Section 4.6. Group (ii) is dominated by QCD, W + jets and  $t\bar{t}$  events with a  $\text{jet} \rightarrow \tau_h$  misidentification. This is modelled from data by the fake factor method ( $F_F$ ) and is explained in Section 4.7. Group (iii) is modelled from data to describe the QCD multijet contribution to the background in the  $e\mu$  channel. The method to obtain this background is described in Section 4.5. The data driven

background estimations for (i), (ii) and (iii) contribute  $\approx 98\%$  of all expected background events in the  $\tau_h\tau_h$  channel,  $\approx 90\%$  in  $e\tau_h$  and  $\mu\tau_h$  channels and  $\approx 50\%$  in the  $e\mu$  channel. The final groups, (iv) and (v), are modelled with MC. The  $t\bar{t}$  process is separated due to its large contribution to the phase space where a b jet is required.

In 2016 (2017–2018), the  $W + \text{jets}$  and  $Z \rightarrow ll$  processes are simulated using the MADGRAPH5\_aMC@NLO 2.2.2 (2.4.2) event generator at leading order (LO) [61]. Supplementary samples are generated with up to four outgoing partons in the hard interaction to increase the number of simulated events in regions of high signal purity. For diboson production, MADGRAPH5\_aMC@NLO is used at next-to-LO (NLO) precision [61], and the FxFx [62] (MLM [65]) prescription is used to match the NLO (LO) matrix element calculation with the parton shower model. Samples for  $t\bar{t}$ [66] and (t-channel) single top quark production[67] are generated at NLO precision using PowHEG 2.0 [51, 52, 53, 54], and for single top quark production in association with a  $W$  boson (tW channel)[68], PowHEG version 1.0 at NLO precision is used. When compared with data,  $W + \text{jets}$ ,  $Z \rightarrow ll$ ,  $t\bar{t}$ , and single top quark events in the tW channel are normalised to their cross sections at next-to-next-to-LO (NNLO) precision[69, 70, 71], while single top quark (t-channel) and diboson events are normalized to their cross sections at NLO precision or higher [71, 72, 73].

## 4.5 QCD Estimation in the $e\mu$ Channel

The QCD model in the  $e\mu$  channel, that attempts to model events where two jets are misidentified as an electron muon pair, is taken from data with a same-sign pair with a transfer factor ( $F_T$ ). The transfer factor determines differences from the same-sign to opposite-sign region and is calculated from a sideband region with an anti-isolated muon ( $0.2 < I_{\text{rel}}^\mu < 0.5$ ).  $F_T$  is initially parameterised by the  $\Delta R$  between the electron and muon, and the number of jets in the event, however additional dependencies on the electron and muon  $p_T$  enter via a correction.

Good agreement is observed in events with no b jets, when applying  $F_T$  onto same sign events compared to opposite sign events where both regions have an anti-isolated muon. However, in events with b jets an additional correction is needed. This is determined to be  $\approx 0.75$  (differs very slightly between data taking years). As this correction is large, it is validated by switching the light lepton anti-isolation, so that the electron is required to have  $0.15 < I_{\text{rel}}^e < 0.5$ . Also, events where both light leptons are anti-isolated are looked at. The correction for b tagged events is equivalent

in all three regions, and a global average of the three is taken for the final correction.

To understand the physical reason for the large difference in no b tag and b tag events in same sign and opposite pairs, studies were performed on simulated samples. It was observed that the electron-muon pair is usually produced from pairs of heavy quarks,  $pp \rightarrow b\bar{b}$  ( $c\bar{c}$ ). If the two jets are initiated from the heavy quarks there is a large bias towards opposite sign jets due to the opposite signs of the quark anti-quark pair. However, if one of the heavy quarks is tagged as a b jet, another object has to be the jet initiator (a radiated gluon for example) and there is therefore no charge preference in the pair. As  $F_T$  is originally fit inclusively in numbers of b jets and the 0 bin is dominant, the correction over predicts the opposite sign to same sign ratio and so a large correction is needed as observed.

## 4.6 Embedding Method

The background for genuine di- $\tau$  lepton pairs is modelled via the embedding method. This is a hybrid method that utilises both data and MC techniques to produce high statistic samples, where the bulk of the event comes from data. This minimises both the chance of MC fluctuations and the size of the uncertainties. The background is dominated by  $Z \rightarrow \tau\tau$  decay however there will be smaller contributions from  $t\bar{t}$  and di-boson processes.

The algorithm first selects  $\mu\mu$  events from data. The selection is chosen to naturally target the pure  $Z \rightarrow \mu\mu$  region but still be loose enough to catch events from other processes, so not to introduce a bias on the  $Z$  boson mass. Events are required to pass the double- $\mu$  trigger with minimum requirements on the invariant mass of the two muons ( $m_{\mu\mu}$ ) and the the  $p_T$  of the leading and trailing muon. Also required at the trigger level is a loose association of the track to the PV and a loose isolation in the tracker. Offline objects matched to the trigger muons, are then required to have standard  $d_z$  and  $\eta$  selections and originate from a global muon-track, as defined in Section 3. The muon pair are required to have opposite charge and have  $m_{\mu\mu} > 20$  GeV. The fraction of processes within this selection is tested with MC background samples and a QCD model from same sign muon pairs with an extrapolation factor. Approximately 97% of selected events are expected to come from  $Z \rightarrow \mu\mu$  events with smaller contributions from  $Z \rightarrow \tau\tau$  ( $\tau \rightarrow \mu$ ), di-boson,  $t\bar{t}$  and QCD. The di-boson and  $t\bar{t}$  relative contributions are greater at higher  $m_{\mu\mu}$  and in

events with tagged b jets whilst the QCD contribution is largest at lower  $m_{\mu\mu}$ . The events selected are biased by detector acceptances. Therefore, corrections on the reconstruction and identification efficiencies are performed in muon  $\eta$  and  $p_T$  using the "tag-and-probe" method.

Next, all energy deposits in the detector from the selected muons are removed. This involves removing the hits on global-muon track in the tracker, hits in the muons systems and clusters in the calorimeters that intercept the muon trajectory. Once completed, the selected muons and their kinematic properties are replaced with a  $\tau$  lepton. To account for the difference in mass between the muon and  $\tau$ , the muons are boosted into the center-of-mass frame of the di-muon system and then this 4-vector is taken for the  $\tau$  but boosted back into the laboratory frame. The event simulation is performed from the PV. The  $\tau$  lepton decay is then simulated with PYTHIA [57, 58] and separate samples are produced for different  $\tau\tau$  decay channels. Only the decay of the  $\tau$  leptons are then processed through the detector simulation and the remainder of the  $\mu\mu$  event is added back. A schematic of the process is shown in Figure 4.8.

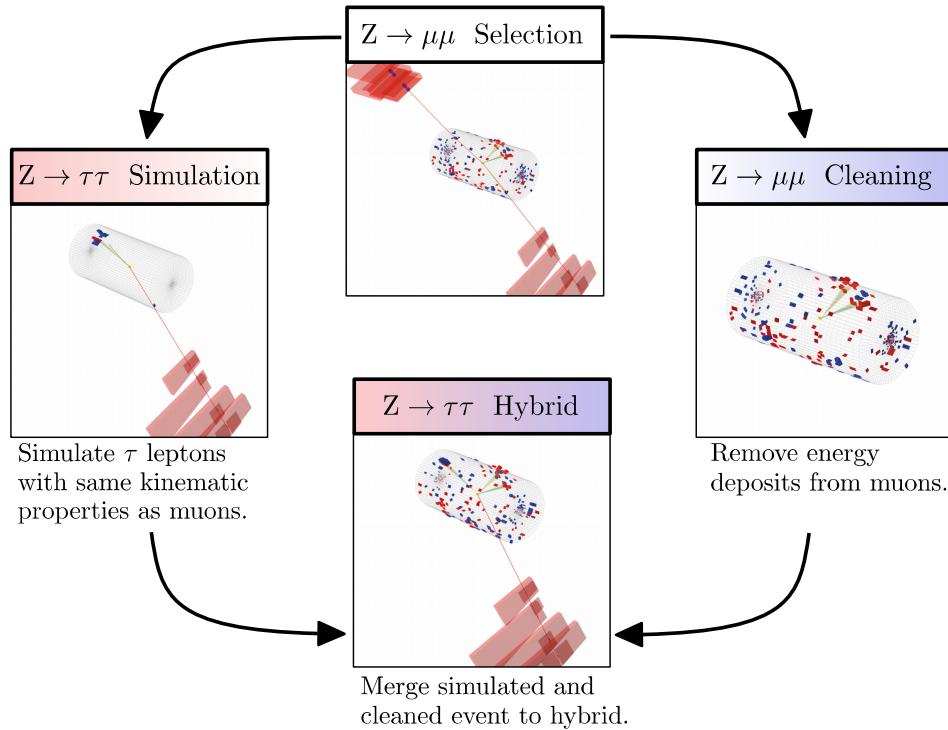


Figure 4.8: Schematic of the embedding method to model genuine di- $\tau$  backgrounds from di-muon events in data.

The embedding method is validated on dedicated samples, where the muons from data are replaced by simulated muons instead of  $\tau$  leptons. A plot of the agreement from these dedicated samples with data is shown in Figure 4.9.

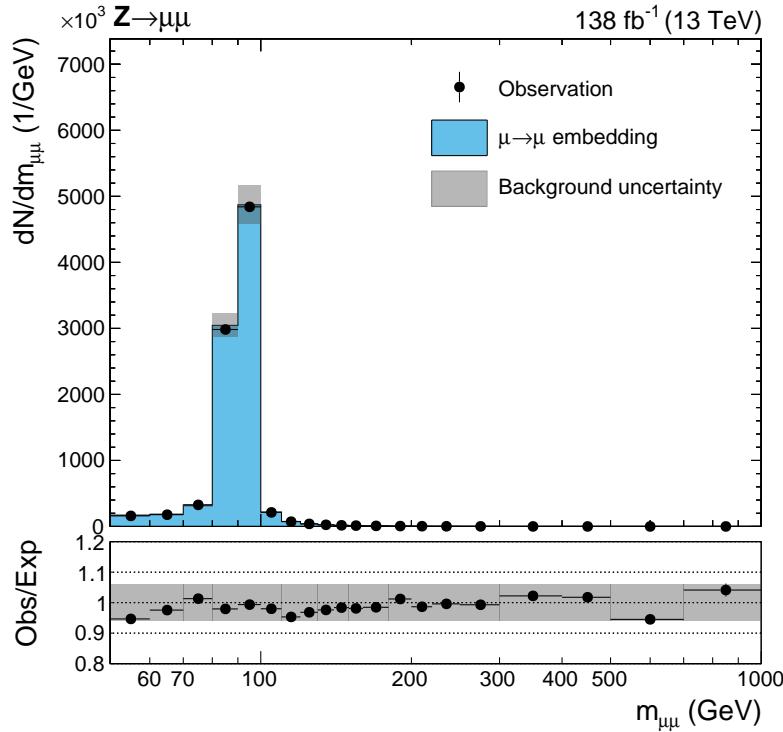


Figure 4.9: Closure plot showing the di-muon mass on the dedicated embedding validation samples.

## 4.7 Fake Factor Method

Backgrounds in which a jet is misidentified as a  $\tau_h$  can be difficult to model using MC due to the poor description of the jet  $\rightarrow \tau_h$  fake rate in simulation. In addition, the small probability of a jet being misidentified as a  $\tau_h$  necessitates the production of high statistics MC samples at a significant computational expense. These shortcomings motivate the use of data-driven estimates for these processes. One such procedure is the fake factor ( $F_F$ ) method.

The  $F_F$  method utilises regions in the data to model the jet  $\rightarrow \tau_h$  background. Firstly, the determination regions, which are jet  $\rightarrow \tau_h$  enriched control regions orthogonal to the signal region. It is used to calculate  $F_F$  by taking the ratio of number of jet  $\rightarrow \tau_h$  events that pass the nominal  $\tau_h$  ID requirement ( $N(\text{Nominal})$ ), to the

number of jet fake events that fail the nominal  $\tau_h$  ID but pass a looser alternative  $\tau_h$  ID requirement ( $N(\text{Alternative} \&\& \text{!Nominal})$ ), as shown in Equation 4.5.

$$F_F = \frac{N(\text{Nominal})}{N(\text{Alternative} \&\& \text{!Nominal})}. \quad (4.5)$$

In the remaining text this numerator and denominator are referred to as the pass and fail regions. The derivation of this ratio is done differentially with respect to key parameters that differ in the two regions. Once  $F_F$  have been derived it is common to calculate corrections in other sideband regions (a region orthogonal to the signal region) and combine  $F_F$  measured from different processes. Finally, the  $F_F$  are applied to the application region (AR). This is defined as the SR but with the criteria that the jet  $\rightarrow \tau_h$  events fail the nominal  $\tau_h$  ID but pass the looser alternative  $\tau_h$  ID requirement. This now models the background from jet  $\rightarrow \tau_h$  events in SR.

The following Sections 4.7.1–4.7.4 detail the complexities of how this method is applied to this analysis. For these searches the nominal  $\tau_h$  ID used is the `Medium DjetWP` and the alternative  $\tau_h$  ID used is the `VVVLoose DjetWP`.

### 4.7.1 Determination Regions

The fake factors are measured separately in each year of data taking period (2016, 2017, 2018), in each channel containing a  $\tau_h$  ( $e\tau_h$ ,  $\mu\tau_h$ ,  $\tau_h\tau_h$ ) and in enriched regions of dominant processes that contribute jet  $\rightarrow \tau_h$  events. In the  $e\tau_h$  and  $\mu\tau_h$  channels  $F_F$  are measured for three processes: QCD, W + jets and  $t\bar{t}$ . In the  $\tau_h\tau_h$  channel  $F_F$  are measured only for the dominant QCD process. The QCD process is assumed to produce two jet fakes and so the fake factors is chosen to be calculated from leading  $p_T$   $\tau_h$  candidate only. Section 4.7.4 discusses how single jet fake events in the  $\tau_h\tau_h$  channel are modelled.

Each separate measurement region is split into three sideband regions based off two cuts that surround the signal region. These regions are named the `Determination Region (A)`, `Alternative Determination Region (B)` and `Correction Region (D)` and are schematically shown in Figure 4.10.

Region A is used to measure and fit fake factors. Region B is an alternative region used to measure and fit fake factor to account for the difference in fake factors between A and C. These alternative fake factors are applied to the fail region in D

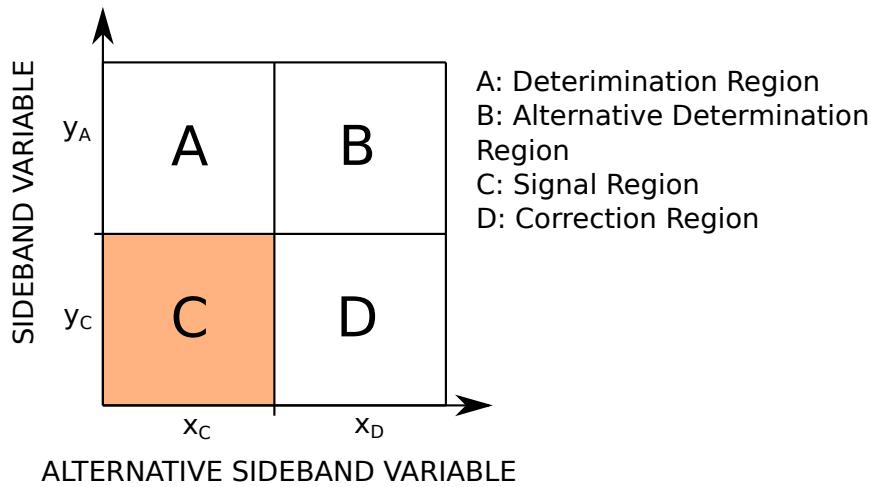


Figure 4.10: Schematic of the regions used for fake factor derivation.

and corrections are calculated comparing it to the pass region in D. The total fake factor per measurement region is calculated as the fake factors derived in region A multiplied by the correction calculated from region B to D.

The selection for  $x_C$ ,  $x_D$ ,  $y_C$  and  $y_A$ , as defined in Figure 4.10, in each separate measurement region are shown below. These are chosen to balance the number of events and the purity of each background in the region.

i)  $\tau_h\tau_h$  QCD

- $y_C$ : The  $\tau_h$  candidates are required to have the opposite sign.
- $y_A$ : The  $\tau_h$  candidates are required to have the same sign.
- $x_C$ : The subleading  $\tau_h$  passes the Medium  $D_{\text{jet}}^{\text{WP}}$ .
- $x_D$ : The subleading  $\tau_h$  fails the VVLoose  $D_{\text{jet}}^{\text{WP}}$  but passes the VVVLoose  $D_{\text{jet}}^{\text{WP}}$ .

ii)  $e\tau_h$  and  $\mu\tau_h$  QCD

- $y_C$ : The  $e/\mu$  and  $\tau_h$  candidates are required to have the opposite sign.
- $y_A$ : The  $e/\mu$  and  $\tau_h$  candidates are required to have the same sign and the  $e/\mu$  to have  $I_{\text{rel}} > 0.05$ .
- $x_C$ : The  $e/\mu$  candidate is required to have  $I_{\text{rel}} < 0.15$ .
- $x_D$ : The  $e/\mu$  candidate is required to have  $0.25 < I_{\text{rel}} < 0.5$ .

iii)  $e\tau_h$  and  $\mu\tau_h$  W + Jets

- $y_C$ : The  $m_T$  between the  $e/\mu$  and the MET  $< 70$  GeV.
- $y_A$ : The  $m_T$  between the  $e/\mu$  and the MET  $> 70$  GeV and no b jets in the event.

$x_C$ : Data.

$x_D$ : W + Jets MC.

iv)  $e\tau_h$  and  $\mu\tau_h$   $t\bar{t}$

$y_C$ : Data.

$y_A$ : MC ( $t\bar{t}$  in B and W + Jets D).

$x_C$ :  $m_T < 70$  GeV.

$x_D$ :  $m_T > 70$  GeV and no b jets.

In the  $\mu\tau_h$  and  $e\tau_h$  channels, W + jets jet $\rightarrow \tau_h$  events are in general the most significant and QCD contributes a smaller fraction.  $t\bar{t}$  inclusively is small but becomes more significant when searching for events with a b jet. The additional  $I_{\text{rel}} > 0.05$  requirement in these channels for QCD is to reduce processes producing genuine leptons and the  $N_{\text{b-jets}} = 0$  requirement for W + Jets is to reduce  $t\bar{t}$  contamination. It is not possible to define a DR that is sufficiently pure in  $t\bar{t}$  events to make a reasonable measurement of  $F_F^{t\bar{t}}$  from data. Therefore  $F_F^{t\bar{t}}$  are derived from MC. A comparison of the  $F_F^{\text{W+jets}}$  measured in data and MC shows only  $\sim 10\text{--}20\%$  differences in the fake rates in data and MC. This observation coupled with the fact that the  $t\bar{t}$  contribution is small compared to the other processes means that any bias introduced by using  $F_F^{t\bar{t}}$  measured in MC is small compared to the uncertainties on the fake factors, discussed in Section 4.9.

#### 4.7.2 Parametrisation

The raw  $F_F^i$  take into account dependencies on  $N_{\text{jets}}$  via the analysis tailed variable  $N_{\text{pre b jets}}$ , the  $p_T$  of the  $\tau_h$  candidate ( $p_T^{\tau_h}$ ) and the  $p_T$  of the jet matched in  $\Delta R$  to the  $\tau_h$  ( $p_T^{\text{jet}}$ ).  $N_{\text{pre b jets}}$  is defined to map the dependence of  $F_F^i$  on  $N_{\text{jets}}$  and describe the categorising variable  $N_{\text{b jets}}$  well. Although not local to the  $\tau_h$ , it helps control other dependencies on the constituents of the event. It is the number of jets in the event with  $|\eta| < 2.4$  and  $p_T > 20$ . These are the same  $\eta$  and  $p_T$  thresholds required for a b jet. The data is split into two bins of  $N_{\text{pre b jets}}$ , equal to 0 and greater than 0. It is then further split by the ratio of  $p_T^{\text{jet}}$  to  $p_T^{\tau_h}$ . An example of the dependence of these two transverse momentums on the fake factor is shown in Figure 4.11.

It is motivated by the observation that the fake factor is largest when the  $p_T^{\text{jet}}$  and

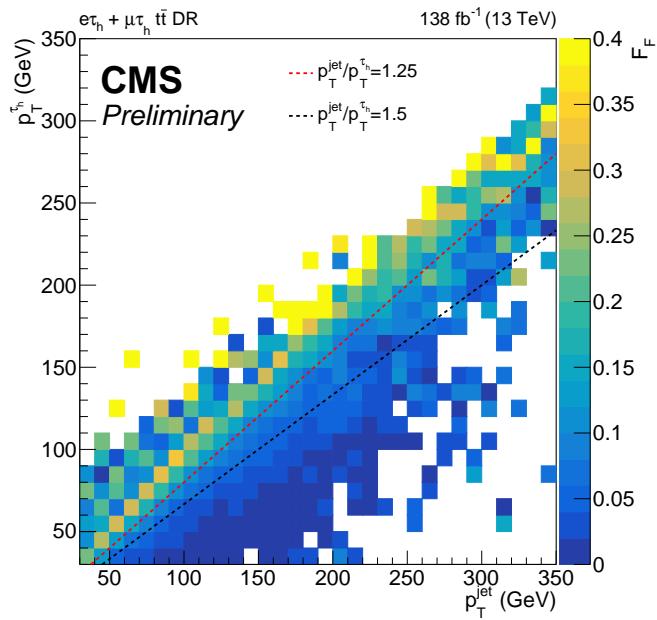


Figure 4.11: A 2D heat map of the fake factors determined from  $t\bar{t}$  MC for the full run-2 dataset in the combined  $e\tau_h$  and  $\mu\tau_h$  channels. This is shown with respect to the  $\tau_h$   $p_T$  and the  $p_T$  of the jet matched to the  $\tau_h$ . The ratio of jet to  $\tau_h$   $p_T$  categorisation used is shown split by the dashed lines.

$p_T^{\tau_h}$  are closest. The physical motivation for this is when they are close, the  $\tau_h$  candidate is likely to be isolated from any other hadronic activity and so more likely to be identified as a  $\tau$ . However, when  $p_T^{\text{jet}}$  is larger than  $p_T^{\tau_h}$ , the candidate is likely surrounded by other hadronic activity and so more likely to be a jet fake. When  $p_T^{\text{jet}}$  is less than  $p_T^{\tau_h}$ , charge pions are likely not close enough to the PV to be clustered into the jet and so the event is more likely to be classified as a jet fake. This will then lead to the fake factor dependence as seen in Figure 4.11.

For all divisions of the phase space, dependence on the  $p_T^{\tau_h}$  is fit using the superposition of a Landau and a zeroth order polynomial in the low- $p_T$  region. The fake factors are seen to rise sharply at high- $p_T$ . This increase happens in either the bin  $140 < p_T^{\tau_h} < 200$  GeV or  $p_T^{\tau_h} > 200$  GeV. To map this effect, binned values are taken based off the algorithm shown in Figure 4.12 and the fit is used below the minimum bin.

The Landau and zeroth order polynomial fits are flattened at  $p_T^{\tau_h}$  values where there is no significant downwards shift or at the final bin. Fake factor fits with respect to  $p_T^{\tau_h}$  are shown in Figures 4.13-4.14. The fake factors are highest in the lowest

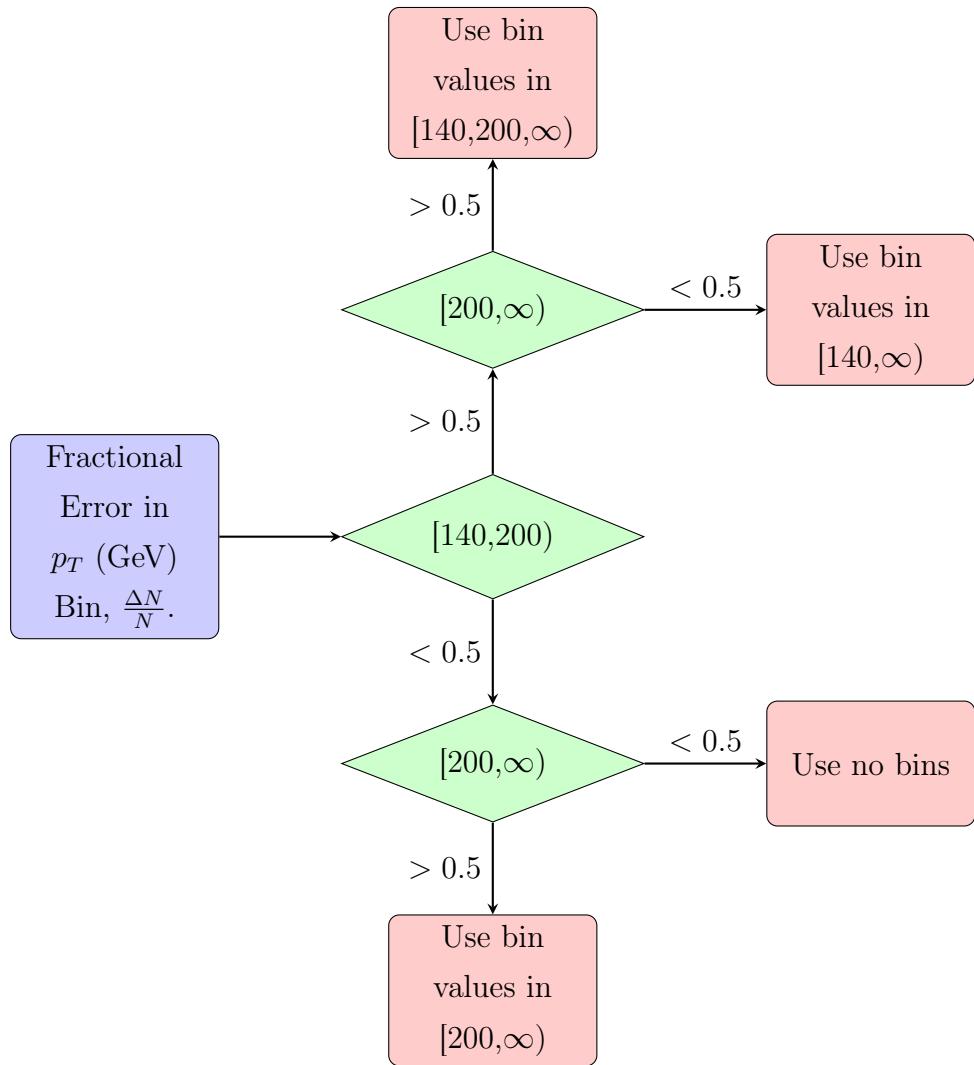


Figure 4.12: Flow chart of the algorithm used to determine where binned values are taken instead of the fit. The blue box represents the input, the green diamonds represent the decisions and the red boxes represent the outputs.

$p_T^{\text{jet}}/p_T^{\tau_h}$  bin and lowest in the highest  $p_T^{\text{jet}}/p_T^{\tau_h}$  bin, as expected. Otherwise the fake factors fall with  $p_T$  in each category until the thresholds used for the high  $p_T$  binning algorithm.

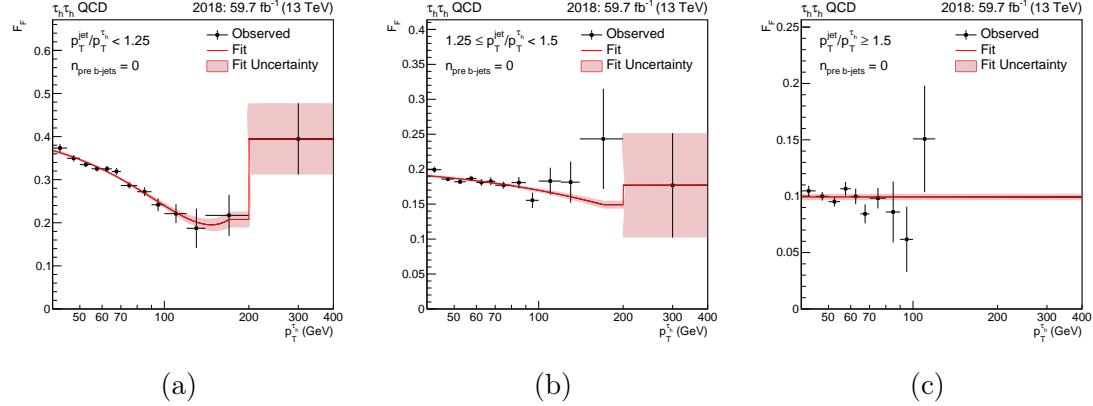


Figure 4.13: Fake factor fits in  $\tau_h\tau_h$  channel for the QCD  $N_{\text{pre b jets}} = 0$  category with 2018 data. The three jet  $p_T$  to  $\tau_h$   $p_T$  categories are shown.

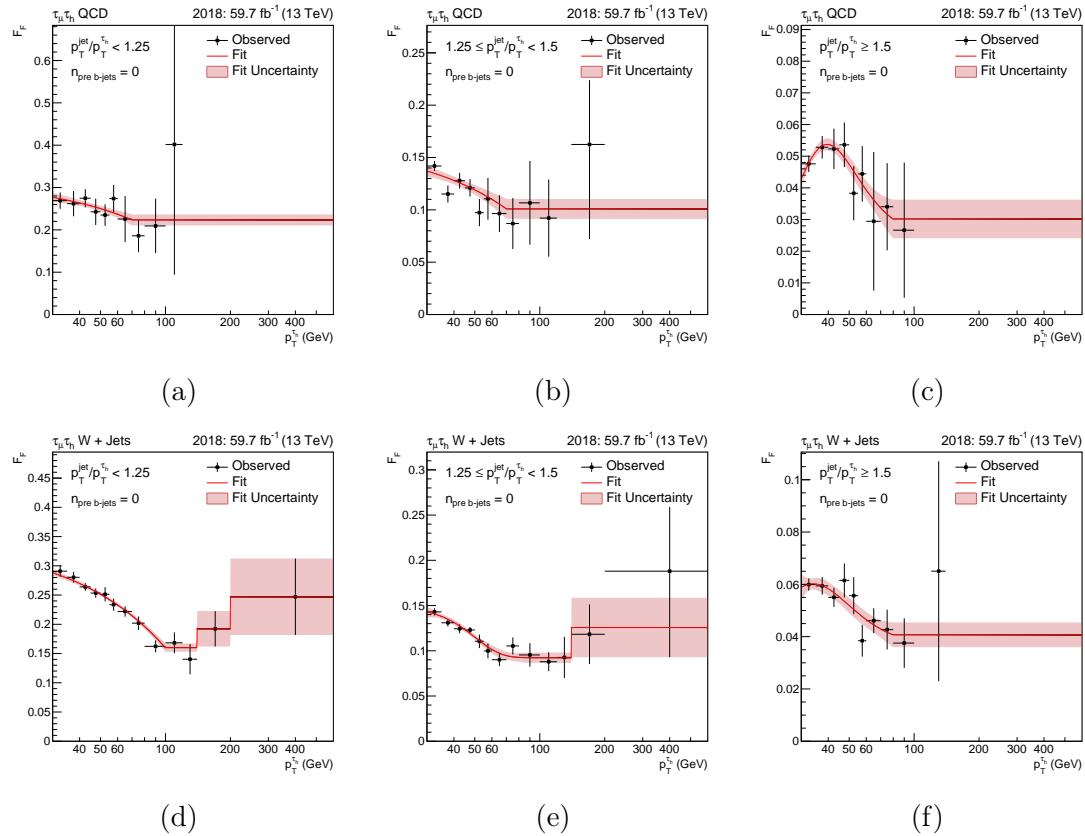


Figure 4.14: Fake factor fits in  $\mu\tau_h$  channel for the QCD and  $W + \text{Jets}$   $N_{\text{pre b jets}} = 0$  category with 2018 data. The three jet  $p_T$  to  $\tau_h$   $p_T$  categories are shown for each process.

### 4.7.3 Corrections

In the  $\tau_h\tau_h$  channel, the measured  $F_F^{\text{QCD}}$  are then corrected to account for non-closures in other variable in the **Determination Region**. The only significance non-closures are observed for  $E_T^{\text{miss}}$  related variables and are largest for events with  $N_{\text{pre b jets}} = 0$ . Closure corrections are performed for the variable  $\Delta R$  in bins of  $N_{\text{b jets}}$ . In the  $\mu\tau_h$  and  $e\tau_h$  channels, the measured  $F_F^{\text{QCD}}$  and  $F_F^{\text{W+jets}}$  are corrected for non-closures observed in the  $E_T^{\text{miss}}$  variables and  $p_T^{e/\mu}$  distributions. A study was performed to determine the nature of these non-closures and it was found that the cause was due to fake  $E_T^{\text{miss}}$  arising from mismeasurement of the energies of particles in a jet. A diagram of this effect is shown in Figure 4.15.

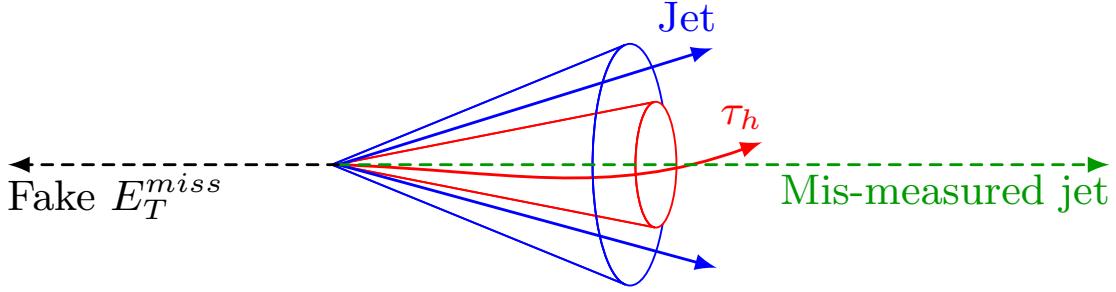


Figure 4.15: Diagram showing how fake  $E_T^{\text{miss}}$  arises from mismodelling jet constituents in  $\tau_h$  identification.

To correct for this effect, the QCD fake factors are corrected as a function of  $C_{\text{QCD}}$ , where  $C_{\text{QCD}}$  is defined as,

$$C_{\text{QCD}} = \frac{E_T^{\text{miss}} \cos \Delta\phi(\vec{p}_T^{\text{miss}}, \vec{p}_T^{\tau_h})}{p_T^{\tau_h}}. \quad (4.6)$$

where  $\Delta\phi(\vec{p}_T^{\text{miss}}, \vec{p}_T^{\tau_h})$  is the separation in the azimuthal angle between the the missing a  $\vec{p}_T^{\text{miss}}$  and  $\vec{p}_T^{\tau_h}$ . The numerator quantifies the missing transverse momentum in the direction of the  $\tau_h$  candidate. Once divided by the  $\tau_h$   $p_T$ ,  $C_{\text{QCD}}$  is a measure of the fraction of missing to visible  $\tau_h$  transverse momentum aligned with the  $\tau_h$ . For  $\text{W} + \text{jets}$  and  $t\bar{t}$  the situation is slightly different due to the presence of genuine missing energy from neutrinos. In this case, the correction variable is modified to approximately subtract the genuine  $E_T^{\text{miss}}$  from the total. This approximation assumes the neutrino is back-to-back and balanced with the light lepton (which is

exactly true for W bosons produce at rest in the transverse direction). The equation then becomes,

$$C_W = \frac{(E_T^{\text{miss}} + p_T^{e/\mu}) \cos \Delta\phi(\vec{p}_T^{\text{miss}} + \vec{p}_T^{e/\mu}, \vec{p}_T^{\tau_h})}{p_T^{\tau_h}}. \quad (4.7)$$

When either correction variable is separated from 0, a larger quantity of fake  $E_T^{\text{miss}}$  is expected in the event. In these regions a large correction is needed due to the mis-measured jet energy spectrum shifting the  $\tau_h$  candidate isolation and so shifting the  $\tau$  identification scores. Examples of these closure corrections are shown in Figure 4.16

After the Determination Region is modelled well for all variables of interest, extrapolation corrections from the fake factors derived in B applied to region D are calculated. In the  $\tau_h\tau_h$  the correction is parameterised by the  $p_T$  of the leading  $\tau_h$  candidate, in the  $e\tau_h$  and  $\mu\tau_h$  channels it is parameterised by the  $p_T$  of the light lepton. Where statistics allow, these corrections are calculated in the high mass optimisation procedure categories. Examples of the extrapolation corrections are shown in Figure 4.17.

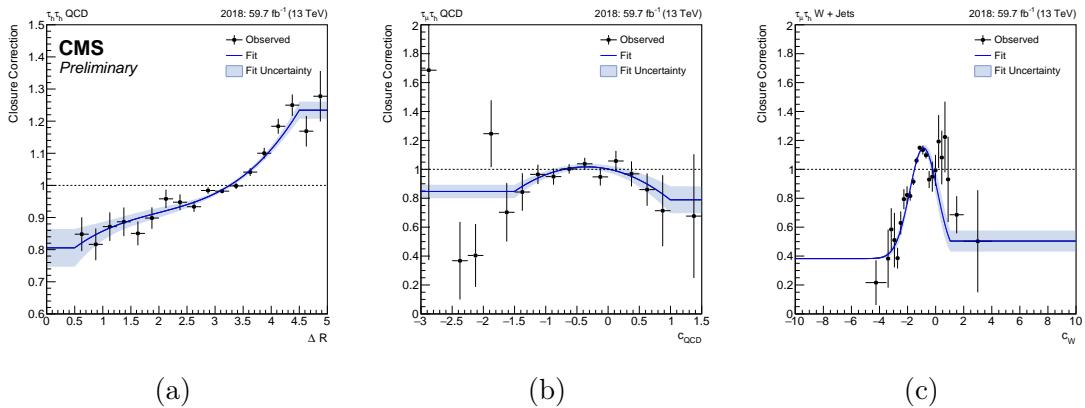


Figure 4.16: Determination region closure correction fits with 2018 data. (a) is the correction parametrised by  $\Delta R$  in events with  $N_b$  jets = 0 in the  $\tau_h\tau_h$  channel. (b) and (c) show the correction for the  $\mu\tau_h$  channel parametrised by the specific correction variables defined in Equation 4.6 and 4.7 for QCD and W + jets processes respectively.

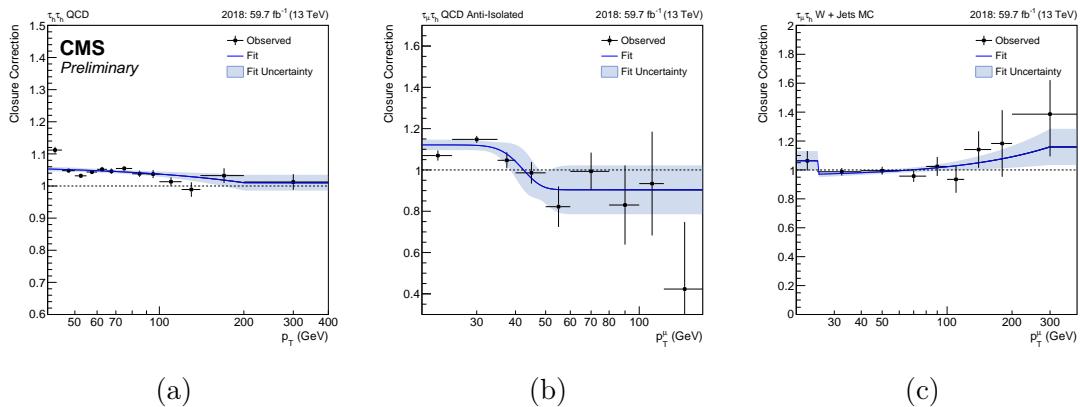


Figure 4.17: Determination region to application region closure correction fits with 2018 data. (a) is the correction moving from same sign to opposite sign  $\tau$  leptons the parameterised by leading  $\tau_h$   $p_T$  in events with  $N_{\text{b-jets}} = 0$  in the  $\tau_h\tau_h$  channel. (b) and (c) show the correction for the  $\mu\tau_h$  channel moving from same sign to opposite sign  $\tau$  leptons and high  $m_T$  to low  $m_T$  both parameterised by the muon  $p_T$  for QCD and W + jets processes respectively.

#### 4.7.4 Applying Fake Factors

In the  $e\tau_h$  and  $\mu\tau_h$  channels the  $F_F^i$  measured for the different processes are combined into an overall factor,  $F_F$ , using

$$F_F = \sum_i f_i \cdot F_F^i, \quad (4.8)$$

where the factor  $f_i$  is defined as

$$f_i = \frac{N_{\text{AR}}^i}{\sum_j N_{\text{AR}}^j}, \quad (4.9)$$

which is the fraction of events with a jet  $\rightarrow \tau_h$  originating from process  $i$  over the total number of jet  $\rightarrow \tau_h$  events for all processes in the application region. These fractions of events are estimated with MC, with a QCD model extrapolated from same sign  $\tau$  pairs. It is observed that  $W + \text{jets}$  is the dominating process in this region, however, there are affects from QCD at low  $m_T$  and from  $t\bar{t}$  in the b tagged categories. These fractions are then multiplied to the relevant corrected fake factor and applied to the fail region in C, with any events which are not jet  $\rightarrow \tau_h$  subtracted off with MC.

For the  $\tau_h\tau_h$  channel there are two  $\tau_h$  candidates that a jet can be misidentified as. For this analysis, the fake factors are only applied the leading  $\tau_h$  candidate failing

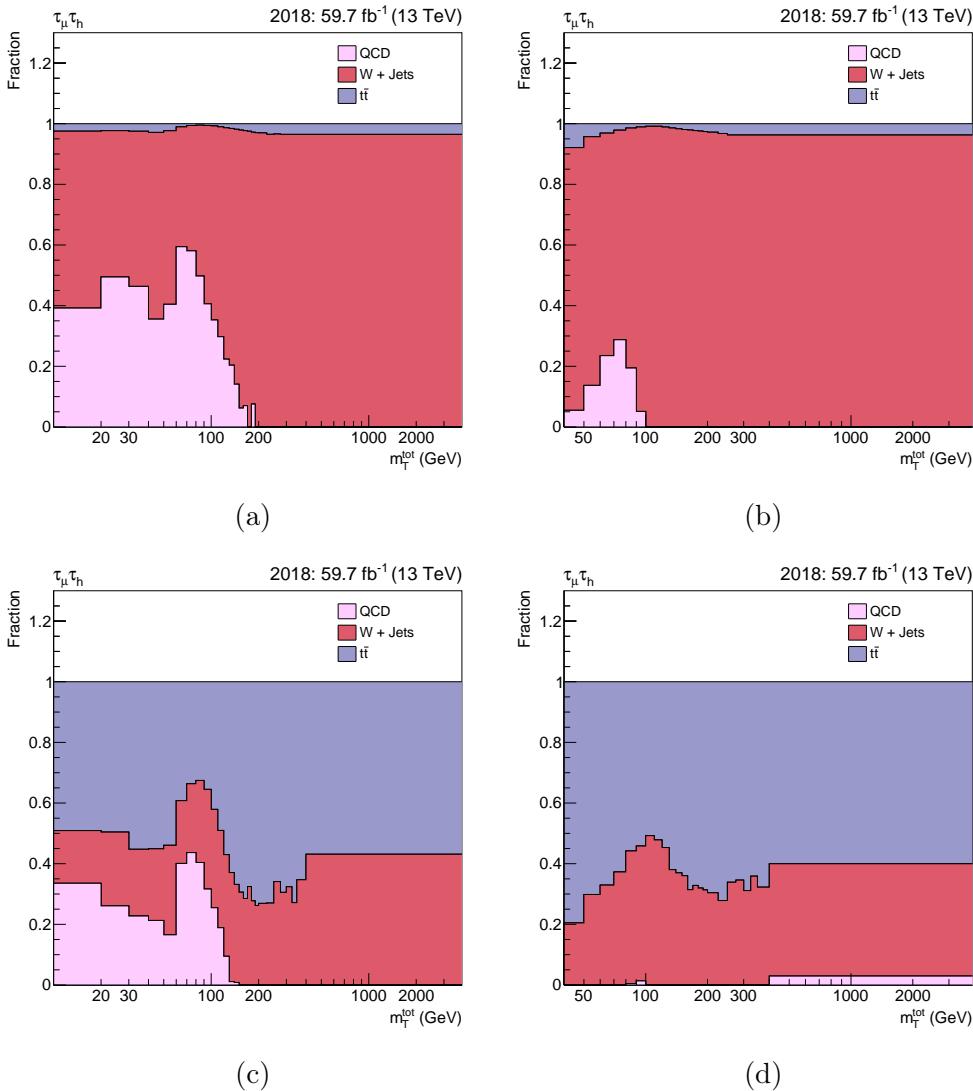


Figure 4.18: The expected application region fractions of the processes in the  $\mu\tau_h$  channel. (a) and (b) show the `no_b_tag` `Tight- $m_T$`  and `Loose- $m_T$`  categories and (c) and (d) show the `b_tag` `Tight- $m_T$`  and `Loose- $m_T$`  categories respectively.

the  $\tau_h$  ID in C. This models all events where the leading  $\tau_h$  candidate is a jet $\rightarrow\tau_h$ . However, this leaves a small fraction of events, where the leading candidate is a genuine  $\tau$  and the sub-leading candidate is a jet $\rightarrow\tau_h$ . This contribution (mostly from W + jets) is added back with MC.

## 4.8 MC Corrections

The corrections apply both to simulated and embedding samples as the  $\tau$  decay is simulated, however they are derived separately. For electrons and muons, corrections are applied to triggers, tracking efficiencies, and the identification and isolation requirements. Using the ‘tag-and-probe’ method, they are obtained in bins of  $p_T$  and  $\eta$  of the corresponding lepton, with  $Z \rightarrow ee$  and  $Z \rightarrow \mu\mu$  events. These corrections are generally no more than a few percent. The electron energy scale is adjusted to the scale measured in data using the Z boson mass peak in  $Z \rightarrow ee$  events.

Similarly, corrections are derived for the efficiency of triggering and identification efficiency of  $\tau_h$  candidates. In the  $e\tau_h$  and  $\mu\tau_h$  channels, trigger efficiency corrections are obtained from the ratio of fits to data versus simulated samples, for the trigger efficiency as a function of  $p_T$ . For the  $\tau_h\tau_h$  channel, this is instead done using the binned values of the  $\tau_h$  decay modes. The identification efficiency corrections are derived as a function of the  $p_T$  of the  $\tau_h$  candidate. Corrections to the energy scale of the  $\tau_h$  candidates and of electrons misidentified as  $\tau_h$  candidates are obtained from likelihood scans of discriminating observables, such as the reconstructed  $\tau_h$  candidate mass. For muons misidentified as  $\tau_h$  candidates, the energy scale correction is negligible.

The correction of the missing transverse energy’s magnitude and resolution in the embedding samples is needed to account for the incomplete removal of energy deposits from muons replaced by simulated  $\tau$  decays during the embedding procedure. These corrections are derived by comparing  $p_T^{\text{miss}}$  in embedded events to fully simulated events.

In fully simulated events, a specific trigger inefficiency caused by a shift in the timing of the inputs of the ECAL L1 trigger in the region at  $|\eta| > 2.0$  during the 2016 and 2017 data taking. This effect is named prefire. This resulted in a loss of efficiency for events containing an electron or jet with  $p_T$  larger than  $\approx 50$  or  $\approx 100$

GeV respectively, in the region of  $2.5 < |\eta| < 3.0$ . Corresponding corrections are derived from data and applied to the simulation.

Corrections on the energy of jets are calculated in bins of the jet  $p_T$  and  $\eta$ . These range from subpercent levels in the central part of the detector to a few percent in the forward region. The energy resolution of the simulated jets is also tuned to match that of the data. A correction is applied to the missing transverse momentum, based on differences in the estimated hadronic recoil between data and simulation. A MC to data correction for a b jet to pass the selection criteria is also determined. This corrections can adjust the number of b jets in a simulated event.

Any differences from simulated events to data, where an electron or muon is reconstructed as a  $\tau_h$ , are corrected from the pure  $Z \rightarrow ee$  and  $Z \rightarrow \mu\mu$  regions. Similarly, a correction is applied to account for residual differences in the  $\mu \rightarrow e$  misidentification rate between data and simulation.

Further MC to data corrections are applied to the dilepton mass and  $p_T$  spectra in simulated  $Z \rightarrow ll$  events. These are derived from  $Z \rightarrow \mu\mu$  events. Additionally, all simulated  $t\bar{t}$  events are weighted to match the top quark  $p_{Tt}$  distribution observed in data.

## 4.9 Uncertainty Model

The statistical uncertainties are taken into account by the Barlow-Beeston method, described in Ref. [74, 75]. The systematic model is split into uncertainties based on the online and offline reconstruction of objects and the background and signal modelling. An uncertainty is correlated across channels when it represents a shift on the reconstruction of an object and decorrelated otherwise. It is decorrelated across era of data taking when the shift is derived independently by era. The embedded samples use the same uncertainty scheme as MC but 50% are correlated and 50% uncorrelated with MC uncertainties, because of the shared real data in the measurement.

## Hadronic Taus

Uncertainties on the  $\tau_h$  triggers are obtained from the fitted scale factors used to derive the corrections for the  $\tau_h$  trigger efficiencies. The legs of the double- $\tau_h$  and  $e/\mu\text{-}\tau_h$  cross triggers in different decay mode bins are treated as uncorrelated. For the single- $\tau_h$  trigger leg, due to limited statistics it is not possible to determine scale factors and uncertainties split by decay mode and therefore a single uncertainty common to all decay modes is applied. The double- $\tau_h$  trigger uncertainties are further split in the  $p_T$  regions  $< 100$  GeV and  $> 100$  GeV to allow the fit more freedom to adjust the high  $p_T$  regions relative to the low  $p_T$  regions. Uncertainties are also applied on the energy scale of the  $\tau_h$  candidates. These uncertainties range between 0.2 and 1.1%. Finally, an uncertainty on the ID efficiency is placed as a function of  $p_T$  in the  $e\tau_h$  and  $\mu\tau_h$  channels, and of the  $\tau_h$  decay mode in the  $\tau_h\tau_h$  channels. These vary between 3-9% and is uncorrelated in each variable bin it is derived in. To account for the different anti-lepton discriminator working points, an uncertainty of 3% per  $\tau_h$  is applied and treated as uncorrelated between the channels where different  $D_{\text{WP}}^{e/\mu}$  are used.

## Light leptons

The uncertainty on the trigger efficiencies amounts to 2% per lepton in the  $e\tau_h$ ,  $\mu\tau_h$  and  $e\mu$  channels. They are basically normalisation uncertainties but implemented as shape uncertainties as they only touch the events triggered by the corresponding cross trigger or single lepton triggers. Uncertainties are also placed on the electron energy scale based off the calibration of ECAL crystals. This information is not reliable for embedding samples and so uncertainties of 0.5-1.25% are placed here. The energy scale variations are negligible and so not included. Another 2% uncertainty is placed on the ID of any electron or muon in the event.

## Jets

Jet energy scale and resolution uncertainties arise from a number of sources. These include limited statistical measurements used for calibration, energy measurement changes due to detector ageing, and bias corrections to address differences between simulation and data. Uncertainty ranges are from sub-percent to  $\mathcal{O}(10\%)$ . Uncertainties are also placed on the tagging of b jets, which vary from 0–3%.

### Leptons misidentified as hadronic taus

Uncertainty shifts are applied for the energy scale of leptons faking  $\tau_h$  parametrised by the  $p_T$  of the  $e/\mu \rightarrow \tau_h$  fake. The magnitude is 1.0% for muons in all eras and for electrons the uncertainties vary between 0.5 and 6.6 %.

### Jets misidentified as hadronic taus

The backgrounds with jets misidentified as  $\tau_h$  are estimated from data with the fake factor method. There are different sources of uncertainty related to this method. The first uncertainties come from subtracting off other background processes with MC to form the determination region. The subtraction is shifted up and down by 10% to determine new weights. Next, statistical uncertainties on all of the fake factor method fits are accounted for, where the binned values are uncorrelated with the rest of the fit. An uncertainty is also placed on the choice of fit function, the shifts are estimated by comparing the fits to a 1st order polynomial fit set to constant above 100 GeV. The final systematic variation are then on the extrapolation corrections by applying the corrections twice and not at all. The size of each systematic uncertainty varies from 0–10%, whilst the statistical element from the fits can be larger in the tails of the distributions.

### Jets misidentified as light leptons

Background with jets misidentified as electrons or muons from QCD are only considered in the  $e\mu$  channel and modelled from data. Uncertainties are placed based on the statistical uncertainties in the determination region which are 2–4% and the extrapolation to the signal region that are  $\mathcal{O}(10\%)$ .

### Muons misidentified as electrons

Background with muons misidentified as electrons are only considered in the  $e\mu$  channel. These events are modelled from MC and any generator matched muon identified as an electron are given a 15%–45% uncertainty, that is derived from the initial corrections.

### MET

The MET uncertainties is different dependent on process. For all processes that are not  $t\bar{t}$  and diboson, the hadronic recoil response and its resolution are varied within

the uncertainties determined during the computation of the recoil corrections. For  $t\bar{t}$  and diboson an uncertainty is derived on the energy carried by an unclustered particle [76]. These uncertainties vary between 0–10%.

### Background process specific uncertainties

Uncertainties on the  $t\bar{t} p_T$  and DY  $m_{ll}\text{-}p_T$  reweighting is placed by applying the correction twice and not at all. An additional uncertainty is placed to cover the  $t\bar{t}$  contamination in embedding, where the removed  $t\bar{t}$  genuine  $\tau$  pair is shifted up and down by 10%. Some non-closures are observed in embedded  $Z \rightarrow \mu\mu$  control samples. Therefore, these non-closures are taken as an additional shape uncertainty as a function of the  $Z p_T$  and  $m_{\tau\tau}$ . Uncertainties on the normalisation background processes with sizes 4% for  $Z \rightarrow ll$  and  $W + \text{jets}$  production [69], 6% for  $t\bar{t}$  production [70, 71], and 5% for diboson and single t quark production [71, 72, 73].

### Signal process specific uncertainties

For the  $gg\phi$  and  $bb\phi$  processes, the variation of `hdamp` parameter of the PowHEG MC generator as well as the  $\mu_R/\mu_F$  scale variations are used to determine the uncertainties on the  $p_T$  spectrum of each contribution at NLO QCD to the Higgs boson production via gluon fusion (top, bottom, top-bottom interference). These are also determined from additional samples produced up to the generator level, and applied as event weights dependent on generator level  $p_T$  after the parton shower simulation. For the VLQ signal samples, the PDF and  $\mu_R/\mu_F$  scale variations are applied on an event-by-event basis. These uncertainties are included as shape uncertainties as they may affect the shapes of the  $m_T^{\text{tot}}$  distribution as well as the predicted signal yields.

### Luminosity

(1.2%, 2.3%, 2.5%) normalisation luminosity uncertainty is applied to the (2016, 2017, 2018) templates which originate from MC simulation.

### Prefiring

Upper and lower bounds are taken from the efficiency maps and propagated on all MC samples as shape uncertainty for 2016 and 2017. The size of the uncertainty as the weight itself depends on the event topology. In general the uncertainty is at the order of 1%.

## 4.10 Signal Extraction

A simultaneous binned maximum likelihood fit over all analysis categories is used to extract the results. The likelihood takes the form,

$$\mathcal{L}(\text{data} \mid \mu, \theta) = \prod_i^{N_i} \text{Poisson}\left(n_i \mid \sum_j^{N_j} g_j(\mu_{ij}) \cdot s_{ij}(\theta) + \sum_k^{N_k} b_{ik}(\theta)\right) \cdot p(\hat{\theta} \mid \theta), \quad (4.10)$$

where  $i$  loops through all histogram bins and analysis categories. The indices  $j$  and  $k$  loop over all signal and background processes for the hypothesis being fit.  $n_i$ ,  $s_i$  and  $b_i$  are the data observed, signal and background expectation respectively in each bin.  $\theta$  represents the set of nuisance parameters (corresponding to the systematic uncertainties as detailed in Section 4.9) that parametrise the signal and background modelling.  $\mu$  are rate parameters and  $g(\mu)$  are scaling functions that scale to a signal to a specific hypothesis. The form of the Poisson probabilities are,

$$\text{Poisson}(n \mid x) = \frac{x^n e^{-x}}{n!}. \quad (4.11)$$

Finally,  $p(\hat{\theta} \mid \theta)$  represents the probability density function (pdf) of each nuisance parameter ( $\theta$ ) with respect to the initial value of the parameter ( $\hat{\theta}$ ).

The pdfs come in two forms, the first is for uncertainties that only affect the normalisation of the process and are modelled by log-normal pdfs. The second is for uncertainties that affect the shape of the distribution, these are assigned Gaussian pdfs. The  $\pm 1\sigma$  shifts for each shape variations are derived and vertical morphing [75] is used to interpolate and extrapolate within and outside the shifts. Both pdfs are dependent on the mean ( $\mu$ ) and standard deviations ( $\sigma$ ) and the functional forms are shown in Table 4.3.

Gaussian	Log-normal
$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$

Table 4.3: pdfs used for nuisance parameters.

The following subsections discuss the results of many such fits. The key fits to understand the results are the background-only fit and the signal-plus-background fits.

The background-only fit is performed with  $N_j$  (number of signal processes) set to 0. For all signal-plus-background fits, the fit is done with respect to a single mass hypothesis, however within this mass hypothesis can be a number of signal processes. The model independent resonance search has separate  $gg\phi$  and  $bb\phi$  signal modes and so two rate parameters  $\mu_{gg\phi}$  and  $\mu_{bb\phi}$  are needed. When the samples are initially scaled to the cross section times branching ratio ( $\sigma \times B(\phi \rightarrow \tau\tau)$ ) of 1 pb and  $g(\mu) = \mu$  for both processes,  $\mu_{gg\phi}$  and  $\mu_{bb\phi}$  represent the  $\sigma \times B(\phi \rightarrow \tau\tau)$  with units of pb. To avoid negative signal strengths,  $\mu$  will only be taken to be positive. Also used in the following subsections, is a signal-plus-background channel/category compatibility fit. In this fit the signal processes and rate parameters are further split in each channel or category utilising index  $i$  in Eq. 4.10. This is used to determine the compatibility of the results in different decay channels and analysis categories. In this case,  $\mu$  is allowed to take negative values to help fully understand the fits to data in each channel or category.

The vector leptoquark search has two signal modes; the t-channel interaction and the interference with Drell-Yan. However, as this is a model-dependent interpretation of these results both these rate parameters scale together. The scaling functions differ between the two processes with  $g_{t\text{-channel}}(\mu) = \mu^4$  and  $g_{\text{interference}}(\mu) = \mu^2$  to mimic how the cross sections of each process scales. When the initial samples are scaled to cross section at  $g_U = 1$ ,  $\mu$  corresponds to the coupling  $g_U$ .

For the MSSM interpretation of the results, there are three Higgs bosons to consider in the signal model (h, H and A) produced via both gluon fusion and in association with b quarks. The  $gg\phi$  samples are also split into the separate loop contributors, so the kinematic properties can be properly scaled to MSSM prediction, as described in Section 4.1.1. The SM-like Higgs boson is considered in the MSSM signal model to monitor differences in the observed Higgs boson prediction between the MSSM and the SM. In each benchmark scenario chosen, the signal prediction depends only on  $m_A$  and  $\tan\beta$  and the scaling to cross section is shown in Eq. 4.1. As the potential scaling functions for MSSM interpretations are not necessarily smooth one-to-one mappings, the likelihood is tested for individual points on the  $m_A$ - $\tan\beta$  parameter space. At each point, the MSSM Higgs bosons are scaled to the theory predicted cross section times branching ratio. To test the MSSM hypothesis over the SM hypothesis, the single rate parameter  $\mu$  is used and only allowed to take values of 1 (MSSM) and 0 (SM) with  $g(\mu) = \mu$ . As the SM Higgs boson is added to the

background modelling and the MSSM prediction of the observed Higgs boson is added to the signal model when  $\mu = 1$ , the SM Higgs boson prediction must then be subtracted from the signal model.

The confidence intervals in the best fit results are given by the  $-2\Delta \ln \mathcal{L}$ , where  $\Delta \ln \mathcal{L}$  is the difference between  $\ln \mathcal{L}$  of the best fit model and the test value of  $\mu$ . The 68% and 95% confidence regions with two degrees of freedom (as in the model independent resonant search) are determined by  $-2\Delta \ln \mathcal{L} = 2.28$  and 5.99 respectively.

Upper limits are placed using the modified frequentest approach [77, 78] with a profile likelihood ratio used for the test statistic, as defined below.

$$q_\mu = -2 \ln \left( \frac{\mathcal{L}(\text{data}|\mu, \hat{\theta}_\mu)}{\mathcal{L}(\text{data}|\hat{\mu}, \hat{\theta}_{\hat{\mu}})} \right), 0 \leq \hat{\mu} \leq \mu, \quad (4.12)$$

where  $\hat{\mu}$  and  $\hat{\theta}_{\hat{\mu}}$  are the best fit values of  $\mu$  and  $\theta$ .  $\hat{\theta}_\mu$  are the values of  $\theta$  that is maximised by the likelihood for a tested value of  $\mu$ . The bounds on  $\hat{\mu}$  are to ensure a positive signal strength with a one-sided confidence interval. The probability of  $q_\mu \geq q_\mu^{\text{obs}}$  is,

$$\text{CL}(\mu) = \int_{q_\mu^{\text{obs}}}^{\infty} f(q_\mu | \mu, \theta_\mu^{\text{obs}}), \quad (4.13)$$

where  $f(q_\mu | \mu, \theta_\mu^{\text{obs}})$  is the pdf of  $q_\mu$ .  $\text{CL}_b$  and  $\text{CL}_{s+b}$  are then defined by the relevant background-only and signal-plus-background fits.  $\text{CL}_s$  is defined as the ratio of  $\text{CL}_{s+b}$  and  $\text{CL}_b$  and then upper limits are placed at the confidence level of  $1 - \text{CL}_s$ . The  $f(q_\mu | \mu, \theta_\mu^{\text{obs}})$  are determined using the asymptotic approximation [79] and results are cross-checked and deemed consistent with toy MC datasets.

If a deviation from the background expectation is observed, the size of the deviation is quantified by a significance. To test rejection of the background-only hypothesis in favour of the signal-plus-background hypothesis,  $\mu$  is replaced with 0 in the test statistic. The  $p$ -value,  $p_0$  is then,

$$p_0 = \int_{q_0^{\text{obs}}}^{\infty} f(q_0 | 0, \theta_\mu^{\text{obs}}). \quad (4.14)$$

$p_0$  is uniformly distributed between 0 and 1 for the background-only hypothesis and so the probability and significance of rejecting the background-only hypothesis can be found.

## 4.11 Postfit Plots

Figures 4.19 and 4.20 show the unblinded distributions in the most sensitive analysis categories. For simplicity, the  $e\tau_h$  and  $\mu\tau_h$  channels have been combined. Figure 4.19 shows the distributions of the  $m_{\tau\tau}$  discriminator in the no b tag low-mass optimisation categories. A signal-plus-background fit for a model-independent gluon fusion resonant mass hypothesis of 100 GeV is shown and the changes in the background modelling when using a background-only fit is displayed in the ratio. Figure 4.20 shows the distributions of the  $m_T^{\text{tot}}$  discriminator in the high mass optimisation categories. A background-only fit is shown for the stacked background and example signal hypotheses for the model independent 1.2 TeV  $gg\phi$  and  $bb\phi$  resonances and VLQ (BM 1) 1 TeV mass points are displayed.

In the low-mass optimisation categories, a small excess of events is observed on the Z boson peak in the no b tag categories and reasonable agreement is observed in the b tag categories. The excess of events are distributed in  $m_{\tau\tau}$  between 80 and 120 GeV. A signal-plus-background hypothesis is best fit with a 100 GeV  $gg\phi$  signal with a cross section times branching ratio of 5.8 pb. In this same fit the  $bb\phi$  process is constrained by the b tag categories to give a signal yield of 0. A background-only fit is also performed on the data, it is observed that this can only partly explain the differences observed between background and data. Even after a background-only fit there is still a small excess of data events over the Z boson peak.

In the high-mass optimisation categories, another small excess is observed in high  $m_T^{\text{tot}}$  bins, particularly in the most sensitive no b tag categories. This excess is best fit by a model independent gluon fusion resonant mass at 1.2 TeV with a cross section times branching ratio of 3.1 fb. There are no considerable differences observed in background modelling between signal-plus-background and background-only fits. This is as the uncertainties in these bins are more statistically dominated and the majority of the systematic uncertainties are constrained in the bulk of the distribution. Good agreement is observed in the rest of the distribution. There is a very small deviation in the b tag categories, but as this can also be explained by a  $gg\phi$  signal, the  $bb\phi$  signal is heavily constrained and so largely does not contribute to the signal-plus-background fit of the excess. Similar to the  $bb\phi$  signal, the VLQ BM 1 signal is constrained by the results in the b tag categories, leading to a small non-zero best fit signal strength, but cannot explain the excess in the no b tag categories.

## 4.12 Model Independent Results

### 4.12.1 Limits

95% CL limits are set on the assumption of absence of a signal for the search for a  $gg\phi$  or  $bb\phi$  resonance and shown in Figure 4.21. In each case, the other process is allowed to float freely in the fit. The excesses observed in the postfit distributions act to weaken the observed limit compared to the expected limit at 100 GeV and 1.2 TeV, as more data was observed than expected. For  $gg\phi$  production the expected limits flatten under 100 GeV, due to difficulty of separating signal from the Z boson at this mass. Both sets of limits vary from  $\mathcal{O}(10 \text{ pb})$  at 60 GeV to 0.3 fb at 3.5 TeV.

95% expected limits are drawn on the fit to each di- $\tau$  decay channel individually and are shown in Figure 4.22. This gives a measure of the sensitivity of each channel. In the high mass optimisation categories, the combined limit is heavily dominated by the  $\tau_h\tau_h$  channel. This is mostly driven by branching fraction, as all channels in this mass range have similar signal separation ability. In the high mass optimisation categories, the combined limit is more a contribution of all channels. In the  $\tau_h\tau_h$  channel in this region, the QCD multijet background is the largest fraction of any non  $Z \rightarrow \tau\tau$  backgrounds in all channels and so the limit for this channel is weakened and the other channels contribute to the combined limit more. The high double- $\tau_h$  trigger  $p_T$  thresholds (chosen because of the QCD multijet background) also lowers the signal acceptance in the  $\tau_h\tau_h$  channel.

A comparison of the limits are also made with the ATLAS experiment and in particular the results presented in Ref. [3]. This ATLAS search looks for the same signal but over a smaller mass range, from 200 GeV to 2.5 TeV. Plots showing the comparison of the expected and observed limits for  $gg\phi$  and  $bb\phi$  are shown in Figure 4.23. The expected limits from the CMS and ATLAS results are roughly compatible over the shared mass range, except at high mass where the extra statistics from the embedded  $Z \rightarrow \tau\tau$  samples compared to MC allow for lower background uncertainties and hence a stronger limit. The ATLAS result observed no excess of events compatible with  $gg\phi$  signal at 1.2 TeV, in fact a small deficit was observed. Also, ATLAS observed local excesses at 400 GeV of  $2.2\sigma$  for  $gg\phi$  and  $2.7\sigma$  for  $bb\phi$ . None of these

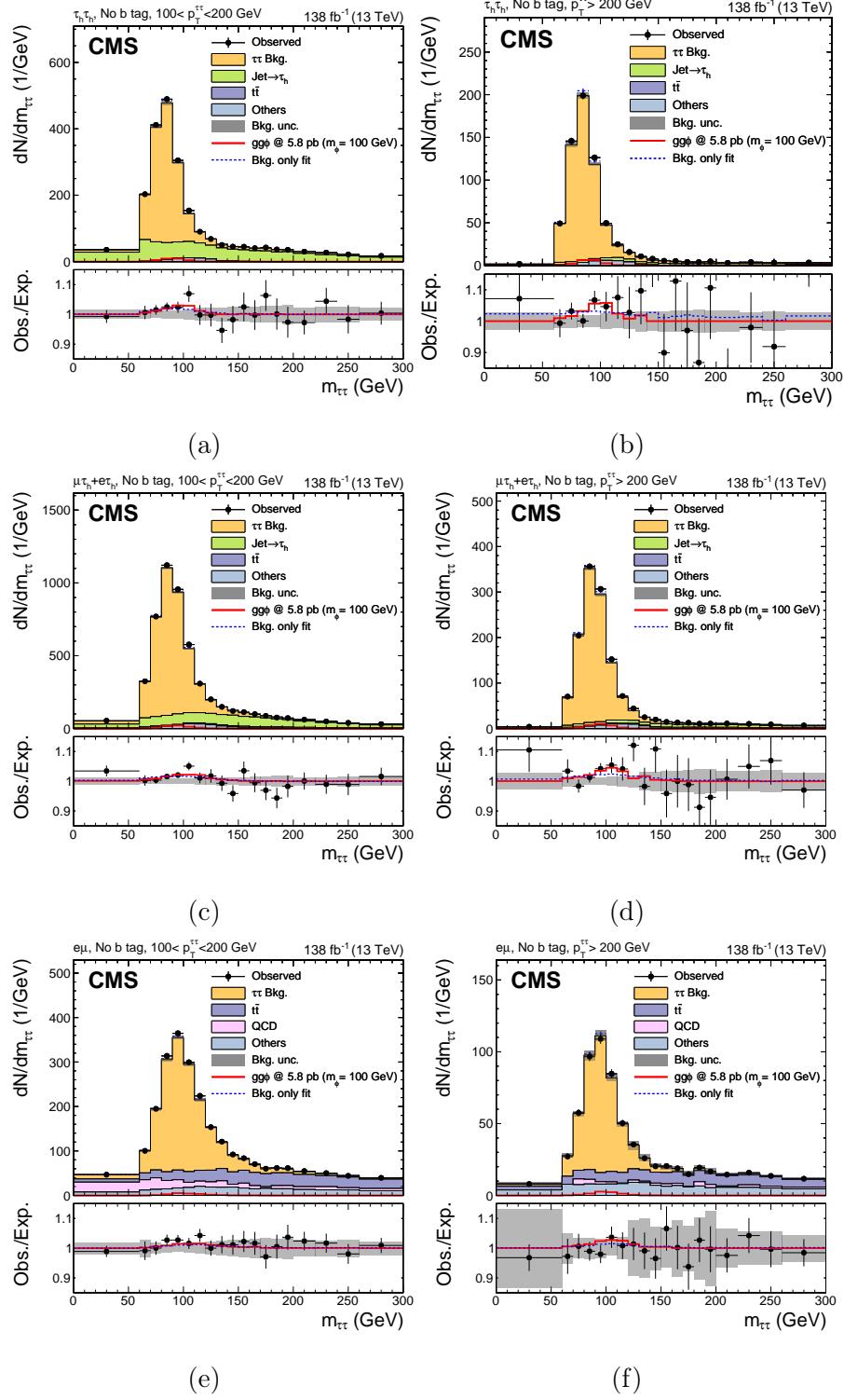


Figure 4.19: Distributions of  $m_{\tau\tau}$  in the no b tag second highest (left) and highest (right)  $p_T$  category for the  $\tau_h\tau_h$  (top), the combined  $e\tau_h$  and  $\mu\tau_h$  (middle) and the  $e\mu$  (bottom) channels. The solid histograms show the stacked background predictions after a signal plus background fit to the data. The best fit gluon fusion signal for  $m_\phi = 100$  GeV is shown by the red line. Also shown by a blue dashed line on the bottom pad is the ratio of the background predictions for the background only fit to the signal plus background fit

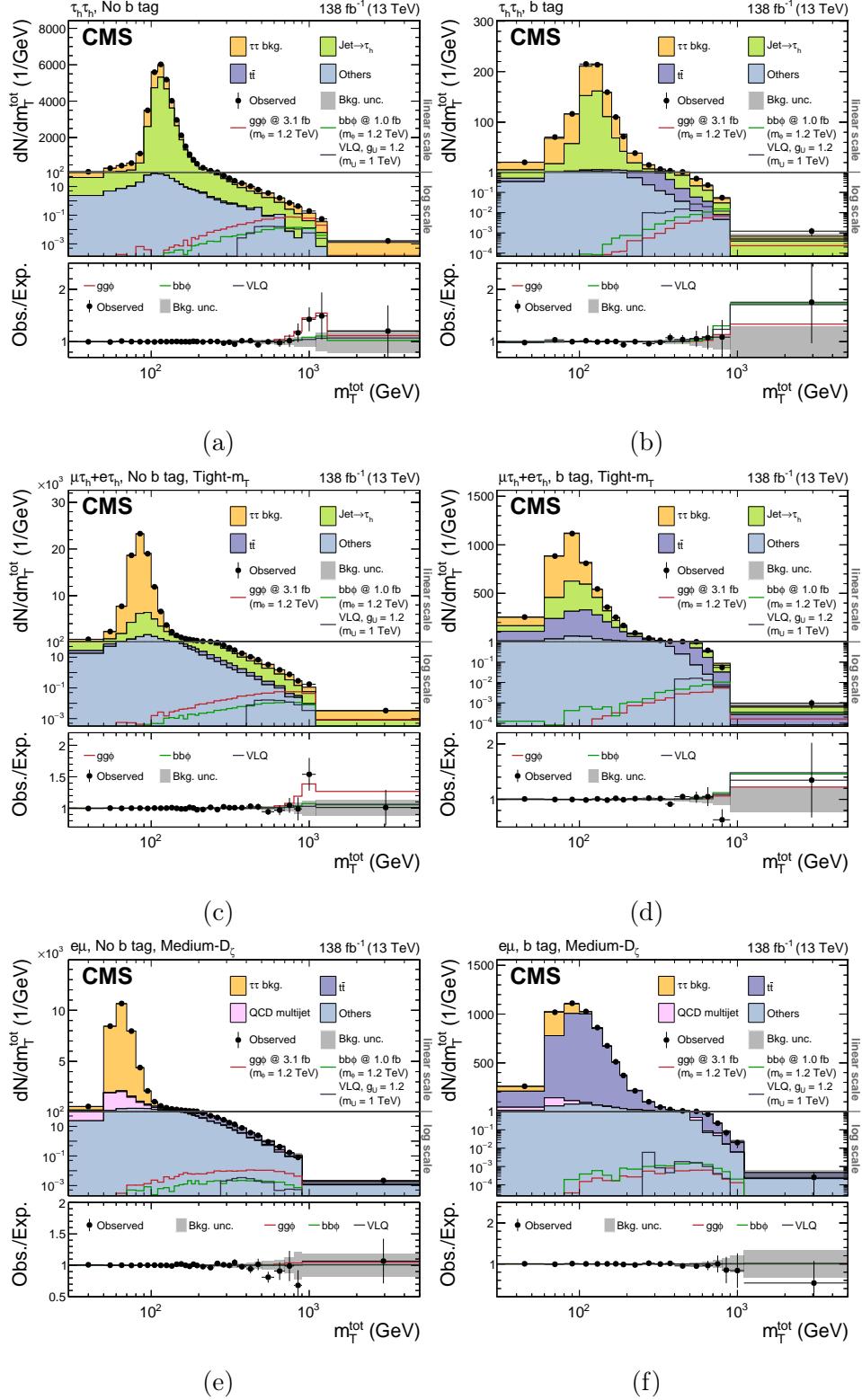


Figure 4.20: Distributions of  $m_T^{\text{tot}}$  in the  $\tau_h\tau_h$  no b tag (a) and b tag (b) categories, the combined  $e\tau_h$  and  $\mu\tau_h$  no b tag (c) and b tag (d) Tight- $m_T$  categories and the  $e\mu$  no b tag (e) and b tag (f) Medium- $D_\zeta$  categories. The solid histograms show the stacked background predictions after a background only fit to the data. The best fit gluon fusion signal for  $m_\phi = 1.2$  TeV is shown by the red line, b associated production and  $U_1$  signals are also shown for illustrative purposes.

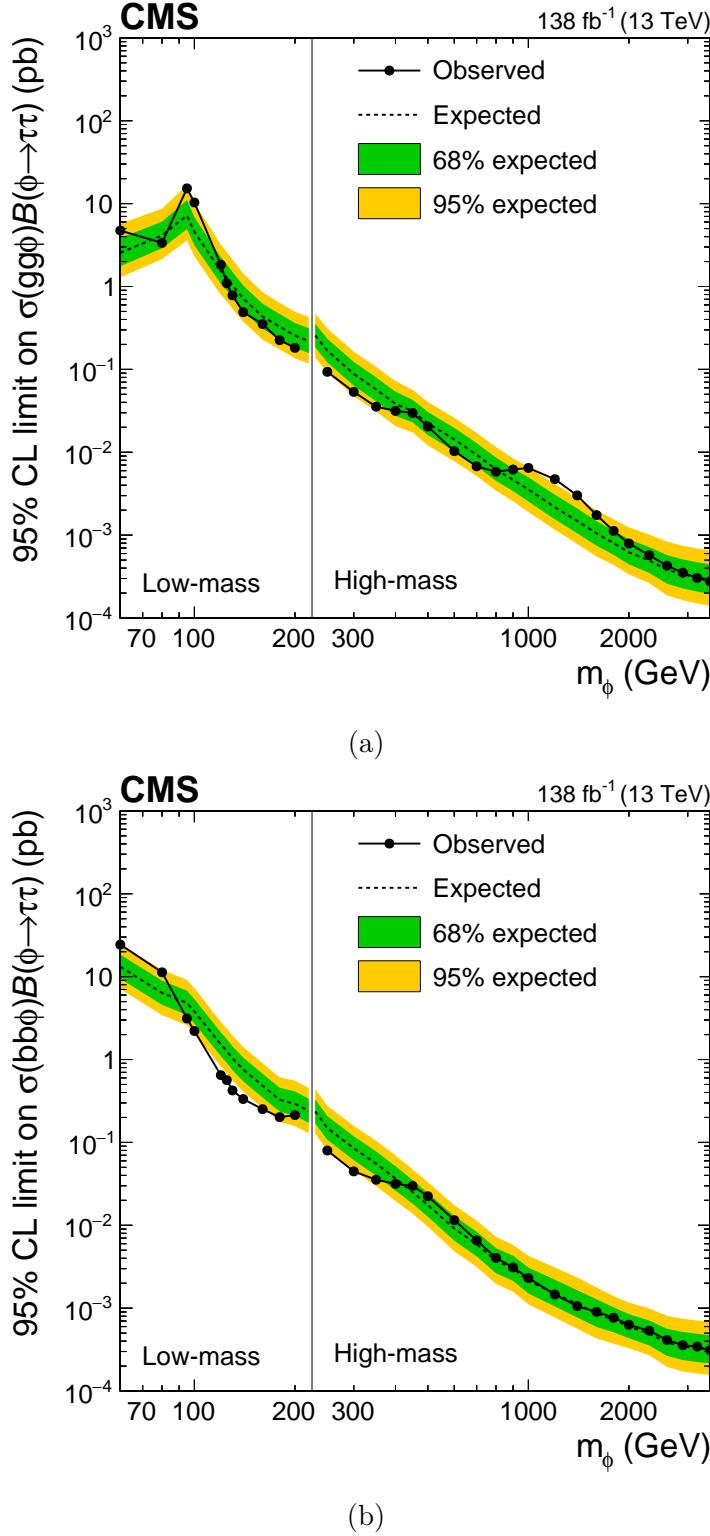


Figure 4.21: Expected (dashed line) and observed (solid line and dots) 95% CL upper limits on the product of the cross sections and branching fraction for the decay into  $\tau$  leptons for (a)  $gg\phi$  and (b)  $bb\phi$  production in a mass range of  $60 \leq m_\phi \leq 3500$  GeV. The dark green and bright yellow bands indicate the central 68% and 95% intervals for the expected exclusion limit.

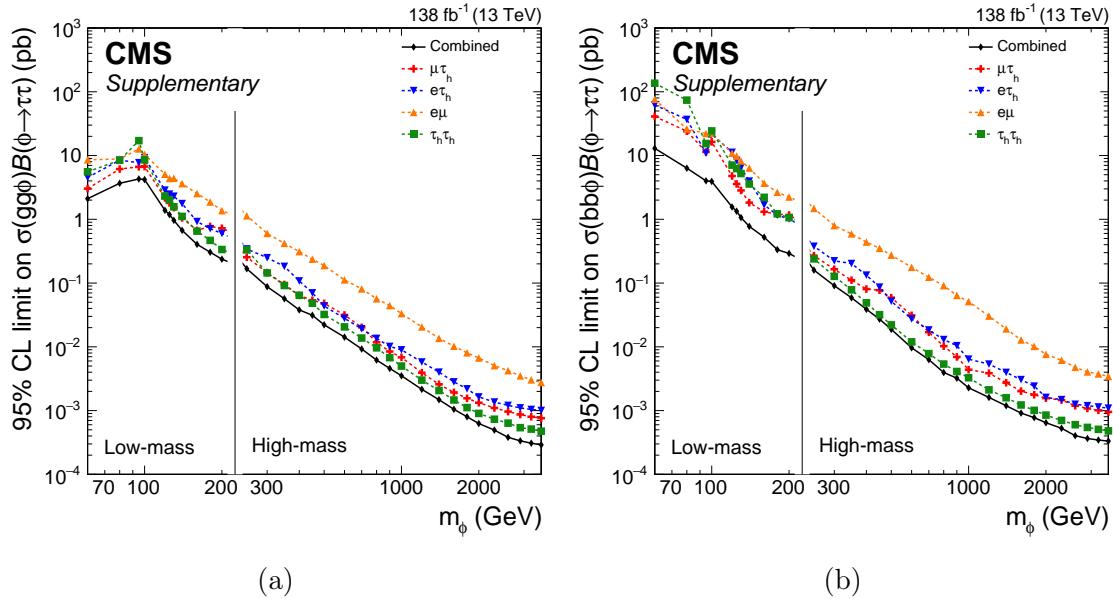


Figure 4.22: Comparison of the expected 95% CL upper limits on the product of the cross sections and branching fraction for the decay into  $\tau$  leptons for (a)  $gg\phi$  and (b)  $bb\phi$  production, split by the  $\tau\tau$  decay products fit individually.

excesses are consistent between the ATLAS and CMS results. The ATLAS search does not stretch to the mass of the low mass CMS excess and so cannot be used a cross-check for this.

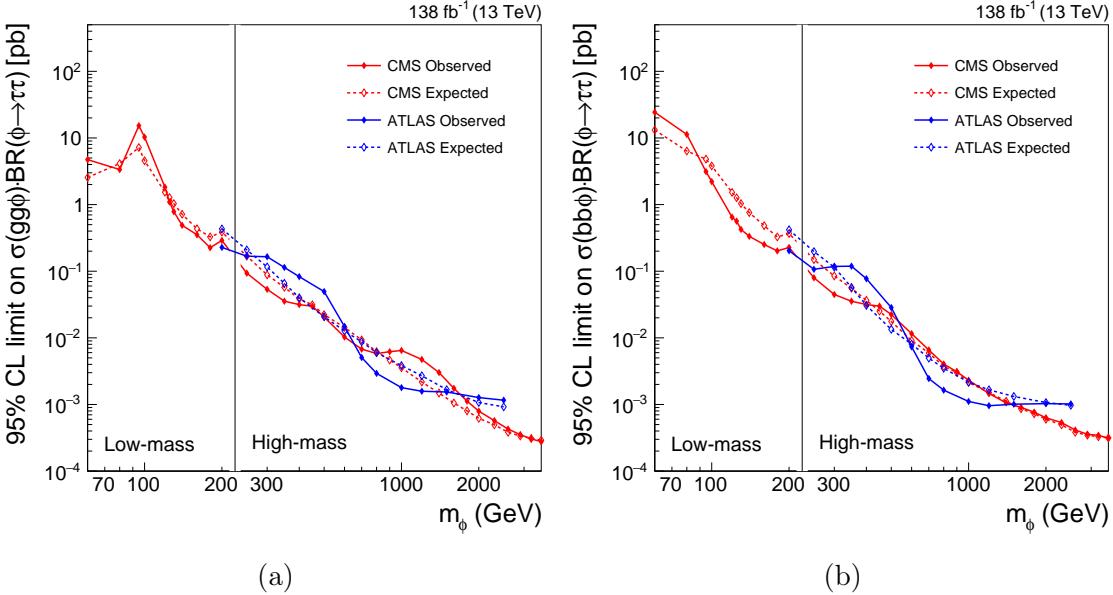


Figure 4.23: Comparison of the expected 95% CL upper limits on the product of the cross sections and branching fraction for the decay into  $\tau$  leptons for (a)  $gg\phi$  and (b)  $bb\phi$  production, split by the CMS result detailed in this thesis and the ATLAS result from Ref. [3].

### 4.12.2 Significance and Compatibility

The  $p$ -values and significances at each model independent signal hypothesis are calculated as described in Section 4.10 and shown in Figure 4.24. Identical to the model independent limits, the  $gg\phi$  or  $bb\phi$  process is allowed to float freely if not the parameter of interest. The excesses for the  $gg\phi$  process peak at 100 GeV and 1.2 TeV and quantify to a local (global) significance of  $3.1\sigma$  ( $2.7\sigma$ ) and  $2.8\sigma$  ( $2.2\sigma$ ) respectively. There are also excesses at neighbouring mass points (particularly at high mass), however this is consistent with the mass resolution of the fitted templates for the central values. No deviations beyond  $2\sigma$  are observed for  $bb\phi$  production.

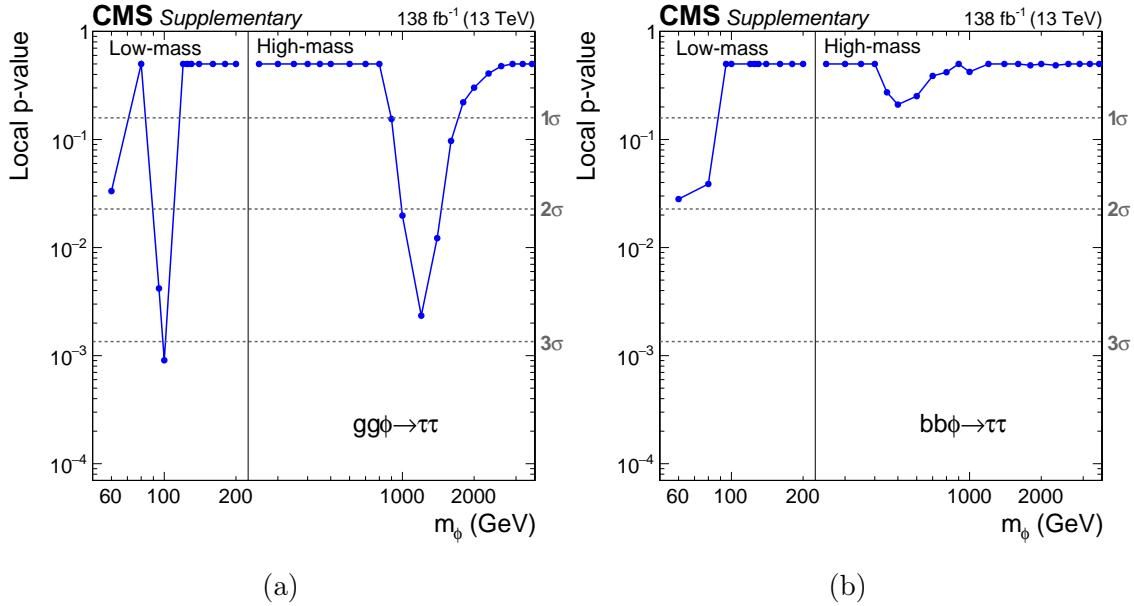


Figure 4.24: Local  $p$ -value and significance of a  $gg\phi$  (left) and  $bb\phi$  (right) signal as a function of  $m_\phi$ .

As many different decay channels and categories are used to extract these significances, the signal strength is studied in each channel and category. This is done via compatibility fits as described in Section 4.10, where the signal strength parameter in each channel/category is decoupled. No statistically significant differences are observed in the best fit signal strength in any decay channel or category fit and  $p$ -values between each channel or category fit are always above 0.05. Figure 4.25 shows the results of the compatibility fits in the low-mass optimisation categories split by di- $\tau$  decay channels and the  $p_T$  bins fit. Figure 4.26 shows the compatibility fits in the high-mass optimisation categories split by di- $\tau$  decay channels. The low-mass signal strengths are no more dominant in any  $p_T$  region than another. In both low-

and high-mass cases, the signal strengths are consistent across di- $\tau$  decay channels. There is a small shift in the high mass  $e\mu$  categories to a negative signal strength, these categories have little to no sensitivity to this signal in comparison to others and a small deficit is observed in data, resulting in fits for a negative signal strength with a large uncertainty.

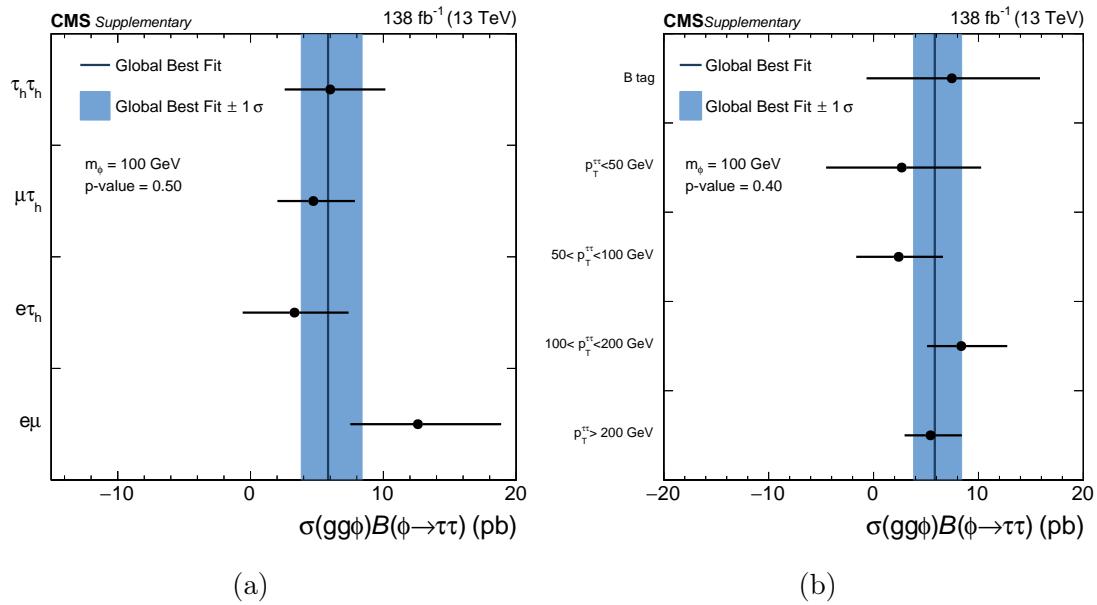


Figure 4.25: Compatibility plots of the 100 GeV excess split into analysis channels (a) and categories (b). In each case the fitted signal strength is decoupled in the bin shown on the plot.

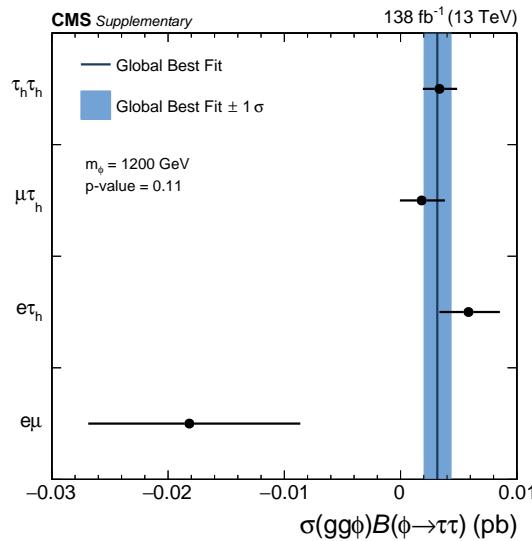


Figure 4.26: Compatibility plots of the 1.2 TeV excess split by analysis channels. In each case the fitted signal strength is decoupled in each channel.

### 4.12.3 2D Likelihood Scans

As the model-independent search looks for two signal modes at each mass point, the results for both processes happening simultaneously are studied. This is done in the form of 2D likelihood scans. The best fit cross section times branching fractions of each process and the 95% and 68% confidence intervals are shown for a number of different mass scenarios in Figure 4.27. The SM prediction in all plots is at (0,0). These results highlight how the excesses at 100 GeV and 1.2 TeV are dominated in the phase space in which  $gg\phi$  and not  $bb\phi$  signals are allowed. In the 60 GeV example, there are smaller deviations in both  $gg\phi$  and  $bb\phi$  and the SM background is again over  $2\sigma$  away. Otherwise, signal strengths are completely compatible with the background expectation.

## 4.13 Model Dependent Limits

The exclusion contours for two benchmark scenarios of the MSSM,  $M_h^{125}$  and  $M_{h,EFT}^{125}$ , are presented in Figure 4.28. The red hatched regions denote areas where  $m_h$  is inconsistent with the observed SM Higgs boson mass within a  $\pm 3$  GeV boundary. For low values of  $\tan \beta$ , higher values of the additional SUSY particle masses, denoted as  $m_{\text{SUSY}}$ , are needed to explain a mass of approximately 125 GeV for the Higgs boson. In the  $M_h^{125}$  scenario,  $m_{\text{SUSY}}$  is fixed, and the predicted value of  $m_h$  is below 122 GeV. In contrast, the  $M_{h,EFT}^{125}$  scenario adjusts  $m_{\text{SUSY}}$  to satisfy the required value of  $m_h$  for each point in  $(m_A, \tan \beta)$  individually, accounting for the logarithmic corrections associated with the large values of  $m_{\text{SUSY}}$  using an effective field theory approach. The red hatched region in Figure 4.28 (b) indicates that the required values of  $m_{\text{SUSY}}$  exceed the GUT scale at very low values of  $m_A$  in this scenario. The Higgs boson masses, mixing angle  $\alpha$ , and effective Yukawa couplings were calculated using FEYNHIGGS, and branching fractions for the decay into  $\tau$  leptons and other final states were obtained from a combination of the FEYNHIGGS and HDECAY, following the prescriptions in Refs.[80, 81, 82], for the scenarios described in Ref.[48].

For the  $M_{h,EFT}^{125}$  scenario, the sensitivity sharply drops at  $m_A = 2m_t$  due to a drop in the branching fractions for the decay of A and H into  $\tau$  leptons, when the A and H decays into two on-shell top quarks becomes kinematically accessible. Both scenarios are excluded at 95% CL for  $m_A \lesssim 350$  GeV. For  $m_A \lesssim 250$  GeV, most of the  $ggH/A$  events do not enter the no b tag categories due to the  $m_{\tau\tau} > 250$  GeV

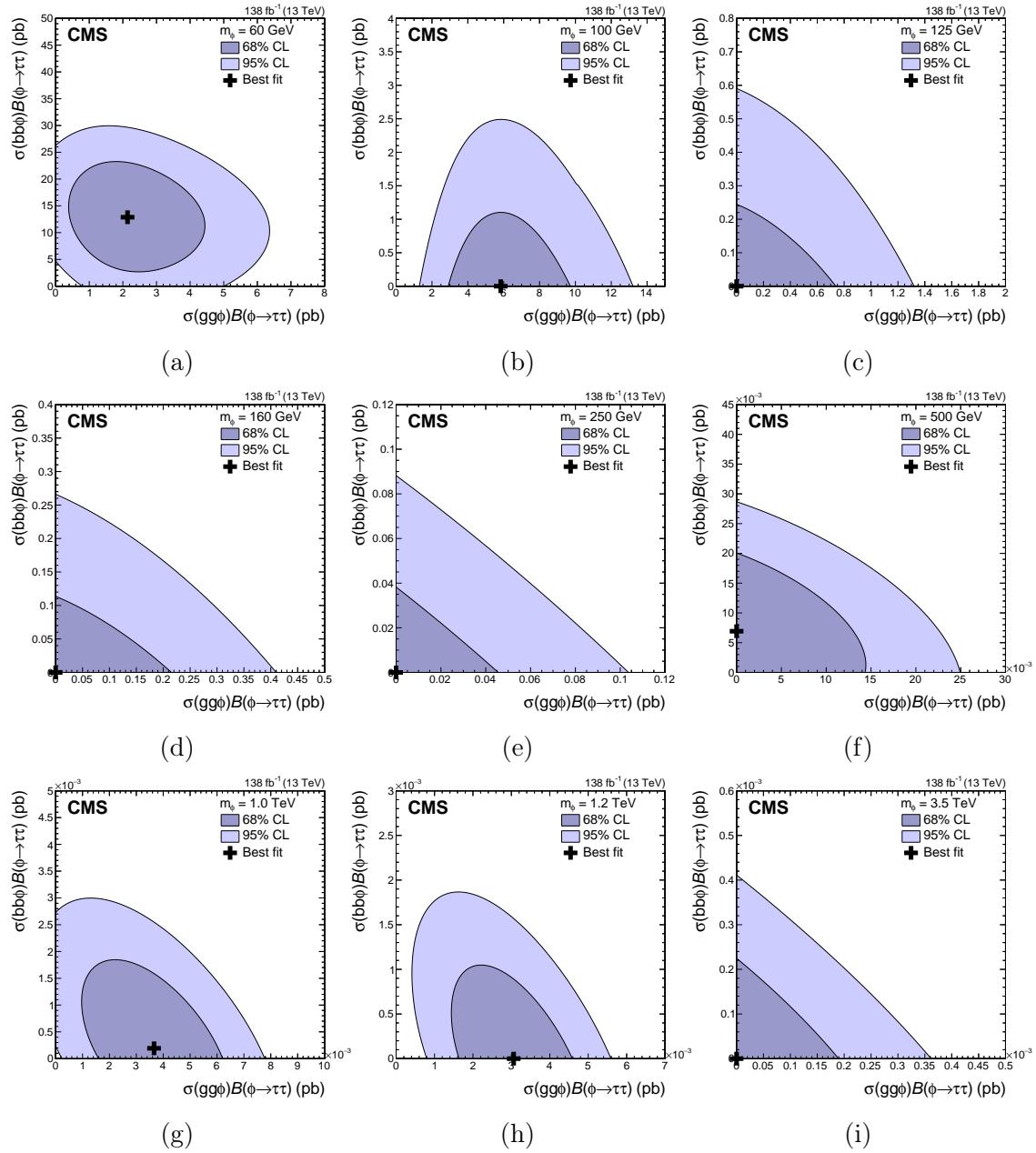


Figure 4.27: Maximum likelihood scans, including 68% and 95% CL contours obtained from the signal likelihood for the model-independent search. The scans are shown for selected values of  $m_\phi$  between 60 GeV and 3.5 TeV.

requirement. In this parameter space, the sensitivity to the MSSM is driven by the measurements of the observed Higgs boson, even though H and A still contribute to the categories here. The sensitivity to the H and A enters mainly via the  $b\bar{b}\phi$  signal in the b tag categories, especially for increasing values of  $\tan\beta$ .

Other MSSM scenarios are tested and detailed in [83]. One scenario of note is the  $M_H^{125}$  scenario, which is the equivalent scenario to the  $M_h^{125}$  but with the observed Higgs boson being the heavier CP-even Higgs boson. Despite the local excess at a resonant mass of 100 GeV, this scenario is entirely excluded by the search. This is mostly due to the sensitivity of the b tag categories to b associated production. The local excess observed at 1.2 TeV is hard to rectify within these MSSM benchmark scenarios. The lack of any excess in the b tag categories strictly constrains the b associated production cross section times branching fraction. It is not possible within these scenarios to predict the excess of gluon fusion events within the constraints placed on b associated production.

Upper limits of 95% CL for VLQ BM 1 and 2 are shown Figure 4.29. These are drawn with respect to the leptoquark mass ( $m_U$ ) and coupling ( $g_U$ ). The limit on  $g_U$  decreases as  $m_U$  increases, with values of  $g_U$  ranging from 1.3 to 5.2 in VLQ BM 1 and 0.8 to 3.2 in VLQ BM 2. VLQ BM 2 has stronger exclusion limits than VLQ BM 1 due to additional right-handed couplings of the leptoquark with a b quark and a  $\tau$  lepton. The observed limits fall within the central 95% intervals of the expected limits when no signal is present. The expected limits are also within the 95% confidence interval of the best fit results reported by Ref.[4] and described in Section ???. This indicates that the search is capable of detecting a part of the parameter space that can explain the anomalies observed in b physics and since no significant excess was observed in this search, new constraints are placed on the vector leptoquark phase space.

Similarly to the MSSM scenarios, the local excess at 1.2 TeV is not consistent with a VLQ BM 1 or 2 vector leptoquark. Again this is due to lack of signal in the b tag categories, where the reduction in backgrounds makes the t-channel signal with initial state radiation the dominant search option.

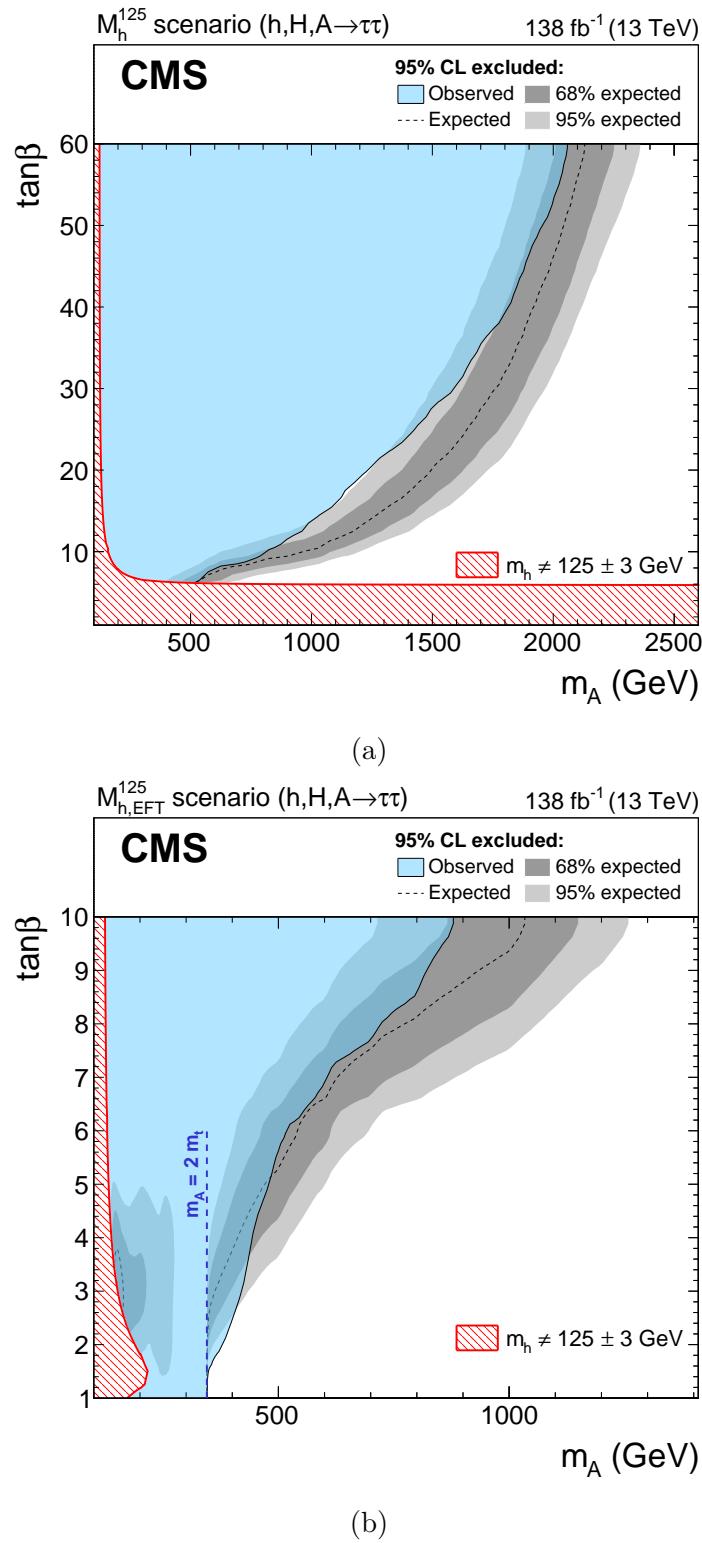


Figure 4.28: Expected and observed 95% CL exclusion contours in the MSSM  $M_h^{125}$  (a) and  $M_{h,EFT}^{125}$  (b) scenarios. The exclusion limit only on background expectation is shown as a dashed black line, the dark and bright grey bands show the 68% and 95% intervals of the expected exclusion and the observed exclusion contour is shown by the blue area. The parameter space where  $m_h$  deviates by more than  $\pm 3$  GeV from the observed SM Higgs boson mass is shown by a red hatched area.

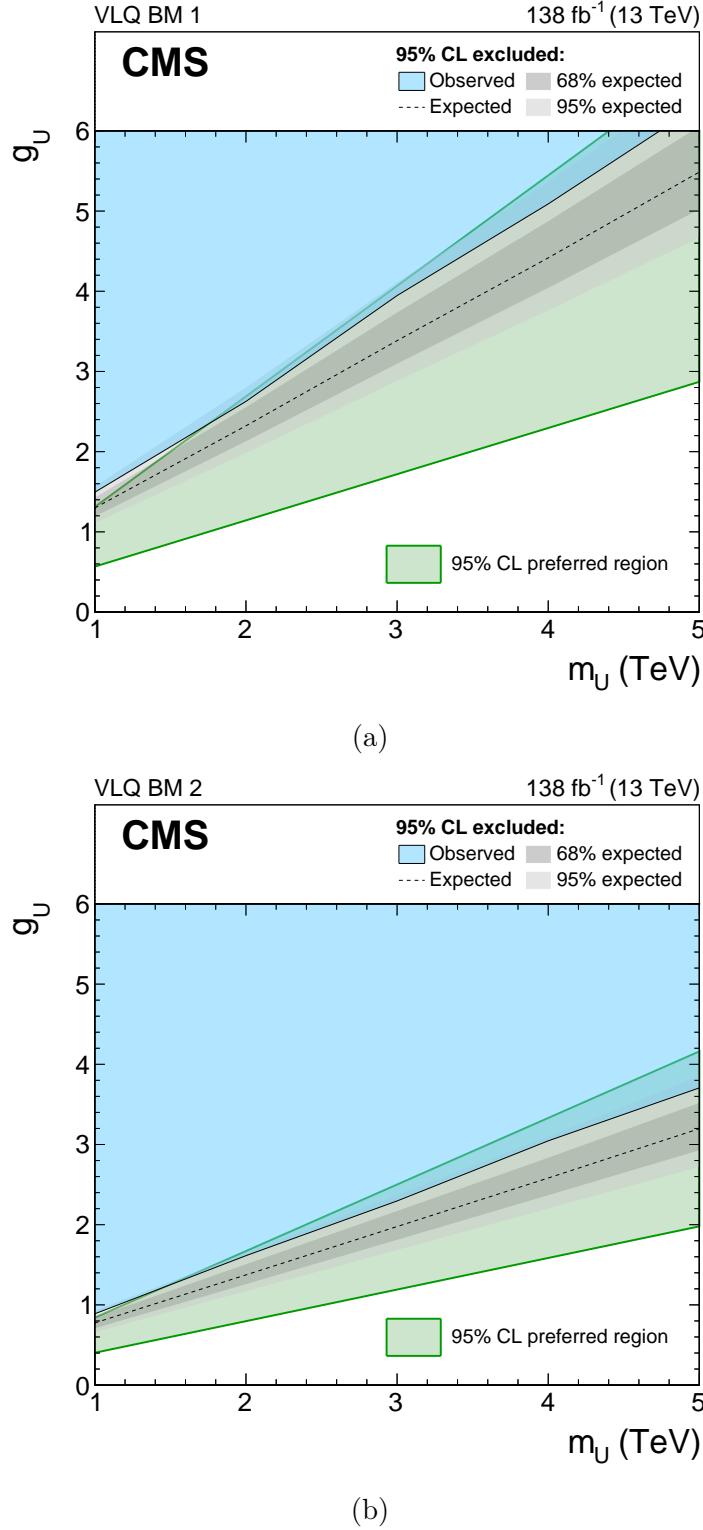


Figure 4.29: Expected and observed 95% CL upper limits on  $g_U$  in the VLQ BM 1 (a) and 2 (b) scenarios, in a mass range of  $1 < m_U < 5$  TeV. The exclusion limit only on background expectation is shown as a dashed black line, the dark and bright grey bands show the 68% and 95% intervals of the expected exclusion and the observed exclusion contour is shown by the blue area. The 95% confidence interval for the preferred region from the global fit presented in Ref. [4] is also shown by the green shaded area.

# Chapter 5

## Search for New Physics in $\tau^+\tau^-\tau^+\tau^-$ Final States

Enhancement from  $\tan\beta$  to up or down-like quark couplings to additional Higgs bosons are essential for the majority of searches for extended Higgs sectors, including in the analysis detailed in Chapter 4. However, in some 2HDMs it can be the case that both up and down-like couplings to additional Higgs bosons are suppressed and this parameter space is left relatively untouched by "MSSM-like" searches. In particular, type X 2HDMs where only lepton couplings are enhanced by  $\tan\beta$ , allow for new physics loop contributions to SM measurements through couplings between leptons and additional Higgs bosons. This is particularly interesting in the context of the g-2 anomaly [] with reasoning explained in Section 1.4.2. This chapter will detail a search for such an extended Higgs sector, that looks for a production mode not suppressed at high  $\tan\beta$ , through the process  $Z^* \rightarrow \phi A \rightarrow 4\tau$ . This search is split up into two sections:

- i) A model independent search for the  $Z^* \rightarrow \phi A \rightarrow 4\tau$  process. Both additional particles are required to have narrow width and no assumptions are made on the production cross-section via an off-shell Z boson or the branching fraction of  $\phi$  and A decaying to a pair of tau leptons.
- ii) A search for the type X 2HDM, motivated by the phase space for possible explanations to the g-2 anomaly. The  $m_A$ - $\tan\beta$  phase space for scenarios of  $m_\phi$  in the alignment limit is scanned, as well as checks outside of this limit on the  $\cos(\beta - \alpha)$ - $\tan\beta$  for specific scenarios of both  $m_\phi$  and  $m_A$ .

These searches are performed with the full run-2 dataset ( $138 \text{ fb}^{-1}$ ) collected by the CMS experiment.

## 5.1 Signal Modelling

Any additional Higgs boson produced in the type X 2HDM at high  $\tan\beta$  will predominantly decay to tau leptons. To probe the type X 2HDM at high  $\tan\beta$ , a production process that is not suppressed is required. Ref. [5] discusses that the following production modes of two additional Higgs bosons are dominant to produce any of these new particles at high  $\tan\beta$ :

- i)  $pp \rightarrow Z^* \rightarrow \phi A \rightarrow (\tau^-\tau^+)(\tau^-\tau^+)$
- ii)  $pp \rightarrow Z^* \rightarrow H^+H^- \rightarrow (\tau^-\nu)(\tau^+\nu)$
- iii)  $pp \rightarrow W^{\pm*} \rightarrow H^\pm A \rightarrow (\tau^\pm\nu)(\tau^-\tau^+)$
- iv)  $pp \rightarrow W^{\pm*} \rightarrow H^\pm\phi \rightarrow (\tau^\pm\nu)(\tau^-\tau^+)$

As the production cross sections are of the similar magnitudes, the search sensitivities depend on the separation of the signals from background. In general, the more objects you can select in the final state, the smaller the background contributions. This is certainly true in tau enriched final states, where backgrounds can be dominated jets misidentified as hadronic taus and so every extra tau selected reduces this background. In particular, (ii) has the production cross sections [] far smaller than the observed limit for gluon fusion production of a single resonance shown in Figure 4.21(a) and it is not possible to use tau decay product and MET alignment to separate the background, so does seem not a viable search option with the run-2 CMS dataset. The increased background from fewer object selections and looser charge sum selection on (iii) and (iv), makes (i) the golden search channel for a type X 2HDM. A Feynman diagram for this process is shown in Figure 5.1.

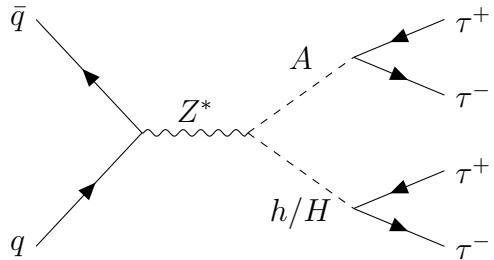


Figure 5.1: Diagram of production of two additional neutral Higgs bosons from an off-shell  $Z$  boson and their decays to tau leptons.

Signal templates for the production of this process with a mass grid for  $\phi$  and A between 100 to 300 and 60 to 160 GeV respectively are generated. These mass ranges are motivated from the results in Table 1.3. The samples are simulated in the five-flavour scheme (5FS) at NLO precision using the MADGRAPH5\_aMC@NLO v2.6.5 event generator [61]. Generation is performed using the parton distribution function (PDF) NNPDF3.1 [55, 56], where  $\tau$  lepton decay, parton showering and hadronisation are all modelled with the PYTHIA event generator with the PU profile matched to data [57, 58]. The events are then passed through the GEANT4-based [59] simulation of the CMS detector and reconstructed in the same way as data. Generator level distributions of di- $\tau$  mass distributions from the decay of  $\phi$  and A are shown in Figure 5.2.

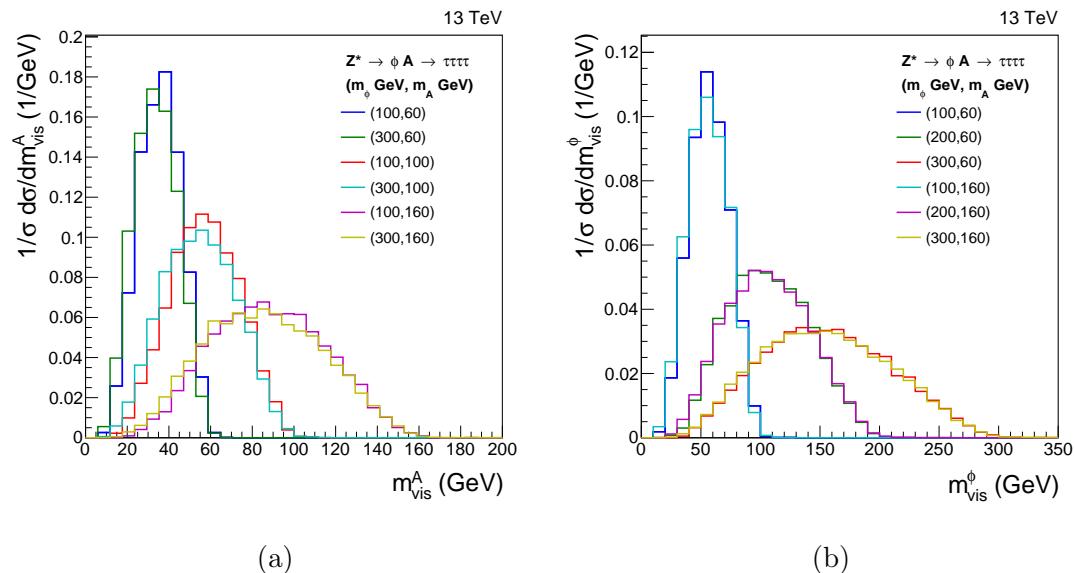


Figure 5.2: Generator level distributions of the visible mass densities for A (a) and  $\phi$  (b) for the signal process.

The cross sections in the alignment scenarios are also determined with this procedure and vary from 10 fb (highest  $m_A$  and  $m_\phi$  scenario) to 650 fb (lowest  $m_A$  and  $m_\phi$  scenario), as shown in Figure 5.3. These are independent of  $\tan\beta$ , however, out of the alignment scenarios the cross sections for H scales with  $\sin^2(\beta - \alpha)$  and h scales with  $\cos^2(\beta - \alpha)$ .

The branching fractions of  $\phi$  and A to pairs of  $\tau$  leptons is dependent on both  $\tan\beta$  and  $\beta - \alpha$ . For this analysis, the branching fractions are calculated using 2HDECAY [84]. In the alignment scenarios, the  $A \rightarrow \tau\tau$  branching fractions are approximately

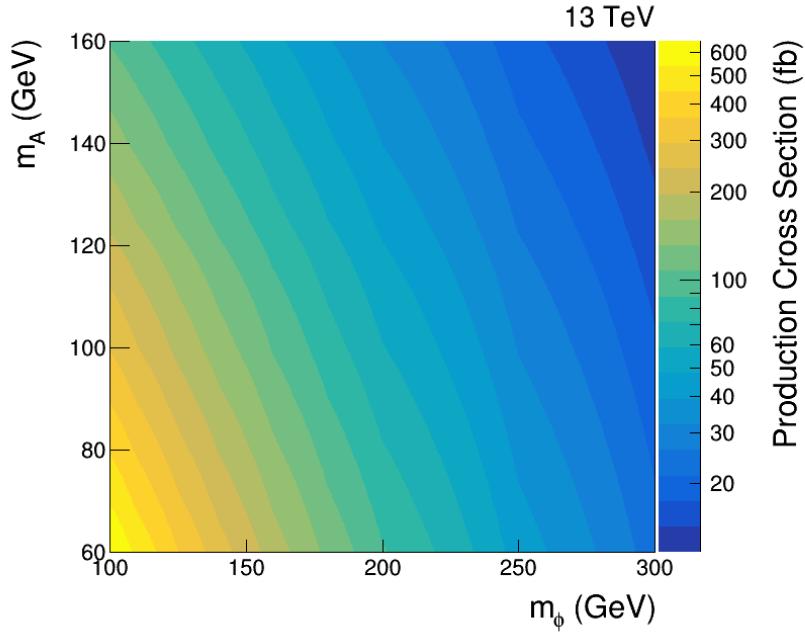


Figure 5.3: Calculated production cross sections for the  $Z^* \rightarrow \phi A$  process, varying the masses of  $\phi$  and  $A$ .

1 above  $\tan\beta \approx 2$ , where below they sharply drop off and other processes such as  $A \rightarrow b\bar{b}$  become dominant. This is also true for  $\phi \rightarrow \tau\tau$  branching fractions, except in the case where  $m_\phi$  is greater than  $m_A$  by more than  $m_Z$  and so the  $\phi \rightarrow ZA$  decay becomes kinematically feasible and can dominate at high  $\tan\beta$ . Examples of this are shown in Figure 5.4. Out of the alignment scenario, the branching fractions of  $\phi$  to tau leptons becomes smaller as the magnitude of the coupling of additional neutral Higgs bosons to taus is reduced, whilst the  $A$  branching fractions, like the couplings, are left unchanged. An example of the  $\phi$  branching fractions out of the alignment scenario is shown in Figure 5.5.

## 5.2 Event Selection

In comparison to Chapter 4, four  $\tau$  leptons produce a much larger number of possible final states. All variations of  $e$ ,  $\mu$  and  $\tau_h$  final state combinations and their branching fractions are shown in Table 5.1. Just under 90% of the branching ratio goes to decay products containing two or more hadronic taus. These are the main final states that are explored in this analysis. In addition to this, an orthogonal  $\tau_h\tau_h\tau_h$  channel is added to target events where reconstruction of the  $\tau_h\tau_h\tau_h\tau_h$  channel loses a single  $\tau_h$  object. This can come about due to low triggering and ID efficiency of  $\tau_h$  candidates,

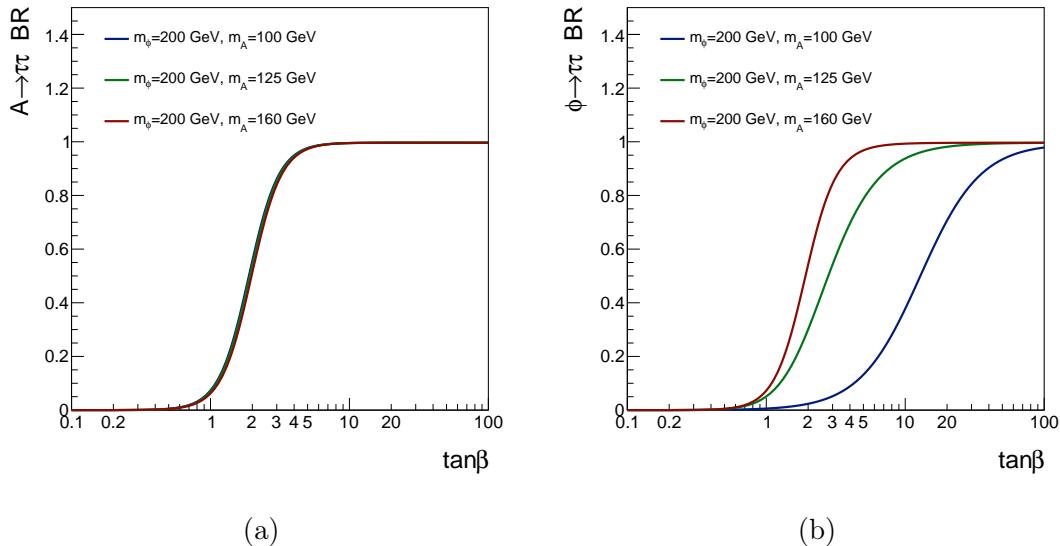


Figure 5.4: Calculated branching fractions of A (a) and  $\phi$  (b) decaying to a pair of  $\tau$  leptons for various mass scenarios.

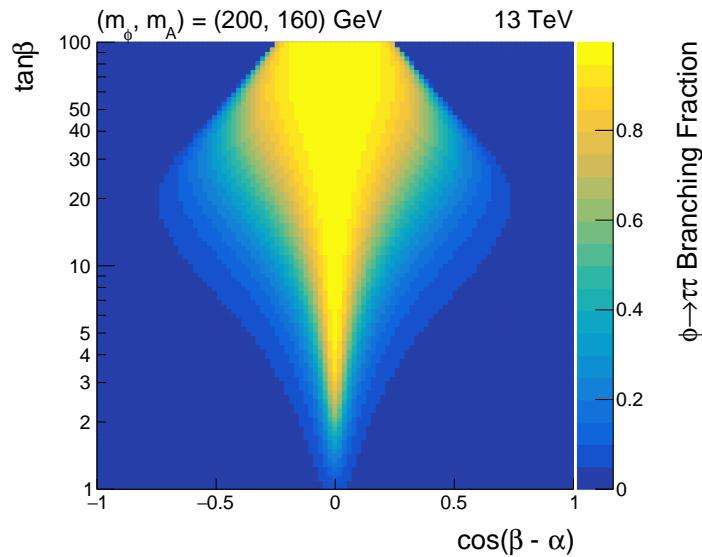


Figure 5.5: Calculated branching fractions in the  $\cos(\beta - \alpha)$ - $\tan\beta$  phase space for  $\phi$  of mass 200 GeV decaying to a pair of  $\tau$  leptons, in the scenario where  $m_A = 160$  GeV.

as well as the high  $p_T$  thresholds required for both.

In total, the analysis consists of 7 channels.

Channel	Branching Fraction
$e\tau_h\tau_h\tau_h$	19.4%
$\mu\tau_h\tau_h\tau_h$	18.9%
$\tau_h\tau_h\tau_h\tau_h$	17.6%
$e\mu\tau_h\tau_h$	15.6%
$ee\tau_h\tau_h$	8.0%
$\mu\mu\tau_h\tau_h$	7.6%
$ee\mu\tau_h$	4.3%
$e\mu\mu\tau_h$	4.2%
$eee\tau_h$	1.5%
$\mu\mu\mu\tau_h$	1.4%
$eee\mu$	1.4%
$ee\mu\mu$	0.6%
$e\mu\mu\mu$	0.4%
$eeee$	0.1%
$\mu\mu\mu\mu$	0.1%

Table 5.1: Branching fractions of four  $\tau$  leptons, where  $e$  and  $\mu$  represent the leptonic decay of the  $\tau$  and  $\tau_h$  represent the hadronic decay of the  $\tau$ .

### 5.2.1 Trigger Requirements

Given that each final state is not exactly triggered on, there is no obvious choice for what triggers to use. A variety of triggers are available for individual and clusters of objects in the final state. The possible triggers for single objects are the single- $e$  and single- $\mu$  triggers. This is not the case for the single- $\tau_h$  trigger as it has a  $p_T$  threshold too high for it to be useful. The possible triggers for clusters of objects are the double- $e$ , double- $\mu$ , double- $\tau_h$ ,  $e$ - $\mu$  cross,  $\mu$ - $\tau_h$  cross and  $e$ - $\tau_h$  cross-triggers. The cross-triggers and double- $e/\mu$  triggers are found to offer little improvement to the signal acceptance and so these events are not included. Any combination of objects in the final state of a channel can be selected by these triggers and the union of events passing each iteration is taken. The trigger  $p_T$  and  $\eta$  thresholds for the remaining triggers (single- $e/\mu$  and double- $\tau_h$ ) are equivalent to what is stated in Section 4.2.1.

### 5.2.2 Offline Requirements

All offline selections stated are in addition the object selection discussed in Section 3. In this analysis,  $\tau_h$  candidates are required to pass the **Loose**  $D_{\text{jet}}^{\text{WP}}$ . The **VVLoose**  $D_e^{\text{WP}}$  and **VLoose**  $D_\mu^{\text{WP}}$  are used in all decay channels. These working points are chosen to maximise the sensitivity of the analysis and looser cuts are used than in Chapter 4.1.1 to ensure there are enough statistics to account for the stricter selection of more objects in the final states.

On top of the object selection, there are few further selections on the total tau collection. As the two tau pairs from the signals originate from two neutral additional Higgs bosons, the sum of charge of the fully reconstructed objects should be 0. This is applied in the decay channels where there are four objects. In the  $\tau_h\tau_h\tau_h\tau_h$  channel, as the assumption is that a hadronic tau has been lost, the absolute value of the sum of the charges of the object is required to be 1. In order to ensure the orthogonality between channels, a number of vetos are needed. Firstly, extra lepton vetos are used in all decay channels, where the absence of additional electrons and muons is required on top of those already part of the selected pair. The kinematic requirements, ID and isolation requirements on these extra leptons are at least as loose as the loosest cuts required for signal electron or muon of the final states. Secondly, an extra  $\tau_h$  veto is applied to the  $\tau_h\tau_h\tau_h\tau_h$  channel, in order to keep it orthogonal to the  $\tau_h\tau_h\tau_h\tau_h$  channel. Here any extra  $\tau_h$  candidate passing the signal selection is vetoed. The constraints set on the number of leptons and  $\tau_h$  candidates for each channel are shown in Table 5.2.

Channel	e	$\mu$	$\tau_h$
$\tau_h\tau_h\tau_h\tau_h$	0	0	-
$\tau_h\tau_h\tau_h$	0	0	3
$\mu\tau_h\tau_h\tau_h$	0	1	-
$e\tau_h\tau_h\tau_h$	1	0	-
$e\mu\tau_h\tau_h$	1	1	-
$\mu\mu\tau_h\tau_h$	0	2	-
$ee\tau_h\tau_h$	2	0	-

Table 5.2: Veto set on the number of leptons for each channel.

Finally, in channels containing an electron or a muon, a veto on events with one or more b tagged jets is placed. This is done to remove the  $t\bar{t}$  background process in a region where little to no signal is expected. This is not done in the fully  $\tau_h$  channels as the  $t\bar{t}$  process is negligible.

### 5.3 Search Optimisation

Due to the limited statistics in each decay channel, only minimal categorisation of events can be performed and the only divisions of the decay channels happen in final states with two light leptons and two  $\tau_h$  candidates, where two categories are made. These separate events where the light leptons have the same charge (**SS Leptons**) and where they have opposite charge (**OS Leptons**). This is motivated to separate regions where specific background processes are dominant. In particular, a large portion of the  $e\mu\tau_h\tau_h$  and  $\mu\mu\tau_h\tau_h$  channel backgrounds come from  $Z \rightarrow ll$  (or  $Z \rightarrow \tau\tau \rightarrow ll$ ) with 2 jets misidentified as  $\tau_h$  candidates, where the two light leptons are of opposite sign. Similarly, the  $e\mu\tau_h\tau_h$  channel contains more background events with an opposite sign electron and muon, from a Z decay via taus or from the  $t\bar{t}$  process. There is no preference for light lepton charge sums in the signal samples and so more sensitivity to the signal is expected in the **SS Lepton** categories. A summary of the categorisation and b tag selection is shown in Figure 5.6.

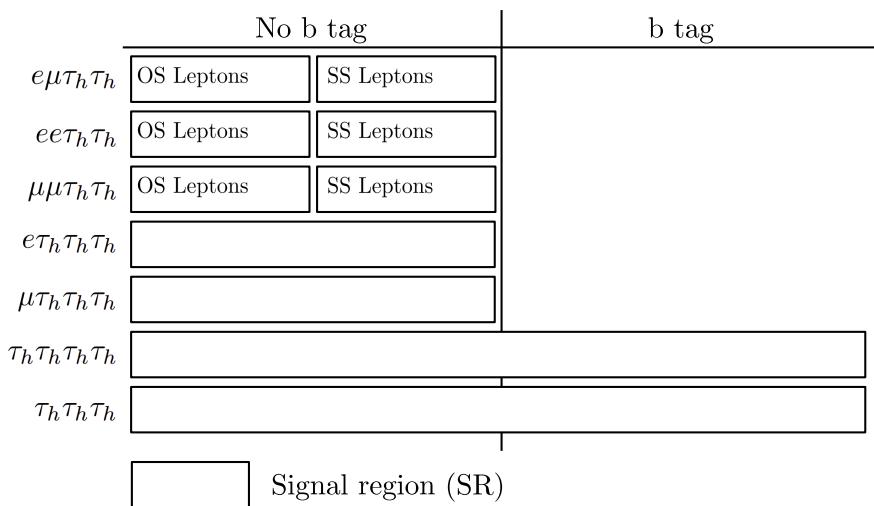


Figure 5.6: Overview of the categories used for the extraction of the signal in the  $Z^* \rightarrow \phi A \rightarrow 4\tau$ .

Similarly to Chapter 4.1.1, events in each 10 analysis channels and categories are drawn out into histogram based off the same discriminating variable, the total transverse mass ( $m_T^{\text{tot}}$ ). However, as there are more objects in the final state, the definition is extended to what is shown below.

$$m_T^{\text{tot}} = \sqrt{\sum_{i=1}^{N_\tau} m_T(\vec{p}_T^{\tau_i}, \vec{p}_T^{\text{miss}})^2 + \sum_{i,j=1; i \neq j}^{N_\tau} m_T(\vec{p}_T^{\tau_i}, \vec{p}_T^{\tau_j})^2}, \quad (5.1)$$

where  $N_\tau$  is the number of objects in the final state, and  $\tau_i$  refers to the visible products of the  $i$ 'th  $\tau$  lepton. This variable again provides excellent discriminating power between resonant signals compared to other non-peaking backgrounds, whilst still maintaining some separation between signal masses. The sensitivity of this discriminating variable was tested against many others, including the visible masses of the two bosons which can be separated to  $\approx 90\%$  efficiency after selection into the correct pairs. As  $m_T^{\text{tot}}$  uses information about the whole event, in particular the higher  $p_T$  objects and high MET from many  $\tau$  decays, a greater sensitivity to the signal is observed.

In this analysis, many of the histograms contain very few background events and in order to minimise statistical fluctuations in the background templates, histogram bins are merged based off the fractional statistical uncertainty of each bin. This leads to a number of single and minimally binned channels/categories where background statistics are low but signal sensitivity is high and a few finely binned and high statistic channels/categories able to better classify the signal, if one is observed.

## 5.4 Background Modelling Overview

The backgrounds are split into a three categories:

- i) Events containing only genuine  $\tau$  leptons.
- ii) Events with one or more jet misidentified as a  $\tau_h$  (jet  $\rightarrow \tau_h$ ).
- iii) Events where all  $\tau_h$  candidates are either correctly identified or a misidentified light lepton.

Background contributions from (i) are mostly from gluon and quark initiated di-Z production and is modelled using MC. The details and validation of this modelling is described in Section 5.5. Background (ii) accounts for a number a different process

with jets misidentified as a  $\tau_h$  candidate. Examples of this are single-Z, single-W and  $t\bar{t}$  productions with additional jets in the events being misidentified. This is modelled with a fake factor method, similar to what is described in Section 4.7, but using machine learning to improve upon method, and discussed further in Section 5.6. The latter contribution is very small in comparison to the others and modelled with MC. Events with jets misidentified as light leptons and no jets misidentified as  $\tau_h$  candidates have been checked with MC and deemed negligible in all channels.

All MC background samples described in Section 4.4 are modelled in the same manner. In addition, tri-boson samples are used that are generated using MADGRAPH5\_aMC@NLO at next-to-LO (NLO) precision [61]. All corrections described in Section 4.8 are applied to MC, with extra corrections are applied on the ZZ $\rightarrow$ 4l process, that are detailed in Section 5.5.

## 5.5 ZZ Modelling

The di-Z background shapes are modelled using MC. The cross-sections for this analysis are scaled to high order precisions than the signal generation using k factors. For quark initiated di-Z production, PowHEG [] is used to determine NNLO/NLO QCD and EW corrections as a function of  $m_{ZZ}$ , which takes an average value of  $\approx 1.2$ . Gluon initiated di-Z production corrections from LO to NNLO is derived with HNNLO v2 [], again as a function of  $m_{ZZ}$ . This gives a much larger correction, equal to  $\approx 2$ . The modelling is validation using a  $\mu\mu\mu\mu$  final state and good agreement is observed between data and simulation. Identical object selection is applied as stated in Section 3, except to ensure better statistics within the channel, the  $I_{\text{rel}}^\mu$  cut is loosened to 0.35. The charges of the  $\mu$  candidates are similarly required to sum to one and no vetos on the number of b jets in the event is applied. Plots of this are shown in Figures 5.7.

## 5.6 Machine Learning Fake Factor Method

The fake factor method, as described in Section 4.7, is an algorithm used to model the jet $\rightarrow\tau_h$  backgrounds. It uses classical reweighting techniques, such as binning the two regions and applying a fit to the ratio in a chosen parametrisation. This method faces difficulties when approached with highly dimensional dependence on

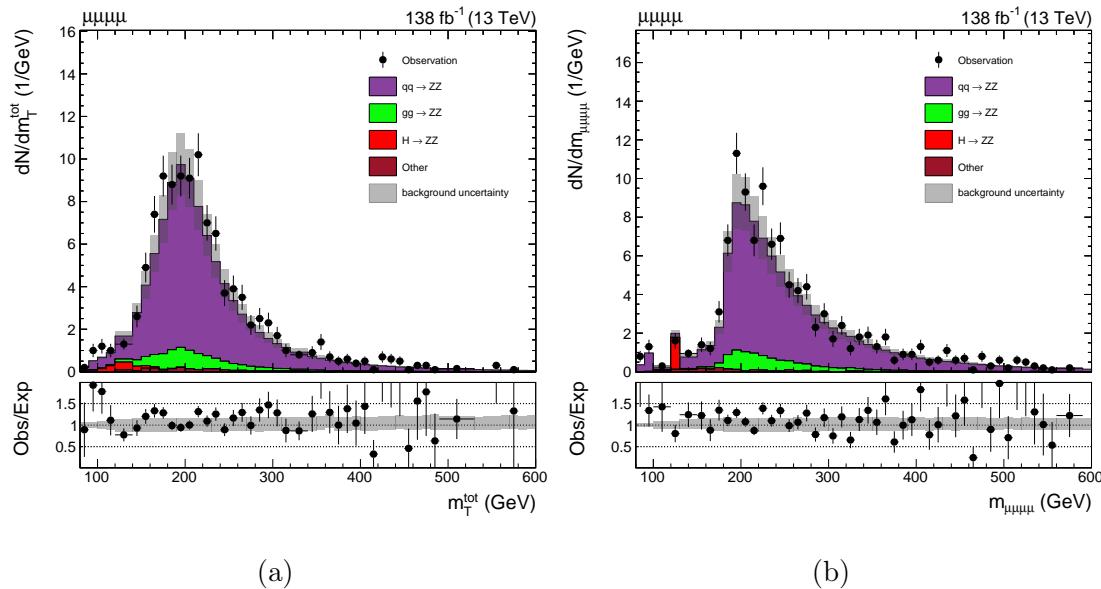


Figure 5.7: Distributions in the  $\mu\mu\mu\mu$  channel are shown for the variables total transverse mass,  $m_T^{\text{tot}}$ , (a) and the total mass,  $m_{\mu\mu\mu\mu}$  (b). The solid histograms show the stacked background predictions.

the parametrisation. In this scenario, classical "bin and fit" methods can break down, as for every new parameter included the statistics of each fit is reduced. A compromise must then be reached between variable dependence and fit statistics and is the case for the jet- $p_T/\tau_h$ - $p_T$  binning described in Section 4.7.2. It is also the case, that the analysis in Section 4.1.1 uses well checked assumptions on for example, the consistency of  $\tau_h$  decay modes from determination to signal regions, in order to average across the dependencies on this variable when calculating fake factors, however this will not always be the case. Also binned upon in the previous search, is the process dependence of the fake factors. The binning used is an admission that the relevant parametrisation of the fake factors between process is too complication to model with "bin and fit" methods. It can also be seen that the initial fits, do not do a good job at modelling all variables as corrections are needed.

There is final problem with using a fake factor method as described, when there are multiple  $\tau_h$  candidates in final states. In the  $\tau_h\tau_h$  channel of the previous analysis, the fake factors were only calculated from the leading- $\tau_h$  candidate. This was valid, as the dominant backgrounds had two jet  $\rightarrow \tau_h$  candidates rather than one genuine  $\tau_h$  and one jet  $\rightarrow \tau_h$  candidate. However, for this search this assumption is not valid.

All of these reason motivate a more generalised and smarter method to model jet

$\rightarrow \tau_h$  backgrounds, as well as the extra importance on modelling these backgrounds where there are more  $\tau_h$  candidates in the final state. This is done utilising machine learning and in particular using a BDT for the purpose of multi-dimensional reweighting.

### 5.6.1 BDT Reweighting

Ref. [85] proposes a new method utilising machine learning techniques to solve the issues with dimensionality. It looks to optimise the regions that most need reweighting. One good way to do this is using a decision tree, as with this the data can be split into "leafs" by checking simple conditions. To best choose the regions that need reweighting the algorithm looks to maximise the symmetrised  $\chi^2$ .

$$\chi^2 = \sum_{\text{leaf}} \frac{(w_{\text{leaf}, 1} - w_{\text{leaf}, 2})^2}{(w_{\text{leaf}, 1} + w_{\text{leaf}, 2})^2} \quad (5.2)$$

The larger the value of  $\chi^2$  the more important reweighting is in this region. The kind of tree is utilised many times in the reweighting algorithm shown below:

- i) Input training dataset with a large number of variables.
- ii) Build a tree as stated above. If not the first loop, use newly determined weights.
- iii) Compute predictions in the leaves  $r_{\text{leaf}} = \log \frac{w_{\text{leaf}, 1}}{w_{\text{leaf}, 2}}$ . The logarithm is taken so weights in different trees can be summed as usually done in boosting.
- iv) In each leaf MC events by  $w = w \times e^{r_{\text{leaf}}}$ .

The final two steps are identical to the first approach except for the use of the logarithm for convenience using boosting. The major difference being how the bins used for reweighting are found, and that this step is repeated multiple times.

### 5.6.2 Fitting Regions

Unlike the standard fake factor method, statistics are not a problem using the BDT reweighter when choosing which variables you can use to parametrise the fake factors. Therefore, rather than fitting two fake factor regions separately and then a correction from sideband to signal to account for missing parametrisation, regions

A,B and D from Figure 4.10 are fit simultaneously to improve statistics. Also, all  $\tau_h$  candidates are fit simultaneously as separate entries in the dataset and for each  $\tau_h$  in the event, the remaining  $\tau_h$  candidates are named the alternative  $\tau_h$  candidates, rather than the  $\tau_h$  that is being fit. The sideband variable definitions and cuts, with respect to Figure 4.10, are shown below.

- i)  $ee\tau_h\tau_h$ ,  $e\mu\tau_h\tau_h$ ,  $\mu\mu\tau_h\tau_h$ ,  $e\tau_h\tau_h\tau_h$  and  $\mu\tau_h\tau_h\tau_h$

$y_C$ : The sum of  $\tau_h$  candidates is required to be 0.

$y_A$ : The sum of  $\tau_h$  candidates is required to not be 0.

$x_C$ : All alternative  $\tau_h$  candidates pass the **Loose**  $D_{\text{jet}}^{\text{WP}}$ .

$x_D$ : At least one alternative  $\tau_h$  candidate fails the **Loose**  $D_{\text{jet}}^{\text{WP}}$  but has  $D_{\text{jet}}^{\text{score}} > 0.1$ .

- ii)  $\tau_h\tau_h\tau_h$

$y_C$ : The absolute value of sum of  $\tau_h$  candidates is required to be 1.

$y_A$ : The absolute value of the sum of  $\tau_h$  candidates is required to not be 1.

$x_C$ : All alternative  $\tau_h$  candidates pass the **Loose**  $D_{\text{jet}}^{\text{WP}}$ .

$x_D$ : At least one alternative  $\tau_h$  candidate fails the **Loose**  $D_{\text{jet}}^{\text{WP}}$  but has  $D_{\text{jet}}^{\text{score}} > 0.1$ .

These selections mimic what is done for the  $\tau_h\tau_h$  channel fake factors from Section 4.7.1, where the definitions of the alternative sideband variable are extended to more than one other  $\tau_h$  candidate. The  $D_{\text{jet}}^{\text{score}} > 0.1$  selection is used as the alternative tau ID selection for this analysis and is also then extended to the alternative  $\tau_h$  candidates for this selection. A cut on the score is used instead of a working point as the loosest defined working point does not provide enough statistics to ensure a good fit.

Extrapolating the regions used for fitting to the  $\tau_h\tau_h\tau_h\tau_h$  channel, there would be some overlap in the fitting region with the  $\tau_h\tau_h\tau_h$  signal region. As this region is expected to be sensitive to signal, this is not used to model jet  $\rightarrow \tau_h$  backgrounds in the  $\tau_h\tau_h\tau_h\tau_h$  channel. Instead the fit from the  $\tau_h\tau_h\tau_h$  channel is used. There are a few variables that are defined differently in the fit between the two channels due to the difference of four to three objects selected, therefore when getting fake factors in the  $\tau_h\tau_h\tau_h\tau_h$ , the lowest  $D_{\text{jet}}^{\text{score}}$  unused  $\tau_h$  candidate is dropped and variables recalculated. This candidate is chosen to be removed to best mimic the initial selection

of the  $\tau_h$  candidates where they are sorted by  $D_{\text{jet}}^{\text{score}}$  and the high scoring candidates are chosen. As shown later in this section, there is no major dependence on the shifted variables and any effect from this removal is covered within the uncertainty model.

### 5.6.3 Variables Used

The variables used are variables that have been shown to have fake factor dependence previously CITE SOME STUFF. Also added are the variables that take you from A to C and D to C, from Figure ???. As all years are fit together, in order to account for any differences in fake fake factor from year to year, this is also added. The final variable added is the  $p_T$  ordered ranking of the  $\tau_h$  candidates in the event. All these variables are shown below.

- The HPS decay mode of the hadronic tau candidate.
- $p_T$  of the hadronic tau candidate.
- The ratio of the  $p_T$  of the matched jet to the  $p_T$  of the hadronic tau candidate.
- $\eta$  of the hadronic tau candidate.
- The charge of the hadronic tau candidate.
- A boolean of whether the hadronic tau candidate passes a leg of the double tau trigger.
- The total charge of the combined objects.
- The boolean of whether  $D_{\text{jet}}^{\text{WP}}$  passes to Loose WP for the other hadronic tau candidates. These are sorted by  $p_T$ .
- Year of data taking.
- $\tau_h$   $p_T$  ordered event rank.

### 5.6.4 Machine Learning Subtraction Method

In the standard fake factor method, histograms are used to fit the fake factors rather than datasets. Using histograms, the small fraction of events which are not jet fakes

can easily be subtracted off. To do this, the data histogram is subtracted from by a stacked MC background produced with generator matching ensuring the event is not a jet  $\rightarrow \tau_h$  and this produces a data-MC hybrid histogram of predicted jet  $\rightarrow \tau_h$  events. However, subtraction is not possible with a full dataset and negative weights do not work with the BDT reweighter. Therefore, the only option is to remove like-for-like events in data compared to the non jet  $\rightarrow \tau_h$  generator matched MC. An example of this is template matching, which takes an event and can find the closest event in another datasets. But as the fitting dataset is highly dimensional, this requires too much computation. The solution proposed for this is to use a BDT to reduce the dimensionality of the datasets effectively the one dimension of an output score of a simple binary classifier.

- i) In each channel, all MC in the fitting region is stacked and scaled to cross section (via weighting) and the variables used for reweighting (Section ??) are put into a dataset.
- ii) Events that are jet fakes and not jet fakes are separated into the two classes that will be used for the binary classification.
- iii) To ensure unbiased training, the weights of the two categories are normalised to one another.
- iv) A BDT is then trained to separate whether the MC is jet  $\rightarrow \tau_h$  candidate or not.
- v) The scores of the BDT for how likely the entry is not a jet  $\rightarrow \tau_h$  candidate is added to the dataset.
- vi) The scores of the non jet  $\rightarrow \tau_h$  candidates are drawn into a histogram with a number of bins suitable for the number of statistics and rescaled to cross section to best match what would be observed in data.
- vii) The output score of the BDT is added to data events
- viii) Each bin of the MC histogram is then looped through:
  - Data events with BDT score within the range of the bin are selected.
  - Events within this bin are then randomly sample and removed.

- This stops when the number of events removed equals to the number of non jet fake events predicted in the MC histogram bin.

This method then gives a data-MC hybrid method to determine a dataset of predicted jet  $\rightarrow \tau_h$  candidate and should give near identical results when drawing out histograms in all variables compared to subtracting off MC non jet  $\rightarrow \tau_h$  candidates from a data histogram as used in the original fake factor method. It allows for non jet  $\rightarrow \tau_h$ -like candidates to be sampled and removed to the correct yield through all variables. It is important to remove events throughout the MC non jet  $\rightarrow \tau_h$  BDT score histogram as if only the highest score events were removed, these events would come primarily from the tails of the distribution where it is easiest to separate.

An uncertainty is placed on the performance of this algorithm with respect to histogram subtraction. This is calculated by drawing each variable used for the method into a histogram with binning chosen to ensure a sensible number of events in each bin. An uncertainty is then derived from the difference in prediction between the histogram and BDT subtraction.

To validate this method, example histograms with this uncertainty are shown for the  $\mu\tau_h\tau_h\tau_h$  pass region in Figures 5.8, comparing the BDT subtraction method and histogram subtraction method for a few of the fitted variables.

### 5.6.5 Fitting

These subtracted datasets are then randomly split 50:50 into a train and test datasets and only the train dataset is fit. The BDT reweighter has a number of hyperparameters and these are tuned with a scan optimising the Kolomogrov-Smirnoff test  $\|$  on the test dataset in each channel separately. Final models are then produced that can optimally model jet  $\rightarrow \tau_h$  candidates for all fitted variables. Examples of the fake factor derived in the  $\mu\tau_h\tau_h\tau_h$  channel are shown in Figure 5.9.

An uncertainty is placed on the performance of this algorithm. This is again calculated by drawing each variable into a histogram and comparing the histograms from reweighted events with the alternative  $\tau_h$  ID selections to the events with the nominal  $\tau_h$  ID in all of the fitted regions simultaneously. Plots of the closure of this method in the  $\mu\tau_h\tau_h\tau_h$  accompanied by this uncertainty are shown in Figure 5.10.

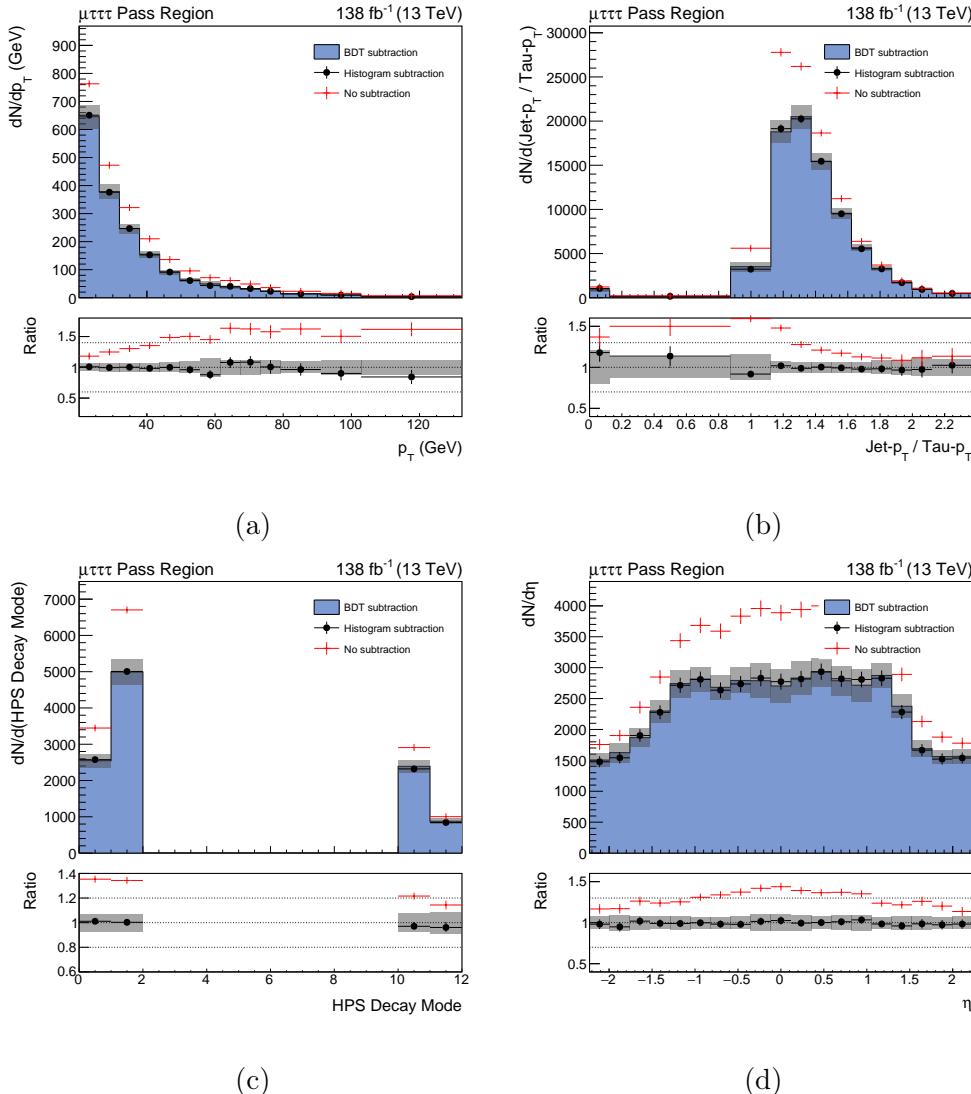


Figure 5.8: Comparison of histograms produced via the BDT subtraction method to histogram subtraction. Also shown is the histogram produced when no subtraction is performed. The uncertainty bands contain statistical uncertainties and uncertainties derived on the non-closure of the method. The figure is shown for four of the fitted variables:  $\tau_h$ - $p_T$ , the ratio of  $\tau_h$ - $p_T$  to jet- $p_T$ , the  $\tau_h$  HPS decay mode and the  $\tau_h$ - $\eta$ .

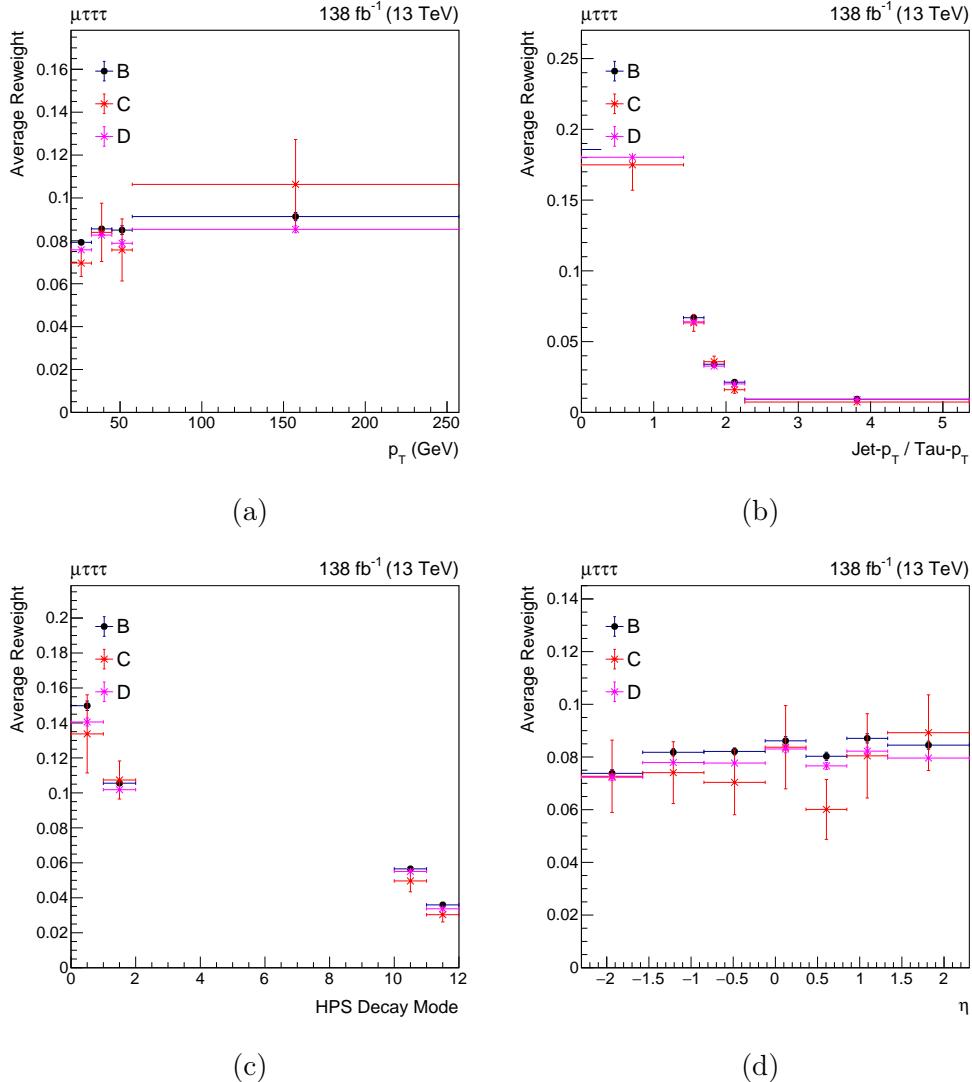


Figure 5.9: Average fake factors (reweights) calculated by the BDT reweighting method shown individually in regions A, B and D NEED TO CHANGE THE REGIONS as defined in Figure 4.10. The is shown for four of the fitted variable:  $\tau_h$ - $p_T$ , the ratio of  $\tau_h$ - $p_T$  to jet- $p_T$ , the  $\tau_h$  HPS decay mode and the  $\tau_h$ - $\eta$ .

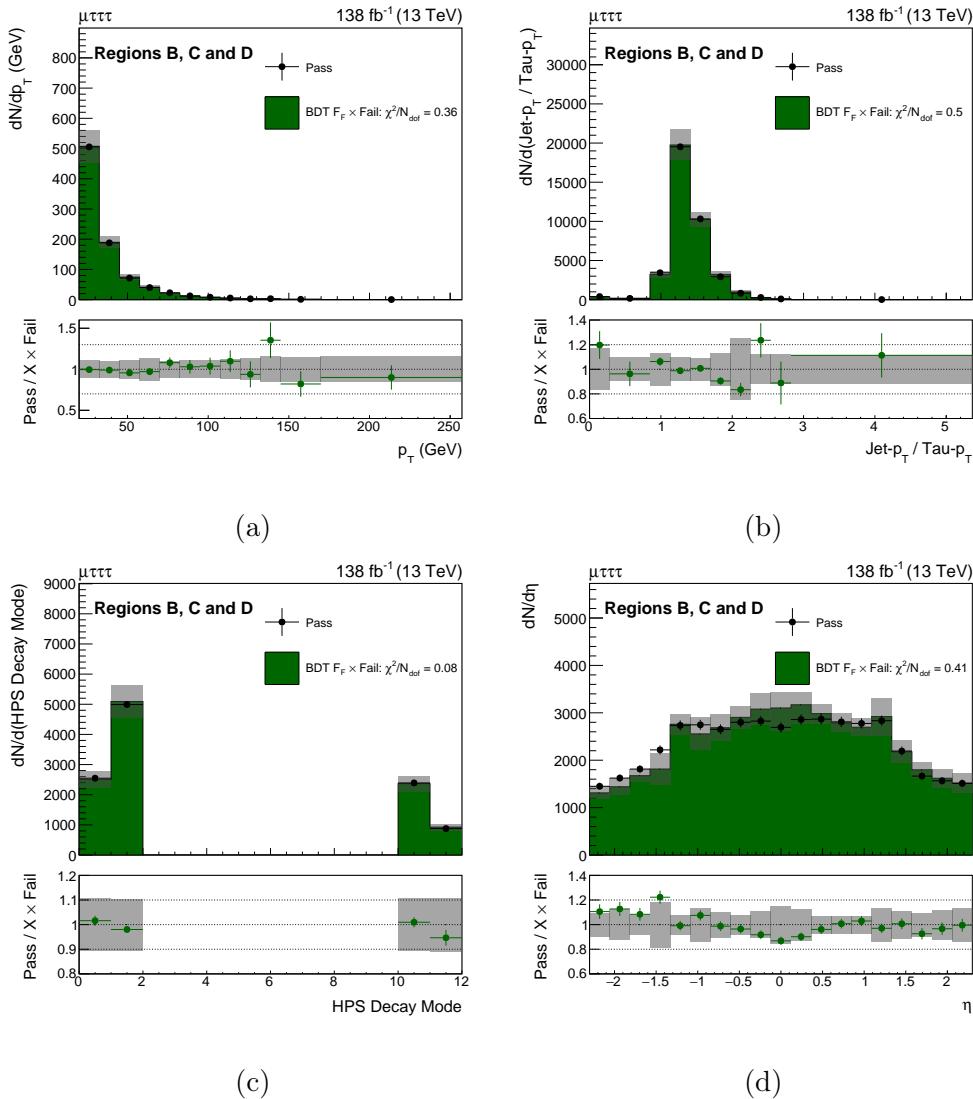


Figure 5.10: Comparison of histograms produced using the fake factors applied to the fitted fail region compared to fitted pass region. The uncertainty bands contain statistical uncertainties and uncertainties derived on the non-closure of the method. This is shown for three of the fitted variables:  $\tau_h$ - $p_T$ , the ratio of  $\tau_h$ - $p_T$  to jet- $p_T$ , the  $\tau_h$  HPS decay mode and the  $\tau_h$ - $\eta$ . The  $\chi^2$  divided by number of degrees of freedom between the two histograms is also shown.

### 5.6.6 Applying Fake Factors

Fake factors,  $F_F^i$ , have now been calculated for each  $\tau_h$  candidate,  $i$ , in the event and uncertainties determined on each weight. However, it is difficult to generate a full description of any number of jet  $\rightarrow \tau_h$  fakes in an event. Taking channels with two  $\tau_h$  candidates as the simplest example, if the jet  $\rightarrow \tau_h$  background is determined purely off the leading  $\tau_h$  candidate, then events the leading  $\tau_h$  is genuine and the sub-leading  $\tau_h$  is a jet, are missed. Similarly, the situation can be flipped if the sub-leading  $\tau_h$  is chosen to determine the jet  $\rightarrow \tau_h$  fake background. A third options can be tried where both  $\tau_h$  candidates are used to determine the background. However, this will only model events where both  $\tau_h$  candidates are jets and not where there is a single jet  $\rightarrow \tau_h$  fake. It is seen that if the first two attempts at calculating this background from individual candidates are added and the contribution from both  $\tau_h$  candidates is subtracted, all possible numbers of jet fakes in the event are accounted for.

Region	$\tau_h^1(\tau)\tau_h^2(j)$	$\tau_h^1(j)\tau_h^2(\tau)$	$\tau_h^1(j)\tau_h^2(j)$
$R_1$ (from $\tau_h^1$ )	0	1	1
$R_2$ (from $\tau_h^2$ )	1	0	1
$R_{12}$ (from both)	0	0	1
$R_1 + R_2 - R_{12}$	1	1	1

Table 5.3: Regions modelled by the fake factor method when using the leading  $\tau_h$  ( $\tau_h^1$ ), the sub-leading  $\tau_h$  ( $\tau_h^2$ ) and both, to model the jet  $\rightarrow \tau_h$  background and whether that specific object is a genuine  $\tau$  lepton ( $\tau$ ) or a jet ( $j$ ). Also shown is a combination of three regions to fully model all possible combinations.

This logic is extended to all channels with one exception. The  $\tau_h\tau_h\tau_h\tau_h$  channel application region would have overlap with the  $\tau_h\tau_h\tau_h$  signal region using this common method, therefore the formula is adjusted to avoid this region. If the overlapped regions are removed from the equation, the scale of the triple and quadruple regions needed to be adjusted to account for this. The caveat to this is that events where there are only one jet  $\rightarrow \tau_h$  candidate are not accounted for. As the majority of the background events in this channel comes from QCD with many jet  $\rightarrow \tau_h$  candidates, this contribution is deemed negligible. The formulae for the total jet  $\rightarrow \tau_h$  backgrounds in each channel are shown below.

i)  $\mu\mu\tau_h\tau_h$ ,  $ee\tau_h\tau_h$  and  $e\mu\tau_h\tau_h$

$$R_1 + R_2 - R_{12} \quad (5.3)$$

ii)  $\mu\tau_h\tau_h\tau_h$ ,  $e\tau_h\tau_h\tau_h$  and  $e\tau_h\tau_h\tau_h$

$$R_1 + R_2 + R_3 - R_{12} - R_{13} - R_{23} + R_{234} \quad (5.4)$$

iii)  $\tau_h\tau_h\tau_h\tau_h$

$$\begin{aligned} & R_{12} + R_{13} + R_{14} + R_{23} + R_{24} + R_{34} \\ & - 2(R_{123} + R_{124} + R_{134} + R_{234}) + 3R_{1234} \end{aligned} \quad (5.5)$$

## 5.7 Uncertainty Model

The uncertainty model follows the schemes detailed in Section 4.9 for the statistical uncertainties and systematic uncertainties for light leptons, jets, leptons misidentified as hadronic taus, MET, luminosity and prefiring. Updates and additions to the previous uncertainty model are shown below.

### Hadronic Taus

An improvement is made to the  $\tau_h$  ID uncertainty correlation scheme applied to simulated events. The fit used to derive scale factors for the  $\tau_h$  MC events consists of both statistical and systematic uncertainties. In the previous search, the ID uncertainties were correlated across decay mode and era despite the statistical components of each fit being orthogonal. Therefore for this search the uncertainty scheme contains correlated and decorrelated part across decay and era. The double- $\tau_h$  trigger uncertainties remain unchanged.

### Jets misidentified as hadronic taus

The backgrounds with jets misidentified as  $\tau_h$  are estimated from data with the ML fake factor method. There are different sources of uncertainty related to this method. All uncertainties are uncorrelated across decay channels, except in the  $\tau_h\tau_h\tau_h\tau_h$  and  $\tau_h\tau_h\tau_h\tau_h$  where the same fit is used, so uncertainties are correlated.

The initial uncertainties come from removal of the non jet $\rightarrow\tau_h$  backgrounds from the fitting region and this is split into two types. Firstly, an uncertainty is placed on the non-closure of the BDT subtraction method, this is done by comparing the distributions in each variable fit using standard histogram subtraction with the BDT subtraction method, and the largest shift is taken for each event. This is done separately in the pass and fail  $\tau_h$  ID regions. The second uncertainty to do with purify the fitting region is motivated by any MC mismodelling of the non jet $\rightarrow\tau_h$  objects predicted. This is shifted up and down by 10%, to represent any MC mismodelling , and the BDT subtraction method and the reweighting is repeated with the differing datasets.

The second source of uncertainties comes from the BDT reweighter fit. In a similar way to the subtraction method, an uncertainty is placed on the non closure of the fit, comparing reweighted events in the fail  $\tau_h$  ID region to the pass  $\tau_h$  ID region. Further uncertainties are placed on the assumption of the variables used to separate the signal region from the fitting region. The assumption is that the fake factors would be identical no matter what combination of these variables are used. Therefore, the uncertainties are placed by taking the largest shift when changing these variables whilst getting the output to the fit. These are decorrelated in each combination of the separating variables.

### **Background process specific uncertainties**

Specific uncertainties are placed on the di-Z simulated events due to application of k factors. Due to the size of the differences between NNLO and lower order predictions for the cross sections, an uncertainty is placed on the size of these yields utilising the k factors derived.

### **Signal process specific uncertainties**

PDF,  $\alpha_s$  and  $\mu_R/\mu_F$  scale variations are applied on an event-by-event basis to the signal samples. The normalisation of these uncertainties are approximately 6%, 1% and 2% respectively.

## 5.8 Signal Extraction

The statistical interpretation of the results is done as described in Section 4.10, but with different signal scaling functions,  $g$ , and parameters of interest,  $\mu$ . The two interpretation of the analysis, as stated at the beginning of chapter, have different parameters of interest and scaling functions.

Firstly, the model independent search uses a linear scaling function,  $g(\mu) = (\mu)$ , and a parameter of interest that represents the cross section of the  $Z^* \rightarrow \phi A$  multiplied by the branching fractions of  $\phi \rightarrow \tau\tau$  and  $A \rightarrow \tau\tau$ . Secondly, the interpretation in the type X 2HDM model is done by testing each point in the parameter space, whether that is  $m_A \cdot \tan \beta$  for the alignment scenario or  $\cos(\beta - \alpha) \cdot \tan \beta$  for scenarios of the remaining parameters. This is done by scaling the samples to the predicted cross sections times branching ratios at that point in the parameter space, and defining a single rate parameter that can only take the values 1 (type X 2HDM) and 0 (SM) with  $g(\mu) = \mu$ .

### 5.8.1 Postfit Plots

Tab. ?? shows the unblinded event yields in each analysis category, alongside the number of bins fit, the background prediction and uncertainties split in statistical and systematic contributions after a background-only fit and a signal yield prediction. Fig. 5.11 show the unblinded distributions in channels where more than one bin is fit. SIGNAL HYPOTHESIS PLOT

DISCUSS RESULTS

## 5.9 Model Independent Results

## 5.10 Model Dependent Limits

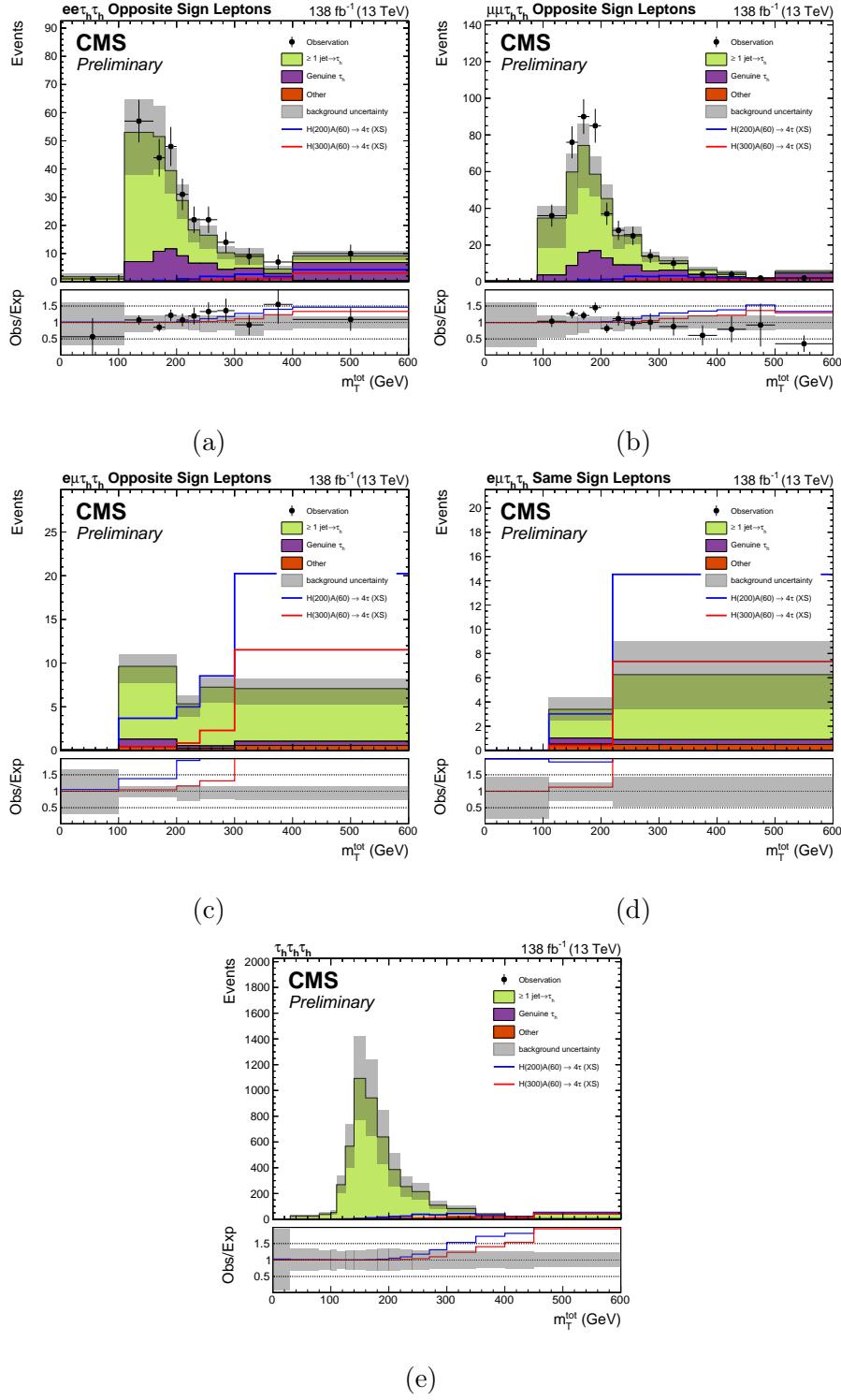


Figure 5.11: Distributions of  $m_T^{\text{tot}}$  in the  $ee\tau_h\tau_h$  OS Leptons (a) and  $\mu\mu\tau_h\tau_h$  OS Leptons (b),  $e\mu\tau_h\tau_h$  OS Leptons (c),  $e\mu\tau_h\tau_h$  SS Leptons (d) and  $\tau_h\tau_h\tau_h$  channel and categories. The solid histograms show the stacked background predictions after a background only fit to the data. The  $m_\phi = 100$  GeV and  $m_A = 100$  GeV signal is also shown for illustrative purposes.

Category	$N_{\text{bins}}$	Total Background	Data Observed	$(m_A, m_\phi) = (160, 300)$ GeV
$\tau_h \tau_h \tau_h \tau_h$	1	$10.9 \pm 3.5 \text{ (stat)} {}^{+1.9}_{-2.0} \text{ (syst)}$	—	26.1
$\tau_h \tau_h \tau_h$	17	$4790.0 \pm 85.7 \text{ (stat)} {}^{+1417.2}_{-1416.0} \text{ (syst)}$	—	94.5
$e \tau_h \tau_h \tau_h$	1	$21.6 \pm 4.2 \text{ (stat)} {}^{+8.6}_{-8.6} \text{ (syst)}$	—	26.8
$\mu \tau_h \tau_h \tau_h$	1	$22.0 \pm 3.5 \text{ (stat)} {}^{+3.3}_{-3.4} \text{ (syst)}$	—	30.7
$ee \tau_h \tau_h$ Opposite Sign Leptons	11	$243.1 \pm 11.1 \text{ (stat)} {}^{+41.7}_{-49.7} \text{ (syst)}$	265	7.0
$ee \tau_h \tau_h$ Same Sign Leptons	1	$3.6 \pm 1.1 \text{ (stat)} {}^{+0.5}_{-0.5} \text{ (syst)}$	—	3.7
$\mu\mu \tau_h \tau_h$ Opposite Sign Leptons	14	$369.0 \pm 13.6 \text{ (stat)} {}^{+52.6}_{-93.5} \text{ (syst)}$	413	9.4
$\mu\mu \tau_h \tau_h$ Same Sign Leptons	1	$4.8 \pm 1.7 \text{ (stat)} {}^{+0.8}_{-1.0} \text{ (syst)}$	—	4.6
$e\mu \tau_h \tau_h$ Opposite Sign Leptons	5	$29.4 \pm 2.6 \text{ (stat)} {}^{+1.3}_{-3.6} \text{ (syst)}$	—	16.9
$e\mu \tau_h \tau_h$ Same Sign Leptons	3	$9.7 \pm 1.4 \text{ (stat)} {}^{+3.1}_{-3.3} \text{ (syst)}$	—	8.5

Table 5.4: Table of background, data and signal integrals, for the  $m_\phi = 100$  GeV and  $m_A = 100$  GeV scenario, in all analysis categories after a background only fit to the data. The background uncertainties are separated into their statistical and systematic and shown along side the background integral. The number of bins used in the fit are also shown.

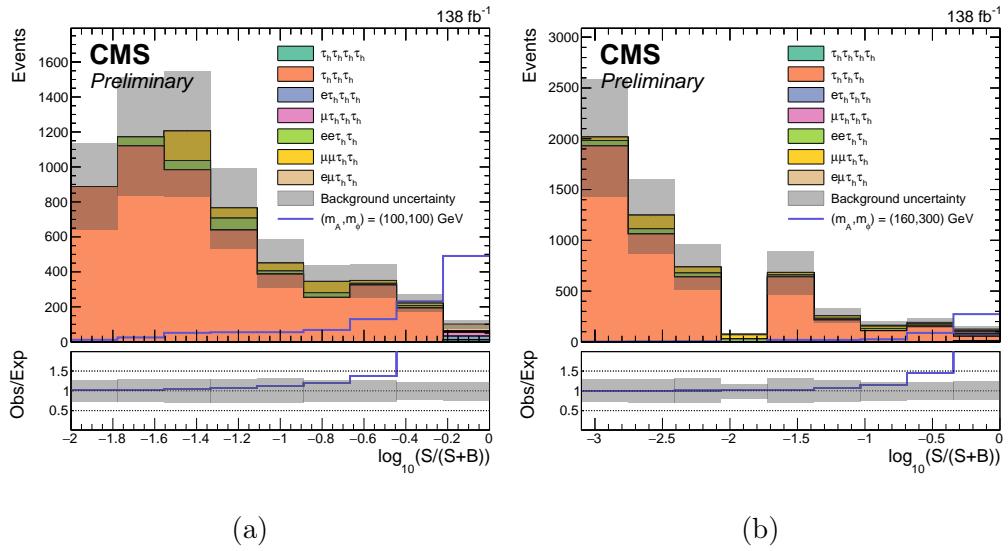


Figure 5.12: Distributions of events with respect to the  $\log_{10}(S/(S + B))$  of all analysis category bins fit for two signal scenarios:  $m_\phi = 100$  GeV and  $m_A = 100$  GeV and  $m_\phi = 160$  GeV and  $m_A = 160$  GeV. The solid histograms show the stacked background predictions after a background only fit to the data. The signal model considered is displayed with a solid line.

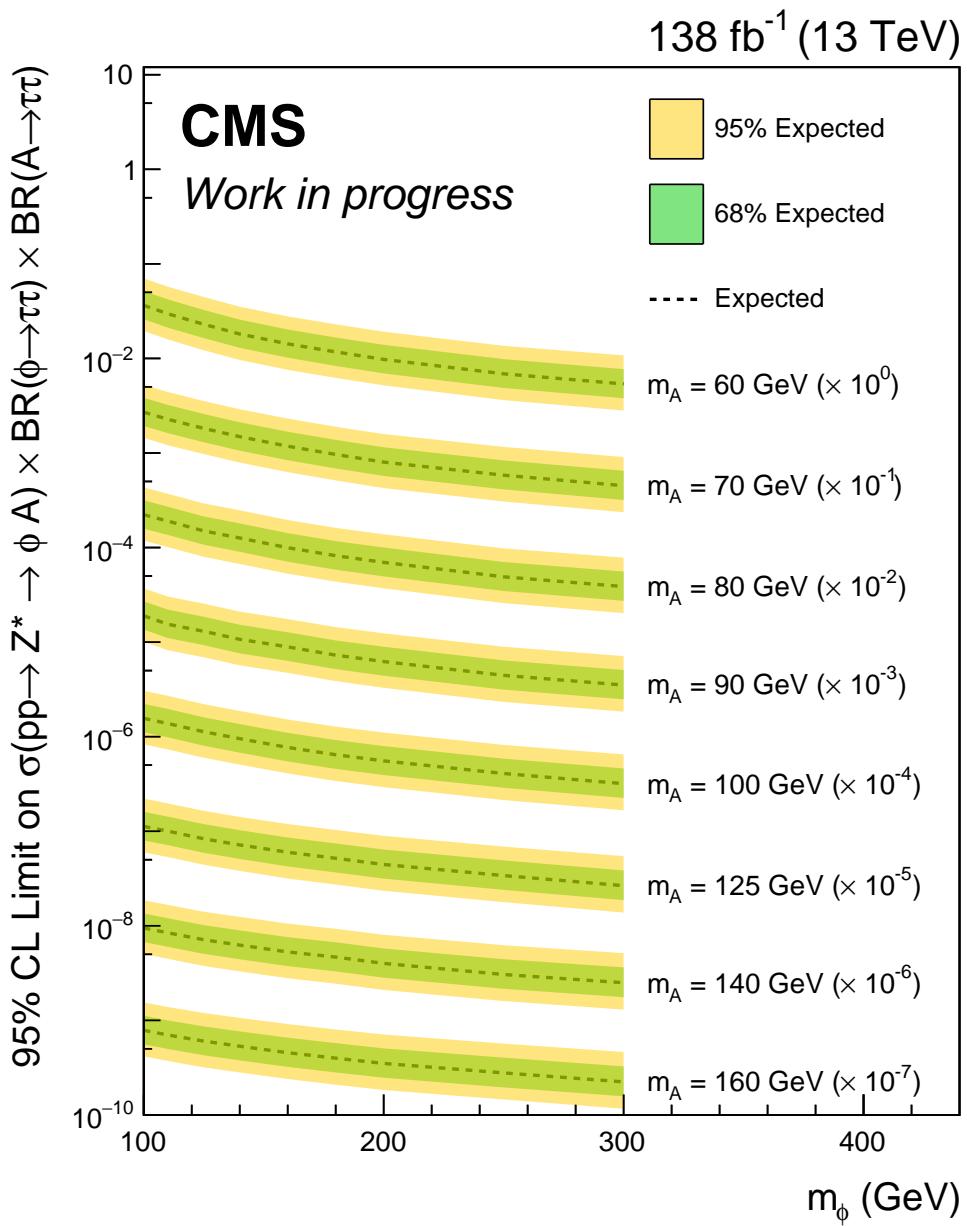


Figure 5.13: Expected (dashed line) and observed (solid line and dots) 95% CL upper limits on the product of the cross sections and branching fractions for the decay of both additional Higgs bosons into  $\tau$  leptons. Different  $m_A$  hypotheses are scaled by different orders of magnitudes (written on plot) in order to make mass points distinguishable. The dark green and bright yellow bands indicate the central 68% and 95% intervals for the expected exclusion limit.

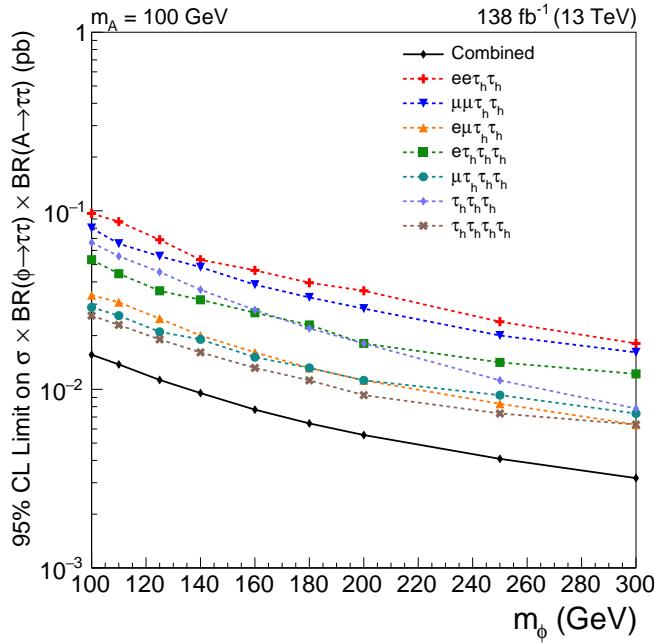


Figure 5.14: Comparison of the expected 95% CL upper limits on the product of the cross sections and branching fractions for the decay into  $\tau$  leptons, split by the  $\tau\tau\tau\tau$  decay products fit individually.

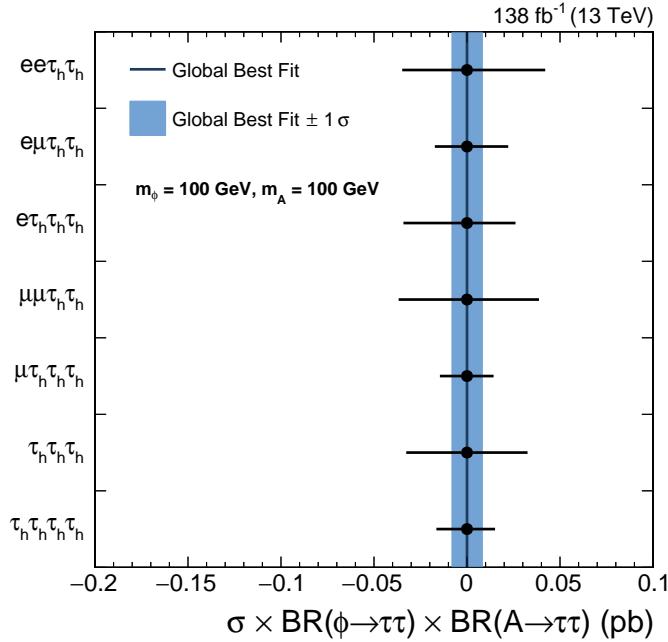


Figure 5.15: Compatibility plots for the  $m_A = 100$  GeV and  $m_\phi = 100$  GeV mass scenario in analysis decay channels. In each case the fitted signal strength is decoupled in the bin shown on the plot.

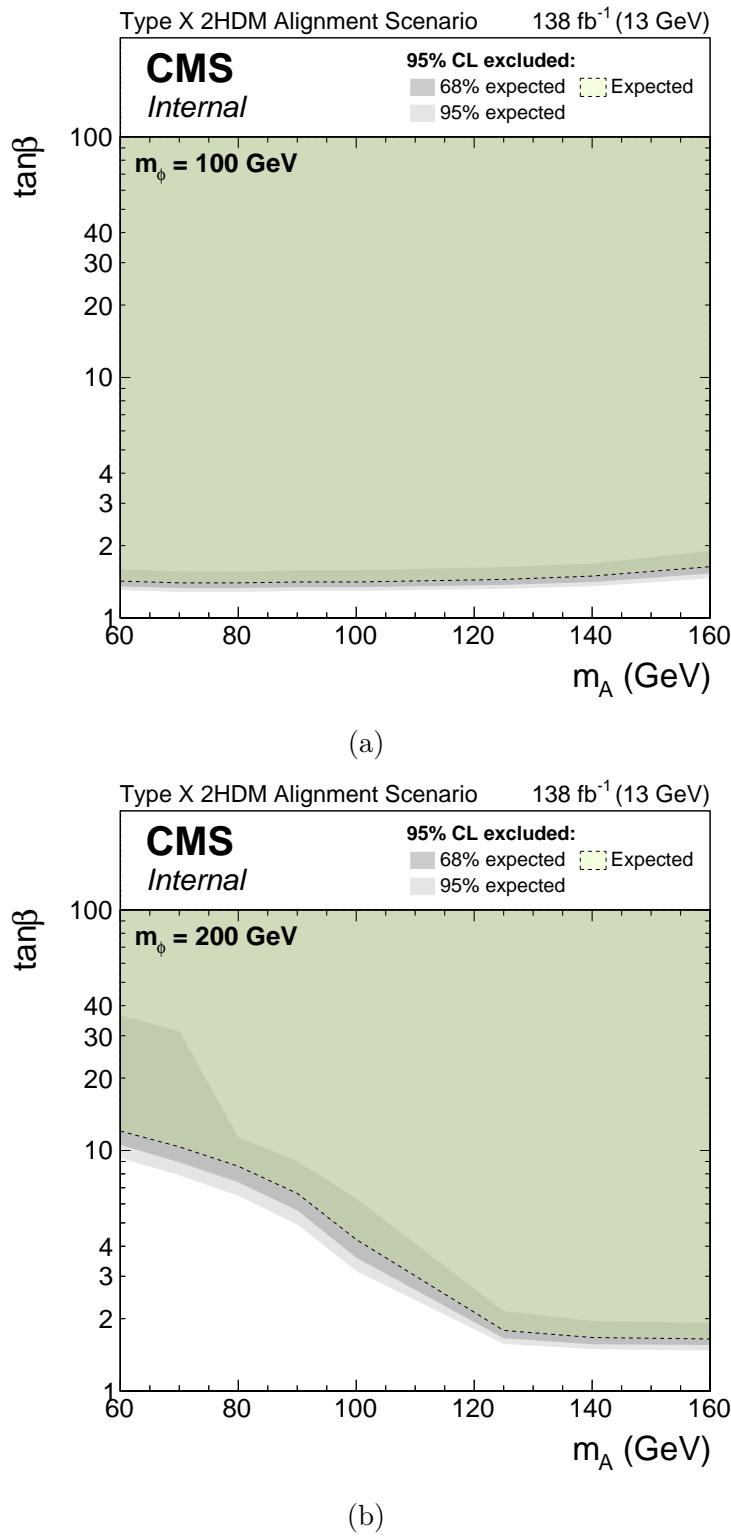


Figure 5.16: Expected and observed 95% CL exclusion contours on the  $m_A$ - $\tan\beta$  phase space in the type X 2HDM alignment scenario for  $m_\phi$  scenarios of 100 GeV (a) and 200 GeV (b). The exclusion limit only on background expectation is shown as a dashed black line, the dark and bright grey bands show the 68% and 95% intervals of the expected exclusion and the observed exclusion contour is shown by the blue area.

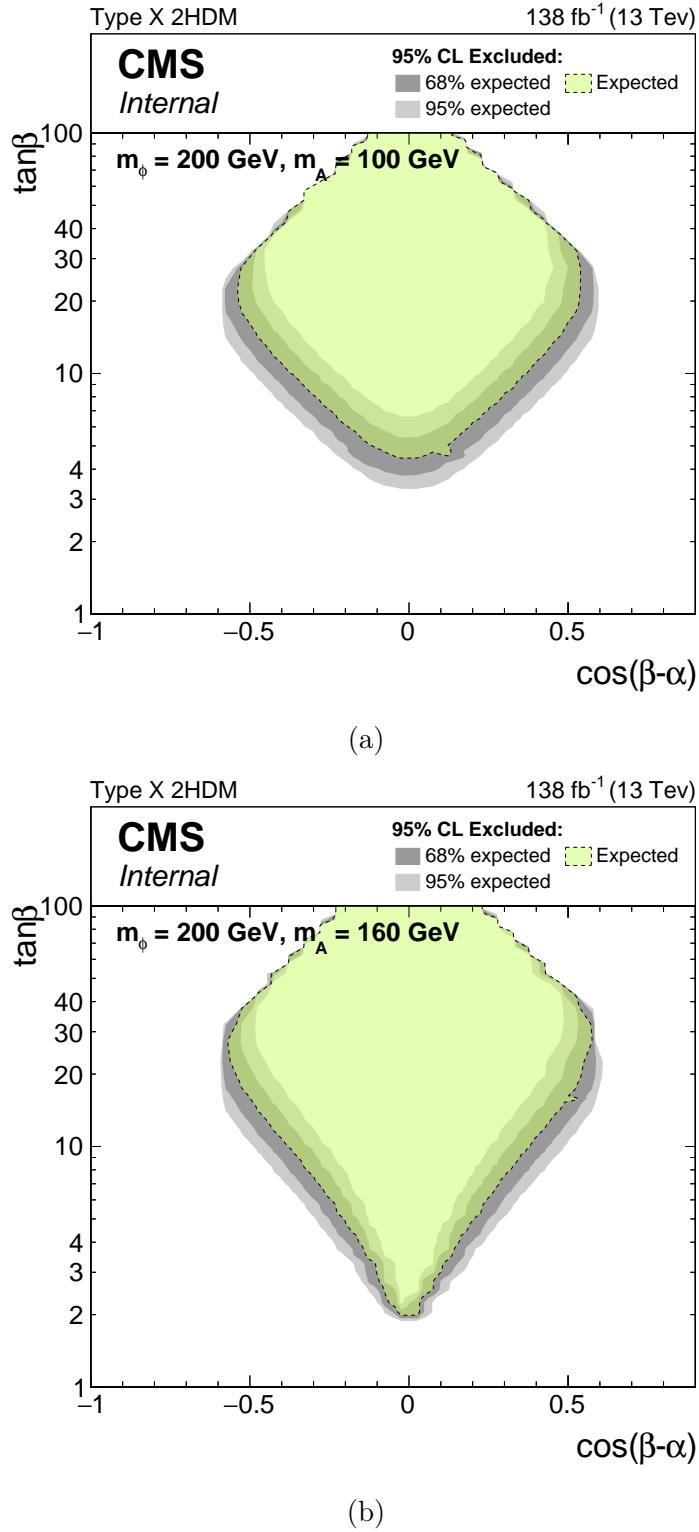


Figure 5.17: Expected and observed 95% CL exclusion contours on the  $\cos(\beta - \alpha)$ - $\tan\beta$  phase space in the type X 2HDM alignment scenario with  $m_\phi$  equal to 200 GeV and  $m_A$  scenarios of 100 GeV (a) and 160 GeV (b). The exclusion limit only on background expectation is shown as a dashed black line, the dark and bright grey bands show the 68% and 95% intervals of the expected exclusion and the observed exclusion contour is shown by the blue area.

# **Chapter 6**

## **Conclusion**

### **6.1 Global Interpretations of Results**

### **6.2 Outlook**

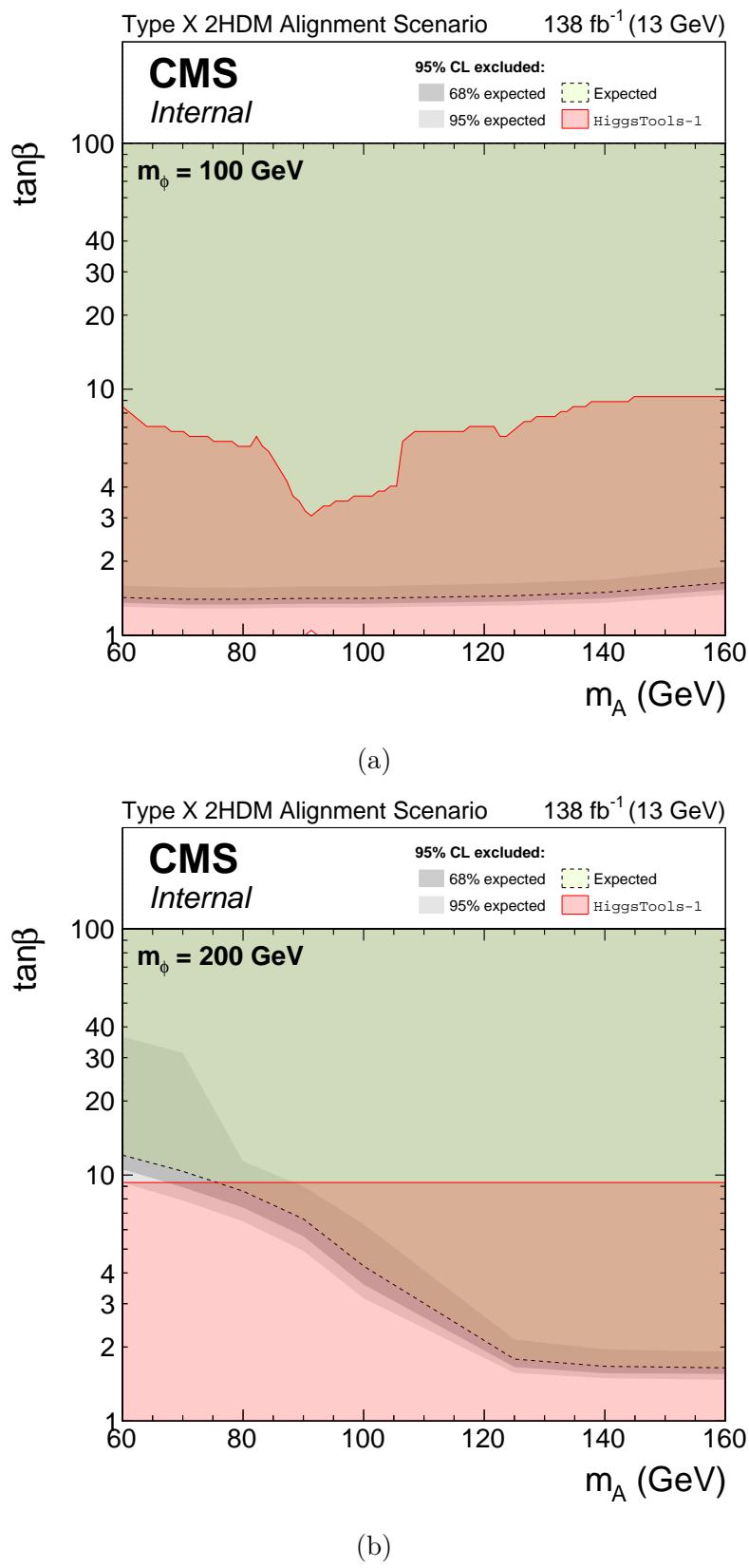


Figure 6.1

# Bibliography

- [1] H. Bartosik and G. Rumolo, “Performance of the LHC injector chain after the upgrade and potential development,” tech. rep., 2022. Contribution to Snowmass 2021.
- [2] CMS Collaboration, “The CMS Experiment at the CERN LHC,” *JINST*, vol. 3, p. S08004, 2008.
- [3] G. Aad *et al.*, “Search for heavy Higgs bosons decaying into two tau leptons with the ATLAS detector using pp collisions at  $\sqrt{s} = 13$  TeV,” *Phys. Rev. Lett.*, vol. 125, p. 051801, 2020.
- [4] C. Cornella, D. A. Faroughy, J. Fuentes-Martin, G. Isidori, and M. Neubert, “Reading the footprints of the B-meson flavor anomalies,” *JHEP*, vol. 08, p. 050, 2021.
- [5] A. Jueid, J. Kim, S. Lee, and J. Song, “Type-X two-Higgs-doublet model in light of the muon g-2: Confronting Higgs boson and collider data,” *Phys. Rev. D*, vol. 104, no. 9, p. 095008, 2021.
- [6] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Phys. Lett. B*, vol. 716, pp. 1–29, 2012.
- [7] CMS Collaboration, “Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC,” *Phys. Lett. B*, vol. 716, pp. 30–61, 2012.
- [8] R. Aaij *et al.*, “Test of lepton universality in beauty-quark decays,” *Nature Phys.*, vol. 18, no. 3, pp. 277–282, 2022.
- [9] B. Kowalewski, “Evidence for an excess of  $B \rightarrow D^{(*)}\tau\nu$  decays,” *J. Phys. Conf. Ser.*, vol. 455, p. 012039, 2013.

- [10] J. P. Lees *et al.*, “Measurement of an excess of  $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$  decays and implications for charged higgs bosons,” *Phys. Rev. D*, vol. 88, p. 072012, 2013.
- [11] M. Huschle *et al.*, “Measurement of the branching ratio of  $\bar{B} \rightarrow D^{(*)}\tau^-\bar{\nu}_\tau$  relative to  $\bar{B} \rightarrow D^{(*)}\ell^-\bar{\nu}_\ell$  decays with hadronic tagging at Belle,” *Phys. Rev. D*, vol. 92, no. 7, p. 072014, 2015.
- [12] R. Aaij *et al.*, “Measurement of the ratio of branching fractions  $\mathcal{B}(\bar{B}^0 \rightarrow D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \rightarrow D^{*+}\mu^-\bar{\nu}_\mu)$ ,” *Phys. Rev. Lett.*, vol. 115, p. 111803, 2015. [Erratum: [doi:10.1103/PhysRevLett.115.159901](https://doi.org/10.1103/PhysRevLett.115.159901)].
- [13] S. Hirose *et al.*, “Measurement of the  $\tau$  lepton polarization and  $R(D^*)$  in the decay  $\bar{B} \rightarrow D^*\tau^-\bar{\nu}_\tau$ ,” *Phys. Rev. Lett.*, vol. 118, p. 211801, 2017.
- [14] R. Aaij *et al.*, “Test of Lepton Flavor Universality by the measurement of the  $B^0 \rightarrow D^{*-}\tau^+\nu_\tau$  branching fraction using three-prong  $\tau$  decays,” *Phys. Rev. D*, vol. 97, no. 7, p. 072013, 2018.
- [15] R. Aaij *et al.*, “Measurement of the ratio of the  $B^0 \rightarrow D^{*-}\tau^+\nu_\tau$  and  $B^0 \rightarrow D^{*-}\mu^+\nu_\mu$  branching fractions using three-prong  $\tau$ -lepton decays,” *Phys. Rev. Lett.*, vol. 120, p. 171802, 2018.
- [16] B. Abi *et al.*, “Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm,” *Phys. Rev. Lett.*, vol. 126, no. 14, p. 141801, 2021.
- [17] F. Englert and R. Brout, “Broken symmetry and the mass of gauge vector mesons,” *Phys. Rev. Lett.*, p. 321, 1964.
- [18] P. W. Higgs, “Broken symmetries, massless particles and gauge fields,” *Phys. Lett.*, p. 132, 1964.
- [19] P. W. Higgs, “Broken symmetries and the masses of gauge bosons,” *Phys. Rev. Lett.*, p. 508, 1964.
- [20] P. W. Higgs, “Spontaneous symmetry breakdown without massless bosons,” *Phys. Rev.*, p. 1156, 1966.
- [21] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global conservation laws and massless particles,” *Phys. Rev. Lett.*, p. 585, 1964.
- [22] T. W. B. Kibble, “Symmetry breaking in non-Abelian gauge theories,” *Phys. Rev.*, p. 1554, 1967.

- [23] J. Goldstone, “Field Theories with Superconductor Solutions,” *Nuovo Cim.*, vol. 19, pp. 154–164, 1961.
- [24] A. M. Sirunyan *et al.*, “A measurement of the Higgs boson mass in the diphoton decay channel,” *Phys. Lett. B*, vol. 805, p. 135425, 2020.
- [25] S. P. Martin, “A Supersymmetry Primer,” *Advanced Series on Directions in High Energy Physics*, p. 1–98, Jul 1998.
- [26] G. W. Bennett *et al.*, “Final Report of the Muon E821 Anomalous Magnetic Moment Measurement at BNL,” *Phys. Rev. D*, vol. 73, p. 072003, 2006.
- [27] V. Ilisie, “New Barr-Zee contributions to  $(g - 2)_\mu$  in two-Higgs-doublet models,” *JHEP*, vol. 04, p. 077, 2015.
- [28] S. M. Barr and A. Zee, “Electric Dipole Moment of the Electron and of the Neutron,” *Phys. Rev. Lett.*, vol. 65, pp. 21–24, 1990. [Erratum: Phys.Rev.Lett. 65, 2920 (1990)].
- [29] “LHC Machine,” *JINST*, vol. 3, p. S08001, 2008.
- [30] “LEP Design Report: Vol.2. The LEP Main Ring,” 6 1984.
- [31] K. Aamodt *et al.*, “The ALICE experiment at the CERN LHC,” *JINST*, vol. 3, p. S08002, 2008.
- [32] G. Aad *et al.*, “The ATLAS Experiment at the CERN Large Hadron Collider,” *JINST*, vol. 3, p. S08003, 2008.
- [33] A. A. Alves, Jr. *et al.*, “The LHCb Detector at the LHC,” *JINST*, vol. 3, p. S08005, 2008.
- [34] M. Malberti, “Operation and performance of the CMS tracker,” *PoS*, vol. TIPP2014, p. 347, 2014.
- [35] “CMS Technical Design Report for the Pixel Detector Upgrade,” 9 2012.
- [36] S. Chatrchyan *et al.*, “Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in  $pp$  Collisions at  $\sqrt{s} = 7$  TeV,” *JINST*, vol. 8, p. P09009, 2013.
- [37] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- $k_t$  jet clustering algorithm,” *JHEP*, vol. 04, p. 063, 2008.

- [38] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet user manual,” *Eur. Phys. J. C*, vol. 72, p. 1896, 2012.
- [39] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector,” *JINST*, vol. 12, no. 10, p. P10003, 2017.
- [40] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV,” *JINST*, vol. 13, no. 05, p. P05011, 2018.
- [41] CMS Collaboration, “Reconstruction and identification of  $\tau$  lepton decays to hadrons and  $\nu_\tau$  at CMS,” *JINST*, vol. 11, no. 01, p. P01019, 2016.
- [42] CMS Collaboration, “Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2,” Tech. Rep. CMS-PAS-TAU-16-002, CERN, Geneva, 2016.
- [43] A. M. Sirunyan *et al.*, “Performance of reconstruction and identification of  $\tau$  leptons decaying to hadrons and  $\nu_\tau$  in pp collisions at  $\sqrt{s} = 13$  TeV,” *JINST*, vol. 13, p. P10005, 2018.
- [44] CMS Collaboration, “Performance of the DeepTau algorithm for the discrimination of taus against jets, electron, and muons,” CMS Detector Performance Note CMS-DP-2019-033, 2019.
- [45] E. Bagnaschi, H. Bahl, E. Fuchs, T. Hahn, S. Heinemeyer, S. Liebler, S. Patel, P. Slavich, T. Stefaniak, C. E. M. Wagner, and G. Weiglein, “MSSM Higgs boson searches at the LHC: Benchmark scenarios for Run 2 and beyond,” *Eur. Phys. J. C*, vol. 79, p. 617, 2019.
- [46] H. Bahl, P. Bechtle, S. Heinemeyer, S. Liebler, T. Stefaniak, and G. Weiglein, “HL-LHC and ILC sensitivities in the hunt for heavy Higgs bosons,” *Eur. Phys. J. C*, vol. 80, p. 916, 2020.
- [47] H. Bahl, S. Liebler, and T. Stefaniak, “MSSM Higgs benchmark scenarios for Run 2 and beyond: the low  $\tan \beta$  region,” *Eur. Phys. J. C*, vol. 79, p. 279, 2019.
- [48] E. A. Bagnaschi, S. Heinemeyer, S. Liebler, P. Slavich, and M. Spira, “Benchmark scenarios for MSSM Higgs boson searches at the LHC,” Tech. Rep. LHCHWG-2021-001, CERN, 2021.

- [49] CMS Collaboration, “Search for additional neutral MSSM Higgs bosons in the  $\tau\tau$  final state in proton-proton collisions at  $\sqrt{s} = 13$  TeV,” *JHEP*, vol. 09, p. 007, 2018.
- [50] V. Khachatryan *et al.*, “Search for neutral MSSM Higgs bosons decaying into a pair of bottom quarks,” *JHEP*, vol. 11, p. 071, 2015.
- [51] P. Nason, “A new method for combining NLO QCD with shower Monte Carlo algorithms,” *JHEP*, vol. 11, p. 040, 2004.
- [52] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with parton shower simulations: the POWHEG method,” *JHEP*, vol. 11, p. 070, 2007.
- [53] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX,” *JHEP*, vol. 06, p. 043, 2010.
- [54] T. Ježo and P. Nason, “On the treatment of resonances in next-to-leading order calculations matched to a parton shower,” *JHEP*, vol. 12, p. 065, 2015.
- [55] R. D. Ball *et al.*, “Parton distributions for the LHC Run II,” *JHEP*, vol. 04, p. 040, 2015.
- [56] R. D. Ball *et al.*, “Parton distributions from high-precision collider data,” *Eur. Phys. J. C*, vol. 77, p. 663, 2017.
- [57] A. M. Sirunyan *et al.*, “Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements,” *Eur. Phys. J. C*, vol. 80, p. 4, 2020.
- [58] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, “An introduction to PYTHIA 8.2,” *Comput. Phys. Commun.*, vol. 191, p. 159, 2015.
- [59] S. Agostinelli *et al.*, “GEANT4—a simulation toolkit,” *Nucl. Instrum. Meth. A*, vol. 506, p. 250, 2003.
- [60] CMS Collaboration, “The search for a third-generation leptoquark coupling to a  $\tau$  lepton and a b quark through single, pair and nonresonant production at  $\sqrt{s} = 13$  TeV,” 2022.

- [61] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, “MadGraph 5: Going beyond,” *JHEP*, vol. 06, p. 128, 2011.
- [62] R. Frederix and S. Frixione, “Merging meets matching in MC@NLO,” *JHEP*, vol. 12, p. 061, 2012.
- [63] L. Bianchini, J. Conway, E. K. Friis, and C. Veelken, “Reconstruction of the Higgs mass in  $H \rightarrow \tau\tau$  events by dynamical likelihood techniques,” *J. Phys. Conf. Ser.*, vol. 513, p. 022035, 2014.
- [64] CMS Collaboration, “Measurements of Higgs boson production in the decay channel with a pair of  $\tau$  leptons in proton-proton collisions at  $\sqrt{s} = 13$  TeV.” Submitted to *Eur. Phys. J. C*, 2022.
- [65] J. Alwall, S. Höche, F. Krauss, N. Lavesson, L. Lönnblad, F. Maltoni, M. L. Mangano, M. Moretti, C. G. Papadopoulos, F. Piccinini, S. Schumann, M. Trecan, J. Winter, and M. Worek, “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions,” *Eur. Phys. J. C*, vol. 53, p. 473, 2008.
- [66] S. Alioli, S.-O. Moch, and P. Uwer, “Hadronic top-quark pair-production with one jet and parton showering,” *JHEP*, vol. 01, p. 137, 2012.
- [67] R. Frederix, E. Re, and P. Torrielli, “Single-top  $t$ -channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO,” *JHEP*, vol. 09, p. 130, 2012.
- [68] E. Re, “Single-top  $Wt$ -channel production matched with parton showers using the POWHEG method,” *Eur. Phys. J. C*, vol. 71, p. 1547, 2011.
- [69] K. Melnikov and F. Petriello, “Electroweak gauge boson production at hadron colliders through  $\mathcal{O}(\alpha_s^2)$ ,” *Phys. Rev. D*, vol. 74, p. 114017, 2006.
- [70] M. Czakon and A. Mitov, “Top++: A program for the calculation of the top-pair cross-section at hadron colliders,” *Comput. Phys. Commun.*, vol. 185, p. 2930, 2014.
- [71] N. Kidonakis, “Top quark production,” in *Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons*, p. 139, 2014.
- [72] J. M. Campbell, R. K. Ellis, and C. Williams, “Vector boson pair production at the LHC,” *JHEP*, vol. 07, p. 018, 2011.

- [73] T. Gehrmann, M. Grazzini, S. Kallweit, P. Maierhöfer, A. von Manteuffel, S. Pozzorini, D. Rathlev, and L. Tancredi, “ $W^+W^-$  production at hadron colliders in next to next to leading order QCD,” *Phys. Rev. Lett.*, vol. 113, p. 212001, 2014.
- [74] R. J. Barlow and C. Beeston, “Fitting using finite Monte Carlo samples,” *Comput. Phys. Commun.*, vol. 77, p. 219, 1993.
- [75] J. S. Conway, “Incorporating Nuisance Parameters in Likelihoods for Multi-source Spectra,” in *PHYSTAT 2011*, pp. 115–120, 2011.
- [76] A. M. Sirunyan *et al.*, “Performance of missing transverse momentum reconstruction in proton-proton collisions at  $\sqrt{s} = 13$  TeV using the CMS detector,” *JINST*, vol. 14, p. P07004, 2019.
- [77] T. Junk, “Confidence level computation for combining searches with small statistics,” *Nucl. Instrum. Meth. A*, vol. 434, p. 435, 1999.
- [78] A. L. Read, “Presentation of search results: The  $CL_s$  technique,” *J. Phys. G*, vol. 28, p. 2693, 2002.
- [79] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based tests of new physics,” *Eur. Phys. J. C*, vol. 71, p. 1554, 2011. [Erratum: [doi:10.1140/epjc/s10052-013-2501-z](https://doi.org/10.1140/epjc/s10052-013-2501-z)].
- [80] LHC Higgs Cross Section Working Group, “Handbook of LHC Higgs cross sections: 3. Higgs properties,” Tech. Rep. CERN-2013-004, 2013.
- [81] LHC Higgs Cross Section Working Group, “Handbook of LHC Higgs cross sections: 4. deciphering the nature of the Higgs sector,” Tech. Rep. CERN-2017-002-M, 2016.
- [82] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, and M. Spira, “Standard model Higgs-boson branching ratios with uncertainties,” *Eur. Phys. J. C*, vol. 71, p. 1753, 2011.
- [83] CMS Collaboration, “Searches for additional Higgs bosons and vector leptophilic quarks in  $\tau\tau$  final states in proton-proton collisions at  $\sqrt{s} = 13$  TeV,” 2022.

- [84] M. Krause, M. Mühlleitner, and M. Spira, “2HDECAY —A program for the calculation of electroweak one-loop corrections to Higgs decays in the Two-Higgs-Doublet Model including state-of-the-art QCD corrections,” *Comput. Phys. Commun.*, vol. 246, p. 106852, 2020.
- [85] A. Rogozhnikov, “Reweighting with Boosted Decision Trees,” *J. Phys. Conf. Ser.*, vol. 762, no. 1, p. 012036, 2016.