Imperial College London

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 (SM) of particle physics is currently the best model of the fundamental particles and their interactions. However, there are significant theoretical issues and experimental tensions with the model. The theoretical issues include the hierarchy problem which forecasts the breakdown of the SM 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 measurement of the muon anomalous magnetic moment. Looking for signatures of theoretical explanations to these issues offers excellent search options for new fundamental particles. This thesis describes searches for new physics that can explain both the theoretical problems and experimental tensions. This is done using tau (τ) leptons seen during Run 2 of the Large Hadron Collider at the Compact Muon Solenoid (CMS) experiment. The beyond SM theories searched for range from the Minimal Supersymmetric SM (MSSM), leptoquarks, and the type X two Higgs doublet models (2HDM). Two local excesses are observed with statistical significances $\approx 3\sigma$ for a resonance produced via gluon fusion in the $\tau\tau$ final states, at masses of 100 GeV and 1.2 TeV. Otherwise, good agreement is observed between the SM background prediction and data. Leading constraints are placed on scenarios of the MSSM phase space and limits that shrink the allowed explanation to the B anomalies are set. The type X 2HDM, as an explanation for the muon anomalous magnetic moment measurement, is fully ruled out and many mass hypotheses in the type X 2HDM are completely excluded by this thesis.

Declaration

I declare that the work in this thesis is mine. Figures and results taken from other sources are indicated by a reference in the text or figure caption. Figures labelled "CMS" are sourced from CMS publications. Figures labelled "CMS Supplementary" have been made public as supplementary material accompanying a publication. The "CMS Preliminary" label is given to CMS public results that are yet to be finalised by the collaboration and the journal. The "Simulation" tag is added when only simulated data is used to generate a CMS public plot. All figures with these labels, including those made by myself, also include the relevant reference in the caption. The analyses presented in this document were developed in collaboration with other members of the CMS experiment. Chapters 1–3 do not contain original work by myself but are intended to describe the theoretical motivation, as well as the apparatus and methods used to generate and collect the data, that are critical to the work presented in the following chapters. The high-mass additional Higgs boson search, described in Chapter 4, and the origin of the background modelling methods were pre-established in Reference [1]. The work described in Chapter 4 was done in collaboration with the other members of the CMS $H \to \tau\tau$ group. For the additional Higgs boson component of this chapter, I contributed to the low-mass optimisation, derivation and parametrisation of fake factors, evaluation of uncertainties and the final statistical fits to data. The vector leptoquark interpretation of the results was solely my work. The results in Chapter 4 were made public in Reference [2]. Chapter 5 contains work performed in collaboration with the Imperial College London $H \to \tau\tau$ group. I was responsible for the full workflow of this analysis including signal simulation, background modelling, optimisation, uncertainty modelling, interpretations and the final statistical fits to data. This analysis is currently not published, but a publication is planned in the near future. Chapter 6 includes interpretations of the results in Chapters 4 and 5, that are entirely my work.

George Uttley

Acknowledgements

I would like to thank the Imperial College London High Energy Physics group and the Science and Technology Facilities Council for giving me the opportunity to conduct this research. Thank you to my supervisor, David Colling, for all the assistance and guidance, in particular, the aid I received whilst working through the pandemic. I also owe a massive thank you to Daniel Winterbottom, for all of the help and advice, as well as the patience he showed with me throughout my PhD. I am also grateful to the remaining members of the $H \rightarrow \tau\tau$ group for all of the interesting discussions and continued motivation whilst writing this thesis. Thank you to my friends and family, who have continually supported me throughout my life and are the reason I was able to complete this work. In particular, thank you to Emmy for her unwavering support and encouragement throughout this journey.

George Uttley

Contents

List of Figures

List of Tables

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-Englet-Higgs (BEH) mechanism for spontaneous symmetry breaking, which allows many fundamental particles to obtain masses. It also predicts a new scalar boson, named the Higgs boson. In 2012, the ATLAS [3] and CMS [4] collaborations discovered a Higgs boson-like particle when colliding protons at high energies at the Large Hadron Collider (LHC). Further measurements of the particle's properties have been consistent with the SM prediction. 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 (SUSY), but no experimental evidence for this theory has yet been found that separates it from the SM. Secondly, results from B physics [5–12] and the measurement of the muon's anomalous magnetic moment [13, 14] have shown deviations from the SM predictions. Although not the statistical significance for discovery, they offer intriguing hints at potential Beyond Standard Model (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 problems and experimental tensions.

One caveat to the B physics results presented in this thesis is the exclusion of the updated R_K and R_{K^*} measurements from the LHCb collaboration [15]. This is due to the interpretation of the B anomalies, described in this chapter and used in Chapter 4, being concluded prior to the release of Reference [15]. The phenomenological effect of Reference [15], on the best fit of new physics to the B anomalies, is yet to be fully determined.

1.1 The Standard Model of particle physics

1.1.1 Fundamental particles and their 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 model the strong, weak and electromagnetic forces, unifying the latter two into one electroweak interaction. The SM consists of 6 quarks, 3 charged leptons and 3 neutrinos, which are grouped into "fermions" because of their shared half-integer spin. Each of these particles has an anti-partner with opposite quantum numbers but the same mass. The SM also consists of a number of particles, named "bosons", with shared 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, which governs the interaction of particles, 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, named Quantum Chromodynamics (QCD).

The quantum number associated with the QCD $SU(3)_C$ symmetry is the colour charge C. Quarks and gluons carry a colour charge and so interact with the strong force. One characteristic feature of QCD is confinement, which requires neutral colour charges and so quarks must be observed in a bound state, named "hadrons". Another key property is asymptotic freedom, which weakens the interaction strength at higher energy or very short distances [18, 19].

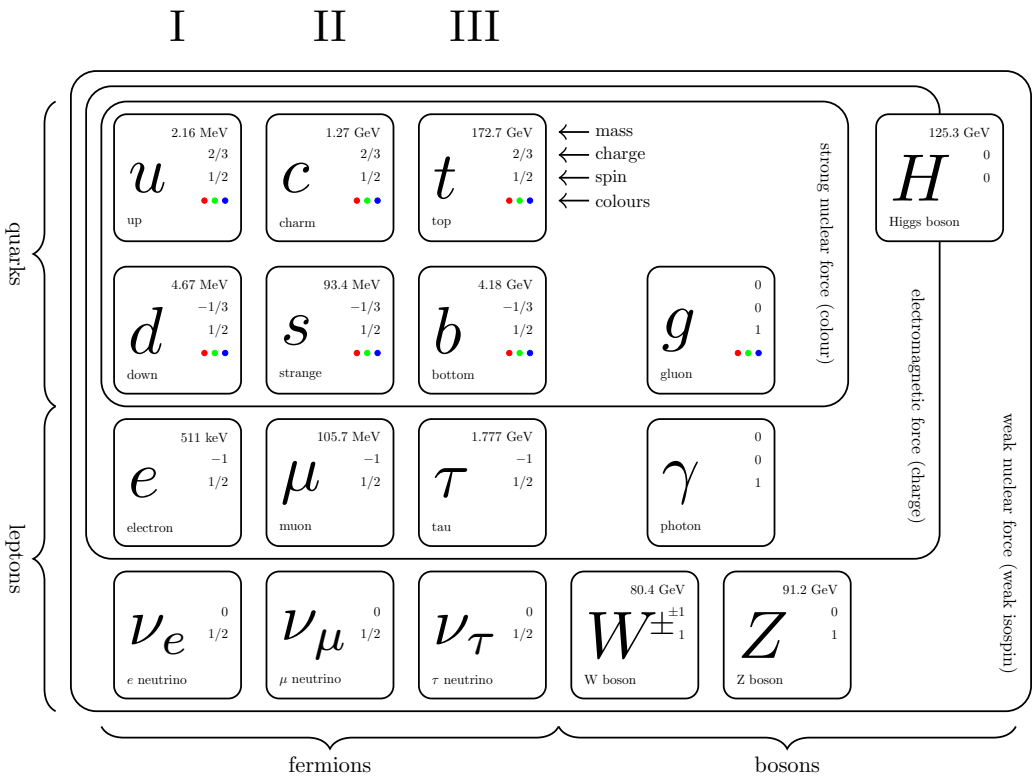


Figure 1.1: Diagram of the fundamental particles that constitute the SM. Also displayed are the fermion generation shown in Roman numerals, the particles measured mass, charge, spin and colours available for the strong interaction. The particle masses used are taken from Reference [16]. The neutrino masses are left blank as the values are unknown. This figure is taken and adjusted from Reference [17].

Electroweak unification was initially proposed by Glashow [20], Weinberg [21] and Salam [22] to combine the theories of the weak and electromagnetic forces into one. It is built on the premise of the Dirac equation. The Dirac Lagrangian for a massless spinor field, ψ, is defined as,

$$\mathcal{L}_{\text{Dirac}} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi, \tag{1.1}$$

where γ^{μ} are the gamma matrices and ∂_{μ} are partial derivatives. The $\mathrm{SU(2)_L}$ transformation operates on the weak isospin, I, only for the left-handed spinors and the $\mathrm{U(1)_Y}$ 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. The reason for the distinction of the SU(2) symmetry to act only on left-handed spinors is from the observed chirality of fermions under the weak interaction [23]. By invoking gauge invariance of the Dirac Lagrangian, four associated gauge fields are present, $\boldsymbol{W}_{\mu} = (W^1_{\mu}, W^2_{\mu}, W^3_{\mu})$ and B_{μ}. The Lagrangian then takes the form,

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

where g and g' are the coupling constants for the SU(2) and U(1) symmetries respectively, the partial derivatives have been replaced with the covariant derivative and the added field tensors $\boldsymbol{W}_{\mu\nu}$ and $B_{\mu\nu}$ are defined as,

$$\boldsymbol{W}_{\mu\nu} = \partial_{\mu}\boldsymbol{W}_{\nu} - \partial_{\nu}\boldsymbol{W}_{\mu} - ig[\boldsymbol{W}_{\mu}, \boldsymbol{W}_{\nu}], \tag{1.3}$$

$$B_{\mu\nu} = \partial_{\mu}B_{\nu} - \partial_{\mu}B_{\mu}. \tag{1.4}$$

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

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

where θ_w represents the weak mixing angle and is defined such that

$$\sin\theta_w = \frac{g'}{\sqrt{g^2+g'^2}}, \qquad \cos\theta_w = \frac{g}{\sqrt{g^2+g'^2}}. \tag{1.6}$$

The W and Z bosons were discovered by the UA1 and UA2 collaborations in 1983 [24, 25], which confirmed the predictions made by electroweak unification. However, the W 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 $\mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ gauge symmetry. The same issue also exists in massive fermions, as the fermion mass term's 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 do not remain invariant under the gauge transformation. To resolve this, the 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 [26], Higgs [27–29] and Guralnik, Hagen and Kibble [30, 31]. It works on the principles of spontaneous symmetry breaking of the $\mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ 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 Φ,

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{1.7}$$

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

$$\mathcal{L}_{\text{Complex Scalar}} = (\partial_\mu\Phi)^\dagger(\partial^\mu\Phi) - V(\Phi), \tag{1.8}$$

with,

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

where μ^2 and λ are two real parameters. λ is required to be positive for the vacuum to be stable. If μ^2 is negative, Φ will have a non-zero VEV and the field will be able to spontaneously break the gauge symmetry. In the vacuum state, the field must satisfy the criteria,

$$\Phi^{\dagger}\Phi = -\frac{\mu^2}{2\lambda}. \tag{1.10}$$

This potential fulfils the criteria of a non-zero VEV as Φ must be non-zero. Choosing a gauge to remove the massless scalar bosons, predicted by Goldstone's theorem [32], by taking ϕ^+ to be zero and ϕ to be real. The ground state of Φ is found as,

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

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

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

Applying the covariant derivative for the electroweak gauge symmetry to produce the Lagrangian for the field, Φ, yields

$$\begin{aligned}\mathcal{L}_{\text{Higgs}} = &\frac{1}{2}(\partial_\mu h)(\partial^\mu h) + \frac{1}{8}g^2(W^1_\mu + iW^2_\mu)(W^{1\mu} - iW^{2\mu})(\nu + h)^2 \\ &+ \frac{1}{8}(gW^3_\mu - g'B_\mu)(gW^{3\mu} - g'B^\mu)(\nu + h)^2 \\ &+ \mu^2(\nu + \mathrm{h})^2 + \lambda(\nu + h)^4.\end{aligned} \tag{1.13}$$

Mass terms for W^1_μ and W^2_μ are now present 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^3_μ and B_μ, 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 Equation 1.5 utilises,

$$\tan\theta_W = \frac{g'}{g}. \tag{1.14}$$

A mass term for the Higgs boson also arises from the potential of the BEH mechanism with $m_{\mathrm{h}} = \sqrt{-2\mu^2}$. 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,

$$\mathcal{L}_{\text{Yukawa}} = \lambda_f(\bar{\psi}_L \Phi \psi_R + \bar{\psi}_R \Phi \psi_L). \tag{1.15}$$

Upon spontaneous symmetry breaking, λ_f becomes proportional to the mass of the fermion, and hence the couplings are stronger between heavier fermions and the Higgs field.

1.2 Extended Higgs sector

There is no theoretical limitation to having only one Higgs doublet in the theory. Therefore, a natural extension to the SM Higgs sector is a two Higgs doublet model (2HDM). The Lagrangian for such a theory is shown below.

$$\mathcal{L}_{\text{2HDM}} = (D_\mu \Phi_1)^\dagger (D_\mu \Phi_1) + (D_\mu \Phi_2)^\dagger (D_\mu \Phi_2) - V_{\text{2HDM}}(\Phi_1, \Phi_2) \tag{1.16}$$

where,

$$\begin{aligned} V_{\text{2HDM}} &= m_{11}^2 |\Phi_1|^2 + m_{22}^2 |\Phi_2|^2 + (m_{12}^2 \Phi_1^\dagger \Phi_2 + \text{h.c.}) \\ &+ \frac{\lambda_1}{2} |\Phi_1|^4 + \frac{\lambda_2}{2} |\Phi_2|^4 + \lambda_3 |\Phi_1|^2 |\Phi_2|^2 + \lambda_4 |\Phi_1^\dagger \Phi_2| \\ &+ \frac{1}{2}\Big[\lambda_5 (\Phi_1^\dagger \Phi_2)^2 + \lambda_6 |\Phi_1|^2 \Phi_1^\dagger \Phi_2 + \lambda_7 |\Phi_2|^2 \Phi_1^\dagger \Phi_2 + \text{h.c.}\Big], \end{aligned} \tag{1.17}$$

where m_{ij} represents the terms in the mass matrix and λ_i parametrises the self-couplings of the Higgs sector. After the BEH mechanism is applied, 2HDMs predict 5 Higgs bosons; 1 lighter and 1 heavier charge-parity (CP)-even (h and H), 1 CP-odd (A) and 2 charged ($H^\pm$) particles. The Lagrangian for the Yukawa interactions with the Higgs bosons in such a theory are,

$$\begin{aligned} \mathcal{L}_{\text{Yukawa}}^{\text{2HDM}} = &- \sum_{f=u,d,l} \Big(\frac{m_f}{\nu} g_h^f \bar{f} f h + \frac{m_f}{\nu} g_H^f \bar{f} f H - i \frac{m_f}{\nu} g_A^f \bar{f} \gamma_5 f A \Big) \\ &- \Big[\frac{\sqrt{2} V_{ud}}{\nu} \bar{u} (m_u g_A^u P_L + m_d g_A^d P_R) d H^+ + \frac{\sqrt{2} m_l g_A^d}{\nu} \bar{\nu}_L l_R H^+ + h.c. \Big], \end{aligned} \tag{1.18}$$

where u, d, l and ν represent up-like quark, down-like quark, charged lepton and neutrino fields, the subscript L and R are the left- and right-handed projections performed via the projection operators, P_L and P_R respectively. m_f are the fermion masses, ν is the vacuum expectation value of the SM Higgs doublet, and g are the couplings (relative to the SM Higgs boson's couplings) of fermion fields to Higgs

fields, h, H, A and H^+. There are four main types of 2HDMs, which are defined based 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 Table 1.1. By convention, Φ_2 is chosen to couple to up-like quarks.

	Type I	Type II	Type X	Type Y
u	Φ_2	Φ_2	Φ_2	Φ_2
d	Φ_2	Φ_1	Φ_2	Φ_1
l	Φ_2	Φ_1	Φ_1	Φ_2

Table 1.1: Table showing which fermion groups couple to which Higgs doublet, in different types of 2HDMs.

The type of 2HDM determines the formulae for the couplings, g, that are functions on two parameters: the CP-even (α) and CP-odd (β) mixing angles of the mass matrices. These relative couplings are shown in Table 1.2.

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

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 α and β. t_x, s_x and c_x represent $\tan x$, $\sin x$ and $\cos x$ respectively.

To match the observed Higgs boson measurements to a CP-even boson predicted by a 2HDM, a linear combination of these two states are taken,

$$h_{\text{obs}} = \sin(\beta - \alpha)h + \cos(\beta - \alpha)H. \tag{1.19}$$

Assuming a non-degeneracy of the observed Higgs boson mass, two possible alignment limits are acquired: the normal scenario where $h_{obs} = h$ and $\cos(\beta-\alpha) = 0$ and the inverted scenario where $h_{obs} = H$ and $\sin(\beta-\alpha) = 0$. In the normal scenario, the values of the coupling ratios from Table 1.2 are $\cos\alpha/\sin\beta = 1$, $\sin\alpha/\cos\beta = -1$, $\sin\alpha/\sin\beta = -1/\tan\beta$ and $\cos\alpha/\cos\beta = \tan\beta$. Whilst in the inverted scenario, the ratios become $\cos\alpha/\sin\beta = 1/\tan\beta$, $\sin\alpha/\cos\beta = \tan\beta$, $\sin\alpha/\sin\beta = 1$ and $\cos\alpha/\cos\beta = 1$.

In the "physical basis", the 2HDM depends on the following parameters,

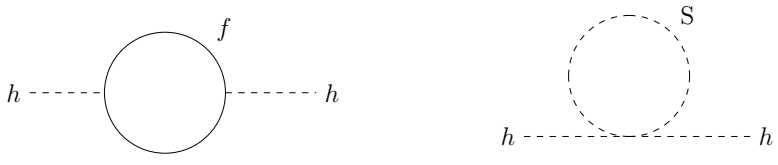
$$m_h, m_H, m_A, m_{H^\pm}, \tan\beta, \cos(\beta-\alpha), m_{12}^2, \lambda_6, \lambda_7 \tag{1.20}$$

It is common to apply a $\mathbb{Z}_2$ symmetry to a 2HDM to avoid quadratic divergences and suppress flavour-changing neutral currents [33, 34]. In this case, the basis of parameters is shrunk, as $\lambda_6 = \lambda_7 = 0$. $\tan\beta$ also represents the ratio of the background expectation values of the two Higgs doublets, Φ_2 and Φ_1. This is generally taken to be greater than 1 because if less than 1 this leads to weaker couplings with down-type fermions, contradicting the observed fermion mass hierarchy. Additionally, taking $\tan\beta < 1$ can result in violations of perturbativity, unitarity and potential vacuum instability in the electroweak sector [36].

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 Compact Muon Solenoid (CMS) experiment is 125.38 ± 0.14 GeV [35] in natural units, which are used throughout this thesis. The lightness is a concern when considering "naturalness" and BSM physics. 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 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 Figure 1.2.



(a) (b)

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

Figure 1.2a shows a 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,

$$\Delta m_{\mathrm{h}}^2 = -\frac{\lambda_f}{16\pi^2}\left[\Lambda_{UV}^2 - 2m_f^2 \ln\left(\frac{\Lambda_{UV}}{m_f}\right)\right] + ..., \tag{1.21}$$

where Λ_{UV} is the ultraviolet momentum cut off [36]. Beyond this, our effective field theory would be expected to break down and for new physics to be found.

Figure 1.2b illustrates a correction to the Higgs boson's 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 boson's mass correction for such a term is derived to be,

$$\Delta m_{\mathrm{h}}^2 = \frac{\lambda_S^2}{16\pi^2}\left[\Lambda_{UV}^2 - 2m_S^2 \ln\left(\frac{\Lambda_{UV}}{m_S}\right)\right] + ... \tag{1.22}$$

Equations 1.21 and 1.22 show that if the mass of the scalar is equivalent to that of the fermion and $\lambda_f = \lambda_S^2$, then the Higgs boson's mass corrections cancel. This offers a solution to the hierarchy problem, by introducing a new symmetry that extends the SM. The symmetry relates fermions and bosons and is known as SUSY. 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. For this to also hold off-shell, an auxiliary field is added to balance the number of degrees of freedom.

If SUSY is an unbroken theory, then it would be expected for the superpartners to have the same mass as the SM particles. This has not been seen experimentally, therefore, SUSY must be a broken theory in the vacuum state. Soft SUSY breaking can be introduced through the addition of a SUSY violating Lagrangian term $\mathcal{L}_{\mathrm{soft}}$ where,

$$\mathcal{L} = \mathcal{L}_{\text{SUSY}} + \mathcal{L}_{\text{soft}}. \tag{1.23}$$

$\mathcal{L}_{\text{soft}}$ contains only mass terms and coupling parameters. Defining m_{soft} as the largest mass scale involved in the soft Lagrangian, m_{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 boson's mass would again become large.

The Minimal Supersymmetric Standard Model (MSSM) is the simplest implementation of SUSY. 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 which are combinations of gauginos and Higgsinos. Additional contributions to the particle content come from the Higgs sector. The Higgs sector of the MSSM is required to be extended to two Higgs doublets to maintain the gauge symmetries and to cancel quantum mechanical inconsistencies [36]. In particular, a type II 2HDM is needed to ensure natural Yukawa couplings, minimal flavour violation and tree-level mass relations, which all cannot be achieved with a different extended Higgs sector [36]. At tree level, the MSSM Higgs sector is only dependent on m_{A} and $\tan\beta$. At high-order accuracies, benchmark scenarios are needed to set the remaining free parameters.

1.4 Experimental tensions and potential solutions

1.4.1 B anomalies

Measurements from the B physics experiments such as LHCb [5,9,11,12], BaBar [6,7] and Belle [8, 10], testing lepton flavour conservation, have found deviations away from the SM expectation of lepton universality. The differences are observed in both neutral current ($b \to s\ell^+\ell^-$) and charged current ($b \to c\tau\nu$) transitions. These B anomalies have prompted the idea for a short-range lepton flavour-violating interaction. This interaction is theorised to be mediated by a new "leptoquark" particle [37, 38]. 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 [39]. Such a leptoquark would couple to fermions by the Lagrangian shown below.

$$\mathcal{L}_U = \frac{g_U}{\sqrt{2}} U^\mu \left[\beta_L^{i\alpha}(\bar{q}_L^i \gamma_\mu l_L^\alpha) + \beta_R^{i\alpha}(\bar{d}_R^i \gamma_\mu e_R^\alpha)\right] + \text{h.c.} \tag{1.24}$$

where g_U is the coupling scaling parameter and β_L and β_R are the left and right-handed mixing matrices,

$$\beta_L = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \beta_L^{s\mu} & \beta_L^{s\tau} \\ 0 & \beta_L^{b\mu} & 1 \end{pmatrix}, \qquad \beta_R = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \beta_R^{b\tau} \end{pmatrix}. \tag{1.25}$$

The coupling g_U is defined such that $\beta_L^{b\tau} = 1$, and the negligible matrix elements for the fit are set to 0. The fit to the B anomalies performed in Reference [39], 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 are named VLQ BM 1 and VLQ BM 2 respectively. These represent no and maximal right-handed contributions. The best-fit results to the matrix parameters are shown in Table 1.3.

Scenario	$\beta_R^{b\tau}$	$\beta_L^{b\mu}$	$\beta_L^{s\tau}$	$\beta_L^{s\mu}$
VLQ BM 1	0	$-0.15^{+0.13}_{-0.11}$	$0.19^{+0.06}_{-0.09}$	$0.014^{+0.01}_{-0.01}$
VLQ BM 2	-1	$-0.14^{+0.12}_{-0.11}$	$0.19^{+0.05}_{-0.08}$	$0.03^{+0.01}_{-0.02}$

Table 1.3: Best-fit values and uncertainties for mixing matrix parameters as given in Reference [39].

The fit also provides a 1σ and 2σ bound on allowed values for the ratio of the vector leptoquark mass m_U, to the coupling constant g_U, and these are,

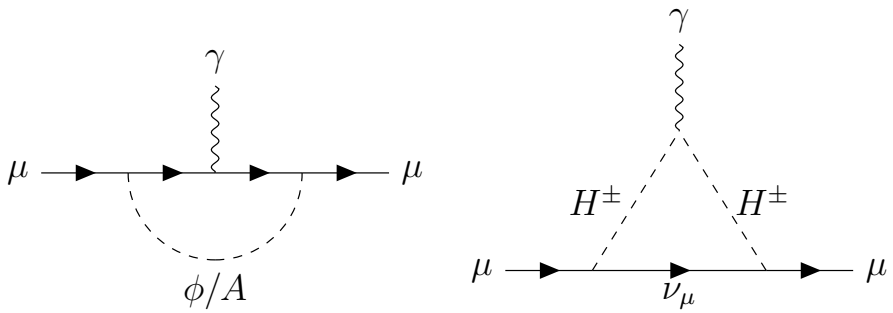
$$\begin{aligned} &\text{VLQ BM 1,}\ \ 1\sigma : 0.70 < g_U/m_U\ (1/\text{TeV}) < 1.09, \\ &\text{VLQ BM 1,}\ \ 2\sigma : 0.57 < g_U/m_U\ (1/\text{TeV}) < 1.38, \\ &\text{VLQ BM 2,}\ \ 1\sigma : 0.49 < g_U/m_U\ (1/\text{TeV}) < 0.67, \\ &\text{VLQ BM 2,}\ \ 2\sigma : 0.39 < g_U/m_U\ (1/\text{TeV}) < 1.25. \end{aligned} \tag{1.26}$$

1.4.2 Muon g-2 anomaly

The measurement of the muon anomalous magnetic moment from the Fermilab National Accelerator Laboratory muon g-2 experiment [14], combined with earlier results from the Brookhaven National Laboratory E821 measurement [13], find the difference of a_μ between the experiment value and SM prediction to be,

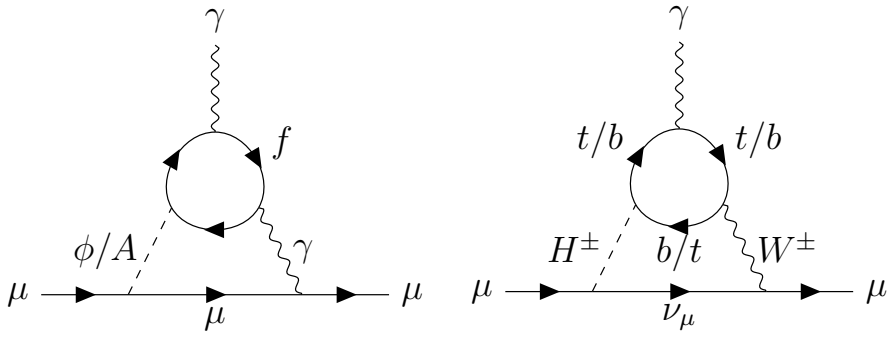
$$\Delta a_\mu^{\text{obs}} = a_\mu^{\text{exp}} - a_\mu^{\text{SM}} = (251 \pm 59) \times 10^{-11}, \tag{1.27}$$

where $a_\mu = (g-2)_\mu/2$. This is a 4.2 σ deviation away from the SM expectation. One potential solution to this deviation is a 2HDM. This contributes in two ways to Δa_μ, which is defined as the difference between the a_μ predicted with the 2HDM included and using just the SM. These are one-loop and two-loop Bar-Zee diagrams [40, 41], that can be mediated by any of the additional Higgs bosons. In this context, ϕ is used as the CP-even additional Higgs boson, h or H, depending on which particle is not matched to the observed Higgs boson.



(a) (b)

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



(a) (b)

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

The one-loop contribution to Δa_μ is positive from ϕ 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 Δa_μ. To fulfil the requirements for the anomaly, enhanced couplings between additional Higgs bosons and muons are needed. This gives two options: a type II or a type X 2HDM, both at large values of $\tan\beta$. The type II 2HDM, which also offers enhanced couplings to down-like quarks, is heavily constrained by Large Electron-Positron (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.

Reference [42] puts an upper limit of $m_A \lesssim 200$ GeV, which is 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 muon g-2 anomaly is shown in Table 1.4.

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

Table 1.4: Regions of interest for muon g-2 anomaly in the type X 2HDM in the normal and inverted alignment scenarios as suggested in Reference [42].

Chapter 2

The LHC and CMS experiment

This chapter will explain the apparatus used to generate and collect the datasets that are utilised for the physics analysed described in Chapters 4 and 5. This is split into two parts. Firstly, an overview of the LHC and a description of how proton-proton collisions at a centre-of-mass energy ($\sqrt{s}$) of 13 TeV are achieved. Secondly, an explanation of the CMS detector is given, including a look at the individual sub-detectors, that are crucial to the reconstruction of particles originating from the proton-proton collisions.

2.1 The LHC

The LHC [43], located at the European Organization for Nuclear Research (CERN) just outside Geneva, is a synchrotron measuring 27 km in circumference, installed in the tunnel previously used by the LEP accelerator [44]. Upon design, its primary purpose was to provide collisions between proton beams, generating centre-of-mass energies of up to 14 TeV and an instantaneous luminosity of approximately 10^{34} cm^{-2}s^{-1}. Figure 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 (PS), and the 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 contains 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

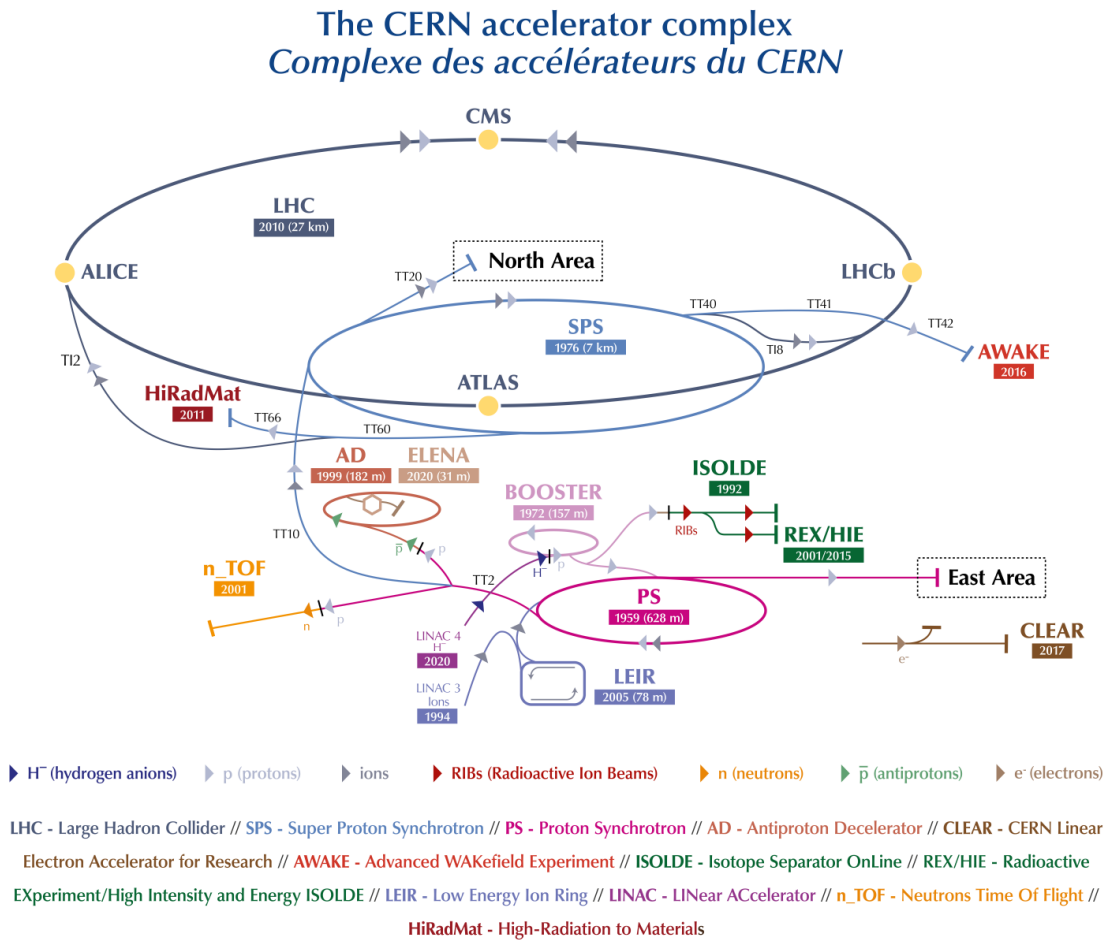


Figure 2.1: A schematic diagram of the CERN accelerator complex [49].

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 intersection points surrounded by the ALICE [45], ATLAS [46], CMS [47] and LHCb [48] detectors, at a collision rate of 40 MHz.

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

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

where σ represents the cross-section of the process and is a function of the centre-of-mass energy, and L_{inst} denotes the LHC machine's instantaneous luminosity, that depends only on the beam parameters and can be calculated by,

$$L_{\text{inst}} = \frac{n_b N_b^2 f_{\text{rev}} \gamma_r}{4\pi \epsilon_n \beta^*} F, \tag{2.2}$$

where n_b is the number of bunches per beam, N_b is the number of particles per bunch, f_{rev} is the revolution frequency, γ_r is the relativistic gamma factor, ϵ_n is the normalised transverse beam emittance, β^* 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 in the instantaneous luminosity is the increase of pileup (PU), defined as the number of additional inelastic proton-proton collisions per bunch crossing.

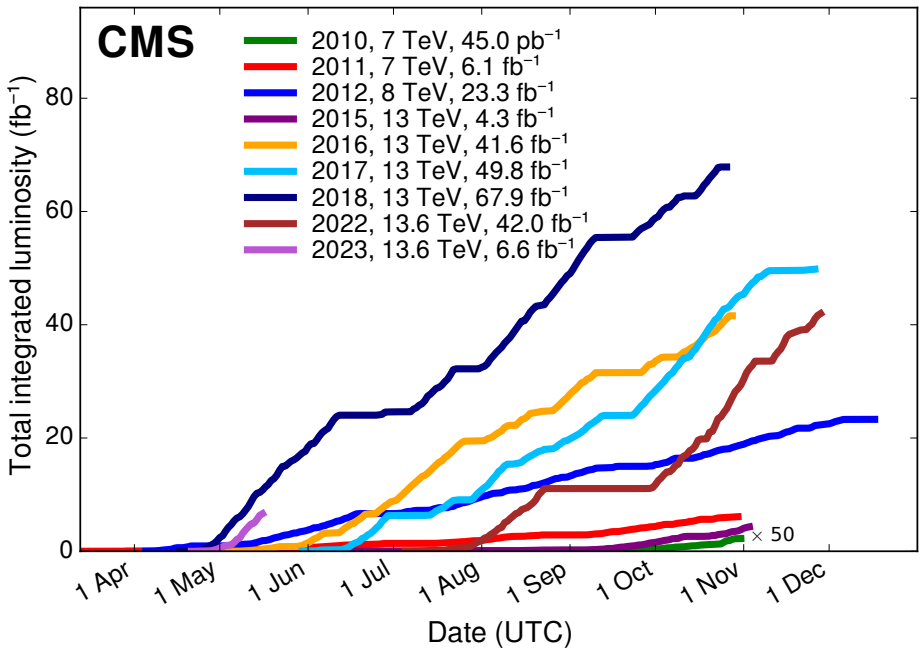


Figure 2.2: The total integrated luminosity for proton-proton collisions collected by the CMS experiment between 2010 and May 2023 [50].

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 2.1×10^{34} cm^{-2}s^{-1} in 2018. Subsequently, the LHC performed a second update period, the Long Shut-

down 2 (LS2), lasting for approximately three years, whereafter the data collection for Run 3 began at $\sqrt{s} = 13.6$ TeV. Figure 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 fb^{-1}, 41.5 fb^{-1} and 59.7 fb^{-1} of data collected in 2016, 2017 and 2018 respectively.

2.2 The CMS detector

The 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 Figure 2.3. A superconducting solenoid, operating at a magnetic field strength of 3.8 T, surrounds 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 centred at the collision point, with the y-axis oriented vertically, the x-axis directed radially inward toward the LHC centre, and the z-axis aligned with the beam direction. The transverse energy (E_T) and transverse momentum (p_T) are defined in the x-y plane. The azimuthal angle (ϕ) and polar angle (θ) are measured relative to the x-axis in the x-y plane and z-axis relative to the z-y plane, respectively. r is used as the radial distance in the x-y plane. Pseudorapidity (η), defined as $\eta = -\ln[\tan(\theta/2)]$, is used due to its gauge invariance, and distances between objects in the ϕ-η 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 [47, 51, 52], which is essential for precise measurements of charged particle trajectories and the determination of the primary vertex (PV) and other vertices, as explained in Sec-

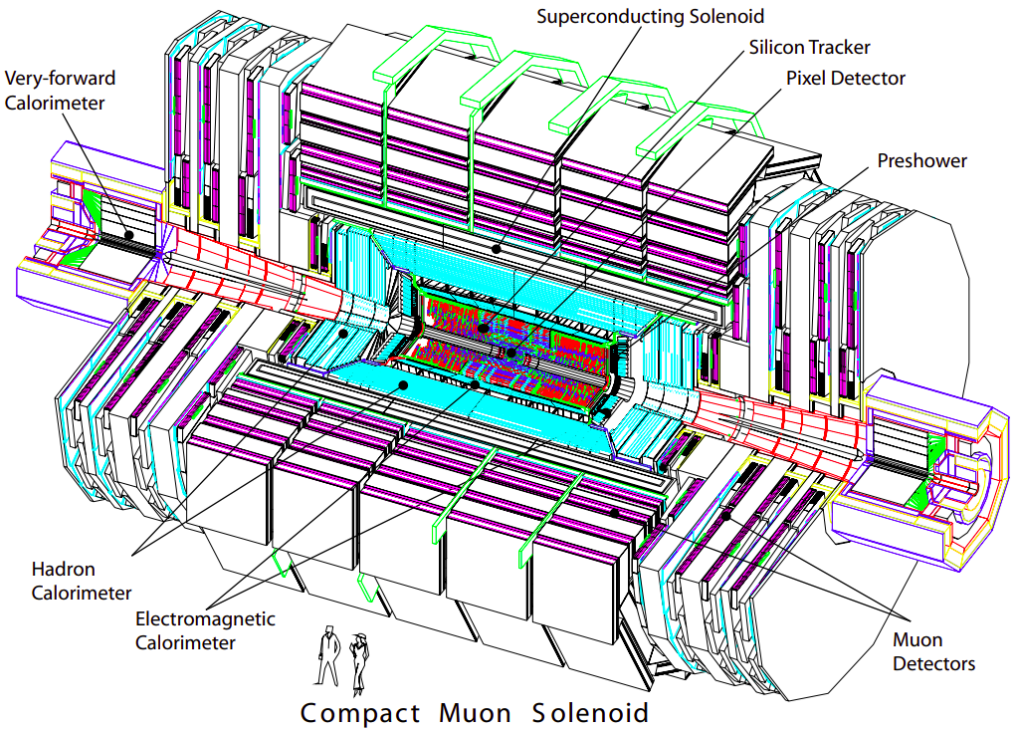


Figure 2.3: A perspective view of the CMS detector [47].

tion 3.1. 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 Figure 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 100×150 μm^2. This configuration provides a spatial resolution of 15–20 μm in both the r-ϕ plane and the z direction, enabling three-dimensional vertex reconstruction. The original tracker was built to handle an instantaneous luminosity of 10^{34} cm^2s^{-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 [52]. 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

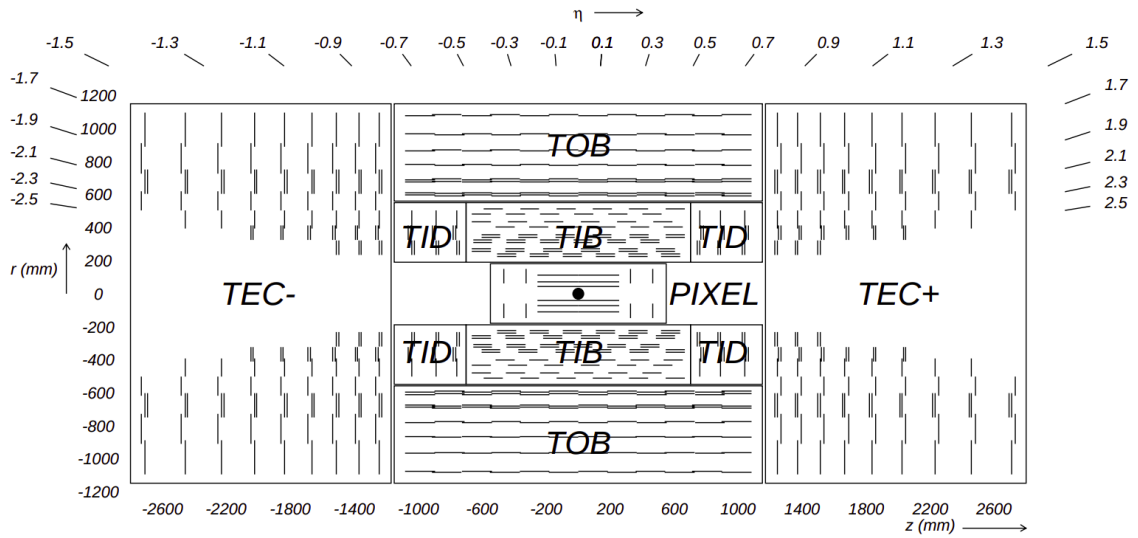


Figure 2.4: Schematic of the CMS tracker (pre-pixel upgrade) in the r-z plane, showing the position of the pixel detector as well as the TIB, TID, TOB, and TEC strip detectors. The lines represent detector modules and the double lines represent back-to-back modules [47].

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 thicknesses and widths, providing multiple measurements of the r-ϕ position with resolutions ranging from 23-35 µm in the TIB to 35-53 µm 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 two-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 [47, 53]. In total, 75,848 lead tungstate ($PbWO_4$) crystals are sorted in a barrel and endcap configuration. $PbWO_4$ was chosen as the crystal material due to its radiation hardness, high den-

sity, 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 Figure 2.5. The barrel region consists of 61,200 crystals with 360-fold granularity in ϕ and 170-fold granularity in η. The crystals are tapered with a front face area of 0.0174×0.0174 in η-ϕ (22×22 mm^2) and a length of 230 mm. The endcaps house 7,324 crystals arranged in a rectangular x-y grid, each with a front face area of 28.62×28.62 mm^2 and a length of 220 mm.

The ECAL system also includes pre-shower 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 parametrised as a function of the incident particle energy E, with terms for the stochastic S, noise N, and constant C contributions.

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

The stochastic term captures fluctuations in lateral shower containment and photon yield, the noise term accounts for electronics' noise and PU, and the constant term arises from the 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 Hadronic calorimeter

The CMS detector includes the hadronic calorimeter (HCAL), which plays a crucial role in measuring the energies of strongly interacting particles and being able to

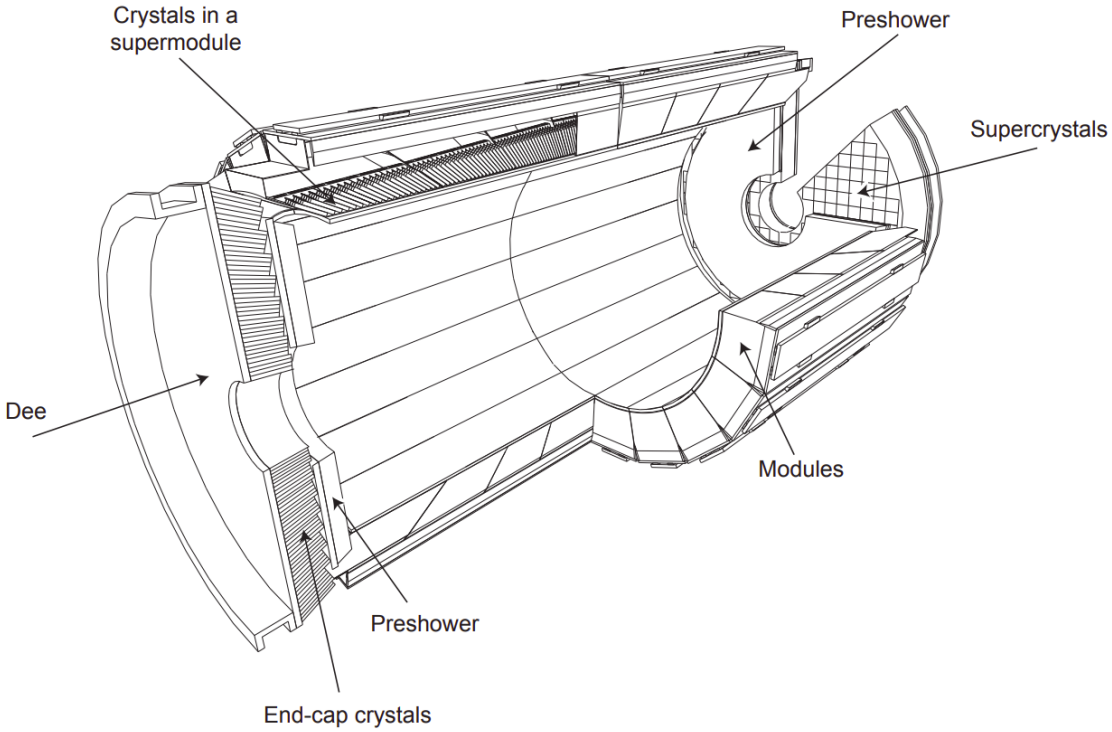


Figure 2.5: A schematic of the CMS ECAL detector [47].

reconstruct missing energy signals. The HCAL is located outside of the ECAL and covers the pseudorapidity region $|\eta| < 5.2$. It consists of four sub-detectors: the hadron barrel (HB), hadron endcap (HE), hadron outer (HO), and hadron forward (HF) calorimeters, arranged as shown in Figure 2.6.

The HB and HE calorimeters are positioned within the magnet bore, covering the pseudorapidity regions $|\eta| < 1.3$ and $1.3 < |\eta| < 3.0$, respectively. They are constructed using alternating layers of brass absorber plates and plastic scintillator tiles. Brass is chosen as the absorber material due to its short nuclear interaction length and non-magnetic properties. The HB provides 5.82 to 10.6 interaction lengths of absorber material, while the HE provides approximately 10 interaction lengths. The plastic scintillator tiles collect scintillation light emitted by charged particles in the hadronic showers, which is then read out using hybrid photodiodes.

To extend the HCAL's containment capability in the central detector region ($|\eta| < 1.3$), the HO calorimeter is utilised. It is placed outside of the solenoid coil and utilises plastic scintillator tiles as the active material. The HO enhances the thickness of the HCAL to a minimum of 11.8 interaction lengths, improving the measurement of late-starting or highly penetrating showers.

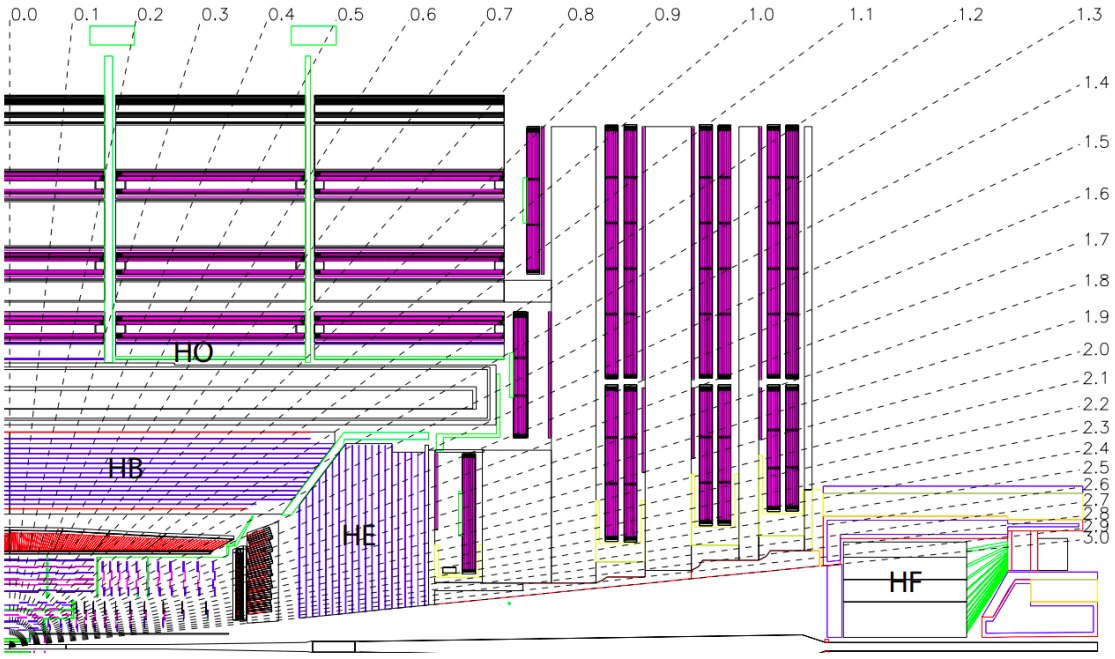


Figure 2.6: A schematic of the CMS HCAL showing the arrangement of the HB, HE, HO, and HF calorimeters. The numbers represent the $|\eta|$ value in the detector [47].

In the forward regions ($|\eta| > 3.0$), the HF detector extends the pseudorapidity coverage of the HCAL up to $|\eta| < 5.2$. The HF is subjected to a high flux of incoming particles, and therefore, requires extremely radiation-hard materials. Quartz fibres embedded in steel absorbers are employed as the active material. Charged particles in the showers generate Cherenkov light in the quartz fibres, which is detected by photomultiplier tubes.

The energy resolution of the HCAL can be parametrised as a function of the incident particle energy E, using the formula,

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + C^2 \tag{2.4}$$

where the stochastic term S is measured to be 0.943 $\text{GeV}^{1/2}$ and the constant term C is measured to be 0.084. During the LS1 period, upgrades were performed on the photomultiplier tubes of the HF.

2.2.4 Muon system

The muon system in the CMS detector plays a crucial role in the accurate identification and measurement of muons, which are essential for various physics analyses.

The system is designed with three primary functions: efficient muon identification, precise momentum measurement, and triggering capability.

Located outside the solenoid coil and HO calorimeter, the muon system is strategically positioned between the iron plates forming the flux return yoke. This arrangement takes advantage of the high-field solenoidal magnet and the yoke's structure to enable the desired functions. The muon system covers a wide pseudorapidity range, specifically $|\eta| < 2.4$. It consists of three types of gaseous detectors: drift tube (DT) chambers, cathode strip chambers (CSC), and resistive plate chambers (RPC).

In the region of $|\eta| < 1.2$, the muon system employs DT chambers, which are organised into four stations. These include inner and outer stations positioned inside and outside the magnet return yoke, as well as two stations inter-spaced between the iron layers of the yoke. The DT chambers are composed of rectangular drift cells filled with a mixture of argon and carbon dioxide gas. Each cell features an anode wire running along its length. The cells are arranged into superlayers, each containing four layers of cells. The DT chambers in the three innermost muon stations consist of three superlayers, whereas the outer station contains only two superlayers. The spatial resolution of the DT chambers is measured to be 77–123 μm for the ϕ coordinate measurement and 133–393 μm for the z coordinate measurement.

In the region of $0.9 < |\eta| < 2.4$, the CSCs are employed. These chambers are trapezoidal-shaped multiwire proportional chambers with six layers of gas. The gas mixture used includes argon, carbon dioxide, and tetrafluoromethane. The CSC modules are organised into four stations positioned between the endcap layers of the magnet's return yoke. They consist of layers of cathode strips and anode wires, allowing for measurements of the muon's ϕ and r coordinates, respectively. The spatial resolution per chamber of the CSCs ranges from 45 μm to 143 μm.

To enhance the system's capabilities in the $|\eta| < 1.6$ region, RPCs are used. These parallel-plate detectors feature a double-gap module design with anode and cathode plates separated by a gas gap. The RPCs are embedded within the barrel and endcap iron yokes. In the barrel, six layers of RPC chambers are present, with two layers in each of the two innermost muon stations and one layer in each of the two outer stations. In the endcap, the RPCs consist of four layers positioned on either side of the three iron disks. While the RPCs exhibit a lower spatial resolution compared

to the DTs and CSCs (0.78–1.38 cm per chamber), they offer a remarkably fast response time, making them suitable for dedicated muon triggering.

2.2.5 Triggering and computing

The LHC collides protons at a rate of 40 MHz during its operation. However, it is impractical to read out and store every event due to the high data volume involved. To address this challenge, the CMS detector employs a dedicated trigger system that selectively chooses the most interesting events, reducing the recorded rate to around 1 kHz.

The CMS trigger system consists of two stages: the Level-1 (L1) trigger and the high-level trigger (HLT). The L1 trigger, implemented with custom-built programmable electronics, operates as the first stage and reduces the event rate to approximately 100 kHz. During the upgrade in 2015/2016, the L1 system was enhanced to accommodate the increased instantaneous luminosity and PU conditions. Upgrades included improved muon momentum resolution, electron/photon isolation, hadronic tau identification and isolation, jet-finding algorithms with PU subtraction, and enhanced trigger menu capabilities.

The L1 trigger utilises a time multiplexed architecture, where energy deposits recorded in the calorimeters and hits from the muon detectors are processed. In the calorimeter trigger, energy deposits from the HCAL and ECAL are passed through two layers (Layer 1 and Layer 2), enabling object identification and the selection of the best candidates based on their p_T. Simultaneously, hits from the muon detectors are combined in the muon trigger's tracking finder and sorting/merging layers, producing a sorted list of muon candidates for the entire detector. The global trigger then integrates the information from both calorimeter and muon triggers to decide on event selection within a maximum storage time of 3.2 µs. A diagrammatic representation of this is shown in Figure 2.7.

The second stage of the trigger system is the HLT, which is a software-based trigger running on a multi-processor farm with thousands of CPU cores. The HLT utilises information from the full detector, including the tracker, calorimeters, and muon systems. By closely following the algorithms and selections used offline, the HLT ensures good momentum resolution and high identification efficiencies for selected

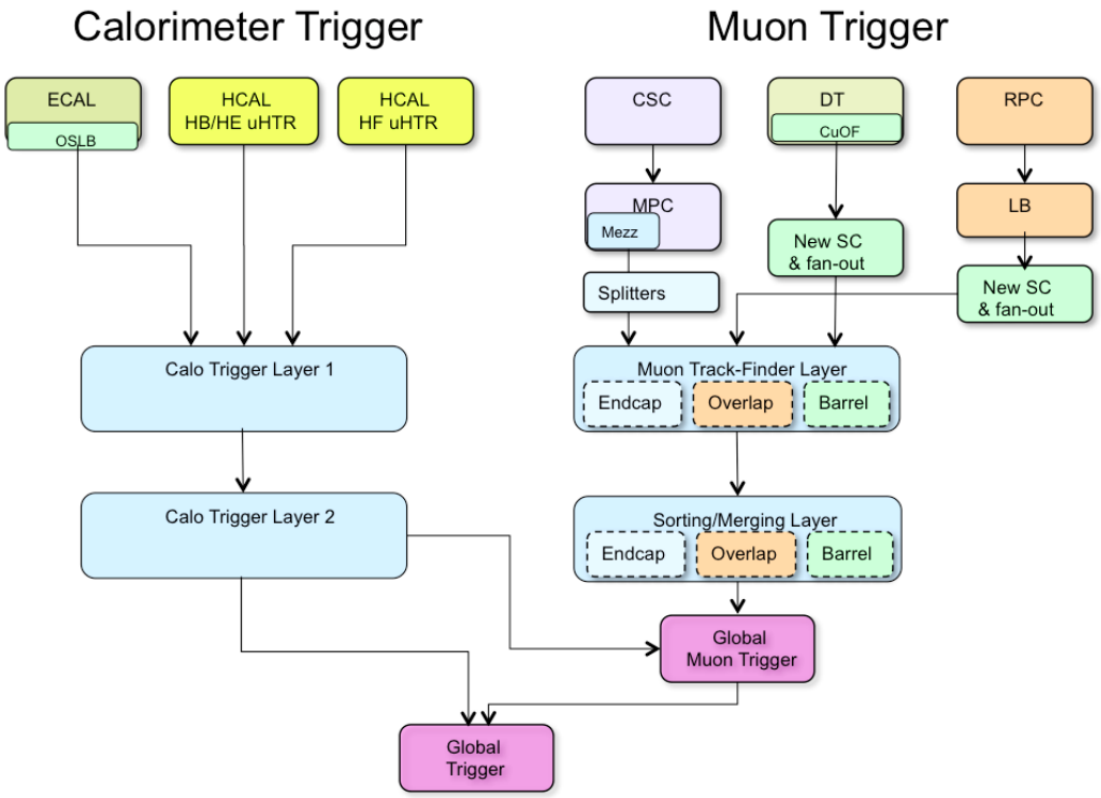


Figure 2.7: A schematic of the L1 trigger workflow for object reconstruction [54].

objects. From the initial event rate of approximately 100 kHz received from the L1 trigger, the HLT further reduces the output rate to about 1 kHz.

The events selected by the HLT produce a significant amount of data, approximately $\mathcal{O}(10\text{ pb})$ per year. To process, store, and facilitate easy access for analysts worldwide, CMS collaborates with other LHC experiments through the Worldwide LHC Computing Grid (WLCG). The WLCG combines computing resources from research institutions globally, organised into three tiers. Tier-0 sites, comprising the CERN Data Center and the Wigner Research Centre in Budapest, perform full data reconstruction and store a copy of the data on tape. The data is then distributed to Tier-1 centres, which handle the storage of raw, reconstructed, and simulated data, as well as the distribution to Tier-2 sites. Tier-2 sites, located at universities and research institutes worldwide, provide computing resources for data production, reconstruction, and analysis by researchers.

Chapter 3

Object reconstruction

This chapter will overview the methods used to reconstruct the objects in each event, that are critical for the searches described in Chapters 4 and 5. The topics discussed here are the reconstruction of tracks and vertices, the particle flow algorithm, the calculation of the missing transverse energy (MET), the measurements of jets and the tagging of their flavours, and the identification of electrons, muons, and tau (τ) leptons.

3.1 Tracks and vertices

The Combinatorial Track Finder (CTF) algorithm [55] is employed by CMS to reconstruct particle tracks. This algorithm consists of four steps to estimate trajectory parameters and uncertainties. Initially, track seeds are formed using hits in the first few layers of the tracker. These seeds provide initial estimates of the trajectory. Next, the track seeds are extrapolated using a Kalman filter (KF) [56], searching for additional hits along the expected trajectory in successive detector layers. The trajectory parameters are continuously updated as hits are added to the track candidate. The extrapolation process continues until the final detector layer is reached.

Afterwards, a KF and smoother are used to fit the final trajectory iteratively. Spurious hits are identified and removed after each iteration until no more spurious hits are found. This fitting process provides the most accurate estimates of the trajectory parameters. Subsequently, tracks must pass a set of quality criteria, and any failing tracks are discarded [56].

The CTF algorithm performs six iterations, gradually reconstructing more challeng-

ing tracks and recovering missed tracks from previous iterations. Associated hits are removed after each iteration, reducing combinatorial complexity and facilitating the search for complicated tracks in later iterations. Tracking efficiencies for muons and pions with p_T greater than 500 MeV have been measured to be over 99.3% and 98.5%, respectively [57].

Once all tracks are reconstructed, the positions of interaction vertices, including the PV and additional vertices from PU, are determined. Vertex reconstruction begins by selecting tracks consistent with being promptly produced near the beamspot. Tracks are then clustered based on the z coordinate of their closest approach to the beamspot, using a deterministic annealing algorithm [58]. Candidate vertices are retained if at least two of their associated tracks are incompatible with other vertices. The candidate vertices are fitted using an adaptive vertex fitter [59] to determine the best estimate of their three-dimensional positions. The PV is determined as the vertex with the highest summed physics-object p_T.
The efficiency of vertex reconstruction has been measured, with values exceeding 98.7% for vertices containing at least two tracks and exceeding 99.9% for vertices with four or more tracks [57].

3.2 Particle flow

The particle flow (PF) algorithm [60–62] is used for the reconstruction of stable particles in an event, such as electrons, photons, muons, and hadrons. All the subdetectors of CMS are used to achieve accurate identification and measurements of particle energies and directions. The final output of the algorithm is a list of particles and their four vectors, that can then be used to construct more complicated objects such as jets, hadronically decaying taus, and the MET.

The tracking and calorimeter clustering information are the initial inputs to the PF algorithm. The tracking information includes the high-resolution momentum and direction of charged hadrons calculated by an iterative-tracking strategy [63], which is not achievable with the calorimeters. The calorimeter clustering is used to achieve multiple objectives, including detecting and measuring the energy and direction of neutral particles, separating neutral particles from charged hadrons, reconstructing electrons and bremsstrahlung photons, and aiding in the energy measurement of charged hadrons with inaccurate track parameters. The clustering algorithm

involves identifying "cluster seeds", growing "topological clusters", and generating "particle flow clusters" based on energy thresholds and cell-cluster distances in each sub-detector, except for the HF, where each cell forms a cluster and is explained in more detail below.

The clustering algorithm used in the calorimeters can be split into three stages. Firstly, "cluster seeds" are identified as local maxima of energy in calorimeter cells above a specified threshold. Secondly, "topological clusters" are formed by combining cells that have at least one side in common with an existing cluster cell and exceed a given energy threshold. These thresholds corresponding to the electronics' noise are 80 MeV in the ECAL barrel, 300 MeV in the ECAL end caps and 800 MeV in the HCAL. The number of seeds dictates how many "particle flow clusters" are formed from a topological cluster. Utilising the granularity of the calorimeter, an iterative process is employed to determine cluster energies and positions by distributing the energy of each cell among all particle flow clusters based on the distance between the cell and the cluster.

Muon reconstruction involves the matching of tracks from the inner tracker to tracks in the muon system [64,65]. Initially, tracks are independently reconstructed in each system. Standalone muon tracks are reconstructed in the muon system using seed groups of CSC or DT segments. These seeds provide an initial trajectory estimate, which is extended using a KF to incorporate hits from DT, CSC, and RPC sub-detectors. In the tracker muon reconstruction, tracks in the inner tracker with p_T greater than 0.5 GeV and total momentum greater than 5 GeV are extrapolated to the muon system. A track qualifies as a tracker muon if it is matched to at least one muon segment based on its positions in a local x-y coordinate frame. The global muon reconstruction starts with standalone muon tracks and matches them to tracks in the inner detector. If a match is found, a combined fit using the KF is performed with hits from both systems.

Electron reconstruction begins by finding matching tracks with ECAL energy clusters [66]. The path of electrons towards the ECAL involves traversing a considerable amount of material, which constitutes the tracker. The thickness of this material varies depending on the η region, ranging from approximately 0.4 to 2 radiation lengths. Within this material, electrons emit bremsstrahlung photons, resulting in a significant energy loss. At regions with maximal material thickness, electrons lose an

average of 86% of their energy, predominantly spreading out in the ϕ direction due to magnetic field-induced bending. To accurately measure the initial electron energy, the energy from all emitted photons is measured through the supercluster (SC) algorithms. These algorithms initiate clustering with seed crystals, expanding the clusters based on energy thresholds and merging them into the SC.

For reconstructing electron tracks, the standard CTF algorithm is not suitable due to the substantial radiative losses in the tracker material. These losses lead to reduced hit collection efficiency and inaccurate trajectory parameter estimation. To address this, a dedicated tracking procedure is employed [66], which identifies seeds consisting of two or three hits in the tracker using two complementary methods: ECAL-based seeding and tracker-based seeding. ECAL-based seeding uses SC energy and position to estimate the trajectory, while tracker-based seeding starts with tracks reconstructed using the CTF algorithm. The selected seeds are then extrapolated and smoothed with the Gaussian sum filter (GSF) instead of the KF. The GSF, with a Gaussian mixture to model bremsstrahlung energy loss, provides a more accurate estimation of the trajectory parameters.

PF elements are connected by a link algorithm that evaluates the quality of the connection between each pair of elements by defining a distance [60]. Elements linked directly or indirectly, form "blocks" that serve as inputs for the particle reconstruction and identification algorithm. The granularity of the CMS sub-detectors ensures that blocks typically contain one to three elements, allowing the algorithm to perform independently of event complexity. The link between a charged-particle track and a calorimeter cluster is established by extrapolating the track to different depths in the ECAL and HCAL, and the track is linked to a cluster if the extrapolated position lies within the cluster boundaries. Similarly, links between calorimeter clusters are established based on their positions in the more granular calorimeter. A link between a charged particle track in the tracker and a muon track in the muon system is established through a global fit, with the link distance determined by the χ^2 value of the fit.

The PF algorithm operates on each block by following a specific set of steps. Initially, global muons within the block are converted into "PF muons" if their combined momentum aligns with that determined solely from the tracker. Afterwards, electron reconstruction and identification take place, where tracks are pre-identified as

potential electrons based on their characteristics. Identified electrons become "PF electrons", and their corresponding tracks and ECAL clusters are removed.

Next, tighter quality criteria are applied to the remaining tracks, ensuring that the measured p_T uncertainty is smaller than the expected calorimetric energy resolution for charged hadrons. Some tracks are rejected based on these criteria, but the energy from genuine tracks is still used in the reconstruction process. The remaining elements in the block can give rise to charged hadrons, photons, neutral hadrons, or additional muons.

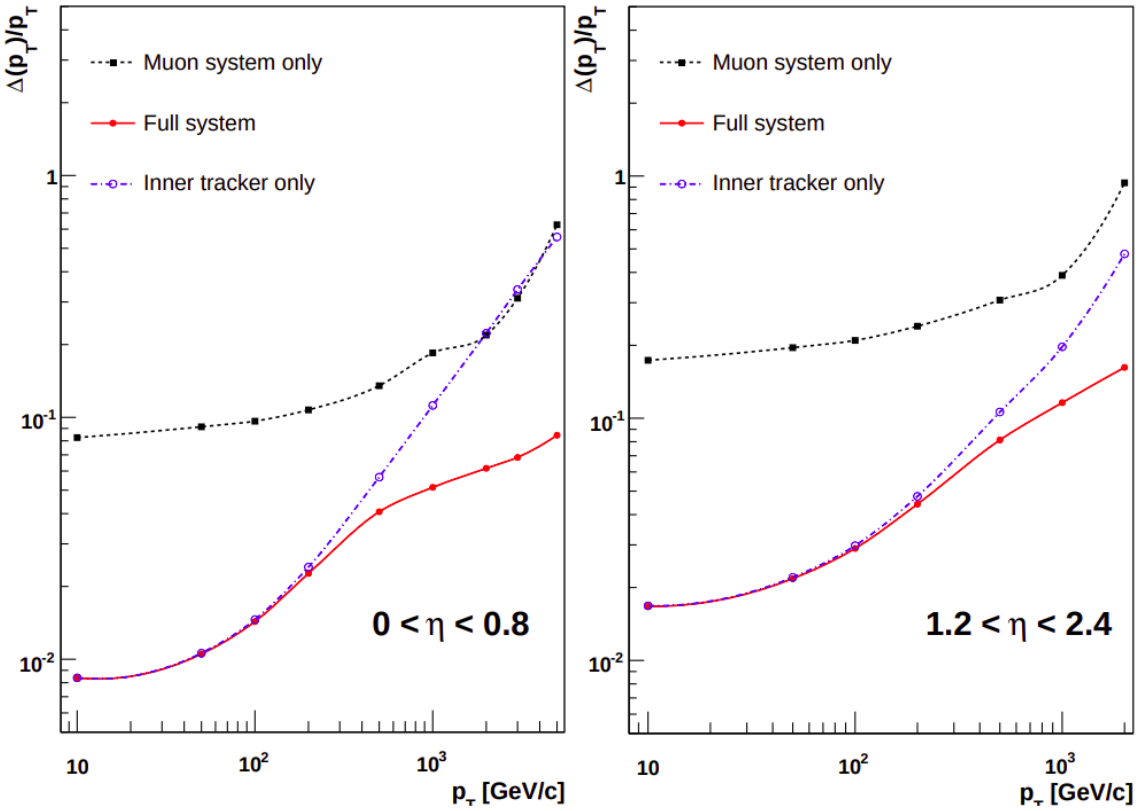
To detect neutral particles, tracks are linked to ECAL and HCAL clusters by comparing momenta and energy measurements. Multiple tracks can be linked to the same HCAL cluster, while the closest cluster is retained when a track is linked to multiple HCAL clusters. Similar considerations apply to ECAL clusters, where ordering based on distance is used to preserve appropriate links.

In cases where the calibrated calorimetric energy is significantly lower than the total track momentum, a search is conducted for additional muons and fake tracks. Tracks are progressively removed based on p_T uncertainty until a certain threshold is reached. The impact of this procedure on the overall track selection is minimal.

Finally, the remaining tracks in the block are identified as "PF charged hadrons", with momentum derived directly from the track measurements. If the calibrated energy of linked ECAL and HCAL clusters exceeds the associated charged particle momentum significantly, "PF photons" and possibly "PF neutral hadrons" are created. The precedence is given to photons over neutral hadrons due to their higher contribution to the jet energy. Unclassified ECAL and HCAL clusters give rise to "PF photons" and "PF neutral hadrons" using the calibration procedure applied to HCAL clusters.

3.3 Muons

Three muon identification techniques are described by the PF algorithm; tracker, global and PF muons. Tracker muons and global muons differ in their requirements for muon segments, with one and two hits required respectively. Subsequently, tracker muon reconstruction efficiency is larger than global reconstruction efficiency,



(a) (b)

Figure 3.1: The p_T resolution for muon identification using the muon system (black), the inner tracker (blue), and the full system (red). This is shown for the pseudorapidity regions $0 < \eta < 0.8$ (a) and $1.2 < \eta < 2.4$ (b) [47].

when dealing with muons with a momentum below 5 GeV. This is because many low-momentum muons are likely to reach only up to the first muon station. However, the muon system reconstruction enhances momentum resolution for muons with a p_T exceeding 200 GeV, whilst below this the inner tracker information is dominant to the resolution and so the tracker muons perform better here, as shown in Figure 3.1.

The `Medium` muon identification [65] is used for the analyses described in subsequent chapters. The requirements for this identification are:

i) The muon is reconstructed by the tracker or global muon reconstruction algorithm.

ii) The impact parameters, d_{xy} and d_z, defined as the distance in the x-y and z plane of the closest approach to the PV, are restricted to $d_{xy} < 0.045$ cm and $d_z < 0.2$ cm to ensure the muon is associated with the PV.

iii) At least 80% of the tracker traversed is required to have hits in.

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

i) The muon is reconstructed by the global muon reconstruction algorithm.

ii) The χ^2 divided by the number of degrees of freedom of the global track fit is smaller than 3.

iii) The χ^2 of the tracker standalone position match [67] is smaller than 12.

iv) The χ^2 of the track kink finder [65] is less than 20.

v) The muon segment compatibility is greater than 0.303.

or,

i) The muon is reconstructed by the tracker muon reconstruction algorithm.

ii) The muon segment compatibility is greater than 0.451

The efficiency of the `Medium` muon identification is approximately 99.5%, with a misidentification probability of approximately 0.1%.

To reduce the contamination from muons originating from heavy-flavoured quark decays, muons are required to be isolated from hadronic activity in the detector. The isolation is based on the p_T of the photon (γ), neutral (h^0) and charged ($\mathrm{h}^\pm$ if originating from the PV and $\mathrm{h}^{\pm,\mathrm{PU}}$ if not originating from PV) hadron PF candidates within a cone size of $\Delta R < 0.4$ of the selected muon (μ). A relative combined isolation variable is defined as,

$$I_{\mathrm{rel}} = \frac{1}{p_T^{\mu}}\Big(\sum p_T^{\mathrm{h}^\pm} + \max\Big(0, \sum p_T^{\mathrm{h}^0} + \sum p_T^{\gamma} - \Delta\beta \sum p_T^{\mathrm{h}^{\pm,\mathrm{PU}}}\Big)\Big), \qquad (3.1)$$

where the sums loop through all objects within the cone and $\Delta\beta$ is the energy estimate of neutral particles due to pileup, which is taken to be 0.5. This represents an estimated fraction of transverse momenta between the other contributions within the cone that originate from the interaction at the PV, and the muon itself. The muon isolation can then be used to separate muons, with a low value of I_{rel}, from other hadronic objects, at higher values of I_{rel}.

3.4 Electrons

In addition to the PF algorithm to reconstruct electrons, they are required to pass an identification variable based on a boosted decision tree (BDT) discriminator to distinguish genuine electrons from backgrounds like misidentified jets, electrons from photon conversions, and electrons produced in heavy-flavour decays [66]. The following variables are used as input to the BDT:

i) Shower shape variables.

ii) The fraction of energy deposited in the HCAL.

iii) The fraction of energy lost through bremsstrahlung.

iv) Track quality variables.

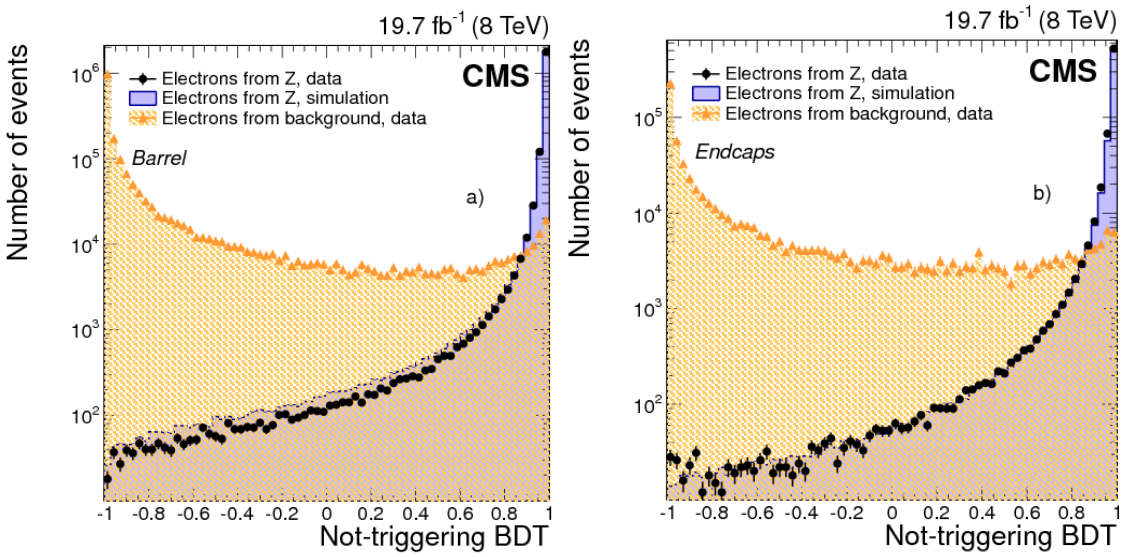
v) The variable $1/E_{\mathrm{SC}} - 1/p$, where E_{SC} is the SC energy and p is the momentum of the track at the point of closest approach to the vertex.

vi) Variables comparing the geometrical matching between the track and the SC.

The BDT was trained on a Z/γ^* monte carlo (MC) sample generated with MadGraph, in three η bins for electrons with $p_T > 10$ GeV. The performance of the BDT is shown in Figure 3.2. This thesis uses the version of the identification BDT without the isolation included in the training, instead using an additional cut on the electron isolation. This allows for easily accessible sideband regions (orthogonal regions to the signal region) used for estimation of the background and optimisation opportunities for τ lepton decays to electrons.

The electrons are required to pass the 90% efficiency working point of the BDT, which corresponds to a misidentification probability of 2% [66]. 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 PV are restricted to $d_{xy} < 0.045$ cm and $d_z < 0.2$ cm to ensure the electron is associated with the PV.

Similarly to muons, an isolation variable is used to further separate electrons from other backgrounds. Different than for muons, the isolation used is based on the photon, neutral and charged hadron PF candidates within a smaller cone size of $\Delta R < 0.3$ of the selected electron. A different method to estimate the neutral particles due to PU is used namely the rho-effective-area method. The PU contribution



(a) (b)

Figure 3.2: Performance of the electron BDT using $Z \to ee$ enriched events from data (dots), background-enriched events from data (triangles), and $Z \to ee$ MC events (purple solid histogram). This is shown for the performance in the barrel (a) and the endcaps (b) for a BDT that does not utilise trigger information during training [66].

in this method is estimated as $\mathrm{PU} = \rho A_{\mathrm{eff}}$, where ρ is the event-specific average PU energy density per unit area in the ϕ-η plane and the A_{eff} is the effective area specific to the given type of isolation. The rho-effective-area subtracted relative combined isolation variable is defined as,

$$I_{\mathrm{rel}} = \frac{1}{p_T^e}\Big(\sum p_T^{\mathrm{h}^\pm} + \max\Big(0, \sum p_T^{\mathrm{h}^0} + \sum p_T^{\gamma} - \rho A_{\mathrm{eff}}\Big)\Big), \tag{3.2}$$

where A_{eff} is measured in bins of η as listed in Table 3.1.

3.5 Jets

Quarks and gluons undergo fragmentation and hadronisation, resulting in collimated sprays of hadrons called jets [68]. The process of combining particles into jets is accomplished through the anti-k_T jet clustering algorithm [69] as implemented in `FastJet` [70]. This algorithm relies on two "distance" parameters: d_{ij}, representing the distance between objects i and j, and d_{iB}, representing the distance between

η range	A_{eff}
$0.0 \leq \lvert\eta\rvert < 1.0$	0.1440
$1.0 \leq \lvert\eta\rvert < 1.479$	0.1562
$1.479 \leq \lvert\eta\rvert < 2.0$	0.1032
$2.0 \leq \lvert\eta\rvert < 2.2$	0.0859
$2.2 \leq \lvert\eta\rvert < 2.3$	0.1116
$2.3 \leq \lvert\eta\rvert < 2.4$	0.1321
$2.4 \leq \lvert\eta\rvert < 5.0$	0.1654

Table 3.1: Electron effective areas used for the ρ-corrected isolation computation [66].

object i and the beam. These objects can be individual particles or clusters of particles. The distance parameters are defined as,

$$\begin{aligned} d_{ij} &= \min(p_{T,i}^{-2}, p_{T,j}^{-2}) \frac{\Delta R_{ij}^2}{R^2}, \\ d_{iB} &= p_{T,i}^{-2}, \end{aligned} \tag{3.3}$$

where ΔR_{ij} is the separation in ΔR of objects i and j, and R is the radius parameter controlling the typical size of the jets. The clustering process unfolds as follows:

i) Compute the minimum of d_{ij} and d_{iB} for all objects.

ii) If the minimum corresponds to d_{ij}, merge objects i and j into a single entity.

iii) If the minimum corresponds to d_{iB}, set object i as a jet and remove it from the object list.

iv) Repeat steps (i) to (iii) until no objects remain.

Chapters 4 and 5 use jets clustered with a radius parameter of $R = 0.4$, often called "AK4 jets". The impact of PU on jet energy and substructure is addressed by using the charged hadron subtraction (CHS) technique [71]. CHS excludes charged hadrons that are not associated with the PV during jet clustering. To reject misidentified jets a discriminator evaluated on variables like energy fractions carried by different types of PF candidates, as well as charged and neutral particle multiplicities, is used [71]. For data collected in 2017, the region with $2.65 < |\eta| < 3.139$ experiences substantial ECAL noise, leading to a notable increase in the number of jets.

To veto this, jets with $p_T < 50$ GeV within this η range are removed and a PU jet identification discriminant is utilised to reject noise [72]. For τ lepton selections in the later chapters, to exclude selected jets from electron, muon and hadronic τ decays, the jets are required to be separated from the selected τ candidate by $\Delta R > 0.5$.

3.6 b jets

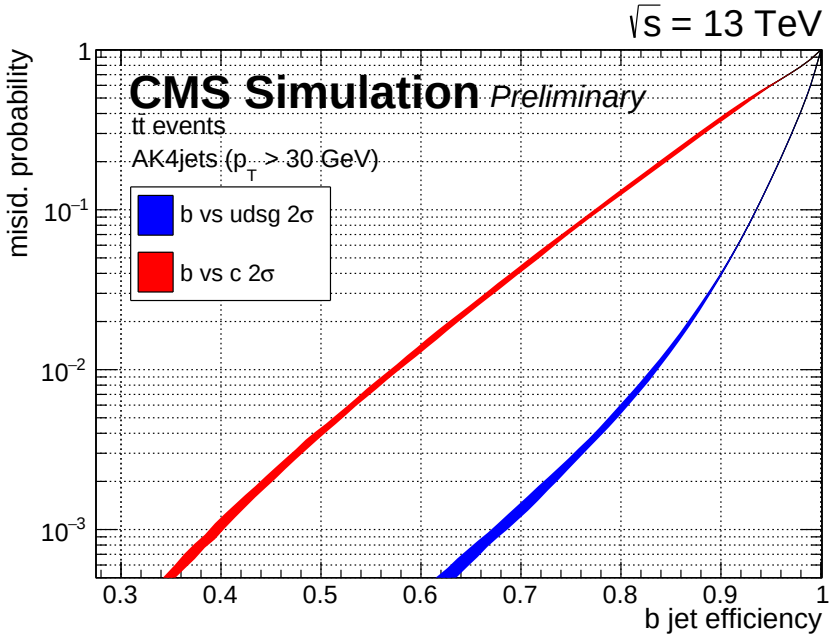
For determining whether a jet is initiated by a b quark, the `DeepJet` algorithm [73, 74] is used. This heavy-flavour jet identifier uses properties of the jet to distinguish jets originating from b quarks, c quarks, the remainder of the lighter quarks grouped, and gluons. Core to the b tagging in this algorithm is the reconstruction of a secondary vertex (SV) due to the lifetime of b quarks, allowing for displaced tracks. The `DeepJet` deep neural network (DNN) uses a combination of high-level variables such as the SV, as well as utilising low-level variables of jet constituents to separate between initiators. The DNN uses a mixture of convolutional and dense layers to perform this categorisation. The performance of the algorithm is shown in Figure 3.3.

For b jet selection, jets are required to have $p_T > 20$ GeV and $|\eta| < 2.4$ (2.5) in 2016 (2017 and 2018), and are considered b tagged if their discriminator value is larger than some threshold that represents a misidentification rate of 1%. This corresponds to a b tagging efficiency of around 80%.

3.7 Missing transverse energy

The MET is used as a feature to understand particles passing through the CMS, such as neutrinos in the SM and weakly interacting particles from hypothetical BSM extensions. This information cannot be pulled from the detector itself but instead, the absence of detection can be used to infer the presence of such an object The MET is nominally calculated as the negative vector sum of all transverse momenta of the PF candidates in the collision event,

$$\vec{E}_T^{\text{miss}} = -\sum_i \vec{p}_T^{\,i}. \tag{3.4}$$



(a)

Figure 3.3: Misidentification probability 2σ bands against efficiency for b tagging using the `DeepJet` algorithm, for AK4jets with $p_T > 30$ GeV and $|\eta|<2.5$, utilising MC with parameters from the 2017 era of data taking. Two bands are displayed, one for the identification of a b quark rather than light quarks and gluons (blue) and one for the identification of a b quark rather than a c quark (red) [75].

However, this raw PFMET is inaccurate due to the p_T thresholds of the calorimeters, the non-linearity of the calorimeter response and the reconstruction inefficiencies [76]. This is fixed with jet energy corrections, determined for jets with $p_T >$ 15 GeV and less than 90% for their energies deposited in the ECAL for both of the problems mentioned. Then the corrected MET is calculated by,

$$\vec{E}_T^{\text{miss,corr}} = \vec{E}_T^{\text{miss}} - \sum_{\text{jets}} (\vec{p}_{T,\text{jet}}^{\text{corr}} - \vec{p}_{T,\text{jet}}^{\text{raw}}), \tag{3.5}$$

where "corr" and "raw" are the corrected and uncorrected values of the jet momenta, respectively.

Further corrections are required to account for PU effects, and the `PUPPI` algorithm is used for this purpose [77]. The algorithm is designed to address the impact of PU on observables involving clustered hadrons, such as jets, missing transverse momentum, and lepton isolation. It achieves this by combining information about the local particle distribution, event PU properties, and tracking data. `PUPPI` operates at the level of individual particle candidates before any clustering is performed. It assigns a weight to each particle, ranging from 0 to 1, based on the information from surrounding particles. A weight of 1 is given to particles believed to originate from the PV. These per-particle weights are then used to rescale the four-momenta of the particles, effectively correcting for the impact of PU and so altering the raw calculation of the MET to,

$$\vec{E}_T^{\text{miss}} = -\sum_i w_i \cdot \vec{p}_T^{\,i}. \tag{3.6}$$

3.8 Taus

Fundamental to this thesis is the identification of τ particles. The τ lepton is measured to have a mean lifetime of 2.9×10^{-13}s. This short lifetime means that the τ lepton is not directly observable in the CMS detector. To detect these particles, it is important to understand how the τ 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 τ lepton is shown in Table 3.2. These decays can be split into three groups: the 17.8% of τ leptons that decay to an electron (e), the 17.4% that decay into a muon (μ), and hadronic tau decays (τ_h) that make up the final 64.8% of τ decays. The leptonic decay of the τ can be accounted for by

the identification of electrons and muons as discussed in the previous subsection.

Decay Mode	Branching Fraction
Leptonic Decay (e, μ)	**35.2%**
$e^-\bar{\nu}_e\nu_\tau$	17.8%
$\mu^-\bar{\nu}_\mu\nu_\tau$	17.4%
Hadronic Decay (τ_h)	**64.8%**
$h^-\pi^0\nu_\tau$	25.9%
$h^-\nu_\tau$	11.5%
$h^-h^+h^+\nu_\tau$	9.8%
$h^-2\pi^0\nu_\tau$	9.3%
$h^-h^+h^-\pi^0\nu_\tau$	4.8%
other	3.2%

Table 3.2: Measured branching fractions for the τ lepton. h represents a charged hadron, typically either a pion or a kaon [16].

A two-step process is used to identify a τ_h. The hadron-plus-strips (HPS) algorithm is used to initially identify τ_h objects based on jets produced by the anti-k_{T} algorithm with a distance parameter of $R = 0.4$. To capture the energy deposits left by π^0 candidates in the ECAL, the photon and electron constituents of the jet responsible for seeding the τ_h reconstruction are assembled into strips. All electrons or photons used are required to have $p_{\mathrm{T}} > 0.5$ GeV. The initial iteration of the HPS algorithm used a fixed strip size of $\Delta\eta \times \Delta\phi$ equal to 0.05×0.20. However, this technique was updated to a dynamical strip size to account for the multiple scatterings of e^+e^- products from a π^0 decay, falling outside of the fixed window. A reliance on the τ_h p_T spectrum of the required strip size was observed and so the following iterative algorithm was proposed to resolve this issue.

i) The highest p_T electron or photon (not previously grouped into a strip) is used to initiate a new strip.

ii) The second highest p_T electron or photon deposition within,

$$\Delta\eta = f(p_T^{e/\gamma}) + f(p_T^{\mathrm{strip}}), \qquad \Delta\phi = g(p_T^{e/\gamma}) + g(p_T^{\mathrm{strip}}) \tag{3.7}$$

of the strip is then combined with the strip. These functions are determined from a fit to τ_h objects from MC so that 95% of electrons and photons are

contained in a single strip. These fits are shown in Figure 3.4 and correspond to,

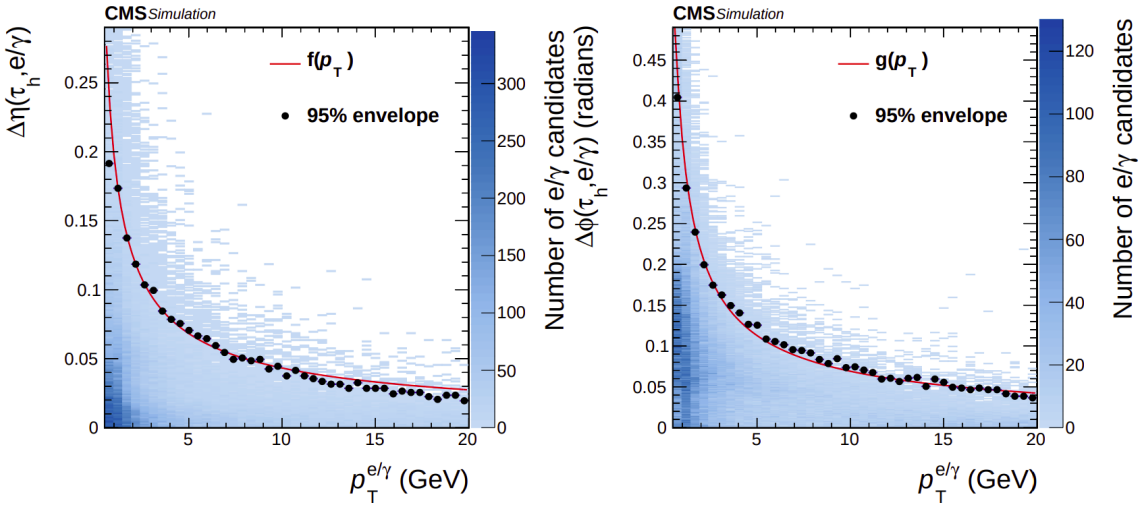
$$f(p_T) = 0.20 p_T^{-0.66}, \qquad g(p_T) = 0.35 p_T^{-0.71} \tag{3.8}$$

where the p_T is in units of GeV. These functions have lower and upper caps of 0.05 to 0.3 for $\Delta\phi$ and 0.05 to 0.15 for $\Delta\eta$.

iii) Recalculate the strip position using a weighted average of the p_T of all the electron and photon strip constituents.

$$\eta_{\text{strip}} = \frac{1}{p_T^{\text{strip}}} \sum p_T^{e/\gamma} \eta_{e/\gamma}, \qquad \phi_{\text{strip}} = \frac{1}{p_T^{\text{strip}}} \sum p_T^{e/\gamma} \phi_{e/\gamma} \tag{3.9}$$

iv) Repeat steps (ii) and (iii) until no other electron or photon candidate fulfilling condition (ii) is found.



(a) (b)

Figure 3.4: The distance between the τ_h and electron or photon η (a) and ϕ (b) with respect the electron or photon p_T. The binned values (points) and the fitted functions f and g (red line), which encapsulates 95% of all electron and photons are shown in both cases [78].

Charged hadrons (prongs) are also required to have $p_T > 0.5$ GeV and originate from the PV, with a loose transverse impact parameter of $d_{xy} < 0.1$ cm. Further constraints are placed on the reconstructed masses of the specific resonances if produced in a τ_h decay. In particular, the visible mass positions and widths of the grouped charged hadrons and strips are optimised to match the ρ (770 MeV) and

a_1 (1260 MeV) decays. This is performed to maximise the fraction of τ_h identification efficiency to the probability of misidentification from jets. The visible decay products of each decay shown in Table 3.2 are reconstructed by:

i) $h^-\pi^0$: One charged hadron candidate and one strip with mass $0.3 < m_\tau < 1.3\sqrt{p_\mathrm{T}/100}$ GeV. The upper mass limit is required to be between 1.3 and 4.2 GeV. The mass selection is performed to target the ρ resonance.

ii) h^-: One charged hadron candidate and no strips.

iii) $\pi^-\pi^-\pi^+$: Three charged hadron candidates 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. The mass selection is performed to target the a_1 resonance.

iv) $h^-2\pi^0$: One charged hadron candidate and two strips. The τ_h mass should be $0.4 < m_\tau < 1.2\sqrt{p_\mathrm{T}/100}$ GeV. The upper mass limit is required to be between 1.2 and 4.0 GeV. The mass selection is performed to target the a_1 resonance.

v) $\pi^-\pi^-\pi^+\pi^0$: Three charged hadron candidates and one strip.

The remaining hadronic decays of the τ leptons are not included in this thesis. The decay mode (DM) of the τ_h object that is reconstructed by HPS is quantified by the following formula relating the number of charged hadron candidates N_C, and the number of strips N_N.

$$\mathrm{DM} = 5(N_C - 1) + N_N \tag{3.10}$$

The second step of the identification comes from a multiclass DNN-based algorithm named `DeepTau`, which seeks to discriminate τ_h decays from electrons, muons, and most importantly quark or gluon jets, which can be misidentified as τ decays [79]. It uses a DNN architecture that consists of multiple interconnected layers of nodes, that attempt to learn whether the input object is a τ_h decay, an electron, a muon or a jet. The algorithm takes inputs from reconstructed particles surrounding the HPS τ_h candidate, including information about energies, momenta, and spatial positions. Convolutional layers are used to efficiently process these inputs by dividing them into smaller regions in η-ϕ space, which allows the algorithm to extract local patterns and features. It also incorporates high-level features of the τ_h candidate calculated from the HPS algorithm, such as the four-momentum, charge, DM, isolation variables used in previous multivariate analysis (MVA) [80], impact parameters,

η and ϕ strip information, as well as event level information such as variables related to the PU. The architecture of the algorithm is shown in Figure 3.5.

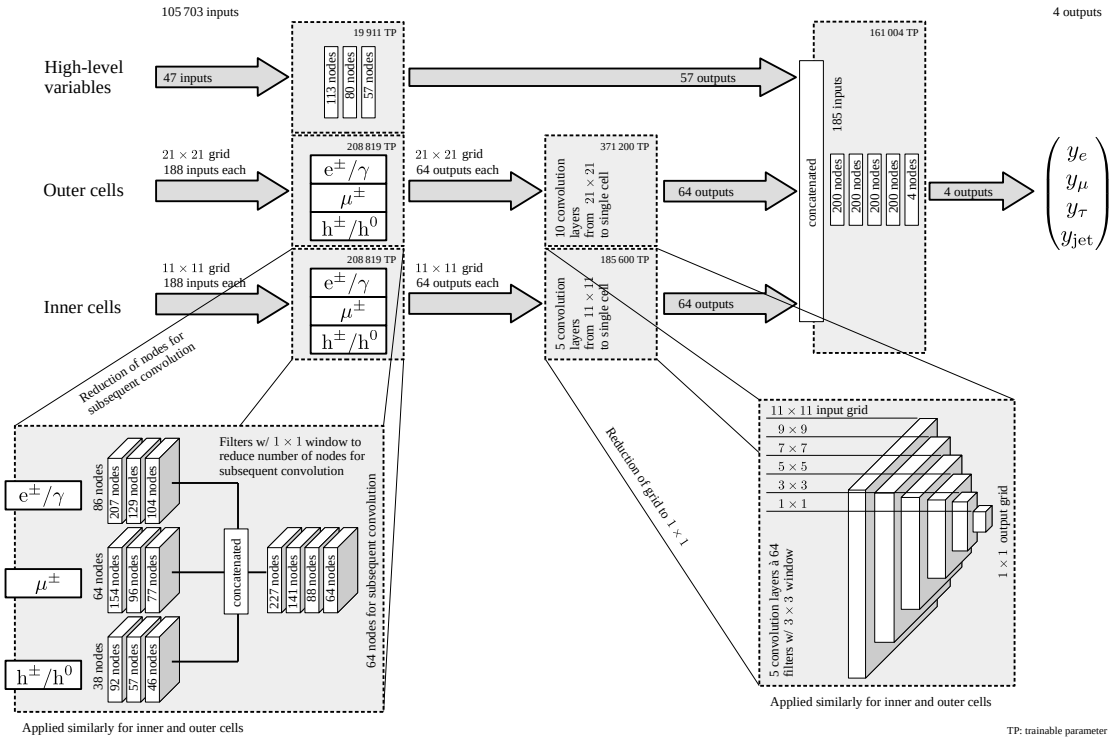


Figure 3.5: The architecture of the `DeepTau` neural network, comprising three sets of input variables: inner cells, outer cells, and high-level features. These sets are processed separately through subnetworks and their outputs are concatenated. Five fully connected layers process the concatenated output to calculate the probabilities for a candidate to be a τ_h, electron, muon, or jet. The high-level input subnetwork consists of three fully connected layers, taking 47 inputs and yielding 57 outputs. Complex subnetworks process the features of inner and outer cells separately, with fully connected layers followed by concatenation and additional fully connected layers. Convolutional layers progressively reduce the grid size for both inner and outer cells. For inner cells, there are 5 convolutional layers, while for outer cells, there are 10 convolutional layers. The number of trainable parameters for the different subnetworks is also provided [79].

The `DeepTau` DNN is trained using a large dataset, incorporating examples of τ_h decays and background processes of electrons, muons and jets. The output of the DNN is four scores, that represent the probability that the object is a τ_h (y_τ), an electron (y_e), a muon (y_μ) or a jet (y_{jet}). From these raw scores, additional scores are

calculated for the probability that an object is a τ_h rather than an electron, muon or jet. These are defined as,

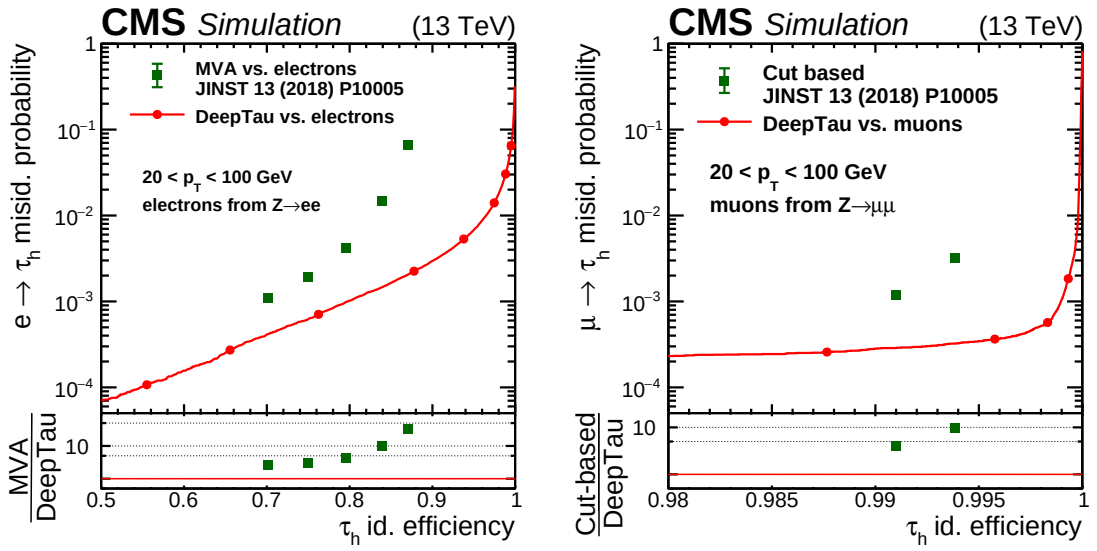
$$D_i^{\text{score}} = \frac{y_\tau}{y_i + y_\tau}, \qquad i \in (e, \mu, \text{jet}). \tag{3.11}$$

From this, working points (WP) are defined to match specific efficiencies of τ_h identification to each of the other three objects, named D_e^{WP}, D_μ^{WP} and $D_{\text{jet}}^{\text{WP}}$. The target efficiencies for different WP are shown in Table 3.3.

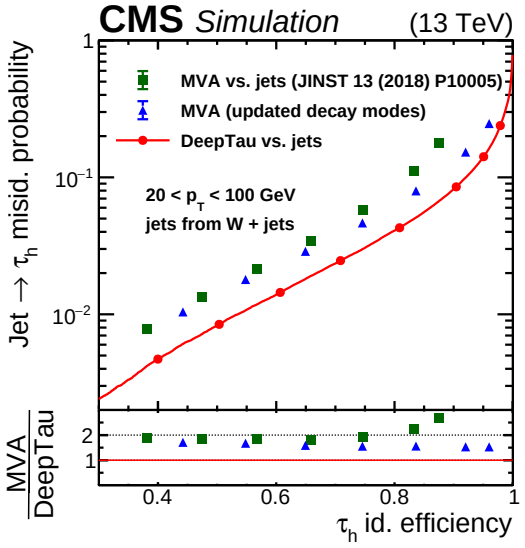
WP	VVTight	VTight	Tight	Medium	Loose	VLoose	VVLoose	VVVLoose
D_e^{WP}	60%	70%	80%	90%	95%	98%	99%	99.5%
D_μ^{WP}	-	-	99.5%	99.8%	99.9%	99.95%	-	-
$D_{\text{jet}}^{\text{WP}}$	40%	50%	60%	70%	80%	90%	95%	98%

Table 3.3: Target efficiencies of the `DeepTau` working points with respect to the electrons, muons and jets discriminators [79].

Due to the lack of muons misidentified as τ_h candidates, high efficiencies can be required for D_μ^{WP}. This is also the reason why the target efficiencies for D_e^{WP} are higher than for $D_{\text{jet}}^{\text{WP}}$, where the large challenge of jets misidentified as τ_h candidates are present. The misidentification probabilities for the different τ_h efficiencies are shown in Figure 3.6. Also shown in these plots, is the performance comparison to an older MVA algorithm for discrimination against electrons and jets, as well as a cut-based algorithm for muons, both described in Reference [80]. Significant improvements compared to the previous algorithm are observed, which results in a significant improvement in analyses utilising τ leptons at the CMS experiment. The different D_e^{WP}, D_μ^{WP} and $D_{\text{jet}}^{\text{WP}}$ can be used to optimise the τ_h identification for analyses attempting to separate τ-enriched signals from backgrounds of misidentified objects. Two examples of analyses that utilise this are described in Chapter 4 and 5.



(a) (b)



(c)

Figure 3.6: Comparisons of the `DeepTau` identification discriminator performance against electrons (a), muons (b) and jets (c) versus the τ_h efficiency. (a) uses $Z \to ee$, (b) uses $Z \to \mu\mu$ and (c) uses W+jets MC, with only objects with $p_T < 100$ GeV used. Working points are indicated by full circles and previous algorithms described in Reference [80] are shown in blue triangles and green squares [79].

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, τ particles are sensitive to resonant production of new neutral particles where the couplings have a mass hierarchy. They are also sensitive to non-resonant effects from new physics mediators. This chapter details 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 a single narrow spin-0 resonance (ϕ), 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 contributes to the gluon fusion loop are set to SM values.

ii) A search for the MSSM Higgs sector in several benchmark scenarios. The benchmark scenarios were proposed in References [81–83] and described fully in Reference [84]. The M_h^{125} and $M_{h,\,\mathrm{EFT}}^{125}$ 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 as detailed in Section 1.4.1.

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

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

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 a Type II 2HDM, such as the MSSM. In these models, $\tan\beta$ enhances couplings of additional Higgs bosons to down-type quarks and charged 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 b quark. Examples of the Feynman diagrams for these processes are shown in Figure 4.1, where ϕ can represent any of the additional neutral Higgs bosons (h, H or A) in these diagrams.

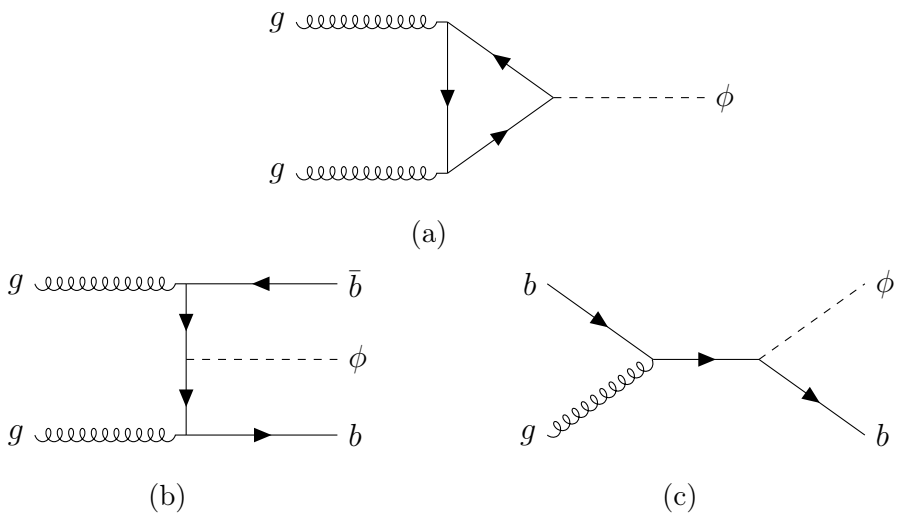


Figure 4.1: Diagram (a) shows the production of neutral Higgs bosons from gluon fusion. The dominant loop contributions to these diagrams are from t-only, b-only and tb-interference. Diagrams (b) and (c) show production in association with b quarks.

With the $\tan\beta$ enhancement, the decays of additional Higgs bosons to τ leptons and b quarks are most dominant. τ leptons are identified with higher purity than b

quarks at the CMS detector. It is also easier to separate $\tau^+\tau^-$ from the large QCD multijet background produced from the high-energy proton-proton collisions. This was confirmed with the 2016 dataset and although no deviations were observed, the strongest limits on the MSSM phase space were placed by the $\tau^+\tau^-$ final states [1,85].

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 next-to-leading-order (NLO) precision using the 2HDM implementation of POWHEG 2.0 [86–89]. The kinematic properties are highly dependent on the contributions to the loop, and these are dependent on the specific signal model. To account for the t quark only, b quark only, and tb-interference loop contributions at the NLO plus parton shower prediction, weights are calculated from the p_T spectra to split these contributions up. 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,

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

where Q_i are resummation scales that depend on the mass of the additional Higgs boson, and σ^i_j and Y^i_j are the determined cross-sections and Yukawa couplings of the relevant theory j and contribution i. Further contributions from any supersymmetric partners were checked and accounted 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 $m_A = 500$ GeV whilst varying $\tan\beta$, is shown in Figure 4.2. The distributions peak at a higher p_T for the t quark loop, therefore at smaller $\tan\beta$, where the t quark contribution is dominant, an additional Higgs boson would be more boosted.

Production in association with b quarks is simulated at NLO precision using the corresponding POWHEG 2.0 implementation in the four flavour scheme (FS) [86–89].

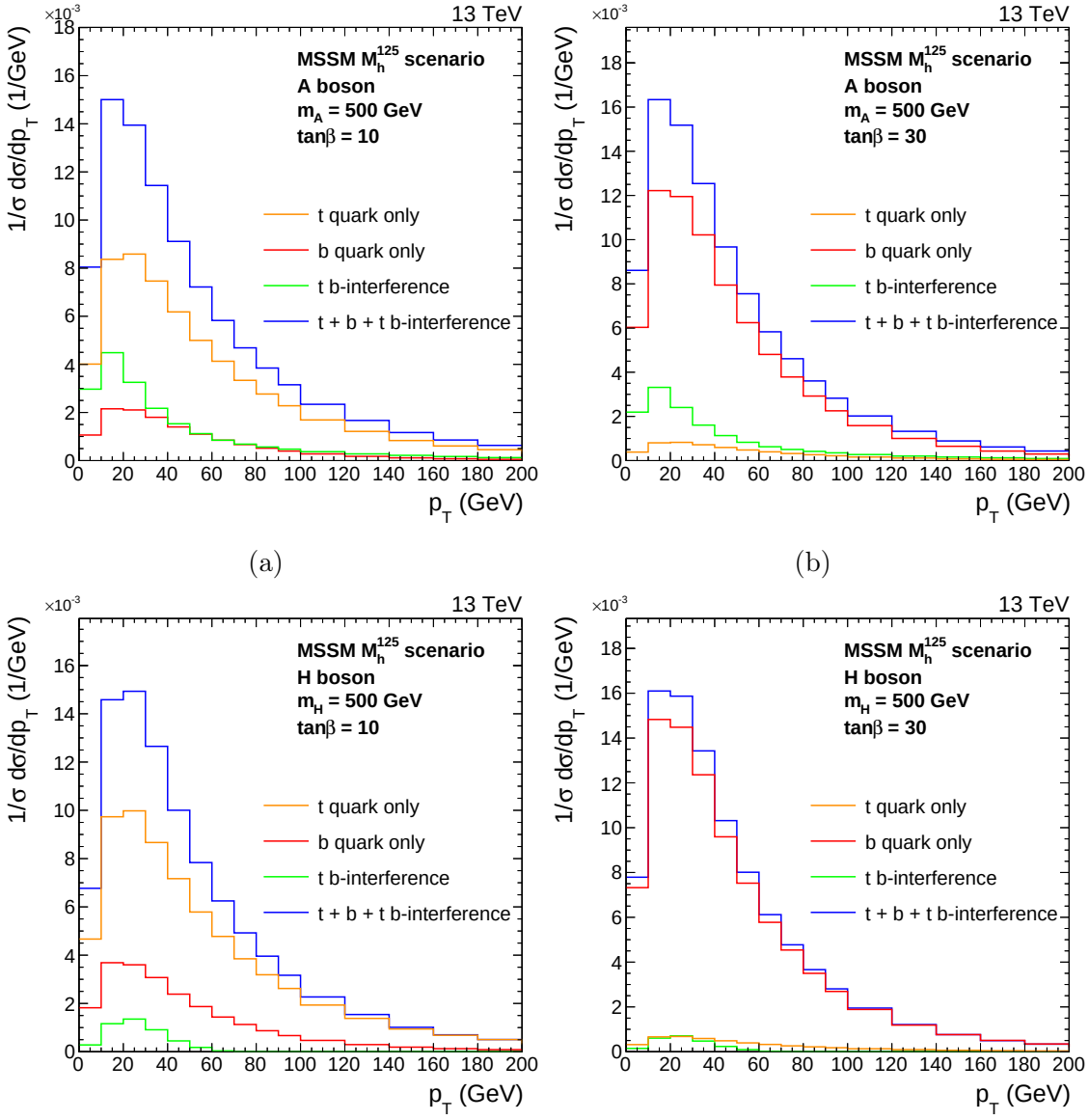


Figure 4.2: p_T density distributions of the A (top) and H (bottom) boson, with contributions to the gluon fusion loop displayed individually and summed. These are shown for tan β values of 10 (left) and 30 (right) where $m_A = 500$ GeV in the MSSM M_h^{125} scenario.

All additional Higgs boson signal generation is performed using the parton distribution function NNPDF3.1 [90, 91]. τ lepton decay, parton showering and hadronisation are all modelled with the PYTHIA event generator where the PU profile is matched to data [92, 93]. All events generated are passed through a GEANT4-based [94] simulation of the CMS detector and reconstructed in the same way as data.

The model-independent and model-dependent additional neutral Higgs boson searches explained in this chapter, utilise the same signal templates generated. The differences lie in the scaling of the gluon fusion cross-section and loop contributions (set to the Yukawa couplings for the model-independent interpretation), and the b associated production cross-section, as well as taking into account the branching fractions in the model-dependent interpretation. The model-dependent search for the MSSM also looks to find differences between the observed SM Higgs boson and the predicted MSSM SM-like Higgs boson. In each MSSM benchmark scenario, an uncertainty of ± 3 GeV is given on the prediction for the SM Higgs boson mass. This uncertainty is to reflect the contribution of any unknown higher-order corrections. The value of the mass is allowed to vary within this window, however, the Yukawa couplings are rescaled to that of the observed mass.

4.1.2 Vector leptoquarks

The best fit to the B anomalies in the available phase space for vector leptoquarks yielded a large b quark and τ lepton coupling to the U(1) particle. The possible production modes of a $\tau^+\tau^-$ final state are shown in Figure 4.3.

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 chapter. A search for single and pair production at the CMS experiment was performed and no statistically significant deviation was observed [95]. Single production places the loosest constraints on the available phase space, out of the processes mentioned. Pair production puts a lower limit on the leptoquark mass, as the process is approximately independent of g_U. However, this limit is heavily dependent on the value chosen for the strong coupling.

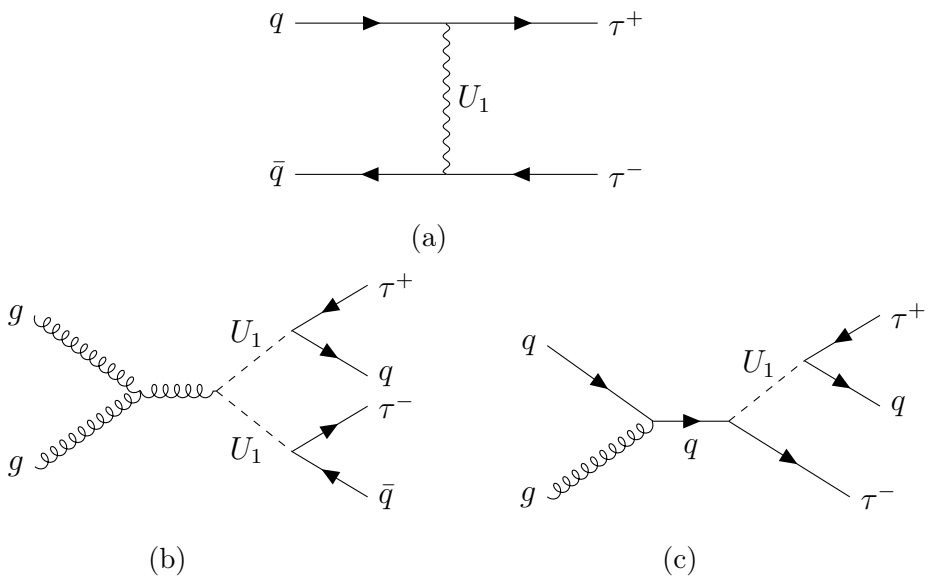


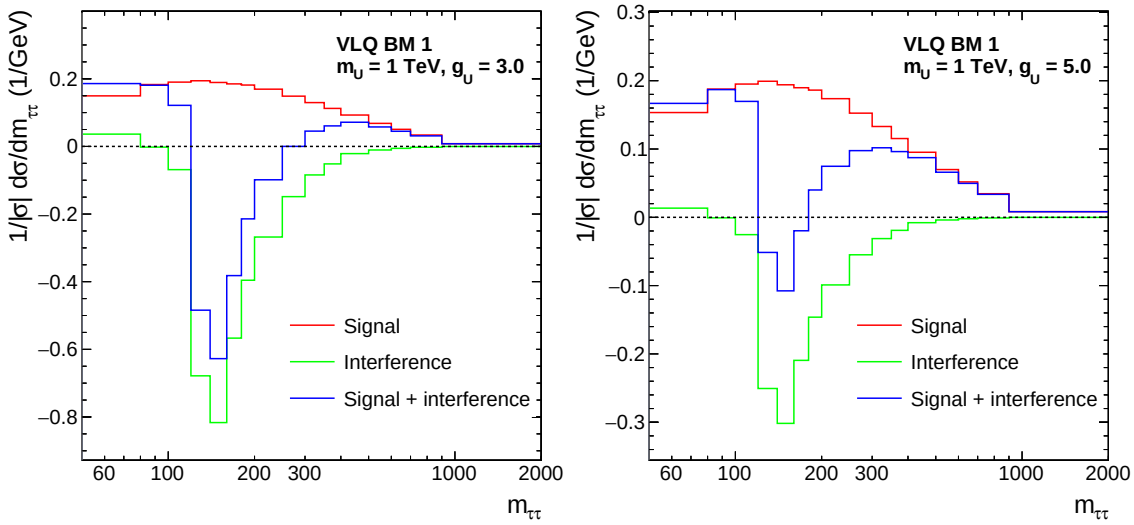
Figure 4.3: Feynman diagrams showing the contribution from U(1) vector leptoquarks to the final state with a pair of oppositely charged τ leptons. q represents any SM quark. The t-channel (a), pair (b) and single (c) productions of a vector leptoquark are shown.

The t-channel process contains two vertices with a U(1) vector leptoquark, a quark and a τ lepton, and hence the cross-section will scale with g_U^4 (g_U^2 for each vertex). 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 leading-order (LO) process, the production of initial state b quarks 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 a t-channel exchange is simulated in the five FS at LO precision using the MadGraph5_aMC@NLO v2.6.5 event generator [96]. Events are generated with one or fewer outgoing partons from the matrix element and the MLM prescription [97] is then used for matching, with a scale set to 40 GeV. Negligible dependence of the decay width is observed. For simulation, this is chosen to approx-

imately match the value predicted by the B anomaly fit [39]. Samples with a mass between 1 and 5 TeV are generated.

The interference between the signal and $Z/\gamma^* \to \tau\tau$ process was checked and a large destructive effect was observed, with magnitude dependent on g_U. To account for this, separate samples are produced for this interference, using the same generators as the t-channel exchange. The interference samples are generated with extra statistics at large di-τ mass values, to have a sufficient number of events in the regions of interest for the signal. The cross-section of these interference samples scales with g_U^2. Examples of the generator level di-τ mass distributions are shown in Figure 4.4.



(a) (b)

Figure 4.4: The generator level $m_{\tau\tau}$ density distributions of the t-channel vector leptoquark signal and the interference with $Z/\gamma^* \to \tau\tau$. 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. This negative signal yield would diminish the Z/γ^* background. Due to the difference in the 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 τ leptons and their branching fractions, where the τ decay is grouped into three categories; e, μ and τ_h as defined in Section 3, are shown in Table 4.1. For this search, the four largest branching fraction channels are used: $\tau_h\tau_h$, $e\tau_h$, $\mu\tau_h$ and $e\mu$. This accounts for approximately 94% of di-τ events. The two same lepton channels are neglected due to the small branching ratio and the dominating $Z \to ee$ and $Z \to \mu\mu$ 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 τ leptons.

4.2.1 Trigger requirements

In these four final state pairs, several different online trigger requirements are needed. In the $\tau_h\tau_h$ channel, two possible triggers are available: the double-τ_h and single-τ_h triggers. The single-τ_h trigger has a high p_T threshold at 120 (180) GeV for events recorded in 2016 (2017-2018), whilst the double-τ_h trigger has a p_T threshold at 40 GeV. Therefore, the double-τ_h trigger is used unaccompanied when the τ_h has p_T below the single-τ_h threshold and the union of single-τ_h and double-τ_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/μ, single-τ_h and the e/μ-τ_h cross-trigger. The p_T thresholds of the light lepton leg of these triggers are shown in Table 4.2. The cross-trigger is used for events where the light lepton has p_T between the thresholds for the cross-trigger and single-e/μ trigger. The light lepton used in the cross-trigger is required to be in the central barrel of the detector within $|\eta| < 2.1$, and the τ_h triggered on is required to have $p_T > 30$ GeV. Above these light lepton cross-trigger p_T thresholds, the single-e/μ

trigger is used. At τ_h p_T above the single-τ_h thresholds, the single-τ_h trigger is used in combination with the single-e/μ trigger.

Year/ Trigger	e-τ_h cross-trigger	single-e	μ-τ_h cross-trigger	single-μ
2016	23	26	20	23
2017	25	28	20	25
2018	25	33	21	25

Table 4.2: Lower trigger p_T thresholds in GeV on light leptons 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-μ and the e-μ cross-trigger. However, only the cross-trigger is used in this analysis, due to the larger efficiencies of correctly selecting light leptons. The electron and muon are required to have $p_T > 15$ GeV and $|\eta| < 2.4$.

4.2.2 Offline requirements

All offline selections stated are in addition to the object selection discussed in Chapter 3. In this analysis, τ_h candidates are required to pass the `DeepTau Medium` $D_{\text{jet}}^{\text{WP}}$, as described in Section 3.8. The D_e^{WP} and D_μ^{WP} choices are dependent on the decay channel. The `VVLoose`, `Tight`, `VVLoose` D_e^{WP} and the `VLoose`, `VLoose` and `Tight` D_μ^{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 a τ_h candidate from the $Z \to \ell\ell$ process. The light lepton isolation requirement is $I_{\text{rel}} < 0.15$ for both electrons and muons, except for in the $e\mu$ channel where the muon is required to have $I_{\text{rel}} < 0.2$.

The selected τ lepton decay candidates are required to have opposite charges and to be separated by more than $\Delta R > 0.5$ in all channels except $e\mu$ where $\Delta R > 0.3$ is required. In events where the number of an object is greater than the required number of objects for the decay channel, the objects are sorted and the leading objects are chosen. This is done by the maximum $D_{\text{jet}}^{\text{score}}$ if a τ_h candidate, or minimum I_{rel} if a light lepton candidate. 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 vetoed, help to suppress the

$Z \to \ell\ell$ background process further, as well as keeping the channels orthogonal.

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

$$m_{\mathrm{T}}(\vec{p}_T^{\,i}, \vec{p}_T^{\,j}) = \sqrt{2p_T^i p_T^j (1 - \cos\Delta\phi)}, \tag{4.2}$$

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

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

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

where $\hat{\zeta}$ is the bisectional direction between the electron and the muon in the transverse plane. A diagram of the inputs is shown in Figure 4.5.

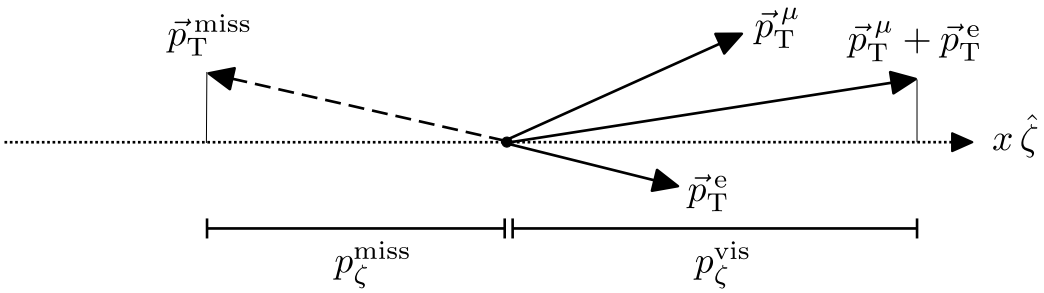


Figure 4.5: Diagram of the inputs to the D_ζ variable [2].

The linear combination is optimised for genuine di-τ events to peak around $D_\zeta = 0$ GeV. It is motivated by the expectation that in di-τ decays from a resonance, the visible and missing (from τ 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_ζ. Therefore, only events with $D_\zeta >$ -35 GeV are considered for signal events. No b tag events failing this cut are vetoed, and events with a b tag also failing this cut are used for a $t\bar{t}$ control region.

4.3 Search optimisation

The optimisation of the signal extraction depends on which of the three scenarios, set out at the beginning of this section, is being 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-τ mass is greater or less than 250 GeV. The presence of more than one neutrino in the event means that it is not possible to fully reconstruct the di-τ mass directly. Instead, the likelihood-based `SVFit` algorithm is used [98]. Finally, the search for vector leptoquarks (iii) uses only the high-mass optimisation procedure. The procedures are discussed in detail below.

4.3.1 High-mass optimisation

The high-mass optimisation procedure follows what was done in Reference [1]. Each event is initially split into categories depending on whether it has 0 or $\geq$1 b tagged jets. Firstly, this helps target the additional Higgs boson production modes gluon fusion and b-associated production respectively. In particular, bbϕ has final state b quarks that if tagged can significantly aid the sensitivity to this signal. Secondly, the production of the dominant initial state b quarks for a t-channel vector leptoquark signal can lead to additional b jets in the final state. The reduced backgrounds in b-tagged events allow 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}: $m_{\mathrm{T}}(\vec{p}_T^{\,e/\mu}, \vec{p}_T^{\,\mathrm{miss}}) < 40$ GeV;
- `Loose-`m_{T}: $40 \leq m_{\mathrm{T}}(\vec{p}_T^{\,e/\mu}, \vec{p}_T^{\,\mathrm{miss}}) < 70$ 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.

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

- `Low-`D_ζ: $-35 \leq D_\zeta < -10$ GeV;
- `Medium-`D_ζ: $-10 \leq D_\zeta < 30$ GeV;
- `High-`D_ζ: $D_\zeta \geq 30$ GeV.

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

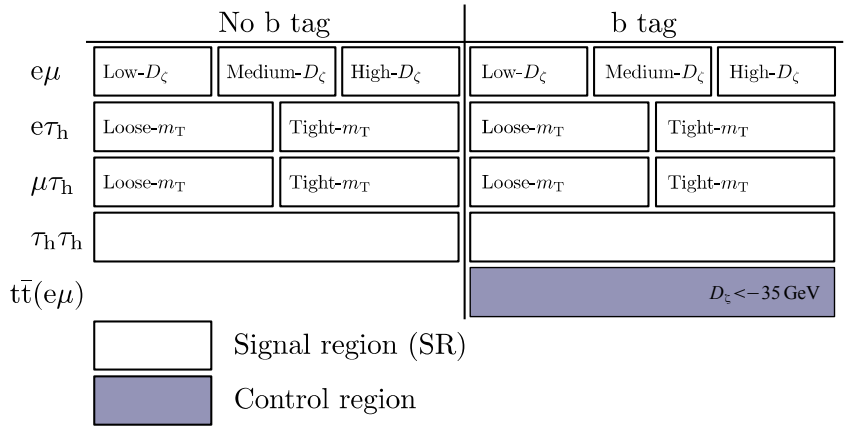


Figure 4.6: Overview of the categories used for the extraction of the signal in the high-mass optimisation procedure [2].

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

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

where τ_1 and τ_2 refer to the visible products of the two τ 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-τ signatures from backgrounds, where a di-τ mass does not represent the mass

of a resonance. 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.

4.3.2 Low-mass optimisation

The low-mass optimisation procedure, loosely follows the high-mass procedure with a few key differences. Categories that are only sensitive to high-mass signals are dropped. This includes the `Low-`D_ζ and `Loose-`m_T categories. Each no b tag subcategory is further divided into four bins of reconstructed di-τ visible p_T with bin edges: 0, 50, 100, 200 and ∞. 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-τ mass is used. This helps to separate signal events from the Z boson peak in this region.

4.3.3 Standard Model Higgs boson optimisation

Finally, the SM Higgs optimisation procedure is taken from the CMS SM $H \to \tau\tau$ analysis and is detailed in Reference [99]. This was previously used for simplified template cross-section measurements. This uses a neural network (NN)-based 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.

4.4 Background modelling overview

The relevant backgrounds for this analysis include Z/γ^*, $t\bar{t}$, W+jets, QCD, di-boson, single-top, and electroweak W and Z bosons production. These are split into five categories:

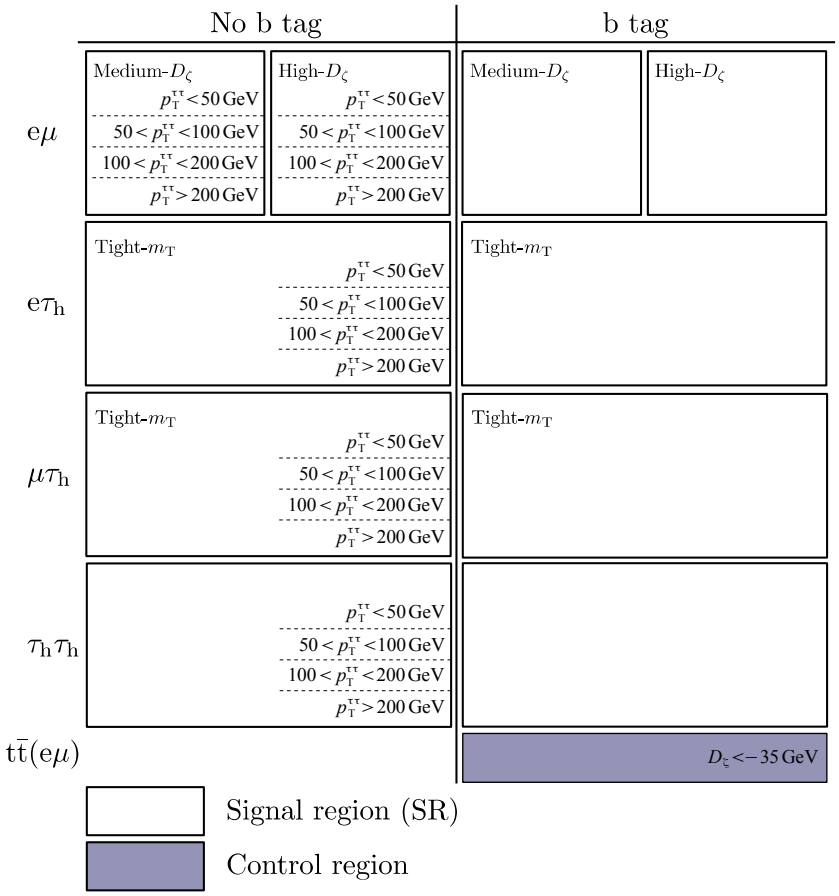


Figure 4.7: Overview of the categories used for the extraction of the signal in the low mass optimisation procedure [2].

i) Events containing only genuine τ leptons.

ii) Events with a jet misidentified as a τ_h (jet $\to \tau_h$) in the $e\tau_h$, $\mu\tau_h$ or $\tau_h\tau_h$ channels.

iii) Events with jets misidentified as both light leptons (jet $\to \ell$) in the $e\mu$ channel.

iv) Events from $t\bar{t}$ with a prompt light lepton (e or μ not from a τ decay) and the other object (if there are not two prompt light lepton) is from a genuine τ lepton.

v) Other events. This is a small contribution and so it is grouped.

- Non $t\bar{t}$ events satisfying (iv).
- Events with a light lepton misidentified as a τ_h and no jet $\to \tau_h$ candidate.
- Events with a jet misidentified as a light lepton and the other object is from genuine τ 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) consist of largely $Z/\gamma^* \to \tau\tau$ events but there are also smaller contributions from other processes. This background is modelled by a data-simulation hybrid method called the "embedding" method [100] 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 jet $\to \tau_h$ misidentification. This is modelled from data by the "fake factor" (F_F) method [101, 102] 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 ≈98% of all expected background events in the $\tau_h\tau_h$ channel, ≈90% in $e\tau_h$ and $\mu\tau_h$ channels and ≈50% in the $e\mu$ channel. The final groups, (iv) and (v), are modelled with MC simulations. The $t\bar{t}$ process from (iv) is separated because of its large contribution to the phase space where a b jet is required.

In 2016 (2017–2018), the W + jets and $Z \to \ell\ell$ processes are simulated using the MadGraph5_aMC@NLO 2.2.2 (2.4.2) event generator at LO [96]. 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 di-boson production, MadGraph5_aMC@NLO is used at NLO precision [96], and the FxFx [97] (MLM [103]) prescription is used to match the NLO (LO) matrix element calculation with the parton shower model. Samples for $t\bar{t}$ [104] and (t-channel) single top quark production [105] are generated at NLO precision using PowHeg 2.0 [86–89], and for single top quark production in association with a W boson (tW channel) [106], PowHeg version 1.0 at NLO precision is used. When compared with data, W + jets, $Z \to \ell\ell$, $t\bar{t}$, and single top quark events in the tW channel are normalised to their cross-sections at next-to-next-to-leading-order (NNLO) precision [107–109], while single top quark and di-boson events are normalised to their cross-sections at NLO precision or higher [109–111].

4.5 QCD estimation in the $e\mu$ channel

The QCD model in the $e\mu$ channel, which attempts to model events where two jets are misidentified as an electron-muon pair, is taken from data where the electron and muon have the same sign, 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 ($0.2 < I_{\text{rel}} < 0.5$). F_T is initially parametrised by the ΔR 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 a b jet, an additional correction is needed. This is determined to be ≈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 election is required to have $0.15 < I_{\text{rel}} < 0.5$. 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, $pp \to b\bar{b}$ ($c\bar{c}$). 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-antiquark 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 overpredicts 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-τ lepton pairs is modelled via the embedding method [100]. 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 genuine di-τ background is dominated by the $Z \to \tau\tau$ process, but 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 $Z \to \mu\mu$ region but still be loose enough to catch events from other processes, so as not to introduce a bias on the Z boson mass. Events are required to pass the double-μ trigger with minimum requirements on the invariant mass of the two muons ($m_{\mu\mu}$) and the p_T of the leading and trailing muons. 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 η selections and originate from a global muon track, as defined in Section 3. The muon pair are required to have an 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 $Z/\gamma^* \to \mu\mu$ events with smaller contributions from $Z/\gamma^* \to \tau\tau$ ($\tau \to \mu$), 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 η and p_T using the "tag-and-probe" method, as described in Reference [112].

Next, all energy deposits in the detector from the selected muons are removed. This

involves removing the hits on global muon tracks in the tracker, hits in the muon systems and clusters in the calorimeters that intercept the muon trajectory. Once completed, the selected muons and their kinematic properties are replaced with simulated τ leptons. To account for the difference in mass between the muon and τ, the muons are boosted into the centre-of-mass frame of the di-muon system and then this four-vector is taken for the τ but boosted back into the laboratory frame. The event simulation is performed from the PV. The τ lepton decay is then simulated with PYTHIA [92,93] and separate samples are produced for different $\tau\tau$ decay channels. Only the decay of the τ leptons is 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.

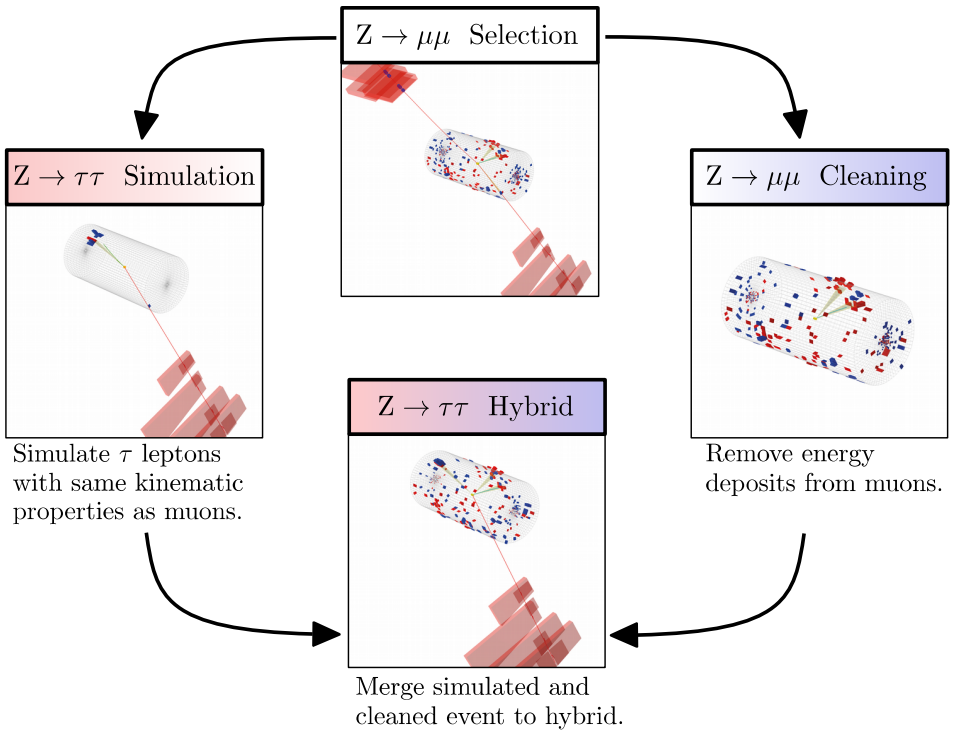


Figure 4.8: Schematic of the embedding method to model genuine di-τ backgrounds from di-muon events in data [100].

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

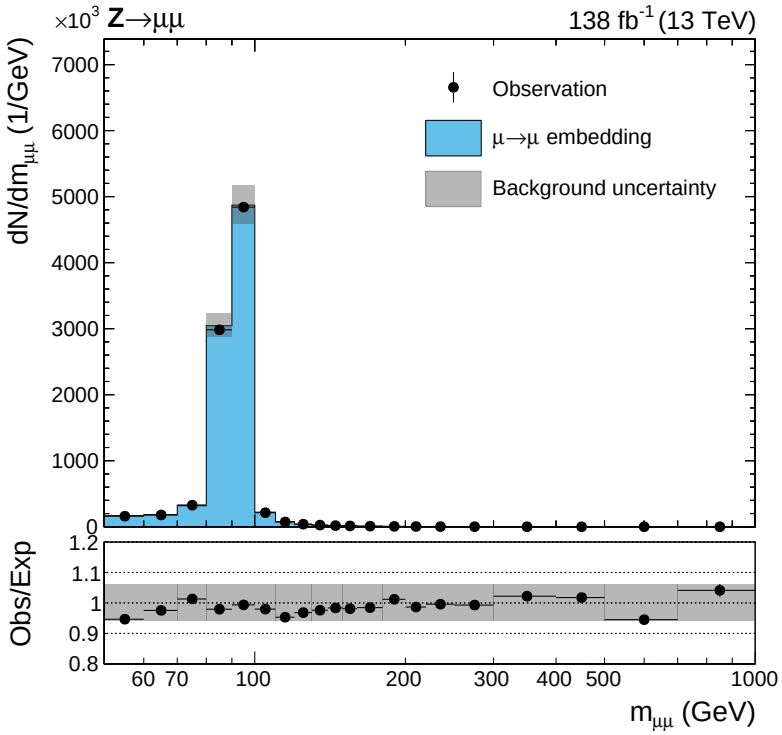


Figure 4.9: Closure plot showing the di-muon mass on the dedicated embedding validation samples.

4.7 Fake factor method

Backgrounds in which a jet is misidentified as a τ_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 τ_h necessitates the production of high statistics MC samples at a significant computational expense. These shortcomings motivate the use of data-driven estimations for these processes. One such procedure is the F_F method [101, 102].

The F_F method utilises regions in the data to model the jet $\rightarrow \tau_h$ background. Firstly, the determination regions are defined, which are jet $\rightarrow \tau_h$ enriched regions orthogonal to the signal region. They are then purified by subtracting off any non jet $\rightarrow \tau_h$ events with MC. It is used to calculate a F_F by taking the ratio of the number of jet $\rightarrow \tau_h$ events that pass the nominal τ_h identification requirement ($N(\texttt{Nominal})$), to the number of jet $\rightarrow \tau_h$ events that fail the nominal τ_h identification but pass a looser alternative τ_h identification requirement ($N(\texttt{Alternative}$ and not $\texttt{Nominal})$), as shown in Equation 4.5.

$$F_F = \frac{N(\texttt{Nominal})}{N(\texttt{Alternative and not Nominal})}. \tag{4.5}$$

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 and combine F_F measured from different processes. Finally, the F_F are applied to the application region. This is defined as the signal region but with the criteria that the jet $\rightarrow \tau_h$ events fail the nominal τ_h identification but pass the looser alternative τ_h identification requirement. This now models the background from jet $\rightarrow \tau_h$ events in the signal region.

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 τ_h identification used is the `Medium` $D_{\text{jet}}^{\text{WP}}$ and the alternative τ_h identification used is the `VVVLoose` $D_{\text{jet}}^{\text{WP}}$.

4.7.1 Determination regions

The F_F are measured separately in each year of data taking period (2016, 2017, 2018), in each channel containing a τ_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 jets misidentified as τ_h candidates, and so the F_F is chosen to be calculated from leading p_T τ_h candidate only. Section 4.7.4 discusses how a single jet $\rightarrow \tau_h$ events in the $\tau_h\tau_h$ channel are modelled.

Each separate measurement region is split into three sideband regions based on 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 the F_F. Region B is an alternative region used to measure and fit the F_F to account for the difference between A and C. These alternative F_F are applied to the fail τ_h identification region in D and corrections are calculated by comparing it to the pass region in D. The total F_F per measurement region is calculated as the product of the F_F derived in region A and the correction

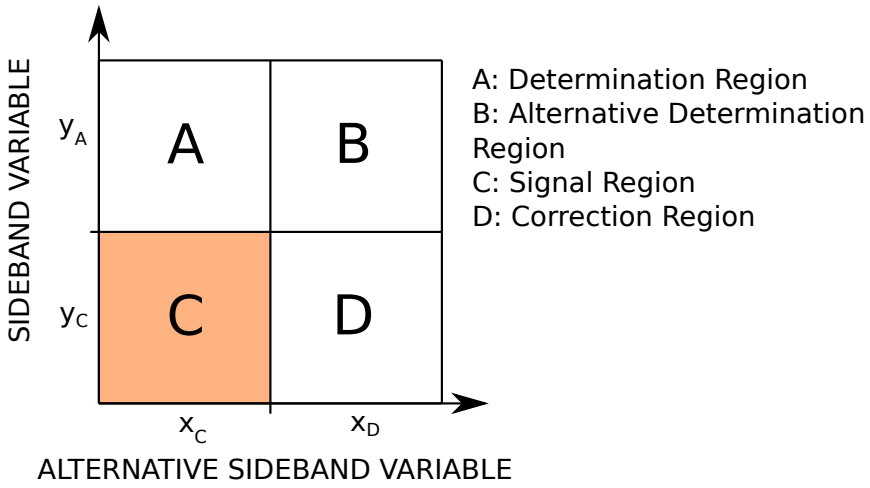


Figure 4.10: Schematic of the regions used for fake factor derivation.

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 τ_h candidates are required to have the opposite sign.
   y_A: The τ_h candidates are required to have the same sign.
   x_C: The subleading τ_h passes the `Medium` $D_{\text{jet}}^{\text{WP}}$.
   x_D: The subleading τ_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/μ and τ_h candidates are required to have the opposite sign.
   y_A: The e/μ and τ_h candidates are required to have the same sign and the e/μ to have $I_{\text{rel}} > 0.05$.
   x_C: The e/μ candidate is required to have $I_{\text{rel}} < 0.15$.
   x_D: The e/μ 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/μ and the MET < 70 GeV.
   y_A: The m_T between the e/μ 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 `Determination Region` that is sufficiently pure in $t\bar{t}$ events to make a reasonable measurement of F_F from data. Therefore, $t\bar{t}$ F_F are derived from MC. A comparison of the W + jets F_F measured in data and MC shows only $\sim$ 10–20% 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 measured in MC is small compared to the uncertainties on the F_F, discussed in Section 4.9. In many cases, selections on the regions B and D are applied which are tighter than required, to purify the regions as much as possible, while maintaining enough statistics to perform the fit.

4.7.2 Parametrisation

The raw F_F take into account dependencies on N_{jets} via an analysis-category tailed variable $N_{\text{pre b jets}}$, the p_T of the τ_h candidate ($p_T^{\tau_h}$) and the p_T of the jet that seeds the HPS algorithm (p_T^{jet}). $N_{\text{pre b jets}}$ is the number of jets in the event with $|\eta| < 2.4$ and $p_T > 20$ GeV, which is the η and p_T selection for a b-tagged jet. It is defined to map the dependence of F_F on the number of jets and describe the categorising variable $N_{\text{b jets}}$ well. Although not local to the τ_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 $p_T > 20$ GeV. These are the same η 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^{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.

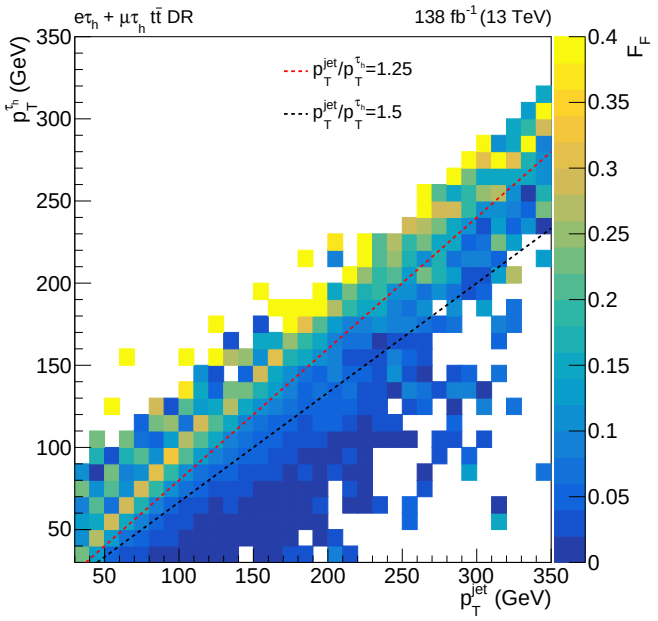


Figure 4.11: A 2D heat map of the F_F determined from $t\bar{t}$ MC for the full Run 2 dataset in the combined $e\tau_h$ and $\mu\tau_h$ channels. This is shown with respect to the τ_h p_T and the p_T of the HPS seeded jet for the τ_h. The ratio of jet to τ_h p_T categorisation used is shown split by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

It is motivated by the observation that the F_F are largest when the p_T^{jet} and $p_T^{\tau_h}$ are closest. The physical motivation for this is when they are close, the τ_h candidate is likely to be isolated from any other hadronic activity and so more likely to be identified as a τ. However, when p_T^{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^{jet} is less than $p_T^{\tau_h}$, charged pions are likely not close enough to the PV to be clustered into the jet and so the event is less likely to be classified as a τ_h. This leads to the F_F 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 function and a constant in the low-p_T region. The F_F 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 on the algorithm shown in Figure 4.12 and the fit is used below the minimum bin.

The fits are flattened at $p_T^{\tau_h}$ values where there is no significant downward shift or at the final bin to avoid extrapolating past the fitted range. F_F fits with respect to $p_T^{\tau_h}$ are shown in Figures 4.13-4.14. The F_F are highest in the lowest $p_T^{\text{jet}}/p_T^{\tau_h}$ bin

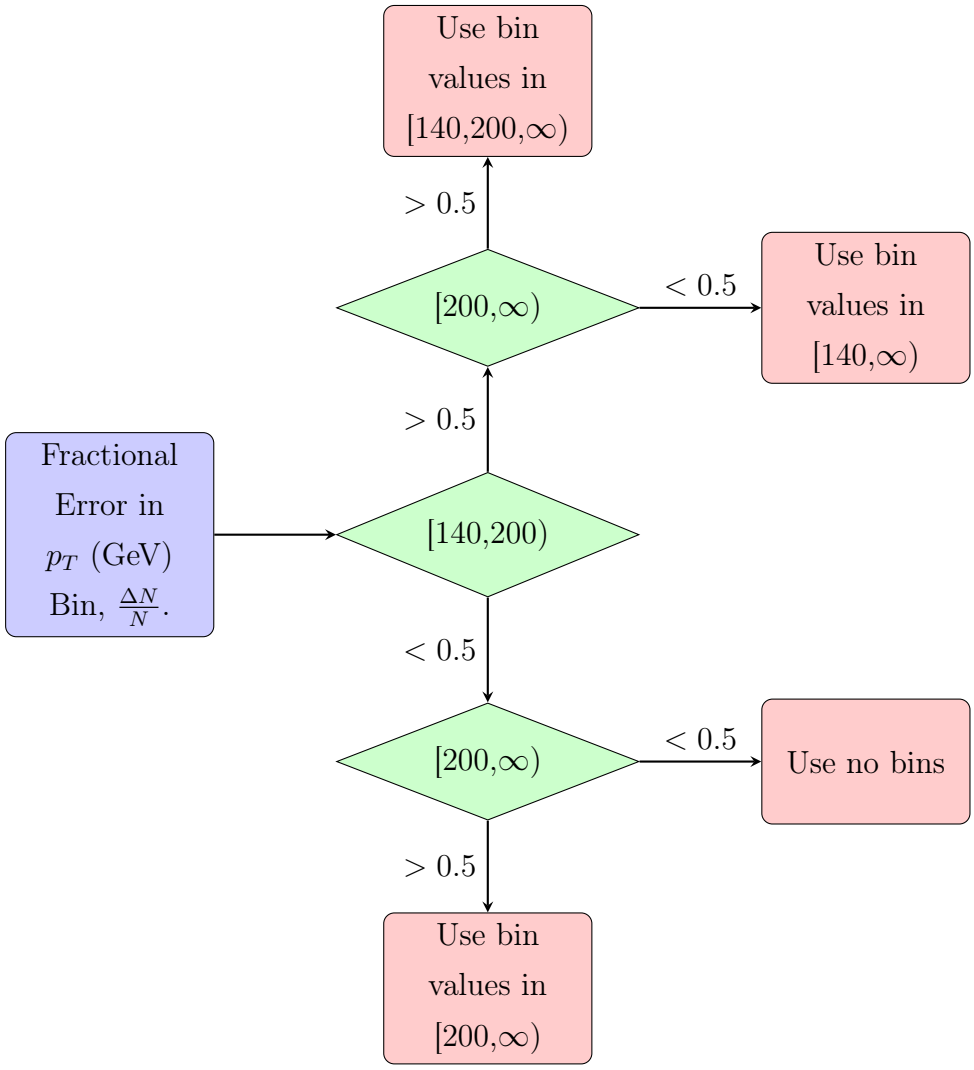
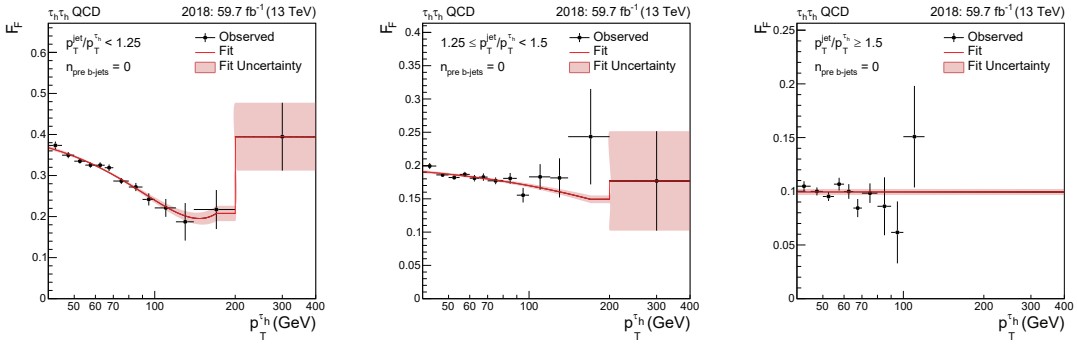


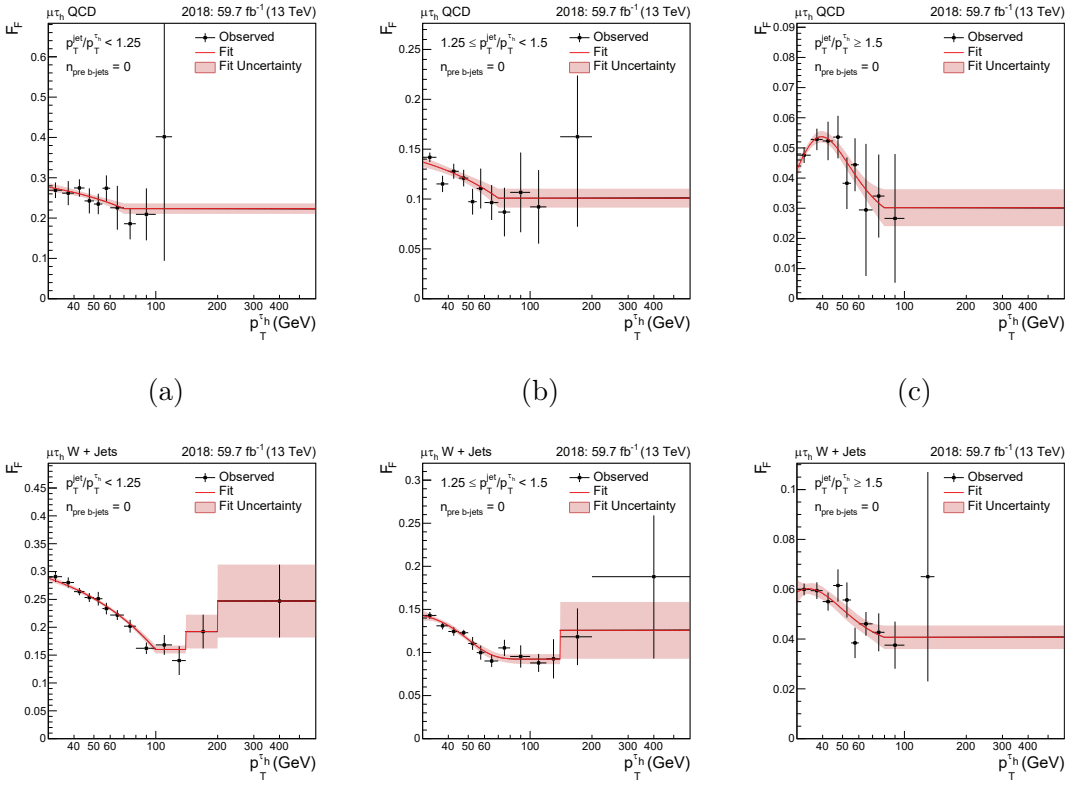
Figure 4.12: Flow chart of the algorithm used to determine where binned values are taken instead of the fit. The blue box represents the input, the green diamonds represent the decisions and the red boxes represent the outputs.

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



(a) (b) (c)

Figure 4.13: F_F 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 τ_h p_T categories are shown.



(a) (b) (c)

(d) (e) (f)

Figure 4.14: F_F 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 τ_h p_T categories are shown for each process.

4.7.3 Corrections

In the $\tau_h\tau_h$ channel, the measured F_F are then corrected to account for non-closures in other variables in the `Determination Region`. The only significant non-closures are observed for E_T^{miss} related variables and are largest for events with $N_{\text{pre b jets}} = 0$. Closure corrections are performed for the variable ΔR between the τ_h candidates in bins of $N_{\text{b jets}}$. In the $\mu\tau_h$ and $e\tau_h$ channels, the measured QCD and W + jets F_F are corrected for non-closures observed in the E_T^{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 MET arising from mismeasurement of the energies of particles in a jet. If a jet's energy is mismeasured, this is also propagated to the reconstruction of the MET and the τ_h candidate and there will be a specific alignment between these objects if no neutrinos are present in the event. A mismeasurement of the jet energy can alter the τ_h isolation and so it can also affect the identification scores, and this shift can be accounted for with some measure of the fake MET. A diagram of this effect is shown in Figure 4.15.

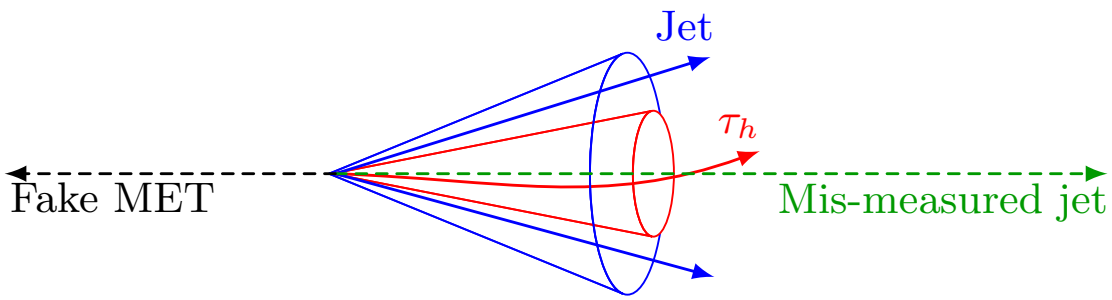


Figure 4.15: Diagram showing how fake MET arises from mis-modelling jet energies and how it can align with the τ_h candidate identified.

To correct for this effect, the QCD F_F are corrected as a function of C_{QCD}, where C_{QCD} is defined as,

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

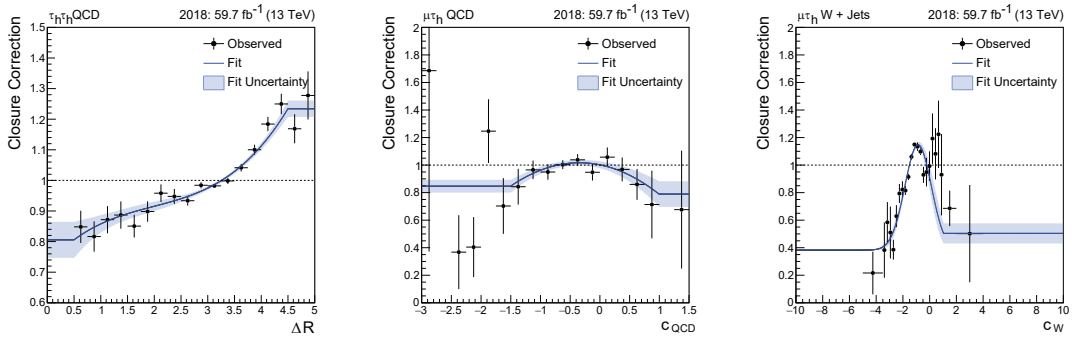
where $\Delta\phi(\vec{p}_T^{\text{miss}}, \vec{p}_T^{\tau_h})$ is the separation in the azimuthal angle between the missing transverse momentum vector $\vec{p}_T^{\text{miss}}$ and $\vec{p}_T^{\tau_h}$. The numerator quantifies the missing transverse momentum in the direction of the τ_h candidate. Once divided by the τ_h p_T, C_{QCD} is a measure of the fraction of missing to visible τ_h transverse momentum

aligned with the τ_h. For W + 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 MET 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 produced at rest in the transverse direction). The equation then becomes,

$$C_{\mathrm{W}} = \frac{(E_T^{\mathrm{miss}} + p_T^{e/\mu}) \cos\Delta\phi(\vec{p}_T^{\,\mathrm{miss}} + \vec{p}_T^{\,e/\mu}, \vec{p}_T^{\,\tau_h})}{p_T^{\tau_h}}. \tag{4.7}$$

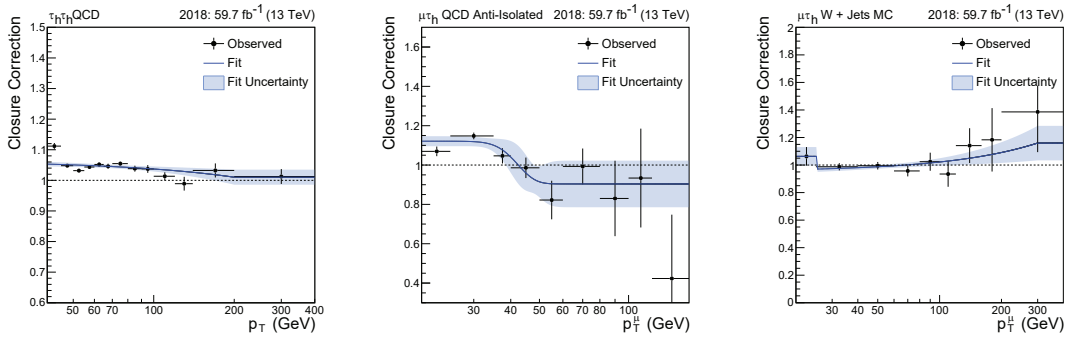
When either correction variable is non-zero, a larger quantity of fake MET is expected in the event. In these regions, a large correction is needed due to the mismeasured jet energy spectrum shifting the τ_h candidate isolation and so shifting the τ identification scores. Examples of these closure corrections are shown in Figure 4.16

After the `Determination Region` is modelled well for all variables of interest, extrapolation corrections from the F_F derived in B applied to region D are calculated. In the $\tau_h\tau_h$ the correction is parametrised by the p_T of the leading τ_h candidate, in the $e\tau_h$ and $\mu\tau_h$ channels it is parametrised 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.



(a) (b) (c)

Figure 4.16: `Determination Region` closure correction fits with 2018 data. (a) is the correction parametrised by ΔR in events with $N_{\mathrm{b\ jets}} = 0$ 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 Equations 4.6 and 4.7 for QCD and W + jets processes respectively.



(a) (b) (c)

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 τ leptons the parametrised by leading τ_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 τ leptons and high m_T to low m_T both parametrised 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, i, are combined into an overall factor F_F using,

$$F_F = \sum_i f_i \cdot F_F^i, \tag{4.8}$$

where the factor f_i is defined as,

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

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 τ pairs, and examples are shown in Figure 4.18. It is observed that W + jets is the dominating process in this region, however, there are effects from QCD at low m_T and from $t\bar{t}$ in the b-tagged categories. These fractions are then multiplied by the relevant corrected F_F and applied to the fail τ_h identification region in C, where this region is purified by subtracting any non jet $\rightarrow \tau_h$ events with MC.

The $\tau_h\tau_h$ channel has two τ_h candidates that a jet can be misidentified as. For this analysis, the F_F are only applied to the leading τ_h candidate failing the τ_h

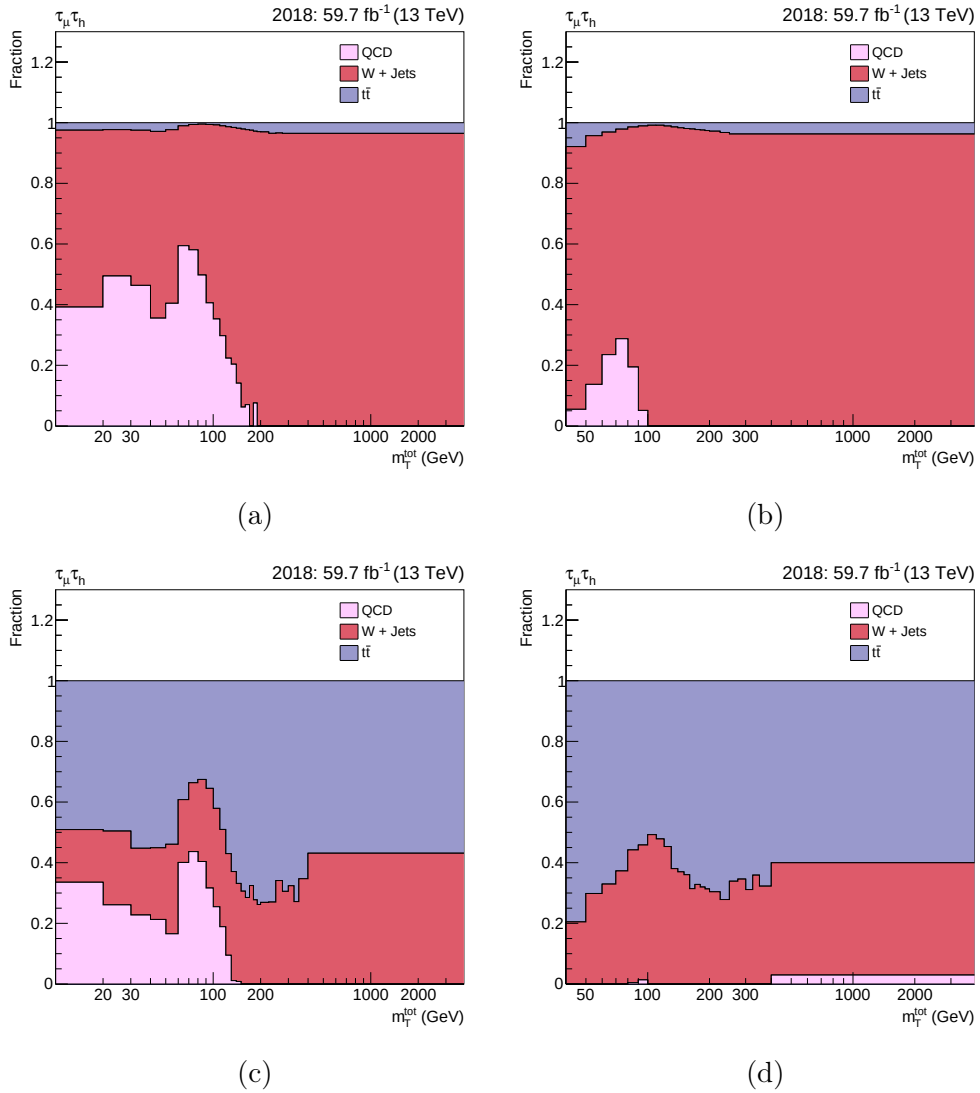


Figure 4.18: The expected `Application Region` fractions of the processes in the $\mu\tau_h$ channel. (a) and (b) show the no b tag `Tight`-m_T and `Loose`-m_T categories and (c) and (d) show the b tag `Tight`-m_T and `Loose`-m_T categories respectively.

identification in C. This models all events where the leading τ_h candidate is a jet $\rightarrow$ τ_h. However, this leaves a small fraction of events, where the leading candidate is a genuine τ 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 stated in this section apply both to simulated and embedding samples, as the τ decay is simulated, however, they are derived separately. For electrons and muons, corrections are applied to triggers, tracking efficiencies, and identification and isolation requirements. Using the tag-and-probe method, they are obtained in bins of p_T and η 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 energy scale of the electron is adjusted to the scale measured in data using the Z boson mass peak in $Z \rightarrow ee$ events. This effect is negligible in muons.

Similarly, corrections are derived for the efficiencies of triggering and identification of τ_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 τ_h decay modes. The identification efficiency corrections are derived as a function of the p_T of the τ_h candidate. Corrections to the energy scale of the τ_h candidates and of electrons misidentified as τ_h candidates are obtained from likelihood scans of discriminating observables, such as the reconstructed τ_h candidate mass. For muons misidentified as τ_h candidates, the energy scale correction is negligible.

The magnitude and resolution of the MET need correcting in embedding events, to take into account the incomplete removal of energy deposits from muons replaced by simulated τ decays during the embedding procedure. These corrections are derived by comparing p_T^{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 $|\eta| > 2.5$ during the 2016 and 2017 data taking [113], is needed to be corrected. This effect is named "prefiring". This resulted in a loss of efficiency for events containing an electron or jet

with p_T larger than approximately 50 or 100 GeV, in 2016 and 2017 respectively. Corresponding corrections are derived from data and applied to the simulation.

Corrections to the energy of jets are calculated in bins of the jet p_T and η. 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. An MC to data correction for a b jet passing the selection criteria is also determined. This correction can alter 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 τ_h, are corrected from the pure $Z \to ee$ and $Z \to \mu\mu$ regions. Similarly, a correction is applied to account for residual differences in the $\mu \to e$ misidentification rate between data and simulation.

Further MC to data corrections are applied to the di-lepton mass and p_T spectra in simulated $Z \to \ell\ell$ events. These are derived from $Z \to \mu\mu$ events. Additionally, all simulated $t\bar{t}$ events are weighted to match the t quark p_T distribution observed in data [114].

4.9 Uncertainty model

The statistical uncertainties are taken into account by the Barlow-Beeston method, described in References [115, 116]. 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 in the reconstruction of an object and is decorrelated otherwise. It is decorrelated across the eras 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% are uncorrelated with MC uncertainties, because of the shared real data in the measurement.

Hadronic taus

Uncertainties on the τ_h triggers are obtained from the fitted scale factors used to derive the corrections for the τ_h trigger efficiencies. The legs of the double-τ_h and e/μ-τ_h cross triggers in different decay mode bins are treated as uncorrelated. For the single-τ_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-τ_h trigger uncertainties are further split into 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 τ_h candidates. These uncertainties range between 0.2-1.1%. Finally, an uncertainty on the identification efficiency is placed as a function of p_T in the $e\tau_h$ and $\mu\tau_h$ channels, and of the τ_h decay mode in the $\tau_h\tau_h$ channels. This varies 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 τ_h is applied and treated as uncorrelated between the channels where different $D^{e/\mu}_{\text{WP}}$ 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 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 on 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 are not included. Another 2% uncertainty is placed on the identification of any electron or muon in the event.

Jets

Jet energy scale and resolution uncertainties arise from several 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 subpercent 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 misidentified as τ_h candidates parametrised by the p_T of the $e/\mu \to \tau_h$ fake. The magnitude is 1.0% for muons in all eras. For electrons the uncertainties vary between 0.5 and 6.6 %.

Jets misidentified as hadronic taus

The backgrounds with jets misidentified as τ_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, which is calculated by comparing the fits to a first-order polynomial fit set to constant above 100 GeV. The final systematic variation is on the `Determination Region` to `Application Region` corrections by applying them twice and not at all to get symmetric shifts. 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

Backgrounds 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 extrapolations to the signal region that are $\mathcal{O}(10\%)$.

Muons misidentified as electrons

Backgrounds 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 is given a 15%–45% uncertainty, which is derived from the calculation of the correction.

MET

The MET uncertainties are different depending on the process. For all processes that are not $t\bar{t}$ or di-boson, the hadronic recoil response and its resolution are varied

within the uncertainties determined during the computation of the recoil corrections. For $t\bar{t}$ or di-boson an uncertainty is derived from the energy carried by an unclustered particle [117]. These uncertainties vary between 0–10%.

Background process-specific uncertainties

Uncertainties on the $t\bar{t}$ p_T and Z boson $m_{\ell\ell}$-p_T reweighting are 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 τ pair is shifted up and down by 10%. Some non-closures are observed in embedded $Z \to \mu\mu$ 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 $Z \to ll$ and W + jets production [107], 6% for $t\bar{t}$ production [108, 109], and 5% for diboson and single t quark production [109–111].

Signal process specific uncertainties

For the ggϕ and bbϕ processes, the variation of the `hdamp` parameter of the POWHEG MC generator as well as the μ_R/μ_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 (t-only, b-only, tb-interference). These are also determined from additional samples produced at generator level and applied as event weights depending on p_T after the parton shower simulation. For the vector leptoquark signal samples, the parton distribution functions and μ_R/μ_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^{tot} distribution as well as the predicted signal yields.

Luminosity

1.2%, 2.3% and 2.5% normalisation uncertainties for the luminosity are applied to the 2016, 2017 and 2018 templates respectively, which originate from MC simulation to data comparisons.

Prefiring

Upper and lower bounds are taken from the efficiency maps and propagated to all MC samples as shape uncertainty for 2016 and 2017. The size of the uncertainty depends on the event topology but averages to a value of 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,

$$\mathcal{L}(\text{data} \mid \mu, \theta) = \prod_{i}^{N_i} \text{Poisson}\Big(n_i \mid \sum_{j}^{N_j} g_j(\mu_{ij}) \cdot s_{ij}(\theta) + \sum_{k}^{N_k} b_{ik}(\theta)\Big) \cdot p(\hat{\theta} \mid \theta), \quad (4.10)$$

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. θ represents the set of nuisance parameters (corresponding to the systematic uncertainties as detailed in Section 4.9) that parametrise the signal and background modelling. μ are rate parameters and $g(\mu)$ are scaling functions that scale 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 distribution function (PDF) of each nuisance parameter (θ) 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 variation are derived and vertical morphing [116] is used to interpolate and extrapolate within and outside the shifts. Both PDFs are dependent on the mean (μ) and standard deviations (σ) and the functional forms are shown in Table 4.3.

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

Table 4.3: Table of 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 to a single mass hypothesis, however, within this mass hypothesis there can be a number of signal processes. The model-independent resonance search has separate ggϕ and bbϕ signal modes and so two rate parameters $\mu_{\mathrm{gg}\phi}$ and $\mu_{\mathrm{bb}\phi}$ are needed. The samples are initially scaled to the cross-section times branching ratio ($\sigma \times B(\phi \to \tau\tau)$) of 1 pb. Setting $g(\mu) = \mu$ for both processes, $\mu_{\mathrm{gg}\phi}$ and $\mu_{\mathrm{bb}\phi}$ represent the $\sigma \times B(\phi \to \tau\tau)$ in units of pb. To avoid negative signal strengths, μ is not allowed to become negative. 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 into each channel or category utilising index i in Equation 4.10. This is used to determine the compatibility of the results in different decay channels and analysis categories. In this case, μ 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 $Z/\gamma^* \to \tau\tau$. 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 $g_{\text{t-channel}}(\mu) = \mu^4$ and $g_{\text{interference}}(\mu) = \mu^2$ to mimic how the cross-sections of each process scale. When the initial samples are scaled to the cross-section at $g_U = 1$, μ 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 ggϕ samples are also split into 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 Equation 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 μ 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 μ. The 68% and 95% confidence level (CL) 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 frequentist approach [118, 119] with a profile likelihood ratio used for the test statistic, as defined below.

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

where $\hat{\mu}$ and $\hat{\theta}_{\hat{\mu}}$ are the best fit values of μ and θ_μ. $\hat{\theta}_\mu$ are the values of θ_μ that are maximised by the likelihood for a tested value of μ. 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,

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

where $f(q_\mu \mid \mu, \theta_\mu^{\text{obs}})$ is the PDF of q_μ. CL_b and CL_{s+b} are then defined by the relevant background-only and signal-plus-background fits. CL_s is defined as the ratio of CL_{s+b} and CL_b and then upper limits are placed at the confidence level of $1 - \text{CL}_s$. The $f(q_\mu \mid \mu, \theta_\mu^{\text{obs}})$ are determined using the asymptotic approximation [120] 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 significance. To test the rejection of the background-only hypothesis in favour of the signal-plus-background hypothesis, μ is replaced with 0 in the test statistic. The p-value, p_0 is then,

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

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 are displayed in the ratio. Figure 4.20 shows the distributions of the m_T^{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 ggϕ and bbϕ resonances and 1 TeV VLQ BM 1 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 ggϕ signal with a cross-section times branching ratio of 5.8 pb. In this same fit, the bbϕ 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 the high m_T^{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 because 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 ggϕ signal, the bbϕ signal is heavily constrained and so largely does not contribute to the signal-plus-background fit of the excess. Similar to the bbϕ 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 the absence of a signal for the search for a ggϕ or bbϕ resonance and these are 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 ggϕ production, the expected limits flatten under 100 GeV, due to the difficulty of separating the signal from the Z boson at this mass. Both sets of limits vary from $\mathcal{O}$(10 pb) at 60 GeV to 0.3 fb at 3.5 TeV.

95% CL expected limits are drawn on the fit to each di-τ 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 the branching fractions, 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 $Z \to \tau\tau$ 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-τ_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 is also made with the ATLAS experiment and in particular the results presented in Reference [121]. 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 ggϕ and bbϕ 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 genuine di-τ samples compared to MC allow for lower background uncertainties and hence a stronger limit. The ATLAS result observed no excess of events compatible with a ggϕ signal at 1.2 TeV, in fact, a small deficit was

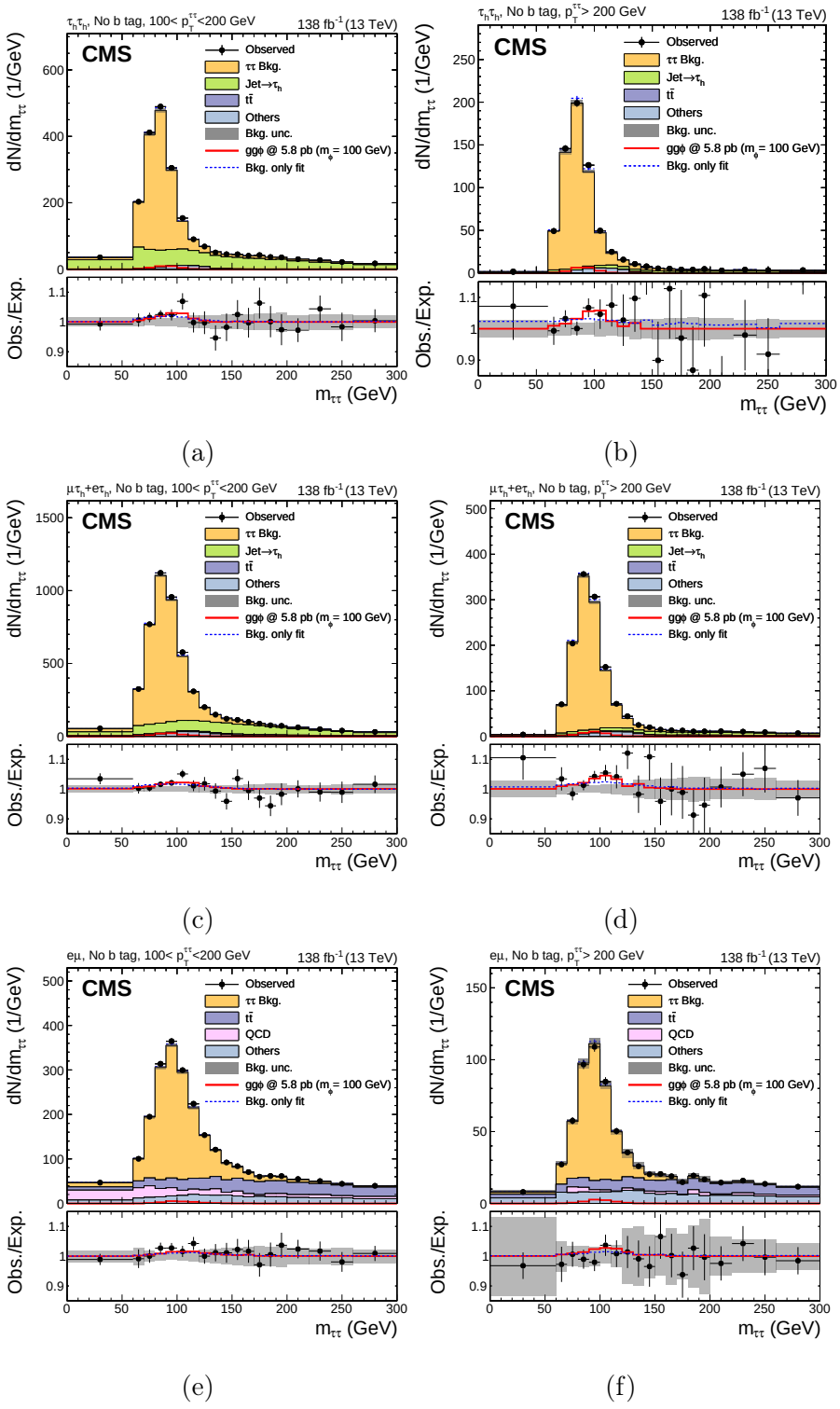


Figure 4.19: Distributions of $m_{\tau\tau}$ in the no b tag second highest (a, c and e) and highest (b, d and e) p_T category for the $\tau_h\tau_h$ (a and b), the combined $e\tau_h$ and $\mu\tau_h$ (c and d) and the $e\mu$ (d and e) channels. The solid histograms show the stacked background predictions 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 [2].

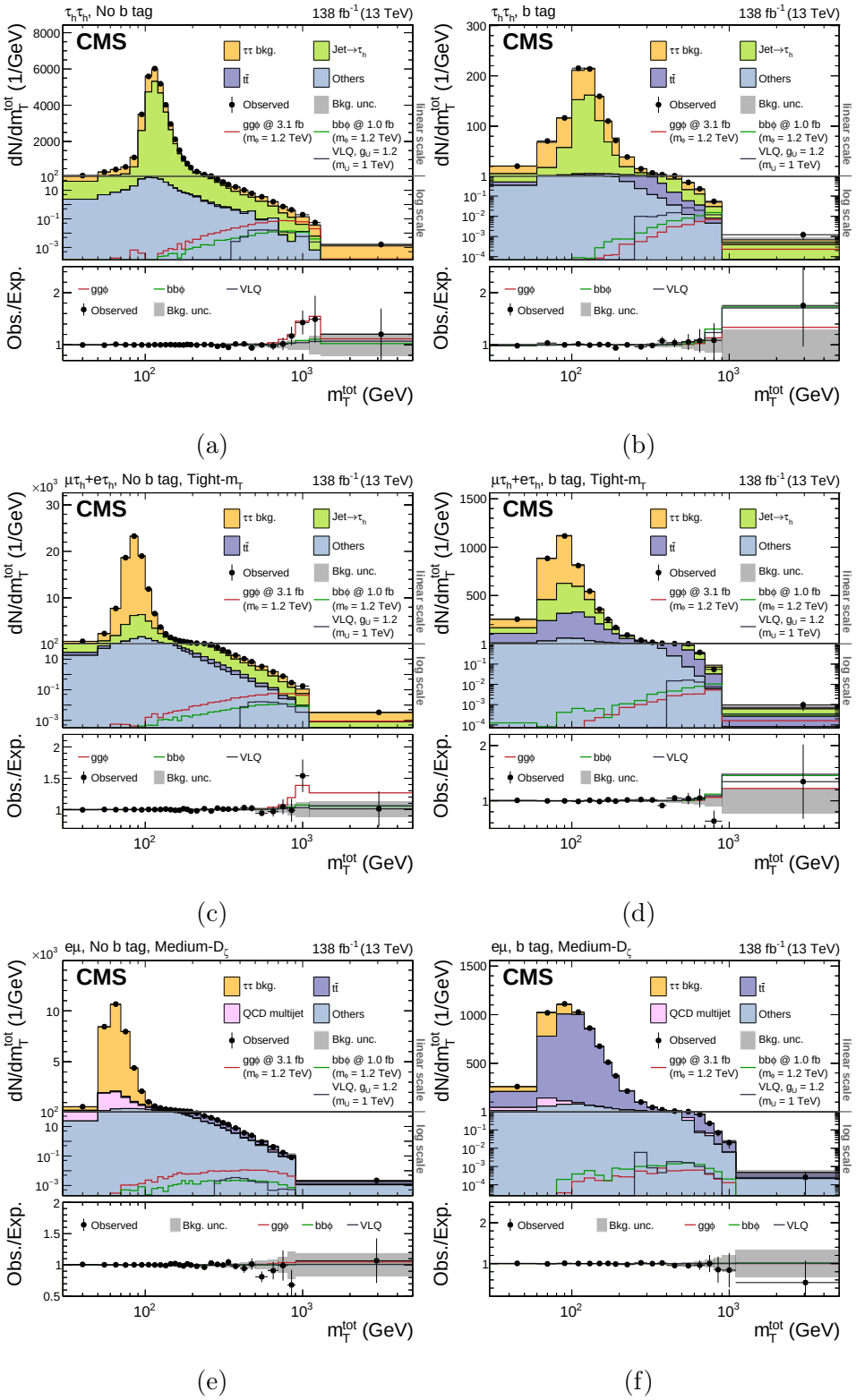
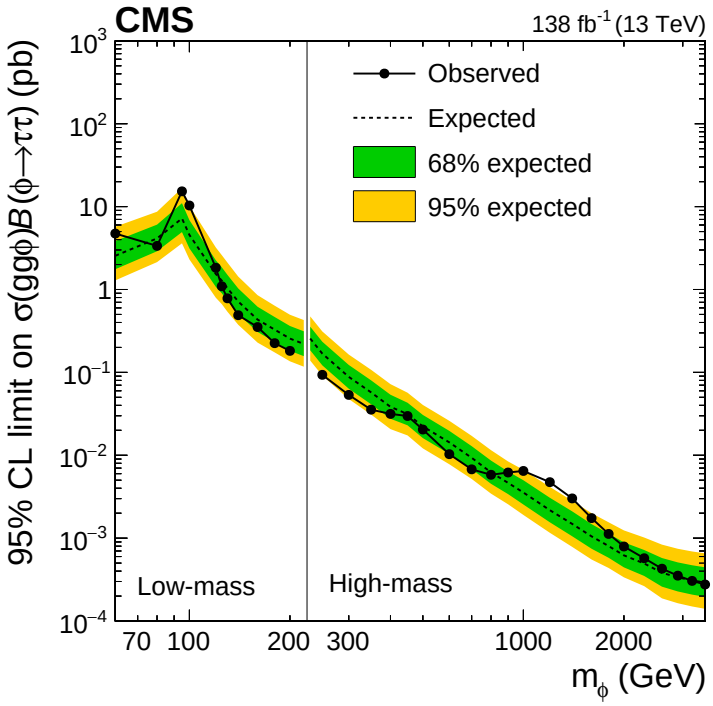
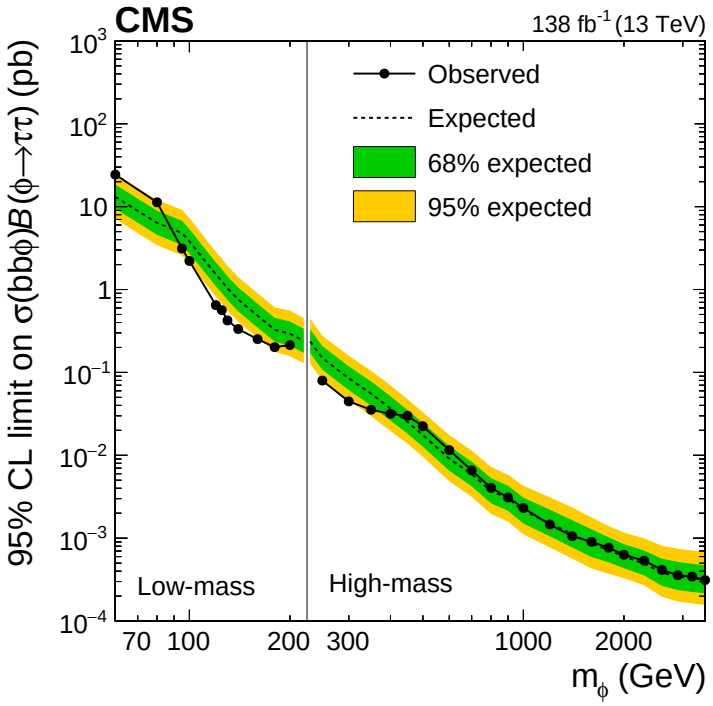


Figure 4.20: Distributions of m_T^{tot} in the $\tau_h\tau_h$ no b tag (a) and b tag (b) categories, the combined $e\tau_h$ and $\mu\tau_h$ no b tag (c) and b tag (d) `Tight-`m_T categories and the $e\mu$ no b tag (e) and b tag (f) `Medium-`D_ζ categories. The solid histograms show the stacked background 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 [2].

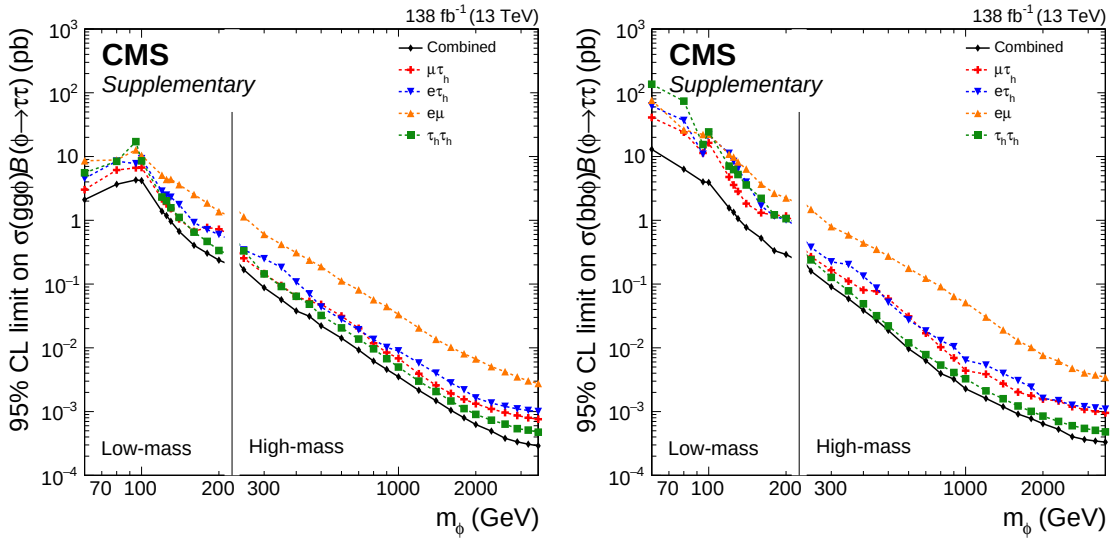


(a)



(b)

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 τ leptons for ggϕ (a) and bbϕ (b) 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 [2].



(a) (b)

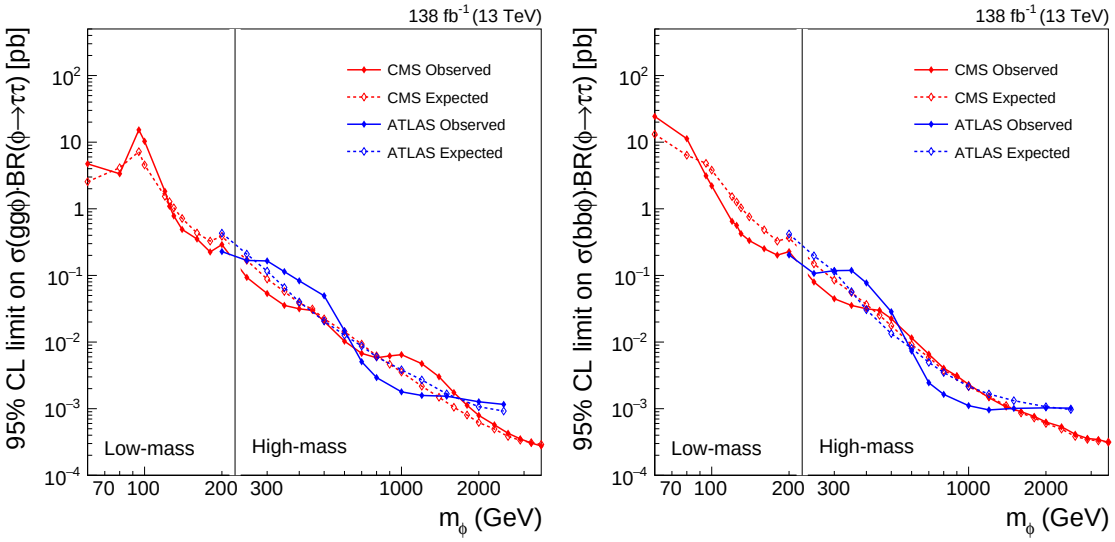
Figure 4.22: Comparison of the expected 95% CL upper limits on the product of the cross-sections and branching fraction for the decay into τ leptons for ggϕ (a) and bbϕ (b) production, split by the $\tau\tau$ decay products fit individually.

observed. Also, ATLAS observed local excesses at 400 GeV of 2.2σ for ggϕ and 2.7σ for bbϕ. None of these 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 as a cross-check for this.

4.12.2 Significance and compatibility

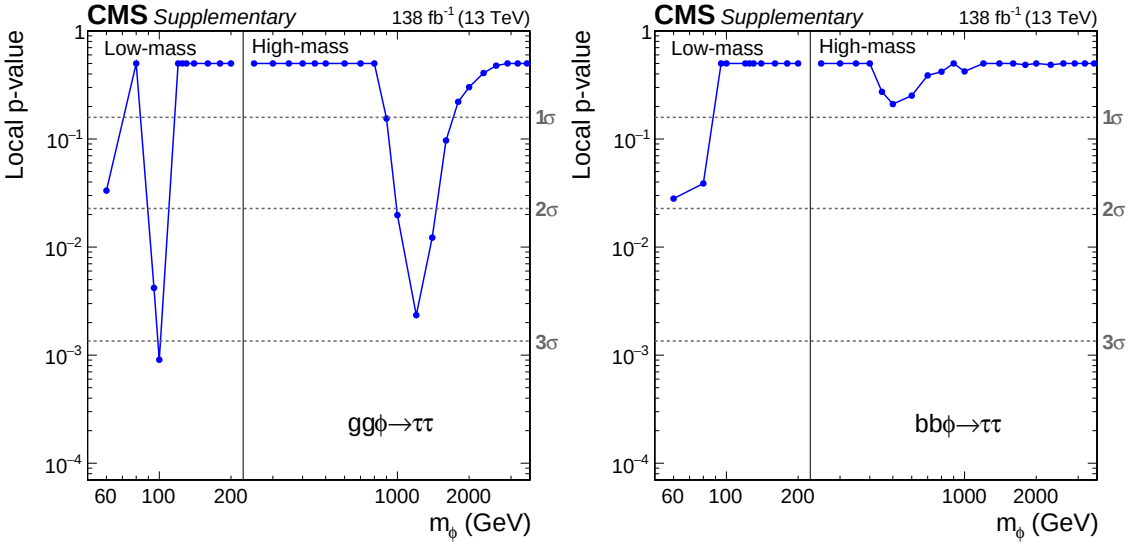
The p-values and significances at each model-independent signal hypothesis are calculated as described in Section 4.10 and shown in Figure 4.24. Identical to the model-independent limits, the ggϕ or bbϕ process is allowed to float freely if not the parameter of interest. The excesses for the ggϕ process peak at 100 GeV and 1.2 TeV and quantify to a local (global) significance of 3.1σ (2.7σ) and 2.8σ (2.2σ) 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 2σ are observed for bbϕ production.

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



(a) (b)

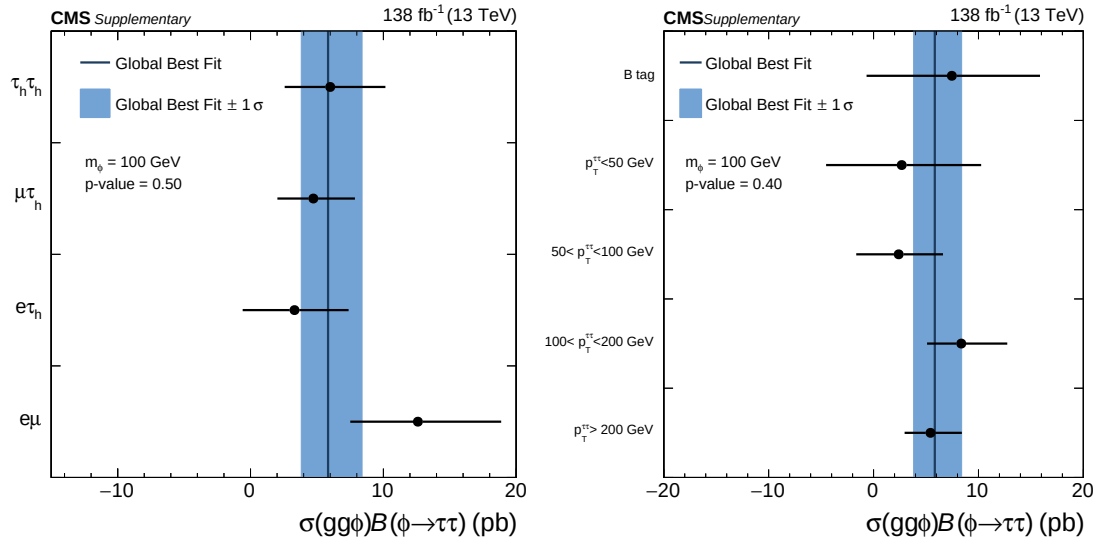
Figure 4.23: Comparison of the expected 95% CL upper limits on the product of the cross-sections and branching fraction for the decay into τ leptons for ggϕ (a) and bbϕ (b) production, split by the CMS result detailed in this thesis and the ATLAS result from Reference [121].



(a) (b)

Figure 4.24: Local p-value and significance of a ggϕ (a) and bbϕ (b) signal as a function of m_ϕ [2].

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-τ decay channels and the p_T bins fit. Figure 4.26 shows the compatibility fits in the high-mass optimisation categories split by di-τ 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-τ 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.



(a) (b)

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 [2].

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 two-dimensional 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

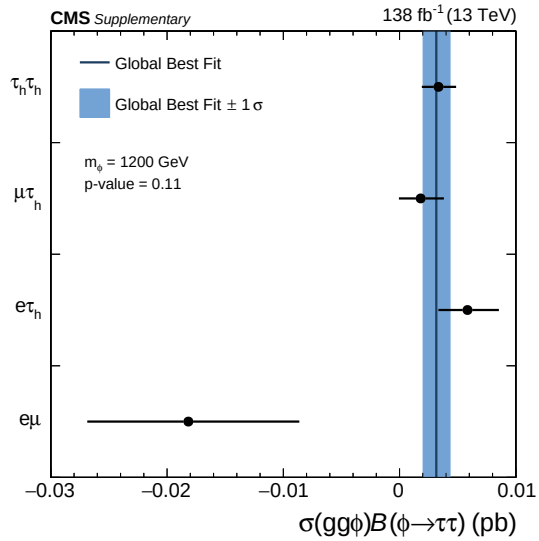


Figure 4.26: Compatibility plots of the 1.2 TeV excess split by analysis channels. In each case, the fitted signal strength is decoupled in each channel [2].

TeV are dominated in the phase space in which ggϕ and not bbϕ signals are allowed. In the 60 GeV example, there are smaller deviations in both ggϕ and bbϕ and the SM background is again over 2σ 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 ± 3 GeV boundary. For low values of $\tan\beta$, higher values of the additional SUSY particle masses, denoted as m_{SUSY}, are needed to explain a mass of approximately 125 GeV for the Higgs boson. In the M_h^{125} scenario, m_{SUSY} is fixed, and the predicted value of m_h is below 122 GeV. In contrast, the $M_{h,EFT}^{125}$ scenario adjusts m_{SUSY} to satisfy the required value of m_h for each point in m_A and $\tan\beta$ individually, accounting for the logarithmic corrections associated with the large values of m_{SUSY} using an effective field theory approach. The red hatched region in Figure 4.28b indicates that the required values of m_{SUSY} exceed the GUT scale at very low values of m_A in this scenario. The Higgs boson masses, mixing angle α, and effective Yukawa couplings were calculated using FeynHiggs, and branching fractions for the decay into τ leptons and other final states were obtained from a combination of the FeynHiggs and HDECAY, following the prescriptions in References [122–124], for the scenarios described in

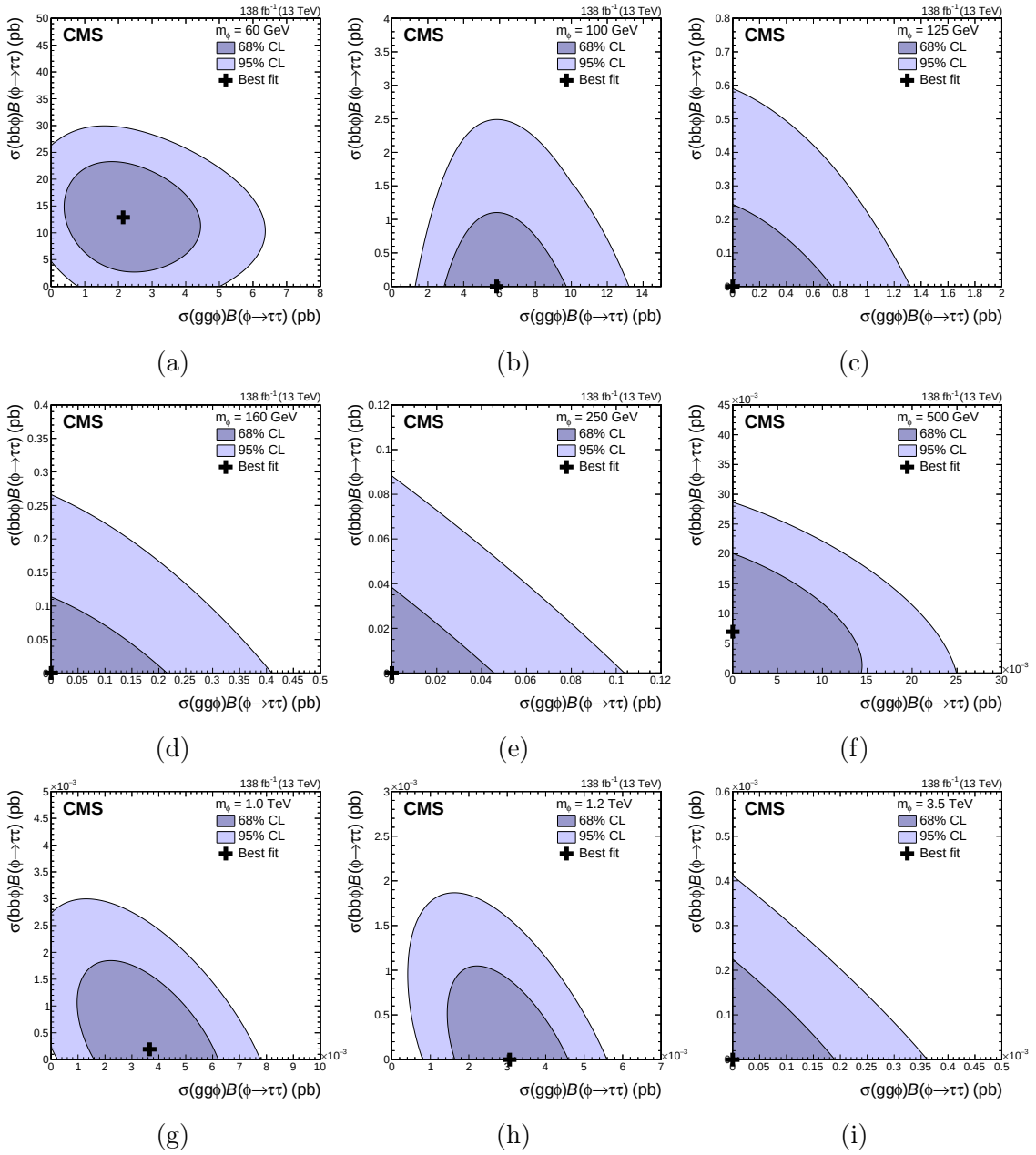


Figure 4.27: Maximum likelihood scans, including 68% and 95% CL contours obtained from the signal likelihood for the model-independent search. The scans are shown form_ϕ values of 60 (a), 100 (b), 125 (c), 160 (d), 250 (e), 500 (f), 1000 (g), 1200 (h) and 3500 (i) GeV [2].

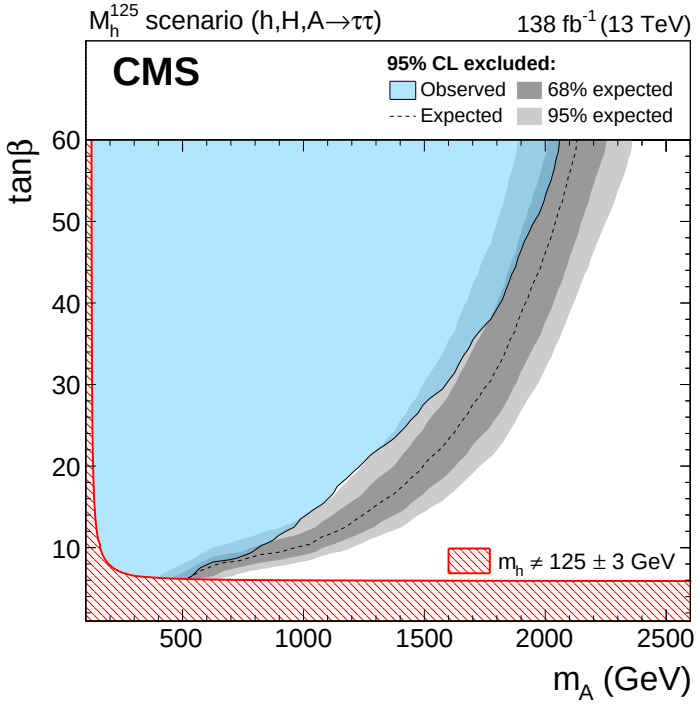
Reference [84].

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 τ leptons, when the A and H decay into two on-shell t 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 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 bbϕ signal in the b tag categories, especially for increasing values of $\tan\beta$.

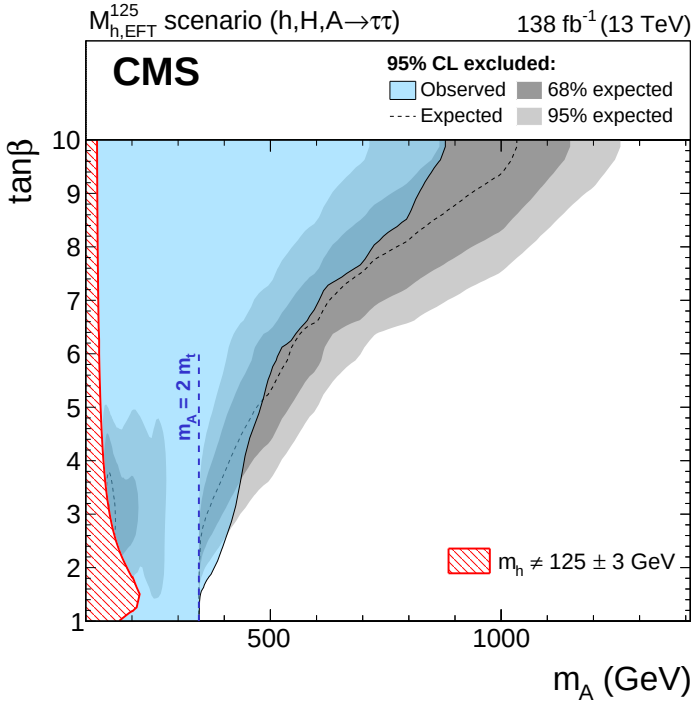
Other MSSM scenarios are tested and detailed in Reference [2]. One scenario of note is the M_H^{125} scenario, which is the equivalent scenario to the M_h^{125} but with the observed Higgs boson being the heavier CP-even Higgs boson. Despite the local excess at a resonant mass of 100 GeV, this scenario 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 the branching fractions. 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 in 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 τ 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 Reference [39] and described in Section 1.4.1. 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

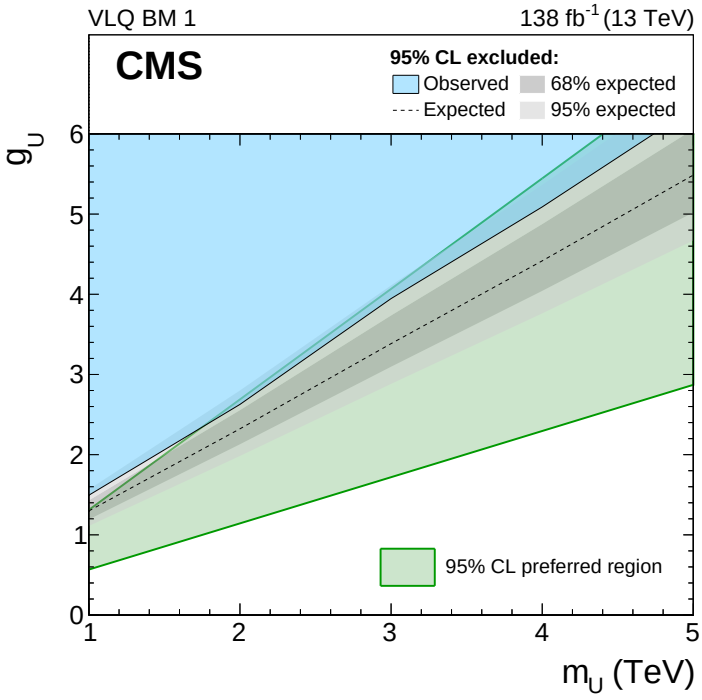


(a)

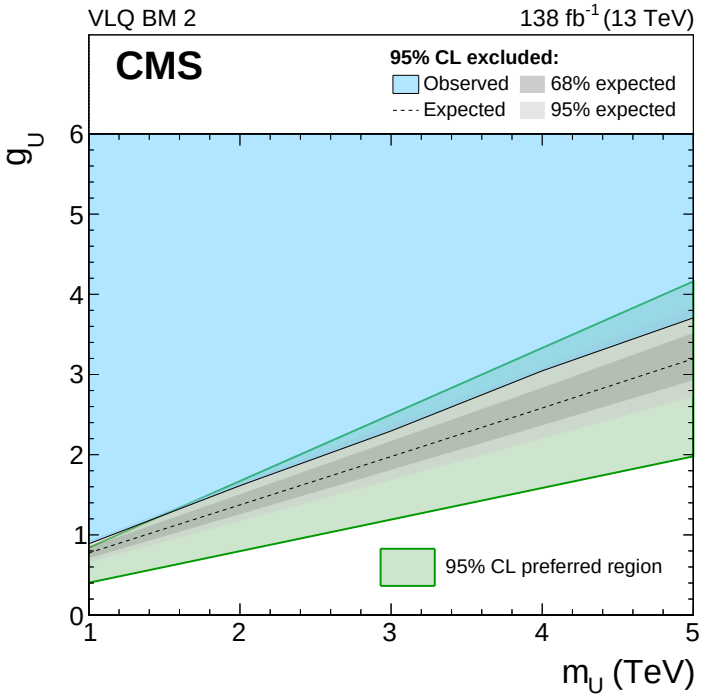


(b)

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 ±3 GeV from the observed SM Higgs boson mass is shown by a red-hatched area. [2]



(a)



(b)

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 that 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 Reference [39] is also shown by the green shaded area [2].

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 the 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.

Chapter 5

Search for new physics in $\tau^+\tau^-\tau^+\tau^-$ final states

Enhancements from $\tan\beta$ to up or down-like quark couplings to additional neutral 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 is the case that both up and down-like couplings to additional Higgs bosons are suppressed. This parameter space is left relatively untouched by "MSSM-like" searches. One example of this, the type X 2HDMs where only lepton couplings are enhanced by $\tan\beta$, allows for BSM loop contributions to SM measurements through couplings between leptons and additional Higgs bosons. This is particularly interesting in the context of the muon g-2 anomaly [13, 14] 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 that is not suppressed at high $\tan\beta$, through the process $Z^* \to \phi A \to 4\tau$, where in this chapter ϕ is defined as the additional CP-even Higgs boson (h if $\mathrm{h_{obs}} = \mathrm{H}$, H if $\mathrm{h_{obs}} = \mathrm{h}$). This search is split up into two sections:

i) A model-independent search for the $Z^* \to \phi A \to 4\tau$ 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 ϕ and A decaying to a pair of τ leptons.

ii) A search for the type X 2HDM, motivated by the phase space for possible explanations for the muon g-2 anomaly. The m_A-$\tan\beta$ phase space for scenarios of m_ϕ 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_ϕ and m_A.

These searches are performed with the full Run 2 dataset (138 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 τ leptons. To probe the type X 2HDM at high $\tan\beta$, a production process that is not suppressed is required. Reference [42] discusses that the following production modes of two additional neutral Higgs bosons are dominant to produce any of these new particles at high $\tan\beta$:

i) $pp \to Z^* \to \phi A \to (\tau^-\tau^+)(\tau^-\tau^+)$

ii) $pp \to Z^* \to H^+H^- \to (\tau^-\nu)(\tau^+\nu)$

iii) $pp \to W^{\pm *} \to H^\pm A \to (\tau^\pm\nu)(\tau^-\tau^+)$

iv) $pp \to W^{\pm *} \to H^\pm \phi \to (\tau^\pm\nu)(\tau^-\tau^+)$

As the production cross-sections of these four processes are of similar magnitudes, the search sensitivities depend on the separation of the signals from the background. In general, the more objects you can select in the final state, the smaller the background contributions. This is certainly true in τ enriched final states, where backgrounds can be dominated by jets misidentified as τ_h objects and so every extra τ selected reduces this background. In particular, (ii) has production cross-sections [42] far smaller than the observed limit for gluon fusion production of a single resonance shown in Figure 4.21a and it is not possible to use τ 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.

Signal templates for the production of this process with a mass grid for ϕ and A between 100 to 300 and 60 to 160 GeV respectively are generated. These mass ranges are motivated by the results in Table 1.4. The samples are simulated in the five FS at NLO precision using the MadGraph5_aMC@NLO v2.6.5 event generator [96]. Generation is performed using the parton distribution function NNPDF3.1 [90, 91], where the τ lepton decay, parton showering and hadronisation are all modelled with

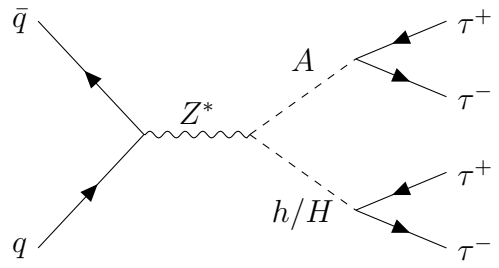
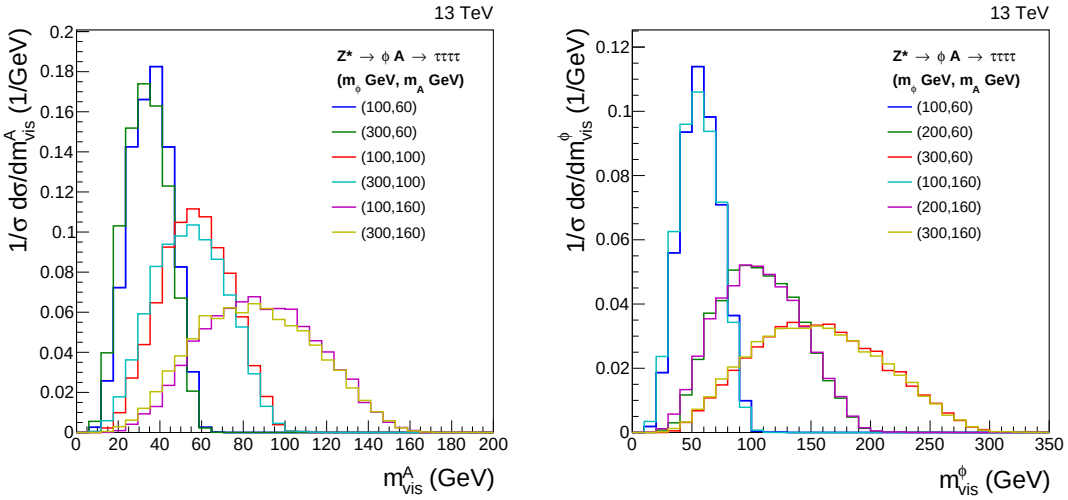


Figure 5.1: Diagram of the production of two additional neutral Higgs bosons from an off-shell Z boson and their decay to τ leptons.

the PYTHIA event generator with the PU profile matched to data [92,93]. The events are then passed through the GEANT4-based [94] simulation of the CMS detector and reconstructed in the same way as data. Generator level distributions of di-τ mass distributions from the decay of ϕ and A are shown in Figure 5.2.



(a) (b)

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

The cross-sections in the alignment scenarios are also determined with this procedure and vary from 10 fb ($m_A = 60$ GeV and $m_\phi = 100$ GeV) to 650 fb ($m_A = 160$ GeV and $m_\phi = 300$ GeV), 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 $\cos^2(\beta - \alpha)$.

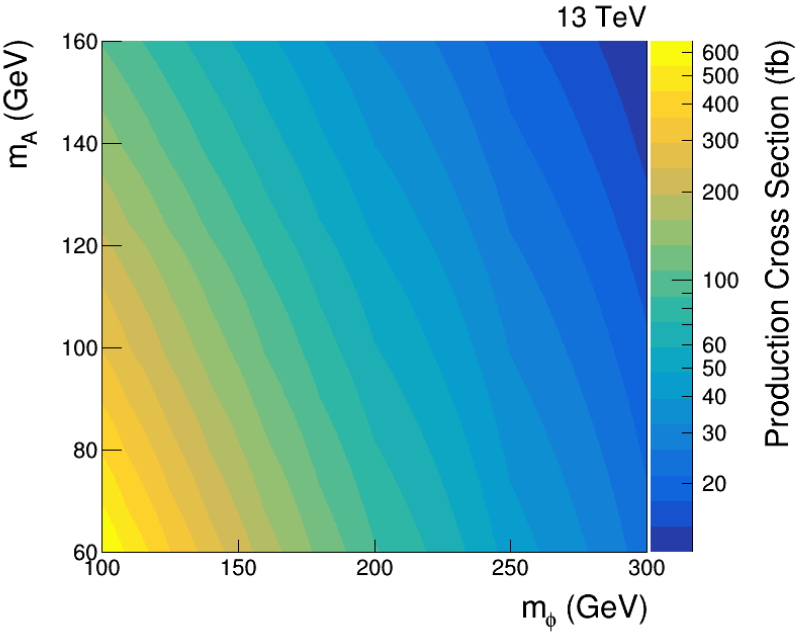
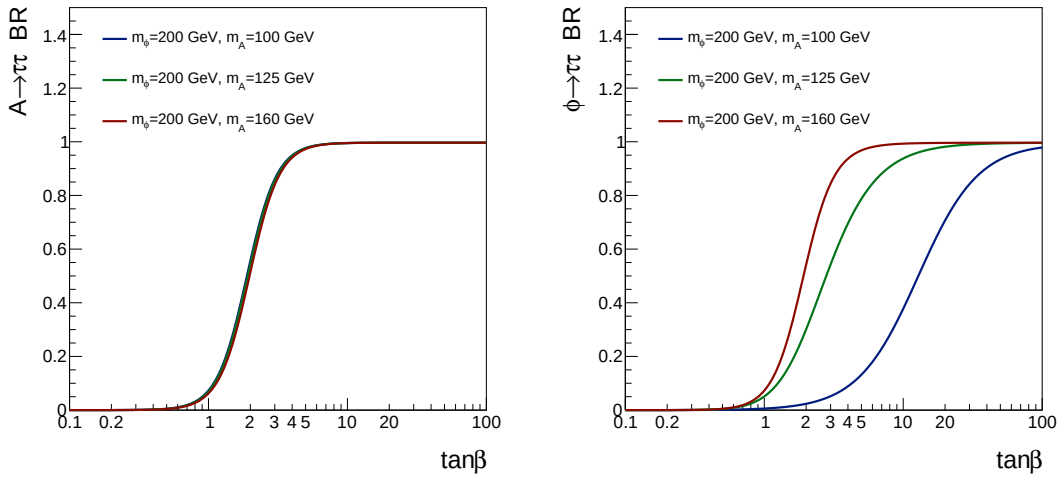


Figure 5.3: Calculated productions cross-sections for the $Z^* \to \phi A$ process, varying the masses of ϕ and A.

The branching fractions of ϕ and A to pairs of τ leptons are dependent on both $\tan\beta$ and $\beta - \alpha$. For this analysis, the branching fractions are calculated using 2HDECAY [125]. In the alignment scenarios, the $A \to \tau\tau$ branching fractions are approximately 1 above $\tan\beta \approx 2$, where below they sharply drop off and other processes such as $A \to b\bar{b}$ become dominant. This is also true for $\phi \to \tau\tau$ branching fractions, except in the case where m_ϕ is greater than m_A by more than m_Z, and so the $\phi \to ZA$ 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 ϕ to τ leptons become smaller as the magnitude of the coupling of additional neutral Higgs bosons to τ leptons is reduced, whilst the A branching fractions, like the couplings, are left unchanged. An example of the ϕ branching fractions out of the alignment scenario is shown in Figure 5.5.

5.2 Event selection

In comparison to Chapter 4, four τ leptons produce a much larger number of possible final states. All variations of e, μ and τ_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



(a) (b)

Figure 5.4: Calculated branching fractions of A (a) and ϕ (b) decaying to a pair of τ leptons for various mass scenarios.

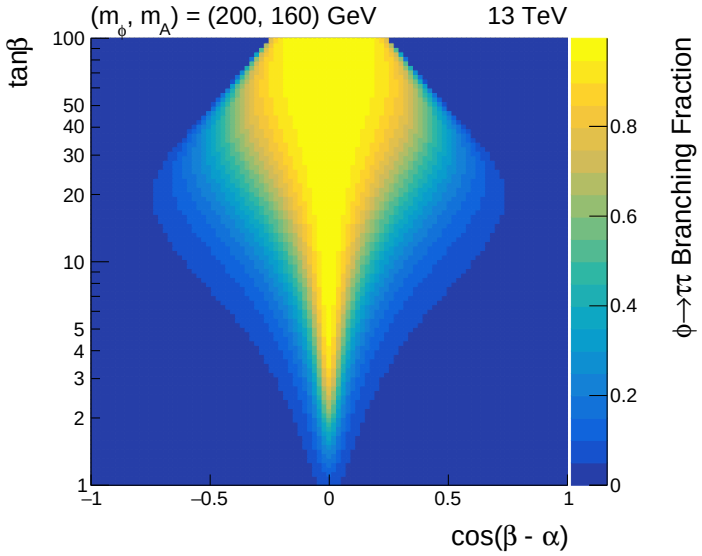


Figure 5.5: Calculated branching fractions in the $\cos(\beta - \alpha)$-$\tan\beta$ phase space for ϕ of mass 200 GeV decaying to a pair of τ leptons, in the scenario where $m_A = 160$ GeV.

are explored in this analysis. In addition to this, an orthogonal $\tau_h\tau_h\tau_h$ channel is added to target events where the reconstruction of the $\tau_h\tau_h\tau_h\tau_h$ channel loses a single τ_h object. This can come about due to low triggering and identification efficiencies of τ_h candidates, as well as the high p_T thresholds required for both.
In total, the analysis consists of seven 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 τ leptons, where e and μ represent the leptonic decay of the τ and τ_h represent the hadronic decay of the τ.

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-μ triggers. This is not the case for the single-τ_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-μ, double-τ_h, e-μ cross, μ-τ_h cross and e-τ_h cross-triggers. The cross-triggers and double-e/μ 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 η thresholds for the remaining triggers (single-e/μ and double-τ_h) are equivalent to what is stated in Section 4.2.1.

5.2.2 Offline requirements

All offline selections stated in this section are in addition to the object selection discussed in Section 3. In this analysis, τ_h candidates are required to pass the `Loose` $D_{\text{jet}}^{\text{WP}}$. The `VVLoose` D_e^{WP} and `VLoose` D_μ^{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 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 a few further selections on the total τ collection. As the two τ pairs from the signal originate from two neutral additional Higgs bosons, the sum of charges of the fully reconstructed objects should be zero. This is applied in the decay channels where there are four objects. In the $\tau_h\tau_h\tau_h$ channel, as the assumption is that a τ_h has been lost, the absolute value of the sum of the charges of the objects is required to be one. 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, identification and isolation requirements on these extra leptons match the loosest cuts required for a nominally selected electron or muon. Secondly, an extra τ_h veto is applied to the $\tau_h\tau_h\tau_h$ channel, to keep it orthogonal to the $\tau_h\tau_h\tau_h\tau_h$ channel. To do this, any extra τ_h candidate passing the signal selection is vetoed. The constraints set on the number of leptons and τ_h candidates for each channel are shown in Table 5.2.

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 τ_h channels as the $t\bar{t}$ process is negligible.

Channel	e	μ	τ_h
$\tau_h\tau_h\tau_h\tau_h$	0	0	$\geq$4
$\tau_h\tau_h\tau_h$	0	0	3
$\mu\tau_h\tau_h\tau_h$	0	1	$\geq$3
$e\tau_h\tau_h\tau_h$	1	0	$\geq$3
$e\mu\tau_h\tau_h$	1	1	$\geq$2
$\mu\mu\tau_h\tau_h$	0	2	$\geq$2
$ee\tau_h\tau_h$	2	0	$\geq$2

Table 5.2: Number of objects required to be selected in each decay channel.

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 τ_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 $ee\tau_h\tau_h$ and $\mu\mu\tau_h\tau_h$ channel backgrounds come from $Z \to \ell\ell$ (or $Z \to \tau\tau \to \ell\ell$) with two jets misidentified as τ_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 τ leptons or from the $t\bar{t}$ process. There is less preference towards the light leptons being of opposite sign in the signal samples compared to in the background, 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.

Similarly to Section 4.3.1, events in each 10 analysis channels and categories are drawn out into a histogram based on the same discriminating variable, the total transverse mass (m_T^{tot}). However, as there are more objects in the final state, the definition is extended to what is shown below.

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

where N_τ is the number of objects in the final state, and τ_i refers to the visible products of the ith τ lepton. This variable again provides excellent discriminating

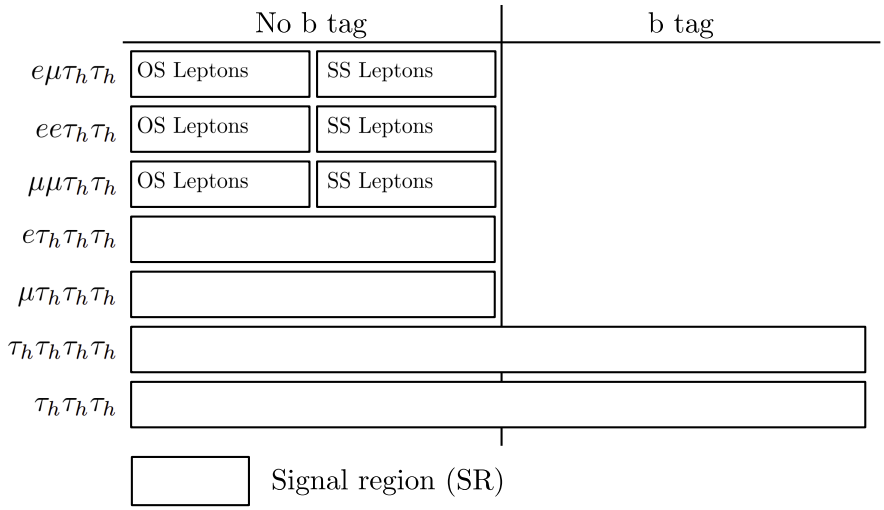


Figure 5.6: Overview of the categories used for the extraction of the signal in the $Z^* \to \phi A \to 4\tau$.

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, when the correct τ objects are selected. As m_T^{tot} uses information about the whole event, in particular, the high p_T objects and high MET from many τ decays, a greater sensitivity to the signal is observed.

In this analysis, many of the histograms contain very few background events and to minimise statistical fluctuations in the background templates, histogram bins are merged based on the fractional statistical uncertainty of each bin. This leads to many 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 three categories:

i) Events containing only genuine τ leptons.

ii) Events with one or more jets misidentified as a τ_h candidate (jet $\to \tau_h$).

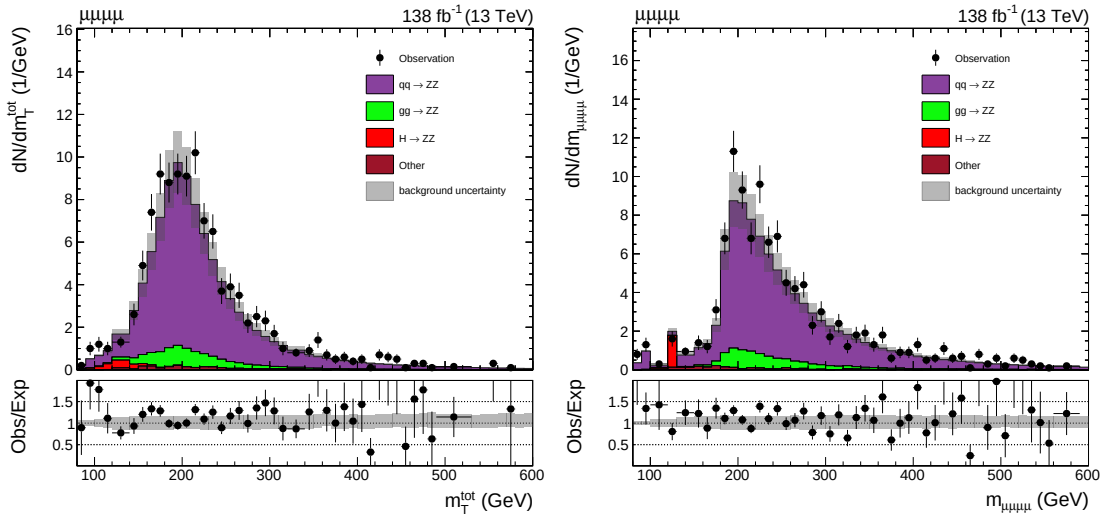
iii) Events with one or more light leptons misidentified as a τ_h candidate and no jet $\rightarrow \tau_h$ objects.

Background contributions from (i) are mostly from gluon and quark-initiated di-Z production and are modelled using MC. The details and validation of this modelling are described in Section 5.5. Background (ii) accounts for a number of different processes with jet $\rightarrow \tau_h$ candidates. 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 (ML) to improve upon the method, and discussed further in Section 5.6. Contribution (iii) is small in comparison to the others and modelled with MC. Events with jets misidentified as light leptons and no jets misidentified as τ_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 NLO precision [96]. All corrections described in Section 4.8 are applied to MC, with extra corrections applied to the ZZ process, which is detailed in Section 5.5.

5.5 ZZ modelling

The di-Z background shapes, where there are no jet $\rightarrow \tau_h$ candidates are modelled using MC. The cross-sections for this analysis are scaled to higher-order precisions than generated using K factors. For quark-initiated di-Z production, PowHeg 2.0 [86–89] is used to determine NNLO/NLO QCD and electroweak corrections as a function of m_{ZZ}, which takes an average value of ≈ 1.2. Gluon initiated di-Z production corrections from LO to NNLO are derived with HNNLO v2 [126], again as a function of m_{ZZ}. This gives a much larger correction, equal to ≈ 2. The modelling is validated 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 muon I_{rel} cut is loosened to 0.35. The charges of the muon 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 Figure 5.7.



(a) (b)

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

5.6 Machine learning fake factor method

The F_F 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 high dimensional dependence on parametrisation. In this scenario, classical "bin and fit" methods can break down, as for every new parameter included the statistics of each fit are reduced. A compromise must then be reached between variable dependence and fit statistics and this is the case for the $p_T^{\text{jet}}/p_T^{\tau_h}$ binning described in Section 4.7.2. It is also the case, that the analysis in Chapter 4 uses well-checked assumptions on, for example, the consistency of τ_h decay modes from determination to signal regions, to average across the dependencies on this variable when calculating F_F, however, this will not always be the case. Also binned upon in the previous search, is the process dependence of the F_F. The binning used is an admission that the relevant parametrisation of the F_F between processes is too complicated to model with "bin and fit" methods. It can also be seen that the initial fits, do not do a good job of modelling all variables as corrections are needed.

There is a final problem with using a F_F method as described when there are mul-

tiple τ_h candidates in final states. In the $\tau_h\tau_h$ channel of the previous analysis, the F_F were only calculated from the leading τ_h candidate. This was valid, as the dominant backgrounds had two jet $\rightarrow \tau_h$ candidates rather than one genuine τ_h and one jet $\rightarrow \tau_h$ candidate. However, for this search, this assumption is not valid.

All of these reasons motivate a more generalised and smarter method to model jet $\rightarrow \tau_h$ backgrounds, as well as the extra importance of modelling these backgrounds where there are more τ_h candidates in the final state. This is done by utilising ML and in particular using a BDT for the purpose of multi-dimensional reweighting.

5.6.1 BDT reweighter

Reference [127] proposes a new method for reweighting utilising ML 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 χ^2,

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

where $w_{\text{leaf, 1}}$ and $w_{\text{leaf, 2}}$ are the entries weights in each leaf of the decision tree from the two datasets. The larger the value of χ^2, the more important reweighting is in this region. This tree is utilised many times in the reweighting algorithm, shown below:

i) Input training datasets 1 and 2 with a large number of variables.

ii) Build a tree as stated above. If not the first loop, use newly determined weights for dataset 2.

iii) Compute predictions in the leafs $r_{\text{leaf}} = \log \frac{w_{\text{leaf, 1}}}{w_{\text{leaf, 2}}}$. The logarithm is taken so weights in different trees can be summed as usually done in boosting.

iv) In each leaf, dataset 2 events are weighted 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 is how the bins used

for reweighting are found, and this step is repeated multiple times.

5.6.2 Fitting regions

Unlike the standard F_F method, statistics are not a problem using the BDT reweighter when choosing which variables you can use to parametrise the F_F. Therefore, rather than fitting two fake factor regions separately and then a correction from the sideband to signal region to account for missing parametrisation, regions A, B and D from Figure 4.10 are fit simultaneously to improve statistics. Also, all τ_h candidates are fit simultaneously as separate entries in the dataset and for each τ_h in the event, the remaining τ_h candidates are named the alternative τ_h candidates. 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 τ_h candidates is required to be 0.
   y_A: The sum of τ_h candidates is required to not be 0.
   x_C: All alternative τ_h candidates pass the `Loose` $D_{\text{jet}}^{\text{WP}}$.
   x_D: At least one alternative τ_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 the sum of τ_h candidates is required to be 1.
   y_A: The absolute value of the sum of τ_h candidates is required to not be 1.
   x_C: All alternative τ_h candidates pass the `Loose` $D_{\text{jet}}^{\text{WP}}$.
   x_D: At least one alternative τ_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 F_F from Section 4.7.1, where the definitions of the alternative sideband variable are extended to more than one other τ_h candidate. The $D_{\text{jet}}^{\text{score}} > 0.1$ selection is used as the alternative τ_h identification selection for this analysis and is also then extended to the alternative τ_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 between the four to three objects selected. Therefore, when getting F_F in the $\tau_h\tau_h\tau_h\tau_h$, the lowest $D_{\text{jet}}^{\text{score}}$ unused τ_h candidate is dropped and the variables are recalculated. This candidate is chosen to be removed to best mimic the initial selection of the τ_h candidates, where they are sorted by $D_{\text{jet}}^{\text{score}}$ and the highest-scoring candidates are chosen. As shown later in this section, there is no major dependence on the shifted variables and any effects from this removal are covered within the uncertainty model.

5.6.3 Variables used

The variables used have been shown to have F_F dependence previously [2,128], and additional properties of the τ_h candidate and the event. Also added are the variables that take you from A to C and D to C, from Figure 4.10. As all years are fit together, to account for any differences in F_F from year to year, this is also added. The final variable added is the p_T-ordered ranking of the τ_h candidates in the event. All the variables used are shown below.

i) The HPS decay mode of the τ_h candidate.

ii) p_T of the τ_h candidate.

iii) The ratio of the p_T of the jet that seeds the HPS reconstruction to the p_T of the τ_h candidate.

iv) η of the τ_h candidate.

v) The charge of the τ_h candidate.

vi) A boolean of whether the τ_h candidate passes a leg of the double-τ_h trigger.

vii) The total charge of the combined objects.

viii) The boolean of whether $D_{\text{jet}}^{\text{WP}}$ passes to `Loose` WP for the alternative τ_h candidates. These are sorted by p_T.

ix) Era of data taking.

x) τ_h p_T ordered event rank.

5.6.4 Machine learning subtraction method

In the standard F_F method, histograms are used to fit the F_F rather than datasets. Using histograms, the small fraction of events which are not jet $\rightarrow \tau_h$ objects 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 $\rightarrow \tau_h$ object 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 dataset. 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, to 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 are put into a dataset.

ii) jet $\rightarrow \tau_h$ and non jet $\rightarrow \tau_h$ objects are separated into the two classes that will be used for 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 a jet $\rightarrow \tau_h$ object or not.

v) The scores of the BDT for how likely the entry is not a jet $\rightarrow \tau_h$ object 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 the 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 entries with BDT score within the range of the bin are selected.
- Entries within this bin are then randomly sampled and removed.
- This stops when the number of events removed equals the number of non jet $\rightarrow \tau_h$ objects predicted in the MC histogram bin.

This method then gives a data-MC hybrid method to determine a dataset of predicted jet $\rightarrow \tau_h$ candidates 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 F_F 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 τ_h identification region in Figure 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 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 Kolmogorov-Smirnov test [129] 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 F_F derived in the $\mu\tau_h\tau_h\tau_h$ channel are shown in Figure 5.9.

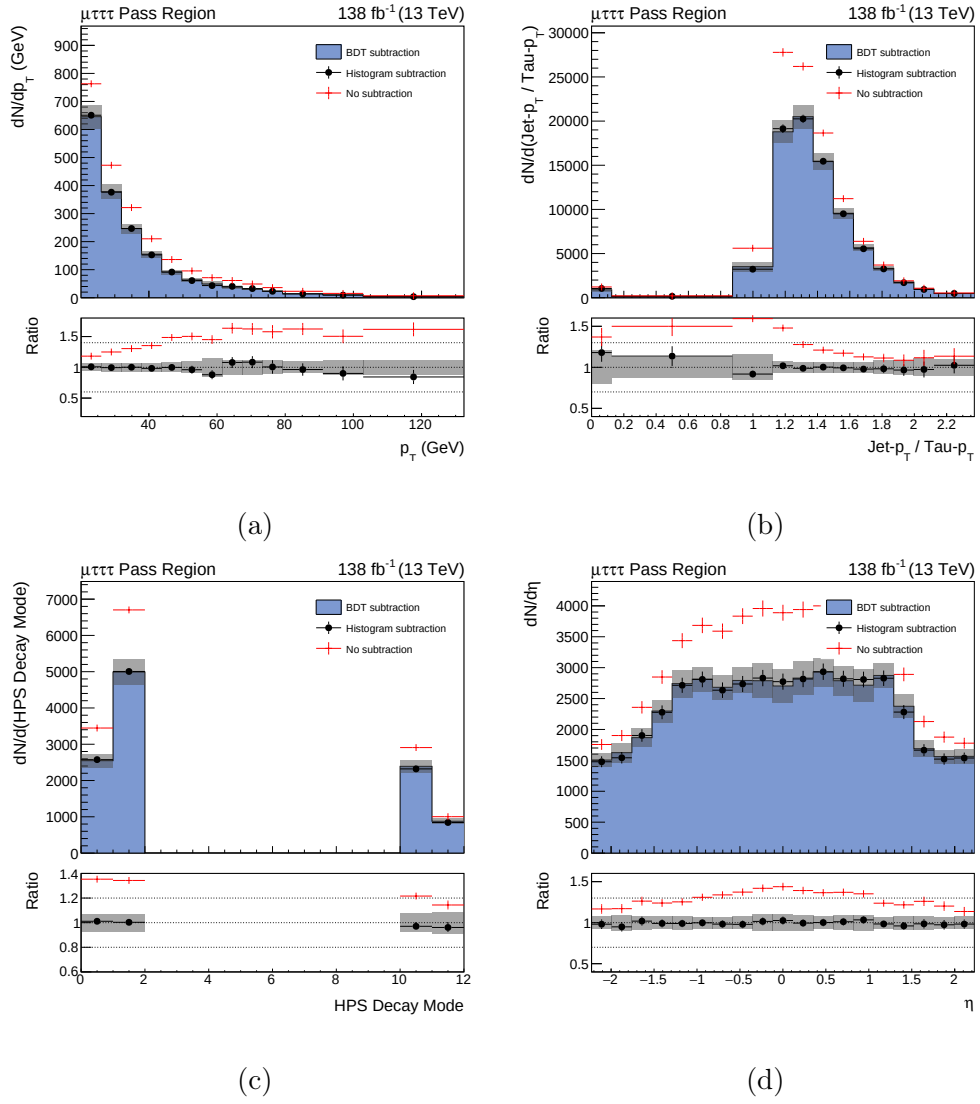
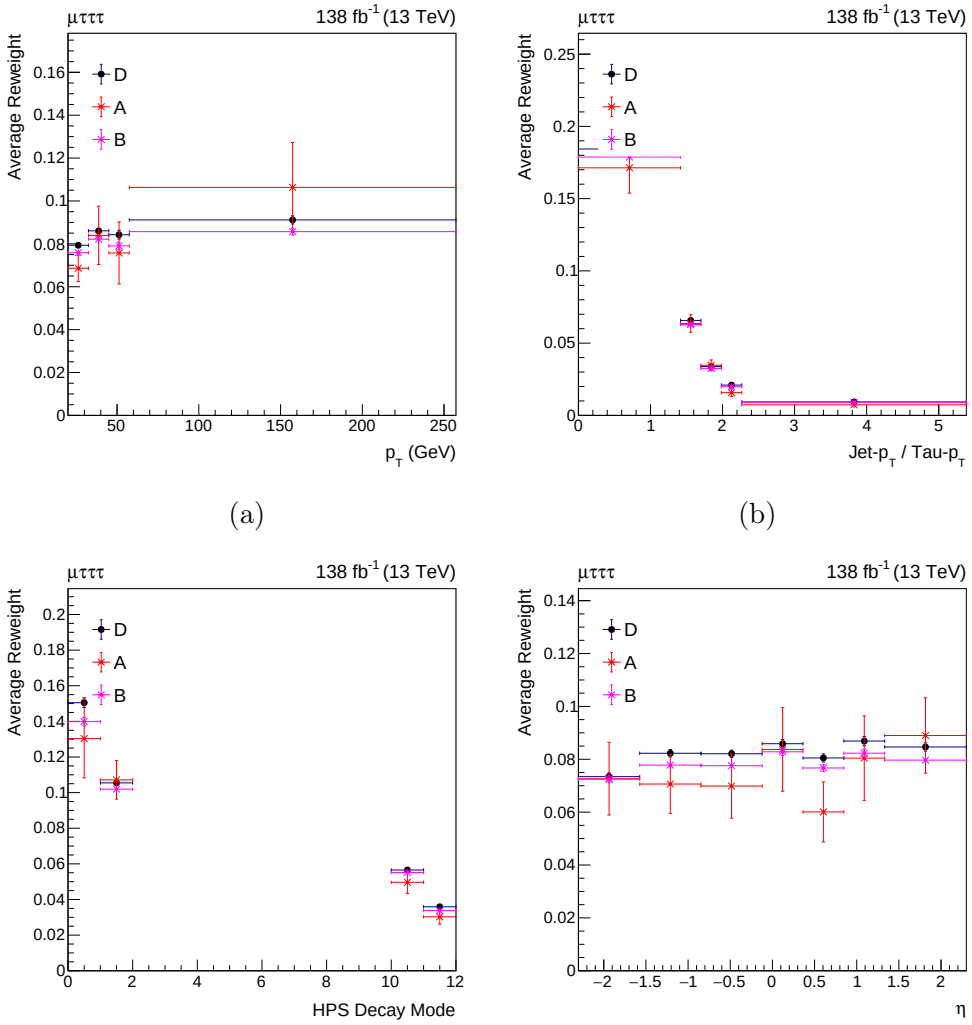


Figure 5.8: Comparison of histograms produced via the BDT subtraction method to histogram subtraction. Also shown is the histogram produced when no subtraction is performed. The uncertainty bands contain statistical uncertainties and uncertainties from on the non-closure of the method. This shown for four of the fitted variable: τ_h-p_T, the ratio of τ_h-p_T to jet-p_T, the τ_h HPS decay mode and the τ_h-η.



(a) (b)

(c) (d)

Figure 5.9: Average fake factors (reweights) calculated by the BDT reweighting method shown individually in regions A, B and D as defined in Figure 4.10. The is shown for four of the fitted variable: τ_h-p_T, the ratio of τ_h-p_T to jet-p_T, the τ_h HPS decay mode and the τ_h-η.

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 τ_h identification selections to the events with the nominal τ_h identification in all of the fitted regions simultaneously. Plots showing the closure of this method in the $\mu\tau_h\tau_h\tau_h$ channel, accompanied by this uncertainty, are shown in Figure 5.10.

5.6.6 Applying fake factors

Fake factors, F_F^i, have now been calculated for each τ_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 $\rightarrow \tau_h$ objects in an event. Taking channels with two τ_h candidates as the simplest example, if the jet $\rightarrow \tau_h$ background is determined purely off the leading τ_h candidate, then events where the leading τ_h is genuine and the sub-leading τ_h is a jet, are missed. Similarly, the situation can be flipped if the sub-leading τ_h is chosen to determine the jet $\rightarrow \tau_h$ fake background. A third option can be tried where both τ_h candidates are used to determine the background. However, this will only model events where both τ_h candidates are jets and not where there is a single jet $\rightarrow \tau_h$ object. It is seen that if the first two attempts at calculating this background from individual candidates are added and the contribution from both τ_h candidates is subtracted, all possible numbers of jet $\rightarrow \tau_h$ objects in the event are accounted for, as shown in Table 5.3.

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 τ_h^1)	0	1	1
R_2 (from τ_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 τ_h (τ_h^1), the sub-leading τ_h (τ_h^2) and both, to model the jet $\rightarrow \tau_h$ background and whether that specific object is a genuine τ lepton (τ) 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 ap-

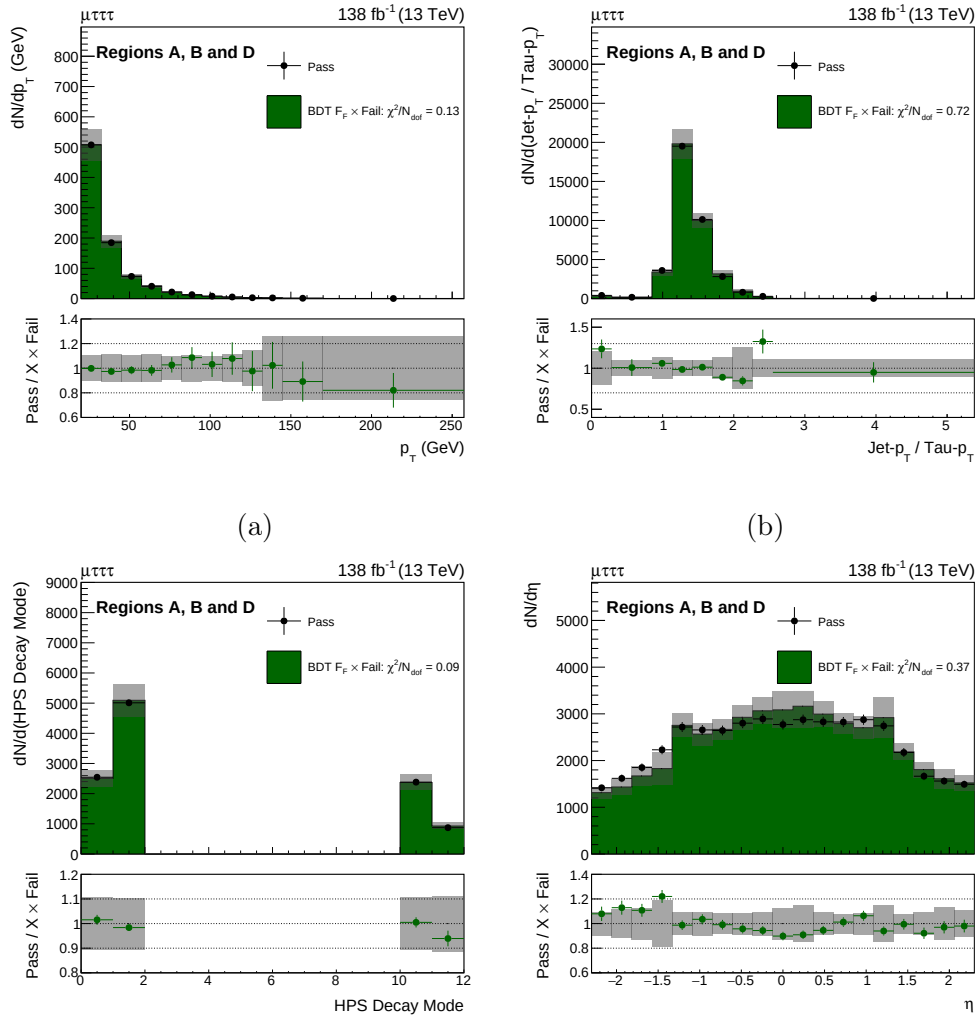


Figure 5.10: Comparison of histograms produced using the fake factors applied to the fitted fail τ_h identification region compared to the fitted pass τ_h identification region. The uncertainty bands contain statistical uncertainties and uncertainties derived from the non-closure of the method. The is shown for three of the fitted variable: τ_h-p_T, the ratio of τ_h-p_T to jet-p_T, the τ_h HPS decay mode and the τ_h-η. The χ^2 divided by the number of degrees of freedom between the two histograms is also shown.

plication 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 is only one jet $\rightarrow \tau_h$ candidate are not accounted for. As the majority of the background events in this channel come 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 $e\mu\tau_h\tau_h$

$$R_1 + R_2 - R_{12} \tag{5.3}$$

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

$$R_1 + R_2 + R_3 - R_{12} - R_{13} - R_{23} + R_{234} \tag{5.4}$$

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

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

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 τ_h identification uncertainty correlation scheme applied to simulated events. The fit used to derive scale factors for the τ_h MC events consists of both statistical and systematic uncertainties. In the previous search, the identification uncertainties were correlated across decay mode and the era of dat taking despite the statistical components of each fit begin orthogonal. Therefore, for this search, the uncertainty scheme contains correlated and decorrelated parts

across the HPS decay mode and the era of data taking. The double-τ_h trigger uncertainties remain unchanged.

Jets misidentified as hadronic taus

The backgrounds with jets misidentified as τ_h are estimated from data with the ML F_F 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 $\tau_h\tau_h\tau_h$ where the same fit is used, so uncertainties are correlated.

The initial uncertainties come from the 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 τ_h identification regions. The second uncertainty to do with purifying 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 are 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 τ_h identification region to the pass τ_h identification 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 is 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 the 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

Parton distribution functions, α_s and μ_R/μ_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. However, these yields are factored out for the model-independent search, where the cross-section (times branching ratios) is searched for.

5.8 Signal extraction

The statistical interpretations of the results are done as described in Section 4.10, but with different signal scaling functions, g, and parameters of interest, μ. The two interpretations of the analysis, as stated at the beginning of the 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 $Z^* \to \phi A$ multiplied by the branching fractions of $\phi \to \tau\tau$ and $\mathrm{A} \to \tau\tau$. Secondly, the interpretation in the type X 2HDM model is done by testing each point in the parameter space, whether that is m_A-$\tan\beta$ for the alignment scenario or $\cos(\beta-\alpha)$-$\tan\beta$ 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

Figure 5.11 shows the distributions of the m_T^{tot} discriminator, after a background-only fit to data, in every bin used in the fit. For visualisation, categories with similar event numbers in each bin are displayed on the same plot. A stacked background of events is shown, separated into three groups: events with 1 or more jet $\to \tau_h$ objects, events where all τ_h candidates are reconstructed correctly, and the remaining events where only light leptons are misidentified as τ_h and not jets. An example signal hypothesis is also shown for a mass hypothesis of $m_A = 160$ GeV and $m_\phi = 200$ GeV scaled to 0.01 pb, which is approximately three times smaller than the predicted

cross-section for this process.

There are no upward deviations of the number of events observed, that would be consistent with any mass hypotheses for the signal model searched for. The combined results are consistent with a background-only fit to data. Within this combined fit, individual bins and categories such as the $e\mu\tau_h\tau_h$ `SS Leptons`, have small deficits of events observed in comparison to events expected. The background prediction yield in this category from the F_F method is checked with a comparison to MC for the non-QCD prediction. The background estimations are compatible when the estimation of the fraction of QCD events from the charge inverted region is taken into account, further validating the F_F background prediction in this category.

5.9 Model-independent results

5.9.1 Limits

95% CL limits are set on the cross-section for two additional neutral Higgs bosons produced via an off-shell Z boson multiplied by the branching fraction of each additional boson decaying to τ lepton pairs and are shown in Figure 5.12. To show this for all 72 mass hypotheses, the limits on higher masses of the A boson are scaled to larger negative powers of 10, as indicated on the plot. The observed limit falls between the lower 1 and 2σ bands of the expected result. This is consistent with the distributions observed in Figure 5.11, as the small deficits in sensitive channels lead to a strengthening of the limit. The observed limit in comparison to the expected limit is relatively consistent across the signal mass range. This is because of the degeneracy of the signal hypothesis in the sensitive category bins fit. Therefore, the strongest constraints on different mass hypotheses mostly arise from the same bins. The observed limits vary from 20 fb at the lowest mass hypothesis of $m_A = 60$ GeV and $m_\phi = 100$ GeV, to 1.4 fb at the highest mass hypothesis of $m_A = 160$ GeV and $m_\phi = 300$ GeV. It is worth noting that each of the observed limits is well below the predicted cross-sections calculated and shown in Figure 5.3, indicating excellent sensitivity to the type X 2HDM in the alignment scenario.

Figure 5.13 shows the effect of individual decay channels, by showing the expected 95% CL limit on fits to each decay channel separately for an $m_A = 100$ GeV scenario. It is observed from this, that the dominant search channels, across the whole m_ϕ

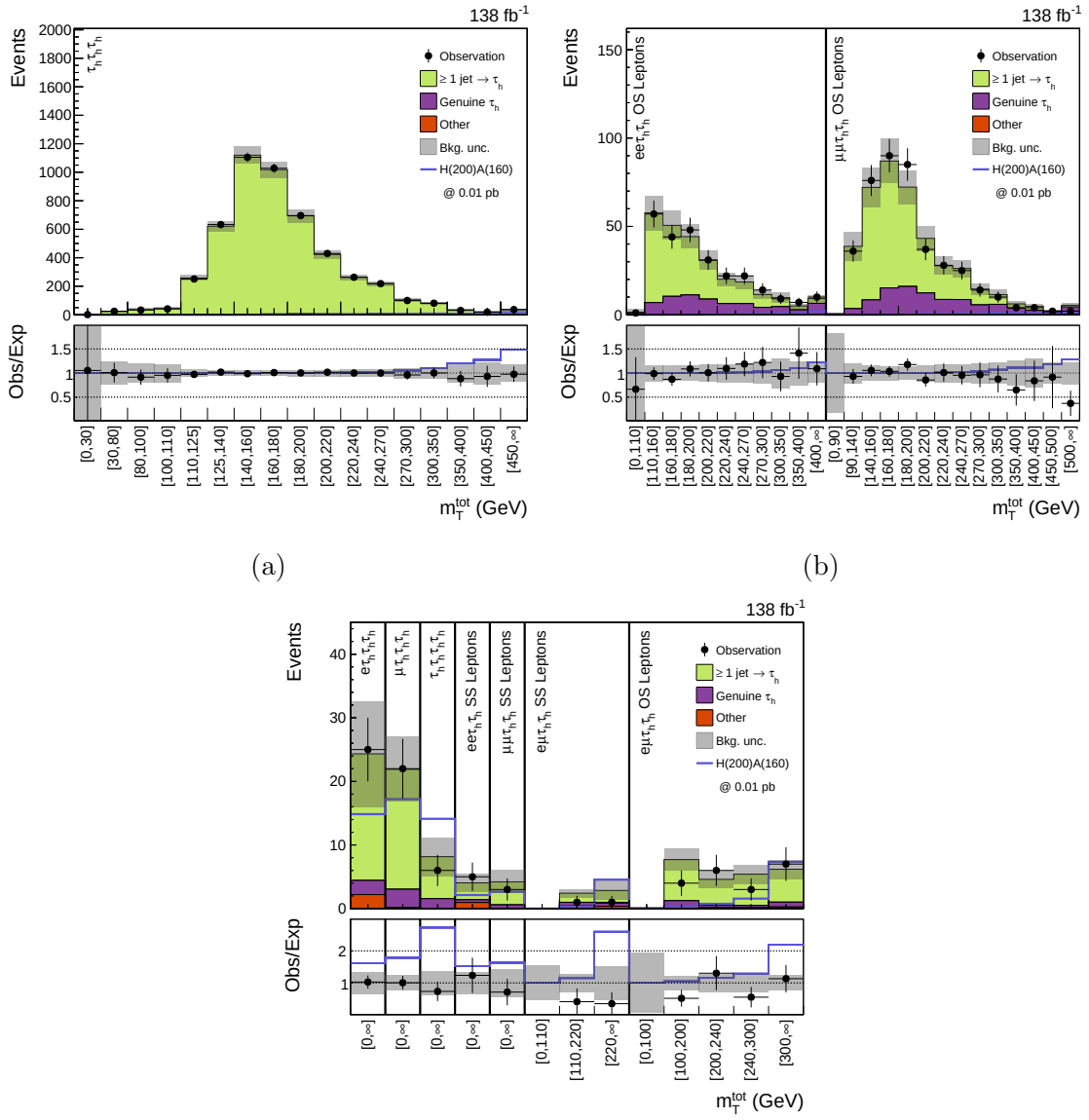


Figure 5.11: Distributions of m_T^{tot} in the high (a), medium (b) and low (c) statistic categories. As the bin sizes vary drastically between channels, the bin sizes are kept constant and the bin intervals are shown on the x-axis. The high statistic categories consist of only the $\tau_h\tau_h\tau_h$ channel, the medium statistic categories include the $ee\tau_h\tau_h$ and $\mu\mu\tau_h\tau_h$ `OS Leptons` categories, and the low statistic categories show the $e\tau_h\tau_h\tau_h$, $\mu\tau_h\tau_h\tau_h$, $\tau_h\tau_h\tau_h\tau_h$, $ee\tau_h\tau_h$ `SS Leptons`, $\mu\mu\tau_h\tau_h$ `SS Leptons`, $e\mu\tau_h\tau_h$ `SS Leptons` and $e\mu\tau_h\tau_h$ `OS Leptons` channels and categories. The solid histograms show the stacked background predictions after a background-only fit to the data. The $m_\phi = 200$ GeV and $m_A = 160$ GeV signal scaled to 0.01 pb is also shown by a blue line for illustrative purposes.

range, are the $\tau_h\tau_h\tau_h\tau_h$, $\mu\tau_h\tau_h\tau_h$ and $e\mu\tau_h\tau_h$. The $\tau_h\tau_h\tau_h$ channel can significantly contribute to the combined limit at higher values of m_ϕ, as the distribution of events in m_T^{tot} peaks higher and so can be more easily separated from the jet $\rightarrow \tau_h$ backgrounds. The remaining channels contribute less to the combined limit. For the $e\tau_h\tau_h\tau_h$ channel, more electrons and jets misidentified as τ_h candidates are present due to worse rejection power for the processes that contribute, and for the $ee\tau_h\tau_h$ and $\mu\mu\tau_h\tau_h$ channels the low branching fractions from four τ leptons and difficult background separation where the light leptons have opposite charge, make it difficult to get a high signal over background acceptance.

5.9.2 Compatibility

Similarly to in Section 4.12.2, the compatibility of the best-fit signal cross-section multiplied by branching fractions in each decay channel is determined, and these are shown in Figure 5.14 for a mass hypothesis of $m_\phi = m_A = 100$ GeV for this search. For this fit the signal strength in each channel is allowed to go negative, although unphysical, to show the data effects in each channel. The best-fit signal strength of the combined fit is between 1 and 2σ below zero. Four categories fit a signal strength slightly below zero, but the combined fit is mostly dominated by the downward fluctuations of the $e\mu\tau_h\tau_h$ channel. The results in each channel are consistent with the zero value within 2σ. These conclusions are consistent across any mass hypotheses, again due to the degenerate nature of the signal shapes in the fitted bins.

5.10 Model-dependent limits

The 95% CL exclusion contours, for the type X 2HDM alignment scenario in the m_A-$\tan\beta$ phase space, are shown in Figure 5.15 for two m_ϕ scenarios of 100 and 200 GeV. These exclusion limits are "top-down" as the cross-sections are unchanged in $\tan\beta$ and the branching fractions to τ leptons are enhanced as $\tan\beta$ increases. The cross-section and branching ratios are calculated as described in Section 5.1. In all cases, as previously observed for the model-independent interpretation, the observed limit lies between the downwards 1 and 2σ expected bands.

The $m_\phi = 100$ GeV scenario has a moderately flat observed $\tan\beta$ limit between 1.2-1.5, across the m_A range. The very slight weakening of the limit at high values

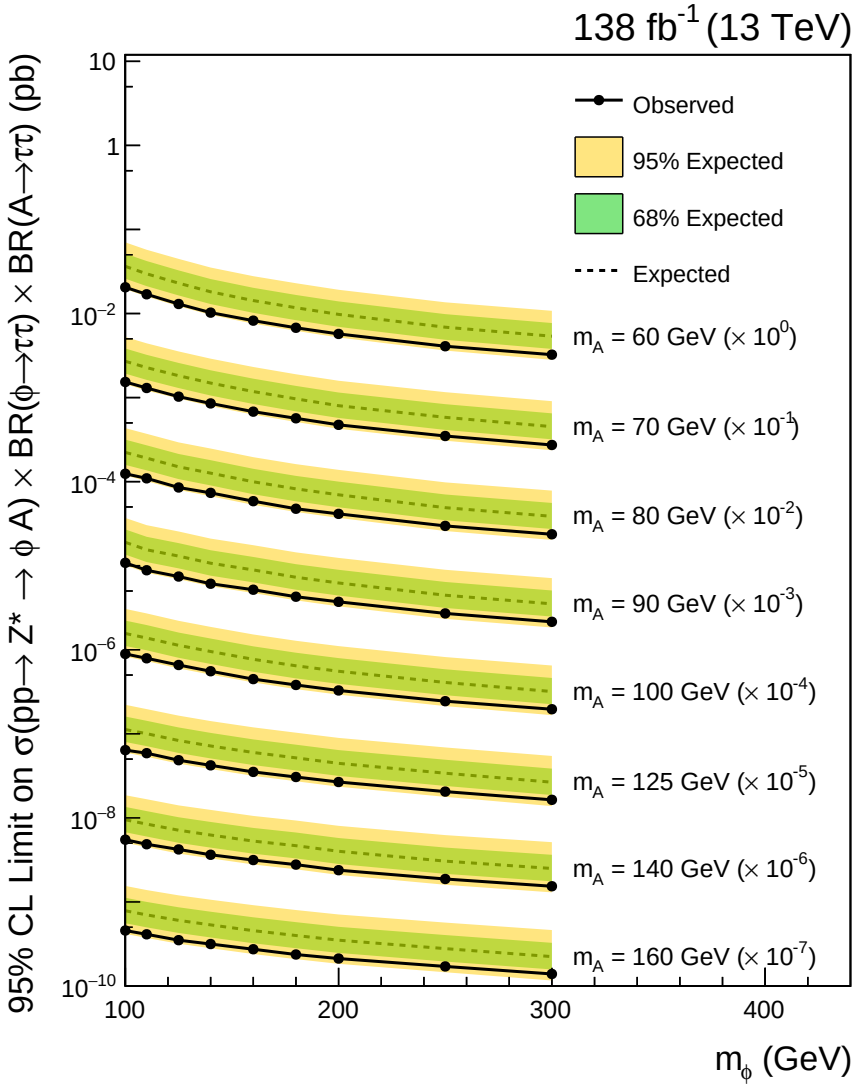


Figure 5.12: Expected (dashed line) and observed (solid line and dots) 95% CL upper limits on the product of the cross-sections and branching fractions for the decay of both additional Higgs bosons into τ leptons. Different m_A hypotheses are scaled by different orders of magnitudes (written on the plot) to make mass points distinguishable. The dark green and bright yellow bands indicate the central 68% and 95% intervals for the expected exclusion limit.

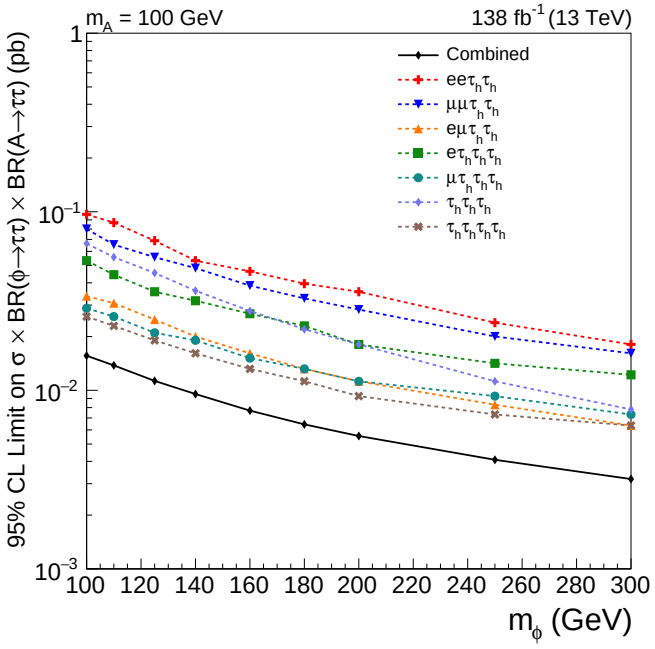


Figure 5.13: Comparison of the expected 95% CL upper limits on the product of the cross-sections and branching fractions for the decay into τ leptons, split by the $\tau\tau\tau\tau$ decay products fit individually.

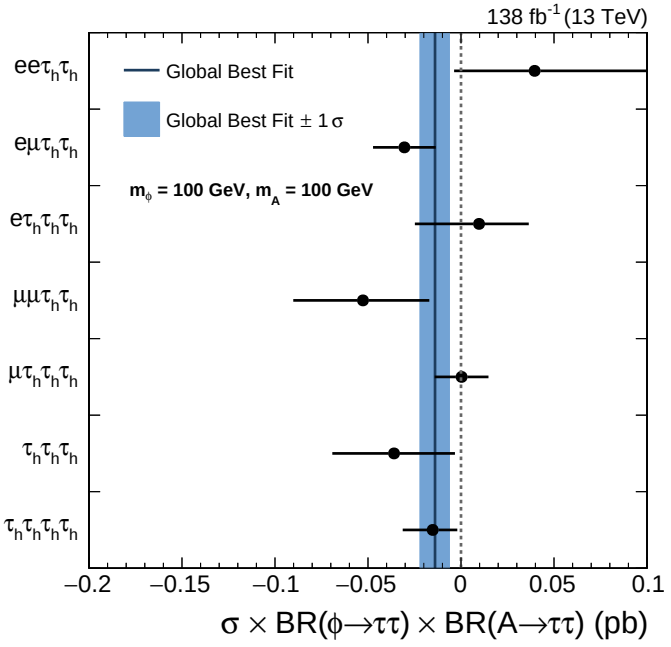
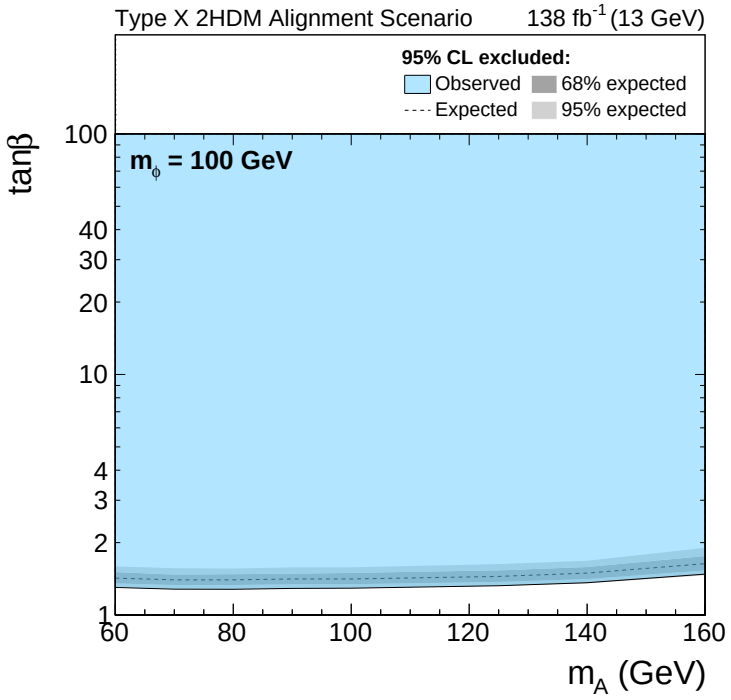


Figure 5.14: Compatibility plots for the $m_A = 100$ GeV and $m_\phi = 100$ GeV mass scenario in analysis decay channels. In each case, the fitted signal strength is decoupled in the bin shown on the plot. The combined best-fit value and its 1σ variation are shown by the blue line and band respectively. The black dashed line indicates a signal strength of zero.

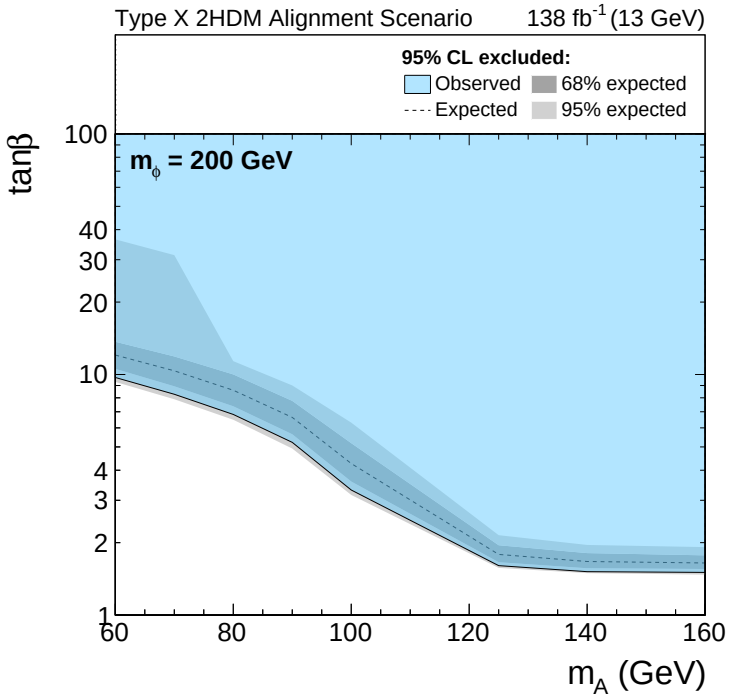
of m_A values is because the $A \to Z$h decay becomes more kinematically feasible and so the $A \to \tau\tau$ branching fraction has to compete with this process. The A $\to Z$h decay will become dominant if m_A is raised much beyond the 160 GeV threshold set in this analysis. The $m_\phi = 200$ GeV scenario's observed limit is weaker at lower values of m_A, with $\tan\beta$ values ranging from 10 to 1.6 between $m_A = 60$ and 125 GeV. This weakening at lower values of m_A happens as the H $\to Z$A decay becomes more kinematically feasible in this region and so competes with H $\to \tau\tau$ for the branching fraction. This was not present in the $m_\phi = 100$ GeV scenario as the ϕ boson is not heavy enough. The limit then flattens for the remainder of the m_A phase space shown, as the H $\to Z$A and A $\to Zh$ decays only minimally hinder the $\tau\tau$ branching fractions of ϕ and A.

Although not shown here, the limits for intermediate ϕ masses see similar trends at low m_A, where the further the ϕ and A masses are separated the weaker the limit, within this mass range. At higher values of m_ϕ than 200 GeV, it becomes difficult to find stable theories across the m_A range. This is because if m_ϕ and m_A are too separated, the values of λ_i as shown in Equation 1.17, can become non-perturbative [42]. Within the m_A allowed regions for this region, the limits follow the same trend as seen at lower values of m_ϕ.

95% CL limits are also set outside of the alignment scenario. This is done by varying relevant alignment parameter, $\cos(\beta-\alpha)$ or $\sin(\beta-\alpha)$ depending on m_ϕ, with respect to $\tan\beta$ for the individual mass hypothesis. Two of these are shown in Figure 5.16 for the mass scenario of $m_\phi = 200$ GeV and $m_A = 100$ GeV, as well as $m_\phi = 200$ GeV and $m_A = 160$. The observed limit lies in the equivalent place compared to the expected as seen in all limit setting for this analysis. The alignment limit at these mass points is equivalent to the limit set at $\cos(\beta - \alpha) = 0$. The shapes of the limits represent the region where the loss of cross-section and branching ratio for ϕ, out of the alignment scenario and at lower $\tan\beta$, is too large that the theory cannot be excluded. The shapes are symmetric in $\cos(\beta - \alpha)$ as no sign dependence is measurable from this process. The widening and narrowing of the limit band in $\tan\beta$ is due to the shape of the branching fractions of ϕ, as shown in Figure 5.5. The widest constraint from this search on $|\cos(\beta - \alpha)|$ is at approximately 0.5 at a $\tan\beta$ around 20.

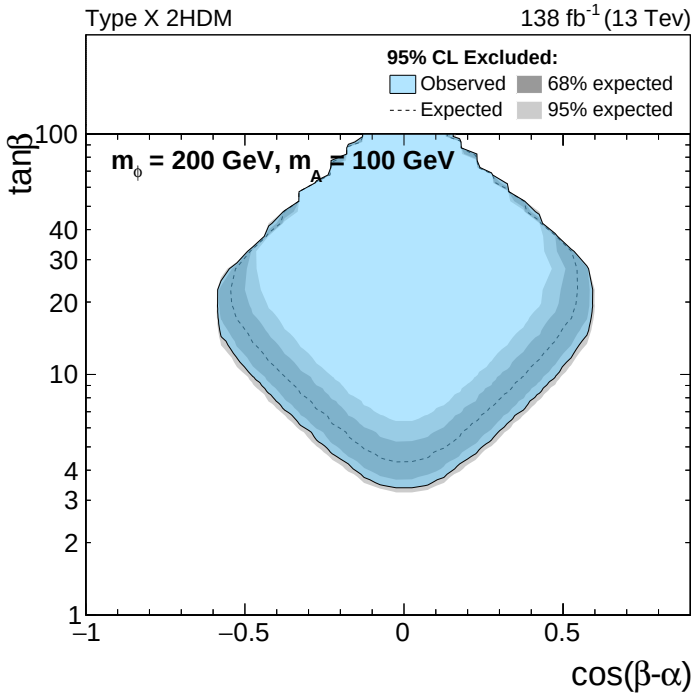


(a)

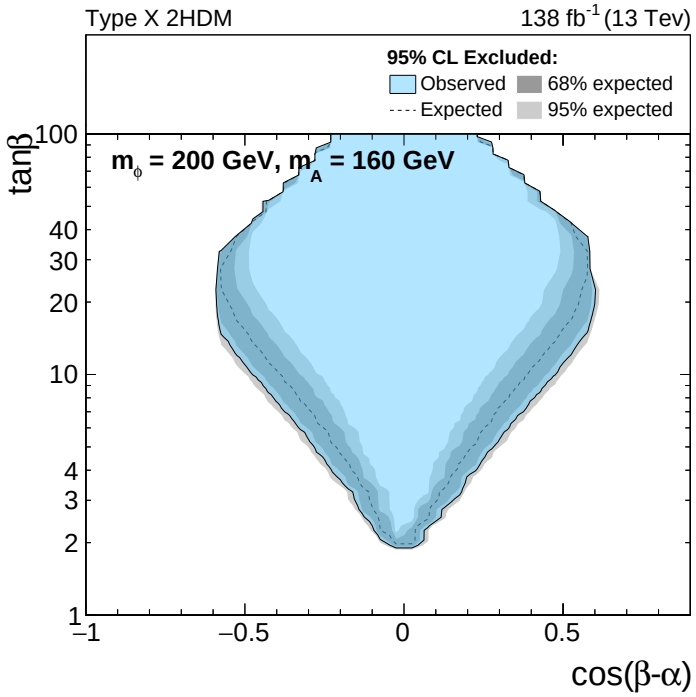


(b)

Figure 5.15: Expected and observed 95% CL exclusion contours on the m_A-tan β phase space in the type X 2HDM alignment scenario for m_ϕ 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.



(a)



(b)

Figure 5.16: 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_ϕ 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

The analyses presented in Chapters 4 and 5, although motivated by different physics, are complementary to one another in the context of the type X 2HDM. The limits on this phase space from the analysis discussed in Chapter 4, are studied using the HiggsTools-1 framework [130]. HiggsTools is a combination of the HiggsBounds [131], HiggsSignals [132] and HiggsPredictions [130] frameworks and these are used for the following purpose:

- HiggsPredictions is used to determine theory production cross-sections and modify decay rates using model parameters.
- HiggsBounds is used to find direct bounds for searches for new particles.
- HiggsSignals is used to find the bounds from shifts to the observed Higgs boson's properties.

HiggsBounds and HiggsSignals contain a database of results from all key measurements from the LHC, LEP and other colliders. The result from Chapter 4 is included in this database.

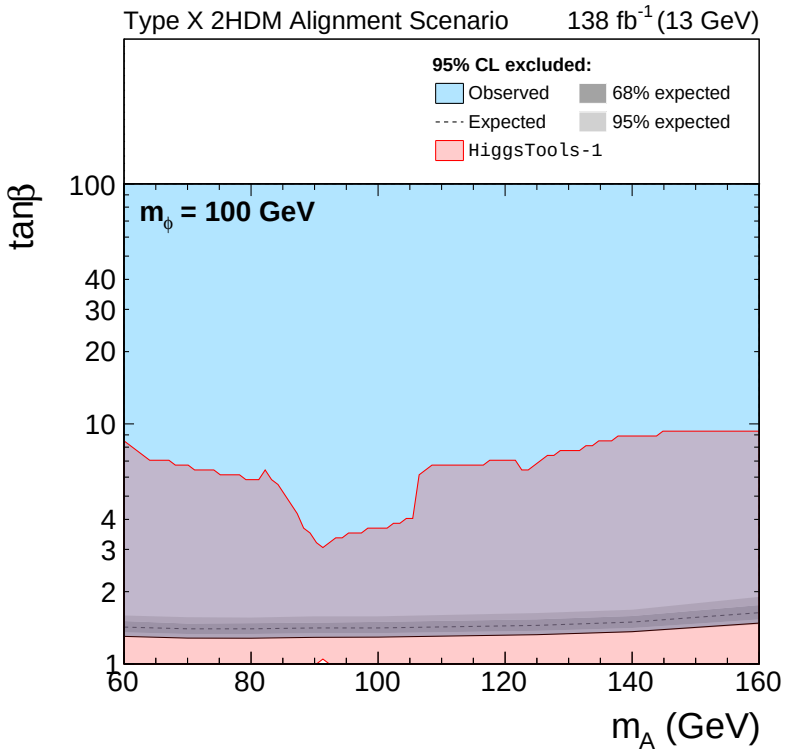
To begin setting constraints on the type X 2HDM, other than that in Chapter 5, the properties of the additional Higgs bosons are required. The widths and branching fractions calculated with 2HDECAY for Chapter 5 are utilised for the scan of the type X 2HDM parameter space. The cross-sections used in each analysis scanned over are scaled to that of the model parameters by the NeutralEffectiveCouplings function [131] implemented in HiggsPredictions. CL_s is calculated at

each point in the parameter space and a 95% CL limit is placed. The limits in the alignment scenario, determined from HiggsBounds, for the m_ϕ equal to 100 and 200 GeV scenarios are overlayed onto the limits shown in Figure 5.15 and shown in Figure 6.1.

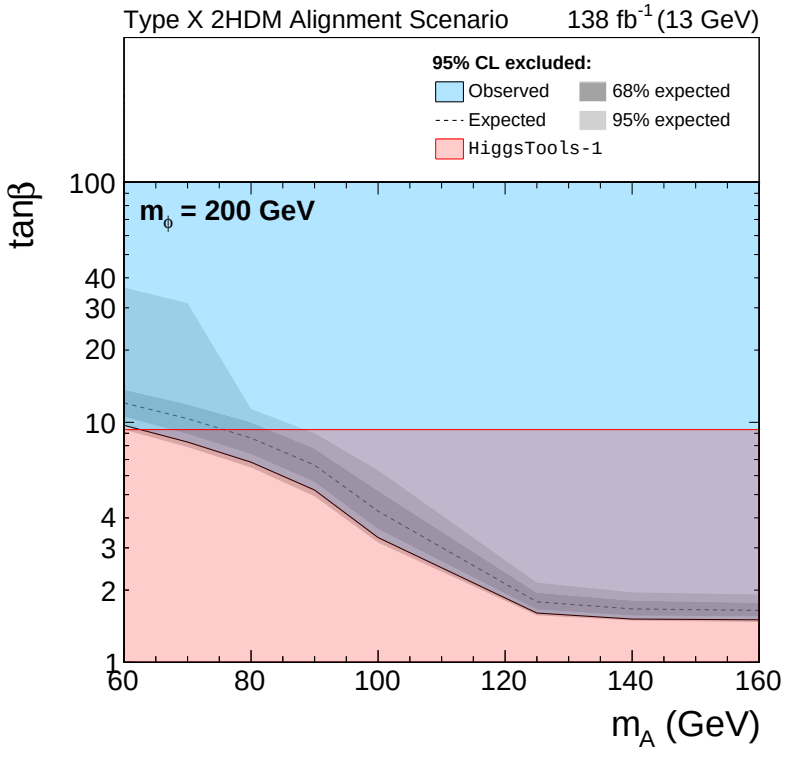
The previous strongest constraints on the type X 2HDM alignment limit parameter space come from the analysis described in Chapter 4. The exclusion limits are at low values of $\tan\beta$ only. As $\tan\beta$ increases, the gluon fusion and b-associated production modes are suppressed but the branching ratios of the additional Higgs bosons to τ leptons are enhanced. Therefore, the exclusion limit represents a compromise between suppressed cross-sections and enhanced branching ratios. The regions where the product is large enough for that parameter point to be excluded happen at low $\tan\beta$. In the type X 2HDM, the b-associated production mode is negligible due to no enhancement of couplings to b quarks. The mass hypotheses change the limit due to the non-flat nature of Figure 4.21a. The limit in Figure 6.1b is flat as the strongest constraint comes from the ϕ(H) boson, and for an additional neutral CP-even Higgs boson $\tan\beta \lesssim 10$ is excluded. However, this is not the case for Figure 6.1a where m_ϕ is lighter and the sensitivity is not always driven by the ϕ(h) boson. In particular, the limit on a 100 GeV resonance, from Figure 4.21, is weakest due to the large background from the Z boson and the local excess observed on top of this peak, so effects from both additional neutral Higgs bosons are present. Together, the exclusion limits from both analyses yield an almost complete coverage of the type X 2HDM alignment scenario within the mass range searched.

Next, the effect of the BSM searches and precision measurements of the SM Higgs boson on the type X 2HDM model, outside of the alignment scenario is studied. Bounds are again calculated with HiggsTools, but the constraints now come from SM Higgs measurements as well as BSM searches such as described in Chapter 4. The bounds determined, overlayed with the results shown in Figure 5.16, are shown in Figure 6.2.

The limits determined from Chapter 4, although stringent across values of $\cos(\beta-\alpha)$, are weaker than the constraints from the SM Higgs boson when moving outside of the alignment limit. They are nonetheless crucial in setting limits, when very close to or on the alignment limit. The constraints from the SM Higgs boson, significantly narrow the region allowed outside of the alignment limit and motivate the continued

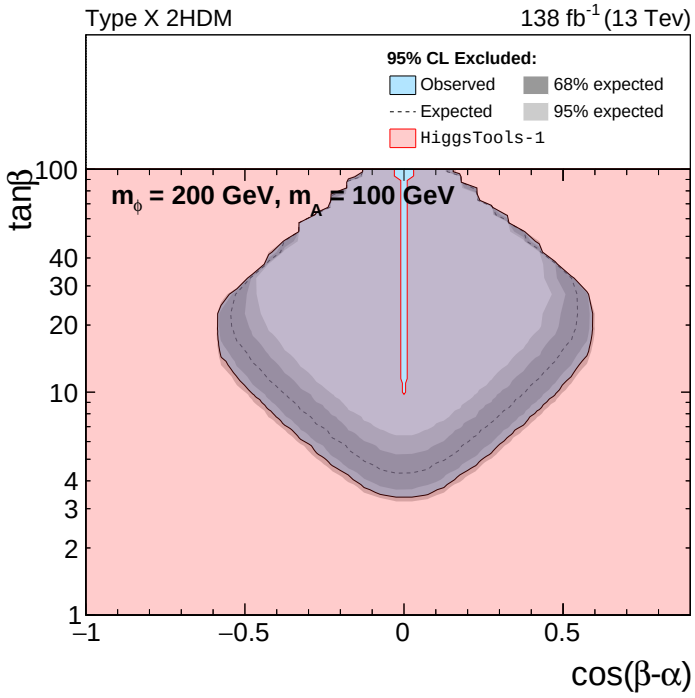


(a)

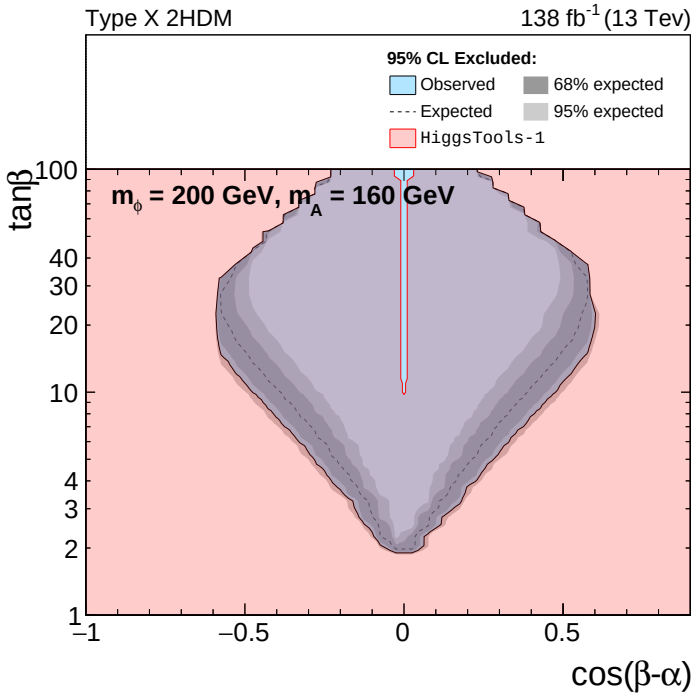


(b)

Figure 6.1: Expected and observed 95% CL exclusion contours on the m_A-tan β phase space in the type X 2HDM alignment scenario for m_ϕ 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. The limit obtained by HiggsTools is shown in the red contour.



(a)



(b)

Figure 6.2: 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_ϕ 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 are shown by the blue area. The limit obtained by HiggsTools is shown in the red contour.

use of alignment scenarios for extended Higgs sector searches. The combination of both BSM and SM Higgs boson results, prior to the work performed in Chapter 5, leaves only a small strip of the phase space for new physics to exist. However, the entirety of the phase space for the two mass points is excluded by the combination of the searches detailed in Chapters 4 and 5, as well as the precision measurement of the SM Higgs boson.

6.2 Summary

This thesis has presented two analyses utilising the full Run 2 dataset collected by the CMS experiment, from the 13 TeV proton-proton collisions at the LHC, targeting final states enriched in τ leptons. The motivation for the searches presented in Chapters 4 and 5 are BSM theories that attempt to resolve the theoretical issue of the hierarchy problem and the experiment tensions of the B anomalies and the muon g-2 anomaly. The preceding chapters, act to motivate the new physics models that could potentially resolve these issues and the apparatus and methods used for the foundation of data taking and reconstruction required for the searches.

Chapter 4 presents a search for two possible areas of new physics in the di-τ final states. The first of these is a search for additional neutral Higgs bosons, motivated by the type II 2HDM of the MSSM, as a consequence of a solution to the hierarchy problem. This targets two production modes: gluon fusion and production in association with a b quark, with the latter being dominant to a search for the MSSM at higher values of $\tan\beta$. No deviation is observed for the search for b-associated production, which targets event categories that require a minimum of one b jet. This makes it very difficult to coincide an MSSM benchmark scenario with any gluon fusion signal. The gluon fusion results yielded two small deviations from the SM expectation, peaking at 100 GeV and 1.2 TeV with a local (global) statistical significance of 3.1σ (2.7σ) and 2.8σ (2.2σ) respectively. The two excesses are present in the no b tag categories fit and are compatible across all decay channels and categories fit. Good agreement between data and the background hypothesis is observed in the rest of the mass hypotheses. Limits are set on the cross-section of both production modes multiplied by the branching fraction of the resonance's decay to τ leptons and both of these vary from $\mathcal{O}(10\text{ pb})$ at 60 GeV to 0.3 fb at 3.5 TeV. These results are also interpreted as exclusion limits in MSSM benchmark scenarios. In the M_h^{125} scenario, values of $m_A < 500$ GeV are excluded and in the remaining phase space as

well as in the $M_{h,EFT}^{125}$ the strongest constraints on the phase spaces are set.

The second area of new physics that is searched for in Chapter 4, is a potential solution to the B anomalies, in vector leptoquarks. The signal model searched for is a non-resonant t-channel interaction producing a di-τ final state, where the initial state is dominated by b quarks. The best-fit vector leptoquark is heavily constrained by the b tag categories where no deviation from the SM expectation is observed. It therefore cannot be used to explain the deviations observed in the no b tag categories. Limits on vector leptoquark phase space are set and constrain the regions allowing for an explanation of the B anomalies.

Chapter 5 details a search for an extended Higgs sector to explain the muon g-2 anomaly. This can be done with type X 2HDM at high values of $\tan\beta$. A different signal strategy is needed than in Chapter 4, due to the suppressed coupling between the additional neutral Higgs bosons and quarks in this model. The preferred signal model here is the production of two additional neutral Higgs bosons via an off-shell Z boson. At high values of $\tan\beta$, the branching fractions of the additional neutral Higgs bosons are dominated by di-τ pairs. Therefore, the $\tau\tau\tau\tau$ final states are used to reconstruct this signal. No significant deviation is observed from the background estimation. Limits on the cross-section multiplied by branching fractions are set and vary from 20 fb at the lowest mass hypothesis, to 1.4 fb at the highest mass hypothesis. The results are interpreted in terms of the type X 2HDM, and it is found that the constraints from the SM Higgs boson measurements limit the phase space to be very close to the alignment limit. The alignment limit exclusions are found to exclude upwards of $\tan\beta$ approximately equal to 1.5 unless the $H \to ZA$ becomes kinematically feasible and the limit is weakened. This is an exclusion way beyond a possible explanation of the muon g-2 anomaly with a type X 2HDM. The results from Chapter 4 are complimentary to Chapter 5 in the type X 2HDM alignment limit, as they exclude downwards of $\tan\beta$ approximately equal to 10, and so together exclude the vast majority of the phase space in mass ranges searched.

Bibliography

[1] 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* **09** (2018) 007 [1803.06553].

[2] CMS Collaboration, *Searches for additional Higgs bosons and vector leptoquarks in $\tau\tau$ final states in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *CMS Physics Analysis Summary* **CMS-PAS-HIG-21-001** (2022) .

[3] 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* **716** (2012) 1 [1207.7214].

[4] CMS Collaboration, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30 [1207.7235].

[5] LHCb Collaboration, *Test of lepton universality in beauty-quark decays*, *Nature Phys.* **18** (2022) 277 [2103.11769].

[6] BaBar Collaboration, *Evidence for an excess of $B \to D^{(*)}\tau\nu$ decays*, *J. Phys. Conf. Ser.* **455** (2013) 012039.

[7] BaBar Collaboration, *Measurement of an excess of $\bar{\mathrm{B}} \to \mathrm{D}^{(*)}\tau^-\bar{\nu}_\tau$ decays and implications for charged higgs bosons*, *Phys. Rev. D* **88** (2013) 072012 [1303.0571].

[8] Belle Collaboration, *Measurement of the branching ratio of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle*, *Phys. Rev. D* **92** (2015) 072014 [1507.03233].

[9] LHCb Collaboration, *Measurement of the ratio of branching fractions $\mathcal{B}(\bar{\mathrm{B}}^0 \to \mathrm{D}^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{\mathrm{B}}^0 \to \mathrm{D}^{*+}\mu^-\bar{\nu}_\mu)$*, *Phys. Rev. Lett.* **115** (2015) 111803 [1506.08614].

[10] Belle Collaboration, *Measurement of the τ lepton polarization and $R(\mathrm{D}^*)$ in the decay $\bar{\mathrm{B}} \to \mathrm{D}^* \tau^- \bar{\nu}_\tau$*, *Phys. Rev. Lett.* **118** (2017) 211801 [1612.00529].

[11] LHCb Collaboration, *Test of Lepton Flavor Universality by the measurement of the $B^0 \to D^{*-} \tau^+ \nu_\tau$ branching fraction using three-prong τ decays*, *Phys. Rev. D* **97** (2018) 072013 [1711.02505].

[12] LHCb Collaboration, *Measurement of the ratio of the $\mathrm{B}^0 \to \mathrm{D}^{*-} \tau^+ \nu_\tau$ and $\mathrm{B}^0 \to \mathrm{D}^{*-} \mu^+ \nu_\mu$ branching fractions using three-prong τ-lepton decays*, *Phys. Rev. Lett.* **120** (2018) 171802 [1708.08856].

[13] Muon g-2 Collaboration, *Final Report of the Muon E821 Anomalous Magnetic Moment Measurement at BNL*, *Phys. Rev. D* **73** (2006) 072003 [hep-ex/0602035].

[14] Muon g-2 Collaboration, *Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm*, *Phys. Rev. Lett.* **126** (2021) 141801 [2104.03281].

[15] *Measurement of lepton universality parameters in $B^+ \to K^+ \ell^+ \ell^-$ and $B^0 \to K^{*0} \ell^+ \ell^-$ decays*, 2212.09153.

[16] R. L. Workman et al., *Review of Particle Physics*, *PTEP* **2022** (2022) 083C01.

[17] D. Galbraith and C. Burgard, "Example: Standard model of physics." https://texample.net/tikz/examples/model-physics/. [Accessed: 25/05/23].

[18] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Nonabelian Gauge Theories*, *Phys. Rev. Lett.* **30** (1973) 1343.

[19] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, *Phys. Rev. Lett.* **30** (1973) 1346.

[20] S. L. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl. Phys.* **22** (1961) 579.

[21] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967) 1264.

[22] A. Salam, *Weak and Electromagnetic Interactions*, *Conf. Proc. C* **680519** (1968) 367.

[23] T. D. Lee and C.-N. Yang, *Question of Parity Conservation in Weak Interactions*, *Phys. Rev.* **104** (1956) 254.

[24] G. Arnison et al., *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at* $\sqrt{s} = 540$ *GeV*, *Phys. Lett. B* **122** (1983) 103.

[25] M. Banner et al., *Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider*, *Phys. Lett. B* **122** (1983) 476.

[26] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, *Phys. Rev. Lett.* (1964) 321.

[27] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, *Phys. Lett.* (1964) 132.

[28] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, *Phys. Rev. Lett.* (1964) 508.

[29] P. W. Higgs, *Spontaneous symmetry breakdown without massless bosons*, *Phys. Rev.* (1966) 1156.

[30] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global conservation laws and massless particles*, *Phys. Rev. Lett.* (1964) 585.

[31] T. W. B. Kibble, *Symmetry breaking in non-Abelian gauge theories*, *Phys. Rev.* (1967) 1554.

[32] J. Goldstone, *Field Theories with Superconductor Solutions*, *Nuovo Cim.* **19** (1961) 154.

[33] S. L. Glashow and S. Weinberg, *Natural conservation laws for neutral currents*, *Phys. Rev. D* **15** (1977) 1958.

[34] I. F. Ginzburg and M. Krawczyk, *Symmetries of two Higgs doublet model and CP violation*, *Phys. Rev. D* **72** (2005) 115013 [hep-ph/0408011].

[35] CMS Collaboration, *A measurement of the Higgs boson mass in the diphoton decay channel*, *Phys. Lett. B* **805** (2020) 135425 [2002.06398].

[36] S. P. Martin, *A Supersymmetry Primer*, *Advanced Series on Directions in High Energy Physics* (1998) 1–98.

[37] B. Diaz, M. Schmaltz and Y.-M. Zhong, *The leptoquark hunter's guide: Pair production*, *JHEP* **10** (2017) 097 [1706.05033].

[38] M. Schmaltz and Y.-M. Zhong, *The leptoquark hunter's guide: Large coupling*, *JHEP* **01** (2019) 132 [1810.10017].

[39] C. Cornella, D. A. Faroughy, J. Fuentes-Martin, G. Isidori and M. Neubert, *Reading the footprints of the B-meson flavor anomalies*, *JHEP* **08** (2021) 050 [2103.16558].

[40] V. Ilisie, *New Barr-Zee contributions to* $(\mathbf{g-2})_\mu$ *in two-Higgs-doublet models*, *JHEP* **04** (2015) 077 [1502.04199].

[41] S. M. Barr and A. Zee, *Electric Dipole Moment of the Electron and of the Neutron*, *Phys. Rev. Lett.* **65** (1990) 21.

[42] 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* **104** (2021) 095008 [2104.10175].

[43] L. Evans and P. Bryant (Eds.), *LHC Machine*, *JINST* **3** (2008) S08001.

[44] LEP Injector Study Group, *LEP Design Report: Vol.2. The LEP Main Ring*, *Technical Design Report* **CERN-LEP-84-01** (1984) .

[45] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *JINST* **3** (2008) S08002.

[46] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003.

[47] CMS Collaboration, *The CMS Experiment at the CERN LHC*, *JINST* **3** (2008) S08004.

[48] LHCb Collaboration, *The LHCb Detector at the LHC*, *JINST* **3** (2008) S08005.

[49] H. Bartosik and G. Rumolo, *Performance of the LHC injector chain after the upgrade and potential development*, 2203.09202.

[50] CMS Collaboration, "CMS Luminosity public results." https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. [Accessed: 25/05/23].

[51] CMS Collaboration, *Operation and performance of the CMS tracker*, *PoS* **TIPP2014** (2014) 347.

[52] CMS Collaboration, *CMS Technical Design Report for the Pixel Detector Upgrade*, *Technical Design Report* **CERN-LHCC-2012-016, CMS-TDR-011** (2012) .

[53] CMS Collaboration, *Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at* $\sqrt{s} = 7$ *TeV*, *JINST* **8** (2013) P09009 [1306.2016].

[54] CMS Collaboration, *CMS Technical Design Report for the Level-1 Trigger Upgrade*, *CMS Technical Design Report* **CERN-LHCC-2013-011, CMS-TDR-12, CMS-TDR-012** (2013) .

[55] CMS Collaboration, *Description and performance of track and primary-vertex reconstruction with the CMS tracker*, *JINST* **9** (2014) P10009 [1405.6569].

[56] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, *Nucl. Instrum. Meth. A* **262** (1987) 444.

[57] CMS Collaboration, *Measurement of Tracking Efficiency*, *CMS Physics Analysis Summary* **CMS-PAS-TRK-10-002** (2010) .

[58] K. Rose, *Deterministic annealing for clustering, compression, classification, regression, and related optimization problems*, *IEEE Proc.* **86** (1998) 2210.

[59] R. Fruhwirth, W. Waltenberger and P. Vanlaer, *Adaptive vertex fitting*, *J. Phys. G* **34** (2007) N343.

[60] CMS Collaboration, *Particle-flow reconstruction and global event description with the CMS detector*, *JINST* **12** (2017) P10003 [1706.04965].

[61] CMS Collaboration, *Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector*, *CMS Physics Analysis Summary* **CMS-PAS-PFT-10-001, CMS-PAS-PFT-10-001** (2010) .

[62] CMS Collaboration, *Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV*, *CMS Physics*

Analysis Summary **CMS-PAS-PFT-10-002, CMS-PAS-PFT-10-002** (2010) .

[63] T. Speer, W. Adam, R. Fruhwirth, A. Strandlie, T. Todorov and M. Winkler, *Track reconstruction in the CMS tracker*, *Nucl. Instrum. Meth. A* **559** (2006) 143.

[64] CMS Collaboration, *Performance of CMS Muon Reconstruction in pp Collision Events at* $\sqrt{s} = 7$ *TeV*, *JINST* **7** (2012) P10002 [1206.4071].

[65] CMS Collaboration, *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at* $\sqrt{s} = 13$ *TeV*, *JINST* **13** (2018) P06015 [1804.04528].

[66] CMS Collaboration, *Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at* $\sqrt{s} = 8$ *TeV*, *JINST* **10** (2015) P06005 [1502.02701].

[67] CMS Collaboration, *Performance of CMS Muon Reconstruction in Cosmic-Ray Events*, *JINST* **5** (2010) T03022 [0911.4994].

[68] G. P. Salam, *Towards Jetography*, *Eur. Phys. J. C* **67** (2010) 637 [0906.1833].

[69] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, *JHEP* **04** (2008) 063 [0802.1189].

[70] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, *Eur. Phys. J. C* **72** (2012) 1896 [1111.6097].

[71] CMS Collaboration, *Jet algorithms performance in 13 TeV data*, *CMS Physics Analysis Summary* **CMS-PAS-JME-16-003** (2017) .

[72] CMS Collaboration, *Pileup Jet Identification*, *CMS Physics Analysis Summary* **CMS-PAS-JME-13-005** (2013) .

[73] CMS Collaboration, *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*, *JINST* **13** (2018) P05011 [1712.07158].

[74] E. Bols, J. Kieseler, M. Verzetti, M. Stoye and A. Stakia, *Jet flavour classification using DeepJet*, *JINST* **15** (2020) P12012 [2008.10519].

[75] CMS Collaboration, "CMS Public Twiki." https://twiki.cern.ch/twiki/bin/view/CMSPublic/BTV13TeV2017DeepJet. [Accessed: 25/05/23].

[76] CMS Collaboration, *Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector*, *CMS Physics Analysis Summary* **CMS-PAS-JME-16-004** (2016) .

[77] CMS Collaboration, *Pileup mitigation at CMS in 13 TeV data*, *JINST* **15** (2020) P09018 [2003.00503].

[78] CMS Collaboration, *Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV*, *JINST* **13** (2018) P10005 [1809.02816].

[79] CMS Collaboration, *Identification of hadronic tau lepton decays using a deep neural network*, *JINST* **17** (2022) P07023 [2201.08458].

[80] CMS Collaboration, *Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV*, *JINST* **13** (2018) P10005 [1809.02816].

[81] E. Bagnaschi, H. Bahl, E. Fuchs, T. Hahn, S. Heinemeyer, S. Liebler et al., *MSSM Higgs boson searches at the LHC: Benchmark scenarios for Run 2 and beyond*, *Eur. Phys. J. C* **79** (2019) 617 [1808.07542].

[82] 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* **80** (2020) 916 [2005.14536].

[83] 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* **79** (2019) 279 [1901.05933].

[84] E. A. Bagnaschi, S. Heinemeyer, S. Liebler, P. Slavich and M. Spira, *Benchmark scenarios for MSSM Higgs boson searches at the LHC*, *LHC Higgs Cross Section Working Group Public Note* **LHCHWG-2021-001** (2021) .

[85] CMS Collaboration, *Search for beyond the standard model Higgs bosons decaying into a* $\mathrm{b\bar{b}}$ *pair in pp collisions at* $\sqrt{s} = 13$ *TeV*, *JHEP* **08** (2018) 113 [1805.12191].

[86] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040 [hep-ph/0409146].

[87] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070 [0709.2092].

[88] 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* **06** (2010) 043 [1002.2581].

[89] T. Ježo and P. Nason, *On the treatment of resonances in next-to-leading order calculations matched to a parton shower*, *JHEP* **12** (2015) 065 [1509.09071].

[90] R. D. Ball et al., *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040 [1410.8849].

[91] R. D. Ball et al., *Parton distributions from high-precision collider data*, *Eur. Phys. J. C* **77** (2017) 663 [1706.00428].

[92] CMS Collaboration, *Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements*, *Eur. Phys. J. C* **80** (2020) 4 [1903.12179].

[93] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten et al., *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159 [1410.3012].

[94] S. Agostinelli et al., GEANT4—*a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250.

[95] CMS Collaboration, *The search for a third-generation leptoquark coupling to a* τ *lepton and a b quark through single, pair and nonresonant production at* $\sqrt{s} = 13$ TeV, *CMS Physics Analysis Summary* **CMS-PAS-EXO-19-016** (2022) .

[96] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, *MadGraph 5: Going beyond*, *JHEP* **06** (2011) 128 [1106.0522].

[97] R. Frederix and S. Frixione, *Merging meets matching in MC@NLO*, *JHEP* **12** (2012) 061 [1209.6215].

[98] L. Bianchini, J. Conway, E. K. Friis and C. Veelken, *Reconstruction of the Higgs mass in $H \to \tau\tau$ events by dynamical likelihood techniques*, *J. Phys. Conf. Ser.* **513** (2014) 022035.

[99] CMS Collaboration, *Measurements of Higgs boson production in the decay channel with a pair of τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *CMS Physics Analysis Summary* **CMS-HIG-19-010, CERN-EP-2022-027** (2022) [2204.12957].

[100] CMS Collaboration, *An embedding technique to determine $\tau\tau$ backgrounds in proton-proton collision data*, *JINST* **14** (2019) P06032 [1903.01216].

[101] 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* **09** (2018) 007 [1803.06553].

[102] CMS Collaboration, *Measurement of the $Z/\gamma^* \to \tau\tau$ cross section in pp collisions at $\sqrt{s} = 13$ TeV and validation of τ lepton analysis techniques*, *Eur. Phys. J. C* **78** (2018) 708 [1801.03535].

[103] J. Alwall, S. Höche, F. Krauss, N. Lavesson, L. Lönnblad, F. Maltoni et al., *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions*, *Eur. Phys. J. C* **53** (2008) 473 [0706.2569].

[104] S. Alioli, S.-O. Moch and P. Uwer, *Hadronic top-quark pair-production with one jet and parton showering*, *JHEP* **01** (2012) 137 [1110.5251].

[105] R. Frederix, E. Re and P. Torrielli, *Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO*, *JHEP* **09** (2012) 130 [1207.5391].

[106] E. Re, *Single-top Wt-channel production matched with parton showers using the POWHEG method*, *Eur. Phys. J. C* **71** (2011) 1547 [1009.2450].

[107] K. Melnikov and F. Petriello, *Electroweak gauge boson production at hadron colliders through $\mathcal{O}(\alpha_s^2)$*, *Phys. Rev. D* **74** (2006) 114017 [hep-ph/0609070].

[108] M. Czakon and A. Mitov, *Top++: A program for the calculation of the top-pair cross-section at hadron colliders*, *Comput. Phys. Commun.* **185** (2014) 2930 [1112.5675].

[109] N. Kidonakis, *Top quark production*, in *Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons*, p. 139, 2014, 1311.0283, DOI.

[110] J. M. Campbell, R. K. Ellis and C. Williams, *Vector boson pair production at the LHC*, *JHEP* **07** (2011) 018 [1105.0020].

[111] T. Gehrmann, M. Grazzini, S. Kallweit, P. Maierhöfer, A. von Manteuffel, S. Pozzorini et al., *W^+W^- production at hadron colliders in next to next to leading order QCD*, *Phys. Rev. Lett.* **113** (2014) 212001 [1408.5243].

[112] CMS Collaboration, *Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **01** (2011) 080 [1012.2466].

[113] CMS Collaboration, *Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JINST* **15** (2020) P10017 [2006.10165].

[114] CMS Collaboration, *Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV*, *Eur. Phys. J. C* **75** (2015) 542 [1505.04480].

[115] R. J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, *Comput. Phys. Commun.* **77** (1993) 219.

[116] J. S. Conway, *Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra*, in *PHYSTAT 2011*, pp. 115–120, 2011, 1103.0354, DOI.

[117] CMS Collaboration, *Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector*, *JINST* **14** (2019) P07004 [1903.06078].

[118] T. Junk, *Confidence level computation for combining searches with small statistics*, *Nucl. Instrum. Meth. A* **434** (1999) 435 [hep-ex/9902006].

[119] A. L. Read, *Presentation of search results: The CL_s technique*, *J. Phys. G* **28** (2002) 2693.

[120] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** (2011) 1554 [1007.1727].

[121] ATLAS Collaboration, *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.* **125** (2020) 051801 [2002.12223].

[122] LHC Higgs Cross Section Working Group Collaboration, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, 1307.1347.

[123] LHC Higgs Cross Section Working Group Collaboration, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, 1610.07922.

[124] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi and M. Spira, *Standard model Higgs-boson branching ratios with uncertainties*, *Eur. Phys. J. C* **71** (2011) 1753 [1107.5909].

[125] 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.* **246** (2020) 106852 [1810.00768].

[126] S. Catani and M. Grazzini, *Next-to-next-to-leading-order subtraction formalism in hadron collisions and its application to higgs-boson production at the large hadron collider*, *Phys. Rev. Lett.* **98** (2007) 222002.

[127] A. Rogozhnikov, *Reweighting with Boosted Decision Trees*, *J. Phys. Conf. Ser.* **762** (2016) 012036 [1608.05806].

[128] CMS Collaboration, *Analysis of the CP structure of the Yukawa coupling between the Higgs boson and τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **06** (2022) 012 [2110.04836].

[129] F. J. Massey, *The kolmogorov-smirnov test for goodness of fit*, *Journal of the American Statistical Association* **46** (1951) 68.

[130] H. Bahl, T. Biekötter, S. Heinemeyer, C. Li, S. Paasch, G. Weiglein et al., *HiggsTools: BSM scalar phenomenology with new versions of HiggsBounds and HiggsSignals*, 2210.09332.

[131] P. Bechtle, D. Dercks, S. Heinemeyer, T. Klingl, T. Stefaniak, G. Weiglein et al., *HiggsBounds-5: Testing Higgs Sectors in the LHC 13 TeV Era*, *Eur. Phys. J. C* **80** (2020) 1211 [2006.06007].

[132] P. Bechtle, S. Heinemeyer, T. Klingl, T. Stefaniak, G. Weiglein and J. Wittbrodt, *HiggsSignals-2: Probing new physics with precision Higgs measurements in the LHC 13 TeV era*, *Eur. Phys. J. C* **81** (2021) 145 [2012.09197].

List of acronyms

2HDM two Higgs doublet model

BDT boosted decision tree

BEH Brout-Englet-Higgs

BSM Beyond Standard Model

CERN the European Organization for Nuclear Research

CHS charged hadron subtraction

CL confidence level

CMS Compact Muon Solenoid

CP charge-parity

CSC cathode strip chambers

CTF Combinatorial Track Finder

DM decay mode

DNN deep neural network

DT drift tube

EB ECAL barrel

ECAL electromagnetic calorimeter

EE ECAL endcap

FS flavour scheme

GSF Gaussian sum filter

HB hadron barrel

HCAL hadronic calorimeter

HE hadron endcap

HF hadron forward

HLT high-level trigger

HO hadron outer

HPS hadron-plus-strips

KF Kalman filter

L1 Level-1

LEP Large Electron-Positron

LHC Large Hadron Collider

LO leading-order

LS1 Long Shutdown 1

LS2 Long Shutdown 2

MC monte carlo

MET missing transverse energy

ML machine learning

MSSM Minimal Supersymmetric Standard Model

MVA multivariate analysis

NLO next-to-leading-order

NN neural network

NNLO next-to-next-to-leading-order

PDF probability distribution function

PF particle flow

PS Proton Synchrotron

PSB Proton Synchrotron Booster

PU pileup

PV primary vertex

QCD Quantum Chromodynamics

RF radio frequency

RPC resistive plate chambers

SC supercluster

SM Standard Model

SPS Super Proton Synchrotron

SUSY supersymmetry

SV secondary vertex

TEC tracker endcaps

TIB tracker inner barrel

TID tracker inner disks

TOB tracker outer barrel

VEV vacuum expectation value

WLCG Worldwide LHC Computing Grid

WP working points