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

Declaration

I did this work I promise

Acknowledgements

Emmeline

Contents

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-Englert-Higgs (BEH) mechanism for spontaneous symmetry breaking, that allows many fundamental particles to obtain masses. It also predicts a new scalar boson, named the Higgs boson. In 2012, the ATLAS [6] and CMS collaboration [7] discovered a Higgs boson-like particle when colliding protons at high energy at the Large Hadron Collider (LHC) and further measurements of this particle's properties have been consistent with such a particle. This discovery experimentally completed the SM particle constituents.

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

1.1 The Standard Model of Particle Physics

1.1.1 Fundamental Particles and Interactions

The SM is a set of fundamental particles, as shown in Figure 1.1, and rules that govern the interactions between particles. The interactions between these particles are able to model the strong, weak and electromagnetic force, unifying the later two into one electroweak interaction. The SM consists of 6 quarks, 3 charged leptons and 3 neutrinos, which are named fermions. Each of these particles contain an anti-partner with opposite quantum numbers but the same mass.

The SM is a renormalisable quantum field theory that is built on the principle of local gauge invariance. The SU(3) ⊗ SU(2) ⊗ U(1) is the gauge symmetry group of the SM. This means that the Lagrangian, that governs the particles interactions, is invariant under such a transformation.

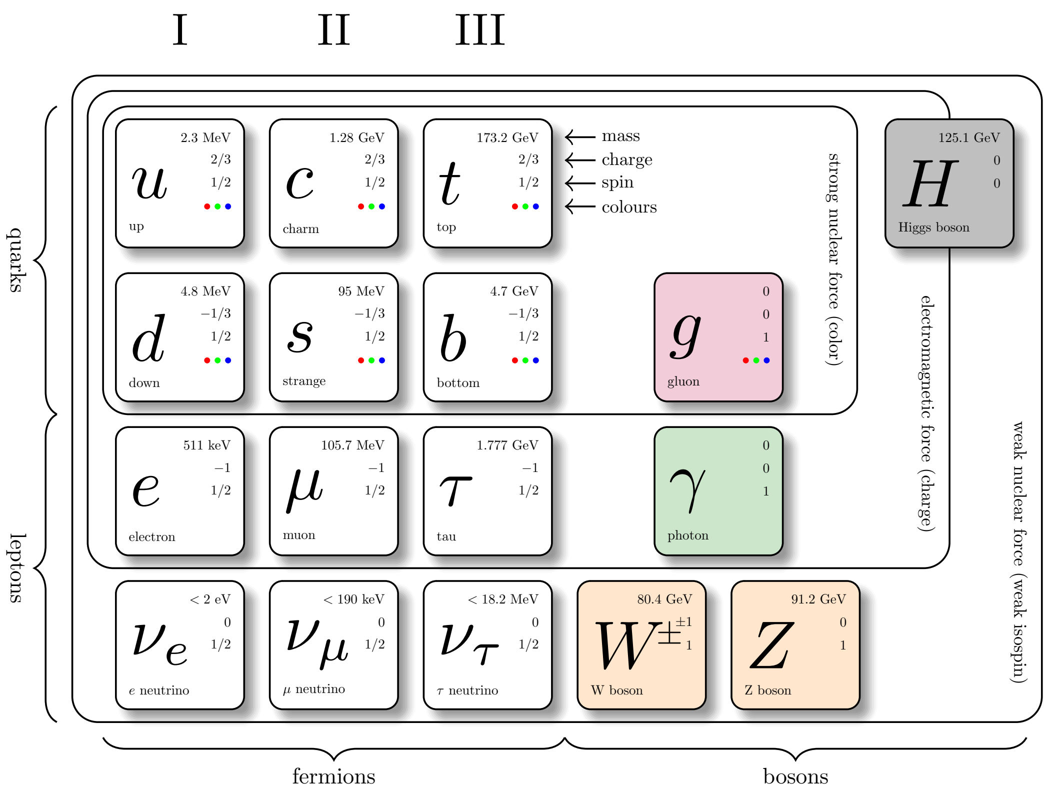


Figure 1.1

1.1.2 Higgs Sector

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{1.1}$$

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - V(\phi) \tag{1.2}$$

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \tag{1.3}$$

Has minima at

$$\phi^\dagger \phi = -\frac{\mu^2}{2\lambda} \tag{1.4}$$

Need massless photon

$$\langle 0|\phi|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu \end{pmatrix} \tag{1.5}$$

Break symmetry

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + h(x) \end{pmatrix} \tag{1.6}$$

To make the mass term gauge invariant

$$\partial_\mu \to D_\mu = \partial_\mu + \frac{i g_W}{2} \vec{\sigma} \cdot \vec{W}_\mu + \frac{i g^{'} Y}{2} B_\mu \tag{1.7}$$

$$D_\mu \phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + h(x) \end{pmatrix} \tag{1.8}$$

1.2 Extended Higgs Sector

There is no theoretical limitation to only have one Higgs doublet in the theory. Therefore, a natural extension to the SM Higgs sector is the two Higgs doublet model (2HDM).

$$\begin{aligned} \mathcal{L}^{\text{2HDM}}_{\text{yukawa}} = &- \sum_{f=u,d,l} \Big(\frac{m_f}{\nu} g^f_h \bar{f} f h + \frac{m_f}{\nu} g^f_H \bar{f} f H - i \frac{m_f}{\nu} g^f_A \bar{f} \gamma_5 f A \Big) \\ &- \Big[\frac{\sqrt{2} V_{ud}}{\nu} \bar{u} (m_u g^u_A P_L + m_d g^d_A P_R) d H^+ + \frac{\sqrt{2} m_l g^d_A}{\nu} \bar{\nu}_L l_R H^+ + h.c. \Big] \end{aligned} \tag{1.9}$$

	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

	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

1.3 Theoretical Problems and Potential Solutions

1.3.1 Hierarchy Problem

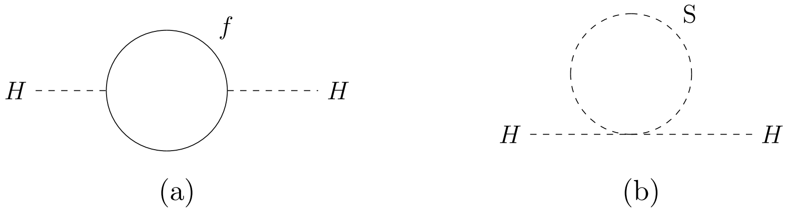


Figure 1.2: One-loop corrections to the Higgs mass by a fermion f (a) and a scalar S (b).

The hierarchy problem can be resolved by adding a symmetry between fermions and bosons that allow cancellations to corrections to the Higgs mass. This theory is known as Supersymmetry. [5] In its simplest form, the Minimal Supersymmetric Standard Model (MSSM), is a type-II 2 Higgs doublet model (2HDM), the predicts the existence of five Higgs bosons: two scalars (h, H), two charged bosons ($H^{\pm}$) and one pseudoscalar (A). At tree level this Higgs sector only depends on two parameters, m_A and $\tan\beta$, where $\tan\beta$ is the ratio of vacuum expectation values.

$$\tan\beta = \frac{\langle H_u^0 \rangle}{\langle H_d^0 \rangle} = \frac{v_u}{v_d} \tag{1.10}$$

The couplings of the neutral Higgs bosons to heavy fermions are different from the Standard Model Higgs couplings. At tree level these differ by the factors shown in the Table **??**.

Due to the hierarchy of the top and bottom masses, it is expected that $\tan\beta$ is greater than 1 and therefore the couplings to tau leptons and bottom quarks would be enhanced and the coupling to top quarks would be suppressed.

Eq.(**??**) and Eq.(**??**) show that if the mass of the scalar is equivalent to that of the fermion and $\lambda_f = \lambda_S^2$, then the Higgs mass corrections cancel. This offers a solution to the hierarchy problem introducing a new symmetry that extends the Standard Model. The symmetry relates fermions and bosons and is known as Supersymmetry. It states that fermions and bosons exist in groups called supermultiplets. Each supermultiplet contains fermion and boson states which are superpartners of one another. On-shell each supermultiplet must have an equivalent number of fermionic and bosonic degrees of freedom. In order for this to also hold off-shell, an auxiliary field is added to balance the number of degrees of freedom. Extending this theory to currently known particles, the simplest set of supermultiplets can be found and are shown in Tables **??** and 1.3.

Two Higgs supermultiplets are needed due to the anomaly cancellation condition, $\mathrm{Tr}[T_3^2 Y] = \mathrm{Tr}[Y^3] = 0$, where Y and T_3 are the third components of weak hypercharge and isospin respectively. The Higgs chiral supermultiplet, the higgsino, makes a significant contribution to the trace as it has a weak hypercharge $Y = +\frac{1}{2}$. This anomaly is solved by introducing another Higgs chiral supermultiplet, with a higgsino of weak hypercharge $Y = -\frac{1}{2}$, so that the traces cancel. This results in five physical Higgs boson states.

Names	Spin $\frac{1}{2}$	Spin 1	$SU(3)_C, SU(2)_L, U(1)_Y$
Gluino, gluon	$\tilde{g}$	g	$(\mathbf{8}, \mathbf{1}, 0)$
Winos, W bosons	$\tilde{W}^{\pm}\ \tilde{W}^0$	$W^{\pm}\ W^0$	$(\mathbf{1}, \mathbf{3}, 0)$
Bino, B boson	$\tilde{B}^0$	B^0	$(\mathbf{1}, \mathbf{1}, 0)$

Table 1.3: The Minimal Supersymmetry Standard Model Vector Supermultiplets [5].

The Z^0 and γ states are found by mixing the W^0 and B^0 states after electroweak symmetry breaking in the Standard Model. There corresponding gauginos that are named zino ($\tilde{Z}^0$) and photino ($\tilde{\gamma}$) are found by the same method from the $\tilde{W}^0$ and $\tilde{B}^0$.

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

$$\mathcal{L} = \mathcal{L}_{\text{SUSY}} + \mathcal{L}_{\text{soft}}. \tag{1.11}$$

$\mathcal{L}_{\text{soft}}$ 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 Standard Model particles and the sparticles. If the mass splitting becomes significant, the hierarchy problem would be reintroduced as corrections to the Higgs mass would again become large.

The MSSM Higgs sector is CP-conserving at tree level. It contains a light, h, and a heavy, H, CP-even Higgs boson, a CP-odd Higgs boson, A, and two charged Higgs bosons $H^{\pm}$. The parameters of the sector can all be obtained from three factors: the Z boson mass, m_Z, the CP-odd Higgs mass, m_A, and $\tan\beta$, where

$$\tan\beta = \frac{\langle H_u^0 \rangle}{\langle H_d^0 \rangle} = \frac{v_u}{v_d}. \tag{1.12}$$

$\tan\beta$ is the ratio of the vacuum expectation values of the neutral components of the Higgs doublets. The masses of the h, H and $H^{\pm}$ Higgs bosons at tree level are calculated to be [8]

$$m_{\mathrm{h}}^2 = \frac{1}{2}\Big(m_{\mathrm{A}}^2 + m_{\mathrm{Z}}^2 - \sqrt{(m_{\mathrm{A}}^2 + m_{\mathrm{Z}}^2)^2 - 4m_{\mathrm{Z}}^2 m_{\mathrm{A}}^2 \cos^2 2\beta}\Big), \tag{1.13a}$$

$$m_{\mathrm{H}}^2 = \frac{1}{2}\Big(m_{\mathrm{A}}^2 + m_{\mathrm{Z}}^2 + \sqrt{(m_{\mathrm{A}}^2 + m_{\mathrm{Z}}^2)^2 - 4m_{\mathrm{Z}}^2 m_{\mathrm{A}}^2 \cos^2 2\beta}\Big), \tag{1.13b}$$

$$m_{\mathrm{H}^\pm}^2 = m_{\mathrm{A}}^2 + m_{\mathrm{W}}^2. \tag{1.13c}$$

All of the tree level couplings can also be expressed in terms of these three parameters but they will not all be stated here. The main couplings of interest for this report are that between the neutral Higgs bosons and heavy fermions. These couplings are the largest of the Higgs sector. The couplings of the three heaviest fermions to the MSSM Higgs sector are shown in the below Table **??**.

Due to the mass hierarchy of the top and bottom quarks, it is expected that $\tan\beta$ lies in the range

$$1 \lesssim \tan\beta \lesssim \frac{m_t}{m_b}. \tag{1.14}$$

1.4 Experimental Tensions and Potential Solutions

1.4.1 B Anomalies

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

$$\mathcal{L}_U = \frac{g_U}{\sqrt{2}} U^\mu \big[\beta_L^{i\alpha}(\bar{q}_L^i \gamma_\mu l_L^\alpha) + \beta_R^{i\alpha}(\bar{d}_R^i \gamma_\mu e_R^\alpha)\big] + \text{h.c.} \tag{1.15}$$

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

$$\beta_L = \begin{pmatrix} 0 & 0 & \beta_L^{d\tau} \\ 0 & \beta_L^{s\mu} & \beta_L^{s\tau} \\ 0 & \beta_L^{b\mu} & 1 \end{pmatrix}, \qquad \beta_R = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \beta_R^{b\tau} \end{pmatrix}. \tag{1.16}$$

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

1.4.2 g-2 Anomaly

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

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

Chapter 2

The LHC and CMS Experiment

2.1 The LHC

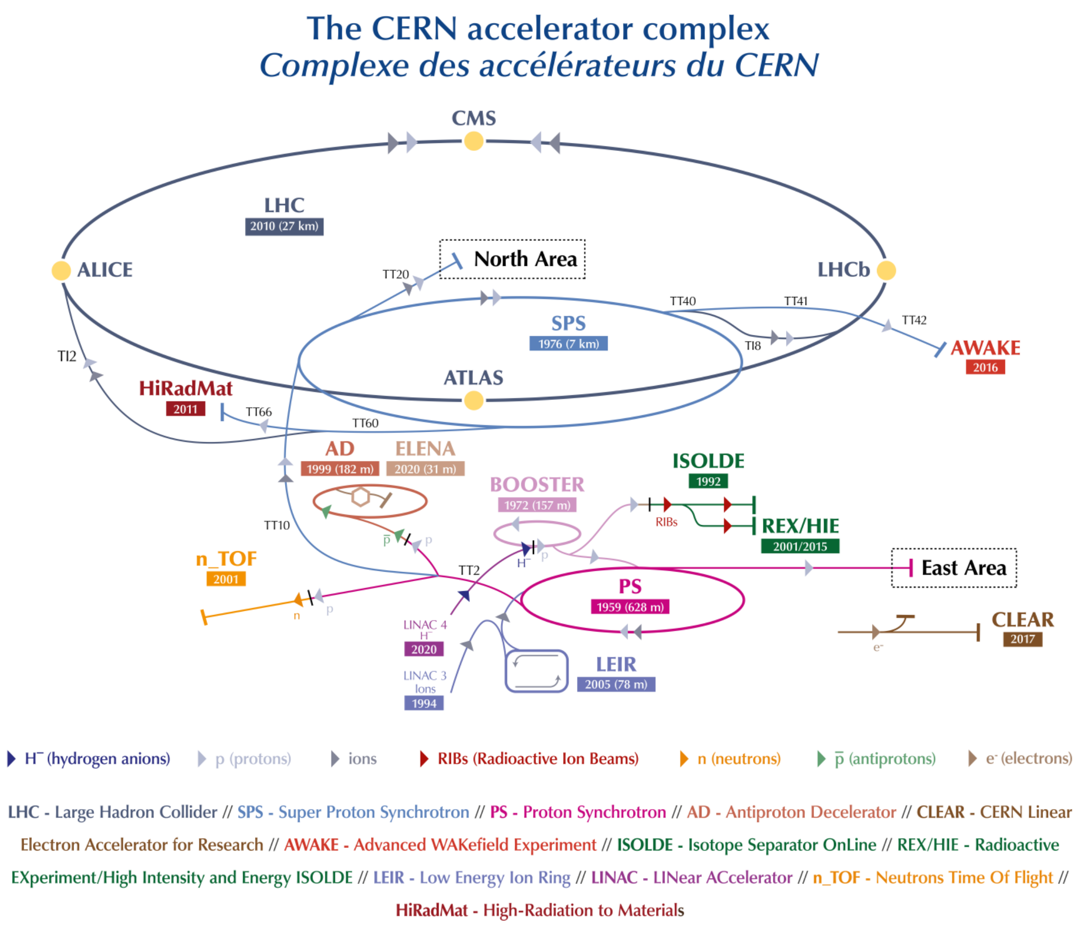


Figure 2.1: A schematic diagram of the CERN [1].

2.2 The CMS Detector

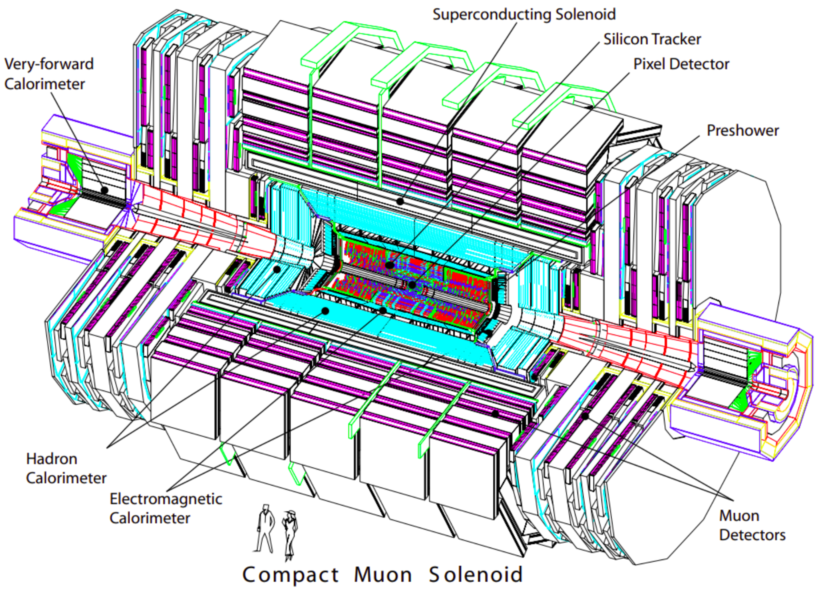


Figure 2.2: A perspective view of the CMS detector [2].

2.2.1 Tracker

2.2.2 Electromagnetic calorimeter

2.2.3 Hadron calorimeter

2.2.4 Muon system

2.2.5 Triggering

Chapter 3

Object Reconstruction

3.1 Tracks and vertices

3.2 Particle flow

The particle-flow (PF) algorithm reconstructs the products of the LHC pp collisions and is described in full in Ref.[9]. It utilises all the information available from the tracker, ECAL, HCAL and muon detectors combined to produce a list of particle candidates. These candidates are either a photon, electron, muon or a neutral or charged hadrons. It begins with defining an event as the data taken per bunch crossing. The PF algorithm then reconstructs the tracks of the particle candidates in order to find the collision vertices. The primary collision vertex is taken to be the one with the largest value of p_T^2 summed over all physics objects originated from that vertex. Physics objects are not only defined to include particle candidate tracks, but also missing tracks represented by the negative vectorial sum of all particle candidate tracks. Other pp collisions vertices are referred to as pileup.

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

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

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

Jets originating from the hadronisation of b quarks, are identified using the combined secondary vertex b-tagging algorithm. This discriminates between jets originating from b quarks from other jets, utilising track impact parameters and secondary vertex related variables [10]. This plays a key part in the analysis, as b-tagging is used for categorisation purposes described in Section **??**. The missing transverse momentum, $\vec{p_T}^{\text{miss}}$, is also used in categorisation of events and is calculated as the negative vector sum of all PF reconstructed transverse momenta.

3.3 Muons

3.4 Electrons

3.5 Jets

3.6 b jets

3.7 Missing transverse energy

3.8 Taus

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

Decay Mode	Branching Fraction
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%
$h^-2\pi^0\nu_\tau$	9.3%
$\pi^-\pi^-\pi^+\nu_\tau$	9.0%
$\pi^-\pi^-\pi^+\pi^0\nu_\tau$	2.7%
other	6.4%

Table 3.1: Measured branching fractions, that are greater than 2%, for the tau lepton. h represents a charged hadron either a pion or a kaon.

These decays can be split into three groups: the 17.8% of taus that decay to an electron (e), the 17.4% that decay into a muon (μ) and hadronic tau decays (τ_h) that make up the final 64.8% of tau decays. The leptonic decays of the tau can be accounted for by the identification of electrons and muons as discussed in the previous subsection. The hadron-plus-strips (HPS) algorithm is used to identify hadronic taus [11, 12]. This algorithm groups electrons, positrons and photons and names this cluster as a "strip". This is defined to represent the decay products of the π^0 meson. The strip size is variable depending on the p_T of its components. In a jet, the number of strips and charged particles are counted. If the numbers are corresponding to the number of π^0 mesons and charged hadrons shown in Table 3.1 for hadronic decays, it is concluded that the jet may originate from a tau lepton.

To reduce misidentification, the tau lifetime is utilised. The tau lepton is expected to travel a small but identifiable distance before it decays. This distance between the decay vertex and the primary vertex is the variable used. To further reduce misidentification from the hadronisation of quarks or gluons, a similar isolation discriminant is used as for electrons and muons with $\Delta R < 0.3$. All of these are combined into a multivariate hadronic tau identification algorithm (MVA) given in Ref.[11]. From this reference the working points Tight, Medium and VeryLoose are used. These refer to the output of the MVA varying the maximum values of the p_T of the hadronic tau candidate. The same MVA (excluding the HPS algorithm) is also used to reduce misidentification of leptonic tau decays. For τ_h identification in this analysis the Tight working point is used. In order to suppress misidentification of leptonic tau decay the Tight (VeryLoose) working point for electrons and Loose (Tight) working points for muons are used in the $e\tau_h(\mu\tau_h)$ channels.

For a final state of two taus, there are six possible final states: $e\mu$, $e\tau_h$, $\mu\tau_h$, $\tau_h\tau_h$, ee and $\mu\mu$. However, ee and $\mu\mu$ are dominated by large backgrounds and have relatively low cross sections, so provide very little sensitivity to this analysis and hence are not included. The other four channels are all utilised.

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 mass hierarchy. They are also sensitive to non-resonant effects from new physics mediators. This chapter will detail the searches for two such areas of new physics: additional Higgs bosons and vector leptoquarks. These searches are split up into three sections:

i) A model independent search for single narrow spin-0 resonance (ϕ), produced via gluon fusion (ggϕ) or in association with a b quark (bbϕ). The SM Higgs boson is treated as a background and the Yukawa couplings of the spin-0 resonance that contribute to the gluon fusion loop are set to SM values.

ii) A search for the MSSM Higgs sector in a number of benchmark scenarios. The benchmark scenarios were proposed in Refs. [13, 14, 15] and summarised in Ref. [16]. 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. These scenarios are detailed in Section 4.1.2.

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 [17].

4.1 Signal Modelling

4.1.1 Additional Higgs Bosons

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

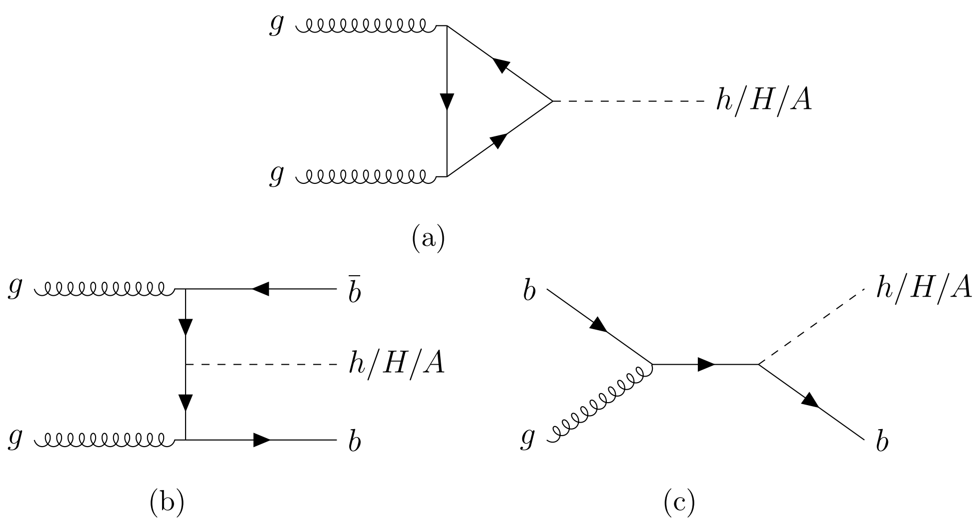


Figure 4.1: Diagram (a) shows the production of neutral Higgs bosons from gluon fusion. The dominant loop contributions to this diagrams are from top-only, bottom-only and top-bottom interference. Diagrams (b) and (c) show production in association with b quarks.

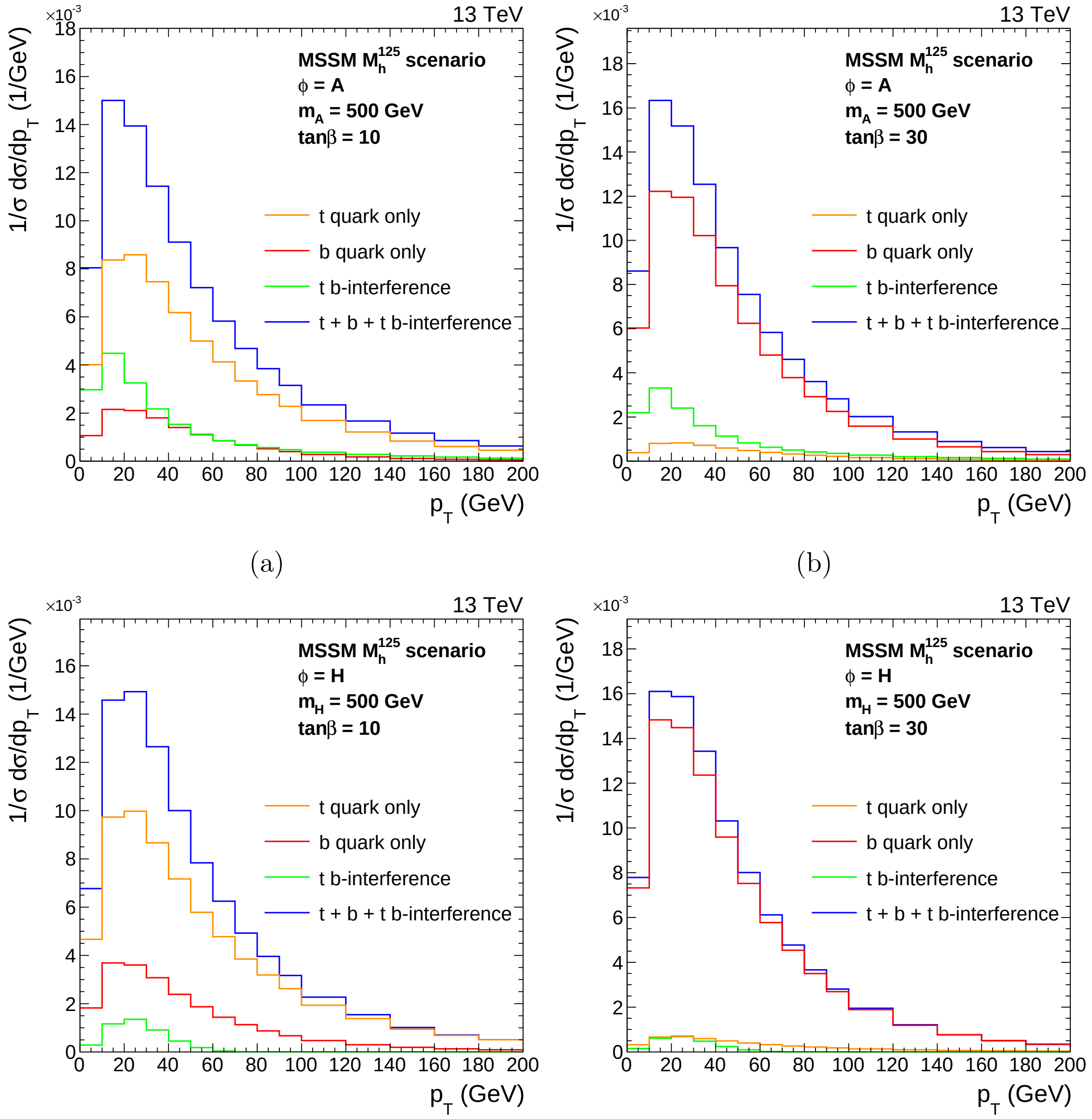
With the $\tan\beta$ enhancement, the decays of additional Higgs bosons to τ leptons and b quarks are most likely. τ leptons are identified with a higher purity than bottom quarks at the CMS detector. It is also easier to separate $\tau^+\tau^-$ from the large QCD multijet background produced from the high energy proton-proton collisions. This hypothesis was tested with the 2016 dataset and although no deviations were observed, the strongest limits on the MSSM phase space was placed by the $\tau^+\tau^-$ final states [17, 18].

For this analysis, signal templates for the production of additional Higgs bosons over a mass range of 60 GeV to 3.5 TeV are generated. Gluon fusion is simulated at NLO precision using the 2HDM implementation of PowHeg 2.0 [19, 20, 21, 22]. The kinematic properties are highly dependent on the contributions to the loop, that are dependent on the specific signal model. To account for the different loop contributions at the NLO plus parton shower prediction, weights based off the p_T spectra are calculated to split the contributions from the t quark only, b quark only, and tb-interference. Once individual templates have been determined for each contribution to the loop, the 2HDM samples can be scaled to the MSSM scenario prediction with the following formula,

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

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

Production in association with bottom quarks is simulated at NLO precision using the corresponding PowHeg 2.0 implementation in the four-flavour scheme [19, 20, 21, 22]. All additional Higgs boson signal generation is performed using the parton



(a) (b)

(c) (d)

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

distribution function (PDF) NNPDF3.1 [23, 24]. τ lepton decay, parton showering and hadronisation are all modelled with the PYTHIA event generator where the PU profile is matched to data [25, 26]. All events generated are passed through a GEANT4-based [27] simulation of the CMS detector and reconstructed in the same way as data.

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

4.1.2 Vector Leptoquarks

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

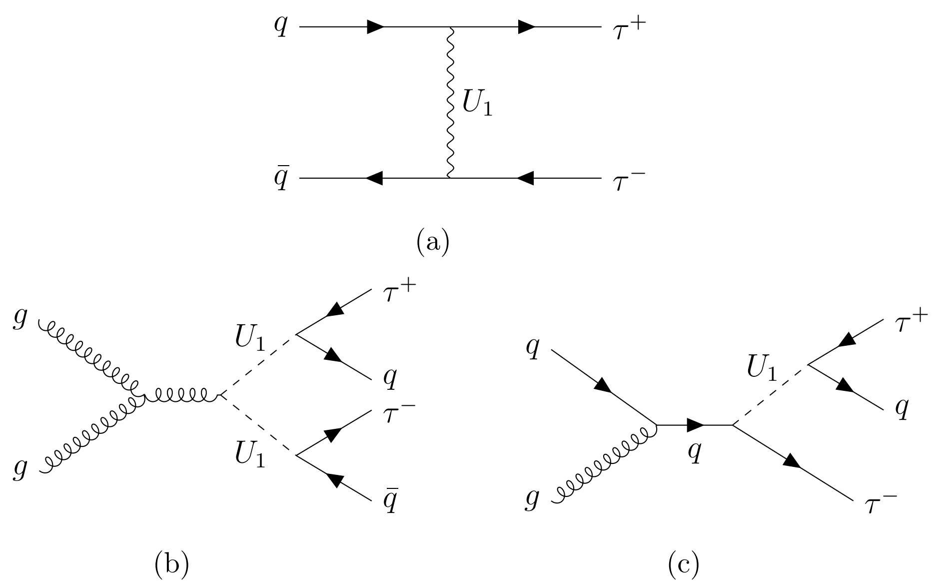


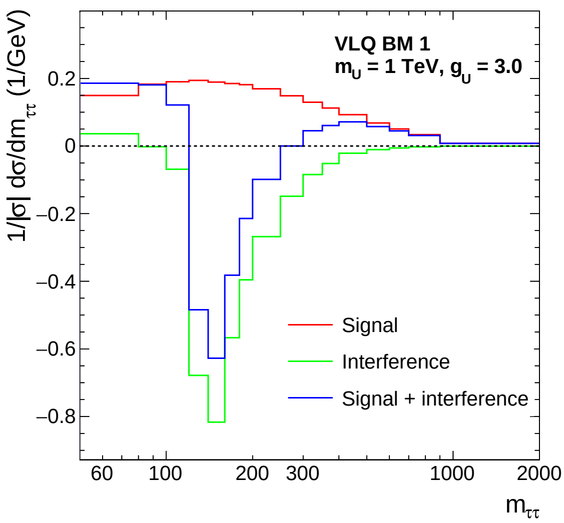
Figure 4.3: Feynman diagrams showing the contribution from U_1 vector leptoquarks to the final state with a pair of oppositely charged τ leptons. Diagram (a) shows t-channel, (b) pair and (c) single production of a vector leptoquark.

Pair and single production of a vector leptoquark are dependent on its strong coupling, which is highly model dependent. For large mass, m_U, the probability of producing an on-shell U_1 singlet or pair is heavily suppressed due to the momentum of the initial partons. These production processes are not discussed further in this search. Further studies have been used to search for single and pair production at the CMS experiment and no statistically significance derivation was observed [28]. Single production places the loosest constraints out of the processes mentioned whilst pair production puts a lower limit on the leptoquark mass, as the process is approximately independent on g_U, and this limit is heavily dependent on the values taken for the strong coupling.

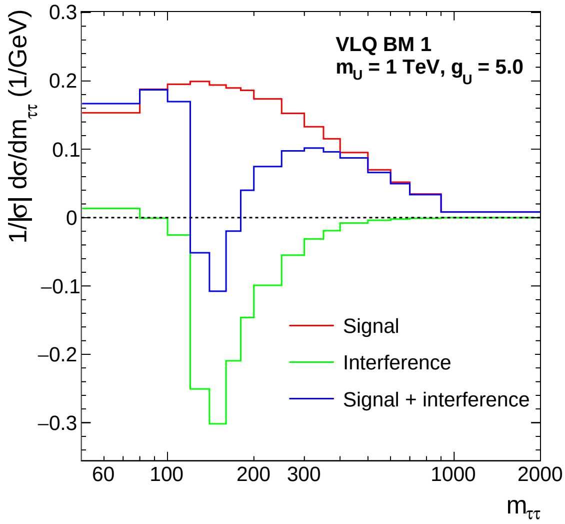
The t-channel process contain two vertices with a U_1 vector leptoquark, a quark and a τ lepton, and hence the cross section will scale with g_U^4. From the best fit to b anomalies, the vertex is dominated by the b quark and hence the initial state will be mostly from $b\bar{b}$, with sub-dominant contributions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

The signal process of the U_1 t-channel exchange is simulated in the five-flavour scheme (5FS) at LO precision using the MadGraph5_aMC@NLO v2.6.5 event generator [29]. Events are generated with one or fewer outgoing partons from the matrix element and the MLM prescription [30] is then used for matching, with a scale set to 40 GeV. Negligible dependence of the U_1 decay width is observed, for simulation this is chosen to approximately match the value predicted by the b anomaly fit. Samples with a mass between 1 and 5 TeV at $g_U = 1$ are generated.

The interference between the U_1 signal and $Z/\gamma^* \to \tau\tau$ production was checked and a large destructive affect was observed, with magnitude dependent on g_U. To account for this, separate samples are produced for this interference, generated in the same way as the t-channel exchange. The interference samples are then split into two with a di-τ mass split in order to have a sufficient number of events in the high di-τ mass regions. The cross section of these interference samples scale 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 Drell-Yan. This is shown in the VLQ BM 1 scenario for a leptoquark of mass 1 TeV for coupling strengths of $g_U = 3$ (a) and $g_U = 5$ (b).

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

4.2 Event Selection

The possible decays of two τ leptons and their branching fractions, where the τ decay is grouped into three categories e, μ and τ_h as defined in Section **??**, are shown in Table 4.1. For this search the four largest branching fraction channels used: $\tau_h\tau_h$, $e\tau_h$, $\mu\tau_h$ and $e\mu$. This accounts for approximately 94% of di-τ events. The two same lepton channels are neglected due to 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 the four final state pairs, a number of different online trigger requirements are needed. In the $\tau_h\tau_h$ channel, two possible triggers are available: the double-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 has a p_T threshold at 40 GeV. Therefore, the double-τ_h trigger is used individually where 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 cross-trigger is used for events where the light lepton has p_T between the thresholds for the cross-trigger and single-e/μ shown in Table 4.2. The light lepton used in the cross-trigger is required to be in the central barrel of the detector within $|\eta| < 2.1$. Above these light lepton p_T thresholds the single-e/μ trigger is used, where it is required that the τ_h has $p_T > 30$ GeV. 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	$e\tau_h$ cross-trigger	single-e	$\mu\tau_h$ cross-trigger	single-μ
2016	23	26	20	23
2017	25	28	20	25
2018	25	33	21	25

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

In the $e\mu$ channel, there are three possible triggers available: the single-e, single-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 e and μ are required to have $p_T > 15$ GeV and $|\eta| < 2.4$.

4.2.2 Offline Requirements

All offline selections stated are in addition the object selection discussed in Section **??**. In this analysis, τ_h candidates are required to pass the `Medium` $D_{\text{jet}}^{\text{WP}}$. D_e^{WP} and D_μ^{WP} 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 τ_h from the $Z \to ll$ process. The light lepton isolation requirement is $I_{\text{rel}}^{e/\mu} < 0.15$ except for in the $e\mu$ channel where the muon is required to have $I_{\text{rel}}^{\mu} < 0.2$.

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

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

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

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 and the MET and light lepton from the W decay are aligned and hence the event has a large $m_T(\vec{p}_T^{\,e/\mu}, \vec{p}_T^{\,\text{MET}})$.

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

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

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

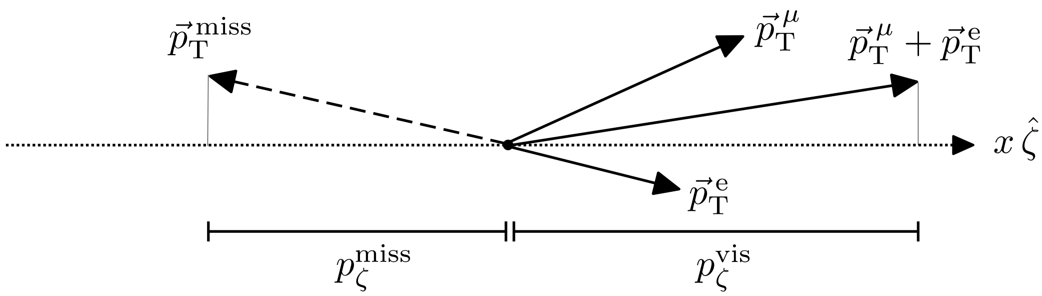


Figure 4.5: Diagram of inputs to the D_ζ variable.

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. Events with no b tag with this cut are vetoed and events with a b tag also with this cut are used for a $t\bar{t}$ control region and discussed further in Section 4.4.

4.3 Search Optimisation

The optimisation of the signal extraction depends on which of the three scenarios, set out at beginning of this section, is searched for. The components of the optimisation are named the high-mass, low-mass and SM Higgs optimisation procedures. For the model independent search (i) the high- or low-mass optimisation procedures are used depending on whether the mass of the resonance is greater or less than 250 GeV. The search for the MSSM Higgs sector (ii) uses the high mass or the SM Higgs optimisation procedures depending on whether the reconstructed di-τ mass is greater or less than 250 GeV. Due to more than one neutrino in the event, it is not possible to fully reconstruct the di-τ mass. To resolve this issue a likelihood-based method named SVFit algorithm is used [31]. Finally, the search for vector leptoquarks (iii) uses only the high mass optimisation procedure. The procedures are discussed in detail below.

The high-mass optimisation procedure follows what was done in Ref. [17]. Each event is initially split into categories depending on whether is has 0 or ≥1 b tagged jets in the event. Firstly, this helps target the additional Higgs boson production modes gluon fusion and b associated production respectively. In particular, $bb\phi$ has final state b quarks that if tagged can significantly aid the sensitivity of this signal. Secondly, the initial state radiation of a t-channel vector leptoquark signal dominated by initial states of b quarks, can lead to additional b jets in the final state. The reduced backgrounds in b tagged events allows for a more sensitive vector leptoquark search in this category.

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

- `Tight`-m_{T}: $m_{\mathrm{T}}(\vec{p}_T^{\,e/\mu}, \vec{p}_T^{\,\mathrm{miss}}) < 40$ 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 improve the signal acceptance for resonant masses of $m_\phi > 700$ GeV.

In the $e\mu$ channel, is subdivided into three signal categories based of 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: $-10 \leq D_\zeta[\mathrm{GeV}] < 30$ GeV;
- `High`-D_ζ: $D_\zeta[\mathrm{GeV}] \geq 30$ 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 are shown in Figure 5.5.

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

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

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 where a di-τ mass is unphysical for the signal. This is due to the use of the transverse momenta and MET in the variable definition. For a t-channel signal where the mediator has high mass, no significant mass separation is expected in any variable.

The low-mass optimisation procedure loosely follows the high-mass procedure with a few key difference. Firstly, categories that are only sensitive to high-mass signals are dropped. This includes the `Low-`D_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.

Finally, the SM Higgs optimisation procedure is taken from the CMS SM $H \to \tau\tau$ analysis and is detailed in Ref. [32]. This was previously used for simplified template cross section measurements. This uses a neural-network-based (NN) categorisation to obtain the most precise estimates from data of the SM Higgs produced via gluon fusion, vector boson fusion or vector boson associated production. The NN based analysis introduces 26 categories, 8 of which are optimised to pull out the Higgs boson signal. Although the NN is trained specifically to target events with an SM-like Higgs boson, signal events with differing masses can also enter the NN categories.

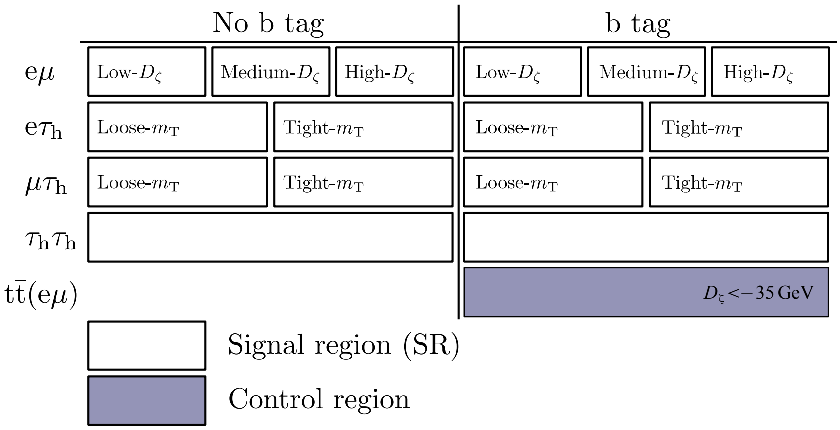


Figure 4.6: Overview of the categories used for the extraction of the signal in the high mass optimisation procedure.

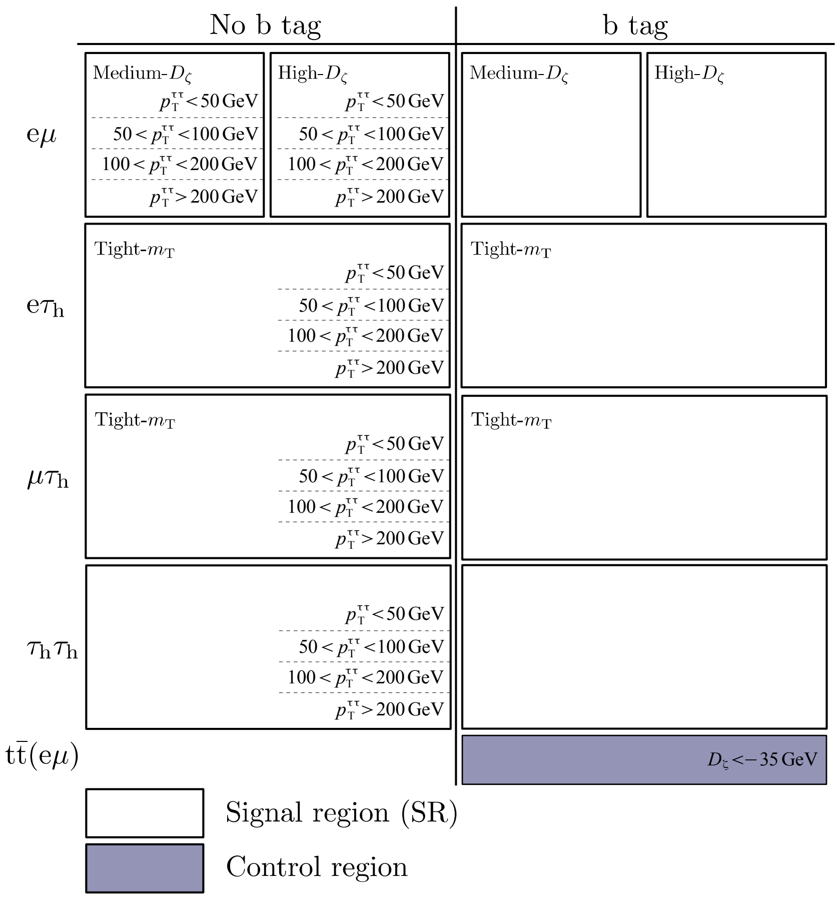


Figure 4.7: Overview of the categories used for the extraction of the signal in the low mass optimisation procedure.

4.4 Background Modelling Overview

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

i) Events containing only genuine τ leptons.

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

iii) Events with jets faking both light leptons (jet$\rightarrow l$) 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 leptons.

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

- Non $t\bar{t}$ events with a prompt light lepton (e or μ not from a τ decay) and the other object (if there are not two prompt light lepton) is from a genuine τ leptons.
- Events with a light lepton misidentified as a τ_h and the other object (if there are not two light leptons faking a τ_h) is reconstructed as prompt light lepton or from genuine τ leptons.
- 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) consists of largely $Z/\gamma^* \rightarrow \tau\tau$ events but there are also smaller contributions from other processes. This background is modelled by a data-simulation hybrid method called the embedding method and this is described in detail in Section 4.6. Group (ii) is dominated by QCD, W + jets and $t\bar{t}$ events with a jet$\rightarrow \tau_h$ misidentification. This is modelled from data by the fake factor method (F_F) and is explained in Section 4.7. Group (iii) is modelled from data to describe the QCD multijet contribution to the background in the $e\mu$ channel. The method to obtain this background is described in Section 4.5. The data driven background estimations for (i), (ii) and (iii) contribute ≈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. The $t\bar{t}$ process is separated due to its large contribution to the phase space where a b jet is required.

In 2016 (2017–2018), the W + jets and $Z \rightarrow ll$ processes are simulated using the MadGraph5_aMC@NLO 2.2.2 (2.4.2) event generator at leading order (LO) [29]. Supplementary samples are generated with up to four outgoing partons in the hard interaction to increase the number of simulated events in regions of high signal purity. For diboson production, MadGraph5_aMC@NLO is used at next-to-LO (NLO) precision [29], and the FxFx [30] (MLM [33]) prescription is used to match the NLO (LO) matrix element calculation with the parton shower model. Samples for $t\bar{t}$[34] and (t-channel) single top quark production[35] are generated at NLO precision using PowHeg 2.0 [19, 20, 21, 22], and for single top quark production in association with a W boson (tW channel)[36], PowHeg version 1.0 at NLO precision is used. When compared with data, W + jets, $Z \rightarrow ll$, $t\bar{t}$, and single top quark events in the tW channel are normalised to their cross sections at next-to-next-to-LO (NNLO) precision[37, 38, 39], while single top quark (t-channel) and diboson events are normalized to their cross sections at NLO precision or higher [39, 40, 41].

4.5 QCD Estimation in the $e\mu$ Channel

The QCD model in the $e\mu$ channel, that attempts to model events where two jets are misidentified as an election muon pair, is taken from data with a same-sign pair with a transfer factor (F_T). The transfer factor determines differences from the same-sign to opposite-sign region and is calculated from a sideband region with an anti-isolated muon ($0.2 < I^{\mu}_{\text{rel}} < 0.5$). F_T is initially parameterised by the ΔR between the election and muon, and the number of jets in the event, however additional dependencies on the electron and muon p_T enter via a correction.
Good agreement is observed in events with no b jets, when applying F_T onto same sign events compared to opposite sign events where both regions have an anti-isolated muon. However, in events with b jets an additional correction is needed. This is determined to be ≈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 $0.15 < I^{e}_{\text{rel}} < 0.5$. 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 anti-quark pair. However, if one of the heavy quarks is tagged as a b jet, another object has to be the jet initiator (a radiated gluon for example) and there is therefore no charge preference in the pair. As F_T is originally fit inclusively in numbers of b jets and the 0 bin is dominant, the correction over predicts the opposite sign to same sign ratio and so a large correction is needed as observed.

4.6 Embedding Method

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

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

Next, all energy deposits in the detector from the selected muons are removed. This involves removing the hits on global-muon track in the tracker, hits in the muons systems and clusters in the calorimeters that intercept the muon trajectory. Once completed, the selected muons and their kinematic properties are replaced with a τ lepton. To account for the difference in mass between the muon and τ, the muons are boosted into the center-of-mass frame of the di-muon system and then this 4-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 [25, 26] and separate samples are produced for different $\tau\tau$ decay channels. Only the decay of the τ leptons are then processed through the detector simulation and the remainder of the $\mu\mu$ event is added back. A schematic of the process is shown in Figure 4.8.

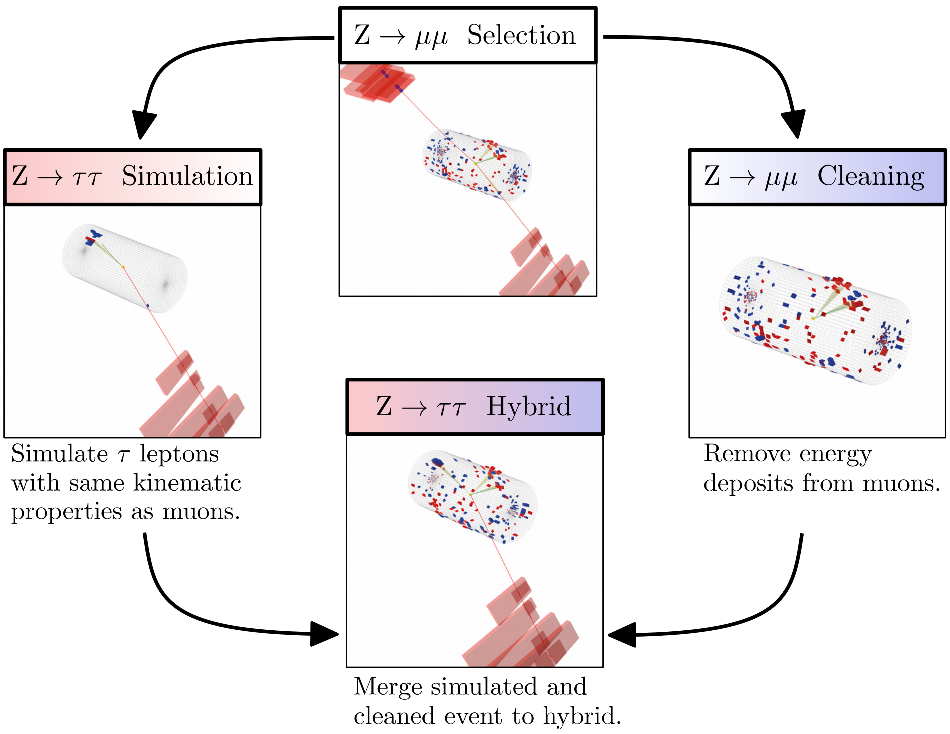


Figure 4.8: Schematic of the embedding method to model genuine di-τ backgrounds from di-muon events in data.

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

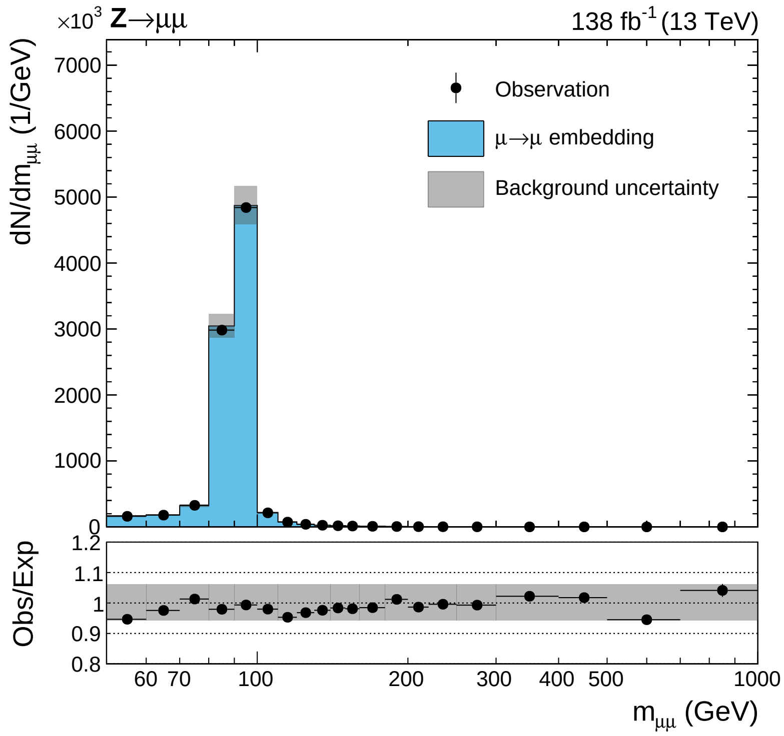


Figure 4.9: Closure plot showing the di-muon mass on the dedicated embedding validation samples.

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 $\to \tau_h$ 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 estimates for these processes. One such procedure is the fake factor (F_F) method.

The F_F method utilises regions in the data to model the jet $\to \tau_h$ background. Firstly, the determination regions, which are jet $\to \tau_h$ enriched control regions orthogonal to the signal region. It is used to calculate F_F by taking the ratio of number of jet$\to \tau_h$ events that pass the nominal τ_h ID requirement ($N(\texttt{Nominal})$), to the number of jet fake events that fail the nominal τ_h ID but pass a looser alternative τ_h ID requirement ($N(\texttt{Alternative \&\& !Nominal})$), as shown in Equation 4.5.

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

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

The following Sections 4.7.1–4.7.4 detail the complexities of how this method is applied to this analysis. For these searches the nominal τ_h ID used is the `Medium` $D_{\text{jet}}^{\text{WP}}$ and the alternative τ_h ID used is the `VVVLoose` $D_{\text{jet}}^{\text{WP}}$.

4.7.1 Determination Regions

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

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

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

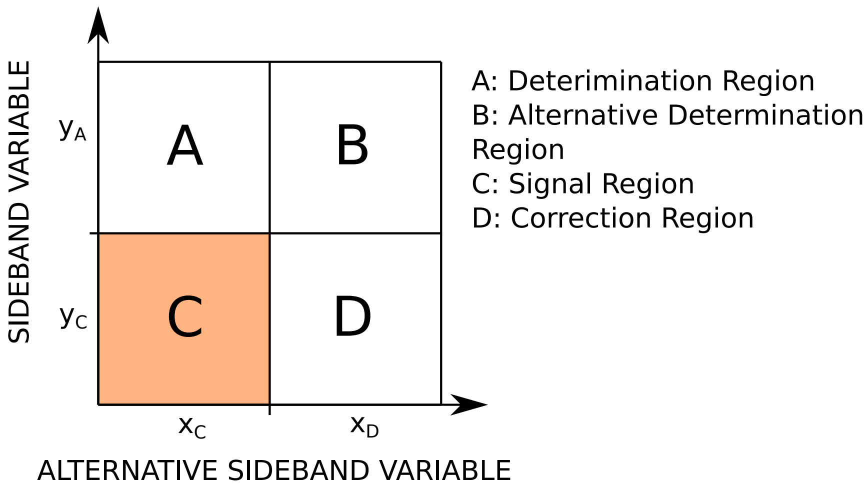


Figure 4.10: Schematic of the regions used for fake factor derivation.

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

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

i) $\tau_h\tau_h$ QCD

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

4.7.2 Parametrisation

The raw F_F^i take into account dependencies on N_{jets} via the analysis 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 matched in ΔR to the τ_h (p_T^{jet}). $N_{\text{pre b jets}}$ is defined to map the dependence of F_F^i on N_{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 $p_T > 20$. 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.

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

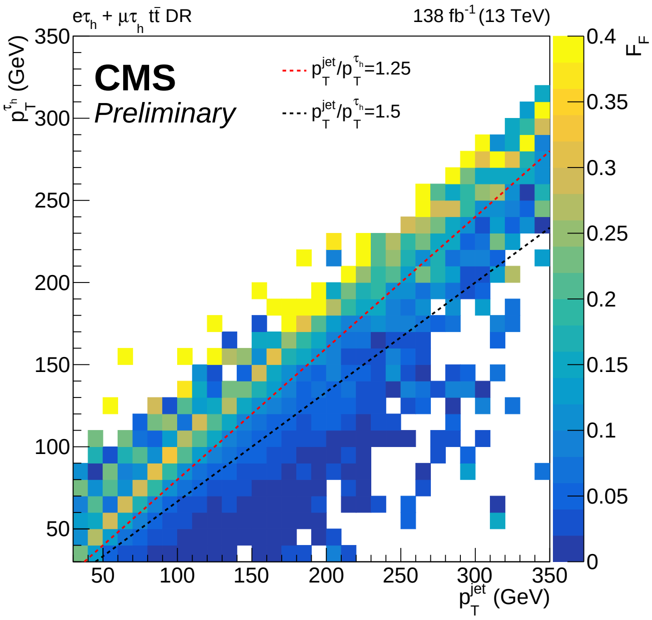


Figure 4.11: A 2D heat map of the fake factors determined from $t\bar{t}$ MC for the full run-2 dataset in the combined $e\tau_h$ and $\mu\tau_h$ channels. This is shown with respect to the τ_h p_T and the p_T of the jet matched to the τ_h. The ratio of jet to τ_h p_T categorisation used is shown split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

$p_T^{\tau_h}$ are closest. The physical motivation for this is when they are close, the 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 $p_T^{\tau_h}$, charge pions are likely not close enough to the PV to be clustered into the jet and so the event is more likely to be classified as a jet fake. This will then lead to the fake factor dependence as seen in Figure 4.11.

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

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

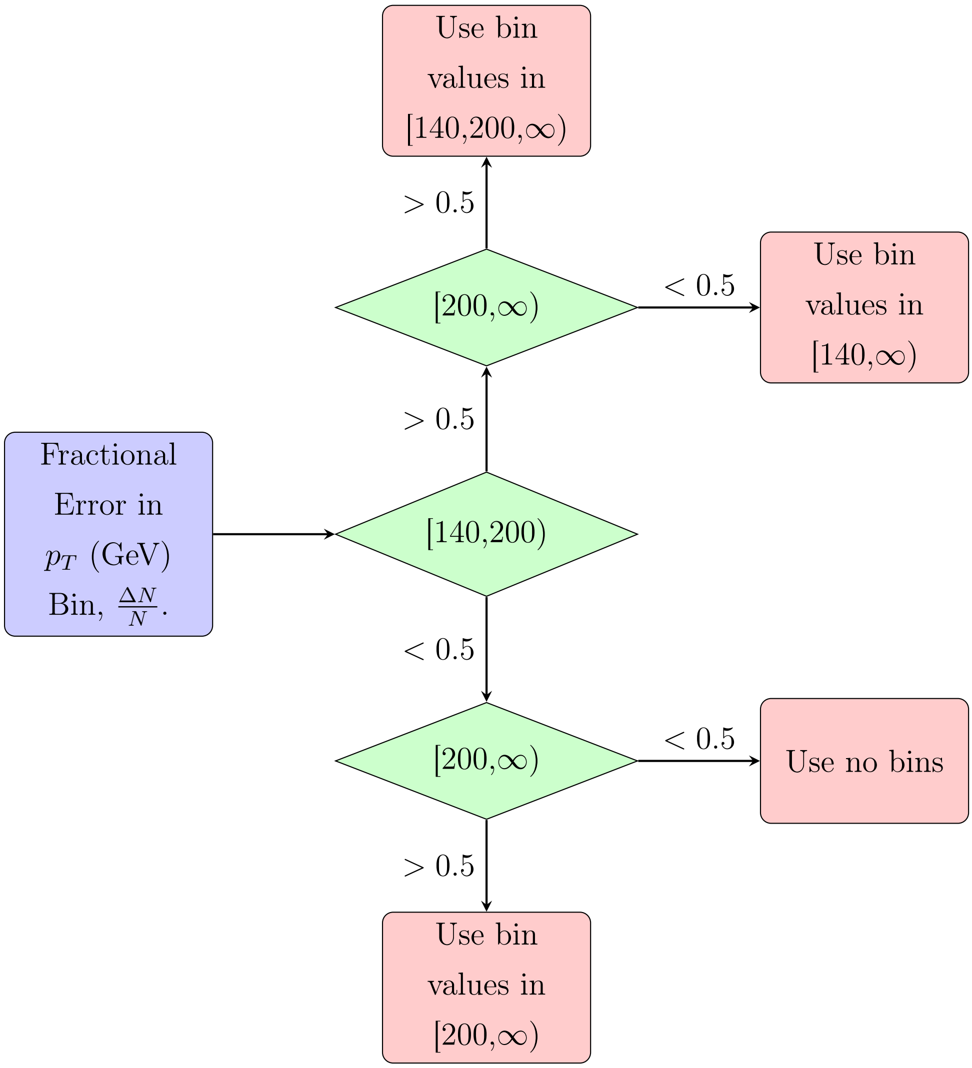
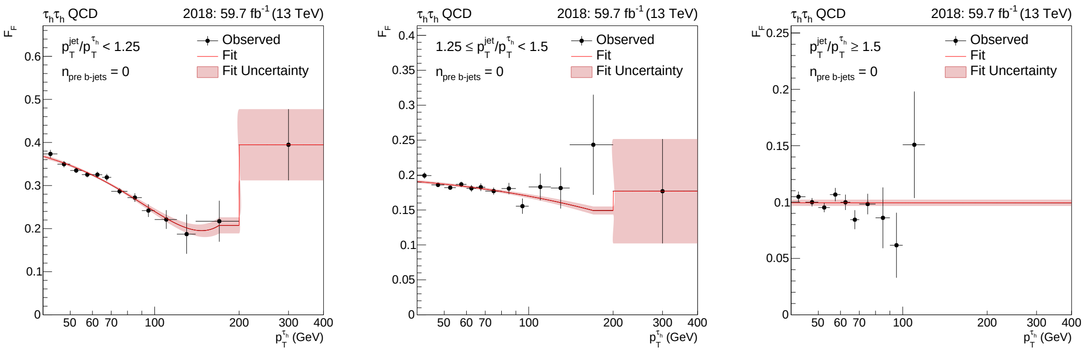


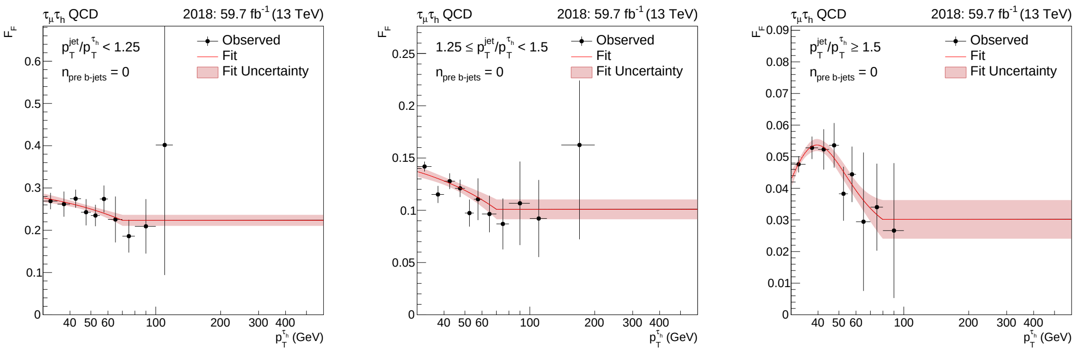
Figure 4.12: Flow chart of the algorithm used to determine where binned values are taken instead of the fit. The blue box represents the input, the green diamonds represent the decisions and the red boxes represent the outputs.

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

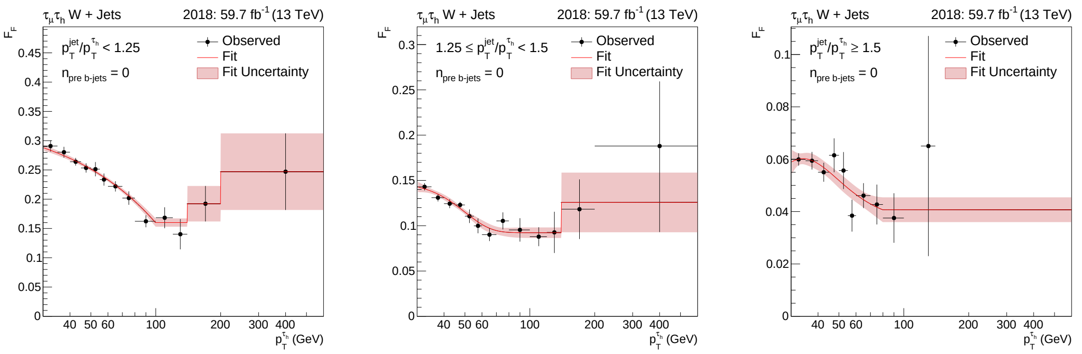


(a) (b) (c)

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



(a) (b) (c)



(d) (e) (f)

Figure 4.14: Fake factor fits in $\mu\tau_h$ channel for the QCD and W + Jets $N_{\text{pre b jets}} = 0$ category with 2018 data. The three jet p_T to 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^{QCD} are then corrected to account for non-closures in other variable in the `Determination Region`. The only significance non-closures are observed for E_T^{miss} related variables and are largest for events with $N_{\text{pre b jets}} = 0$. Closure corrections are performed for the variable ΔR in bins of $N_{\text{b jets}}$. In the $\mu\tau_h$ and $e\tau_h$ channels, the measured F_F^{QCD} and $F_F^{\text{W+jets}}$ 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 E_T^{miss} arising from mismeasurement of the energies of particles in a jet. A diagram of this effect is shown in Figure 4.15.

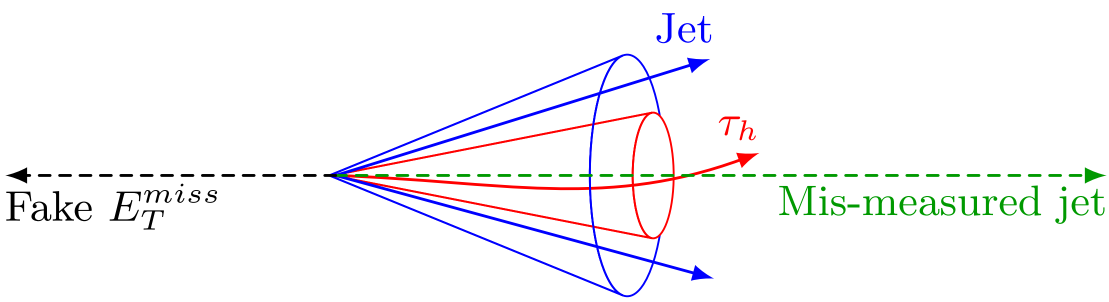


Figure 4.15: Diagram showing how fake E_T^{miss} arises from mismodelling jet constituents in τ_h identification.

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

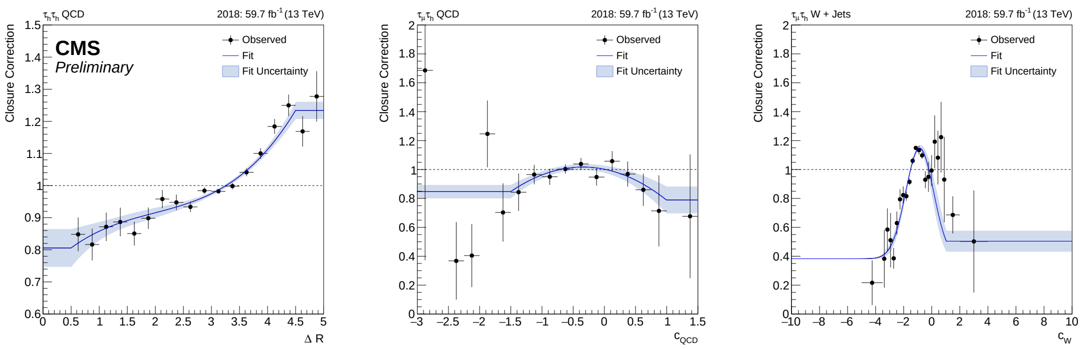
$$C_{\text{QCD}} = \frac{E_T^{\text{miss}} \cos \Delta\phi(\vec{p}_T^{\,\text{miss}}, \vec{p}_T^{\,\tau_h})}{p_T^{\tau_h}}. \tag{4.6}$$

where $\Delta\phi(\vec{p}_T^{\,\text{miss}}, \vec{p}_T^{\,\tau_h})$ is the separation in the azimuthal angle between the the missing a $\vec{p}_T^{\,\text{miss}}$ and $\vec{p}_T^{\,\tau_h}$. The numerator quantifies the missing transverse momentum in the direction of the 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 E_T^{miss} from the total. This approximation assumes the neutrino is back-to-back and balanced with the light lepton (which is exactly true for W bosons produce at rest in the transverse direction). The equation then becomes,

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

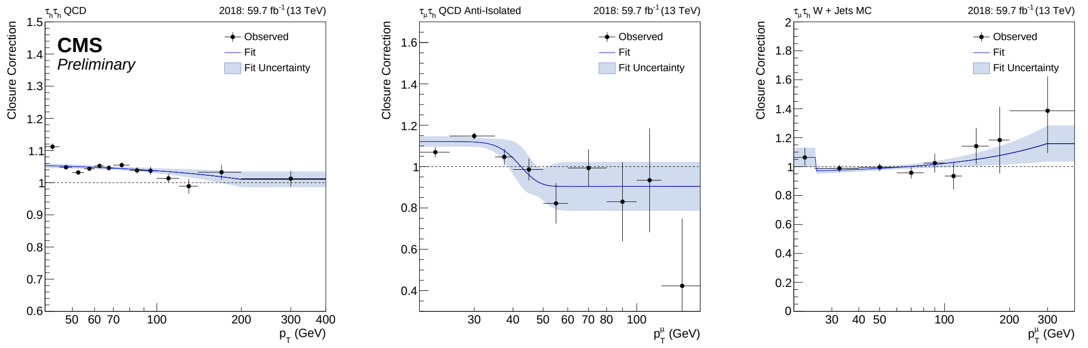
When either correction variable is separated from 0, a larger quantity of fake E_T^miss is expected in the event. In these regions a large correction is needed due to the mis-measured jet energy spectrum shifting the τ_h candidate isolation and so shifting the τ identification scores. Examples of these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

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



(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 Equation 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 parameterised 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 parameterised by the the muon p_T for QCD and W + jets processes respectively.

4.7.4 Applying Fake Factors

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

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

where the factor f_i is defined as

$$f_i = \frac{N^i_{\text{AR}}}{\sum_j N^j_{\text{AR}}}, \tag{4.9}$$

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

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

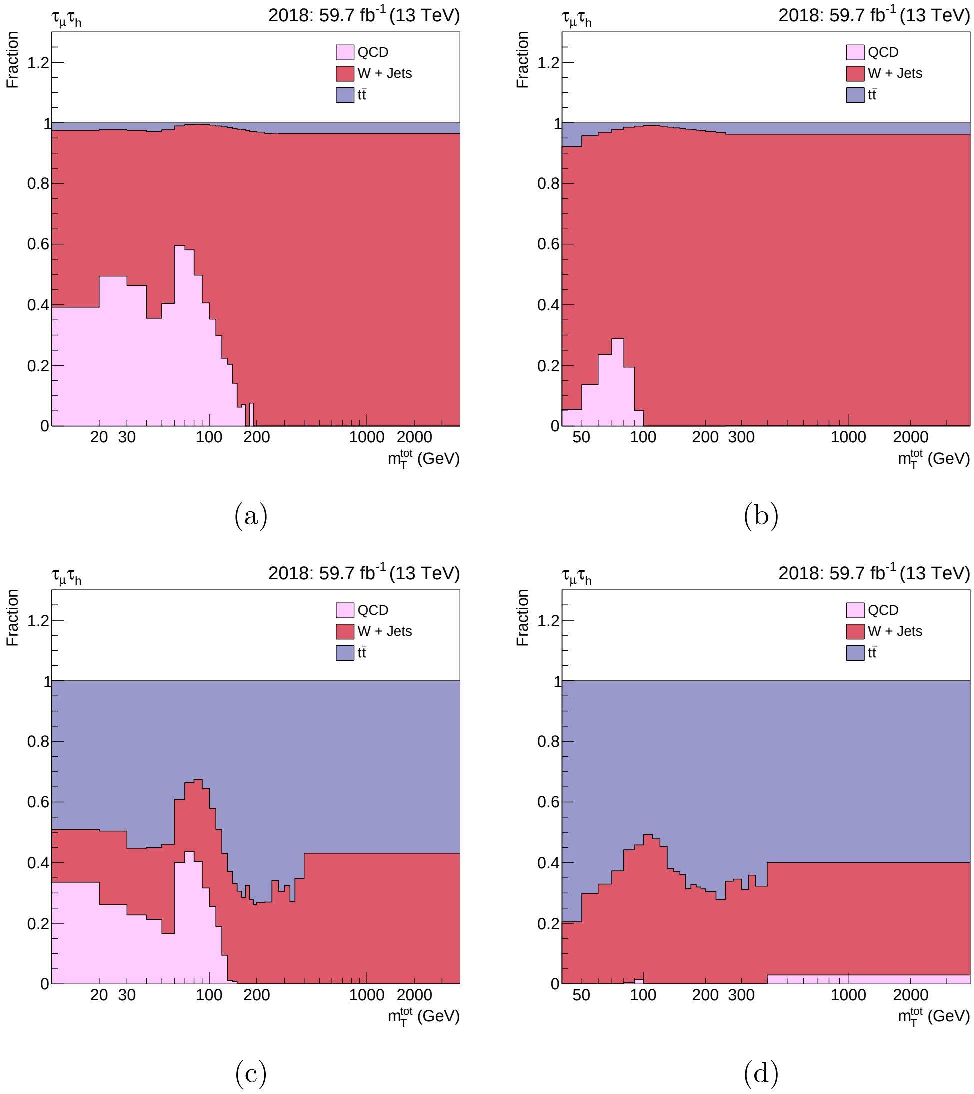


Figure 4.18: The expected application region fractions of the processes in the $\mu\tau_h$ channel. (a) and (b) show the `no b tag` `Tight`-m_T and `Loose`-m_T categories and (c) and (d) show the `b tag` `Tight`-m_T and `Loose`-m_T categories respectively.

the τ_h ID in C. This models all events where the leading τ_h candidate is a jet$\rightarrow \tau_h$. 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 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 the 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 electron energy scale is adjusted to the scale measured in data using the Z boson mass peak in $Z \rightarrow ee$ events.

Similarly, corrections are derived for the efficiency of triggering and identification efficiency of 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 correction of the missing transverse energy's magnitude and resolution in the embedding samples is needed to account for the incomplete removal of energy deposits from muons replaced by simulated 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 $|\eta| > 2.0$ during the 2016 and 2017 data taking. This effect is named prefiring. This resulted in a loss of efficiency for events containing an electron or jet with p_T larger than $\approx$50 or $\approx$100)

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

Corrections on the energy of jets are calculated in bins of the jet p_T and 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. A MC to data correction for a b jet to pass the selection criteria is also determined. This corrections can adjust the number of b jets in a simulated event.

Any differences from simulated events to data, where an electron or muon is reconstructed as a 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 dilepton mass and p_T spectra in simulated $Z \to ll$ events. These are derived from $Z \to \mu\mu$ events. Additionally, all simulated $t\bar{t}$ events are weighted to match the top quark $p_T t$ distribution observed in data.

4.9 Uncertainty Model

The statistical uncertainties are taken into account by the Barlow-Beeston method, described in Ref. [42, 43]. The systematic model is split into uncertainties based on the online and offline reconstruction of objects and the background and signal modelling. An uncertainty is correlated across channels when it represents a shift on the reconstruction of an object and decorrelated otherwise. It is decorrelated across era of data taking when the shift is derived independently by era. The embedded samples use the same uncertainty scheme as MC but 50% are correlated and 50% uncorrelated with MC uncertainties, because of the shared real data in the measurement.

Hadronic Taus

Uncertainties on the 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 in the p_T regions < 100 GeV and > 100 GeV to allow the fit more freedom to adjust the high p_T regions relative to the low p_T regions. Uncertainties are also applied on the energy scale of the τ_h candidates. These uncertainties range between 0.2 and 1.1%. Finally, an uncertainty on the ID efficiency is placed as a function of p_T in the $e\tau_h$ and $\mu\tau_h$ channels, and of the τ_h decay mode in the $\tau_h\tau_h$ channels. These vary between 3-9% and is uncorrelated in each variable bin it is derived in. To account for the different anti-lepton discriminator working points, an uncertainty of 3% per τ_h is applied and treated as uncorrelated between the channels where different $D_{\mathrm{WP}}^{e/\mu}$ are used.

Light leptons

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

Jets

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

Leptons misidentified as hadronic taus

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

Jets misidentified as hadronic taus

The backgrounds with jets misidentified as 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, the shifts are estimated by comparing the fits to a 1st order polynomial fit set to constant above 100 GeV. The final systematic variation are then on the extrapolation corrections by applying the corrections twice and not at all. The size of each systematic uncertainty varies from 0–10%, whilst the statistical element from the fits can be larger in the tails of the distributions.

Jets misidentified as light leptons

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

Muons misidentified as electrons

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

MET

The MET uncertainties is different dependent on process. For all processes that are not $t\bar{t}$ and diboson, the hadronic recoil response and its resolution are varied within the uncertainties determined during the computation of the recoil corrections. For $t\bar{t}$ and diboson an uncertainty is derived on the energy carried by an unclustered particle [44]. These uncertainties vary between 0–10%.

Background process specific uncertainties

Uncertainties on the $t\bar{t}$ p_T and DY m_{ll}-p_T reweighting is placed by the 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 [37], 6% for $t\bar{t}$ production [38, 39], and 5% for diboson and single t quark production [39, 40, 41].

Signal process specific uncertainties

For the ggϕ process, in particular for low mass hypotheses, the variation of `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 (top, bottom, top-bottom interference). These are also determine from additional samples produced up to the generator level, and applied as event weights dependent on generator level p_T after the parton shower simulation. 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 %, 2.5 %) normalisation luminosity uncertainty is applied to the (2016, 2017, 2018) templates which originate from MC simulation.

Prefiring

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

4.10 Signal Extraction

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

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

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 to a signal to a specific hypothesis. The form of the Poisson probabilities are,

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

Finally, $p(\hat{\theta} \mid \theta)$ represents the probability density function (pdf) of each nuisance parameter (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 variations are derived and vertical morphing [43] 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: pdfs used for nuisance parameters.

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

The background-only fit is performed with N_j (number of signal processes) set to 0. For all signal-plus-background fits, the fit is done with respect to a single mass hypothesis, however within this mass hypothesis can be a number of signal processes. The model independent resonance search has separate $gg\phi$ and $bb\phi$ signal modes and so two rate parameters $\mu_{gg\phi}$ and $\mu_{bb\phi}$ are needed. When the samples are initially scaled to the cross section times branching ratio ($\sigma \times B(\phi \to \tau\tau)$) of 1 pb and $g(\mu) = \mu$ for both processes, $\mu_{gg\phi}$ and $\mu_{bb\phi}$ represent the $\sigma \times B(\phi \to \tau\tau)$ with units of pb. To avoid negative signal strengths, μ will only be taken to be positive. Also used in the following subsections, is a signal-plus-background channel/category compatibility fit. In this fit the signal processes and rate parameters are further split in each channel or category utilising index i in Eq. 4.10. This is used to determine the compatibility of the results in different decay channels and analysis categories. In this case, 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 Drell-Yan. However, as this is a model-dependent interpretation of these results both these rate parameters scale together. The scaling functions differ between the two processes with $g_{\text{t-channel}}(\mu) = \mu^4$ and $g_{\text{interference}}(\mu) = \mu^2$ to mimic how the cross sections of each process scales. When the initial samples are scaled to cross section at $g_U = 1$, 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 the separate loop contributors, so the kinematic properties can be properly scaled to MSSM prediction, as described in Section 4.1.1. The SM-like Higgs boson is considered in the MSSM signal model to monitor differences in the observed Higgs boson prediction between the MSSM and the SM. In each benchmark scenario chosen, the signal prediction depends only on m_A and $\tan\beta$ and the scaling to cross section is shown in Eq. 4.1. As the potential scaling functions for MSSM interpretations are not necessarily smooth one-to-one mappings, the likelihood is tested for individual points on the m_A-$\tan\beta$ parameter space. At each point, the MSSM Higgs bosons are scaled to the theory predicted cross section times branching ratio. To test the MSSM hypothesis over the SM hypothesis, the single rate parameter 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 regions with two degrees of freedom (as in the model independent resonant search) are determined by $-2\Delta \ln \mathcal{L} = 2.28$ and 5.99 respectively.

Upper limits are placed using the modified frequentest approach [45, 46] 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. $\hat{\theta}_\mu$ are the values of θ that is 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 [47] and results are cross-checked and deemed consistent with toy MC datasets.

If a deviation from the background expectation is observed, the size of the deviation is quantified by a significance. To test rejection of the background-only hypothesis in favour of the signal-plus-background hypothesis, 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 is 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 VLQ (BM 1) 1 TeV mass points are displayed.

In the low-mass optimisation categories, a small excess of events is observed on the Z boson peak in the no b tag categories and reasonable agreement is observed in the b tag categories. The excess of events are distributed in $m_{\tau\tau}$ between 80 and 120 GeV. A signal-plus-background hypothesis is best fit with a 100 GeV 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 an small excess of data events over the Z boson peak.

In the high-mass optimisation categories, another small excess is observed in high m_T^{tot} bins, particularly in the most sensitive no b tag categories. This excess is best fit by a model independent gluon fusion resonant mass at 1.2 TeV with a cross section times branching ratio of 3.1 fb. There are no considerable differences observed in background modelling between signal-plus-background and background-only fits. This is as the uncertainties in these bins are more statistically dominated and the majority of the systematic uncertainties are constrained in the bulk of the distribution. Good agreement is observed in the rest of the distribution. There is a very small deviation in the b tag categories, but as this can also be explained by a 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 absence of a signal for the search for a ggϕ or bbϕ resonance and shown in Figure 4.21. In each case, the other process is allowed to float freely in the fit. The excesses observed in the postfit distributions act to weaken the observed limit compared to the expected limit at 100 GeV and 1.2 TeV, as more data was observed than expected. For ggϕ production the expected limits flatten under 100 GeV, due to difficulty of separating signal from the Z boson at this mass. Both sets of limits vary from $\mathcal{O}$(10 pb) at 60 GeV to 0.3 fb at 3.5 TeV.

95% 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 branching fraction, as all channels in this mass range have similar signal separation ability. In the high mass optimisation categories, he 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 are also made with the ATLAS experiment and in particular the results presented in Ref. [3]. This ATLAS search looks for the same signal but over a smaller mass range, from 200 GeV to 2.5 TeV. Plots showing the comparison of the expected and observed limits for 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 $Z \to \tau\tau$ samples compared to MC allow for lower background uncertainties and hence a stronger limit. The ATLAS result observed no excess of events compatible with ggϕ signal at 1.2 TeV, in fact a small deficit was observed. Also, ATLAS observed local excesses at 400 GeV of 2.2σ for ggϕ and 2.7σ for bbϕ. None of these

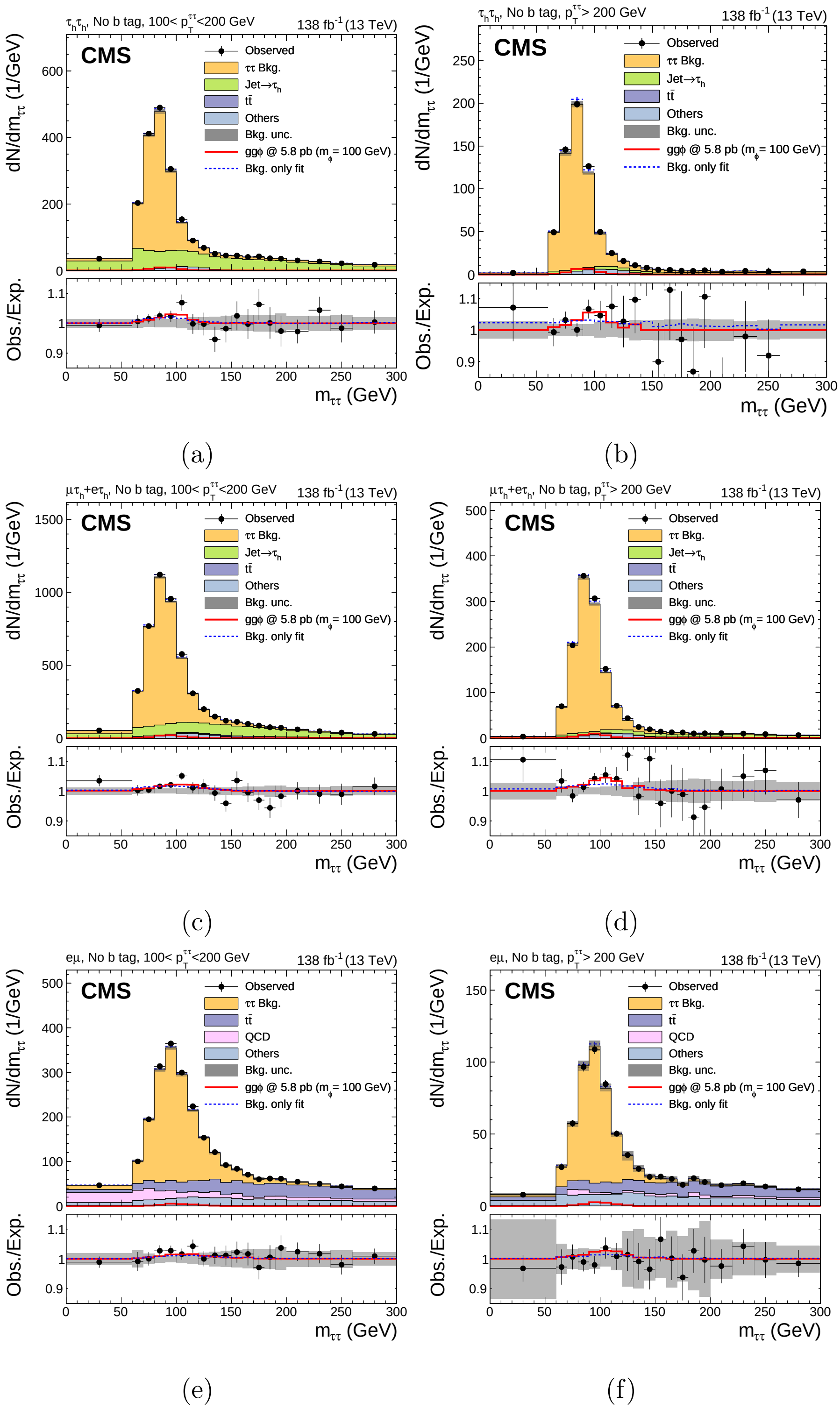


Figure 4.19: Distributions of $m_{\tau\tau}$ in the no b tag second highest (left) and highest (right) p_T category for the $\tau_h\tau_h$ (top), the combined $e\tau_h$ and $\mu\tau_h$ (middle) and the $e\mu$ (bottom) channels. The solid histograms show the stacked background predictions after a signal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

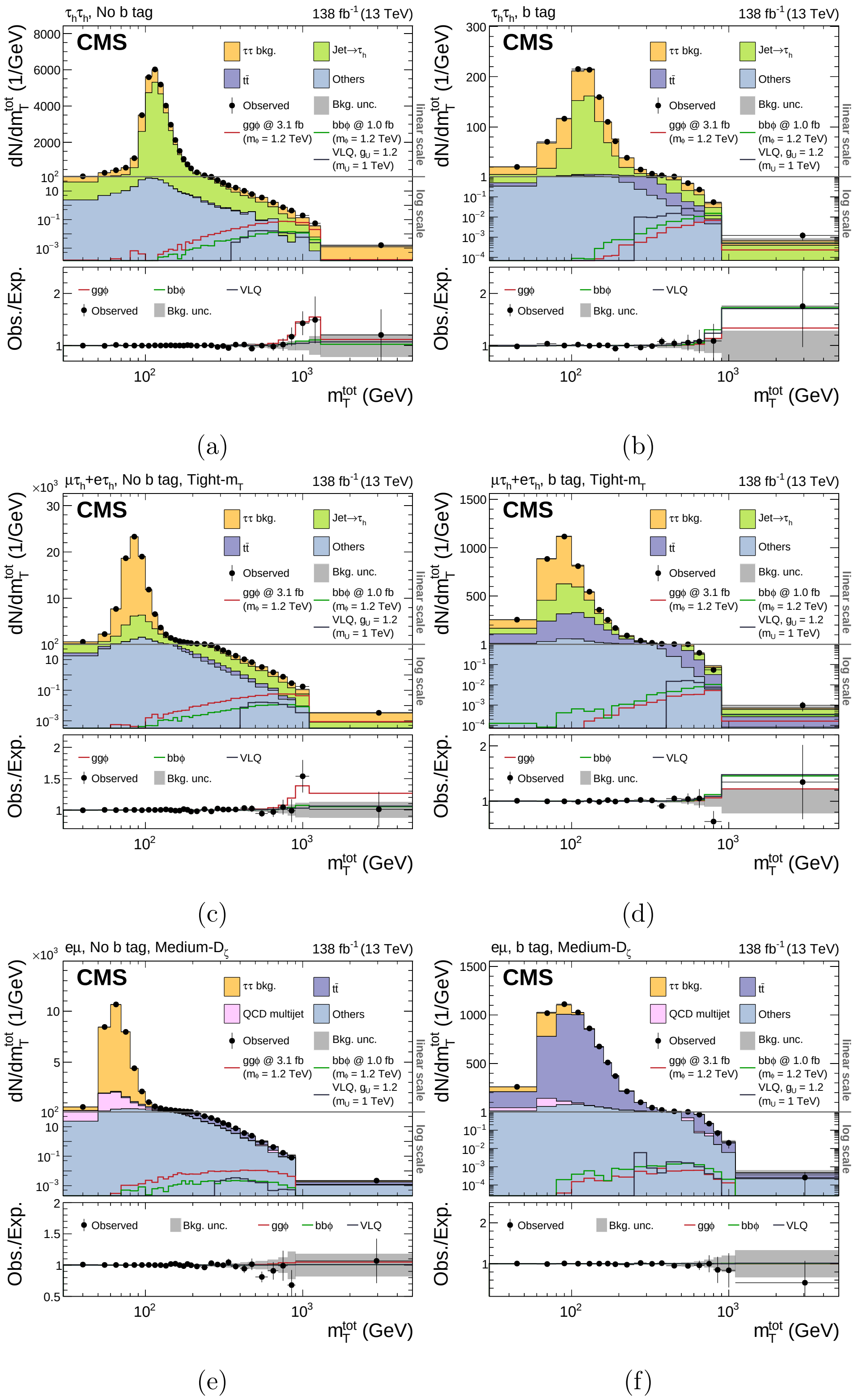
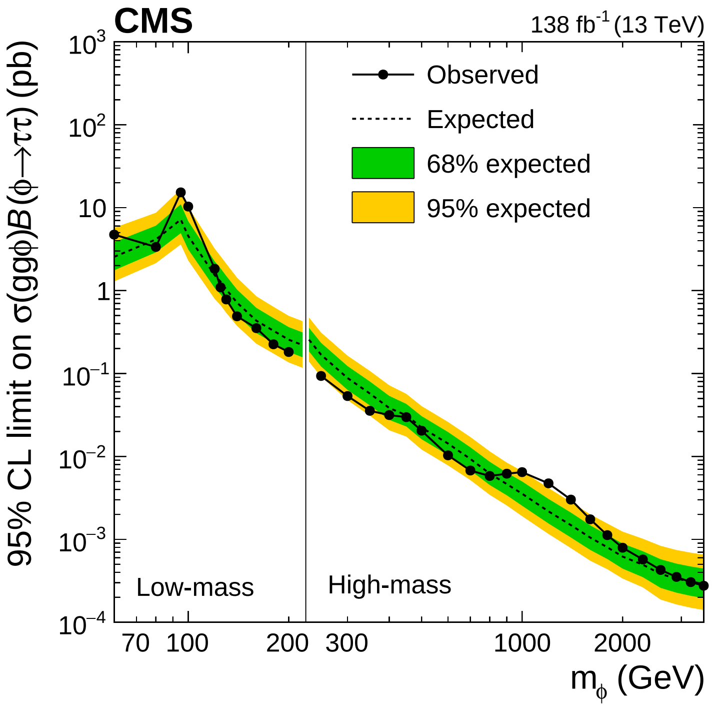
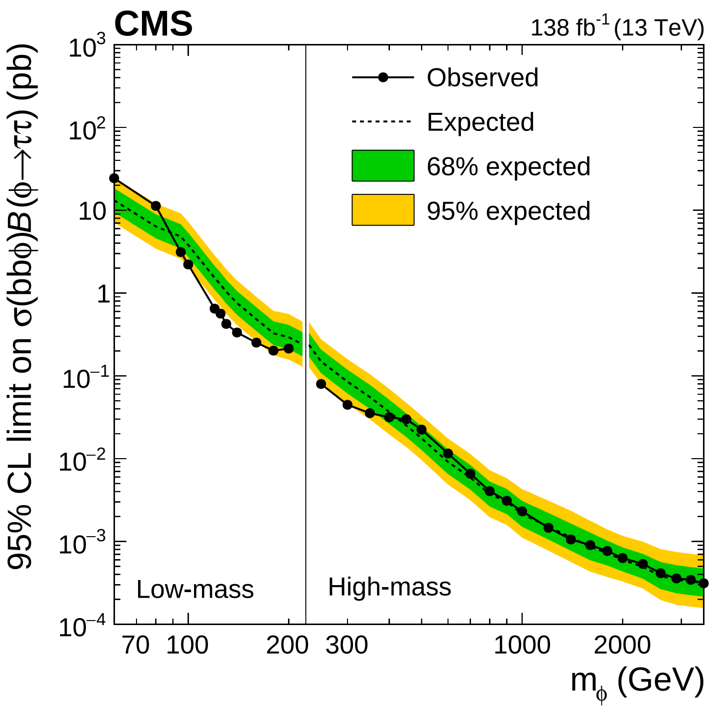


Figure 4.20: Distributions of m_T^{tot} in the $\tau_h\tau_h$ no b tag (a) and b tag (b) categories, the combined $e\tau_h$ and $\mu\tau_h$ no b tag (c) and b tag (d) Tight-m_T categories and the $e\mu$ no b tag (e) and b tag (f) Medium-D_ζ categories. The solid histograms show the stacked background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

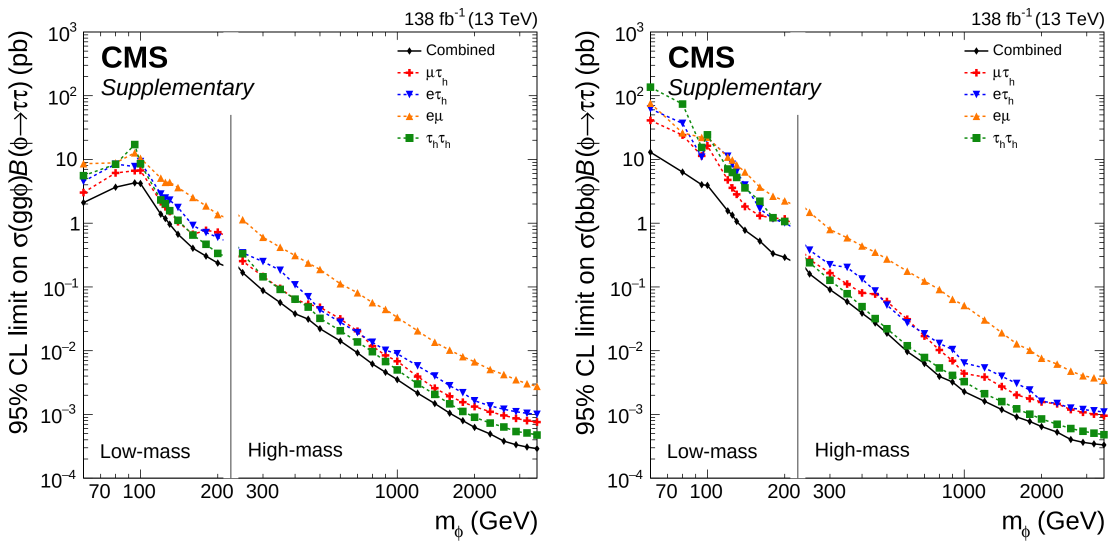


(a)



(b)

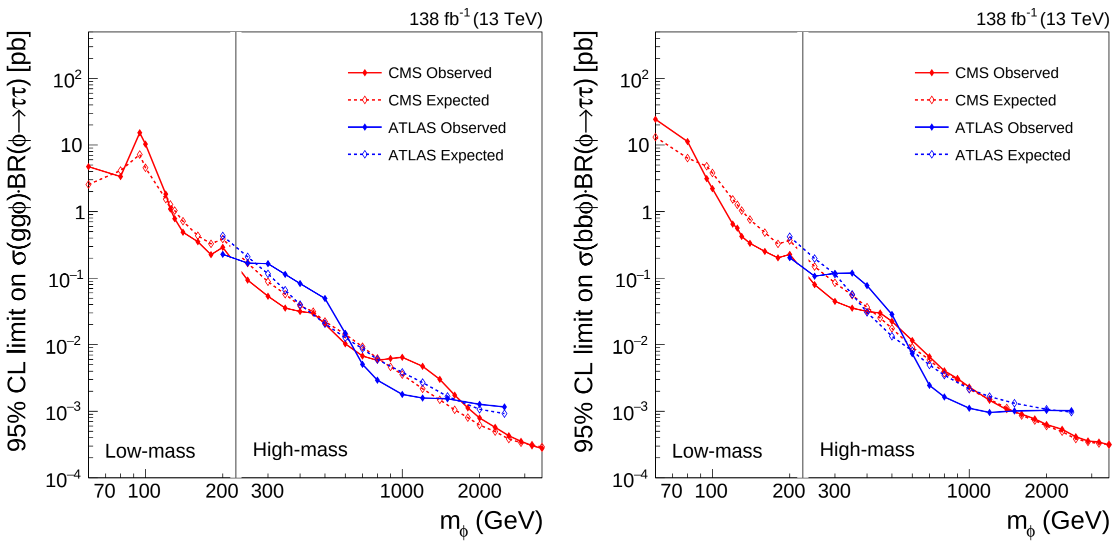
Figure 4.21: Expected (dashed line) and observed (solid line and dots) 95% CL upper limits on the product of the cross sections and branching fraction for the decay into τ leptons for (a) ggϕ and (b) bbϕ production in a mass range of 60 $\leq m_\phi \leq$ 3500 GeV. The dark green and bright yellow bands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



(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 (a) ggϕ and (b) bbϕ production, split by the $\tau\tau$ decay products fit individually.

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

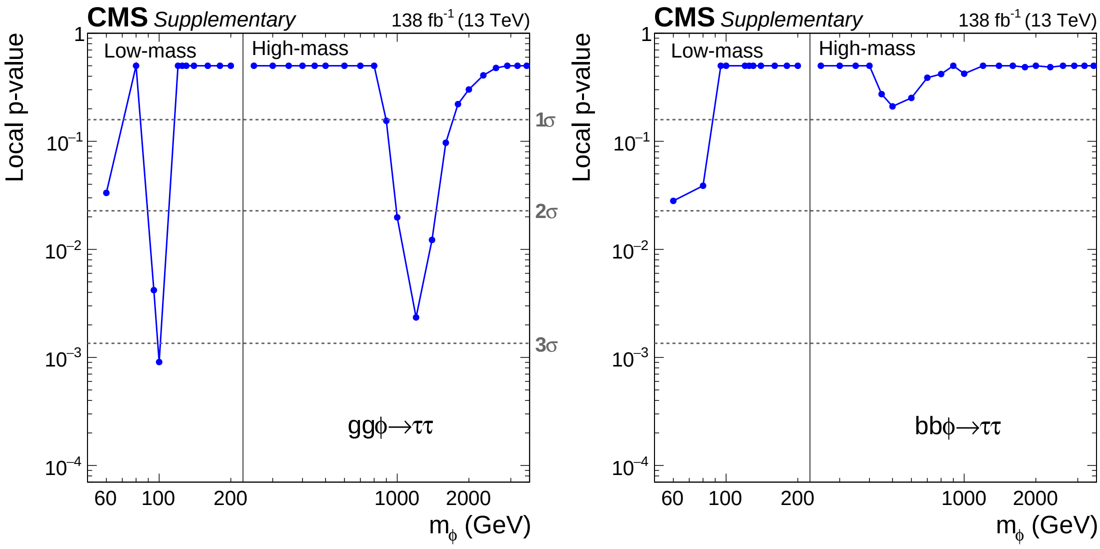


(a) (b)

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 (a) ggϕ and (b) bbϕ production, split by the CMS result detailed in this thesis and the ATLAS result from Ref. [3].

4.12.2 Significance and Compatibility

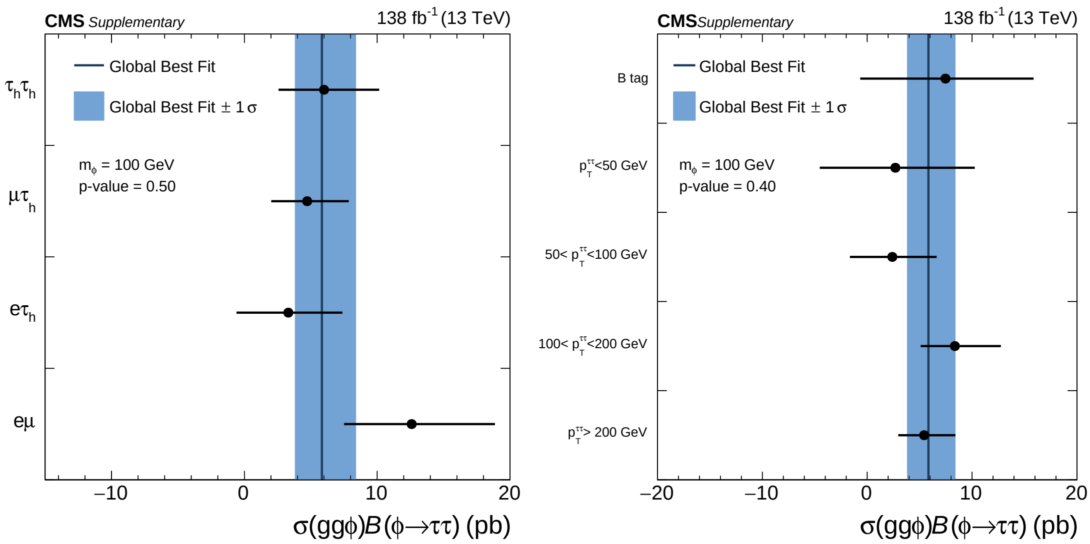
The p-values and significances at each model independent signal hypothesis are calculated as described in Section 4.10 and shown in Figure 4.24. Identical to the model independent limits, the ggϕ or bbϕ process is allowed to float freely if not the parameter of interest. The excesses for the ggϕ process 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.



(a) (b)

Figure 4.24: Local p-value and significance of a ggϕ (left) and bbϕ (right) signal as a function of m_ϕ.

As many different decay channels and categories are used to extract these significances, the signal strength is studied in each channel and category. This is done via compatibility fits as described in Section 4.10, where the signal strength parameter in each channel/category is decoupled. No statistically significant differences are observed in the best fit signal strength in any decay channel or category fit and p-values between each channel or category fit are always above 0.05. Figure 4.25 shows the results of the compatibility fits in the low-mass optimisation categories split by di-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.

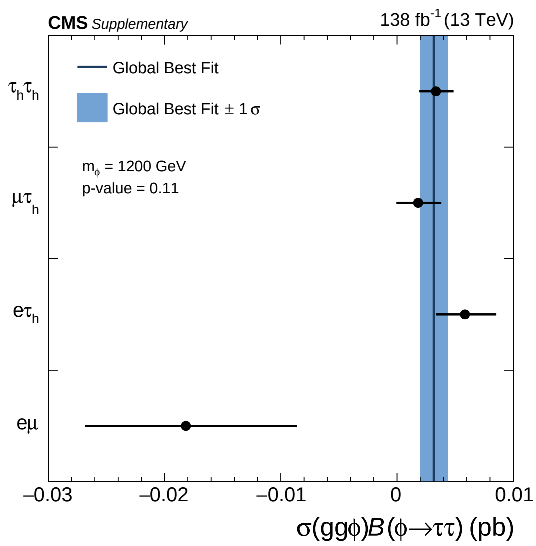


Figure 4.26: Compatibility plots of the 1.2 TeV excess split by analysis channels. In each case the fitted signal strength is decoupled in each channel.

4.12.3 2D Likelihood Scans

As the model-independent search looks for two signal modes at each mass point, the results for both processes happening simultaneously are studied. This is done in the form of 2D likelihood scans. The best fit cross section times branching fractions of each process and the 95% and 68% confidence intervals are shown for a number of different mass scenarios in Figure 4.27. The SM prediction in all plots is at (0,0). These results highlight how the excesses at 100 GeV and 1.2 TeV are dominated in the phase space in which 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, $\tan\beta$) 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 4.28 (b) 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 Refs.[48, 49, 50], for the scenarios described in Ref.[16].

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 decays into two on-shell top quarks becomes kinematically accessible. Both scenarios are excluded at 95% CL for $m_A \lesssim 350$ GeV. For $m_A \lesssim 250$ GeV, most of the ggH/A events do not enter the no b tag categories due to the $m_{\tau\tau} > 250$ GeV

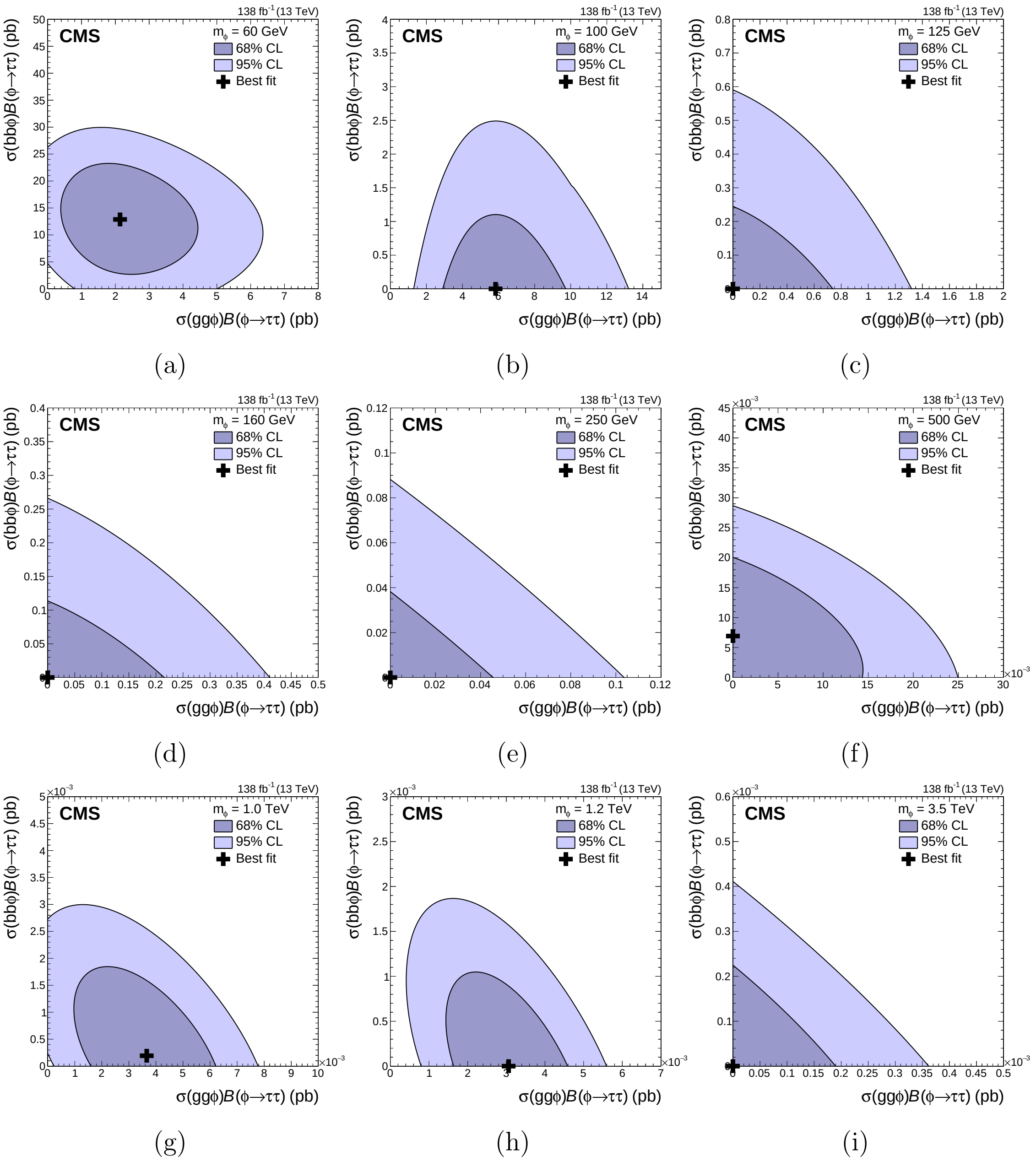


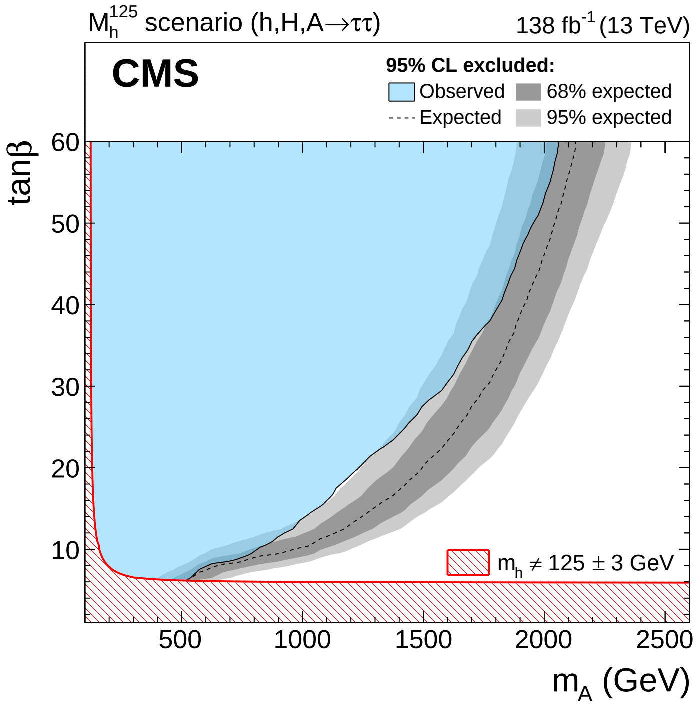
Figure 4.27: Maximum likelihood scans, including 68% and 95% CL contours obtained from the signal likelihood for the model-independent search. The scans are shown for selected values of m_ϕ between 60 GeV and 3.5 TeV.

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

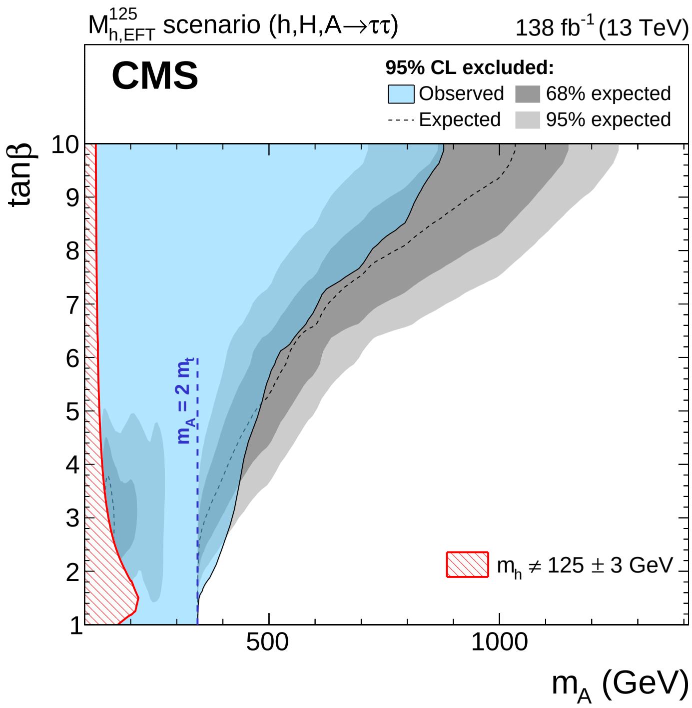
Other MSSM scenarios are tested and detailed in [51]. One scenario of note is the M_H^{125} scenario, which is the equivalent scenario to the M_h^{125} but with the observed Higgs boson being the heavier CP-even Higgs boson. Despite the local excess at a resonant mass of 100 GeV, this scenario is entirely excluded by the search. This is mostly due to the sensitivity of the b tag categories to b associated production. The local excess observed at 1.2 TeV is hard to rectify within these MSSM benchmark scenarios. The lack of any excess in the b tag categories strictly constrains the b associated production cross section times branching fraction. It is not possible within these scenarios to predict the excess of gluon fusion events within the constraints placed on b associated production.

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

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

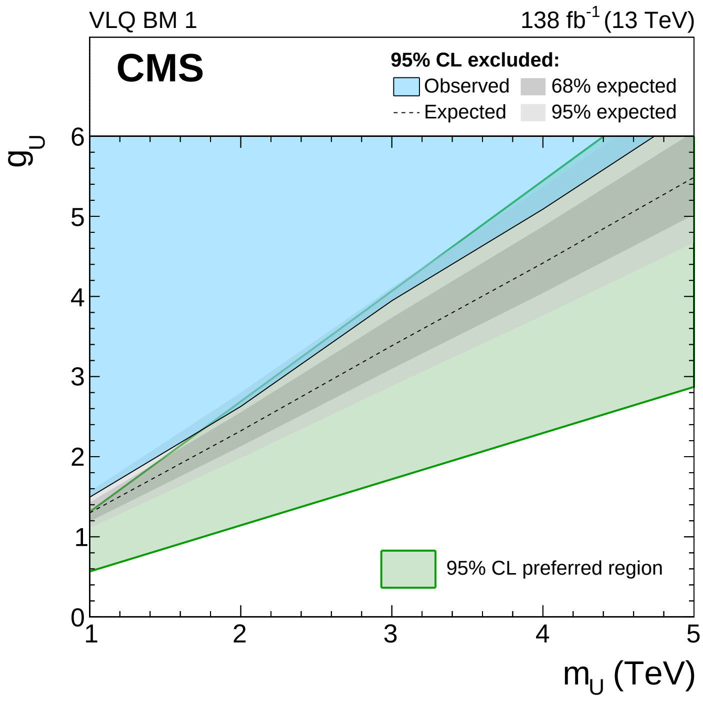


(a)

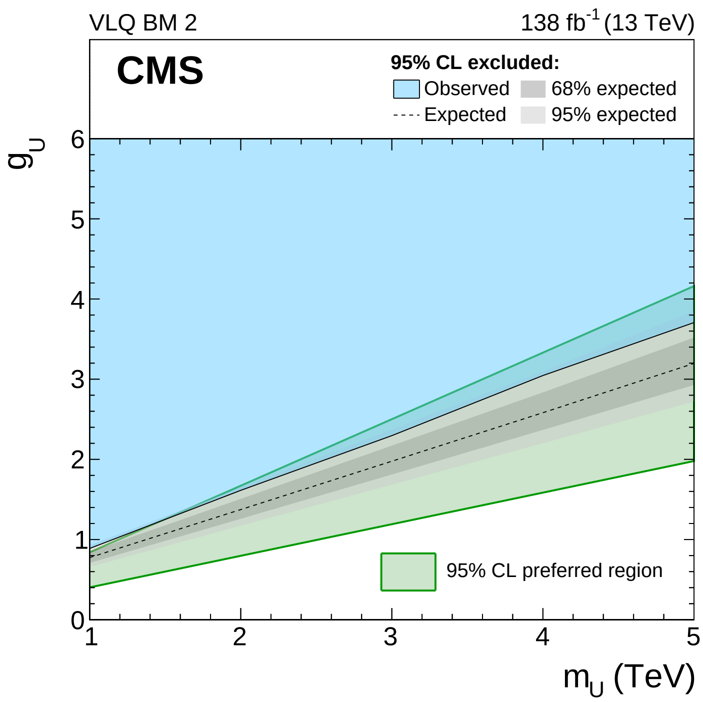


(b)

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 then ± 3 GeV from the observed SM Higgs boson mass it shown by a red hatched area.



(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 background expectation is shown as a dashed black line, the dark and bright grey bands show the 68% and 95% intervals of the expected exclusion and the observed exclusion contour is shown by the blue area. The 95% confidence interval for the preferred region from the global fit presented in Ref. [4] is also shown by the green shaded area.

Chapter 5

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

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

ii) A search for the type X 2HDM, motivated by the phase space for possible explanations to the g-2 anomaly. The m_A-$\tan\beta$ phase space for scenarios of m_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 tau leptons. To probe the type X 2HDM at high $\tan\beta$, a production process that is not suppressed is required. Ref. [?] discusses that the following production modes of two additional Higgs bosons are dominant to produce any of these new particles at high $\tan\beta$:

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

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

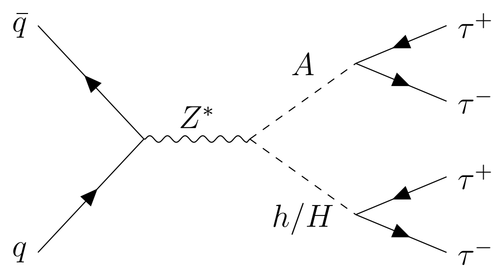


Figure 5.1: Diagram of production of two additional neutral Higgs bosons from an off-shell Z boson and their decays to tau leptons.

Signal templates for the production of this process with a mass grid for ϕ and A between 100 to 300 and 60 to 160 GeV respectively are generated. These mass ranges are motivated from the results in Table 1.4. The samples are simulated in the five-flavour scheme (5FS) at NLO precision using the MadGraph5_aMC@NLO v2.6.5 event generator [29]. Generation is performed using the parton distribution function (PDF) NNPDF3.1 [23, 24], where τ lepton decay, parton showering and hadronisation are all modelled with the Pythia event generator with the PU profile matched to data [25, 26]. The events are then passed through the GEANT4-based [27] 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 **??**.

FIGURE OF GENERATOR LEVEL MASSES

The cross sections in the alignment scenarios are also determined with this procedure and vary from 10 fb (highest m_A and m_ϕ scenario) to 650 fb (lowest m_A and m_ϕ scenario), as shown in Figure 5.2. 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 $\cos^2(\beta-\alpha)$.

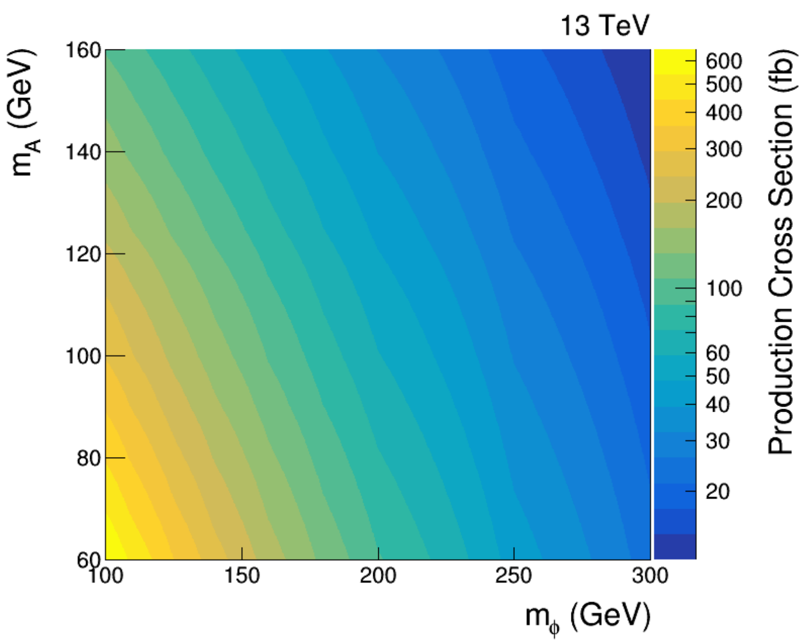
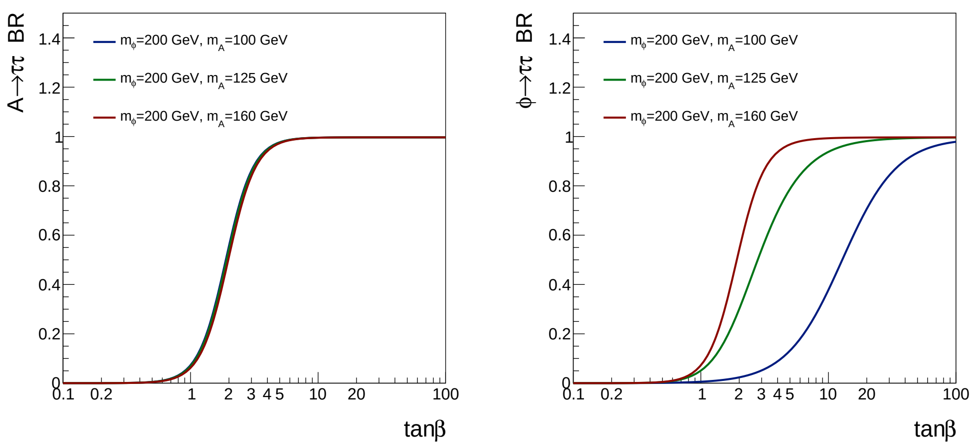


Figure 5.2: Calculated productions cross sections for the $Z^* \to \phi A$ process, varying the masses of ϕ and A.

The branching fractions of ϕ and A to pairs of τ leptons is dependent on both $tan\beta$ and $\beta - \alpha$. For this analysis, the branching fractions are calculated using `2HDECAY` [?]. 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 5.3. Out of the alignment scenario, the branching fractions of ϕ to tau leptons becomes smaller as the magnitude of the coupling of additional neutral Higgs bosons to taus is reduced, whilst the A branching fractions, like the couplings, are left unchanged. An example of the ϕ branching fractions out of the alignment scenario is shown in Figure 5.4.



(a) (b)

Figure 5.3: Calculated branching fractions of A (a) and ϕ (b) decaying to a pair of τ leptons for various mass scenarios.

5.2 Event Selection

In comparison to Chapter 4, four τ leptons produce a much larger number of possible final states. All possible variation 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 are explored in this analysis. In addition to this, an orthogonal $\tau_h\tau_h\tau_h$ channel is added to target events where reconstruction of the $\tau_h\tau_h\tau_h\tau_h$

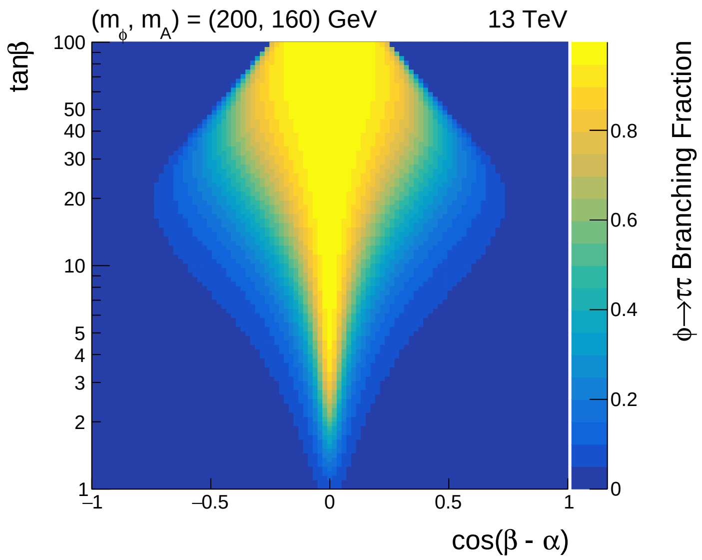


Figure 5.4: Calculated branching fractions in the $\cos(\beta - \alpha)$-$\tan\beta$ phase space for ϕ of mass 200 GeV decaying to a pair of τ leptons, in the scenario where $m_A = 160$ GeV.

channel loses a single τ_h object. This can come about due to low triggering and ID efficiency of τ_h candidates, as well as the high p_T thresholds required for both.

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

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 SingleElectron and SingleMuon triggers. This is not the case for the SingleTau trigger as it has a p_T threshold too high for it to be useful. The possible triggers for clusters of objects are the DoubleElectron, DoubleMuon, DoubleTau, Electon-Muon cross, Muon-Tau cross and Electron-Tau cross. The p_T thresolds of the mentioned triggers are shown in Table **??** and the corresponding HLT paths used for each year are shown in Tables **??**-**??**.

The decision on which trigger to use is a balance between simplicity, due to the corrections that need to be applied, and trigger acceptance. The trigger acceptance of a number of trigger combinations in a number of different channels were calculated and shown in Table **??**. Firstly, the SingleElectron and SingleMuon triggers were chosen for the signal acceptance table over the DoubleElectron and DoubleMuon triggers in the $ee\tau_h\tau_h$, $\mu\mu\tau_h\tau_h$. This is motivated by the lack of sensitivity we expect to have in these channels, as they have the small branching fractions, and backgrounds from Z + 2 jet fakes and ZZ events including leptons would be larger than in other channels. Using the single triggers in this channels would allow them to be more consistent with other more sensitive channels including leptons and correlate relevant trigger uncertainties.

5.2.2 Offline Requirements

Deeptau working points
On top of the object selection, there are few further selections on the total tau collection.
Charge
As the two tau pairs from the signals originate from two neutral additional Higgs bosons, the sum of charge of the fully reconstructred objects should be 0. This is applied in the decay channels where there are four objects. In the $\tau_h\tau_h\tau_h$ the assumption is that a hadronic tau has been lost, so the charge selection required here is that the absolute value of the sum of the charges of the object must be 1.
Event Vetos
In order to ensure the orthogonality between channels, a number of vetos are needed. The constraints set on the number of leptons for each channel are shown in Table **??**.

- Extra Lepton Veto
  - Extra lepton vetos are used in all decay channels.
  - Orthogonality is achieved by requiring the absence of additional electrons and muons on top of those already part of the selected pair.
  - The kinematic requirements, ID and isolation requirements on these extra leptons are at least as loose as the loosest cuts required for signal electron or muon of the final states.
- Extra Hadronic Tau Veto
  - The $\tau_h\tau_h\tau_h\tau_h$ and $\tau_h\tau_h\tau_h$ are the only channels now not orthogonal after extra lepton vetos.
  - In the $\tau_h\tau_h\tau_h$ channel, any extra hadronic tau candidate passing the signal selection is vetoed.

5.3 Search Optimisation

TALK ABOUT CATEGORY OPTIMISATION

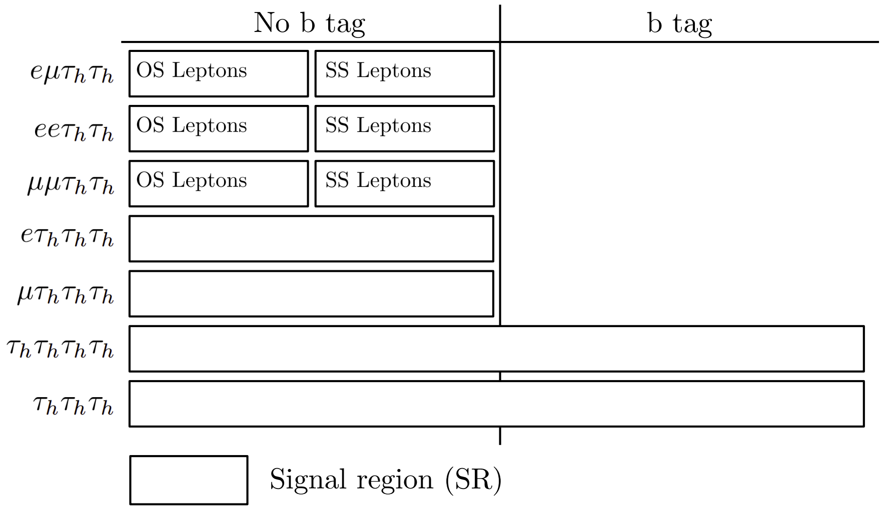


Figure 5.5: Overview of the categories used for the extraction of the signal in the $Z^* \to \phi A \to 4\tau$.

The variable m_T^{tot} is defined as:

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

5.4 Background Modelling Overview

The background is split into two components in each decay channel. These are events where at least one hadronic tau is a jet fake and the remaining contribution dominated by events where all objects are correctly selected. The former is derived using a machine learning fake factor approach as described in Section **??** and the latter is modelled by MC and modelling checked with a control region as described in Section **??**.

5.5 ZZ Modelling

5.6 Machine Learning Fake Factor Method

5.6.1 BDT Reweighter

Ref. [?] proposes a new method utilising machine learning techniques to solve the issues with the previous approach. It looks to optimise the regions that most need reweighting. A 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.

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

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

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

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

5.6.2 Fitting Regions

Unlike the standard fake factor method, statistics are not a problem using the BDT reweighter when choosing which variables you can use to parameterise the fake rate. Therefore rather than fitting two fake factor regions separately and then a correction from sideband to signal to account for missing parameterisation, we can fit the fake rate in regions B,C and D from Figure **??** simultaneously to improve statisitics.

The variable used for the sideband variable where we have four objects is whether the total charge is 0 or not. In the $\tau_h\tau_h\tau_h$ channel this is whether the absolute value of the total charge is 1 or not. More than one alternative sideband variables are used in this analysis, this matches what is done for the di-tau fully hadronic fake factors, using whether other hadronic taus (that you are not fitting the fake rate for) pass or fail the tau ID. The split is defined as whether all hadronic taus pass the tau ID and where one or more taus fail the tau ID.

Every hadronic tau in the event is treated as a separate entry in the BDT reweighter. As the Loose DeepTauVsJets working point is used for the tau ID, an additional looser baseline selection of rawDeepTauVsJetsScore $\geq$ 0.1 for each hadronic tau is applied to keep the fail and pass tau ID regions not too far separated from one another. Each channel is fitted separately but all years are combined into a single fit.

Using this method for deriving the regions used for fitting there would be some overlap in the $\tau_h\tau_h\tau_h\tau_h$ channel 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 fakes background. Instead the fake factors for the $\tau_h\tau_h\tau_h\tau_h$, use the fit from the $\tau_h\tau_h\tau_h$ channel. There are a few variables that are defined differently in the fit between the two channels due to the difference of 4 to 3 objects selected, therefore when getting fake factors in the $\tau_h\tau_h\tau_h\tau_h$, the lowest DeepTauVsJets scoring unused hadronic tau candidate is dropped and variables recalculated. This candidate is chosen to be removed to best mimic the initial selection of the hadronic tau candidates where they are sorted by DeepTauVsJets score and the high scoring candidates are chosen. As shown later in this section, there is no major dependence on the shifted variables and any effect from this removal is covered within the uncertainty model.

After this looser baseline tau ID is applied, following the logic stated above the below regions are reweighted onto one another to provide the fake factors. Each hadronic tau is labelled numerically and whether it passes the Loose WP of hadronic tau i as P_i. The total bracket with its index shown, means the features of that hadronic tau is used.

5.6.3 Machine Learning Subtraction Method

In the standard fake factor method, histograms are used to fit the fake factors rather than datasets. Using histograms, the small fraction of events which are not jet fakes can easily be subtracted off. To do this, the data histogram is subtracted from by a stacked MC background produced with generator matching ensuring the event is not a jet fake and this produces a data-MC hybrid histogram of predicted jet fake 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 fake generator matched MC. An example of this is template matching, which takes an event and can find the closest event in another datasets. But as the fitting dataset is highly dimensional, this requires too much computation. The solution proposed for this is to use a BDT to reduce the dimensionality of the datasets effectively the one dimension of an output score of a simple binary classifier.

- In each channel, all MC in the fitting region is stacked and scaled to cross section (via weighting) and the variables used for reweighting (Section **??**) are put into a dataset.
- Events that are jet fakes and not jet fakes are separated into the two classes that will be used for the binary classification.
- To ensure unbiased training, the weights of the two categories are normallised to one another.
- A BDT is then trained to separate whether the MC is jet fake or not.
- The scores of the BDT for how likely the event is not a jet fake is added to the dataset.
- The scores of the non jet fake events are drawn into a histogram with a number of bins suitable for the number of statistics and rescaled to cross section to best match what would be observed in data.
- The output score of the BDT is added to data events
- Each bin of the MC histogram is then looped through:
  - Data events with BDT score within the range of the bin are selected.

  - Events within this bin are then randomly sample and removed.
  - This stops when the number of events removed equals to the number of non jet fake events predicted in the MC histogram bin.

This method then gives a data-MC hybrid method to determine a dataset of predicted jet fake events and should given near identical results when drawing out histograms in all variables compared to subtracting off MC non jet fakes events from a data histogram as used in the original fake factor method. It allows for non jet fake-like events to be sampled and removed to the correct yield through all variables. It is important to remove events throughout the MC non jet fake 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 jet fake and non jet fakes. To validate this method, histograms are shown in Figures **??**–**??**, comparing the BDT subtraction method and histogram subtraction method for a few of the fitted variables.

5.6.4 Fitting

The variables used to fit the BDT reweighter on the subtracted datasets are variables that have been shown to have fake rate dependence previously. Also added are the variables that take you from B to A and C to A, from Figure **??**. As all years are fit together, in order to account for any differences in fake rate from year to year, this is also added. All these variables are shown below.

- The HPS decay mode of the hadronic tau candidate.
- p_T of the hadronic tau candidate.
- The ratio of the p_T of the matched jet to the p_T of the hadronic tau candidate.
- η of the hadronic tau candidate.
- The charge of the hadronic tau candidate.
- A booleon of whether the hadronic tau candidate passes a leg of the double tau trigger.
- The total charge of the combined objects.
- The DeepTauVsJets score of the other hadronic tau candidates. These are sorted by p_T.

- Year of data taking.

These datasets are finally randomly split 50:50 into a train and test datasets and only the train dataset is fit. The BDT reweighter has a number of hyperparameters and these are tuned with a grid scan optimising the Kolomogrov-Smirnoff test on the test dataset in each channel separately.

5.6.5 Applying Fake Factors

Fake factors, F_F^i, have now been calculated for each hadronic tau candidate, i, in the event and uncertainties determined on each weight. The question is how to combine them to generate a full jet fakes background in the signal region. Taking channels with two hadronic taus as the simplest example, if the jet fakes background is determined purely off the leading hadronic tau candidate, i.e. the jet fakes background is determined by $F_F^1(q_{sum} = 0\&!P_1\&P_2)$, events where the leading hadronic tau is genuine and the sub-leading hadronic tau is a jet fake are missed. Similarly, the situation can be flipped if the sub-leading hadronic tau is chosen to determine the fake rate. A third options can be tried where both hadronic taus are used to determine the jet fakes background, i.e. $F_F^1 F_F^2(q_{sum} = 0\&!P_1\&!P_2)$. However, this will only model events where both hadronic taus are jet fakes and not where there is a single fake. It is seen that if the first two attempts at calculating a jet fakes background from individual candidates are added and the contribution from both hadronic taus is subtracted, all possible numbers of jet fakes in the event are accounted for. The following logic is used on all channels and total jet fakes formula is shown below for each channel. Again in the $\tau_h\tau_h\tau_h\tau_h$ channel, there would be overlap in this determination region with the $\tau_h\tau_h\tau_h$ signal region. As 0 events with exclusively 1 jet fake are predicted in MC, this is treated as negligible.

5.7 MC Corrections

5.8 Uncertainty Model

5.9 Signal Extraction

5.9.1 Postfit Plots

5.10 Model Independent Results

5.11 Model Dependent Limits

Chapter 6

Conclusion

6.1 Global Interpretations of Results

6.2 Outlook

Bibliography

[1] H. Bartosik and G. Rumolo, "Performance of the LHC injector chain after the upgrade and potential development," tech. rep., 2022. Contribution to Snowmass 2021.

[2] CMS Collaboration, "The CMS Experiment at the CERN LHC," *JINST*, vol. 3, p. S08004, 2008.

[3] G. Aad *et al.*, "Search for heavy Higgs bosons decaying into two tau leptons with the ATLAS detector using pp collisions at $\sqrt{s} = 13$ TeV," *Phys. Rev. Lett.*, vol. 125, p. 051801, 2020.

[4] C. Cornella, D. A. Faroughy, J. Fuentes-Martin, G. Isidori, and M. Neubert, "Reading the footprints of the B-meson flavor anomalies," *JHEP*, vol. 08, p. 050, 2021.

[5] S. P. Martin, "A Supersymmetry Primer," *Advanced Series on Directions in High Energy Physics*, p. 1–98, Jul 1998.

[6] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys. Lett. B*, vol. 716, pp. 1–29, 2012.

[7] CMS Collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC," *Phys. Lett. B*, vol. 716, pp. 30–61, 2012.

[8] A. Djouadi, L. Maiani, A. Polosa, J. Quevillon, and V. Riquer, "Fully covering the MSSM Higgs sector at the LHC," *JHEP*, vol. 06, p. 168, 2015.

[9] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector," *JINST*, vol. 12, no. 10, p. P10003, 2017.

[10] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV," *JINST*, vol. 13, no. 05, p. P05011, 2018.

[11] CMS Collaboration, "Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS," *JINST*, vol. 11, no. 01, p. P01019, 2016.

[12] CMS Collaboration, "Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2," Tech. Rep. CMS-PAS-TAU-16-002, CERN, Geneva, 2016.

[13] E. Bagnaschi, H. Bahl, E. Fuchs, T. Hahn, S. Heinemeyer, S. Liebler, S. Patel, P. Slavich, T. Stefaniak, C. E. M. Wagner, and G. Weiglein, "MSSM Higgs boson searches at the LHC: Benchmark scenarios for Run 2 and beyond," *Eur. Phys. J. C*, vol. 79, p. 617, 2019.

[14] H. Bahl, P. Bechtle, S. Heinemeyer, S. Liebler, T. Stefaniak, and G. Weiglein, "HL-LHC and ILC sensitivities in the hunt for heavy Higgs bosons," *Eur. Phys. J. C*, vol. 80, p. 916, 2020.

[15] H. Bahl, S. Liebler, and T. Stefaniak, "MSSM Higgs benchmark scenarios for Run 2 and beyond: the low $\tan\beta$ region," *Eur. Phys. J. C*, vol. 79, p. 279, 2019.

[16] E. A. Bagnaschi, S. Heinemeyer, S. Liebler, P. Slavich, and M. Spira, "Benchmark scenarios for MSSM Higgs boson searches at the LHC," Tech. Rep. LHCHWG-2021-001, CERN, 2021.

[17] CMS Collaboration, "Search for additional neutral MSSM Higgs bosons in the $\tau\tau$ final state in proton-proton collisions at $\sqrt{s} = 13$ TeV," *JHEP*, vol. 09, p. 007, 2018.

[18] V. Khachatryan *et al.*, "Search for neutral MSSM Higgs bosons decaying into a pair of bottom quarks," *JHEP*, vol. 11, p. 071, 2015.

[19] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms," *JHEP*, vol. 11, p. 040, 2004.

[20] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method," *JHEP*, vol. 11, p. 070, 2007.

[21] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX," *JHEP*, vol. 06, p. 043, 2010.

[22] T. Ježo and P. Nason, "On the treatment of resonances in next-to-leading order calculations matched to a parton shower," *JHEP*, vol. 12, p. 065, 2015.

[23] R. D. Ball *et al.*, "Parton distributions for the LHC Run II," *JHEP*, vol. 04, p. 040, 2015.

[24] R. D. Ball *et al.*, "Parton distributions from high-precision collider data," *Eur. Phys. J. C*, vol. 77, p. 663, 2017.

[25] A. M. Sirunyan *et al.*, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements," *Eur. Phys. J. C*, vol. 80, p. 4, 2020.

[26] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, "An introduction to PYTHIA 8.2," *Comput. Phys. Commun.*, vol. 191, p. 159, 2015.

[27] S. Agostinelli *et al.*, "GEANT4—a simulation toolkit," *Nucl. Instrum. Meth. A*, vol. 506, p. 250, 2003.

[28] 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," 2022.

[29] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, "MadGraph 5: Going beyond," *JHEP*, vol. 06, p. 128, 2011.

[30] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO," *JHEP*, vol. 12, p. 061, 2012.

[31] 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.*, vol. 513, p. 022035, 2014.

[32] 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." Submitted to *Eur. Phys. J. C*, 2022.

[33] J. Alwall, S. Höche, F. Krauss, N. Lavesson, L. Lönnblad, F. Maltoni, M. L. Mangano, M. Moretti, C. G. Papadopoulos, F. Piccinini, S. Schumann, M. Treccani, J. Winter, and M. Worek, "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions," *Eur. Phys. J. C*, vol. 53, p. 473, 2008.

[34] S. Alioli, S.-O. Moch, and P. Uwer, "Hadronic top-quark pair-production with one jet and parton showering," *JHEP*, vol. 01, p. 137, 2012.

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

[36] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method," *Eur. Phys. J. C*, vol. 71, p. 1547, 2011.

[37] K. Melnikov and F. Petriello, "Electroweak gauge boson production at hadron colliders through $\mathcal{O}(\alpha_s^2)$," *Phys. Rev. D*, vol. 74, p. 114017, 2006.

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

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

[40] J. M. Campbell, R. K. Ellis, and C. Williams, "Vector boson pair production at the LHC," *JHEP*, vol. 07, p. 018, 2011.

[41] T. Gehrmann, M. Grazzini, S. Kallweit, P. Maierhöfer, A. von Manteuffel, S. Pozzorini, D. Rathlev, and L. Tancredi, "W^+W^- production at hadron colliders in next to next to leading order QCD," *Phys. Rev. Lett.*, vol. 113, p. 212001, 2014.

[42] R. J. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples," *Comput. Phys. Commun.*, vol. 77, p. 219, 1993.

[43] J. S. Conway, "Incorporating nuisance parameters in likelihoods for multisource spectra," in *PHYSTAT 2011*, p. 115, 2011.

[44] A. M. Sirunyan *et al.*, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ using the CMS detector," *JINST*, vol. 14, p. P07004, 2019.

[45] T. Junk, "Confidence level computation for combining searches with small statistics," *Nucl. Instrum. Meth. A*, vol. 434, p. 435, 1999.

[46] A. L. Read, "Presentation of search results: The $\mathrm{CL_s}$ technique," *J. Phys. G*, vol. 28, p. 2693, 2002.

[47] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," *Eur. Phys. J. C*, vol. 71, p. 1554, 2011. [Erratum: doi:10.1140/epjc/s10052-013-2501-z].

[48] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs cross sections: 3. Higgs properties," Tech. Rep. CERN-2013-004, 2013.

[49] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs cross sections: 4. deciphering the nature of the Higgs sector," Tech. Rep. CERN-2017-002-M, 2016.

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

[51] CMS Collaboration, "Searches for additional Higgs bosons and vector leptoquarks in $\tau\tau$ final states in proton-proton collisions at $\sqrt{s} = 13$ TeV," 2022.