Search for supersymmetric neutral Higgs bosons decaying to τ pairs in the $e + \tau$-jet final state with calibration using $Z \to \tau\tau$ events at CMS

Michael Cutajar

Department of Physics, Imperial College London
2012

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

Abstract

A search for the neutral Higgs bosons, Φ, of the Minimal Supersymmetric Standard Model (MSSM) decaying to τ pairs in proton-proton collisions at $\sqrt{s} = 7$ TeV is performed. The $e + \tau$-jet final state is considered in data corresponding to 2.1 fb^{-1} of integrated luminosity collected by the Compact Muon Solenoid detector. The search is performed separately in the inclusive channel, sensitive to both $gg \to \Phi$ and $gg \to b\bar{b}\Phi$ production, and the b-tagged channel which maximises sensitivity to $gg \to b\bar{b}\Phi$. No significant excess of events over the predicted Standard Model backgrounds is observed. Upper limits are set on the $\Phi \to \tau^+\tau^-$ production cross section, $\sigma(pp \to \Phi + X) \times \mathcal{B}(\Phi \to \tau^+\tau^-)$. Cross sections greater than 18 and 1.3 pb are excluded at 95% confidence level for $m_A = 100$ and $m_A = 400$ GeV respectively. The limits are interpreted as constraints on tan β in the m_h^{max} scenario of the MSSM: Values of tan β greater than 12 and 40 are excluded at 95% confidence level for $m_A = 100$ and $m_A = 400$ GeV respectively. The inclusive channel is found to have the greatest tan β sensitivity at large m_A and the b-tagged channel to have greatest sensitivity at small m_A. The τ-jet reconstruction is calibrated using $Z \to \tau^+\tau^-$ events. A measurement of the $Z \to \tau^+\tau^-$ cross section in the $e + \tau$-jet final state using the same dataset as the Higgs search yields $\sigma(pp \to Z + X) \cdot \mathcal{B}(Z \to \tau^+\tau^-) = 1.06^{+0.05}_{-0.04}(\text{stat.}) \pm 0.07(\text{syst.}) \pm 0.05(\text{lumi.})$ nb. The result is in good agreement with the Standard Model prediction of 0.972 ± 0.042 nb at next-to-next-to-leading order. This measurement provides a cross check of the methods used in the $\Phi \to \tau^+\tau^-$ search.

Declaration

The studies presented in Chapters 4 and 5 of this thesis are my own work and represent my contributions to the CMS $Z \to \tau\tau$ and $H \to \tau\tau$ analyses (entries [52], [55], [93], [95] and [100] in the Bibliography). I was responsible for the development of the event selections for the $e + \tau_h$ channel in these analyses in collaboration with other members of the CMS $H \to \tau\tau$ group. The optimisation of the b-tagging selections in Section 5.2 is entirely my own work. I performed the first measurements of the OS/SS ratio and the QCD background shape and developed the data driven background estimation method for the $e + \tau_h$ channel as described in Section 4.4. I also performed the first measurement of the electron to τ-jet fake rate however the fake rate given in Section 4.3.4 was evaluated by other members of the CMS $H \to \tau\tau$ group with a larger dataset. In addition to these contributions, I developed independent code to cross check the entire $e + \tau_h$ analysis including the statistical interpretation.

The study of the $gg \to \Phi$ modelling in Section 5.4.1 represents my contribution to the Handbook of LHC Higgs Cross Sections (entry [40] in the Bibliography).

The measurement of the τ trigger efficiency in Section 4.3.2 and the correction factors given in Section 4.3.3 were performed by other members of the CMS $H \to \tau\tau$ group as referenced in the text. The measurement of the efficiency of the electron legs of the triggers in Section 4.3.2 is my own work. The information presented in Chapters 1, 2 and 3 is obtained from the Bibliography as referenced in the text. Graphics obtained from the Bibliography are referenced in the figure captions. In Chapter 6 I present the latest MSSM Higgs results from CMS which are the work of the CMS $H \to \tau\tau$ group.

Michael Cutajar

March 2012

Preface

The progress of Particle Physics towards the end of the 20th century was defined by the success of the Standard Model (SM) in predicting and explaining a plethora of experimental data. The one aspect of the SM to remain without experimental verification, the spontaneous breaking of the electroweak symmetry, has driven the construction of the Large Hadron Collider (LHC). The LHC will elucidate the nature of electroweak symmetry breaking and confirm or exclude the existence of the SM Higgs boson.

The LHC also offers the exciting prospect of discovering new physics beyond the SM. An array of new physics models have been postulated to address several shortcomings of the SM and the LHC provides an essential tool for their testing. One such model, the Minimal Supersymmetric Standard Model (MSSM), was proposed to resolve the question of why the electroweak scale is so much smaller than the scale of gravity. The MSSM predicts the existence of five Higgs bosons; three neutral and two charged; the observation of which would provide evidence for the Model and for supersymmetry in general. A search for the neutral MSSM Higgs bosons using the Compact Muon Solenoid (CMS) detector at the LHC is the subject of this thesis.

The MSSM neutral Higgs bosons, Φ, have large couplings to τ leptons so that the optimal search channel at the LHC is $\Phi \to \tau^+\tau^-$. The search methods are validated by studying the $Z \to \tau^+\tau^-$ process whereby the τ identification techniques are calibrated against the well measured properties of the Z boson.

The SM and the Higgs sector of the MSSM are outlined in Chapter 1. The background and motivation for the experimental search strategy is also given in Chapter 1 together with a summary of the LHC's operation. The CMS detector is described in Chapter 2. The reconstruction of particles produced in the LHC collisions using CMS detector measurements is the subject of Chapter 3. A measurement of the $Z \to \tau^+\tau^-$ production cross section at CMS is presented in Chapter 4. This measurement serves as an important cross check of the methods used in the $\Phi \to \tau^+\tau^-$ search presented in Chapter 5. The

results of these studies are summarised in Chapter 6 where the latest constraints on the MSSM resulting from CMS $\Phi \to \tau^+\tau^-$ searches are also given.

Throughout this thesis, natural units with $c = 1$ are used so that energies, momenta and masses are all given in units of electronvolts. In the writing of particle reactions, charge labels are dropped when the charge of a particle is unambiguous, for example in $Z \to \tau\tau$. For clarity, the neutrinos produced in τ decays are not written explicitly, for example in $Z \to \tau\tau \to \mu + \mu$ where the presence of four neutrinos in the final state is implied.

For convenience, the acronyms used in this thesis are defined in the Appendix as well as in the text at the point of first use.

Acknowledgements

I would like to thank the members of the STFC, the Imperial College HEP group and the CMS collaboration who have made this project possible and given me the dream opportunity of working at CERN.

Special thanks go to my supervisor David Colling for his guidance from the beginning to the end of this endeavour and for proof reading the thesis; to Sasha Nikitenko for his insights into Higgs$\rightarrow \tau$ physics and for providing me with many opportunities to present my work; and to Mónica Vázquez Acosta for her invaluable help and direction without which I would have been quite lost.

I thank my friends and fellow CMS collaborators Alex Sparrow and Arlo Guneratne Bryer who have provided assistance and cheer on countless occasions as well as a rather heuristic approach to understanding the arcane knowledge of Particle Physics.

Most importantly I thank my parents Moira and Abraham and my sister Mary for their unwavering love and their support for my pursuit of physics; my girlfriend Natasha for all of her encouragement and understanding; and my friends for always being there and providing a most necessary distraction from work.

Michael Cutajar

March 2012

Contents

List of Figures

List of Tables

"As you develop more and more powerful microscopic instruments, the universe becomes smaller and smaller in order to escape the investigation."

— Alan Watts, The Tao of Philosophy.

Chapter 1

Introduction

The Standard Model (SM) provides a highly successful explanation of a wide range of phenomena in Particle Physics and has undergone extensive experimental scrutiny [1]. The SM does however possess several deficiencies which motivate searches for new physics such as the search presented in this thesis. An outline of the SM and the background to the experimental studies presented in Chapters 4 and 5 are given in this introduction.[1]

The SM is a Lorentz invariant renormalisable quantum field theory in which interactions between spin-$\frac{1}{2}$ chiral fermions and spin-1 bosons are described by a Lagrangian which is invariant under the SU(3) and SU(2)×U(1) groups of gauge transformations of the fields [5][6][7][8]. For each gauge symmetry there is a conserved charge. The strength of each interaction is parameterised by a coupling constant which varies with energy scale as a result of renormalisation.

The SU(3) symmetry of the SM Lagrangian gives rise to three colour charges and the strong interaction, described by Quantum Chromodynamics (QCD) [7]. The strong interaction acts between coloured fermions, the quarks, and the SU(3) gauge bosons, the gluons g. As SU(3) is a non-Abelian group, the gluons self-interact and are massless as a result of the gauge symmetry. Consequently, the QCD coupling constant α_s decreases with increasing energy scale and the interaction becomes asymptotically free at high energies [9][10]. As large energy scales correspond to small distance scales, the quarks are confined in hadrons with the exception of the top quark which decays more rapidly than the time scale of hadronisation [1]. As QCD is strong at small energy scales, soft interactions between quarks and gluons cannot be described perturbatively. Even at high energies, higher order perturbative terms are often large so that calculations in

[1] Comprehensive introductions to the SM are given in References [2], [3] and [4]. Here the features most relevant to the experimental studies presented in this thesis are summarised.

QCD are typically presented at Leading Order (LO), Next-to-Leading Order (NLO) or Next-to-Next-to-Leading Order (NNLO).

The SU(2)×U(1) electroweak symmetry describes the unified electromagnetic and weak interactions [11][5][12]. The gauge fields associated with the SU(2) and U(1) groups are the three weak isospin fields, W^i_μ, $i = 1, 2, 3$, and the hypercharge field, B_μ, respectively. The coupling constants are denoted g and g' respectively. The left handed fermions transform as doublets,

$$\begin{pmatrix} u_i \\ d_i \end{pmatrix}_L , \begin{pmatrix} \nu_i \\ l_i \end{pmatrix}_L , \tag{1.1}$$

under SU(2) where u_i are the up-type quarks (u, c, t), d_i are the down type quarks (d, s, b), ν_i are the neutrinos (ν_e, ν_μ, ν_τ), l_i are the charged leptons (e, μ, τ) and L denotes the left handed projections of these fermions. The right handed fermions are weak isospin singlets. The physical $W^\pm$ and Z^0 bosons and the photon, γ, which mediate the weak charged, weak neutral and electromagnetic currents respectively, arise from the mixing of the gauge fields;

$$W^\pm_\mu = \frac{1}{\sqrt{2}}(W^1_\mu \mp \mathrm{i}W^2_\mu), \tag{1.2}$$

$$\begin{pmatrix} A_\mu \\ Z^0_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix}; \tag{1.3}$$

where A_μ is the photon field and

$$\cos\theta_W = g/(g^2 + g'^2)^{1/2} \tag{1.4}$$

$$\sin\theta_W = g'/(g^2 + g'^2)^{1/2}. \tag{1.5}$$

The mixing of quark flavours in charged current interactions is parameterised by the non-diagonal Cabibbo-Kobayashi-Maskawa (CKM) matrix [13][14].

The $W^\pm$ and Z^0 bosons have experimentally determined masses of approximately 80.4 and 91.2 GeV respectively [1]. As gauge boson mass terms cannot be added directly to the Lagrangian without violating gauge invariance, the electroweak symmetry must be spontaneously broken. The mechanism of electroweak symmetry breaking in the SM, described in Section 1.1, predicts the existence of a massive scalar Higgs boson which remains unobserved [15][16].

The SM does not describe gravitational interactions so that the natural energy scale up to which the SM can be expected to be valid is the Planck scale; approximately 10^{19} GeV. Extreme fine tuning of SM parameters is therefore required to explain why the scale of electroweak symmetry breaking is less than 1 TeV [17]. This problem of naturalness is discussed further in Section 1.2 and, together with the absence of evidence for the existence of the Higgs boson, is the primary motivation for the experimental studies presented in Chapters 4 and 5.

Further shortcomings of the SM include the incomplete unification of QCD with the electroweak interactions and the lack of a particle candidate to account for the observed dark matter energy density of the universe [18]. These problems may be resolved to varying degrees of satisfaction in models of new physics which in turn require experimental verification. The existence of Higgs bosons or new physics at the TeV scale will be confirmed or excluded by the Large Hadron Collider (LHC) which is described in Section 1.4 [19].

1.1 The Higgs mechanism

The Lagrangian of the electroweak gauge fields of the SM, $\mathcal{L}_G$, is given by [4][5]:

$$\mathcal{L}_G = -\frac{1}{4}(\boldsymbol{F}_{\mu\nu} \cdot \boldsymbol{F}^{\mu\nu} + G_{\mu\nu}G^{\mu\nu}). \tag{1.6}$$

$\boldsymbol{F}_{\mu\nu}$ and $G_{\mu\nu}$ are the weak isospin and hypercharge field strength tensors respectively;

$$\boldsymbol{F}_{\mu\nu} = \partial_\mu \boldsymbol{W}_\nu - \partial_\nu \boldsymbol{W}_\mu - g\boldsymbol{W}_\mu \times \boldsymbol{W}_\nu, \tag{1.7}$$

$$G_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{1.8}$$

where $\boldsymbol{W}_\mu$ has components W^i_μ, $i = 1, 2, 3$.

The masses of the $W^\pm$ and Z^0 bosons are accounted for by the Higgs mechanism of spontaneous symmetry breaking which preserves the invariance of the Lagrangian under SU(2)×U(1) gauge transformations but results in an asymmetric vacuum state [15][16][5]. The introduction of a complex scalar weak isospin doublet

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{1.9}$$

gives rise to additional terms

$$\mathcal{L}_\phi = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi) \tag{1.10}$$

in the electroweak Lagrangian with the covariant derivative

$$D_\mu = \partial_\mu - \frac{1}{2}(igT_i W_\mu^i - ig' B_\mu) \tag{1.11}$$

where T_i are the SU(2) group generators. The general form of the potential term $V(\phi)$ is given by

$$V(\phi) = \lambda(\phi^\dagger \phi)^2 - \mu_{SM}^2 \phi^\dagger \phi \tag{1.12}$$

where λ and μ_{SM} are constants parameterising the self-interactions and masses of the scalar fields. The minima of $V(\phi)$ correspond to vacuum states with expectation values $\langle 0|\phi|0\rangle$. For $\mu_{SM}^2 > 0$, a non-zero Vacuum Expectation Value (VEV),

$$\langle 0|\phi|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{1.13}$$

where

$$v = \sqrt{\frac{\mu_{SM}^2}{\lambda}}, \tag{1.14}$$

may be chosen such that the vacua are asymmetric under SU(2)×U(1) gauge transformations. The electroweak symmetry is then spontaneously broken when a vacuum corresponding to arbitrary arg ϕ is chosen. The physical particles associated with ϕ may be identified by an expansion of ϕ about its VEV:

$$\phi = \exp(-i\theta_i T^i / 2\nu) \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + H_{SM} \end{pmatrix}. \tag{1.15}$$

The phase fields θ_i may be reduced to zero by an appropriate gauge transformation leaving only the physical Higgs boson H_{SM}. By combining Equations 1.2, 1.3, 1.6, 1.10 and 1.15, H_{SM} is identified as a scalar field with mass $\sqrt{2}\mu_{SM}$ and the $W_\mu^\pm$ and Z_μ^0 fields

acquire mass terms with respective masses m_W and m_Z related by

$$m_W = m_Z \cos\theta_W = \frac{g\nu}{2}. \tag{1.16}$$

The ϕ^+ field does not acquire a non-zero VEV in Equation 1.13 so that the electromagnetic symmetry is unbroken and the photon field remains massless.

Fermion mass terms are introduced to the SM by Yukawa interactions between the fermion and Higgs fields. The down-type fermions (the down-type quarks and the charged leptons) gain mass terms through the coupling of ϕ with the left handed fermion weak isospin doublets and the right handed down-type singlets while the up-type quarks gain masses through the coupling of $\tilde{\phi}$ with the left handed quark doublets and right handed singlets.

1.2 The Hierarchy problem and supersymmetry

The Higgs mass parameter μ_{SM} is not predicted by the SM and the physical mass of the Higgs boson, $m_{H_{SM}}$, must be determined experimentally. An upper limit on $m_{H_{SM}}$ is obtained by considering the WW scattering process which violates unitarity for $m_{H_{SM}} \gtrsim 1$ TeV [20]. Furthermore, since Higgs bosons contribute to radiative corrections to the W and Z masses and couplings, measurements of electroweak observables yield indirect constraints on $m_{H_{SM}}$. A fit to such measurements, including the W and Z masses and decay widths, constrains $m_{H_{SM}}$ to be less than 152 GeV at 95% Confidence Level (CL) [21].

The bare, i.e. prerenormalisation, Higgs boson mass $\sqrt{2}\mu_{SM}$ receives radiative corrections as shown in Figure 1.1 and is related to $m_{H_{SM}}$ by [17]

$$m^2_{H_{SM}} = 2\mu^2_{SM} - \Delta m^2_H. \tag{1.17}$$

The size of the correction, Δm^2_H, is quadratically dependent on the ultraviolet cut-off Λ, corresponding to the scale at which new physics is introduced:

$$\Delta m^2_H \sim \mathcal{O}(\Lambda^2) \tag{1.18}$$

The SM may be valid up to the Planck scale where gravitational interactions become significant. If Λ is the Planck scale then Δm^2_H is more than 30 orders of magnitude

greater than $m^2_{H_{SM}}$ for $m_{H_{SM}} < 1$ TeV. The bare Higgs boson mass then requires extreme fine tuning raising a problem of naturalness known as the hierarchy problem.

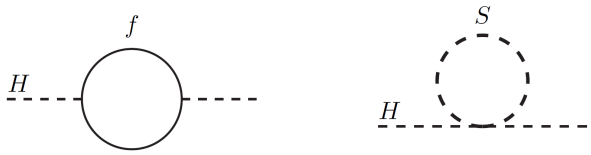


Figure 1.1: One-loop radiative corrections to the mass of a Higgs boson H via a fermion f and a scalar S [17].

Boson contributions to Δm^2_H have opposite sign to the fermion contributions. This motivates the proposition of Supersymmetry (SUSY), a symmetry between bosons and fermions [22]. Such a symmetry permits cancellation of Δm^2_H terms thus avoiding the hierarchy problem. In supersymmetric models, each SM fermion has a boson superpartner and each SM boson has a fermion superpartner. The operator defining the transformation between the boson and fermion partners commutes with the square of the spacetime translation operator and the generators of the SM symmetry groups so that the superpartners have equal mass and quantum numbers, except spin, to their SM counterparts. Since superpartners remain unobserved experimentally, their masses must be at a higher scale than that of the SM particles [1][23]. SUSY is therefore a broken symmetry if it exists in nature. The dynamics of SUSY breaking are unknown, however their effects can be parameterised in supersymmetric models [17].

In a supersymmetric model $\Delta m^2_H \sim \mathcal{O}(m^2_{boson} - m^2_{fermion})$ where $m_{fermion}$ and m_{boson} are the characteristic masses of the SM fermions and their superpartners respectively. SUSY therefore provides a natural solution to the hierarchy problem when $m^2_{boson} - m^2_{fermion} \lesssim 1$ TeV. Furthermore, supersymmetric models offer the possibility of weakly interacting massive particles as a candidate for dark matter and, in the context of grand unification, the convergence of the coupling constants of the SM interactions at high energies. Searches for SUSY are therefore an experimental priority for Particle Physics.

1.3 The Higgs sector of the Minimal Supersymmetric Standard Model

The Minimal Supersymmetric Standard Model (MSSM) is the supersymmetric extension of the SM which minimises the numbers of additional fields and interactions and conserves

R-parity [24].[2] R-parity is an operator with eigenvalues of 1 and -1 for SM particles and superpartners respectively. Its conservation is imposed in order to explain the stability of the proton. Fermion masses cannot be introduced to the model by the Yukawa interactions as described in Section 1.1 without breaking SUSY. Instead, two complex scalar weak isospin doublets, ϕ_u and ϕ_d, are included in the model to generate masses for the up-type and down-type fermions [25]:

$$\phi_u = \begin{pmatrix} \phi_u^+ \\ \phi_u^0 \end{pmatrix}, \tag{1.19}$$

$$\phi_d = \begin{pmatrix} \phi_d^0 \\ \phi_d^- \end{pmatrix}. \tag{1.20}$$

The MSSM Lagrangian contains potential terms

$$V = \mu(\phi_u^+ \phi_d^- - \phi_u^0 \phi_d^0) \tag{1.21}$$

where μ is the MSSM equivalent of the SM Higgs mass parameter μ_{SM} in Equation 1.12. The VEVs of the scalar fields are parameterised by

$$\langle 0|\phi_u|0\rangle = \begin{pmatrix} 0 \\ v_u \end{pmatrix}, \tag{1.22}$$

$$\langle 0|\phi_d|0\rangle = \begin{pmatrix} v_d \\ 0 \end{pmatrix} \tag{1.23}$$

and are related to m_W by

$$v_u^2 + v_d^2 = \frac{2m_W^2}{g}. \tag{1.24}$$

in analogy with Equation 1.16. Given Equation 1.24, the phenomenology of the MSSM Higgs sector is conveniently described in terms of the ratio of the VEVs, $\tan\beta = v_u/v_d$, which is not predicted by the model.

After electroweak symmetry breaking, i.e. the acquisition of non-zero v_u and v_d, three of the degrees of freedom associated with the scalar fields may be reduced to zero by gauge

[2] A detailed introduction to the MSSM is given in Reference [17].

transformations in analogy to the fields θ_i in Equation 1.15. Five physical Higgs bosons remain; a neutral pseudoscalar A, two neutral scalars h, H and two charged scalars $H^{\pm}$. These bosons are defined by the mixing of ϕ_u and ϕ_d which is parameterised by a mixing angle α in addition to β:

$$A = \sqrt{2}\left(\mathrm{Im}(\phi_u^0)\cos\beta + \mathrm{Im}(\phi_d^0)\sin\beta\right), \tag{1.25}$$

$$\begin{pmatrix} h \\ H \end{pmatrix} = \sqrt{2}\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}\begin{pmatrix} \mathrm{Re}(\phi_u^0) - v_u \\ \mathrm{Re}(\phi_u^0) - v_d \end{pmatrix}, \tag{1.26}$$

$$H^+ = \phi_u^+\cos\beta + \phi_d^{-\dagger}\sin\beta, \tag{1.27}$$

$$H^- = \phi_u^{+\dagger}\cos\beta + \phi_d^-\sin\beta. \tag{1.28}$$

The self interactions of the scalar fields are not independent parameters in the MSSM and can be expressed in terms of g and g'. As a result, the masses and couplings of the MSSM Higgs bosons are determined by only two free parameters at tree level; tan β and one of the Higgs boson masses which is conventionally chosen to be the mass of the pseudoscalar, m_A. The tree level masses obey the relations

$$m_h^2 = \frac{1}{2}\left(m_A^2 + m_Z^2 - \sqrt{(m_A^2 + m_Z^2)^2 - 4m_Z^2 m_A^2 \cos^2 2\beta}\right), \tag{1.29}$$

$$m_H^2 = \frac{1}{2}\left(m_A^2 + m_Z^2 + \sqrt{(m_A^2 + m_Z^2)^2 - 4m_Z^2 m_A^2 \cos^2 2\beta}\right), \tag{1.30}$$

$$m_{H^\pm}^2 = m_A^2 + m_W^2 \tag{1.31}$$

where m_h, m_H and $m_{H^\pm}$ are the masses of the h, H and $H^\pm$ bosons respectively, yielding a constraint on α;

$$\cos^2(\beta - \alpha) = \frac{m_h^2(m_Z^2 - m_h^2)}{m_A^2(m_H^2 - m_h^2)}. \tag{1.32}$$

From Equation 1.29 it can be shown that there is an upper bound on m_h;

$$m_h \leq m_Z|\cos 2\beta|. \tag{1.33}$$

The tree level Higgs boson masses and couplings are modified by higher order corrections. The dominant effect arises from the incomplete cancellation of top and stop (the superpartner of top) loops. These corrections, illustrated in Figure 1.1, do not

completely cancel because SUSY is broken. The size of the corrections are therefore strongly dependent on the top and stop masses where the stop mass depends on μ and the parameters describing SUSY breaking. After higher order corrections, the upper bound on m_h is increased and is maximised for large m_A ($m_A \gg m_Z$) and large tan β (tan $\beta \gg 1$). The maximum value of m_h for given m_A and tan β is dependent on μ and the SUSY breaking parameters. The maximum possible m_h across the whole parameter space is denoted m_h^{max}. If the scale of SUSY breaking, M_{SUSY}, is at 1 TeV, $m_h^{max} \approx 130$ GeV for a top quark mass of approximately 173 GeV [1]. For $M_{SUSY} = 2$ TeV, m_h^{max} is increased to approximately 135 GeV.

Discovery of the MSSM Higgs bosons would provide important evidence for the existence of SUSY, and verification of the MSSM in particular. It is therefore important that the phenomenology of the Higgs bosons, accounting for higher order corrections, is understood in order to optimise experimental search strategies. This is complicated by the number of SUSY breaking parameters in the MSSM. It is therefore customary to study benchmark scenarios in which the SUSY breaking parameters and μ are fixed while m_A and tan β are varied. One such scenario is the m_h^{max} scenario in which the SUSY breaking parameters are chosen such that $m_h = m_h^{max}$ at large m_A and tan β with $M_{SUSY} = 1$ TeV and $\mu = \pm 200$ GeV [1][25]. Searches for the neutral Higgs bosons, $\Phi = A, h, H$, of the MSSM are discussed in Section 1.5.

The masses of the MSSM Higgs bosons in the m_h^{max} scenario are shown as a function of m_A in Figure 1.2 for tan $\beta = 3$, 30. At large tan β, one of the neutral scalars is approximately degenerate in mass with the pseudoscalar and has similar couplings. Two mass regimes are identified depending on m_A. At large tan β and for $m_A \ll m_h^{max}$, $m_A \approx m_h$. In this regime the A and h bosons have approximately equal couplings and their couplings to down-type fermions are enhanced by a factor of tan β relative to the SM Higgs boson while their couplings to vector bosons are negligible. The H boson has similar couplings to the SM Higgs boson in this regime. At large tan β and for $m_A \gg m_h^{max}$, $m_A \approx m_H$ and A and H have approximately equal couplings with couplings to down-type fermions enhanced by a factor of tan β relative to the SM and negligible couplings to vector bosons. In this regime the h has similar couplings to the SM Higgs boson.

The effect of the enhanced couplings to down-type fermions is illustrated by the branching ratios of the pseudoscalar Higgs boson, shown in Figure 1.3 for tan $\beta = 3$, 30 in the m_h^{max} scenario. At large tan β, the dominant decays are to the heaviest down-type fermions, the b quark and the τ lepton, with branching ratios $0.8 \lesssim \mathcal{B}(A \to bb) \lesssim 0.9$

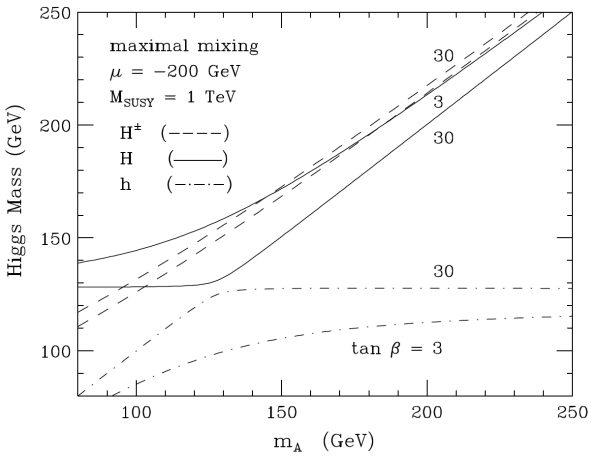
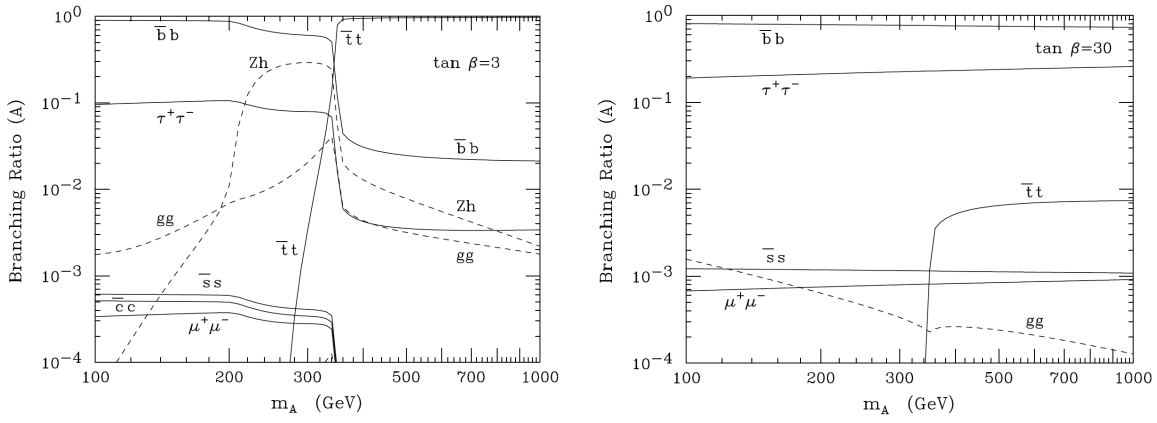


Figure 1.2: Higgs boson masses as a function of m_A for $\tan\beta = 3$, 30 in the m_h^{max} scenario of the MSSM [25].

and $0.1 \lesssim \mathcal{B}(A \to \tau\tau) \lesssim 0.2$ for $0.1 < m_A < 1$ TeV. For small $\tan\beta$ and $m_A \lesssim 200$ GeV, decays to bb and $\tau\tau$ also dominate. Overall, at large $\tan\beta$, the coupling of the neutral Higgs bosons to down-type fermions is enhanced by a factor of approximately $2 \cdot \tan\beta$ relative to the SM Higgs boson. This motivates the search for the neutral Higgs bosons in final states containing down-type fermions as discussed in Section 1.5.



(a) (b)

Figure 1.3: Branching ratios of the pseudoscalar Higgs boson A in the m_h^{max} scenario of the MSSM with $\tan\beta = 3$ (a) and $\tan\beta = 30$ (b) [25].

1.4 The Large Hadron Collider

The deficiencies of the SM and the absence of evidence for the existence of the Higgs boson prompt the exploration of physics at the TeV scale. Such exploration is facilitated by the LHC, a synchrotron designed to collide counter rotating beams of protons with center-of-mass energy $\sqrt{s} = 14$ TeV [19]. The objectives of the LHC are to test the SM at the TeV scale, elucidate the nature of electroweak symmetry breaking and search for new physics such as SUSY. The LHC also collides Pb nuclei at 2.76 TeV per nucleon in order to investigate the properties of matter at high temperatures and densities. Searches for supersymmetric neutral Higgs bosons at the LHC are discussed in Section 1.5. A summary of the LHC's operation is given in this Section.[3]

Proton beams are injected into the LHC at 450 GeV in bunches of approximately 10^{11} protons each before undergoing acceleration in radio frequency cavities [26]. The beams are bent and focused by superconducting dipole and quadrupole magnets respectively which are cooled by superfluid He. There are four interaction points at the LHC as illustrated in Figure 1.4 and a detector is situated at each point in order to study the products of the collisions. Of these, A Toroidal Large Hadron Collider Apparatus (ATLAS) and the Compact Muon Solenoid (CMS) are general purpose detectors intended to study a wide range of physics, Large Hadron Collider beauty (LHCb) is optimised to study the physics of b-flavoured hadrons and A Large Ion Collider Experiment (ALICE) is a specialised nucleus collision experiment [27][28][29][30]. The CMS detector is described in Chapter 2.

The LHC is currently operated below design specification. The studies presented in Chapters 4 and 5 analyse collision data recorded by CMS during Run 2011A of the LHC at $\sqrt{s} = 7$ TeV. The instantaneous luminosity, $\mathcal{L}$, during this run increases from approximately 10^{32} to 10^{33} cm^{-2}s^{-1}. The cumulative distribution of the integrated luminosity, $\mathcal{L}_{int}$, in Run 2011A is shown in Figure 1.5. Of the total 2.63 fb^{-1} delivered by the LHC, 2.1 fb^{-1} are certified for physics analyses by CMS.

The total inelastic proton-proton (pp) interaction cross section at $\sqrt{s} = 7$ TeV is approximately 74 mb leading to an interaction rate of $\mathcal{O}(100$ MHz) at each interaction point at $\mathcal{L} \sim \mathcal{O}(10^{33}$ cm^{-2}s$^{-1})$ [33]. The vast majority of these interactions are QCD scattering processes from which rare events such as Higgs boson production must be distinguished. Furthermore, multiple interactions occur per bunch crossing so that the

[3] A comprehensive technical description of the LHC is given in Reference [26].

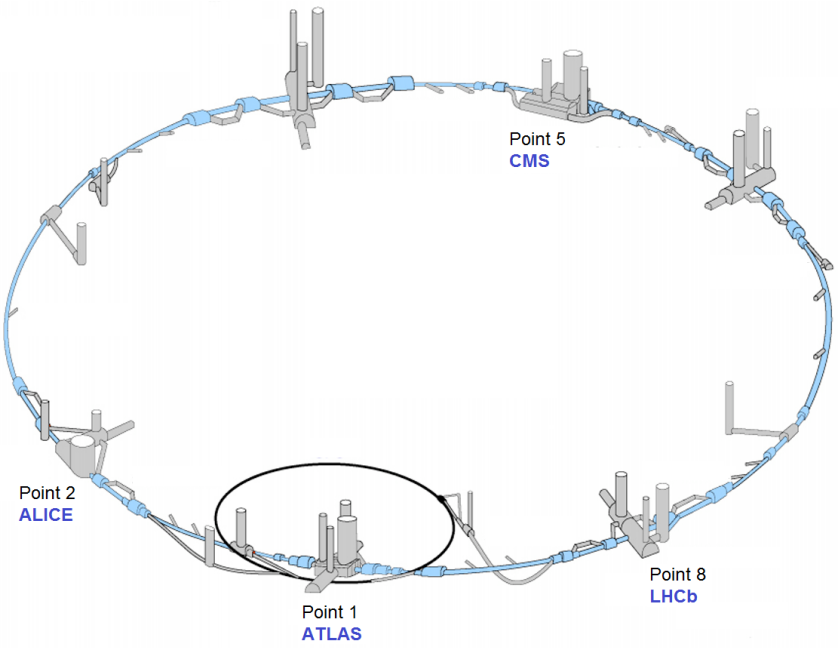


Figure 1.4: Illustration of the LHC and associated facilities. The ATLAS, ALICE, CMS and LHCb detectors are situated at interaction Points 1, 2, 5 and 8 respectively. Image adapted from Reference [31].

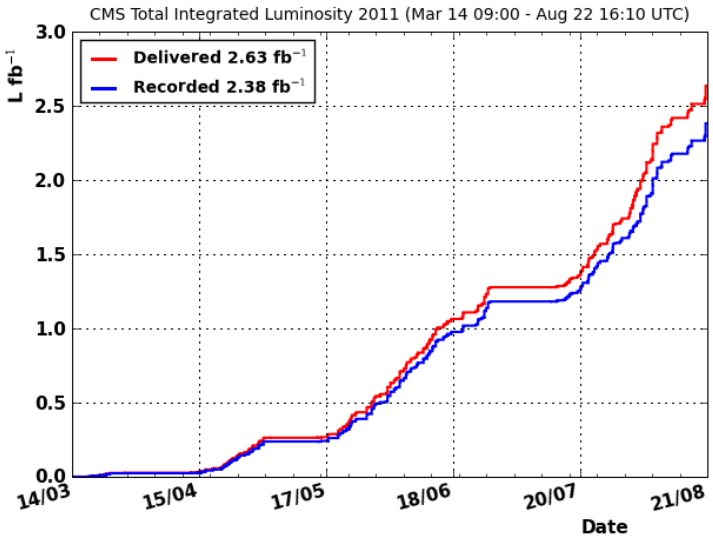


Figure 1.5: Cumulative integrated luminosity versus time in Run 2011A of the LHC [32]. Of the total 2.63 fb^{-1} delivered by the LHC, 2.38 fb^{-1} were recorded by CMS.

experimental signature of any particular interaction is contaminated by the presence of these additional Pile-Up (PU) interactions. The LHC experiments are therefore required to account for the effects of PU in the detectors when measuring the properties of events of interest. In Run 2011A, an average of six PU interactions occur per bunch crossing.

The cross section for any particular scattering process depends on the fraction of the total proton momenta carried by the interacting partons (quarks or gluons) [34]. Parton Distribution Functions (PDFs), describing the probability densities for finding partons with given momentum fractions, are therefore required in order to relate theoretical cross section calculations to measurements conducted at the LHC. As the QCD processes within a proton are non-perturbative, PDFs are estimated empirically. At the LHC the largest average momentum fractions are carried by gluons [35]. The momenta of the partons in any given interaction are unknown. When studying the kinematics of particles produced in a collision, energies and momenta transverse to the beam direction, E_T and p_T respectively, are therefore generally considered.

1.5 Searches for supersymmetric neutral Higgs bosons

While discovery of the MSSM neutral Higgs bosons would provide evidence for the existence of SUSY and resolve ambiguity over the nature of electroweak symmetry breaking, negative search results allow constraints to be placed on the MSSM parameters in benchmark scenarios such as the m_h^{max} scenario described in Section 1.3.

Searches for the MSSM neutral Higgs bosons were conducted at the Large Electron-Positron Collider (LEP) in the $e^+e^- \rightarrow hZ, HZ, hA, HA$ production channels. The search results are negative and yield the limits $m_A > 93.4$ GeV and $0.7 < \tan\beta < 2.0$ at 95% CL in the m_h^{max} scenario [36]. Searches conducted at hadron colliders such as the LHC are complementary and provide upper limits on $\tan\beta$.[4]

[4]Searches performed at the Tevatron proton-antiproton collider are described in References [37] and [38].

1.5.1 Neutral Higgs boson production at the Large Hadron Collider

Due to the large fractions of proton momentum carried by gluons at the LHC, the dominant MSSM neutral Higgs boson production mechanism at small tan β is the gluon fusion process $gg \to \Phi$ [39][40]. At small tan β the dominant contribution to $gg \to \Phi$ is via a top quark loop, however at large tan β this process proceeds predominantly via b quark loops as a result of the enhanced couplings to down-type fermions. At large tan β, the dominant production mechanism is Higgs radiation by b quarks, $gg \to bb\Phi$. The $gg \to \Phi$ and $gg \to bb\Phi$ processes are shown in Figure 1.6. Cross sections for these processes have been evaluated by the LHC Higgs Cross Section Working Group as summarised in this Section.[5] The masses and couplings of the Higgs bosons are calculated in the m_h^{max} scenario using FEYNHIGGS [41]. The parton level cross sections are combined with Martin-Stirling-Thorne-Watt-2008 (MSTW08) sets of PDFs to give cross sections $\sigma(pp \to \Phi + X)$ for the LHC at $\sqrt{s} = 7$ TeV (X denotes the production of any additional particles) [35][42].

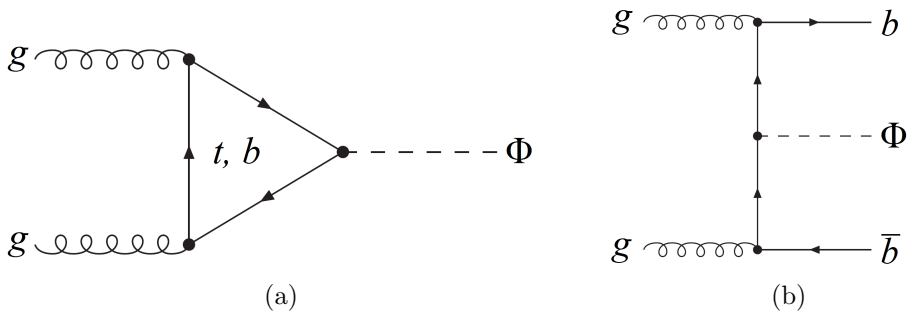
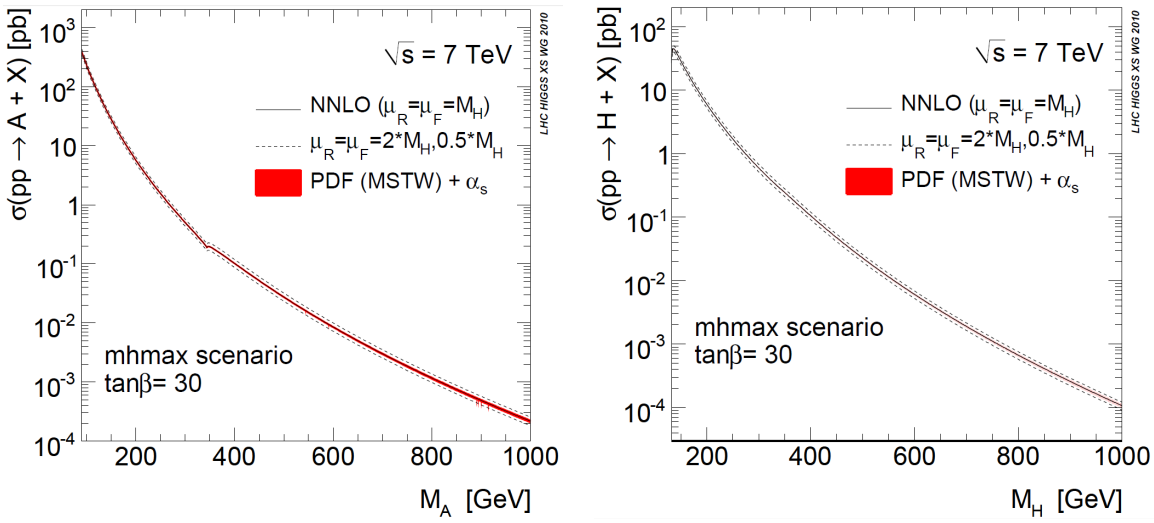


Figure 1.6: LO diagrams contributing to the gluon fusion $gg \to \Phi$ (a) and Higgs radiation by b quarks $gg \to bb\Phi$ (b) processes [39].

The $gg \to \Phi$ process is calculated at NLO in QCD using HIGLU [43]. NNLO corrections in the limit of infinite quark mass are evaluated with GGH@NNLO and applied to the top quark loop contributions to $gg \to \Phi$ [44][45]. The predicted cross sections for A and H production by gluon fusion are shown in Figure 1.7 for tan $\beta = 30$. Uncertainties on the cross sections arise from the choice of the QCD renormalisation scale and the scale at which infrared QCD radiation is cut off, the factorisation scale. These uncertainties are estimated by varying these scales around their nominal values which are chosen to be the Higgs boson mass at the point of calculation. Uncertainties on the PDFs, including

[5] The cross section calculations are described in detail in References [39] and [40].

uncertainties on α_s, are added linearly to the renormalisation and factorisation scale uncertainties resulting in total uncertainties of approximately 15% on the cross sections.



(a) (b)

Figure 1.7: Predicted cross sections for $gg \to A$ (a) and $gg \to H$ (b) production at the LHC at $\sqrt{s} = 7$ TeV as functions of m_A and m_H respectively for $\tan\beta = 30$ in the m_h^{max} scenario of the MSSM [39]. The cross sections are calculated by the LHC Higgs Cross Section Working Group at NLO using HIGLU with NNLO corrections from GGH@NNLO and MSTW08 PDF sets. The red bands indicate the PDF uncertainties including α_s uncertainties and the dotted lines the uncertainties due to choice of renormalisation and factorisation scales, μ_R and μ_F respectively.

The $gg \to bb\Phi$ process shown at LO in Figure 1.6(b) is known to NLO in QCD when four active parton flavours in the colliding protons are considered [46][47]. An alternative scheme considers the process $bb \to \Phi$ at LO, shown in Figure 1.8, together with b quark probability densities in the proton. This five flavour calculation is performed with BBH@NNLO to NNLO in QCD at which the full $gg \to bb\Phi$ process is manifested [48]. The two schemes in principle converge at higher perturbative orders however such calculations are unavailable. Figure 1.9 shows the evaluated cross sections for H and A production in the four and five flavour schemes together with their uncertainties. While the cross section uncertainties of the two schemes overlap for Higgs boson masses between 100 and 500 GeV, the central values of the cross sections differ by up to 30% for large masses.

The four and five flavour schemes provide the most accurate descriptions of the $gg \to bb\Phi$ cross sections in the limits $m_\Phi/m_b \to 1$ and $m_\Phi/m_b \to \infty$ respectively where m_Φ and m_b are the Higgs boson and b quark masses respectively. The Santander matching procedure is used to interpolate between these limits and combine the four and five

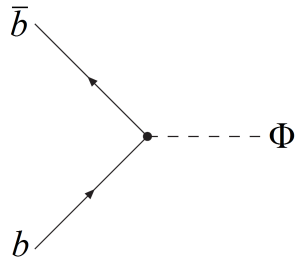


Figure 1.8: The $bb \to \Phi$ process at LO [39].

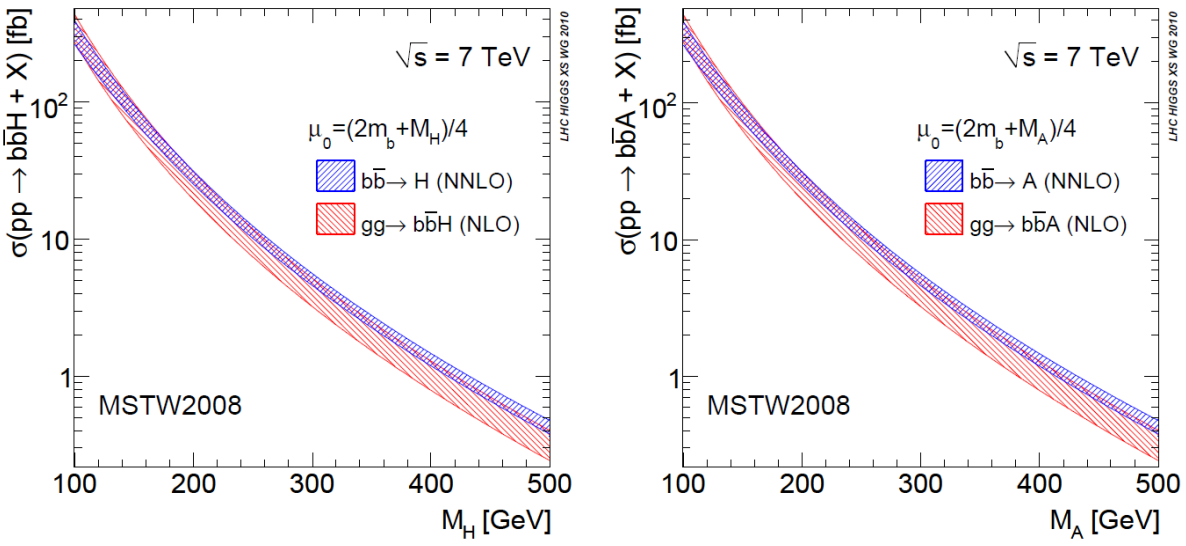


Figure 1.9: Cross sections for $gg \to bbH$ and $gg \to bbA$ production at the LHC at $\sqrt{s} = 7$ TeV as functions of m_H and m_A respectively with SM-like couplings between the b quarks and Higgs bosons [39]. The cross sections are calculated by the LHC Higgs Cross Section Working Group in the four (red bands) and five (blue bands) flavour schemes using MSTW08 PDF sets. The width of the four flavour scheme bands represents the renormalisation and factorisation scale uncertainties while the five flavour scheme bands also include scale, PDF and α_s uncertainties.

flavour calculations with appropriate weights [40]. Figure 1.10 shows the central values of the H production cross section predictions in the four and five flavour schemes relative to the matched cross sections. The total uncertainties on the matched $gg \to bb\Phi$ cross section predictions are between approximately 15 and 25%.

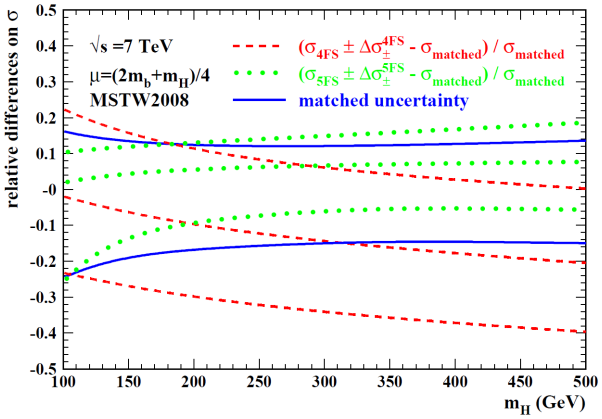


Figure 1.10: $gg \to bbH$ production cross sections and uncertainties for the LHC at $\sqrt{s} = 7$ TeV in the four (red curves) and five (green curves) flavour schemes relative to the central values of the Santander matched $gg \to bbH$ cross section predictions [40]. The relative uncertainties on the Santander matched predictions are also shown (blue curves).

Figure 1.11 shows a comparison of the $gg \to \Phi$ and $gg \to bb\Phi$ cross sections for $\tan\beta = 5, 30$. At $\tan\beta = 30$, the $gg \to bb\Phi$ process dominates while at $\tan\beta = 5$ the $gg \to \Phi$ cross section is greater than or comparable to the $gg \to bb\Phi$ cross section.

1.5.2 Search channels at the Large Hadron Collider

The dominant decays of the MSSM neutral Higgs bosons at large $\tan\beta$ are to $b\bar{b}$ pairs, as shown in Figure 1.3(b), however searches in the $b\bar{b}$ final state at the LHC are complicated by the large hadronic backgrounds resulting from QCD scattering events. The second most common decays, to $\tau^+\tau^-$, are easier to distinguish from background hadronic activity so that searches in the $\Phi \to \tau\tau$ channel are favoured [49][50]. Analyses of specific final states, depending on the decays of the τ leptons, are typically performed and combined to increase overall sensitivity. The decays of τ leptons are discussed in Section 3.6. Approximately 65% of τ decays are hadronic, producing a jet of hadrons; a τ-jet (denoted τ_h); such that the $\Phi \to \tau\tau$ channel with the greatest total branching fraction is $\Phi \to \tau\tau \to \tau_h + \tau_h$. The leptonic decays of τ leptons, to final states including an electron or a muon, have smaller branching ratios however the final state leptons

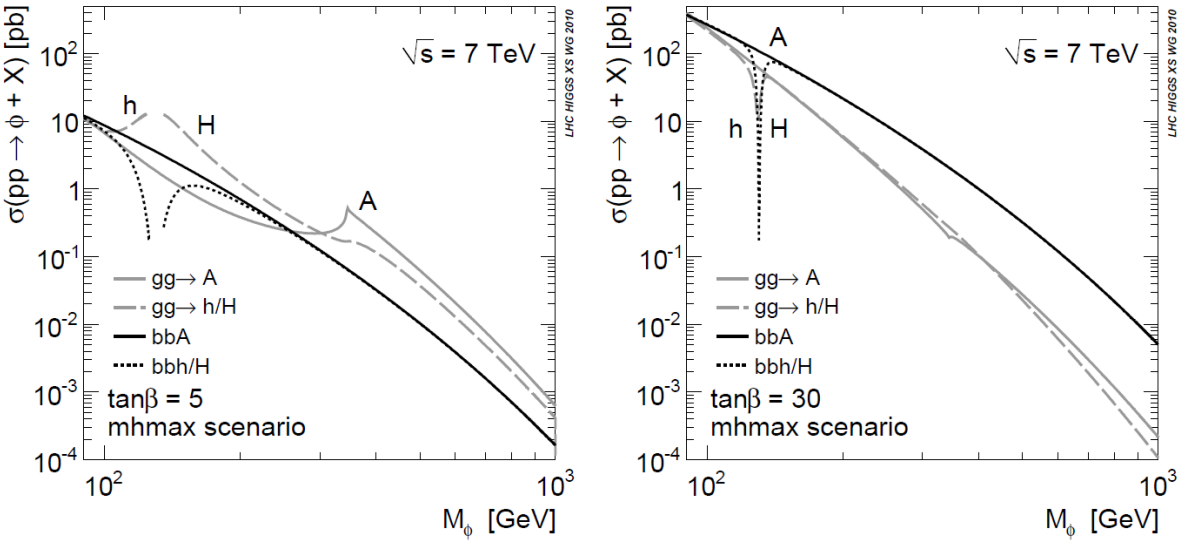


Figure 1.11: Central cross section predictions for MSSM neutral Higgs boson production by $gg \to \Phi$ and $gg \to bb\Phi$ at the LHC at $\sqrt{s} = 7$ TeV with $\tan\beta = 5$, 30 in the m_h^{max} scenario [39]. The cross sections are calculated by the LHC Higgs Cross Section Working Group with the $gg \to bb\Phi$ calculations performed in the five flavour scheme.

are easier to isolate from hadronic activity and are typically reconstructed with greater efficiency than τ-jets. The $\Phi \to \tau\tau \to e + \tau_h$ and $\Phi \to \tau\tau \to \mu + \tau_h$ channels therefore typically offer the greatest sensitivity to neutral Higgs boson production. Fully leptonic final states are most distinct from hadronic backgrounds however sensitivity in such channels is impeded by the relatively small leptonic branching ratios of τ leptons. Of the fully leptonic channels, the $\Phi \to \tau\tau \to e + \mu$ channel typically provides the greatest sensitivity as it is straightforwardly distinguished from $Z \to ee$ and $Z \to \mu\mu$ backgrounds. In all cases, the invariant mass of the τ decay products is generally reconstructed in order to test for the presence of a Higgs boson signal.

Analyses of the $\tau_h + \tau_h$, $e + \tau_h$ and $\mu + \tau_h$ final states require efficient reconstruction of the τ-jets and discrimination against large QCD scattering backgrounds. The Identification (ID) of τ-jets in the CMS experiment is the subject of Section 3.6. τ_h reconstruction methods are tested and calibrated in studies of $Z \to \tau\tau$ events as described in Section 1.5.3. $Z \to \tau\tau$ production also constitutes the dominant irreducible background to $\Phi \to \tau\tau$ events further motivating detailed studies of this process. Additional backgrounds to $\Phi \to \tau\tau$ events arise from production of W bosons and photons in association with hadronic jets, $Z \to ee$ and $Z \to \mu\mu$ events and pair production of W and Z bosons and top quarks. The properties of these backgrounds are discussed in Section 4.4.

Searches may be performed inclusive of both the $gg \to \Phi$ and $gg \to bb\Phi$ production mechanisms. Alternatively, by requiring the additional presence of jets containing b-flavoured hadrons, b-jets, in the final state, SM backgrounds to $\Phi \to \tau\tau$ events are suppressed thus maximising sensitivity to $gg \to bb\Phi$ production. The former and latter approaches are denoted the inclusive and b-tagged channels respectively.

Searches performed by the ATLAS and CMS collaborations with approximately 1 fb^{-1} of Run 2011A data each yield negative results, constraining tan β in the m_h^{max} scenario as shown in Figures 1.12 and 1.13 [51][52]. The search conducted by the ATLAS collaboration is performed in the inclusive channel considering the $\tau_h + \tau_h$, $e + \tau_h$, $\mu + \tau_h$ and $e + \mu$ final states. The CMS analysis considers only the $e + \tau_h$, $\mu + \tau_h$ and $e + \mu$ final states and combines both the inclusive and b-tagged channels. The complementary constraints on tan β resulting from the LEP searches are also shown in Figures 1.12 and 1.13. The $\Phi \to \tau\tau$ search using the full CMS Run 2011A dataset, corresponding to $\mathcal{L}_{int} = 2.1$ fb^{-1}, considering the inclusive and b-tagged channels in the $e + \tau_h$ final state, is presented in Chapter 5. The most recent $\Phi \to \tau\tau$ search results from CMS with 4.6 fb^{-1} of integrated luminosity are presented in Chapter 6.

1.5.3 Calibration with $Z \to \tau\tau$ events

The analysis methods used in $\Phi \to \tau\tau$ searches are calibrated by reconstructing $Z \to \tau\tau$ events which have a very similar experimental signature. In particular, the methods used in the reconstruction of τ-jets and the estimation of backgrounds to $\tau\tau$ events may be validated by measuring the $Z \to \tau\tau$ production cross section. Such a measurement also provides a test of the SM and a benchmark for other new physics searches in final states containing τ leptons.

Measurement of the $Z \to \tau\tau$ cross section must be performed under the assumption that there is no $\tau\tau$ signal from new physics processes such as $\Phi \to \tau\tau$. If the measured $Z \to \tau\tau$ cross section is in agreement with measured $Z \to ee$ and $Z \to \mu\mu$ cross sections, and SM predictions, then the presence of a large $\tau\tau$ contribution from new physics can be ruled out. The presence of a small contribution from new physics will still bias the $Z \to \tau\tau$ cross section measurement but not measurements of the τ_h reconstruction efficiency or the estimation of backgrounds to genuine $\tau\tau$ events. The measurement of the τ_h reconstruction efficiency using $Z \to \tau\tau$ events is described in Section 3.6.3. For these reasons, the measured $Z \to \tau\tau$ cross section cannot be used in the estimation of the $Z \to \tau\tau$ background in $\Phi \to \tau\tau$ searches. Instead, the size of the $Z \to \tau\tau$ background

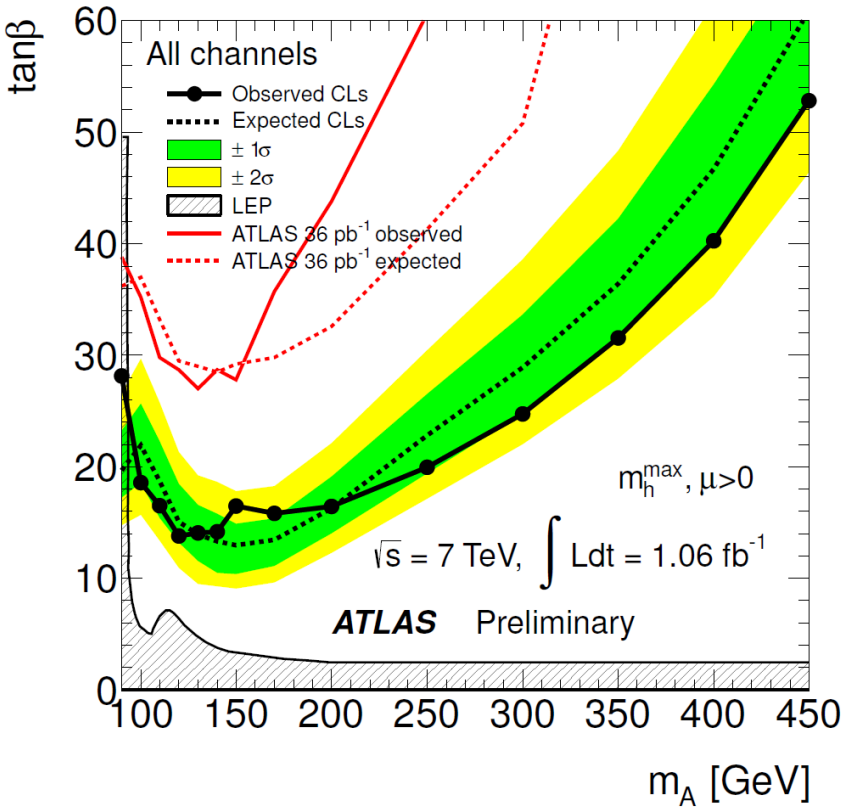


Figure 1.12: Expected and observed upper limits on $\tan\beta$ at 95% CL as a function of m_A in the m_h^{max} scenario of the MSSM resulting from $\Phi \to \tau\tau$ searches conducted by ATLAS with 1.06 fb^{-1} of integrated luminosity at $\sqrt{s} = 7$ TeV [51]. The results are derived from searches in the $\tau_h + \tau_h$, $e + \tau_h$, $\mu + \tau_h$ and $e + \mu$ final states. The results of a previous ATLAS search with a smaller dataset and the lower limits derived from searches at LEP are also shown.

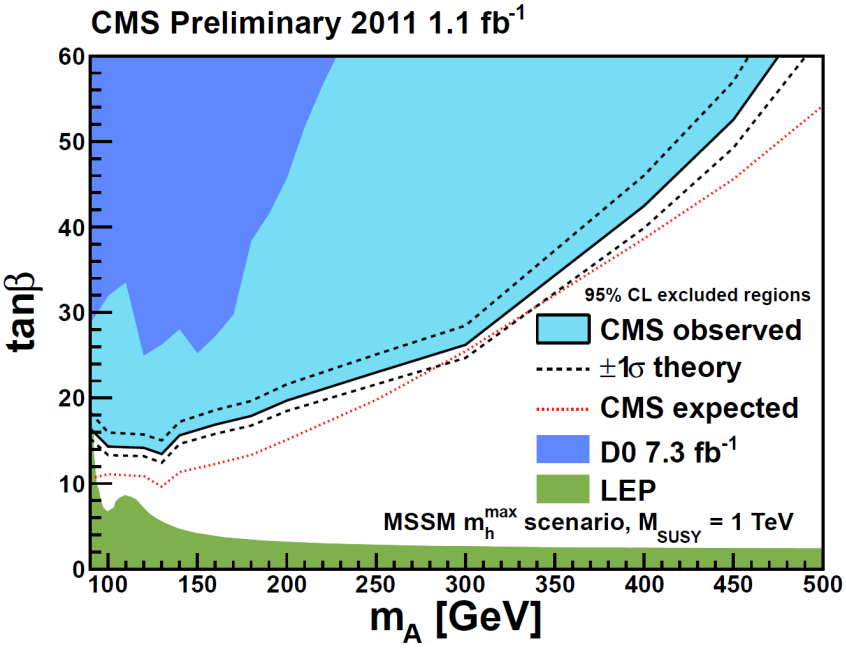


Figure 1.13: Expected and observed upper limits on tan β at 95% CL as a function of m_A in the m_h^{max} scenario of the MSSM resulting from $\Phi \to \tau\tau$ searches conducted by CMS with 1.1 fb^{-1} of integrated luminosity at $\sqrt{s} = 7$ TeV [52]. The results are derived from searches in the $e + \tau_h$, $\mu + \tau_h$ and $e + \mu$ final states. The theoretical uncertainties on the observed limits include renormalisation and factorisation scale, PDF and α_s uncertainties. The results of $\Phi \to \tau\tau$ searches performed by the D$\emptyset$ collaboration and the lower limits derived from searches at LEP are also shown.

may be constrained by either the theoretical cross section or the measured cross sections for $Z \to ee$ and $Z \to \mu\mu$ production. The measured $Z \to \tau\tau$ cross section is then only used as a cross check of the analysis methods.

The dominant production mechanism for Z bosons with dilepton final states at the LHC is the Drell-Yan process $qq \to Z/\gamma^* \to ll$, shown at LO in Figure 1.14 [53]. The Drell-Yan process consists of interfering Z and γ terms. Z boson production is therefore typically studied in events where the invariant mass of the final state leptons is between 60 and 120 GeV so that the dominant contribution is via the Z boson.

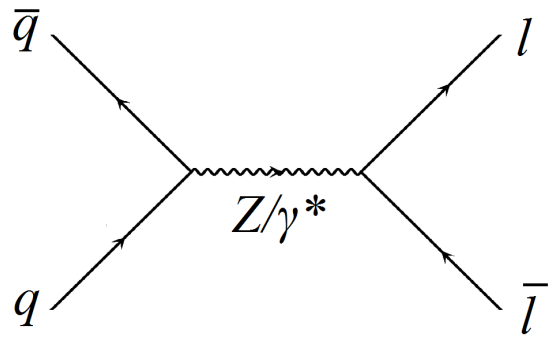


Figure 1.14: The Drell-Yan process at LO.

The $Z \to \tau\tau$ cross section, $\sigma(pp \to Z + X) \cdot \mathcal{B}(Z \to \tau\tau)$, for the LHC at $\sqrt{s} = 7$ TeV is known to NNLO accuracy. The prediction of 0.972 ± 0.042 nb is calculated with FEWZ using the MSTW08 PDF set in the 60 to 120 GeV τ pair invariant mass range [54][35]. The CMS measurement of the cross section combining the $e + \tau_h$, $\mu + \tau_h$, $e + \mu$ and $\mu + \mu$ final states in a dataset corresponding to $\mathcal{L}_{int} = 36$ pb^{-1} is in good agreement with the theoretical prediction and CMS measurements of the $Z \to ee$ and $Z \to \mu\mu$ cross sections as shown in Figure 1.15 [55]. This provides an important validation of the CMS τ_h reconstruction and a reference point from which $\Phi \to \tau\tau$ searches at CMS can be performed. Calibration of the τ_h reconstruction methods using $Z \to \tau\tau \to \mu + \tau_h$ events is discussed further in Section 3.6.

Measurement of the $Z \to \tau\tau$ cross section in the $e + \tau_h$ final state using the full CMS Run 2011A dataset is presented in Chapter 4 and serves as a calibration of the $\Phi \to \tau\tau$ search presented in Chapter 5.

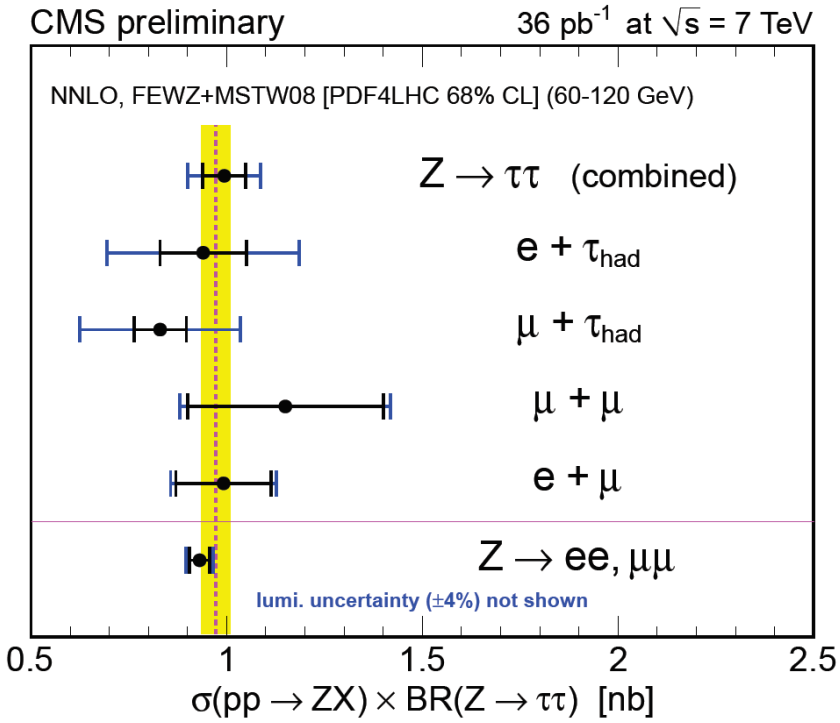


Figure 1.15: Measurements of the $Z \to \tau\tau$ cross section $\sigma(pp \to Z + X) \cdot \mathcal{B}(Z \to \tau\tau)$ at CMS with 36 pb^{-1} of integrated luminosity at $\sqrt{s} = 7$ TeV in the τ pair invariant mass range 60 to 120 GeV [55]. The measurements in the $e + \tau_h$, $\mu + \tau_h$, $\mu + \mu$ and $e + \mu$ final states are combined and compared with CMS measurements of the $Z \to ee$ and $Z \to \mu\mu$ cross sections and the theoretical $Z \to \tau\tau$ cross section prediction at NNLO. The inner error bars represent statistical uncertainties only. The outer error bars include all systematic uncertainties except uncertainties on the integrated luminosity.

Chapter 2

The Compact Muon Solenoid detector

The CMS experiment is designed to perform comprehensive tests of the SM and searches for Higgs bosons and new physics at the LHC [28][56]. The CMS detector, illustrated in Figure 2.1, is a hermetic cylindrical apparatus centred on the nominal pp interaction point at Point 5 of the LHC. The detector and its performance are summarised in this Chapter.[1]

Particles produced in the pp collisions traverse various subdetector elements which record the trajectories and energy depositions of the particles. Closest to the interaction point, the Inner Tracking System, described in Section 2.1, records the trajectories of charged particles. The trajectories bend in a 3.8 T magnetic field generated by a superconducting solenoid allowing measurement of the charged particle momenta. The Inner Tracking System is surrounded by an Electromagnetic Calorimeter (ECAL), described in Section 2.2, in which photons and electrons produce electromagnetic showers, permitting measurement of their energies. Hadron calorimeters surrounding the ECAL provide hermetic coverage and measure the energies of strongly interacting particles as described in Section 2.3. The calorimeters in principle absorb all particles except muons and weakly interacting particles and are situated inside the circumference of the solenoid to maximise containment of showers within their volume. The outermost subdetectors are integrated into the flux return of the magnet and comprise the Muon System, described in Section 2.4, which measures the trajectories of muons escaping the calorimeters. Read-out of the detector is triggered by a hardware Level-1 (L1) stage followed by a software High Level Trigger (HLT) as described in Section 2.5.

[1] A comprehensive technical description of the CMS detector is given in Reference [28].

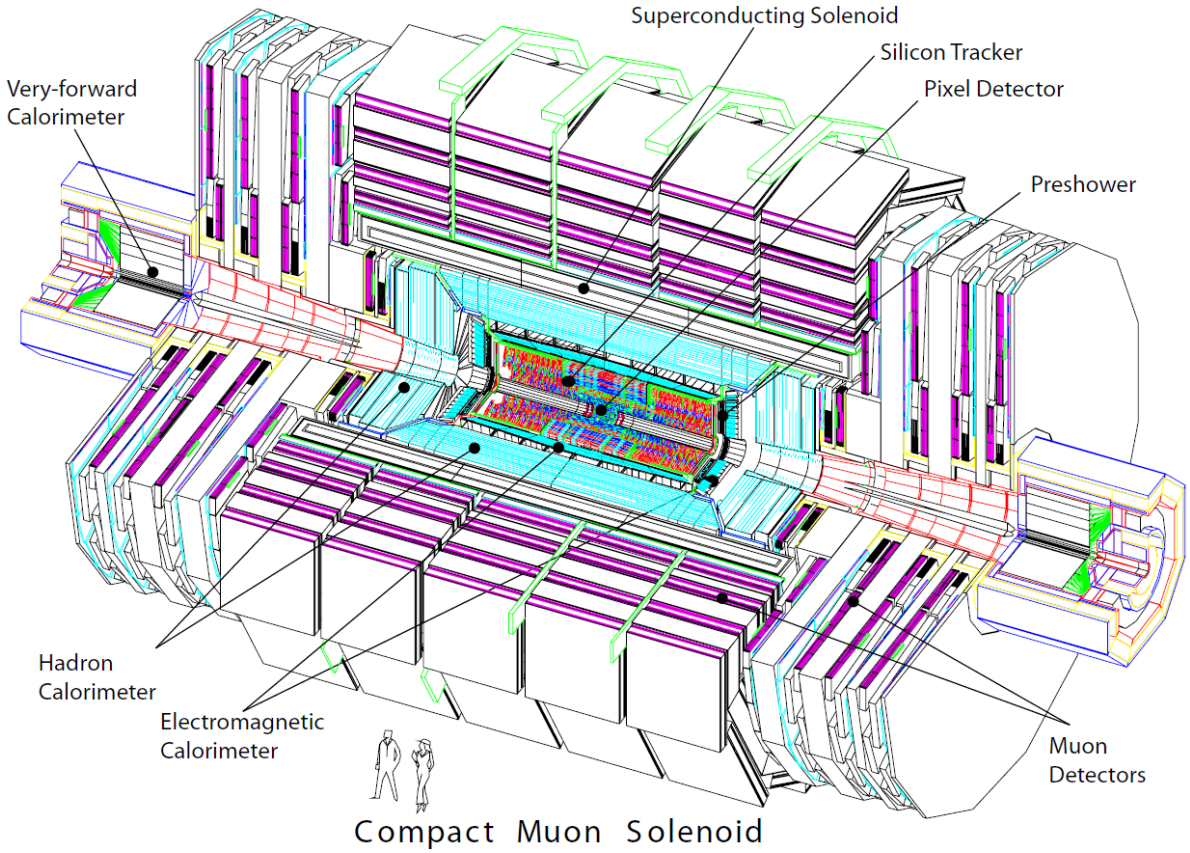


Figure 2.1: Illustration of the CMS detector with labelled subdetector elements [28]. The LHC beams pass through the center of the detector along its major axis.

The performance of the CMS detector is characterised by good momentum and electromagnetic energy resolutions, high muon detection efficiency and robustness in the large particle multiplicity environments produced at the LHC. The reconstruction and ID of particles using CMS detector measurements are described in Chapter 3.

The origin of the CMS coordinate system is at the nominal interaction point with the z-axis aligned with the beam direction. The x-axis is directed towards the centre of the LHC and the y-axis points vertically up. The azimuthal angle ϕ is defined in the x-y plane with respect to the x-axis and r is the radial coordinate in this plane. Pseudorapidity is defined as $\eta = -\ln(\tan(\theta/2))$ where θ is the polar angle defined with respect to the z-axis. Distances in the η-ϕ plane are given in terms of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ where $\Delta\eta$ and $\Delta\phi$ are distances in the η and ϕ directions respectively.

As illustrated in Figure 2.1, the CMS subdetectors consist of cylindrical barrel sections aligned with the z-axis and endcap disks perpendicular to the barrel and centred on the z-axis. The flux of particles produced in pp collisions at the LHC is approximately constant per unit η so that the endcap sections, covering the largest range in η, receive the greatest incident particle fluxes. The technologies used in the endcaps are therefore required to perform in these high flux environments and are in general different to the technologies used in the barrel.

2.1 Inner Tracking System

The Inner Tracking System measures the trajectories of charged particles that traverse it and together with the magnetic field provided by the solenoid allows measurement of their momenta and charge. The 25 ns design bunch spacing of the LHC and the large particle multiplicities produced in its pp collisions require fast response and radiation hard tracking detectors. This motivates the use of a silicon based tracking system in CMS. Hits are recorded when charged particles produce electron-hole pairs in the silicon elements which then drift in an applied electric field and are detected at the electrodes.

The Inner Tracking System, illustrated in Figure 2.2, consists of the Pixel Detector, situated closest to the interaction point where particle fluxes are greatest, and the Silicon Strip Tracker which surrounds it. The Pixel Detector uses silicon pixel tracking elements to provide precise location of interaction vertices and efficient tracking in a large particle

multiplicity environment. Particle fluxes are smaller in the Silicon Strip Tracker which has silicon microstrip tracking elements.

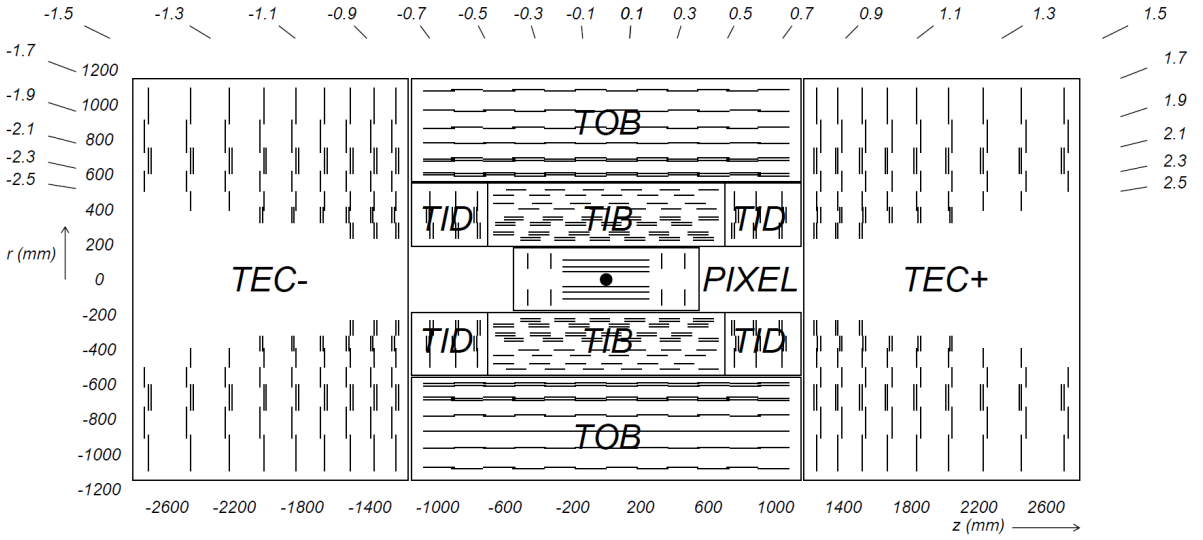


Figure 2.2: Illustration of the Inner Tracking System in the r-z plane showing the arrangement of its components [28]. Dimensions in η are shown in addition to r and z.

The Pixel Detector covers the pseudorapidity region $|\eta| < 2.1$ and has spatial resolutions of 10 and 20 µm in the r-ϕ plane and z direction respectively. The Pixel Detector has three barrel layers of pixels with the innermost layer at a radius of 4.4 cm from the z-axis. At each end of the barrel are two endcap disks of pixels arranged in a turbine geometry with blades rotated 20° out of the x-y plane. This geometry allows spatial resolution to be improved by interpolation of the charge sharing between pixels.

The Silicon Strip Tracker covers the pseudorapidity region $|\eta| < 2.4$ and consists of inner and outer barrel and endcap components as shown in Figure 2.2. In total there are ten barrel layers of microstrips of which the innermost two, the fifth and the sixth are stereo layers. The stereo layers have microstrips on both sides with an angle of 100 mrad between them. These layers allow for a precise two dimensional position measurement. There are twelve endcap layers of which the innermost two, the fourth, the fifth, and the eighth are stereo layers. The position resolution is best in the innermost components; the Tracker Inner Barrel (TIB) and Tracker Inner Disks (TID); between 23 and 34 µm in the r-ϕ plane depending on the layer and 320 µm in the z direction. In the Tracker Outer Barrel (TOB) and Tracker Endcaps (TEC) the position resolution is between 35 and 52 µm in the r-ϕ plane depending on the layer and 530 µm in the z direction. Better resolution is required in the r-ϕ plane for the measurement of p_T.

Track reconstruction is seeded by triplets of hits in the inner tracking layers [57]. The trajectory of a seed is extrapolated to the outer tracking layers using the Kalman filter method [58]. Compatible hits in each layer are added to the trajectory and its uncertainties are updated until no more compatible hits are found. To maintain high tracking efficiency while minimising misreconstruction of tracks, tracking is performed iteratively by removing hits unambiguously assigned to a track and then reconstructing tracks with the remaining hits in a subsequent iteration.

Track reconstruction efficiency for muons with $1 < p_T < 100$ GeV and $|\eta| < 2.1$ is greater than 95% and drops to around 75% outside of the Pixel Detector acceptance. The efficiency to reconstruct charged pion tracks with $1 < p_T < 100$ GeV varies between approximately 60% and 95% depending on η. The p_T resolution for muons with $p_T = 1$, 10, 100 GeV is shown as a function of $|\eta|$ in Figure 2.3. Momentum resolution degrades with increasing momentum due to decreasing curvature of the tracks. The resolutions of track displacement from the nominal pp interaction point in the x-y plane and the z direction; the transverse and longitudinal impact parameters respectively; are shown as a function of η in Figure 2.4 for tracks with $p_T = 1$, 3, 8 GeV and are less than 0.2 and 1 mm respectively [59].

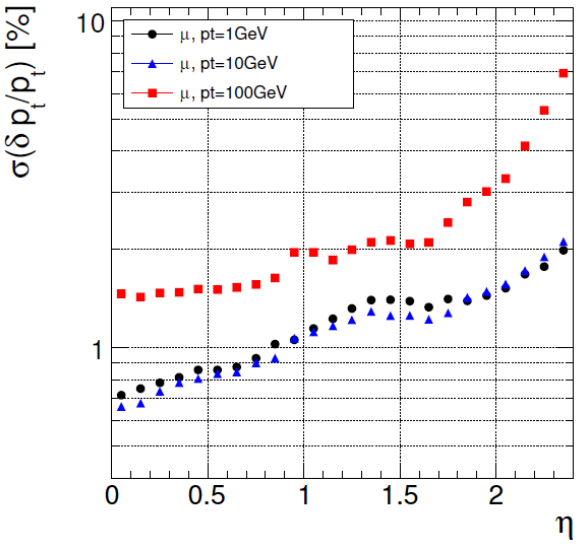


Figure 2.3: p_T resolution of muon tracks reconstructed in the Inner Tracking System as a function of η for simulated muons with $p_T = 1$, 10, 100 GeV [28].

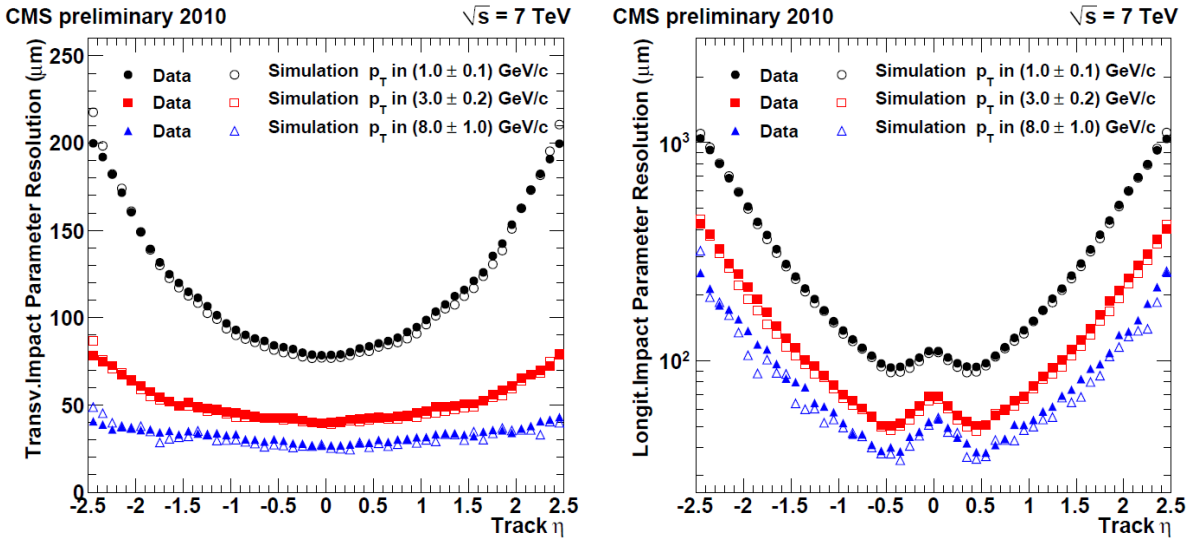


Figure 2.4: Transverse and longitudinal track impact parameter resolutions as a function of η for tracks with p_T = 1, 3, 8 GeV in CMS data and simulation [59].

2.2 Electromagnetic Calorimeter

The ECAL is a homogeneous $PbWO_4$ crystal calorimeter covering the pseudorapidity region $|\eta| < 3$. $PbWO_4$ is a fast and radiation hard scintillator with a short radiation length X_0 and a short Molière radius. The ECAL has high granularity, excellent energy resolution and is more than 25 X_0 deep with respect to the nominal pp interaction point.

The ECAL consists of Electromagnetic Calorimeter Barrel (EB) and Electromagnetic Calorimeter Endcaps (EE) sections in which the crystals are arranged in modules as illustrated in Figure 2.5. Photons and electrons shower in the ECAL crystals resulting in scintillation from which the energy of the shower is measured. The scintillation signals in the EB crystals are read out by avalanche photodiodes while more radiation hard vacuum phototriodes read out the crystals in the EE. Preshower detectors consisting of lead absorbers and silicon microstrips are positioned between the TEC and the EE to aid neutral pion (π^0) ID in the forward, i.e. large $|\eta|$, region where the pions are typically energetic enough to make individual photon resolution difficult.

The crystals in the EB have segmentation $\Delta\eta \times \Delta\phi = 0.017 \times 0.017$ and are mounted 3° off the axis projecting to the nominal pp interaction point to avoid energy loss between crystals. Uninstrumented regions exist between the EB and EE, $1.442 < |\eta| < 1.566$, through which cables and services pass. Electrons and photons traversing these regions may not undergo significant showering resulting in poor measurement of their energies.

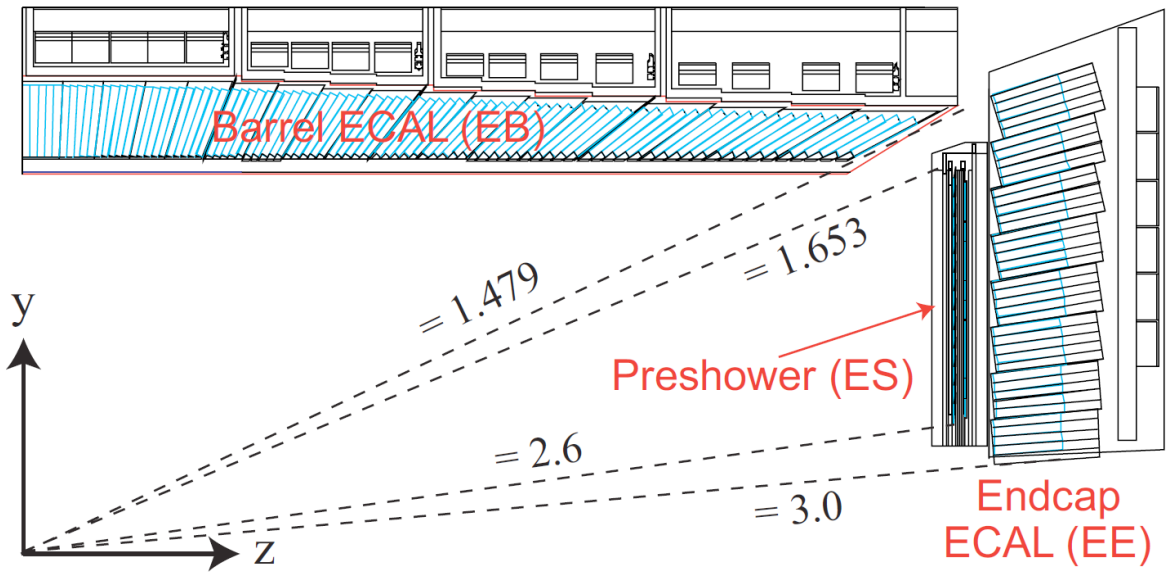


Figure 2.5: Illustration of the ECAL in the r-z plane showing the arrangement of the crystals (blue rectangles) in the EB and EE, and the location of the Preshower detector [56]. The dotted lines at constant η show the boundaries of the fiducial volumes of the ECAL components.

The crystals in the endcap are mounted in an x-y grid with 28.6×28.6 mm^2 segmentation and are also off axis with respect to the nominal interaction point.

The energy resolution of the ECAL, σ_E/E where E is measured energy and $\sigma(E)$ its uncertainty, is given by

$$\left(\frac{\sigma(E)}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{2.1}$$

where S parameterises stochastic fluctuations in scintillation and shower shape, N the electronics noise and C the non-uniformity of read-out, shower penetration beyond the ECAL and calibration errors. $\sigma(E)/E$ is measured as a function of E using electron beams. From a fit to the measurements, shown in Figure 2.6, the resolution parameters are evaluated to be S=0.028 GeV$^{1/2}$, N = 0.12 GeV and C = 0.003 [28].

2.3 Hadron Calorimeters

A sampling Hadron Calorimeter (HCAL) consisting of brass absorber layers interspersed with active plastic tiles covers the pseudorapidity region $|\eta| < 3$ as illustrated in Figure 2.7.

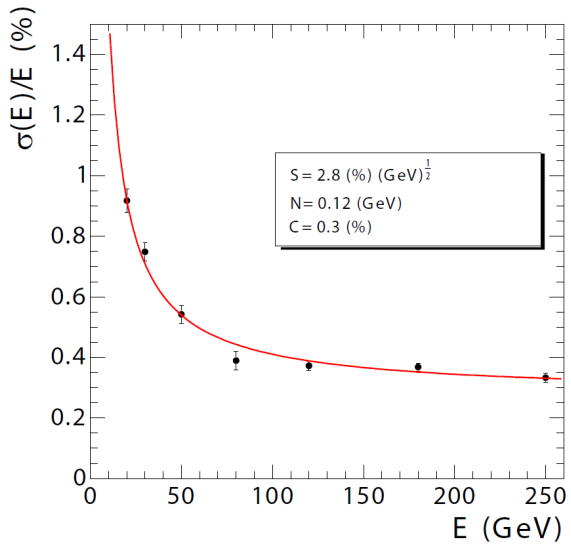


Figure 2.6: ECAL energy resolution measured as a function of E with electron test beams [28]. The stochastic, noise and constant parameters of the energy resolution are determined by a fit to the measurements (red line).

Hadron showers initiated in the absorber layers induce scintillation in the tiles which is converted by wavelength shifting fibres for read-out by hybrid photodiodes. The HCAL is enclosed by the solenoid, maximising shower containment and reducing non-Gaussian tails in energy resolution due to energy loss. The Hadron Calorimeter Barrel (HB), covering $|\eta| < 1.4$, is read out in towers of size $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. The Hadron Calorimeter Outer (HO), a layer of scintillator tiles lining the outside of the solenoid, samples the tails of highly penetrating showers. The Hadron Calorimeter Endcaps (HE) at each end of the barrel are segmented into towers with dimensions varying between $\Delta\eta \times \Delta\phi = 0.087 \times 0.8$ and 0.35×0.8.

The energy resolution of the HCAL is significantly larger than the electromagnetic energy resolution of the ECAL and its granularity is coarser. Furthermore, the responses to the electromagnetic and hadronic fractions of hadron showers differ leading to a non-linear overall energy response. The combined energy resolution of the EB and HB for charged pions ($\pi^{\pm}$) is measured using test beams to be [28]

$$\left(\frac{\sigma(E)}{E}\right)^2 = \left(\frac{1.2\text{GeV}^{1/2}}{\sqrt{E}}\right)^2 + 0.069^2. \tag{2.2}$$

The Forward Calorimeter (HF) extends the pseudorapidity coverage of the hadron calorimetry to $|\eta| < 5$ as shown in Figure 2.7. The HF uses steel absorber and Cerenkov

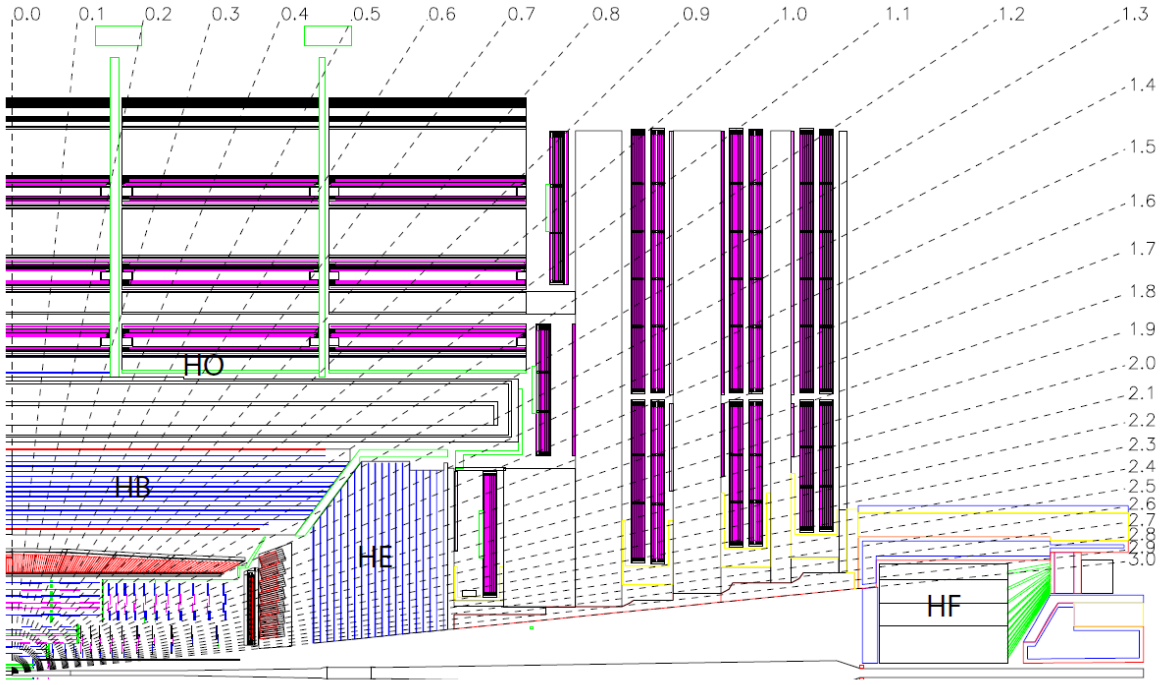


Figure 2.7: Illustration of the hadron calorimeters in the r-z plane showing the locations of the HCAL components and the HF [28]. The dotted lines are at constant η.

radiating quartz fibre active layers read out by photomultipliers. The HF samples primarily the electromagnetic fraction of hadron showers resulting in relatively narrow visible shower shapes. This improves spatial resolution in the large particle flux environment of the forward region. The segmentation of the HF towers is $\Delta\eta \times \Delta\phi = 0.175 \times 0.175$. In addition to the measurement of forward particle production, the HF is used to measure the LHC luminosity delivered at CMS by inferring the mean number of inelastic pp interactions per bunch crossing from the average fraction of empty HF towers in an event. The relative uncertainty on the CMS luminosity measurement is 4.5% [60].

The hadron calorimeters provide hermetic coverage, have a depth with respect to the nominal pp interaction point varying between 10 and 15 interaction lengths depending on η and contain no uninstrumented regions. These qualities enable the measurement of the E_T balance in an event and subsequently the missing transverse energy, $\not\!\!E_T$, from which the production of neutrinos, or other weakly interacting stable particles, may be inferred as described in Section 3.8.

2.4 Muon System

The iron flux return of the magnet is instrumented with a system of muon detectors which covers the pseudorapidity region $|\eta| < 2.4$. These detectors operate on the principle of the wire chamber whereby hits are recorded when traversing muons ionise the gas in the detector elements. These hits are combined with measurements from the Inner Tracking System to reconstruct muon trajectories and measure their momenta as described in Section 3.5.

As illustrated in Figure 2.8, there are four layers of detectors in both the barrel and endcap sections of the Muon System. In the barrel, $|\eta| < 1.2$, each layer stations Ar+CO_2 Drift Tubes (DTs) and Resistive Plate Chambers (RPCs). The DTs have a position resolution of approximately 200 μm and a resolution on track direction of approximately 1 mrad in ϕ. The RPCs consist of a gas gap enclosed by cathode plates and are operated in avalanche mode. The RPCs have inferior position resolution to the DTs but have a fast time resolution of approximately 1 ns which aids identification of the pp bunch crossing time. Multiwire Cathode Strip Chambers (CSCs) which have finer segmentation and greater radiation hardness than the DTs are used in the endcaps. Each CSC has six gas gaps, a position resolution of 100-200 μm depending on η, and approximately 10 mrad resolution on track direction in ϕ. For $|\eta| < 1.6$ the CSCs are complemented by RPCs.

2.5 Trigger

At design specification, the proton bunch crossing rate in the LHC, corresponding to the event rate in CMS, is 40 MHz. Recording all such events would require an unfeasible data rate so a highly selective two stage trigger is employed to identify events of interest and reduce the recorded event rate to hundreds of events per second.

A hardware L1 Trigger constructed with custom electronics reduces the event rate to approximately 100 kHz. The complete detector read-out for each event is stored in a pipeline of 128 events, corresponding to a period of 3.2 μs at design luminosity. To produce a decision within this time constraint the L1 Trigger uses information from the Muon System and calorimeters only, with the calorimeters analysed in coarsely grained towers corresponding to 5×5 arrays of ECAL crystals and individual HCAL towers.

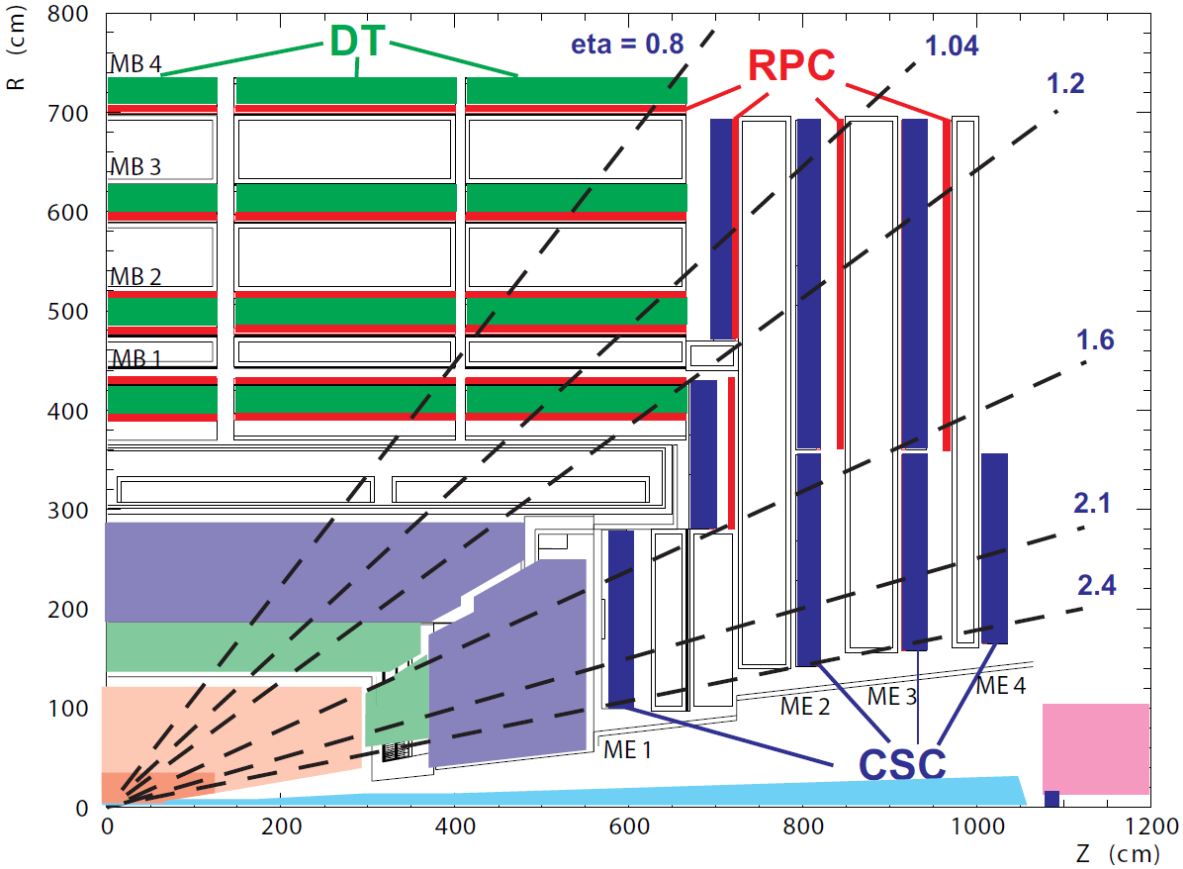


Figure 2.8: Illustration of the Muon System in the r-z plane showing the arrangement of the DT, RPC and CSC detectors [56].

Events accepted by the L1 Trigger are then filtered by a processor farm running the HLT algorithms which reconstruct and identify various particle candidates in an event. The software implementation of the HLT allows for maximum flexibility in online event selection. Events are rejected as early as possible by first using only calorimeter and Muon System information and then using the more computing intensive track reconstruction.

The L1 and HLT algorithms used in the analyses presented in Chapters 4 and 5 are described in Sections 3.4.4 and 3.6.5.

Chapter 3

Physics object reconstruction

The track momenta and energy deposits measured by the CMS detector are used to construct higher level objects for use in physics analyses. The algorithms which identify the signatures of lepton, photon and hadron production are described in this Chapter. Many CMS analyses, including the analyses presented in Chapters 4 and 5, employ the CMS Particle Flow (PF) algorithm which combines measurements from all CMS subdetectors with the aim of reconstructing and identifying all particles produced in an event as described in Section 3.2. The analyses presented in Chapters 4 and 5 involve primarily the reconstruction of electrons, jets, τ-jets and b-jets. The ID and calibration of these objects are described in Sections 3.4, 3.3, 3.6 and 3.7 respectively. The analyses also indirectly require the reconstruction of muons, pp interaction vertices and $\not\!\!E_T$ as outlined in Sections 3.5, 3.1 and 3.8 respectively.

3.1 Primary vertices

The tracking of charged particles by the Inner Tracking System allows the location of the hard pp interaction, the Primary Vertex (PV), to be identified by reconstructing the common vertex of the tracks [56]. This is necessary in order to distinguish prompt production of particles, i.e. production at the PV, from in-flight decays of hadrons and photon conversions, and to account for the effects of PU in which multiple pp interactions occur in the same bunch crossing. The primary hard pp scattering in an event, distinct from PU interactions, is typically assumed to be the PV with the largest scalar sum of associated track momenta, Σp_T.

PV reconstruction is seeded by groups of two or more tracks separated in z by less than 1 cm at their points of closest approach to the z-axis [59]. The tracks in each group are clustered into vertex candidates by a deterministic annealing algorithm which allows for separation of vertices closely spaced in z, as is the case in large PU multiplicity events [61]. The three-dimensional position of each vertex candidate is then evaluated by the Adaptive Vertex Finder (AVF) algorithm which is robust with respect to misassociated tracks or mismeasured track errors [62]. Each track is assigned a weight w_{track} which takes values of unity for tracks most compatible with the fitted vertex position and zero for tracks least compatible with the vertex. The number of degrees of freedom of the vertex candidate, defined as $2 \cdot \sum_{track} w_{track} - 3$, is strongly correlated with the number of tracks compatible with a common vertex and is used to distinguish real pp interactions from misreconstructed vertices.

3.2 The Particle Flow algorithm

Calorimeter based reconstruction of hadronic jets and $\not\!\!E_T$ in CMS is largely limited by the granularity and energy resolution of the HCAL [56]. The PF algorithm provides an alternative approach by combining information from all CMS subdetectors thereby exploiting the precision trajectory measurements of the Inner Tracking System and the high granularity and energy resolution of the ECAL [63]. The outputs of the PF algorithm are all particle candidates in the event which are exclusively classified as electrons, muons, photons, charged hadrons and neutral hadrons. These particle candidates may then be used to reconstruct higher level physics objects such as jets and $\not\!\!E_T$. PF reconstruction achieves superior spatial and energy resolutions for jets and more accurate measurement of the physical $\not\!\!E_T$ in an event than calorimeter based reconstruction [64][65].

The input elements to the PF algorithm are tracks, clusters of calorimeter energy deposits and hits in the Muon System. PF contains a dedicated algorithm for clustering energy deposits in the calorimeters, designed to aid separation of spatially close deposits and for efficient detection of low energy particles as necessary in a global event reconstruction. The clustering is performed separately in the ECAL and HCAL. Cluster seeds are identified as calorimeter cells; crystals in the ECAL and towers in the HCAL; which are local maxima in energy. Topological clusters are formed by combining seed cells with adjacent cells that have energy greater than two standard deviations of read-out noise. Each seed cell in a topological cluster gives rise to a PF cluster. The energy of each cell

in the topological cluster is shared between the PF clusters according to the cell to PF cluster distance. In the HF, each cell is a PF cluster.

The input elements are linked into blocks: Each link in the algorithm is made between two elements however a block may contain more than two elements, for example when a track is linked to an ECAL cluster and the ECAL cluster is also linked to an HCAL cluster. Tracks and clusters are linked when the track trajectory, extrapolated to the calorimeter, is within the boundaries of the cluster. An example of this linking is illustrated in Figure 3.1. ECAL and HCAL clusters are linked when the centre of the ECAL cluster is within the boundaries of the HCAL cluster. Tracks are linked to hits in the Muon System when a global track fit including the muon hits returns an acceptable χ^2.

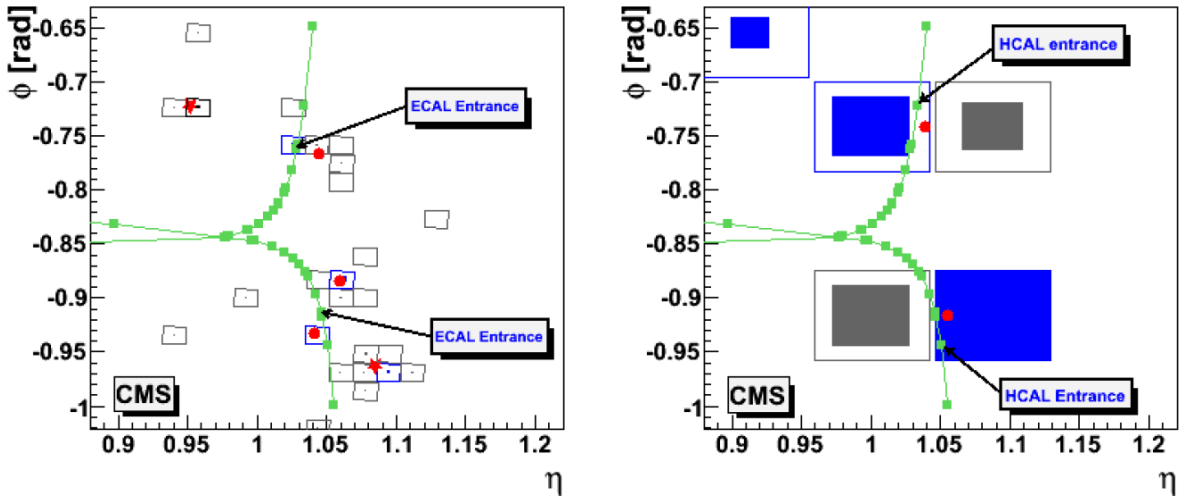


Figure 3.1: Illustration of track-cluster linking in the PF algorithm in the η-ϕ plane [66]. The squares represent ECAL (left) and HCAL (right) cells where the blue squares are seed cells and adjacent cells comprise clusters. The clusters marked with a red dot are linked to tracks (green lines) with trajectories passing within their boundaries. The clusters marked with a red star are unlinked.

The PF particle ID algorithms are applied to each reconstructed block in order to determine their particle content. First, links between tracks and Muon System hits are identified as PF muons when the momentum returned by the global fit is compatible with the momentum of the track only to within three standard deviations. PF muons and their energy deposits in the calorimeters, estimated from studies of cosmic ray events, are then removed from the block. Muon reconstruction is discussed further in Section 3.5.

Electron candidates are considered after removal of PF muons from the blocks. Electrons lose energy through interactions with the material of the Inner Tracking System and electron tracks are preidentified by a loose selection based on the momentum difference

between the beginning and the end of the track. Tracks passing this preselection are then refitted by the Gaussian Sum Filter (GSF) algorithm [67]. The GSF algorithm more accurately accounts for the non-Gaussian tails in radiative energy loss with respect to the Kalman filter. The interaction of electrons with the CMS detector is described in further detail in Section 3.4. GSF tracks are then extrapolated to the ECAL and are linked to ECAL clusters when their extrapolated trajectories pass within the boundaries of a cluster. Tangents to the GSF tracks are projected to the ECAL and if such a tangent is within the boundaries of an ECAL cluster not already linked to a track, the cluster is linked to the GSF track as a potential bremsstrahlung photon candidate. Considering such combinations of GSF tracks and clusters, electrons are distinguished from hadrons by a multivariate discriminator ξ constructed from variables describing the geometric and kinematic matching of the track and the clusters, the estimated fraction of radiated energy and the electromagnetic shower profile. Electron ID variables are described in further detail in Section 3.4. The distribution of ξ for PF electron candidates in data collected at $\sqrt{s} = 900$ GeV; mostly fake electrons misreconstructed from hadrons; and for Monte Carlo (MC) simulated electrons is shown in Figure 3.2. If $\xi >$ -0.1, the GSF track and linked clusters are identified as a PF electron and are removed from the block.

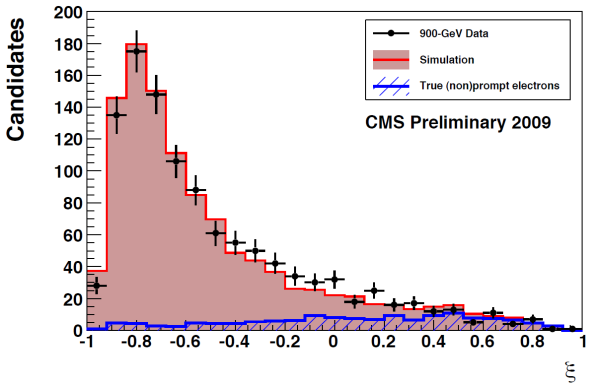
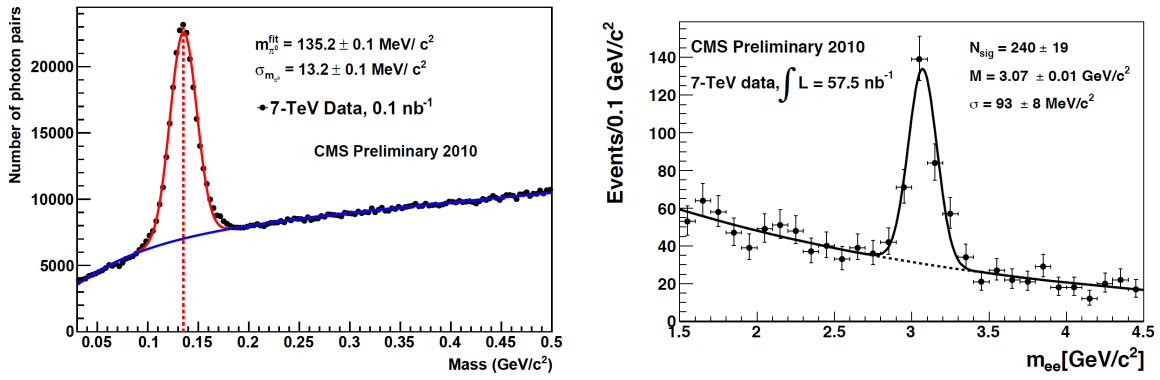


Figure 3.2: Distribution of the PF multivariate electron discriminator ξ for PF electrons reconstructed in CMS data collected at $\sqrt{s} = 900$ GeV and in MC [66]. The PF electrons in the data are predominantly fake electrons misreconstructed from hadrons. The fraction of true electrons in the MC distribution is shown as the blue shaded region.

The remaining tracks in a block give rise to PF charged hadrons. If the cluster energy and track momentum of a block are compatible under the charged pion mass hypothesis, the PF charged hadrons' momenta and energies are derived by a fit to the track and calorimeter measurements. If the cluster energy of the block is in excess of the track momenta, the presence of neutral particles in the block is inferred. If such an excess

is larger than the total ECAL energy of the block, a PF photon is reconstructed and assigned the ECAL energy and a PF neutral hadron is reconstructed with the remaining excess energy. If the excess is less than the total ECAL energy, a PF photon only is reconstructed. Remaining ECAL and HCAL clusters in an event not linked to tracks give rise to PF photons and neutral hadrons respectively.

The performance of physics object reconstruction using the PF algorithm has been studied in CMS data at $\sqrt{s} = 7$ TeV [68][69]. Figure 3.3 shows the invariant mass spectrum of reconstructed PF photon pairs; the observed resonance at 135.2 MeV is in agreement with the π^0 mass to within 1% [1]. Figure 3.3 also shows the invariant mass distribution of reconstructed PF electron pairs displaying a resonance at 3.07 GeV; compatible with the J/ψ mass [1]. The reconstruction of jets, hadronic τ decays and $\not\!\!E_T$ using PF is described in Sections 3.3, 3.6 and 3.8 respectively.



(a) (b)

Figure 3.3: Invariant mass distribution of PF photon (a) and electron (b) pairs in data [68][69]. The observed resonances at 135.2 MeV (a) and 3.07 GeV (b) are compatible with the π^0 and J/ψ masses respectively.

3.3 Jets

Quarks and gluons are produced copiously at the LHC in the hard scattering of partons. Quarks and gluons fragment and hadronise to form jets of strongly interacting particles and their decay products [34]. Analysis of jet production at CMS allows for testing of perturbative QCD predictions as well as searches for new physics, including searches for SUSY [50]. Understanding of reconstructed jet energies with respect to the physical jet energies, the Jet Energy Scale (JES), is essential in order to minimise and accurately estimate systematic uncertainties in such analyses. Reconstruction of jets also serves

as the starting point for the ID of hadronic τ decays and b-hadrons as described in Sections 3.6 and 3.7 respectively.

Jets contain a broad spectrum of particle types which interact with the CMS detector in different ways. The PF algorithm, combining all of the CMS subdetectors, is therefore highly suited to the reconstruction of jets [64]. PF candidates are clustered into jets with the infrared- and collinear-safe Anti-k_t algorithm so that the jet shapes are not influenced by soft QCD radiation [70]. The clustering size parameter is 0.5.

Jet ID criteria are applied to suppress the reconstruction of jets due to calorimeter and electronics noise [64]. The jet energy is required to have contributions from both electromagnetic and hadronic depositions with fractions of total jet energy carried by photons and neutral hadrons less than 0.99 separately. All jets are required to consist of more than one PF constituent. Jets within the acceptance of the Inner Tracking System are further required to contain at least one charged object and must have a charged hadron energy fraction greater than zero and an electron energy fraction less than 0.99. These requirements retain more than 99% of real jets while rejecting jets arising from spurious energy readings in a single subdetector. The requirements also prevent the reconstruction of jets from isolated photons and electrons.

Figure 3.4 shows the average fractions of total jet energy carried by electrons, charged hadrons, neutral hadrons and photons as reconstructed by the PF algorithm in minimum bias events, i.e. events triggered only by the requirement of an inelastic pp scattering. This information is used to account for the effects of PU in the reconstruction of electrons and τ-jets as described in Sections 3.4.2 and 3.6.1. This is possible because the signatures of PU interactions in the CMS detector, in events triggered by high p_T physics objects, are the same as those of minimum bias events. In the fiducial region of the Inner Tracking System, $|\eta| < 2.4$, approximately 65% of the average jet energy is carried by charged hadrons with photons accounting for 25% and neutral hadrons 10% [68].

3.3.1 Energy scale

Due to uninstrumented regions in the ECAL, non-linear HCAL response and contamination of jets by particles produced in PU interactions, the reconstructed energy of a jet is in general different from the physical energy of the jet [71]. Reconstructed jets are therefore calibrated so that their measured energies more accurately represent the physical energies.

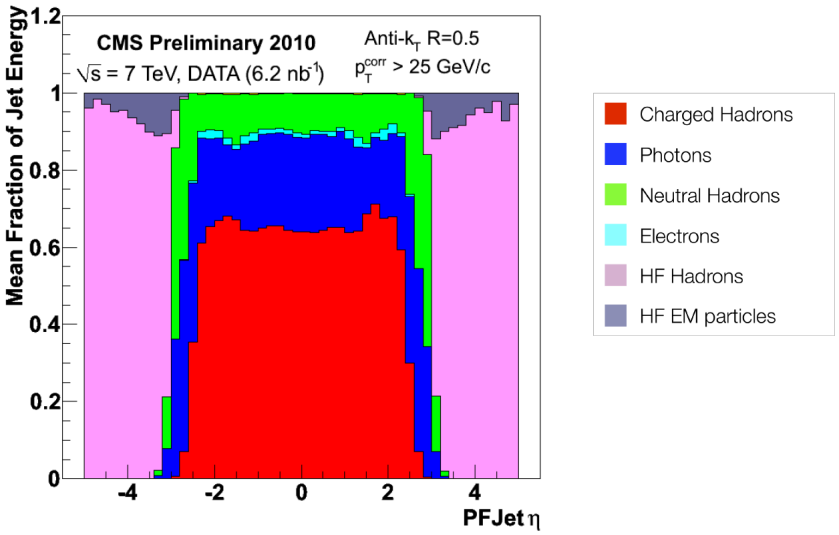


Figure 3.4: Average PF jet energy fractions carried the various types of PF object as a function of η in minimum bias events [68].

Corrections to the raw JES are applied as separate subcorrections [71]: A PU correction is first derived by simulating the average p_T density in each event due to PU and subtracting the corresponding p_T contribution to each jet. Relative and absolute corrections then correct the variation of energy response as a function of jet η and p_T respectively. The corrected jet energy, E_{corr}, is given by

$$E_{corr} = \left(E_{raw} - C_{PU}(p_T^{jet})\right) \cdot C_{rel}(\eta^{jet}, p_T^{jet}) \cdot C_{abs}(p'_T) \tag{3.1}$$

where E_{raw} is the uncorrected jet energy and C_{PU}, $C_{rel}(\eta^{jet}, p_T^{jet})$ and $C_{abs}(p'_T)$ are the PU, relative and absolute corrections respectively which are functions of the uncorrected jet transverse momentum p_T^{jet}. The relative correction is also a function of the jet pseudorapidity and $p'_T = C_{rel}(\eta^{jet}, p_T^{jet}) \cdot p_T^{jet}$.

The relative correction is determined by the dijet p_T balance method which exploits momentum conservation in dijet events containing a central jet with $|\eta| < 1.3$, corresponding to a region of relative uniformity in the detector, and a probe jet at arbitrary η [72][28]. The relative response $r(\eta^{probe}, p_T^{dijet})$, i.e. the ratio of the probe jet p_T to the central jet p_T, is measured in bins of probe jet pseudorapidity η^{probe} and average p_T of the two jets p_T^{dijet}. The response is shown in Figure 3.5 in the $40 < p_T^{dijet} < 55$ and $85 < p_T^{dijet} < 100$ GeV bins. The relative correction is given by $C_{rel}(\eta^{jet}, p_T^{jet}) = 1/r(\eta^{jet}, p_T^{jet})$. Deviations in r from unity are small and do not exceed 10%. The responses measured in data and

MC are also largely in agreement with small differences for jets with $|\eta| > 1.5$. The response in data is further corrected for these residual differences.

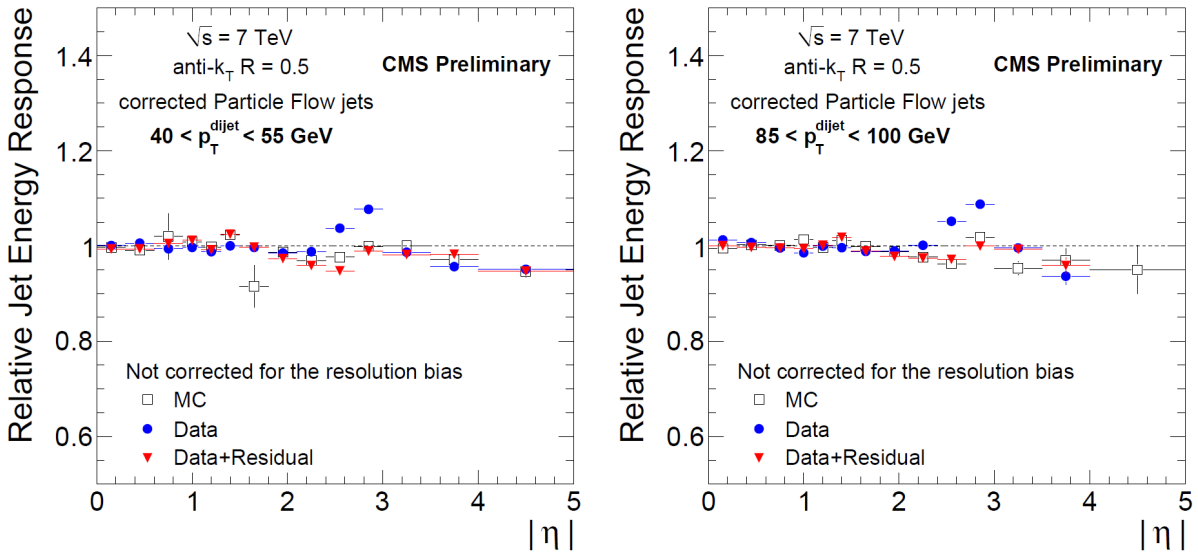


Figure 3.5: Relative jet energy response as a function of jet η measured in dijet events with $40 < p_T^{dijet} < 55$ and $85 < p_T^{dijet} < 100$ GeV in data and MC [71]. The distributions in data after residual corrections are also shown.

Due to the steeply falling p_T^{dijet} spectrum and finite jet energy resolution, each p_T^{dijet} bin is contaminated by jets with smaller physical energies and upward fluctuated detector response. This resolution bias results in the dominant systematic uncertainty on the relative response measurement and is estimated by modelling shifts in the jet energy resolution. For jets with $p_T = 50$ GeV, the size of the uncertainty is estimated to be less than 2% at $\eta = 0$ and rises to 5% at $|\eta| = 4.5$.

The absolute JES is measured in the central reference region of the relative corrections, defined by $|\eta| < 1.3$, using the Missing transverse energy Projection Fraction (MPF) method [73]. Similarly to the measurement of the relative response, the absolute response measurement exploits momentum conservation in γ+jet events together with the accuracy of the photon energy measurement provided by the ECAL [74]. Considering the momenta of the photon p_T^{γ} and the recoil jet p_T^{rec} and the respective detector responses, R_{γ} and R_{rec}, the reconstructed $\not{E}_T$ in the event is given by

$$\not{E}_T = -R_{\gamma} p_T^{\gamma} - R_{rec} p_T^{rec}. \tag{3.2}$$

$\not\!\!E_T$ is reconstructed using the PF algorithm as described in Section 3.8. Given that the physical $\not\!\!E_T$ in γ+jet events is zero, $p_T^{\gamma} + p_T^{rec} = 0$ so that

$$R_{rec} = R_{\gamma} + \frac{\not\!\!E_T \cdot p_T^{\gamma}}{(p_T^{\gamma})^2}. \tag{3.3}$$

By considering the reconstructed $\not\!\!E_T$ in the recoil jet response, the presence of additional soft jets and unclustered energy in the event is accounted for. R_{rec} is measured in bins of photon p_T and the absolute correction in a given p_T bin is $C_{abs} = 1/R_{rec}(p_T^{\gamma})$. R_{rec} varies between approximately 0.85 and 0.95 depending on the photon p_T. The response is also measured in MC γ+jet events and the data is corrected for the small residual differences as in the case of the relative corrections. The ratio of the response in data to the response in MC as a function of photon p_T is shown in Figure 3.6. To cross check the MPF method, a simpler p_T balance method is employed, in analogy to the dijet balance method used for measurement of the relative JES, with which consistent results are obtained.

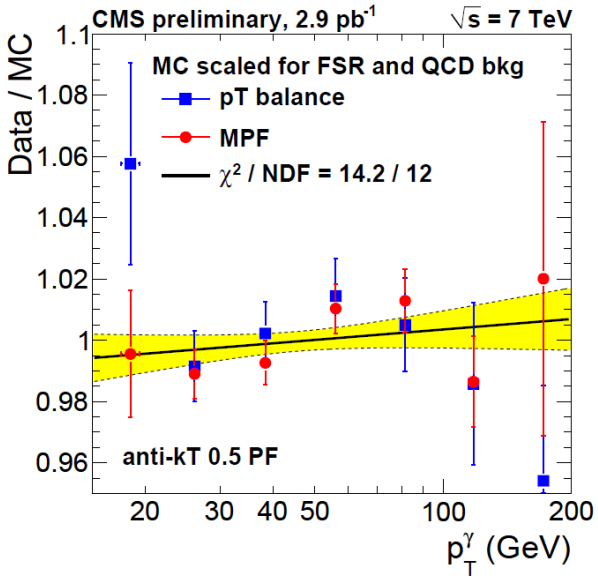
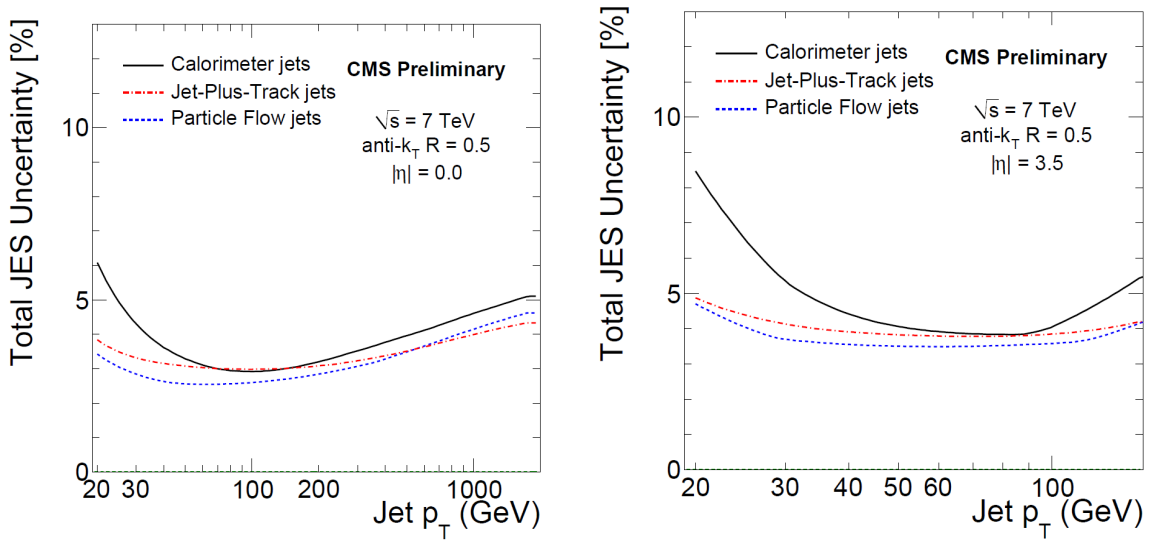


Figure 3.6: Ratio of absolute jet energy response in data to the response MC as a function of photon p_T as measured by the p_T balance and MPF methods in γ+jet events [71]. A fit to the results obtained by both methods (black line) with associated statistical uncertainty (yellow band) is also shown.

Dominant uncertainties on the absolute JES measurement arise from the uncertainty on the photon energy scale, which is less than 1%, and uncertainties inherent to the MPF method: The MPF method is sensitive to the production of multiple hard jets in γ+jets

events and also to multijet background events. The MPF uncertainty varies between 1 and 4% depending on photon p_T.

The total JES uncertainty after applying the jet energy corrections according to Equation 3.1 is given by the quadratic sum of uncertainties on the PU, relative and absolute corrections and is shown in Figure 3.7 as a function of jet p_T for jets with $\eta = 0$ and $|\eta| = 3.5$. The uncertainty on the PU correction is determined by varying the average p_T density in the simulations used to derive the correction and this is less than 5% for jets with $p_T > 20$ GeV. The total JES uncertainty varies between approximately 3 and 5% depending on jet η and p_T .



(a) (b)

Figure 3.7: JES uncertainty as a function of jet p_T after applying jet energy corrections for jets with $\eta = 0$ (a) and $|\eta| = 3.5$ (b) [71]. The blue curves represent the JES uncertainty for jets reconstructed using the PF algorithm while the red and black curves represent the uncertainties for calorimeter based jet reconstruction with and without corrections derived from track measurements respectively.

3.4 Electrons

Reconstruction of electrons is important for the measurement of many SM processes at CMS as well as for searches for Higgs bosons and new physics [50]. Electrons are light with a rest mass of approximately 500 keV and produce electromagnetic showers in the ECAL [1][56]. Electrons are reconstructed by matching ECAL energy deposits to

tracks in the Inner Tracking System. This is challenging because of the interactions of electrons with the material of the tracking detectors. Electrons traversing this material undergo bremsstrahlung with the total energy radiated dependent on the amount of material traversed. This varies from approximately 0.35 X_0 at $\eta = 0$ to 1.4 X_0 at the end of the EB. The radiated energy arrives at the ECAL spread in ϕ as a result of the 3.8 T magnetic field in the detector. A further complication arises from the conversion of radiated photons in the tracking material.

Electrons produced in τ decays in $Z \to \tau\tau$ and $\Phi \to \tau\tau$ events, as studied in Chapters 4 and 5, are produced with significant p_T and are typically well isolated in the detector. Backgrounds to such electrons in CMS arise from jet production in which the overlap of a charged hadron and a π^0, early showering of a charged hadron or real electron production in a heavy flavour decay can fake the prompt electron signature. Misreconstructed electrons also arise from real electrons produced in conversions of prompt photons. These backgrounds are suppressed by applying a combination of ID and isolation criteria to electron candidates as well as searching for photon conversion vertices. The measurement of the efficiency of the electron selection used in the $Z \to \tau\tau$ and $\Phi \to \tau\tau$ analyses is presented in Section 4.3.3.

3.4.1 Supercluster and track reconstruction

Electron reconstruction is seeded by superclusters of ECAL energy deposits into which the electron radiation deposits are collected [56]. This method is optimised for isolated electrons with transverse momenta typical of those produced in vector boson decays [75]. Superclusters are constructed by combining energy deposits in the ECAL that are spread narrowly in η and more widely in ϕ so as to collect radiated energy. The different geometries of the EB and EE require the use of two distinct superclustering algorithms. In both cases the clustering is seeded by the most energetic crystal and continues until there are no more crystals above a certain energy threshold that are not already clustered. Total supercluster E_T is required to be greater than 4 GeV.

In the EB, a seed cluster, centred on the seed crystal, is defined by a collection of groups of strips of three to five crystals in the η direction separated from other such groups by strips with E_T less than 100 MeV. Other such clusters within $\Delta\phi = 0.3$ of the seed crystal, assumed to be radiation deposits, are combined with the seed cluster to form the supercluster. In the EE a cluster corresponding to a 5×5 array of crystals around the seed crystal is constructed. The outermost crystals of the array may seed a new 5×5

cluster allowing closely overlapping radiation deposits to be collected. The supercluster is formed by collecting further clusters within $\Delta\eta = 0.14$ and $\Delta\phi = 0.6$ of the seed cluster.

Superclusters are used to locate electron hits in the Pixel Detector by calculating the energy weighted average position of the supercluster. This position coincides with the impact point at the ECAL of a non-radiating electron with energy equal to the supercluster energy. This position is then propagated to the first pixel layer under both charge hypotheses and a compatible hit is searched for in a loose ϕ-z window. If found, the hit in the first layer is used to update the estimated electron trajectory with which hits in the outer pixel layers are located. The electron candidate pixel hits seed the reconstruction of the electron trajectory through the Silicon Strip Tracker using the GSF algorithm, described in Section 3.2.

3.4.2 Identification, conversion rejection and isolation

Given a GSF track matched to an ECAL supercluster, electron ID, isolation and photon conversion rejection variables are used to suppress background electrons. The cuts placed on these variables are tuned to specific analyses. The cuts applied in the analyses presented in Chapter 4 and 5 are given in Section 4.2.

Electron ID variables include quantities describing the electromagnetic shower profile, the geometric matching between the supercluster and the GSF track and the quantity of energy deposited in the HCAL towers behind the supercluster [56]. The shower profile of the supercluster in η, $\sigma_{i\eta i\eta}$, enables discrimination against broader and more fluctuant hadron showers. $\sigma_{i\eta i\eta}$ is defined as

$$\sigma_{i\eta i\eta} = \frac{\sum_i^{5\times5} w_i(\eta_i - \bar{\eta}_{5\times5})}{\sum_i^{5\times5} w_i} \tag{3.4}$$

where

$$w_i = 4.2 + \ln(E_i/E_{5\times5}) \tag{3.5}$$

and the index i runs over the ECAL crystals in a 5×5 array centered on the seed crystal [75]. η_i is the η position of the ith crystal, $\bar{\eta}_{5\times5}$ is the energy weighted mean η position of the 5×5 array, and w_i is the weight of the ith crystal where E_i and $E_{5\times5}$ are the energies of the ith crystal and 5×5 array respectively. The width in η is considered,

rather than the width in ϕ, because it is largely unaffected by the spread of electron energy deposits caused by the magnetic field.

The geometric matching between electron superclusters and tracks is quantified by the variables $\Delta\eta_{in}$ and $\Delta\phi_{in}$ defined by the distances in η and ϕ respectively between the energy weighted average supercluster position and the track trajectory as measured at the PV. Requiring small $\Delta\eta_{in}$ and $\Delta\phi_{in}$ provides strong suppression of backgrounds due to photon conversions, heavy flavour decays and overlapping charged hadrons and neutral pions. Hadrons typically deposit most of their energy in the HCAL so that H/E, the ratio of the HCAL energy behind the supercluster to the supercluster energy may be used to further suppress hadron backgrounds.

The distributions of the electron ID variables in isolated dielectron and QCD dijet MC events are shown in Figure 3.8 where the distinction between the electron and background distributions is clearly visible [76].

Electron candidates arising from photon conversions may be suppressed by considering the properties of the electron track and by reconstructing candidate conversion vertices. An electron produced in a photon conversion is on average expected to originate with some displacement from the PV so that its track will have missing hits in the pixel layers. The number of pixel layers with missing hits before the first hit of the electron track, N_{miss}, may therefore be used to suppress conversion electrons [75]. To identify candidate conversion vertices, Opposite Sign (OS) pairs of tracks with a point of closest approach with absolute distance from the nominal interaction point greater than 0.9 cm are considered [52]. For each pair of tracks a vertex fit is performed using CTVMFT where the tracks are constrained to be parallel in both the r-ϕ and r-z planes at the vertex [77]. Candidate conversion vertices are identified when the fit probability is greater than 10^{-6} and the transverse distance between the conversion vertex and the largest Σp_T PV is greater than 2 cm. The density of identified vertices in minimum bias events is found to be well matched to the density of material in the Inner Tracking System where conversions are most likely to take place [75].

As fake electron candidates reconstructed in jets are typically produced in close proximity to other particles, applying isolation to electron candidates is a powerful means of suppressing QCD backgrounds. To combine information from the CMS subdetectors a PF based isolation is applied where the isolation variable is constructed from the p_T sums of PF charged hadrons, neutral hadrons and photons within a cone of size $\Delta R = 0.4$ centred on the electron candidate [52]. As the isolation cone may be contaminated

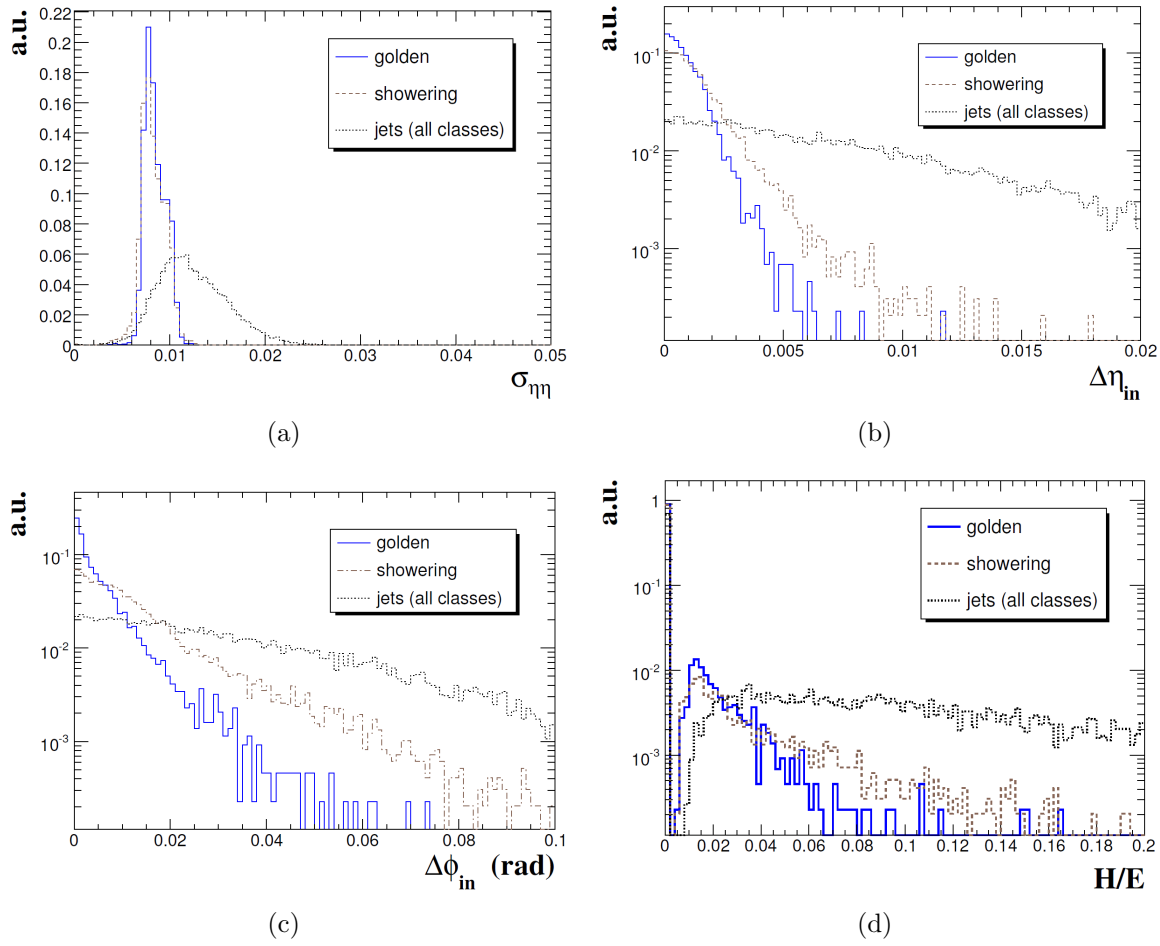


Figure 3.8: Distributions of the electron ID variables $\sigma_{i\eta i\eta}$ (a), $\Delta\eta_{in}$ (b), $\Delta\phi_{in}$ (c) and H/E (d) in dielectron MC events with electron p_T between 5 and 50 GeV and dijet MC events with jet p_T between 25 and 170 GeV [76]. The golden and showering electron types refer to electrons which undergo minimal and large numbers of interactions in the Inner Tracking System respectively.

by particles originating from PU interactions, only charged hadrons consistent with production at the largest Σp_T PV are considered. Furthermore, the p_T sum of charged hadrons originating from other vertices, assumed to be PU vertices, is used to estimate and subtract the contribution to the isolation variable from neutral particles produced in PU interactions. The isolation variable I_e is defined as

$$I_e = \frac{\Sigma p_T^{charged} + \Sigma p_T^{neutral} + \Sigma p_T^{photon} - \Delta\beta_e}{p_T^e} \tag{3.6}$$

where $\Sigma p_T^{charged}$, $\Sigma p_T^{neutral}$ and Σp_T^{photon} are the p_T sums of charged hadron, neutral hadron and photons respectively within the isolation cone and p_T^e is the transverse momentum of the electron candidate. The PU correction $\Delta\beta_e$ is the p_T sum of charged hadron candidates in the isolation cone which are not consistent with production at the PV multiplied by the ratio of neutral to charged particles in minimum bias events. This ratio is approximately 0.5 as shown in Figure 3.4 [68].

3.4.3 Energy scale

Reconstructed electron energies and momenta are determined by combining measurements of the track momentum and supercluster energy weighted by their respective uncertainties [56]. The two measurements are complementary as the track measurement is more precise for low p_T electrons and the resolution of the energy measured in the ECAL increases with electron p_T.

The difference between physical and reconstructed electron energy is estimated by reconstructing $Z \to ee$ events. A fit to the reconstructed dielectron mass distribution is performed with the Z resonance modelled by a Breit-Wigner function [78][52]. The peak of the mass distribution is expected to be at approximately 91.2 GeV [1]. Fits are performed in different electron pseudorapidity bins. In all cases the difference between the fitted electron energy scale and the raw reconstructed scale is less than 2% for electrons within the acceptance of the Inner Tracking System, $|\eta| < 2.4$.

The reconstruction of the Z resonance in the dielectron channel is shown in Figure 3.9 after correcting the electron energy scale and is used to measure the inclusive $Z \to ee$ cross section in CMS. The measured cross section, for dielectron invariant masses between 60 and 120 GeV, using a data sample collected at $\sqrt{s} = 7$ TeV corresponding to $\mathcal{L}_{int} =$

36 pb^{-1},

$$\sigma(pp \to Z + X) \times \mathcal{B}(Z \to ee) = 0.992 \pm 0.011 \text{ (stat.)} \pm 0.018 \text{ (syst.)} \pm 0.016 \text{ (th.)} \pm 0.040 \text{ (lumi.) nb}, \quad (3.7)$$

is in good agreement with theoretical predictions [78].

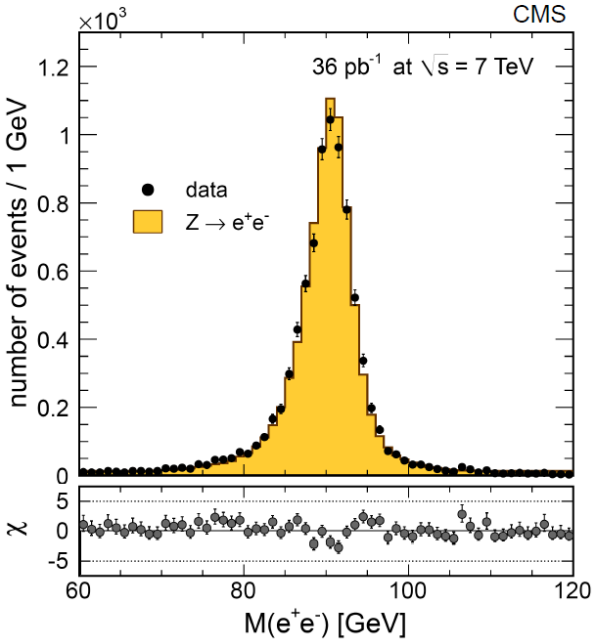


Figure 3.9: Reconstructed dielectron invariant mass distribution in data and in $Z \to ee$ MC events [78]. The bin-by-bin normalised difference between data and MC, χ, is also shown.

3.4.4 Electron triggers

In the $\Phi \to \tau\tau$ and $Z \to \tau\tau$ analyses presented in Chapters 4 and 5, events are selected online by requiring the reconstruction of electron objects at trigger level. The reconstruction of electromagnetic objects at L1 considers ECAL trigger towers which are 5×5 arrays of ECAL crystals [56]. The E_T deposited in each trigger tower is computed in 250 MeV steps. Starting with seed trigger towers defined as the most energetic trigger towers, the E_T in the most energetic adjacent trigger tower is added to the seed trigger tower to form pairs which constitute the L1 electromagnetic trigger objects. The L1 trigger decision is generated by applying E_T thresholds to these objects. Two such trigger selections, EG12 and EG15, are employed in the $\Phi \to \tau\tau$ and $Z \to \tau\tau$ analyses with E_T thresholds of 12 and 15 GeV respectively.

Reconstruction of electrons in the HLT is seeded by superclusters with E_T greater than a specified threshold that are spatially matched to a L1 electromagnetic trigger object [56][52]. A compatible hit in the Pixel Detector is required which then seeds the reconstruction of the electron track by the Kalman filter method. The track trajectory is required to be matched to the supercluster position under looser criteria than those used offline. A simpler form of isolation is applied with respect to the isolation described in Section 3.4.2 by evaluating p_T and E_T sums of tracks and energy deposits in a cone of size $\Delta R = 0.3$ centred on the electron candidate. A strip relatively narrow in η and extended in ϕ is excluded so that radiated electron energy is accounted for. The HLT selection criteria are then defined by cuts on the supercluster E_T and the isolation sums.

3.5 Muons

Muons have a mean lifetime of approximately 2.2 μs and a mass of approximately 106 MeV so that they typically traverse the CMS calorimeters with minimal energy deposition and are then tracked by the muon detectors [1][56]. As a result, muons in CMS are relatively straightforward to isolate from hadronic activity and are thus powerful signatures of SM and new physics processes with leptons in the final state [50].

There are two primary approaches to muon reconstruction in CMS, both using tracks reconstructed from hits in the Inner Tracking System and the Muon System [79]. Global reconstruction considers tracks reconstructed in the Muon System and searches for compatible tracks in the Inner Tracker System. If a matching inner track is found a global muon candidate is reconstructed with momentum determined by a fit to both the hits in the Inner Tracking System and the Muon System. For muon $p_T \gtrsim 200$ GeV the global fit improves the momentum resolution compared to the track only measurement. Tracker muon reconstruction extrapolates tracks in the Inner Tracker System to the Muon System, accounting for expected energy loss and uncertainties due to multiple scattering, and searches for compatible muon hits. The reconstruction of tracker muons is more efficient than global reconstruction for muons with momenta less than 5 GeV.

Backgrounds to prompt muons arise from in-flight decays of hadrons and punch-through of charged hadrons through the calorimeters to the Muon System. These backgrounds may be suppressed by applying muon ID selections based on the properties of the muon track. Requiring hits in the Pixel Detector suppresses in-flight decays that take place beyond the pixel layers. Muons produced in in-flight decays are not in general consistent

with production at the PV and punch-through hadrons lose more energy in the HCAL than muons. As a result by requiring small impact parameters of muon tracks with respect to the PV and good quality global track fits, both backgrounds may be suppressed. Furthermore, since hadrons are typically produced in jets at CMS, applying isolation to reconstructed muons further suppresses the hadron backgrounds. This is done by requiring small sums of track p_T and calorimeter energy in a cone centred on the muon candidate.

The performance of CMS muon reconstruction is measured by reconstructing $J/\psi \to \mu\mu$ and $Z \to \mu\mu$ events. The efficiency of CMS muon reconstruction is greater than 99% for high p_T muons and with loose ID and isolation criteria. Misreconstruction rates are smaller than 1% [79]. The muon momentum resolution is approximately 1% for muons with momentum less than 100 GeV. The CMS measurement of the $Z \to \mu\mu$ cross section, for dimuon invariant masses between 60 and 120 GeV, with $\mathcal{L}_{int} = 36$ pb^{-1} at $\sqrt{s} = 7$ TeV,

$$\begin{aligned}\sigma(pp \to\ Z + X) \times \mathcal{B}(Z \to \mu\mu) = 0.968 \pm 0.008\ \text{(stat.)} \\ \pm 0.007\ \text{(syst.)} \pm 0.018\ \text{(th.)} \pm 0.039\ \text{(lumi.)}\ \text{nb},\end{aligned} \tag{3.8}$$

is in good agreement with theoretical predictions [78].

3.6 τ-jets

As the heaviest of the leptons with a mass of approximately 1.78 GeV, the τ lepton is an important signature in Higgs boson and new physics searches at the LHC [1][50]. The τ decays weakly with a mean lifetime of approximately 291 fs so that prompt τ leptons in CMS decay before reaching the Pixel Detector [1]. τ lepton production must therefore be inferred from the detection of the decay products.

Approximately 35% of τ lepton decays are leptonic with a final state consisting of a ν_τ plus a light lepton (electron or muon) and its corresponding antineutrino [1]. The reconstruction of electrons and muons in CMS is described in Sections 3.4 and 3.5 respectively. The neutrinos do not interact with the CMS detector. The p_T spectra of light leptons produced in τ decays in $Z \to \tau\tau$ events are softer than those in $Z \to ee$ and $Z \to \mu\mu$ events due to the momentum carried by the neutrinos [55][78].

Decay mode	Intermediate resonance	Branching fraction
$\tau^{\pm} \to h^{\pm}\pi^0\nu_\tau$	$\rho^{\pm}(770)$	26.0%
$\tau^{\pm} \to h^{\pm}\nu_\tau$	N/A	11.6%
$\tau^{\pm} \to h^{\pm}h^{\pm}h^{\mp}\nu_\tau$	$a_1^{\pm}(1200)$	9.8%
$\tau^{\pm} \to h^{\pm}\pi^0\pi^0\nu_\tau$	$a_1^{\pm}(1200)$	9.5%
$\tau^{\pm} \to h^{\pm}h^{\pm}h^{\mp}\pi^0\nu_\tau$	N/A	4.5%

Table 3.1: Dominant hadronic τ decay modes and corresponding branching ratios [80][1]. $h^{\pm}$ denotes a charged hadron. The $\tau^{\pm} \to h^{\pm}\pi^0\nu_\tau$, $\tau^{\pm} \to h^{\pm}h^{\pm}h^{\mp}\nu_\tau$ and $\tau^{\pm} \to h^{\pm}\pi^0\pi^0\nu_\tau$ modes proceed via the intermediate resonances $\rho^{\pm}(770)$ and $a_1^{\pm}(1200)$ as shown.

Approximately 65% of τ leptons decay hadronically to the $\nu_\tau + \tau_h$ final state [1]. The largest fraction of these decays proceed via the $\rho^{\pm}(770)$ resonance which decays to a charged hadron and a π^0. Other common modes produce one or three charged hadrons and neutral pions. The charged hadrons are predominantly pions although decays to heavier hadrons also occur. The dominant hadronic τ decay modes are summarised in Table 3.1. Decays with more than three charged hadrons in the final state are rare with branching ratios of less than 0.1%.

The τ decay products are constrained by the τ mass so the hadronic system is highly collimated compared to a quark or gluon jet [80]. Together with the relatively low multiplicity of the decay products, this characteristic allows τ-jets to be differentiated from quark and gluon jets. This is essential in order to fully exploit τ signatures because analyses considering only leptonic τ decays suffer from the smaller leptonic branching ratios. The identification of τ-jets in CMS is described in this Section. Reconstruction of $Z \to \tau\tau$ events and measurement of the $Z \to \tau\tau$ cross section provides a validation of the τ_h reconstruction as described in Chapter 4 and Reference [55].

3.6.1 The Hadron Plus Strips algorithm

In the analyses presented in Chapters 4 and 5, τ_h candidates are reconstructed by the Hadron Plus Strips (HPS) algorithm with which the individual hadronic τ decay modes are identified [80]. The HPS algorithm employs PF to reconstruct the charged hadrons and the photons produced in π^0 decays that comprise τ-jets. The kinematic properties

of the reconstructed τ decay products are then required to be consistent with one of the modes given in Table 3.1.

The HPS algorithm is seeded by PF jets reconstructed as described in Section 3.3. The algorithm is designed to account for spread in ϕ of π^0 energy deposits caused by photon conversions in the Inner Tracking System and the subsequent bending of the conversion electrons in the magnetic field. PF electromagnetic (electron and photon) candidates are clustered into strips in order to collect the converted photon energy. Strips of size $\Delta\eta = 0.05$ and $\Delta\phi = 0.2$ are initially centered on the most energetic electromagnetic candidate in the PF jet. If additional electromagnetic candidates are found within the strip, the most energetic candidate is associated with the strip and strip four-momentum and position are recalculated. This procedure is repeated until no further electromagnetic candidates can be associated with the strip. The invariant mass of the strip is required to be between 50 and 200 MeV so as to be consistent with the π^0 mass.

Only strips with $p_T > 1$ GeV are considered and these strips are combined with PF charged hadrons in the jet to reconstruct the candidate τ decay. All charged hadrons are assumed to be pions for the purpose of four-momentum calculation. Four decay types are considered:

- The Single Hadron type consists of a single charged hadron and no strips and corresponds to either the $\tau^\pm \to h^\pm\nu_\tau$ mode, or the $\tau^\pm \to h^\pm\pi^0\nu_\tau$ mode if the π^0 is not energetic enough to be reconstructed as a strip.
- The Hadron Plus Strip type consists of a single charged hadron and one strip and corresponds to the $\tau^\pm \to h^\pm\pi^0\nu_\tau$ mode, or the $\tau^\pm \to h^\pm\pi^0\pi^0\nu_\tau$ mode if the two neutral pions do not have significant spatial separation and are not distinguished by the PF algorithm. Hadron Plus Strip combinations are required to have an invariant mass compatible with either the $\rho^\pm(770)$ or the $a_1^\pm(1200)$ mass, between 300 and 1300 MeV.
- The Hadron Plus Two Strips type corresponds to the $\tau^\pm \to h^\pm\pi^0\pi^0\nu_\tau$ mode in the case that the neutral pions are well separated.
- τ_h candidates comprising three charged hadrons, required to originate from a common vertex, constitute the Three Hadron type and correspond to the $\tau^\pm \to h^\pm h^\pm h^\mp\nu_\tau$ and $\tau^\pm \to h^\pm h^\pm h^\mp\pi^0\nu_\tau$ modes.

The invariant masses of Hadron Plus Two Strips and Three Hadron combinations are required to be compatible with either the $\rho^{\pm}(770)$ or the $a_1^{\pm}(1200)$ mass, between 800 and 1500 MeV.

In all cases, the system of charged hadrons and strips is required to be within a cone of size $\Delta R = (2.8\text{ GeV})/p_T^{\tau_h}$, where $p_T^{\tau_h}$ is the transverse momentum of the τ_h candidate. This dynamically varying cone size allows for the fact that more boosted decay products are more highly collimated. If a τ_h candidate satisfies the identification criteria of more than one decay type, the type which results in the largest $p_T^{\tau_h}$ is chosen.

In order to exploit the collimated nature of τ-jets and further discriminate against quark and gluon jets, an isolation variable I_τ is constructed from the p_T sums of charged hadron and photon candidates within a cone of size $\Delta R = 0.5$ centred on the τ_h candidate [52]. Neutral hadrons are not considered for isolation due to inadequate HCAL resolution. The effect of PU on the isolation p_T sums is accounted for in the same way as in the construction of the I_e variable used to isolate electron candidates as described in Section 3.4. Only charged hadron candidates consistent with production at the largest Σp_T PV are considered. The p_T sum of charged hadrons originating from other vertices is used to estimate and subtract the contribution to the isolation variable from photons produced in PU interactions. I_τ is given by

$$I_\tau = \frac{\Sigma p_T^{charged} + \Sigma p_T^{photon} - \Delta\beta_\tau}{p_T^{\tau_h}} \tag{3.9}$$

where $\Sigma p_T^{charged}$ and Σp_T^{photon} are the p_T sums of charged hadron and photon candidates respectively within the isolation cone. The PU correction $\Delta\beta_\tau$ is the p_T sum of charged hadron candidates in the isolation cone which are not consistent with production at the PV multiplied by the ratio of photons to charged hadrons in minimum bias events which is approximately 0.4 as shown in Figure 3.4 [68]. Requiring $I_\tau < 0.2$ results in an efficiency of approximately 1% for quark and gluon jets to be reconstructed as τ_h candidates [80].

$Z \to \tau\tau \to \mu + \tau_h$ events are used to study the performance of the τ_h reconstruction [80]. The muon, well separated from hadronic activity, provides a clean reference from which to probe the properties of the τ_h. Figure 3.10 shows the τ decay modes reconstructed by the HPS algorithm for τ-jets in candidate $Z \to \tau\tau \to \mu + \tau_h$ events in data and MC. As expected, the dominant reconstructed modes are of the Hadron Plus Strip and Hadron

Plus Two Strips types with smaller contributions of Single Hadron and Three Hadron types. The rates of reconstruction of each mode in data and MC are in good agreement.

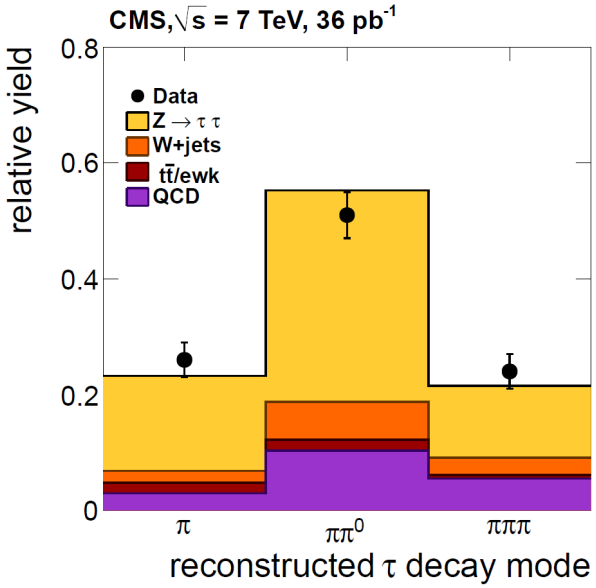


Figure 3.10: Distribution of hadronic τ decay modes reconstructed by the HPS algorithm in candidate $Z \to \tau\tau \to \mu + \tau_h$ events in data and MC [80]. π denotes the Single Hadron decay type, $\pi\pi^0$ the Hadron Plus Strip and Hadron Plus Two Strips types and $\pi\pi\pi$ the Three Hadron type.

3.6.2 Electron and muon rejection

Electrons such as those produced in $Z \to ee$ events leave isolated energy deposits in the calorimeters and may therefore result in fake τ_h signatures [80]. In these cases, the electron is misidentified as a charged hadron produced in a τ decay. The PF reconstruction of electrons is highly suited to the identification of electrons within jets [69]. Such electrons are therefore identified and rejected using the PF electron discriminator ξ described in Section 3.2. τ_h candidates in which the leading, i.e. highest p_T, charged hadron constituent has $\xi > -0.1$ are rejected [80]. Approximately 2% of electrons in $Z \to ee$ events pass this selection so that further electron rejection is generally required.

Electrons usually deposit all of their energy in the ECAL while τ-jets deposit much of their energy in the HCAL [56]. The fraction of τ_h candidate energy deposited in the HCAL may then be used to discriminate against electrons. For this purpose, the variable H/P is defined as the ratio of HCAL energy associated with the τ_h candidate to the momentum of the leading charged hadron [52]. If the leading charged hadron candidate

is an electron that has passed the $\xi < -0.1$ requirement, it is a badly reconstructed electron likely to have undergone significant bremsstrahlung. Additional discrimination against electrons may therefore be achieved by estimating the fraction of energy radiated by the leading charged hadron candidate under the assumption that it is an electron. The estimated bremsstrahlung fraction, f_{brem}, is given by the ratio of ECAL energy within $\Delta\eta = 0.03$ of the leading charged hadron to the total photon energy associated with the τ_h candidate [52]. The invariant mass of the τ_h candidate, m_{τ_h}, is also used to discriminate against electrons. τ_h candidates misreconstructed from electrons have a falling m_{τ_h} spectrum, while the spectrum of real τ-jets peaks at the $\rho^{\pm}(770)$ mass [52].

To achieve strong suppression of electrons, τ_h candidates of the Single Hadron type are required to satisfy $H/P > 0.08$ [52]. Candidates of all other types are required to satisfy either $f_{brem} < 0.99$ or $m_{\tau_h} > 0.55$ GeV. Electrons passing through the uninstrumented regions between the EB and EE deposit most of their energy in the HCAL so that the H/P and f_{brem} variables may fail to discriminate against them. All τ_h candidates with leading charged hadron constituents passing through these regions are therefore rejected. The efficiency for electrons to pass all of these rejection criteria and still fake a τ_h signature, the e-τ_h fake rate, is measured in data as described in Section 4.3.4 [52].

In rare cases, muon energy depositions in the calorimeters may give rise to fake τ_h signatures [80]. Such signatures are suppressed by requiring that the track of the leading charged hadron constituent is not also reconstructed as a tracker muon as described in Section 3.5.

3.6.3 Identification efficiency

The efficiency for τ-jets in MC $Z \to \tau\tau$ events to pass the HPS ID and isolation criteria described in Section 3.6.1 is shown in Figure 3.11 as a function of τ_h p_T for τ-jets with $|\eta| < 2.3$. For τ-jets with $p_T > 20$ GeV the efficiency is approximately 50% [80].

The difference between ID efficiency in data and MC is generally required to be understood in physics analyses. The efficiency is measured in data using the tag and probe method [80]. A sample of events is selected by a single muon trigger. Candidate $Z \to \tau\tau \to \mu + \tau_h$ events are selected offline by requiring the presence of a global muon reconstructed as described in Section 3.5, the tag, that is spatially matched to the muon trigger object. The probe is defined by a a PF jet candidate with $p_T > 20$ GeV and $|\eta| < 2.3$. The leading charged hadron in the jet is required to have p_T greater than 5 GeV in

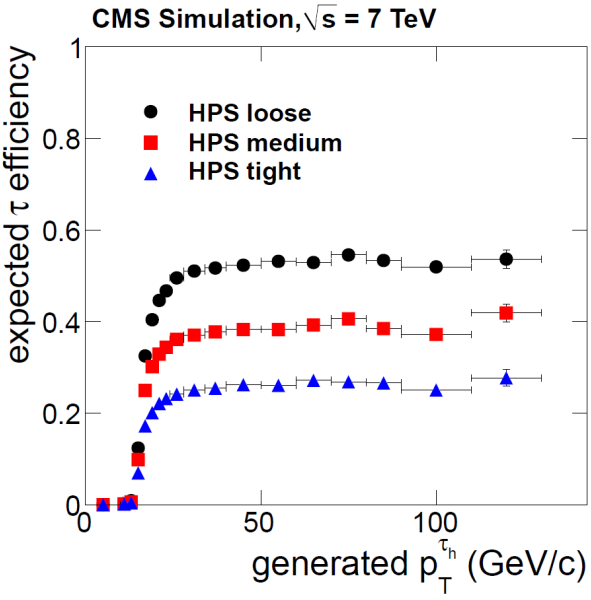


Figure 3.11: τ_h ID efficiency as a function of τ_h p_T for τ-jets with $|\eta| < 2.3$ in MC $Z \to \tau\tau$ events [80]. HPS loose refers to the isolation selection described in Section 3.6.1 and used in the analyses presented in Chapters 4 and 5. HPS medium and HPS tight refer to tighter isolation requirements.

order to increase the fraction of real τ-jets in the sample. This requirement biases the efficiency measurement but still permits comparison between data and MC. Pass and fail subsamples are defined by events in which the probe passes and fails respectively the HPS ID and isolation criteria. The tag and probe invariant mass distributions in the pass and fail subsamples are shown in Figure 3.12 with estimates of the $Z \to \tau\tau$ and background contributions derived from MC. The pass subsample is dominated by $Z \to \tau\tau$ events containing real τ-jets while the fail subsample is dominated by backgrounds of fake τ_h candidates. The $Z \to \tau\tau$ yields in each subsample are determined by fits to the tag and probe mass distributions with $Z \to \tau\tau$ and background shapes derived from MC. The measured efficiency is given by the ratio of $Z \to \tau\tau$ events in the pass subsample to the total number of $Z \to \tau\tau$ events in both samples.

The efficiency measurement is repeated for MC events. The ratio of the efficiency measured in data to the efficiency in MC is found to be unity with a 6% uncertainty [80][52]. The dominant systematic contributions to the uncertainty arise from uncertainties on track reconstruction and on the efficiency for jets to pass the leading charged hadron p_T requirement.

The efficiency for quark and gluon jets to pass the τ_h ID criteria, the misidentification rate, is measured in data using dijet and W+jet events [80]. In both cases the measured

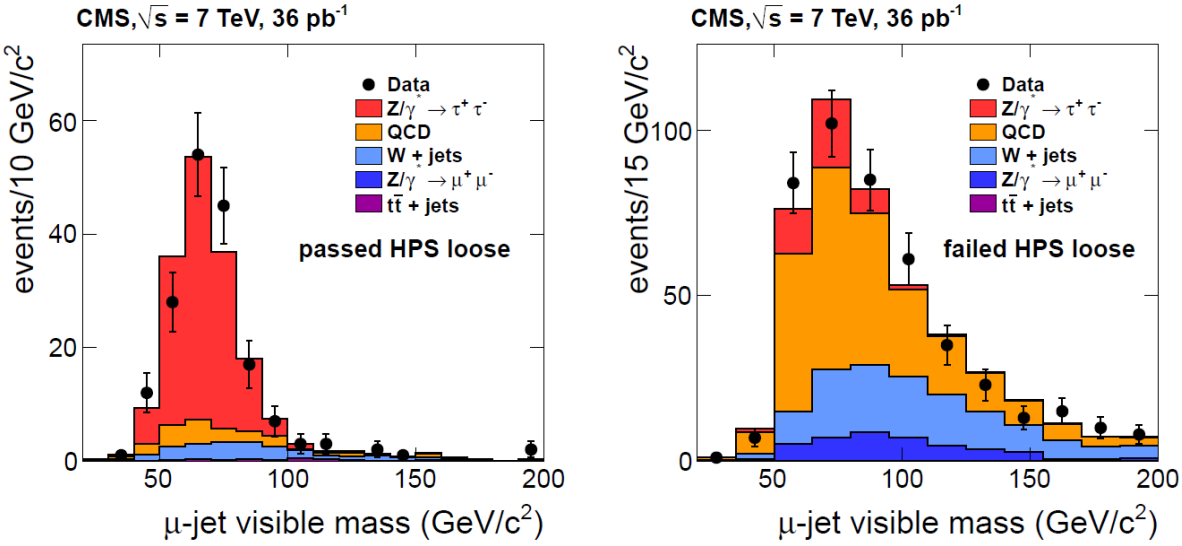


Figure 3.12: $\mu + \tau_h$ invariant mass distributions in candidate $Z \to \tau\tau \to \mu + \tau_h$ events in which the τ_h candidate passes or fails the HPS ID and isolation criteria, in data and MC [80]. The τ_h ID efficiency is measured by evaluating the ratio of $Z \to \tau\tau$ events in the pass sample to the total number of $Z \to \tau\tau$ events in both samples.

misidentification rate is $\mathcal{O}(1\%)$ for jets with $p_T = 20$ GeV and $\mathcal{O}(0.1\%)$ for jets with $p_T = 150$ GeV. The misidentification rates in MC events are compatible with those measured in data within uncertainties of approximately 5%. In the analyses presented in Chapters 4 and 5 the MC is relied upon to predict the misidentification rates except in the cases of the W+jets and QCD multijet backgrounds which are normalised by data driven methods.

3.6.4 Energy scale

As τ-jets are reconstructed using the PF algorithm, the measurement of their energies benefits from the high momentum and energy resolutions of the Inner Tracking System and ECAL respectively. The uncertainty on the reconstructed τ_h energy is estimated to be 3% from studies of $Z \to \tau\tau \to \mu + \tau_h$ events [80]. The reconstructed $\mu + \tau_h$ mass is very sensitive sensitive to the τ_h energy scale as the muon momenta are measured with high precision. $Z \to \tau\tau$ and background shapes derived from MC are fitted to the observed $\mu + \tau_h$ invariant mass distribution with the τ_h energy scale allowed to vary. The best fit is obtained by scaling the MC τ_h energy scale by a factor of 0.97 ± 0.03.

3.6.5 τ-jet triggers

In the analyses presented in Chapters 4 and 5, events are selected based on the reconstruction of τ-jets by the HLT algorithms. In the HLT, τ-jets are identified by a simpler cone algorithm which allows for greater flexibility in offline selections [56][52]. The cone algorithm identifies τ-jets as narrow, isolated jets of PF candidates as in the case of the HPS algorithm, but does not attempt to reconstruct the individual τ decay modes [63]. The jets are required to contain a leading charged hadron with $p_T > 5$ GeV. The remaining candidate τ-jet constituents are required to be contained within a narrow signal cone centred on the leading charged hadron with a radius shrinking with increasing leading charged hadron p_T. Isolation is applied by requiring no additional charged hadron or photon candidates above certain p_T thresholds in an isolation annulus of radius $\Delta R = 0.5$ centred on the leading charged hadron. Varying tightness of isolation may be applied by varying the p_T thresholds of the particle candidates considered in the isolation annulus.

3.7 b-jets

The decays of b quarks are suppressed by small CKM matrix elements. As a result, the lifetimes of b-flavoured hadrons, produced in the fragmentation of b quarks, are relatively long; $\mathcal{O}(1$ ps) [1]. Due to the relatively large mass of the b quark, approximately 4 GeV, the decay products of b-hadrons typically have large momenta perpendicular to the b-hadron momentum. This results in the production of a relatively wide jet with respect to a light (up, down or strange quark or gluon) jet. These two features as well as the significant semileptonic branching ratios of b-hadrons allow such b-jets to be distinguished experimentally from light jets. The identification of b-jets is important for measurements of matter-antimatter asymmetries, studies of top quark production and searches for new physics including searches for Higgs boson production in association with b quarks [50].

Due to the long lifetime of b-hadrons, their decay vertices are significantly displaced from the PV allowing for reconstruction of the secondary decay vertex [56]. This method is used to tag b-jets in the $\Phi \to \tau\tau$ search presented in Chapter 5. The decays of b-hadrons to final states containing muons are exploited to provide an independent measurement of the b-tagging efficiency [81].

Secondary Vertex (SV) candidates are reconstructed from the tracks associated with a jet reconstructed as described in Section 3.3 [82]. The distance of closest approach between a track and the jet direction is required to be less than $\Delta R = 0.3$ in order to suppress tracks originating from PU interactions. The position of the SV is evaluated by the AVF algorithm, described in Section 3.1. SV candidates that share more than 65% of their tracks with the PV or are displaced by greater than 4 cm from the PV are rejected. SV candidates inconsistent with the jet direction with respect to the PV are also rejected.

The flight distance significance of an SV candidate is defined by the ratio of the three-dimensional distance between the PV and the SV to its uncertainty. The distribution of flight distance significance for SV candidates reconstructed in data is shown in Figure 3.13 together with the distributions in MC events containing light jets, charm flavoured jets and b-jets. SV candidates with large flight distance significance are dominated by b-jets in the simulation.

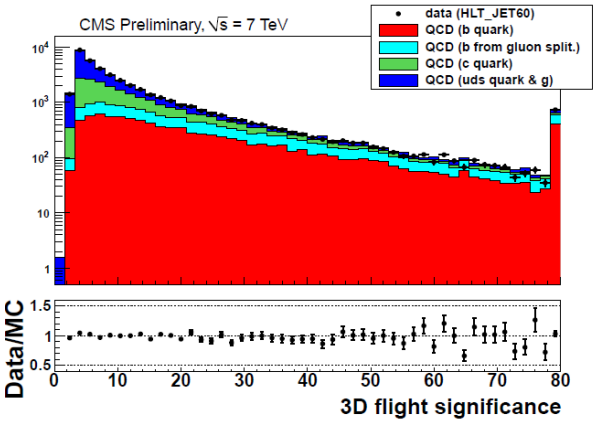


Figure 3.13: SV flight distance significance of SV candidates reconstructed in data and MC [82]. The MC events are divided into samples containing b quarks, c quarks and light quarks (up, down or strange) or gluons only. The bin-by-bin ratio of data to MC is also shown.

A Simple Secondary Vertex (SSV) discriminator is defined as a monotonic function of the SV flight distance significance and is used to tag b-jets in physics analyses. Simple Secondary Vertex High Efficiency (SSVHE) and Simple Secondary Vertex High Purity (SSVHP) variants are defined by considering only SV candidates with more than one or two tracks respectively. The outputs of these discriminators are denoted d_{SSVHE} and d_{SSVHP} respectively and take values between one and five. Outputs of five correspond to the most b-jet like candidates and outputs of one correspond to the most light jet like candidates. The distributions of d_{SSVHE} and d_{SSVHP} are shown in Figure 3.14 for jets reconstructed in data and for MC b, c and light jets. Depending on the requirements of

a physics analysis, b-jets may be tagged by requiring $d_{SSVHE} > 1.74$ for which light jets have an approximate 1% efficiency, or by requiring $d_{SSVHP} > 2$ for which light jets have an approximate 0.1% efficiency. Using the SSV algorithms, measurements of the cross section for Z boson production in association with b quarks at CMS are found to be in agreement with theoretical predictions [83].

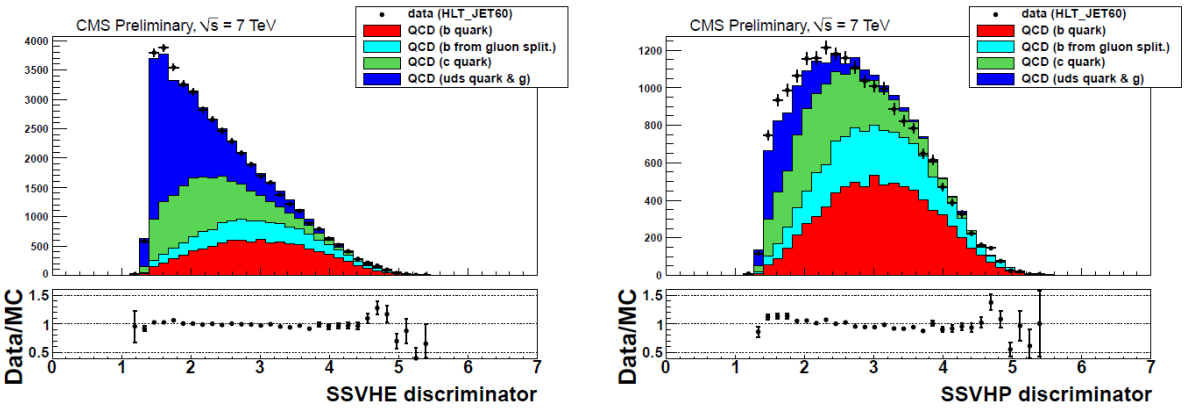


Figure 3.14: Distributions of the SSVHE and SSVHP discriminators for jets reconstructed in data and MC [82]. The MC events are divided into samples containing b quarks, c quarks and light quarks (up, down or strange) or gluons only. The bin-by-bin ratio of data to MC is also shown

3.7.1 *b*-tagging efficiency

Understanding of the b-tagging efficiency is essential in order to minimise systematic uncertainties in physics analyses that employ b-tagging [50]. Typically the ratio between the efficiency measured in data and the efficiency in MC and the uncertainty on this ratio are the quantities of interest. b-hadrons have relatively large semileptonic branching ratios (approximately 40% when cascade decays via charm flavoured states are considered) and muons in b-jets have larger average momentum perpendicular to the jet momentum, $p_{\perp}^{rel}$, than muons in light jets [1]. Muon jets, i.e. jets containing muons, therefore offer an alternative method with which to identify b-jets [81]. Furthermore, the efficiencies of b-jet ID using muon jets and using SVs are uncorrelated so that muon jets can be used to measure the b-tagging efficiency of the SSV algorithms.

The efficiency of the SSV b-tagging algorithms is measured in data using dijet events [81]. One of the jets is required to contain a global muon, reconstructed as described in Section 3.5, with $p_T > 5$ GeV and the second jet is required to be tagged as a b-jet by one of the SSV algorithms. The resulting event sample has a high b-jet purity. $p_{\perp}^{rel}$

distributions of muons in b-jets and light jets are obtained from MC. These distributions are then fitted to the $p_{\perp}^{rel}$ distribution of the muon jets in the dijet sample in order to determine their b-jet fraction. The events are then divided into two subsamples, one in which the muon jets are tagged as b-jets by the SSV algorithm and one in which they fail the b-tagging requirements. The b-tagging efficiency, ϵ_b, for the SSV algorithm is then given by

$$\epsilon_b = \frac{f_{tag} \cdot N_{tag}}{f_{tag} \cdot N_{tag} + f_{fail} \cdot N_{fail}} \tag{3.10}$$

where f_{tag} and f_{fail} are the extracted b-jet fractions in the muon jets which pass and fail the SSV b-tagging respectively, and N_{tag} and N_{fail} are the total numbers of events in which the muon jets pass and fail the SSV b-tagging respectively.

The dominant systematic uncertainties on the measurement of the b-tagging efficiency arise from the muon p_T threshold and the modelling of gluon splitting, $g \to bb$, in the MC used to estimate the $p_{\perp}^{rel}$ distributions. The value of the muon p_T cut changes the shape of the $p_{\perp}^{rel}$ distribution and the size of this effect is estimated by repeating the efficiency measurement with different p_T thresholds. The size of the uncertainty due to the modelling of gluon splitting is estimated by changing the fraction of events that contain gluon splitting in the MC.

The b-tagging efficiency is measured in bins of muon jet p_T and and is compared to the efficiencies for MC b-jets to give data/MC correction factors. The correction factors and uncertainties for the SSVHP algorithm with the b-tag criterion $d_{SSVHP} > 2$ are are given in Table 3.2. These correction factors are used in the $\Phi \to \tau\tau$ analysis presented in Chapter 5.

3.8 Missing transverse energy

Neutrinos and hypothetical weakly interacting stable particles do not interact with the CMS detector. Their production must be inferred from the resultant $\not\!\!E_T$ in an event [56]. The hermeticity of the CMS detector is sufficient to measure the $\not\!\!E_T$ in an event to the accuracy required to establish a physics signature involving weakly interacting stable particles [50]. This is important in many searches for new physics as well as in the

Jet p_T / GeV	Correction factor	Stat. uncertainty	Syst. uncertainty
$20 \leq p_T < 30$	1.000	$\pm$0.048	$\pm$0.057
$30 \leq p_T < 40$	0.995	$\pm$0.042	$\pm$0.073
$40 \leq p_T < 50$	0.948	$\pm$0.037	$\pm$0.069
$50 \leq p_T < 60$	0.912	$\pm$0.036	$\pm$0.060
$60 \leq p_T < 70$	0.829	$\pm$0.037	$\pm$0.055
$70 \leq p_T < 80$	0.948	$\pm$0.042	$\pm$0.063
$80 \leq p_T < 90$	0.900	$\pm$0.047	$\pm$0.125
$100 \leq p_T < 120$	0.863	$\pm$0.056	$\pm$0.120
$p_T \geq 120$	0.794	$\pm$0.036	$\pm$ 0.167

Table 3.2: data/MC correction factors and uncertainties for the SSVHP b-tagging efficiency [81].

reconstruction of leptonically decaying W bosons which constitute a background to $Z \to \tau\tau$ and $\Phi \to \tau\tau$ events.

The PF algorithm is highly suited to the reconstruction of $\not\!\!E_T$ as it provides a global reconstruction of all particles produced in an event with an optimal combination of the CMS subdetectors. PF reconstructed $\not\!\!E_T$ is defined as $-\Sigma E_T$ where ΣE_T is the sum of the transverse energies of all particles reconstructed in the event [66]. The E_T contributions from jets with $p_T > 10$ GeV are corrected by the jet energy corrections described in Section 3.3.1 in order to minimise the effects of inhomogeneous and non-linear calorimeter response [84]. Jets below the 10 GeV p_T threshold are not corrected because the uncertainties on the corrections for such jets are very large.

Figure 3.15 shows the reconstructed $\not\!\!E_T$ distribution in dijet events. While such events have no intrinsic $\not\!\!E_T$, calorimeter noise, uncorrected calorimeter response in the case of unclustered energy, and particles passing through poorly instrumented regions of the detector lead to the reconstruction of some artificial $\not\!\!E_T$ [65]. It is important that this artificial $\not\!\!E_T$ is understood. Reasonable agreement is seen between the $\not\!\!E_T$ distribution in data and MC. The distribution in MC is somewhat narrower than that observed in data due to imperfect modelling of calorimeter response in the simulation.

Figure 3.16 shows the reconstructed $\not\!\!E_T$ distribution in events containing a high p_T isolated electron candidate. The presence of a $W \to e\nu$ signal is clearly observed demonstrating the ability of the PF $\not\!\!E_T$ reconstruction to isolate events containing high

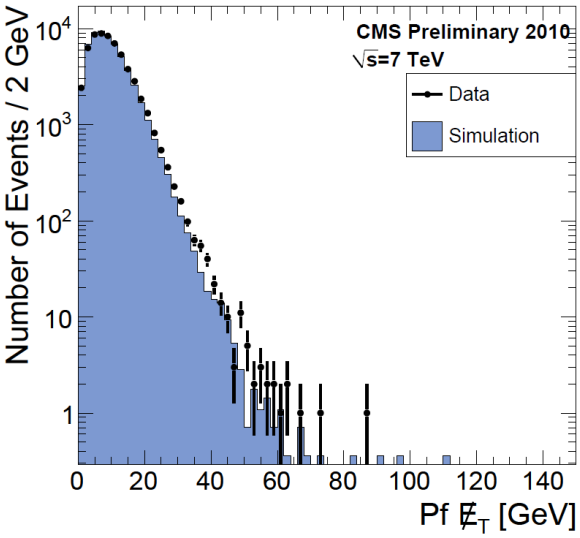


Figure 3.15: Reconstructed $\not\!\!E_T$ distribution in dijet events in data and MC [65].

p_T neutrinos [78]. This is important for the analyses presented in Chapters 4 and 5 where $\not\!\!E_T$ is used to both suppress and estimate backgrounds arising from W production.

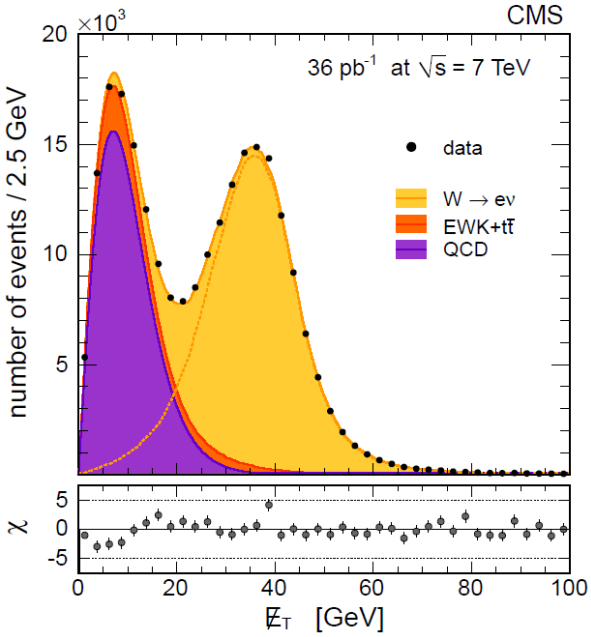


Figure 3.16: Reconstructed $\not\!\!E_T$ in events containing a high p_T isolated electron candidate [78]. The estimates of the $W \to e\nu$ and background contributions are also shown. The normalised difference between data and the combined $W \to e\nu$ plus background estimate is shown as a function of $\not\!\!E_T$.

Chapter 4

Measurement of the $Z \to \tau\tau$ cross section

Measurement of the $Z \to \tau\tau$ cross section provides a test of the SM as well as a benchmark for new physics searches using τ leptons as discussed in Section 1.5.3. Measurement of a $Z \to \tau\tau$ cross section in agreement with SM predictions validates the τ_h reconstruction as well as the background estimation methods used in the Higgs search presented in Chapter 5. The measurement presented in this Chapter considers the $e + \tau_h$ final state in the CMS Run 2011A dataset. It is based on a previous CMS analysis conducted with a smaller dataset corresponding to $\mathcal{L}_{int} = 36$ pb^{-1} [55].

Details of the online event selection and the MC samples used to study the signal and background processes are given in Section 4.1. Corrections applied to the MC are discussed in Section 4.3. The measurement is challenging due to large backgrounds arising primarily from QCD multijet, $Z \to ee$ and W production. The event selection described in Section 4.2 suppresses the QCD background by applying ID and isolation criteria to the electron and τ_h candidates. Additional dedicated selections are applied to suppress the $Z \to ee$ and W backgrounds. Data driven methods are used to estimate the QCD and W background contributions as described in Section 4.4. The signal is extracted by a fit to the reconstructed visible $e + \tau_h$ mass, $m_{e+\tau_h}$, distribution as described in Section 4.5. The evaluation of the cross section and discussion of systematic uncertainties are presented in Section 4.6.

	L1 seed	HLT e p_T threshold / GeV	HLT τ_h isolation	$\mathcal{L}_{int}$ / fb^{-1}
Trigger 1	EG12	15	Loose	1.10 $\pm$ 0.05
Trigger 2	EG12	15	Tight	0.75 $\pm$ 0.03
Trigger 3	EG15	18	Medium	0.25 $\pm$ 0.01

Table 4.1: Properties of the trigger selections. Three isolation criteria with varying tightness are applied to the τ_h candidate depending on $\mathcal{L}$ and the τ_h p_T is required to be greater than 20 GeV in all cases. The corresponding integrated luminosity of the data collected with each trigger is shown in the right hand column.

4.1 Data and Monte Carlo samples

The data comprise the CMS Run 2011A dataset and correspond to $\mathcal{L}_{int} = 2.1 \pm 0.1$ fb^{-1} at $\sqrt{s} = 7$ TeV. Events are selected by the EG12 and EG15 triggers at L1 as described in Section 3.4.4. The events are required to pass HLT selections requiring the presence of loosely isolated electron and τ_h candidates. The available single electron triggers are unsuitable due to high p_T thresholds which reduce the acceptance of $Z \to \tau\tau$ events. As $\mathcal{L}$ increases during Run 2011A, different HLT selections with increasing tightness are applied in order to maintain an acceptable trigger rate. Due to momentum loss to the neutrinos in a τ decay the visible τ decay products are relatively soft, thus tightening the isolation requirements on the trigger objects as $\mathcal{L}$ increases is favoured over raising the p_T thresholds.

The HLT selection requires an electron candidate, reconstructed as described in Section 3.4.4, with p_T greater than 15 or 18 GeV depending on $\mathcal{L}$. The isolation applied to the electron trigger object is nominally looser than the offline selection given in Section 4.2. A τ_h candidate with $p_T > 20$ GeV is also required, reconstructed as described in Section 3.6.5. The isolation applied to the τ_h trigger object is tightened with increasing $\mathcal{L}$ and is nominally looser than the isolation applied in the offline selection. The properties of the trigger selections are summarised in Table 4.1. Measurement of the trigger efficiencies is described in Section 4.3.2. Additional triggers are used to study the QCD background and are described in Section 4.4.3.

MC samples of signal and background events are generated in order to study the respective processes. $Z \to \tau\tau$ and $Z \to ee$ events are simulated with the POWHEG generator which achieves NLO accuracy for inclusive quantities including the Z boson p_T and rapidity [85][86][87]. This approach is therefore more suited to the $Z \to \tau\tau$ cross section

Sample	Generator	Final state	Phase space	Generated events
$Z \to \tau\tau$	POWHEG	$\tau\tau + X$	$m_{\tau\tau} > 20$ GeV	19,937,479
$Z \to ee$	POWHEG	$ee + X$	$m_{ee} > 20$ GeV	29,497,207
W+jets	MADGRAPH	$l\nu + X$	inclusive	81,352,581
$t\bar{t}$ +jets	MADGRAPH	inclusive	inclusive	3,701,947
WW	PYTHIA	inclusive	inclusive	4,225,916
WZ	PYTHIA	inclusive	inclusive	4,265,243
ZZ	PYTHIA	inclusive	inclusive	4,187,885

Table 4.2: Properties of the MC samples.

measurement than a LO matrix element implementation as the $Z \to \tau\tau$ acceptance is expected to be more accurately described [88]. The background contributions due to W and $t\bar{t}$ production are significantly dependent on the presence of additional jets in the event. Samples of W+jets and $t\bar{t}$ +jets events are therefore produced with the MADGRAPH LO matrix element generator where the associated jet production is described with LO accuracy [89]. Diboson (WW, WZ and ZZ) events constitute a small background and are generated with PYTHIA [90]. All processes are simulated at $\sqrt{s} = 7$ TeV. Hadronisation and particle decays are simulated with PYTHIA with the exception of τ decays which are simulated with TAUOLA [91]. Interactions with the CMS detector are simulated with GEANT [92]. The properties of the MC samples are summarised in Table 4.2. Samples of multijet and γ+jets events with adequate statistics are unavailable and these backgrounds are studied entirely by data driven methods.

The results of the trigger decisions are not available in the MC samples and no trigger selection is applied to the simulated events. Instead the MC event yields are corrected by the trigger efficiencies measured in data as described in Section 4.3.2. For the purpose of comparing data to MC, the simulated $Z \to \tau\tau$, $Z \to ee$ and W+jets events are normalised to the integrated luminosity of the data using NNLO cross sections, times branching ratios to $\tau\tau$, ee and $l\nu$ respectively. The $Z \to \tau\tau$ and $Z \to ee$ cross sections are calculated with FEWZ for $m_{ee}, m_{\tau\tau} > 20$ GeV, where m_{ee} and $m_{\tau\tau}$ are the invariant masses of the generator level lepton pairs, and are equal to 1.666 nb [54]. The $W \to l\nu$ cross section is also calculated with FEWZ and is equal to 31.314 nb. The $t\bar{t}$ and diboson samples are normalised according to the background estimates given in Section 4.4.

4.2 Offline event selection

The event selection is identical to that of the inclusive $\Phi \to \tau\tau$ search presented in Chapter 5 and is optimised for sensitivity to a Higgs boson signal as described in Section 5.2 [93]. As $Z \to \tau\tau$ production constitutes an irreducible background to the $\Phi \to \tau\tau$ process these selections maintain high efficiency for $Z \to \tau\tau$ events. Their application in the measurement of the $Z \to \tau\tau$ cross section allows for the most rigorous cross check of the methods used in the Higgs search. The selections require a well reconstructed pair of electron and τ_h candidates which are isolated so as to suppress the QCD background. Additional requirements are imposed to suppress the $Z \to ee$ and W backgrounds. To ensure genuine pp collision events are selected, events are required to contain at least one PV, reconstructed as described in Section 3.1, with $|z| < 24$ cm, transverse impact parameter less than 2 cm and at least five degrees of freedom.

Electron candidates are reconstructed as described in Section 3.4.1 and are required to have $p_T > 20$ GeV and $|\eta| < 2.1$. The p_T threshold is below the plateau of the trigger turn-on curve but is favoured over a higher threshold in order to maintain efficiency for Higgs events. The variation of the trigger efficiency with electron p_T is accounted for as described in Section 4.3.2. The pseudorapidity cut ensures the electrons are within the acceptance of the Pixel Detector. Only candidates with transverse and longitudinal impact parameters less than 0.045 cm and 0.2 cm respectively with respect to the largest Σp_T PV are considered in order to suppress non-prompt electrons. The cuts applied to the electron ID and isolation variables described in Section 3.4.2 are given in Table 4.3. An additional selection is applied to suppress electrons produced in photon conversions; if a conversion vertex, identified as described in Section 3.4.2, is reconstructed from the electron candidate track and another track in the event, and the electron track has no hits between the PV and the candidate conversion vertex, the electron candidate is rejected. The efficiency of the electron selection is measured in data as described in Section 4.3.3.

τ_h candidates are reconstructed by the HPS algorithm as described in Section 3.6.1 and are required to have $p_T > 20$ GeV and $|\eta| < 2.3$. As with the electron candidates, the p_T threshold is below the plateau of the trigger turn-on curve and the variation of the trigger efficiency with τ_h p_T is accounted for as described in Section 4.3.2. The pseudorapidity cut ensures the τ_h candidates are within the acceptance of the Silicon Strip Tracker. The longitudinal impact parameter of the τ_h candidate is required to be less than 0.2 cm with respect to the largest Σp_T PV in order to suppress events

	Candidate electron selection		Veto electron selection	
Variable	Barrel	Endcap	Barrel	Endcap
$\sigma_{i\eta i\eta}$	< 0.01	< 0.03	< 0.012	< 0.031
$\Delta\eta_{in}$	< 0.004	< 0.007	< 0.007	< 0.011
$\Delta\phi_{in}$	< 0.06	< 0.03	< 0.8	< 0.7
H/E	< 0.04	< 0.025	N/A	N/A
N_{miss}	< 1	< 1	< 1	< 1
I_e	< 0.1	< 0.1	< 0.3	< 0.3

Table 4.3: Cuts applied to the electron ID variables. The candidate electron selections apply to the electrons considered as candidate τ daughters in the $Z \to \tau\tau$ decay. The veto electron selections apply to additional electrons in the event which are considered in the dielectron veto.

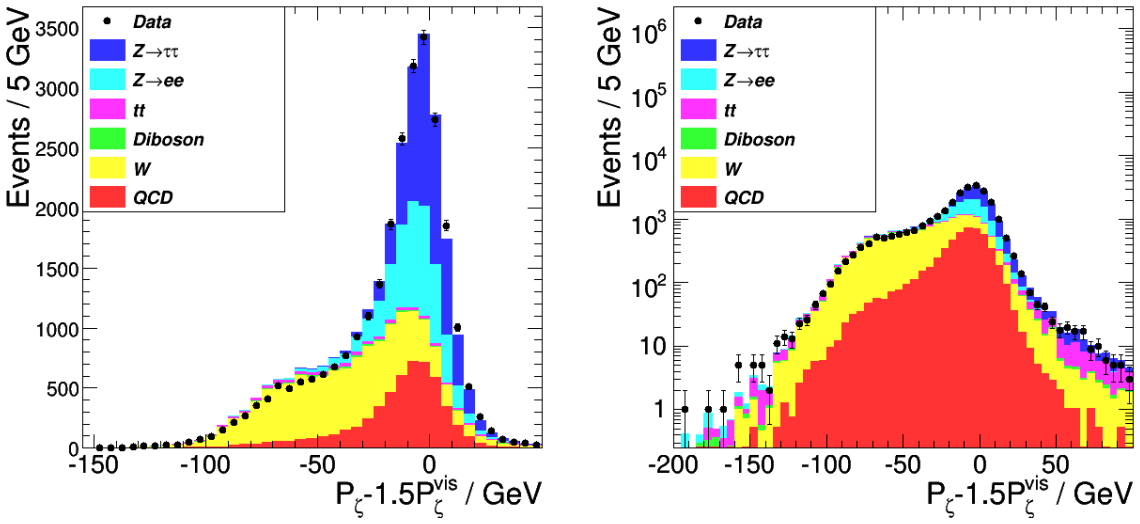
in which the electron and τ_h candidates are produced at different PVs. The isolation, electron rejection and muon rejection selections described in Section 3.6 are applied to suppress fake τ_h candidates reconstructed from quark or gluon jets, electrons and muons respectively. The measurement of the τ_h ID efficiency is described in Section 3.6.3.

Events containing an electron and a τ_h candidate satisfying these identification criteria are selected. The electron and τ_h candidates are further required to be separated by $\Delta R > 0.5$ and to be of opposite charge. These selections ensure the reconstructed electron and τ_h do not correspond to the same physical object and that they are consistent with a Z boson decay.

A dielectron veto is applied to suppress the $Z \to ee$ background; events containing a second well reconstructed electron of opposite charge to the candidate signal electron are rejected. The electron candidates considered in the veto are required to pass the same kinematic selections as the candidate signal electron but with looser ID and isolation criteria as given in Table 4.3.

The topological difference between a W+jet event and a $Z \to \tau\tau$ event is exploited to suppress the W background. In a W+jet event in which the jet fakes the electron or τ_h signature, the jet is recoiling against the W boson such that the system of the candidate visible τ decay products is expected to have significant azimuthal separation from the $\not\!\!E_T$. In a $Z \to \tau\tau$ event the neutrinos produced in the τ decays are approximately collinear with the visible τ decay products so that the $\not\!\!E_T$ is expected to bisect the azimuthal angle between the electron and τ_h candidates. The variable $P_\zeta - 1.5P_\zeta^{vis}$ is constructed to

discriminate between the two topologies. P_ζ is the projection of the visible transverse τ decay momentum and $\not{E}_T$ onto the bisector of the azimuthal angle between the visible τ decay products. P_ζ^{vis} is the projection of the visible transverse τ decay momentum only onto this bisector [94]. $\not{E}_T$ is reconstructed as described in Section 3.8. Events with $P_\zeta - 1.5P_\zeta^{vis} < -20$, which are dominated by W events, are rejected. The distributions of $P_\zeta - 1.5P_\zeta^{vis}$ in data and MC are shown in Figure 4.1.



(a) (b)

Figure 4.1: Distribution of $P_\zeta - 1.5P_\zeta^{vis}$ in events passing all selections except the cut on this variable, in linear (a) and logarithmic (b) scale. The $Z \to \tau\tau$ and W MC distributions are normalised with their respective NNLO cross sections and the other background contributions are normalised according to the methods given in Section 4.4. Events with $P_\zeta - 1.5P_\zeta^{vis} < -20$ are rejected.

17,848 events in the data pass all the event selections. The p_T and η distributions of the electron and τ_h candidates after all selections are shown Figures 4.2 and 4.3 respectively. The $Z \to \tau\tau$ MC in combination with the background estimates given in Section 4.4 accurately describes the observed p_T and η distributions of the selected electron candidates. A deficiency of τ_h candidates with $20 < p_T < 25$ GeV in the MC results from differences in the τ_h ID efficiency between data and MC that are not accounted for in the efficiency measurement described in Section 3.6.3. For $20 < p_T < 25$ GeV, the τ_h ID efficiency is not constant as a function of τ_h p_T as shown in Figure 3.11. While the unbinned ID efficiency in MC is found to be compatible with the efficiency in data, the same cannot be assumed for the turn-on region $20 < p_T < 25$ GeV. This discrepancy is discussed further in Section 4.4.4.

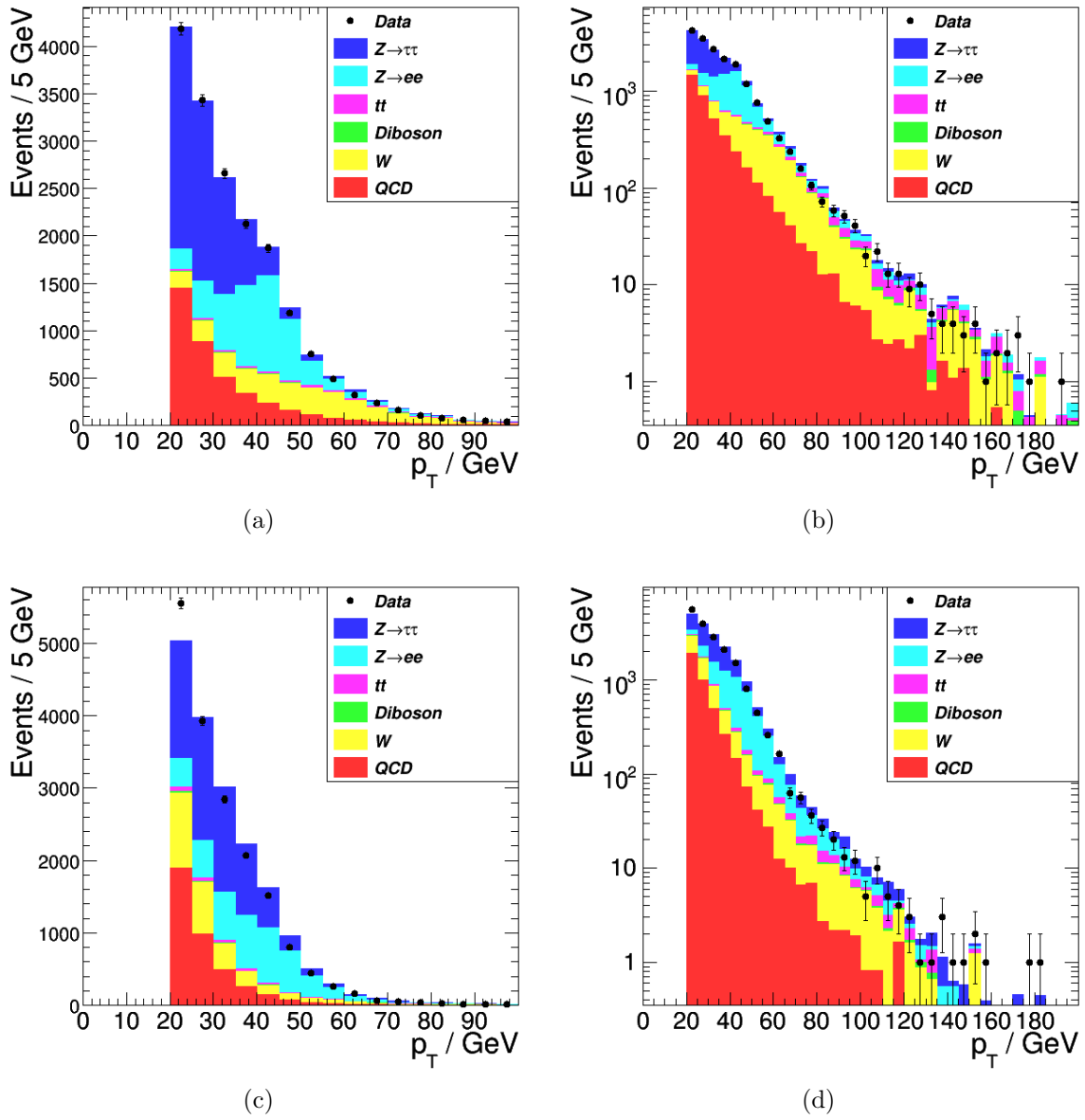
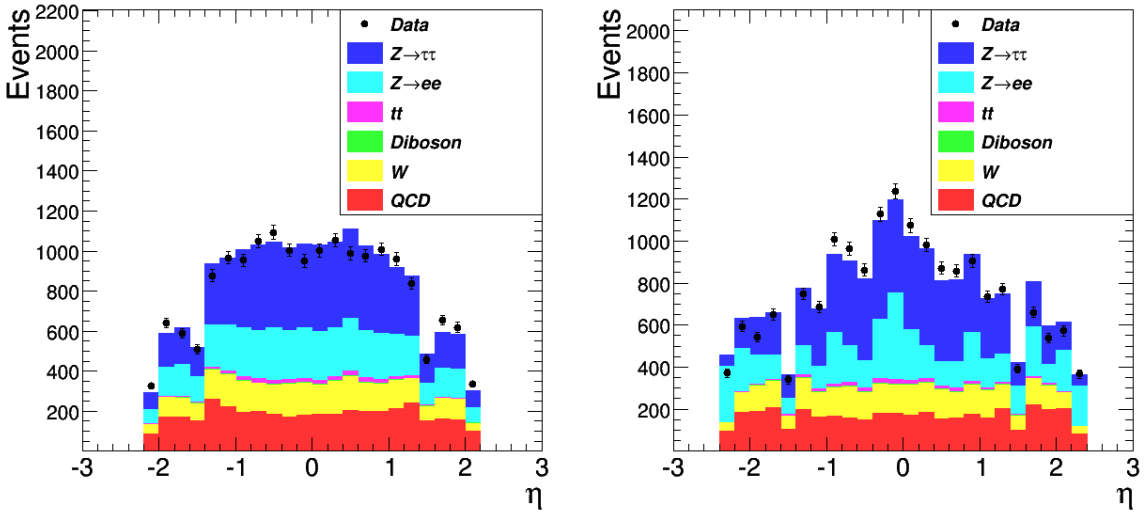


Figure 4.2: p_T distributions of the selected electron (a, b) and τ_h (c, d) candidates in linear (a, c) and logarithmic (b, d) scale. The $Z \to \tau\tau$ MC is normalised with the NNLO cross section and the background contributions correspond to the the estimates given in Section 4.4.



(a) (b)

Figure 4.3: η distributions of the selected electron (a) and τ_h (b) candidates. The $Z \to \tau\tau$ MC is normalised with the NNLO cross section and the background contributions correspond to the the estimates given in Section 4.4.

4.3 Monte Carlo correction factors

The MC simulation does not perfectly reproduce the data with regard to the efficiencies of the event selections. Efficiencies are therefore measured in data and the MC event yields are corrected to agree with measurement. The MC is also corrected to account for the difference between generated PU rates and those observed in the data.

4.3.1 Pile-Up reweighting

To simulate the effect of PU, the MC samples are generated with minimum bias events superimposed on the nominal event with a PU distribution flat in PU interaction multiplicity up to ten PU interactions. The MC is corrected for the difference between this distribution and that observed in the data to prevent the MC selection efficiencies being biased by erroneous rates of reconstruction of particles produced in PU interactions. The correction is applied by assigning weights w_i^{PU} to the MC events where the index i is the number of PU interactions in the MC event.

$$w_i^{PU} = \frac{N_i^{data} N_{MC}}{N_i^{MC} N_{data}} \tag{4.1}$$

N_i^{data} and N_i^{MC} are the numbers of events in the data and MC with i PU interactions and N_{data} and N_{MC} are the total numbers of data and MC events respectively. The distribution of PU multiplicity in the data is known from the measurement of $\mathcal{L}$ as described in Section 2.3. The reweighting procedure is seen to accurately reproduce the distribution of reconstructed PV multiplicity, N_{PV}, in the data as shown in Figure 4.4.

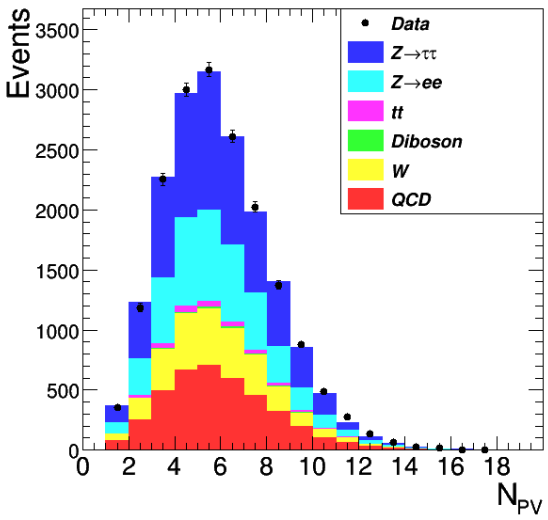


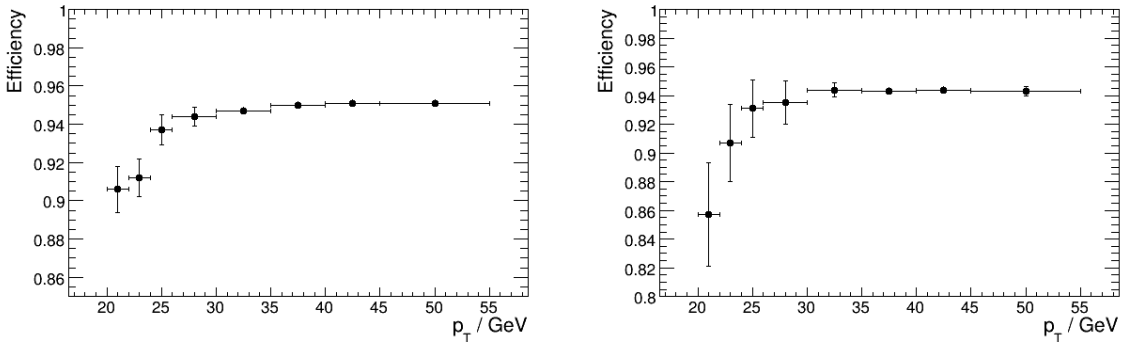
Figure 4.4: Reconstructed PV multiplicity in selected events after applying PU reweighting to the MC. The $Z \to \tau\tau$ MC is normalised with the NNLO cross section and the background contributions correspond to the the estimates given in Section 4.4.

4.3.2 Trigger efficiency

The efficiencies of the trigger selections are measured in data using the tag and probe method [78]. As the offline p_T thresholds are not on the plateaux of the trigger turn-on curves it is particularly important to measure the efficiencies in bins of electron and τ_h p_T. The online electron and τ_h selection efficiencies are measured separately and are applied directly to the MC as event weights dependent on the p_T and η of the offline electron and τ_h candidates.

To measure the efficiency of the online electron selections events are selected by a dedicated tag and probe trigger, seeded by EG12 at L1, which requires a HLT electron

candidate with $p_T > 17$ GeV passing loose ID and isolation requirements. An additional supercluster with $E_T > 8$ GeV is required to be reconstructed at HLT level and the invariant mass of the electron-supercluster pair is required to be greater than 30 GeV. The offline tag and probe electrons are required to pass all the electron selections given in Section 4.2. The tag electron is required to be spatially matched to the electron trigger object with $\Delta R < 0.3$. The selection of only an additional supercluster at trigger level, as opposed to an electron candidate, ensures the efficiency measurement is unbiased by trigger level electron selections. The tag and probe pair are required to have a mass between 60 and 120 GeV so that the event sample is dominated by $Z \to ee$ events. Two subsamples are defined by events in which the the probe electron passes or fails the trigger selections given in Section 4.1. The trigger efficiency is derived as the ratio of events in the pass subsample to the number of events in the pass and fail subsamples. As all the offline electron selections are applied to both the tag and the probe and the pair are constrained to be on the Z resonance, background contributions to the event sample are negligible. The measurement is performed separately for each of the online electron selections with 15 and 18 GeV p_T thresholds, for probes in the barrel and endcap, and in bins of probe p_T. Tighter binning is used at low p_T in order to measure the behaviour of the trigger turn-on. To obtain the total efficiency for given probe p_T and η the average efficiency of the 15 and 18 GeV online selections, weighted by the fractions of the total integrated luminosity of the data selected by each trigger, are evaluated. The measured total efficiencies are shown in Figure 4.5.



(a) (b)

Figure 4.5: Measured efficiency of the electron trigger selection, in bins of electron p_T, in the barrel (a) and endcap (b). The plotted efficiencies are the $\mathcal{L}_{int}$ weighted averages of the efficiencies of the different electron trigger selections given in Table 4.1.

To measure the efficiencies of the online τ_h selections events are selected by a single muon trigger [93]. Candidate $Z \to \tau\tau \to \mu + \tau_h$ events are selected offline by requiring the

presence of a muon, the tag, that is spatially matched to the muon trigger object. The probe τ_h candidate is required to pass all the τ_h selections given in Section 4.2. The tag and probe pair are required to have a mass between 45 GeV and 70 GeV so that the event sample is dominated by $Z \to \tau\tau$ events. The pass and fail subsamples are defined by events in which the probe passes and fails the trigger selections given in Section 4.1 respectively. The $Z \to \tau\tau$ yields in each subsample are determined by fits to the tag and probe mass distributions with the $Z \to \tau\tau$ and background shapes derived from MC. The efficiency, given by the ratio of the $Z \to \tau\tau$ yield in the pass subsample to the yield in the pass and fail subsamples, is measured separately for each of the three online τ_h selections and in bins of probe p_T. Tighter binning is used at low p_T in order to measure the behaviour of the trigger turn-on. As with the electron trigger efficiencies, the $\mathcal{L}_{int}$ weighted averages of the efficiencies of the three online τ_h selections are evaluated to give total efficiencies which are shown in Figure 4.6.

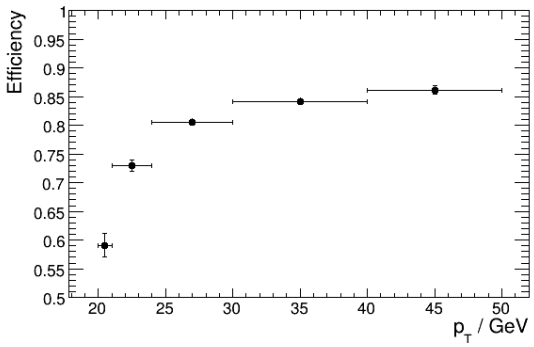


Figure 4.6: Measured efficiency of the τ_h trigger selection in bins of τ_h p_T [93]. The plotted efficiencies are the $\mathcal{L}_{int}$ weighted averages of the efficiencies of the different τ_h trigger selections given in Table 4.1.

4.3.3 Electron identification and isolation efficiencies

The tag and probe method is also used to measure the efficiency of the offline electron ID and isolation selections [78][93]. Events are selected by the tag and probe trigger described in Section 4.3.2. The offline tag electron is required to pass all the electron selections given in Section 4.2 and to be spatially matched to the electron trigger object. The probe electron is not required to pass ID or isolation criteria. The tag and probe pair are required to have a mass between 60 and 120 GeV so that the event sample is dominated by $Z \to ee$ events. The pass and fail subsamples are defined by events in which the probe passes and fails the electron ID and isolation requirements respectively.

p_T bin	Data	MC	Correction factor
$20 < p_T < 25$	0.542 ± 0.007	0.552 ± 0.003	0.982 ± 0.014
$25 < p_T < 30$	0.626 ± 0.005	0.641 ± 0.002	0.976 ± 0.009
$p_T \geq 30$	0.761 ± 0.002	0.772 ± 0.001	0.986 ± 0.003

Table 4.4: Measured efficiency of the electron ID and isolation selections in bins of electron p_T together with the efficiencies in MC $Z \to ee$ events and the resultant MC correction factors [93].

The $Z \to ee$ yields in each subsample are determined by fits to the tag and probe mass distribution with the $Z \to ee$ shape derived from MC and the background shape modelled by an exponential function. The efficiency is given by the ratio of the $Z \to ee$ yield in the pass subsample to the yield in the pass and fail subsamples and is measured in three bins of probe p_T; $20 < p_T < 25$, $25 < p_T < 30$ and $p_T > 30$ GeV. Data/MC correction factors are derived as the ratios of the measured efficiencies to the efficiencies in $Z \to ee$ MC events and are shown in Table 4.4. The correction factors are applied to MC event yields in the background estimation procedure described in Section 4.4.

4.3.4 Electron to τ-jet fake rate

The efficiency for electrons to pass the τ_h ID criteria, the e-τ_h fake rate, must be known in order to accurately model the $Z \to ee$ background. The e-τ_h fake rate is measured in data using the tag and probe method [80][52]. A sample of $Z \to ee$ events is defined by events containing two OS electron candidates, passing loose ID and isolation criteria, with an invariant mass between 60 and 120 GeV. One of the electron candidates, the probe, is required to be also reconstructed as an HPS τ_h candidate passing the isolation criteria given in Section 3.6.1. The pass and fail subsamples are defined by events in which the probe passes and fails the τ_h electron rejection selections respectively. The $Z \to ee$ yields in each subsample are determined by fits to the tag and probe mass distribution with the $Z \to ee$ and background shapes derived from MC. The measured e-τ_h fake rate, 0.0081 ± 0.0005, is defined by the ratio of the $Z \to ee$ yield in the pass subsample to the yield in the pass and fail subsamples. The fake rate in the $Z \to ee$ MC is 0.0088 ± 0.0001 giving a data/MC correction factor of 0.92 ± 0.06 which is applied to the $Z \to ee$ MC yield in the background estimation procedure described in Section 4.4.

4.4 Background estimation

Due to momentum lost to neutrinos in τ decays the visible mass shape of $Z \to \tau\tau \to e{+}\tau_h$ events is relatively broad with a peak extending across a range of approximately 30 to 100 GeV. The $Z \to \tau\tau$ signal is extracted by a fit to the $m_{e+\tau_h}$ distribution so that accurate estimation of the large backgrounds which also occupy the same mass range is essential.[1] The dominant backgrounds are due to $Z \to ee$, W and QCD multijet production with smaller contributions of $t\bar{t}$ and diboson events. The characteristics of these backgrounds and the methods used to estimate their sizes and mass shapes are described in this Section.

4.4.1 $Z \to ee$, $t\bar{t}$ and diboson

$Z \to ee$, $t\bar{t}$ and diboson production at CMS has been shown to be well described by MC. MC is therefore relied upon for the estimation of these backgrounds in conjunction with the MC correction factors given in Section 4.3 [78][96][97].

$Z \to ee$ events can fake an $e + \tau_h$ signature in two ways. The second electron may be reconstructed as a τ_h, or this electron can escape acceptance or be badly reconstructed in which case an additional jet in the event may fake the τ_h signature. The former type constitutes 86 $\pm$ 1% of MC $Z \to ee$ events passing the full event selection. The $Z \to ee$ MC is normalised to the integrated luminosity of the data using the NNLO cross section given in Section 4.1. The event yield is corrected for the trigger and electron selection efficiencies. The NNLO calculation is in excellent agreement with the CMS measurement and the 3% uncertainty of the CMS measurement is taken as the systematic uncertainty on the normalisation together with the 4.5% uncertainty on $\mathcal{L}_{int}$ [78][60]. The $Z \to ee$ sample is divided into two fractions, one in which the second electron fakes the τ_h and the other in which the τ_h is faked by an additional jet (POWHEG describes the production of one associated jet at tree level) [87]. The former fraction is further corrected for the e-τ_h fake rate. Uncertainties on the event yield from the trigger efficiency (1%), electron selection efficiency (1%) and e-τ_h fake rate (6%) are considered. The estimate of the $Z \to ee$ background is 4,280 $\pm$ 660 events.

[1] Mass reconstruction and resolution may be improved by estimating the τ pair invariant mass, $m_{\tau\tau}$. $m_{\tau\tau}$ may be reconstructed by maximising a likelihood function with respect to the neutrino momenta with constraints from the measured visible τ decay kinematics, $\not{E}_T$, and the τ mean lifetime. Such methods may provide increased separation between the signal and background mass distributions [95].

Electrons and τ-jets produced in semileptonic top decays in $t\bar{t}$ events can fake a $Z \to \tau\tau \to e{+}\tau_h$ signature. This background is relatively small and is estimated by normalising the $t\bar{t}$ +jets MC events to the integrated luminosity of the data using CMS measurements of the $t\bar{t}$ cross section in the single lepton, dilepton and all hadronic channels. The combined measurement of the inclusive $t\bar{t}$ cross section is 165.8 $\pm$ 2.2 (stat.) $\pm$ 10.6 (syst.) $\pm$ 7.8 (lumi.) pb [96]. The MC event yield is corrected for the trigger and electron selection efficiencies. Uncertainties on the integrated luminosity (4.5%), cross section (8%, not including the luminosity uncertainty), trigger efficiency (1%), electron selection efficiency (1%) and τ_h ID efficiency (6%) are considered. The uncertainty on the τ_h ID efficiency is discussed in Section 3.6.3. The estimate of the $t\bar{t}$ background is 290 $\pm$ 60 events.

Diboson events with final states containing electrons and τ-jets can also fake a $Z \to \tau\tau \to e + \tau_h$ signature. These backgrounds are also relatively small and are estimated by normalising the diboson MC events to the integrated luminosity of the data using CMS inclusive cross section measurements given by [97]

$$\sigma(pp \to WW + X) = 55.4 \pm 3.3 \text{ (stat.)} \pm 6.9 \text{ (syst.)} \pm 3.3 \text{ (lumi.) pb}, \quad (4.2)$$

$$\sigma(pp \to WZ + X) = 17.0 \pm 2.4 \text{ (stat.)} \pm 1.1 \text{ (syst.)} \pm 1.0 \text{ (lumi.) pb}, \quad (4.3)$$

$$\sigma(pp \to ZZ + X) = 3.8^{+1.5}_{-1.2} \text{ (stat.)} \pm 0.2 \text{ (syst.)} \pm 0.2 \text{ (lumi.) pb}. \quad (4.4)$$

The diboson MC yields are corrected for the trigger and electron selection efficiencies. Uncertainties on the cross sections and the trigger, electron and τ_h selection efficiencies are considered. The estimate of the total diboson background is 70 $\pm$ 20 events.

The $m_{e+\tau_h}$ shapes of the $Z \to ee$, $t\bar{t}$ and diboson backgrounds are taken from the MC and their uncertainties are discussed in Section 4.5.1. The estimation procedure is repeated for Same Sign (SS) events which pass all event selections with the exception that the electron-τ_h pair are required to be of same charge. This information is then used to correct the estimate of the QCD background as described in Section 4.4.4.

4.4.2 *W*

$W \to e\nu$ and $W \to \tau\nu$ events can fake an $e + \tau_h$ signature when an additional jet in the event is reconstructed as an electron or τ_h. A W enriched control region is defined by $P_\zeta - 1.5 P_\zeta^{vis} < -40$ with all other event selections applied. This control region is dominated by W events as shown in Figure 4.7. The disagreement between data and

MC in this region demonstrates the need to estimate the W background in a data driven way. Using the ratio of MC W+jets events in the control region to events in the signal region, R_W, the W contribution in the signal region is extrapolated from the event yield in data in the control region. The rate of W events passing the electron and τ_h selection criteria is therefore estimated using data while the MC is relied upon only to describe the $P_\zeta - 1.5P_\zeta^{vis}$ distribution of the events. By considering SS $e + \tau_h$ combinations the same procedure is used to correct the estimate of the QCD background as described in Section 4.4.4.

The estimate of the W background, N_W^{OS}, and the corresponding estimate for SS events, N_W^{SS}, are given by

$$N_W^{OS} = \left(N_{data}^{controlOS} - N_{t\bar{t}}^{controlOS} - N_{Z\to ee}^{controlOS}\right) \cdot R_W, \tag{4.5}$$

$$N_W^{SS} = \left(N_{data}^{controlSS} - N_{t\bar{t}}^{controlSS} - N_{Z\to ee}^{controlSS}\right) \cdot R_W \tag{4.6}$$

where $N_{data}^{controlOS}$ and $N_{data}^{controlSS}$ are the numbers of OS and SS events in the data in the control region respectively. $N_{t\bar{t}}^{controlOS}$, $N_{t\bar{t}}^{controlSS}$, $N_{Z\to ee}^{controlOS}$ and $N_{Z\to ee}^{controlSS}$ are the small contributions of $t\bar{t}$ and $Z \to ee$ events in the OS and SS control regions respectively. These contributions are estimated by the procedure described in Section 4.4.1. The W background is normalised by this method in all plots of the signal region (i.e. $P_\zeta - 1.5P_\zeta^{vis} > -20$) shown in this Chapter and in Chapter 5.

The uncertainty on R_W is determined by rescaling the energies of the reconstructed electron and τ_h candidates in the W+jets MC by their respective energy scale uncertainties as given in Sections 3.4.3 and 3.6.4. A 5% shift in the $\not{E}_T$ scale is also considered. The deviation of R_W under each energy scale shift is taken as a systematic uncertainty and the uncertainties derived from the energy scale shifts of each of the three objects are added in quadrature to give a 6% total uncertainty on R_W. The estimate of the W background is 2,730 $\pm$ 200 events where the uncertainty includes the statistical uncertainty on the data yield in the control region and $R_W = 0.576 \pm 0.035$. The shape of the W background is taken from the MC.

4.4.3 QCD

Jets and non-prompt leptons in QCD multijet and photons in γ+jets events may fake an $e + \tau_h$ signature. The multijet and γ+jets backgrounds are considered together as one QCD background, the size and shape of which is estimated with data using SS $e + \tau_h$

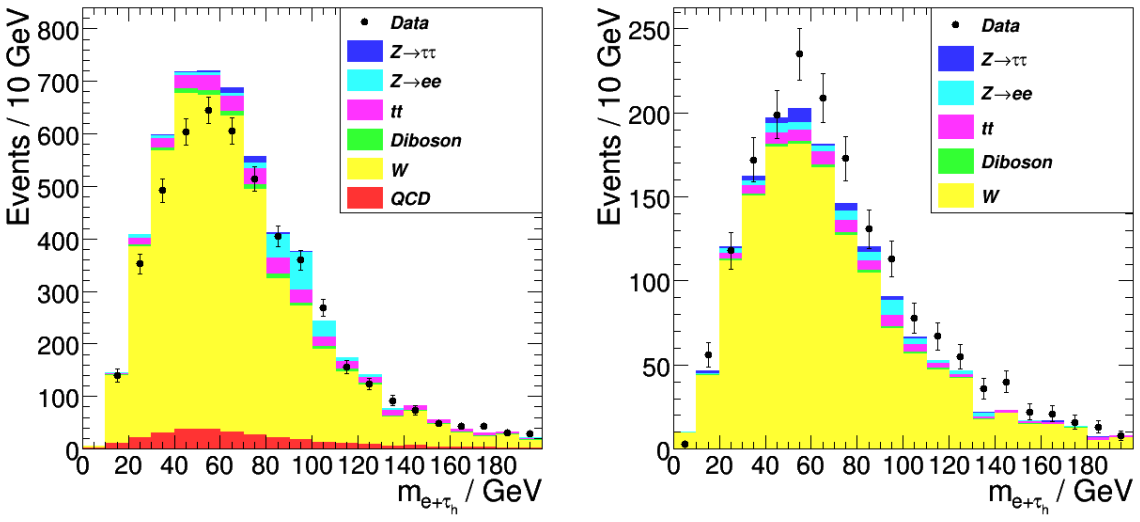


Figure 4.7: $m_{e+\tau_h}$ distributions in the OS (a) and SS (b) W control regions defined by $P_\zeta - 1.5P_\zeta^{vis} < -40$ GeV after all other selections. The $Z \to \tau\tau$ and W MC are normalised with their respective NNLO cross sections and the other background contributions are normalised according to the methods given in Sections 4.4.1 and 4.4.3. A reliable estimate for the QCD contribution to the SS region is not available due to inadequate MC statistics.

events. The selection of SS events suppresses the signal and $Z \to ee$ processes but not QCD processes as there is no charge correlation between fake electron and τ_h candidates in QCD events. SS $e + \tau_h$ events are therefore expected to be dominated by QCD events. The visible mass distribution of these events is shown in Figure 4.8. The size of the QCD background is estimated by subtracting the significant contamination of other backgrounds from the SS region and correcting the resultant yield by the ratio of OS to SS events, $R_{OS/SS}$, measured in data with inverted electron isolation. Signal and other background events are also suppressed in the inverted isolation region so that the OS/SS ratio measured in data in this region approximates the ratio in QCD events.

The inverted electron isolation region is defined by $I_e > 0.2$. This region is studied in data selected by triggers identical to those described in Section 4.1 except that no isolation is applied to the electron trigger objects. The data selected by these triggers correspond to an effective integrated luminosity of 38 ± 2 pb^{-1} after accounting for the trigger prescaling. Figure 4.9 shows the electron isolation variable I_e for OS and SS events after all other selections. Contamination of the inverted isolation region by $Z \to \tau\tau$, $Z \to ee$, $t\bar{t}$, diboson and W events is seen to be negligible in the MC. The OS/SS ratio measured

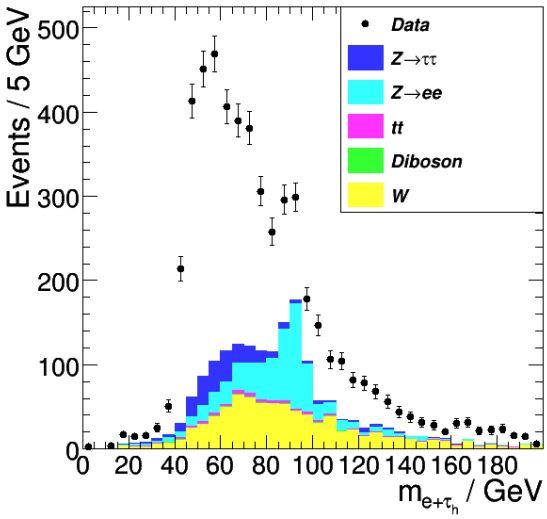


Figure 4.8: Visible mass distribution of SS $e+\tau_h$ events after all other selections. The $Z \to \tau\tau$ MC is normalised with the NNLO cross section and the background contributions correspond to the the estimates given in Sections 4.4.1 and 4.4.2. The SS region is expected to be dominated by QCD events which are assumed to account for the difference between data and MC.

in the inverted isolation region in the data is $R_{OS/SS} = 1.13 \pm 0.05$. Figure 4.10 shows the behaviour of this ratio as the inverted isolation requirement is tightened. The OS/SS ratio is not significantly sensitive to the value at which the isolation cut is inverted for cuts of $I_e > 0.1$ or tighter. For looser inverted isolation cuts, the OS/SS ratio is biased by the presence of signal events with isolated electron candidates.

The estimate of the QCD background, N_{QCD}, is given by

$$N_{QCD} = \left(N^{SS}_{data} - N^{SS}_{W} - N^{SS}_{t\bar{t}} - N^{SS}_{Z\to ee}\right) \cdot R_{OS/SS} = 3,990 \pm 350 \qquad (4.7)$$

where N^{SS}_{data} is the event yield in data in the SS region and $N^{SS}_{Z\to ee}$ and $N^{SS}_{t\bar{t}}$ are the $Z \to ee$ and $t\bar{t}$ contributions in the SS region which are estimated by the procedure given in Section 4.4.1. The uncertainty on N_{QCD} includes the uncertainty on $R_{OS/SS}$ and the statistical uncertainties on the data yields in the SS and W control regions.

Figure 4.11 shows the OS and SS $e + \tau_h$ mass distributions in data in inverted isolation regions defined by four different inverted isolation cuts. The OS and SS shapes are compatible with each other and do not depend significantly on the value at which the isolation cut is inverted. This allows the QCD background shape in the signal region to be estimated with data using SS events and loose electron isolation. Figure 4.12 shows the $e+\tau_h$ visible mass distribution in the SS region with a loose electron isolation of $I_e <$

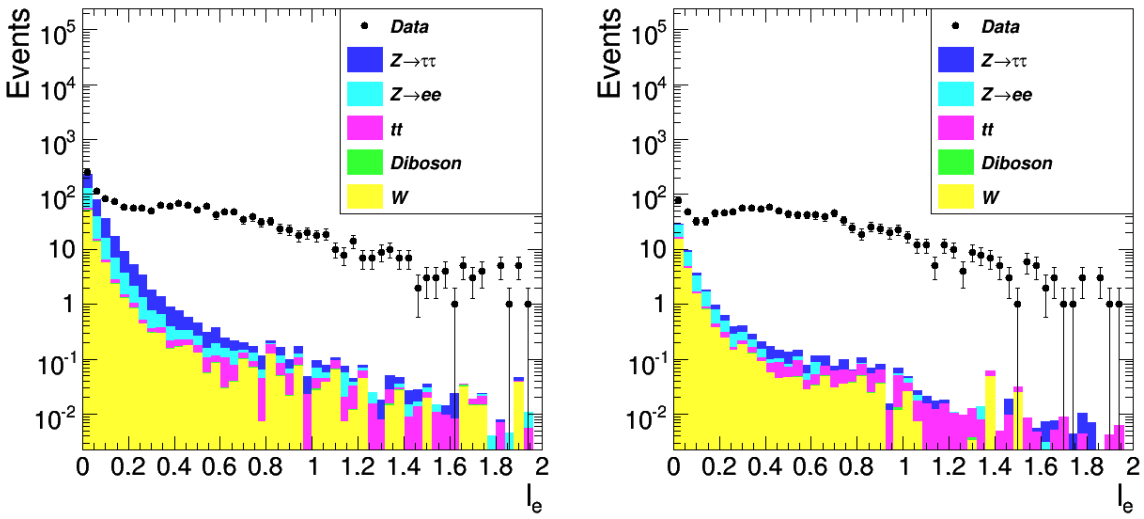


Figure 4.9: Distributions of the electron isolation variable I_e for OS (a) and SS (b) $e+\tau_h$ pairs after all selections except the cut on I_e. The data are selected by triggers which do not require the electron trigger candidate to be isolated and correspond to $\mathcal{L}_{int}$ = 38 pb^{-1}. The $Z \to \tau\tau$ MC is normalised with the NNLO cross section and the background contributions correspond to the the estimates given in Sections 4.4.1 and 4.4.2. A reliable estimate for the QCD contribution is not available due to inadequate MC statistics. The inverted isolation region defined by $I_e > 0.2$ is expected to be dominated by QCD events.

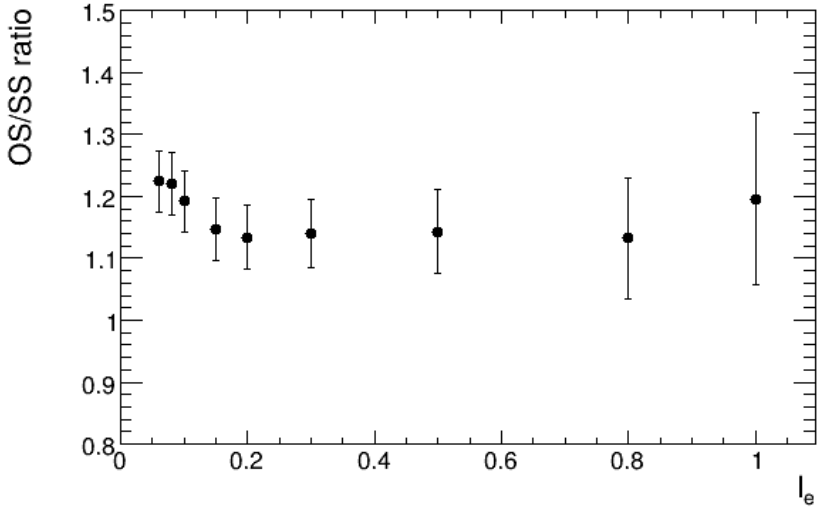


Figure 4.10: The ratio of OS to SS $e + \tau_h$ events as a function of inverted electron isolation after all other selections. The OS/SS ratio is measured in data selected by triggers which do not require the electron trigger candidate to be isolated, corresponding to $\mathcal{L}_{int}$ = 38 pb^{-1}. The OS/SS ratio used in the estimation of the QCD background is defined at $I_e > 0.2$.

0.3 applied. With this selection the contamination from $Z \to \tau\tau$ and other backgrounds is reduced and the distribution of the data is used as the estimate of the QCD background shape.

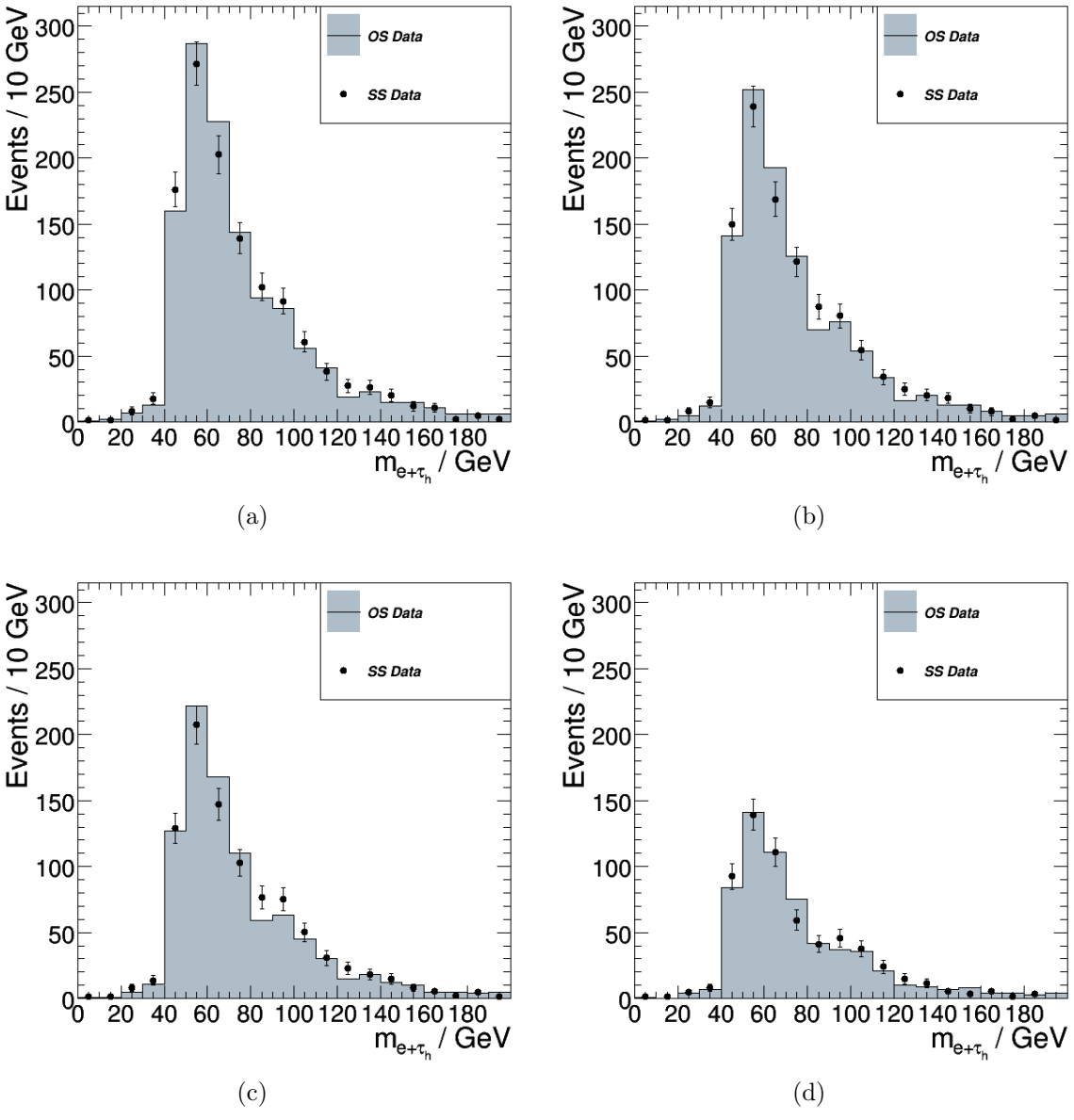


Figure 4.11: Visible mass distributions of OS and SS $e + \tau_h$ events with $I_e > 0.1$ (a), $I_e > 0.2$ (b), $I_e > 0.3$ (c) and $I_e > 0.5$ (d) after all other selections. The data are selected by triggers which do not require the electron trigger candidate to be isolated and correspond to $\mathcal{L}_{int} = 38$ pb^{-1}. The SS distributions are normalised to the number of OS events to allow for comparison of the OS and SS $m_{e+\tau_h}$ shapes. The OS/SS ratio used in the estimation of the QCD background is defined at $I_e > 0.2$.

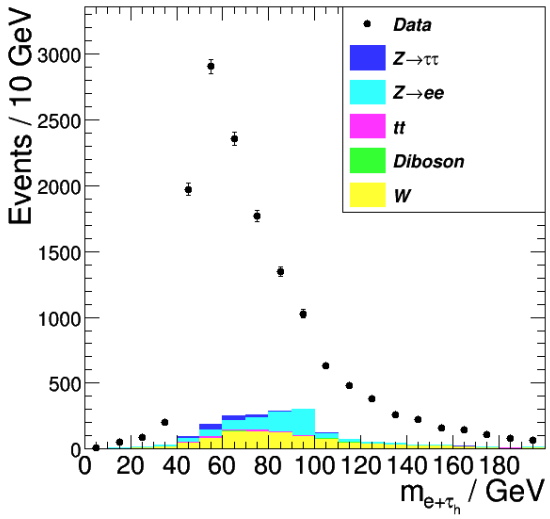


Figure 4.12: Visible mass distribution of SS $e + \tau_h$ events with $I_e < 0.3$ after all other selections. The $Z \to \tau\tau$ MC is normalised with the NNLO cross section and the background contributions correspond to the the estimates given in Sections 4.4.1 and 4.4.2. The distribution of the data is an estimate of the $m_{e+\tau_h}$ shape of the QCD background.

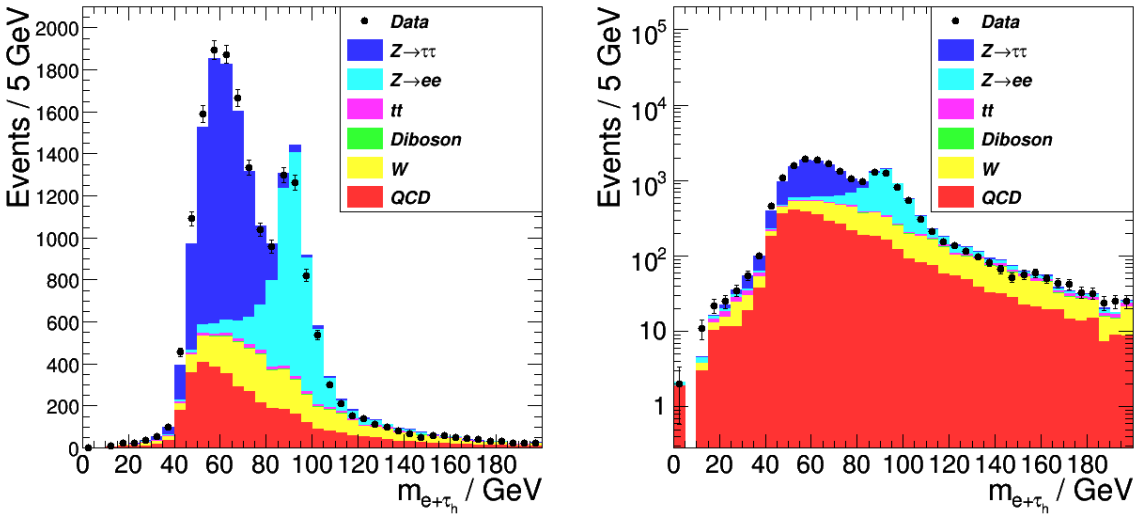
4.4.4 Background estimation results

The estimates of the background contributions are summarised in Table 4.5 and, taken together with the $Z \to \tau\tau$ yield predicted by MC, are in good agreement with the observed data. The estimate of the total background contribution is 11,360 $\pm$ 780 events. The $e + \tau_h$ mass distribution, with background shapes and sizes corresponding to the estimates described in this section, is shown in Figure 4.13.

In $m_{e+\tau_h}$ region in which the $Z \to \tau\tau$ signal peaks, approximately 40 to 80 GeV, the data is in excess of the sum of the background estimate and the $Z \to \tau\tau$ MC prediction. This excess is compatible with the data within the uncertainties on the background estimates and the uncertainty on the τ_h ID effiency, and results from the deficit of MC τ_h candidates with $20 < p_T < 25$ GeV as discussed in Section 4.2. For events with candidate τ_h $p_T > 25$ GeV, good agreement of the background estimate and $Z \to \tau\tau$ MC with the data is observed for $40 < m_{e+\tau_h} < 80$ GeV as shown in Figure 4.14.

	Estimate
$Z \to ee$	4,280 ± 660
$t\bar{t}$	290 ± 60
Diboson	70 ± 20
W	2,730 ± 200
QCD	3,990 ± 350
Total background	11,360 ± 780
Data	17,848
$Z \to \tau\tau$	6,610

Table 4.5: Estimates of backgrounds to $Z \to \tau\tau \to e + \tau_h$ events passing the event selections and the observed yield in data. The $Z \to \tau\tau$ prediction from MC is normalised with the NNLO cross section after applying corrections for the trigger and electron selection efficiencies.



(a) (b)

Figure 4.13: $m_{e+\tau_h}$ distribution after all event selections in linear (a) and logarithmic (b) scale. The $Z \to \tau\tau$ MC is normalised with the NNLO cross section after applying corrections for the trigger and electron selection efficiencies and the background contributions are given by the estimates detailed in Section 4.4. Uncertainties on the background contributions are given in Table 4.5.

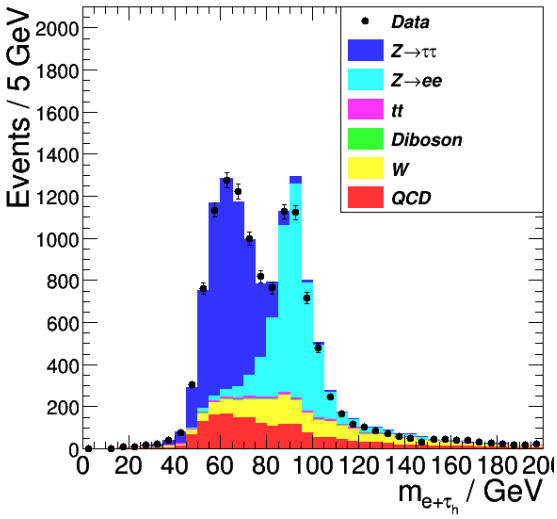


Figure 4.14: $m_{e+\tau_h}$ distribution after all event selections for events with candidate τ_h $p_T >$ 25 GeV. The $Z \to \tau\tau$ MC is normalised with the NNLO cross section after applying corrections for the trigger and electron selection efficiencies and the background contributions are estimated according to the method described in Section 4.4.

4.5 Signal extraction

The $Z \to \tau\tau$ yield in data is extracted by fitting to the $m_{e+\tau_h}$ distribution with background contributions constrained by the estimates given in Section 4.4. A binned maximum likelihood fit is performed in the range $0 < m_{e+\tau_h} < 200$ GeV in bins of 10 GeV. The QCD background shape is determined as described in Section 4.4 and the shapes of the signal and other backgrounds are taken from MC. The shapes of the W and diboson backgrounds may not be sufficiently distinct to be differentiated in a fit so these backgrounds are combined in a single Electroweak (EWK) contribution with a shape given by the addition of the separate background shapes with the relative normalisation of each shape weighted according to its estimated contribution. The EWK background is dominated by W events with an approximate 3% contribution from diboson events. Uncertainties on the shapes are discussed in Section 4.5.1.

The background normalisations are parameterised as functions of a set of nuisance parameters, θ, which represent the uncertainties discussed in Section 4.4.[2] Each nuisance

[2] A Nuisance parameter is a parameter which is not the subject of measurement but which has a finite uncertainty that must be accounted for in the measurement of the parameter of interest. In this case the $Z \to \tau\tau$ yield is the parameter of interest. The integrated luminosity, signal and background cross sections, selection efficiencies and yields in background control regions are nuisance parameters.

Nuisance parameter	$Z \to ee$	$t\bar{t}$	EWK	QCD
Luminosity	4.5%	4.5%	-	-
$Z \to ee$ cross section	3%	-	-	-
$t\bar{t}$ cross section	-	8%	-	-
Diboson cross sections	-	-	1%	-
Trigger efficiency	1%	1%	-	-
Electron selection efficiency	1%	1%	-	-
τ_h ID efficiency	-	6%	-	-
e-τ_h fake rate	6%	-	-	-
R_W	-	-	6%	-
$R_{OS/SS}$	-	-	-	5%
W estimate stat. uncertainty	-	-	1%	-
QCD estimate stat. uncertainty	-	-	-	2%

Table 4.6: Nuisance parameters in the $m_{e+\tau_h}$ maximum likelihood fit of which each background contribution is parameterised as a function. The numbers in the table constitute the uncertainties on the background estimates and correspond to the width parameters of the *pdfs* describing the constraints on the nuisance parameters. An entry '-' denotes no dependence of a background on the corresponding nuisance parameter. The EWK background is dominated by W events with an approximate 3% contribution from diboson events such that luminosity and efficiency uncertainties on the diboson contribution are neglected.

parameter is constrained by a probability density function (*pdf*) with a width parameter corresponding to the relative size of the uncertainty. Table 4.6 summarises the dependence of the background contributions on the nuisance parameters. The $Z \to \tau\tau$ yield, N, is unconstrained.

The likelihood function to be maximised with respect to N and θ is given by

$$\mathcal{L}(d|N,\theta) = \prod_n p_n(d_n|N_n,\theta) \cdot \prod_i \rho_i(\theta_i|\bar{\theta}_i,\delta\bar{\theta}_i) \cdot \prod_j \rho_j(\theta_j|\bar{\theta}_j,\delta\bar{\theta}_j) \tag{4.8}$$

where d denotes the observed data. $p_n(d_n|N_n,\theta)$ is the Poisson probability to observe d_n data in the nth bin of the $m_{e+\tau_h}$ histogram given a $Z \to \tau\tau$ yield N_n and a background

estimate $b_n(\theta)$ [98]:

$$p_n = \frac{(N_n + b_n(\theta))^{d_n} \exp\left(-N_n - b_n(\theta)\right)}{d_n!}. \tag{4.9}$$

The nuisance parameters are implemented as scale factors with expected values $\bar{\theta}$ equal to unity, uncertainties $\delta\bar{\theta}$ corresponding to the numbers in Table 4.6 and constraining *pdfs* $\rho(\theta|\bar{\theta}, \delta\bar{\theta})$. $b(\bar{\theta})$ is equal to the total background estimate given in Table 4.5. The i nuisance parameters representing the statistical uncertainties on the data yields in the background control regions are constrained by gamma distributions with shape and scale parameters k_i and s_i respectively [99]:

$$\rho_i(\theta_i|\bar{\theta}_i, \delta\bar{\theta}_i) = \frac{\theta_i^{k_i-1}\exp\left(-\theta_i/s_i\right)}{s_i^{k_i}(k_i-1)!}, \tag{4.10}$$

$$k_i = \frac{\bar{\theta}_i^2}{(\delta\bar{\theta}_i^2)}, \tag{4.11}$$

$$s_i = \frac{(\delta\bar{\theta}_i^2)}{\bar{\theta}_i}. \tag{4.12}$$

The j nuisance parameters representing luminosity, cross section and selection efficiency uncertainties are constrained by log-normal distributions with means and standard deviations $\bar{\theta}_j$ and σ_j respectively [99]:

$$\rho_j(\theta_j|\bar{\theta}_j, \delta\bar{\theta}_j) = \frac{\exp\left(-(\ln\theta_j - \bar{\theta}_j)^2/(2\sigma_j^2)\right)}{\theta_j\sigma_j\sqrt{2\pi}}, \tag{4.13}$$

$$\sigma_j = \ln\left(1 + \frac{\delta\bar{\theta}_j}{\bar{\theta}_j}\right). \tag{4.14}$$

The fitted signal and background $m_{e+\tau_h}$ distributions are shown in Figure 4.15. The fitted $Z \to \tau\tau$ yield is $N = 7140 \pm 170$ events. Good agreement between the fit and the data is seen in all bins except the $40 < m_{e+\tau_h} < 50$ and $80 < m_{e+\tau_h} < 90$ GeV bins. These discrepancies result from imperfect modelling of the signal and background $m_{e+\tau_h}$ shapes. Uncertainties on these shapes are discussed in Section 4.5.1.

4.5.1 Shape uncertainties

In addition to the nominal signal shape, four additional shapes are derived by shifting up and down the energies of electron and τ_h candidates by their respective energy scale

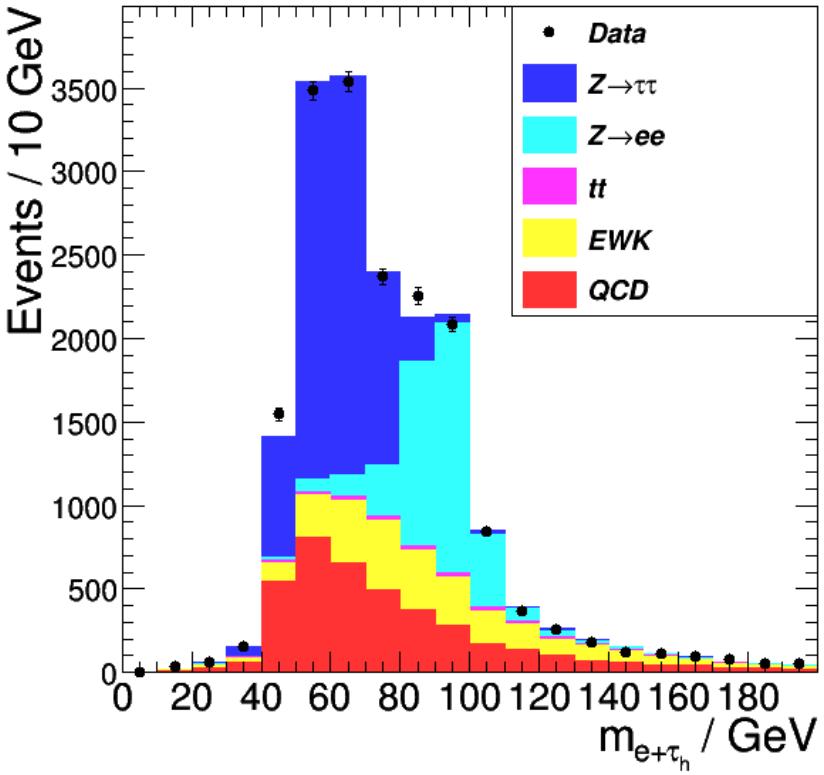


Figure 4.15: $m_{e+\tau_h}$ distribution with signal and background normalisations given by the results of the maximum likelihood fit.

uncertainties and repeating the event selection with the shifted energies. The electron and τ_h energy scale uncertainties are discussed in Sections 3.4.3 and 3.6.4 respectively. The effect on the signal shape due to these energy scale shifts is shown in Figure 4.16. Uncertainties on the shapes of the $Z \to ee$, $t\bar{t}$ and EWK backgrounds are determined in the same way and are also shown in Figure 4.16. The shape of the QCD background is estimated from data as described in Section 4.4 so that the energy scale uncertainties do not apply.

The $m_{e+\tau_h}$ fit is repeated for each energy scale shift. The deviations in the $Z \to \tau\tau$ yield with respect to the yield extracted from the fit with the nominal shapes are taken as the shape uncertainties on N. The shape uncertainties due to each energy scale shift are added in quadrature with the uncertainty on N extracted from the fit with the nominal shapes to give the total uncertainty on the $Z \to \tau\tau$ yield. The $Z \to \tau\tau$ yield after accounting for shape uncertainties is $N = 7140^{+310}_{-280}$ events.

4.6 Results

The $Z \to \tau\tau$ cross section is given by

$$\sigma(pp \to Z + X) \cdot \mathcal{B}(Z \to \tau\tau) = \frac{N}{\mathcal{L}_{int} \cdot \mathcal{B}_{e+\tau_h} \cdot \mathcal{A} \cdot \epsilon} \tag{4.15}$$

where N is the $Z \to \tau\tau$ yield determined by the $m_{e+\tau_h}$ fit. $\mathcal{B}_{e+\tau_h}$ is the branching ratio to the $e + \tau_h$ final state, $\mathcal{B}_{e+\tau_h} = 2 \cdot \mathcal{B}(\tau \to e\nu_e\nu_\tau) \cdot \mathcal{B}(\tau \to \tau_h\nu_\tau)$. ϵ is the efficiency for signal events to pass the online and offline event selections and $\mathcal{A}$ is the kinematic acceptance of $Z \to \tau\tau \to e + \tau_h$ events. The inputs to the cross section calculation are summarised in Table 4.7.

$\mathcal{A}$ and ϵ are evaluated using the $Z \to \tau\tau$ MC, considering only the generated $Z \to \tau\tau \to e + \tau_h$ events. $\mathcal{A} \cdot \epsilon$ is given by the ratio of MC events that pass all the event selections, corrected for the trigger and electron selection efficiencies as described in Section 4.3.3, to the total number of generated $Z \to \tau\tau \to e + \tau_h$ events. The result is quoted for events with $60 < m_{\tau\tau} < 120$ GeV, where $m_{\tau\tau}$ is the true generated $\tau\tau$ mass, so as to give the cross section for events on the Z resonance only. 98% of events in the $Z \to \tau\tau$ MC that pass the full event selection have $60 < m_{\tau\tau} < 120$ GeV and $\mathcal{A}$ is corrected for this.

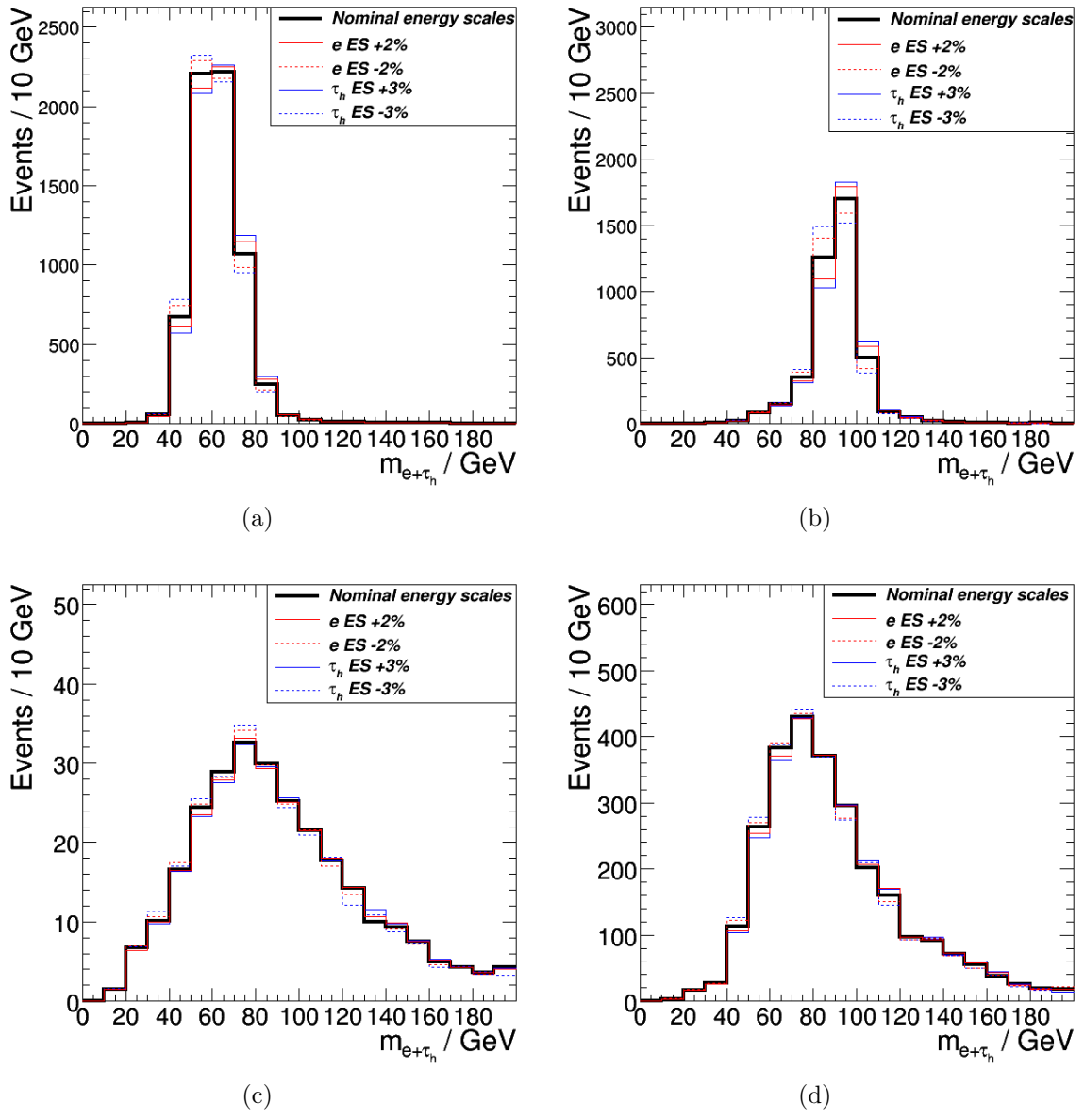


Figure 4.16: Effect of energy scale uncertainties on the $Z \to \tau\tau$ (a), $Z \to ee$ (b), $t\bar{t}$ (c) and EWK (d) shapes used in the $m_{e+\tau_h}$ fit. The black lines are the nominal shapes while the red and blue lines are the shapes derived with the electron and τ_h energy scales shifted by their respective 2 and 3% uncertainties.

Input	Value
N	7140^{+310}_{-280}
$\mathcal{L}_{int}$	2.1 fb^{-1}
$\mathcal{B}_{e+\tau_h}$	0.231
$\mathcal{A} \cdot \epsilon$	0.014

Table 4.7: Inputs to the $Z \to \tau\tau$ cross section calculation. $\mathcal{B}_{e+\tau_h}$ is the branching ratio to the $e + \tau_h$ final state, $\mathcal{B}_{e+\tau_h} = 2 \cdot \mathcal{B}(\tau \to e\nu_e\nu_\tau) \cdot \mathcal{B}(\tau \to \tau_h\nu_\tau)$, and is taken from Reference [1].

Source	Relative uncertainty
Integrated luminosity	4.5%
Acceptance due to e energy scale	1%
Acceptance due to τ_h energy scale	3%
Trigger efficiency	1%
Electron ID and isolation efficiency	1%
τ_h ID efficiency	6%

Table 4.8: Systematic uncertainties on the $Z \to \tau\tau$ cross section.

The systematic uncertainties on the $Z \to \tau\tau$ cross section are given in Table 4.8. $\mathcal{L}_{int}$ and its uncertainty are measured as described in Section 2.3 [60]. The uncertainty on $\mathcal{B}_{e+\tau_h}$ is smaller than 1% and is neglected [1]. The uncertainty on $\mathcal{A}$ is estimated by shifting the MC electron and τ_h energies by their respective energy scale uncertainties as given in Sections 3.4.3 and 3.6.4 and repeating the event selection. For each energy scale shift the difference between the resultant event yield and the nominal yield is taken as the systematic uncertainty due to that energy scale shift. The systematic uncertainties on ϵ arise from the uncertainties on the trigger, electron and τ_h selection efficiencies which are determined by the tag and probe studies described in Section 4.3.

The measured cross section, for $60 < m_{\tau\tau} < 120$ GeV,

$$\sigma(pp \to Z + X) \cdot \mathcal{B}(Z \to \tau\tau) = 1.06^{+0.05}_{-0.04}(\text{stat.}) \pm 0.07(\text{syst.}) \pm 0.05(\text{lumi.})\ \text{nb} \quad (4.16)$$

is in good agreement with the SM prediction of 0.972 $\pm$ 0.042 nb at NNLO [54][35]. The result is also in good agreement with the CMS measurements of the $Z \to ee$ and $Z \to \mu\mu$ cross sections given in Equations 3.7 and 3.8. The agreement of the result with

the SM provides an important validation for the methods used in the Higgs boson search presented in Chapter 5, most of which are shared with the analysis presented in this Chapter.

Chapter 5

Search for supersymmetric neutral Higgs bosons

$\Phi \to \tau\tau$ searches allow constraints to be placed on the MSSM parameter space as discussed in Section 1.5. The $\Phi \to \tau\tau$ search presented in this Chapter is performed in the $e + \tau_h$ final state using the CMS Run 2011A dataset corresponding to $\mathcal{L}_{int} = 2.1 \pm 0.1$ fb^{-1} at $\sqrt{s} = 7$ TeV. The search forms part of the developing CMS $\Phi \to \tau\tau$ analyses [95][52][100][93]. The search is performed separately in the inclusive and b-tagged channels, discussed in Section 1.5.2, where the inclusive channel is expected to be sensitive to both $gg \to \Phi \to \tau\tau$ and $gg \to bb\Phi \to bb\tau\tau$ production while the b-tagged channel suppresses $gg \to \Phi \to \tau\tau$ and background processes to achieve increased sensitivity to $gg \to bb\Phi \to bb\tau\tau$ events. Both searches employ the methods validated by the $Z \to \tau\tau$ cross section measurement described in Chapter 4 which provides an important cross check of the search results. In the absence of evidence for a $\Phi \to \tau\tau$ signal, upper limits on the $\Phi \to \tau\tau$ cross section are evaluated. These limits are then interpreted as constraints on tan β in the m_h^{max} scenario of the MSSM.

The MC samples used to study the signal and background processes are described in Section 5.1. Details of event selections are given in Section 5.2. $Z \to \tau\tau$ events are the dominant and irreducible background to the $\Phi \to \tau\tau$ signal and the understanding of this background is demonstrated by the measurement of the $Z \to \tau\tau$ cross section. $t\bar{t}$ production also constitutes a large background in the b-tagged channel in which dedicated cuts are applied to suppress it. The adaptation of the background estimation methods of the $Z \to \tau\tau$ analysis is described in Section 5.3. The estimation of the signal acceptance and efficiency is described in Section 5.4. Upper limits on the neutral Higgs boson

production cross section are derived from fits to the $m_{e+\tau_h}$ distributions of events in the inclusive and b-tagged channels as described in Section 5.5.

5.1 Monte Carlo samples

$gg \to \Phi \to \tau\tau$ and $gg \to bb\Phi \to bb\tau\tau$ events are generated with PYTHIA in order to optimise the event selections and test for the presence of a signal in the data [90]. Samples of 220,000 events each are generated for m_A = 100, 120, 140, 160, 180, 200, 250, 300 and 400 GeV and the τ leptons are decayed with TAUOLA [91]. The simulated signal events are normalised to the integrated luminosity of the data using the cross sections given in Section 1.5.1.

The $Z \to \tau\tau$ and $Z \to ee$ background contributions in the b-tagged channel depend on the production of jets in association with the Z boson. A sample of Z+jets events generated with MADGRAPH is therefore used to study these backgrounds instead of the POWHEG samples described in Section 4.1 [89]. The associated jet production in the Z+jets sample is described with LO accuracy while the POWHEG generation describes only the production of one additional jet at LO with further jets generated in the collinear approximation [88]. The Z+jets sample consists of 34,500,750 events in the $ll + X$ final state with $m_{ll} > 50$ GeV. τ decays are simulated by TAUOLA. The sample is divided into $Z \to \tau\tau$ and $Z \to ee$ events at the generator level. The Z+jets sample is also used to study the $Z \to \tau\tau$ and $Z \to ee$ backgrounds in the inclusive channel, allowing for a direct comparison of signal sensitivity between the two channels. The samples of W+jets, $t\bar{t}$+jets and diboson events described in Section 4.1 are also used in the $\Phi \to \tau\tau$ analysis.

5.2 Event selection

The event selections are the result of ongoing studies at CMS and are optimised to give the most stringent expected upper limits on the $\Phi \to \tau\tau$ cross section [50][95][52][100][93]. Limits derived from observation of the data are not considered for optimisation in order to minimise bias in the analysis methodology.

The online and offline selections in the inclusive channel are identical to those used in the $Z \to \tau\tau$ cross section measurement and are given in Section 4.2. Additional requirements on the presence of jets are made in the b-tagged channel. Jet candidates

are reconstructed as described in Section 3.3. Only jet candidates with $p_T > 20$ GeV and $|\eta| < 4.5$ are considered where the pseudorapidity cut is applied to ensure the jets are contained within the fiducial volume of the detector. Jet candidates are required to be separated from the electron and τ_h candidates by $\Delta R > 0.5$ in order to ensure they do not correspond to the same physical objects. The multiplicity, p_T and η distributions of jet candidates satisfying these criteria, after applying the inclusive selections, are shown in Figures 5.1, 5.2 and 5.3 respectively. Reasonable agreement is seen between the distributions in data and those given by the background estimates detailed in Section 5.3. The background estimates in the $20 < p_T < 25$ GeV bin and the $2.2 < |\eta| < 3.8$ region are in excess of the observed data. These discrepancies are caused by imperfections in the residual corrections to the relative JES described in Section 3.3.1.

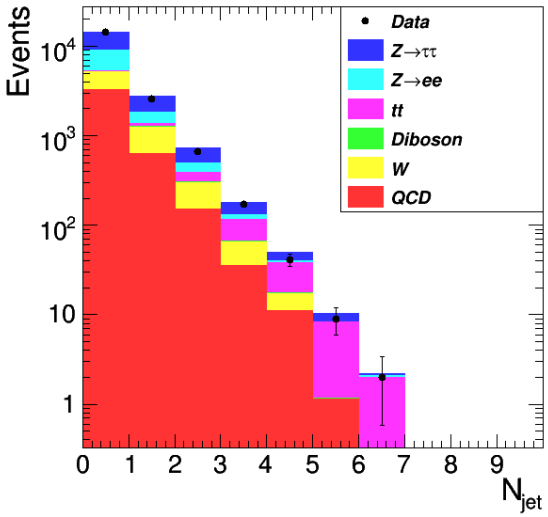
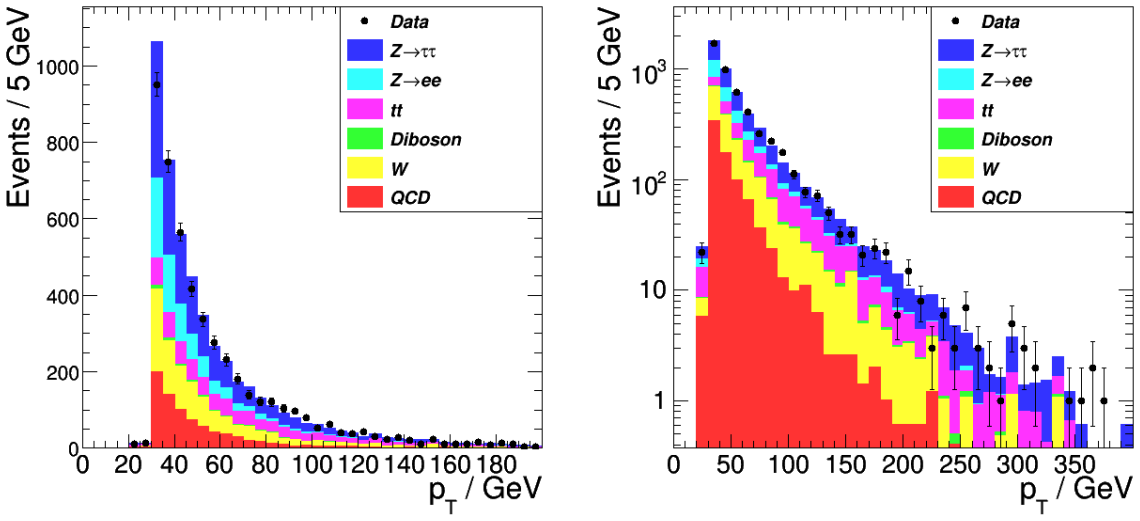


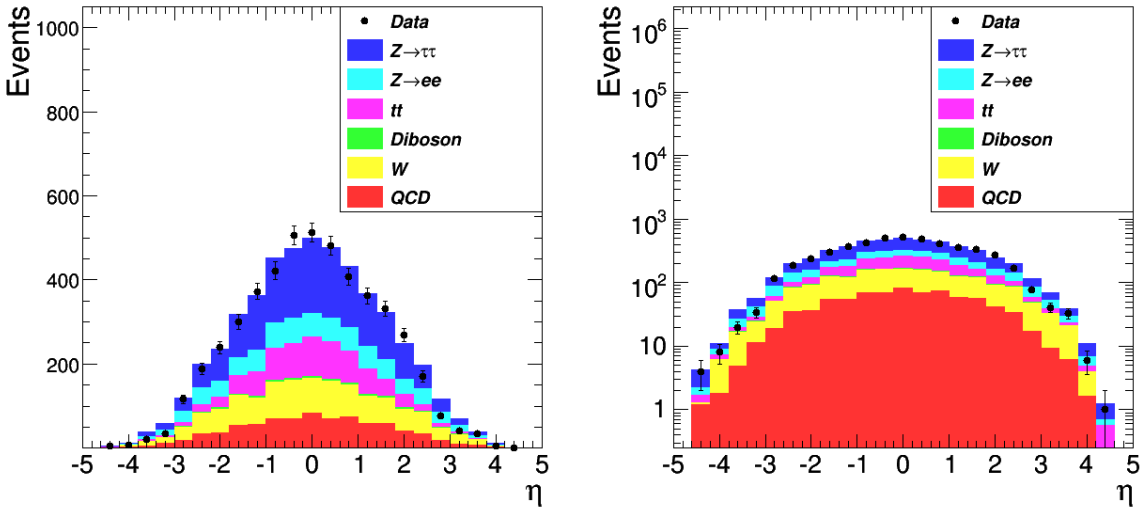
Figure 5.1: Multiplicity, N_{jet}, distribution of reconstructed jets in events passing the inclusive selections. The background contributions correspond to the estimates given in Section 5.3.

b-jets are identified as jet candidates tagged by the SSVHP algorithm, described in Section 3.7. b-jet candidates are required to satisfy $d_{SSVHP} > 2$ and $|\eta| < 2.4$ to ensure they are within the acceptance of the Silicon Strip Tracker. The d_{SSVHP} distribution of all tagged jet candidates with $|\eta| < 2.4$ after applying the inclusive selections is shown in Figure 5.4. The multiplicity distribution of b-jet candidates is shown in Figure 5.5. Good agreement between the data and background estimates is observed in both distributions. Events in the b-tagged channel are required to contain exactly one b-jet candidate. An inclusive selection on the presence of b-jets is disfavoured as events containing more than one b-jet candidate are expected to be dominated by $t\bar{t}$ events as shown in Figure 5.5. To further suppress the $t\bar{t}$ background, events containing additional jet candidates with



(a) (b)

Figure 5.2: p_T distribution of reconstructed jets in events passing the inclusive selections in linear (a) and logarithmic (b) scale. Jet candidates with $20 < p_T < 30$ GeV are required to be tagged as b-jets by the SSVHP algorithm and are used to isolate signal events in the b-tagged channel. Jets with $p_T > 30$ GeV are used to discriminate against the $t\bar{t}$ background. The background contributions correspond to the estimates given in Section 5.3.



(a) (b)

Figure 5.3: η distribution of reconstructed jets in events passing the inclusive selections in linear (a) and logarithmic (b) scale. The background contributions correspond to the estimates given in Section 5.3.

$p_T > 30$ GeV are rejected. This requirement does not excessively degrade the efficiency for $gg \to bb\Phi \to bb\tau\tau$ events as the b-jets in these events are relatively soft. The event yield in data in the b-tagged channel is 72 events.

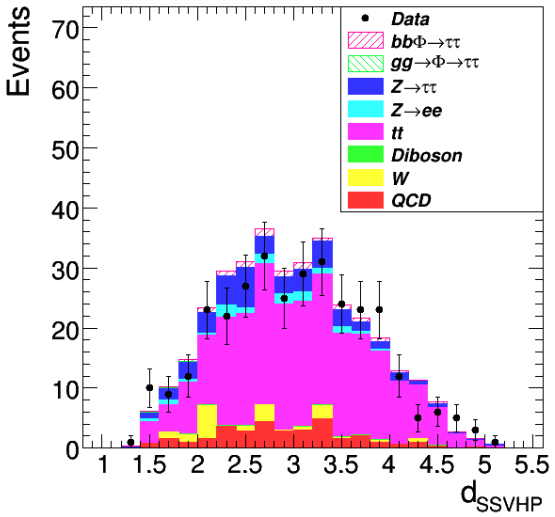


Figure 5.4: d_{SSVHP} distribution of tagged jet candidates with $|\eta| < 2.4$ in events passing the inclusive selections. The background contributions correspond to the estimates given in Section 5.3. The signal contributions expected from MC for $m_A = 200$ GeV and $\tan\beta = 15$ are also shown. b-jet candidates are required to have $d_{SSVHP} > 2$.

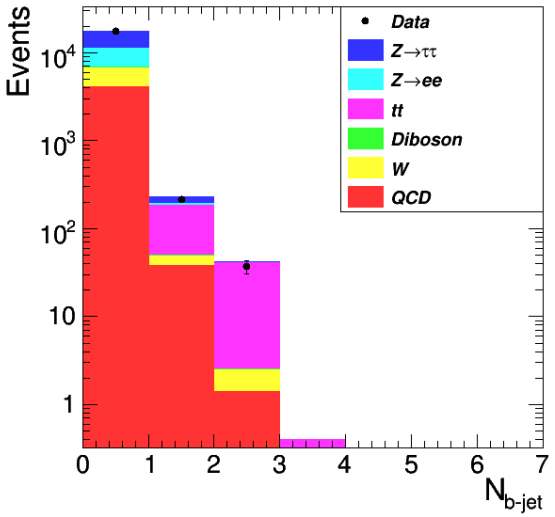


Figure 5.5: Multiplicity, $N_{b\text{-}jet}$, distribution of b-jet candidates in events passing the inclusive selections. The background contributions correspond to the estimates given in Section 5.3. Events in the b-tagged channel are required to contain exactly one b-jet candidate.

5.3 Background estimation

As the $\Phi \to \tau\tau$ signal overlaps the $m_{e+\tau_h}$ region occupied by the backgrounds for most of the m_A range considered, the size and shape of the backgrounds must be accurately estimated in order to maximise sensitivity to signal events. The background estimation follows the method described in Section 4.4 with the addition of estimates for the irreducible $Z \to \tau\tau$ background. Given the agreement of CMS $Z \to ee$ and $Z \to \mu\mu$ observations with MC predictions together with the validation of the τ_h ID provided by the $Z \to \tau\tau$ cross section measurement, MC is relied upon to estimate the $Z \to \tau\tau$ background [78]. The validity of this approach in the b-tagged channel is supported by CMS measurements of the cross section for Z boson production in association with b quarks which are also in agreement with MC predictions [83].

The Z+jets MC is normalised to the integrated luminosity of the data using the NNLO cross section times $Z \to ll$ branching ratio of 3.048 nb calculated with FEWZ for $m_{ll} >$ 50 GeV [54]. This prediction is in excellent agreement with CMS measurement and the 3% uncertainty on the CMS measurement is taken as the systematic uncertainty on the normalisation together with the 4.5% uncertainty on $\mathcal{L}_{int}$ [78][60]. The rate of MC $Z \to \tau\tau$ events passing the inclusive selections is corrected for the trigger and electron selection efficiencies as described in Section 4.3. The estimate of the $Z \to \tau\tau$ background in the inclusive channel is 6,480 $\pm$ 1,010 events where the uncertainty includes the uncertainties on the MC normalisation as well as the trigger (1%), electron selection (1%) and τ_h ID (6%) efficiencies. The estimate is in good agreement with the $Z \to \tau\tau$ yield extracted from the $m_{e+\tau_h}$ fit given in Section 4.5.

The Z+jets MC events passing the b-tagged channel selections are dominated by $b\bar{b}Z$ events. The $Z \to \tau\tau$ event yield in the b-tagged channel is therefore further corrected for the b-tagging efficiency using the correction factors given in Table 3.2. The uncertainty on the event yield due to the JES uncertainties discussed in Section 3.3.1 is estimated by shifting the energies of all reconstructed jets by 5% and repeating the event selection. The change in the $Z \to \tau\tau$ yield under this energy scale shift is taken as the JES uncertainty on the yield and amounts to 3.5%. The estimate of the $Z \to \tau\tau$ background in the b-tagged channel is 22 $\pm$ 6 events where the uncertainty includes the uncertainties considered in the inclusive estimate as well as the JES and b-tagging efficiency (10%) uncertainties [81].

The $Z \to ee$, $t\bar{t}$, W and diboson backgrounds in the inclusive channel are estimated by the methods given in Sections 4.4.1 and 4.4.2 with the difference that the $Z \to ee$ events in the Z+jets sample are used instead of the POWHEG $Z \to ee$ sample. In the b-tagged channel the $Z \to ee$ and $t\bar{t}$ MC event yields are further corrected for the b-tagging efficiency and the estimates of their contributions are 7 ± 2 and 17 ± 8 events respectively. The uncertainties include the uncertainties on the b-tagging efficiency and the JES uncertainties which are evaluated by the same method as for the $Z \to \tau\tau$ background. The estimate of the diboson background in the b-tagged channel is $0.1^{+0.2}_{-0.1}$ events. The diboson background is neglected in the limit extraction procedure described in Section 5.5.

The ratios R_W, in the W+jets MC, and $R_{OS/SS}$, in the data with inverted electron isolation, do not change significantly after applying the b-tagged channel selections. The W and QCD backgrounds in both the inclusive and b-tagged channels are estimated using the methods described in Sections 4.4.2 and 4.4.3 with the difference that the W and SS control regions in the b-tagged channel are defined after applying the b-tagged channel selections. The estimate of the W background in the b-tagged channel is 18 ± 4 events where the uncertainty includes the 15% statistical uncertainty on the data yield in the W control region. Using the estimate of the $Z \to \tau\tau$ background, the estimate of the QCD background in both the inclusive and b-tagged channels is further corrected for the contamination of the SS control region by $Z \to \tau\tau$ events. The estimate of the QCD background in the b-tagged channel is 13 ± 5 events where the uncertainty includes the 25% statistical uncertainty on the data yield in the SS control region.

The estimates of the backgrounds in the inclusive and b-tagged channels are summarised in Table 5.1. The estimates are in good agreement with the observed event yields in data. The estimates of the total background contributions in the inclusive and b-tagged channels are $17{,}780 \pm 1{,}250$ and 77 ± 12 respectively.

5.3.1 Estimation of background shapes

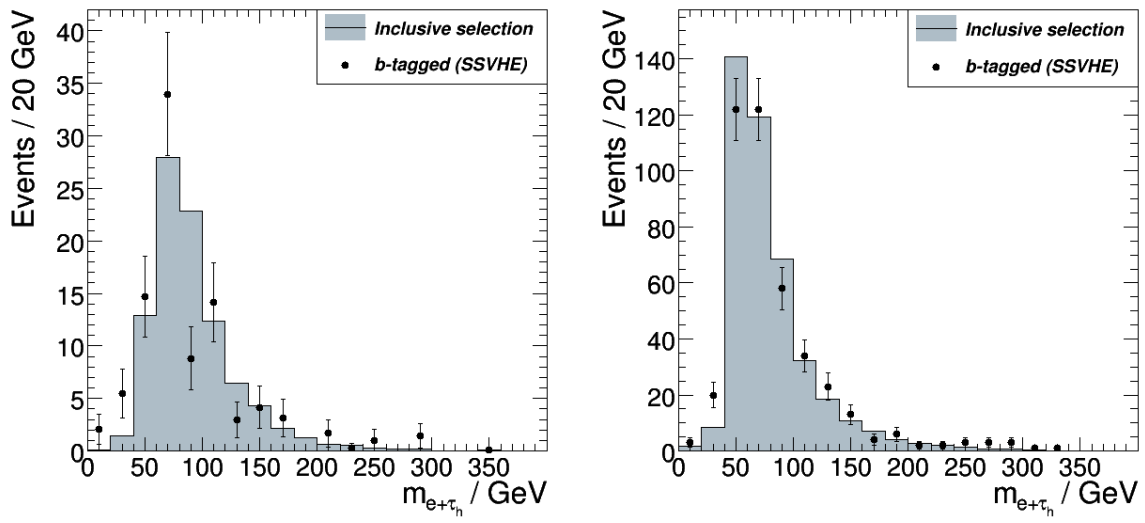
The estimates of the background shapes in the inclusive channel are the same as those used in the $Z \to \tau\tau$ cross section measurement except that the $Z \to \tau\tau$ and $Z \to ee$ shapes are taken from the Z+jets sample instead of the POWHEG samples. The shapes of the $Z \to \tau\tau$, $Z \to ee$ and $t\bar{t}$ backgrounds in the b-tagged channel are also taken from the MC. The uncertainties on these shape estimates are discussed in Section 5.5.1.

	Inclusive estimate	b-tagged estimate
$Z \to \tau\tau$	6,480 ± 1,010	22 ± 6
$Z \to ee$	4,320 ± 670	7 ± 2
$t\bar{t}$	290 ± 60	17 ± 8
Diboson	70 ± 20	$0.1^{+0.2}_{-0.1}$
W	2,740 ± 200	18 ± 4
QCD	3,880 ± 250	13 ± 5
Total background	17,780 ± 1,250	77 ± 12
Data	17,848	72
$gg \to \Phi \to \tau\tau$	70	0
$gg \to bb\Phi \to bb\tau\tau$	70	6

Table 5.1: Estimates of backgrounds to $\Phi \to \tau\tau \to e + \tau_h$ events in the inclusive and b-tagged channels and the observed yields in data. The signal predictions from MC are for $m_A = 200$ GeV and $\tan\beta = 15$ and are normalised with the cross sections given in Section 1.5.1 after applying corrections for trigger, electron and b-jet selection efficiencies.

The event yields of W+jets MC and SS data in the b-tagged channel are small and cannot provide precise estimates of the QCD and W background shapes. The more efficient SSVHE b-tagging algorithm, described in Section 3.7, is used to study the shapes of these backgrounds in the b-tagged channel. The b-tagged channel selections, with the SSVHP b-tag replaced with the SSVHE tag, are applied to W+jets MC and SS data events and the $m_{e+\tau_h}$ distributions of the events passing the selections are compared with the inclusive shapes. These distributions are shown in Figure 5.6 where the electron candidates in the SS data events are loosely isolated as in the estimate of the QCD background shape in the inclusive channel. In both the W+jets MC and SS data events, the inclusive $m_{e+\tau_h}$ shapes are consistent with the shapes obtained in the b-tagged channel with the SSVHE b-tag, within statistical uncertainties. This demonstrates that the b-tagged channel selections do not significantly change the shapes of the W and QCD backgrounds with respect to the inclusive channel. The estimates of the W and QCD background shapes in the inclusive channel are therefore also taken as the estimates of the corresponding shapes in the b-tagged channel. These shapes are normalised by the same method as used in the inclusive channel as described in Section 5.3.

Figures 5.7 and 5.8 show the $m_{e+\tau_h}$ distributions after the inclusive and b-tagged channel selections respectively with background shapes and sizes corresponding to the estimates



(a) (b)

Figure 5.6: $m_{e+\tau_h}$ distributions of W+jets MC events after the inclusive channel selections and after the b-tagged channel selections with the SSVHP b-tag replaced with the SSVHE tag (a), and the corresponding distributions for SS $e + \tau_h$ events in data with $I_e < 0.3$ (b). The distributions of the b-tagged events are normalised to the event yields after the inclusive selections only to allow for comparison of the shapes. The distributions after the inclusive selections only are estimates of the $m_{e+\tau_h}$ shapes of the QCD and W backgrounds in the b-tagged channel.

given in this Section. The distributions of the data and the background estimates agree within the uncertainties on the background estimates.

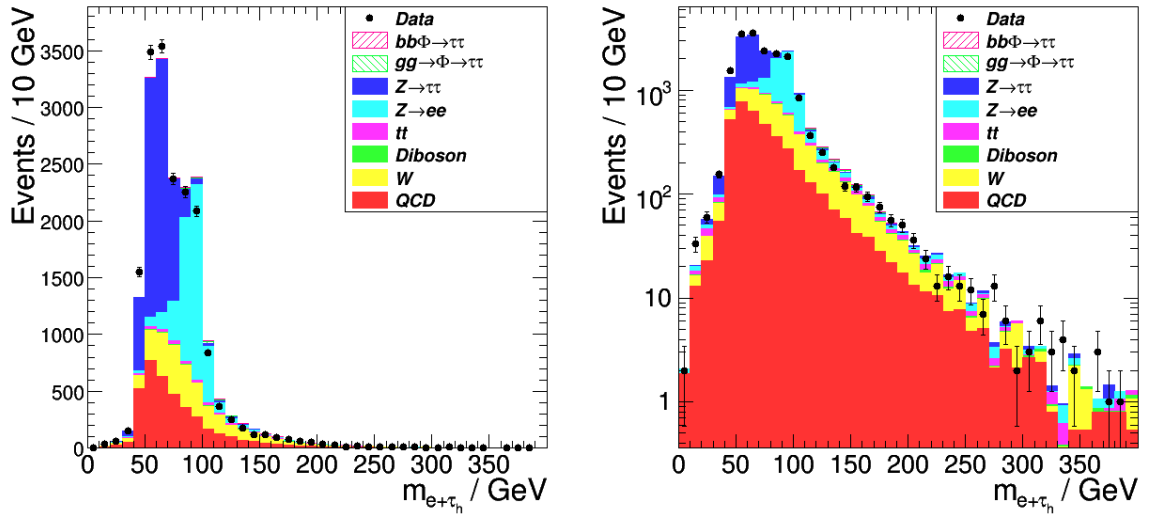


Figure 5.7: $m_{e+\tau_h}$ distribution after the inclusive event selections in linear (a) and logarithmic (b) scale. The background contributions are given by the estimates detailed in Section 5.3 and summarised in Table 5.1. The signal predictions from MC for $m_A = 200$ GeV and $\tan\beta = 15$ are normalised with the cross sections given in Section 1.5.1 after applying corrections for trigger and electron selection efficiencies. Fits to this distribution are performed to test for the presence of a $\Phi \to \tau\tau$ signal in the data.

5.4 Signal acceptance and efficiency

The $\Phi \to \tau\tau$ acceptance and efficiency, as well as the $\Phi \to \tau\tau$ $m_{e+\tau_h}$ shapes, are estimated using the MC in order to set limits on the $\Phi \to \tau\tau$ production cross section times branching ratio, $\sigma(pp \to \Phi) \cdot \mathcal{B}(\Phi \to \tau\tau)$. The product of the acceptance and efficiency, including $\mathcal{B}_{e+\tau_h}$, is given by the ratio of MC $\Phi \to \tau\tau$ events passing the event selections to the number of generated events, after correcting for trigger, electron and b-jet selection efficiencies. The estimates are determined separately for the inclusive and b-tagged selections and the $gg \to \Phi \to \tau\tau$ and $gg \to bb\Phi \to bb\tau\tau$ processes. The estimates for each m_A point considered are shown in Table 5.2. As expected, the efficiencies for $gg \to \Phi \to \tau\tau$ and $gg \to bb\Phi \to bb\tau\tau$ events in the inclusive channel are similar while in the b-tagged channel the efficiencies for $gg \to bb\Phi \to bb\tau\tau$ events are larger than

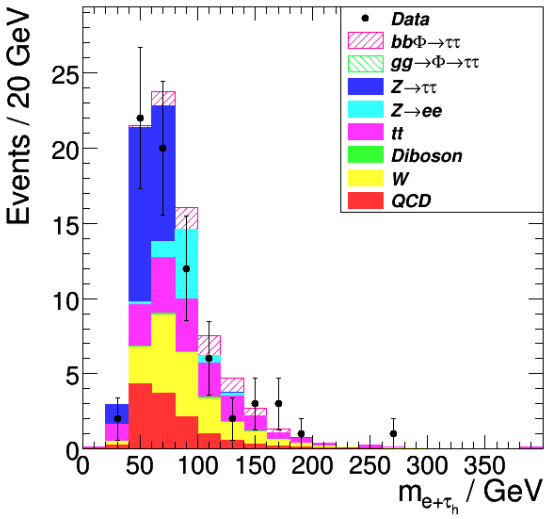


Figure 5.8: $m_{e+\tau_h}$ distribution after the b-tagged channel selections. The background contributions are given by the estimates detailed in Section 5.3 and summarised in Table 5.1. The signal predictions from MC for $m_A = 200$ GeV and $\tan\beta = 15$ are normalised with the cross sections given in Section 1.5.1 after applying corrections for trigger, electron and b-jet selection efficiencies. Fits to this distribution are performed to test for the presence of a $\Phi \to \tau\tau$ signal in the data.

those for $gg \to \Phi \to \tau\tau$ events by one to two orders of magnitude depending on m_A. The uncertainties on the efficiencies due to trigger, electron, τ_h and b-jet selections are incorporated into the limit extraction procedure described in Section 5.5.2. The uncertainties on the efficiencies due to the the JES uncertainty are also accounted for and are estimated in the same way as for the $Z \to \tau\tau$ background.

5.4.1 Effect of b-loop contributions to $gg \to \Phi$ on the $gg \to \Phi \to \tau\tau$ acceptance

The $gg \to \Phi \to \tau\tau$ MC samples are generated with the assumption of infinite top quark mass which is applicable in the SM, and the MSSM for small $\tan\beta$, where the dominant contribution to the gluon fusion process proceeds via top quark loops [39]. At large $\tan\beta$, $gg \to \Phi$ production in the MSSM proceeds predominantly via b quark loops producing a softer Higgs boson p_T spectrum than that predicted in the infinite top mass approximation [101]. As a result the kinematic acceptance of $gg \to \Phi \to \tau\tau$ events as given in Table 5.2 may be overestimated. To determine the size of this effect, the generated Pythia Higgs boson p_T spectrum is reweighted to match the shape predicted

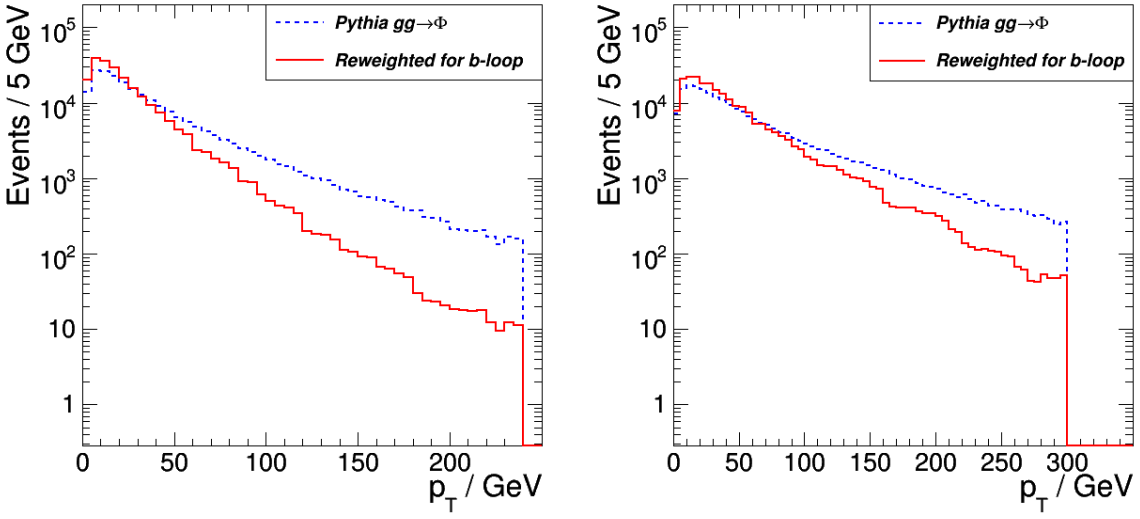
	Inclusive acceptance × efficiency		b-tagged acceptance × efficiency	
m_A / GeV	$gg \to \Phi \to \tau\tau$	$gg \to bb\Phi \to bb\tau\tau$	$gg \to \Phi \to \tau\tau$	$gg \to bb\Phi \to bb\tau\tau$
100	6.81×10^{-3}	6.90×10^{-3}	1.47×10^{-5}	4.35×10^{-4}
120	1.03×10^{-2}	1.08×10^{-2}	1.80×10^{-5}	8.64×10^{-4}
140	1.34×10^{-2}	1.50×10^{-2}	2.08×10^{-5}	1.22×10^{-3}
160	1.64×10^{-2}	1.75×10^{-2}	3.56×10^{-5}	1.25×10^{-3}
180	1.99×10^{-2}	2.10×10^{-2}	3.14×10^{-5}	1.79×10^{-3}
200	2.18×10^{-2}	2.28×10^{-2}	2.79×10^{-5}	1.88×10^{-3}
250	2.74×10^{-2}	2.84×10^{-2}	1.00×10^{-4}	2.39×10^{-3}
300	2.98×10^{-2}	3.22×10^{-2}	9.69×10^{-5}	2.74×10^{-3}
400	3.24×10^{-2}	3.63×10^{-2}	8.14×10^{-5}	3.18×10^{-3}

Table 5.2: Products of acceptance and efficiency for $gg \to \Phi \to \tau\tau$ and $gg \to bb\Phi \to bb\tau\tau$ events in the inclusive and b-tagged channels for different m_A. The products are evaluated using the MC with corrections for the trigger, electron and b-jet selection efficiencies applied and include $\mathcal{B}_{e+\tau_h}$. Uncertainties on the JES and trigger, electron selection, τ_h ID and b-tagging efficiencies are incorporated in to the limit extraction procedure described in Section 5.5.2.

when considering only the b-loop contributions to $gg \to \Phi$ as given in Reference [101]. The $gg \to \Phi \to \tau\tau$ acceptance is then estimated before and after reweighting of the events.

The reweighting procedure is performed with the $m_A = 140$ and 400 GeV $gg \to \Phi \to \tau\tau$ samples in order to test the effect at the low and high ends of the considered m_A range. The generated events are assigned weights $w_j = N_j^{b-loop}/N_j^{PYTHIA}$ where N_j^{PYTHIA} and N_j^{b-loop} are the normalised event rates predicted by Pythia and with only the b-loop contribution to $gg \to \Phi$ respectively in j bins of Higgs boson p_T. Events with Higgs p_T < 240 GeV and $p_T < 300$ GeV are considered for $m_A = 140$ and 400 GeV respectively. Figure 5.9 shows the generated Higgs boson p_T distributions for the two mass points before and after applying the event weights.

Figure 5.10 shows the generator level electron and τ_h p_T distributions for the two mass points before and after applying the event weights. As expected, the p_T spectra of the τ decay products are softer when only the b-loop contribution to the $gg \to \Phi$ process is considered, compared to the infinite top mass approximation implemented in Pythia. To estimate the acceptance, MC events containing a generator level electron and τ_h passing the kinematic requirements given in Section 4.2 and separated by $\Delta R > 0.5$ are selected.



(a) (b)

Figure 5.9: Higgs boson p_T distributions in PYTHIA $gg \to \Phi \to \tau\tau$ events before and after reweighting to account for only b-loop contributions to the $gg \to \Phi$ process, for $m_A = 140$ (a) and 400 (b) GeV.

	Acceptance	
m_A / GeV	PYTHIA, infinite top mass	Reweighted PYTHIA, b-loop only
140	0.072 ± 0.001	0.070 ± 0.001
400	0.149 ± 0.001	0.152 ± 0.001

Table 5.3: $gg \to \Phi \to \tau\tau \to e + \tau_h$ acceptance predicted by PYTHIA with the infinite top mass approximation and after reweighting to account for only b-loop contributions to the $gg \to \Phi$ process, for $m_A = 140$ and 400 GeV. The quoted uncertainties are statistical only.

The acceptance, defined as the ratio of selected to generated events, is determined before and after applying the reweighting procedure and is shown in Table 5.3.

When only the b-loop contribution is considered the acceptance is approximately 3% smaller than predicted under the infinite top mass approximation for $m_A = 140$ GeV and approximately 2% larger for $m_A = 400$ GeV. The small size of the effect results from the relatively low p_T thresholds applied to the visible τ decay products and is much smaller than the theoretical uncertainties on the $gg \to \Phi$ production cross section given in Section 1.5.1. The effect is therefore neglected in the limit extraction procedure described in Section 5.5.2.

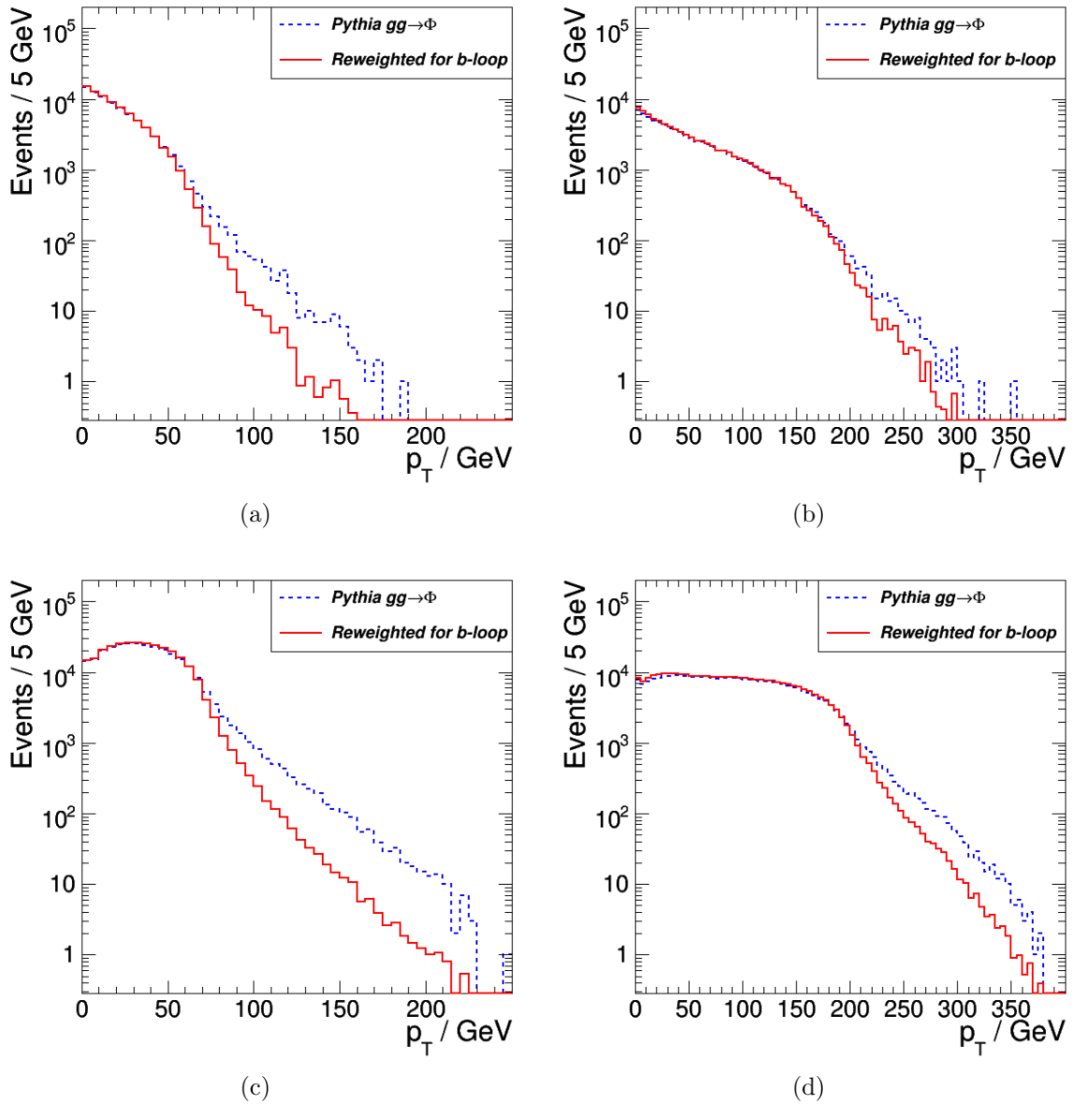


Figure 5.10: Generator level electron (a, b) and τ_h (c, d) p_T distributions in PYTHIA $gg \to \Phi \to \tau\tau$ events before and after reweighting to account for only b-loop contributions to the $gg \to \Phi$ process, for $m_A = 140$ (a, c) and 400 (b, d) GeV.

5.5 Results

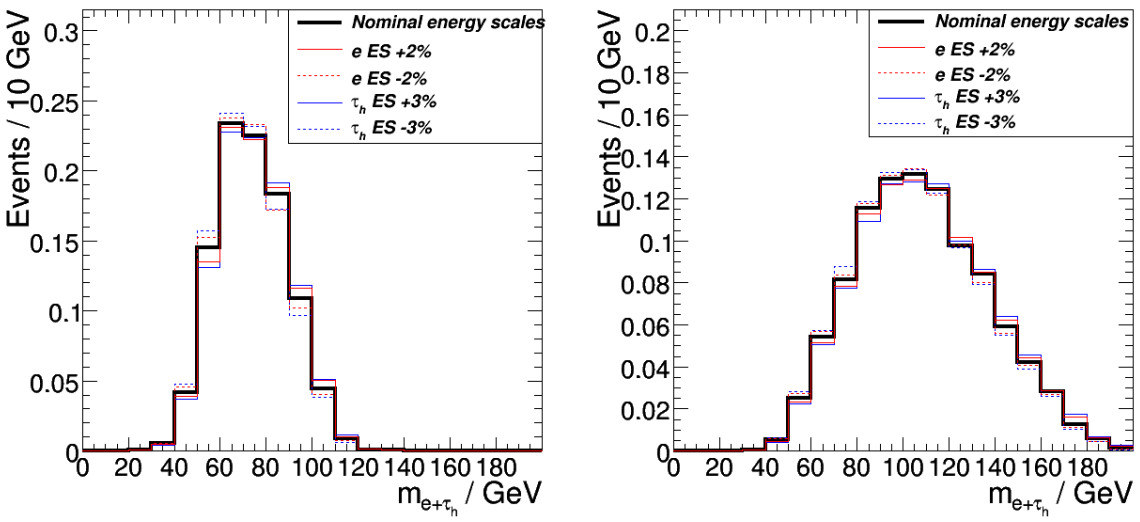
No significant excess of events in the data is observed over the predicted SM backgrounds in either the inclusive or b-tagged channels. Upper limits on $\sigma(pp \to \Phi) \cdot \mathcal{B}(\Phi \to \tau\tau)$ are set at 95% CL using the modified frequentist method $\mathrm{CL_s}$ which is chosen to avoid excluding signals to which the search has no sensitivity [102]. The limits are computed separately in each channel and for each m_A point considered and are interpreted as 95% CL upper limits on $\tan\beta$ in the m_h^{max} scenario of the MSSM.

5.5.1 Shape uncertainties

The test statistic used in the limit computation is constructed from fits to the $m_{e+\tau_h}$ distributions shown in Figures 5.7 and 5.8. Uncertainties on the estimated signal and background shapes are incorporated by a template morphing method [95]. As described in Section 4.5.1 in the case of the $Z \to \tau\tau$ cross section measurement, signal and background shapes in addition to the nominal shapes are derived by shifting up and down the electron and τ_h energy scales by their respective uncertainties. The signal shapes in the inclusive channel are derived by adding the $gg \to \Phi \to \tau\tau$ and $gg \to bb\Phi \to bb\tau\tau$ shapes from MC with each contribution weighted by its yield estimated with MC normalised using the cross sections given in Section 1.5.1. The $gg \to bb\Phi \to bb\tau\tau$ shape only is used in the b-tagged channel where the estimated yield of $gg \to \Phi \to \tau\tau$ events is negligible. In the inclusive channel, the W and diboson contributions are combined in a single EWK background as described in Section 4.5. The diboson background is neglected in the b-tagged channel.

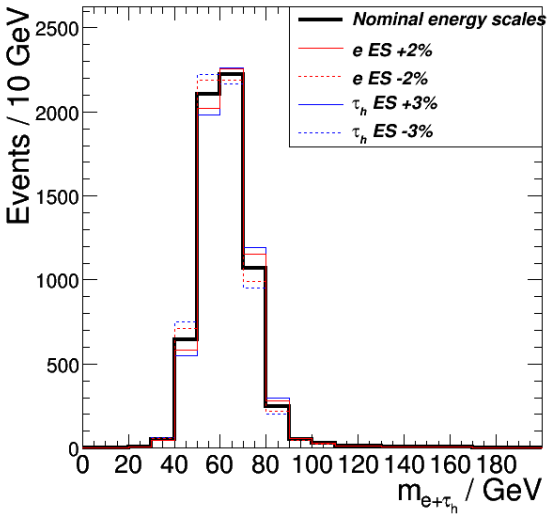
The nominal and shifted $m_{e+\tau_h}$ shapes for the signal at the $m_A = 120$ and 200 GeV points and the $Z \to \tau\tau$ and $Z \to ee$ backgrounds in the inclusive channel are shown in Figure 5.11. The $t\bar{t}$ and EWK background shapes in the inclusive channel are identical to those used in the $Z \to \tau\tau$ cross section measurement. The corresponding nominal and shifted shapes in the b-tagged channel are shown in Figures 5.12 and 5.13. In both channels the QCD background shape is estimated with data as described in Section 5.3.1.

Each shape, $S(\theta)$, to which energy scale uncertainties apply is parameterised in terms of morphing parameters θ_k which represent the variation of energy scale k (electron or τ_h)

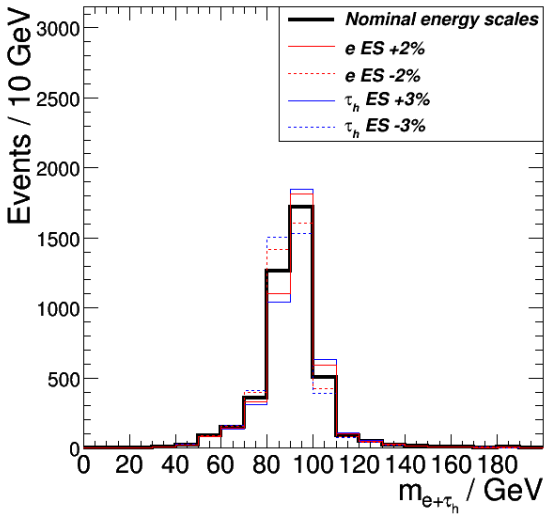


(a)

(b)

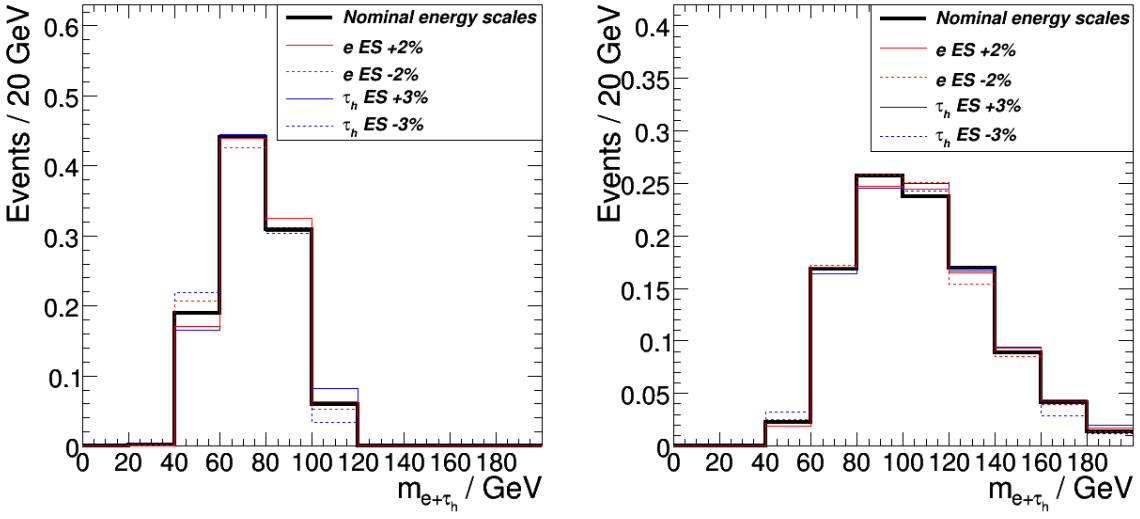


(c)



(d)

Figure 5.11: Effect of energy scale uncertainties on the $m_{e+\tau_h}$ shapes in the inclusive channel for the signal at $m_A = 120$ (a) and 200 (b) GeV and the $Z \to \tau\tau$ (c) and $Z \to ee$ (d) backgrounds. The black lines are the nominal shapes while the red and blue lines are the shapes derived with the electron and τ_h energy scales shifted by their respective 2 and 3% uncertainties. The uncertainties on the $t\bar{t}$ and EWK background shapes are shown in Figure 4.16.



(a) (b)

Figure 5.12: Effect of energy scale uncertainties on the $m_{e+\tau_h}$ shapes in the b-tagged channel for the signal at $m_A = 120$ (a) and 200 (b) GeV. The black lines are the nominal shapes while the red and blue lines are the shapes derived with the electron and τ_h energy scales shifted by their respective 2 and 3% uncertainties.

with respect to the nominal energy scale [95]. The shape is then given by

$$S(\theta) = S^0 + \sum_k \left(a(\theta_k) S_k^+ + b(\theta_k) S^0 + c(\theta_k) S_k^- \right) \tag{5.1}$$

where S^0 is the nominal shape and $S_k^\pm$ are the shapes under 1σ positive and negative shifts of energy scale k corresponding to 2% and 3% shifts of the electron and τ_h energy scales respectively. For $\theta_k = \bar{\theta}_k = 0$, $S(\theta) = S^0$ and for $\theta_k = \bar{\theta}_k \pm \delta\bar{\theta}_k = \pm 1$, $S(\theta) = S_k^\pm$. As shapes are determined only for the nominal and 1σ shifted energy scales, and the θ_k are continuous variables, interpolation between the nominal and shifted shapes is necessary to estimate $S(\theta)$ for $\theta_k \neq \bar{\theta}_k, \bar{\theta}_k \pm \delta\bar{\theta}_k$. Quadratic interpolation defined by

$$a = \theta_k \cdot (\theta_k + 1)/2 \tag{5.2}$$

$$b = \theta_k \cdot (\theta_k - 1)/2 \tag{5.3}$$

$$c = -\theta_k^2 \tag{5.4}$$

is used for for $|\theta_k| \leq 1$ and linear interpolation defined by $a = -c = \theta_k$, $b = -|\theta_k|$ for $|\theta_k| > 1$ in order to correspond approximately to the Gaussian distribution of θ_k. The *pdf* of θ_k which constrains the shape variations in the $m_{e+\tau_h}$ fits as described in Section 5.5.2

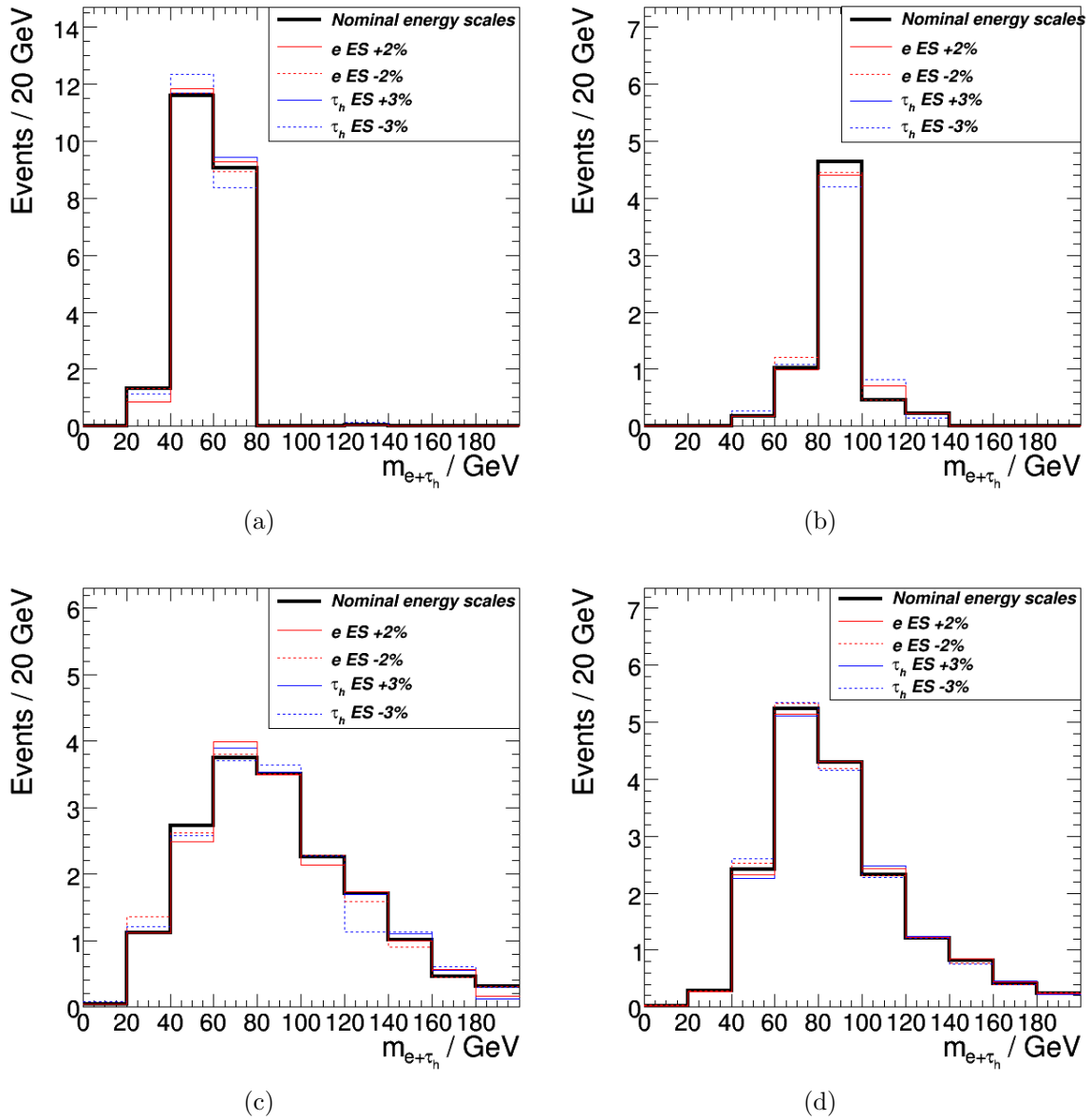


Figure 5.13: Effect of energy scale uncertainties on the $m_{e+\tau_h}$ shapes in the b-tagged channel for the $Z \to \tau\tau$ (a), $Z \to ee$ (b), $t\bar{t}$ (c) and W (d) backgrounds. The black lines are the nominal shapes while the red and blue lines are the shapes derived with the electron and τ_h energy scales shifted by their respective 2 and 3% uncertainties.

is given by $\rho_k(\theta_k|\bar{\theta}_k, \delta\bar{\theta}_k)$ [98]:

$$\rho_k(\theta_k|\bar{\theta}_k, \delta\bar{\theta}_k) = \frac{\exp\left(-\frac{(\theta_k-\bar{\theta}_k)^2}{2(\delta\bar{\theta}_k)^2}\right)}{\sqrt{2\pi(\delta\bar{\theta}_k)^2}}. \tag{5.5}$$

5.5.2 Limit extraction

For each m_A point considered, binned maximum likelihood fits to the $m_{e+\tau_h}$ distributions of events passing the selections are performed in the range $0 < m_{e+\tau_h} < 400$ GeV in 10 GeV bins in the inclusive channel and 20 GeV bins in the b-tagged channel where the event yields are smaller. The signal and background shapes and their uncertainties are given in Section 5.5.1. The signal and background normalisations, $s(\theta)$ and $b(\theta)$ respectively, are parameterised as functions of a set of nuisance parameters, θ, which represent the uncertainties on the yields. As in the fit described in Section 4.5, the nuisance parameters are implemented as scale factors with central values $\bar{\theta}$ equal to unity and uncertainties $\delta\bar{\theta}$ given in Tables 5.4 and 5.5. $b(\bar{\theta})$ is equal to the total background estimate and $s(\bar{\theta}) = 1$.

The limit extraction procedure follows the prescription of the LHC Higgs Combination Group [103]. The combined signal plus background yield is given by $\mu \cdot s(\theta) + b(\theta)$ where μ is a signal strength parameter quantifying the size of the signal contribution. The test statistic used in the $\mathrm{CL_s}$ computation is the profile likelihood ratio

$$q_\mu = -2 \cdot \ln\frac{\mathcal{L}_\mu(d|\mu, \hat{\theta}_\mu)}{\mathcal{L}(d|\hat{\mu}, \hat{\theta})}, \tag{5.6}$$

$$0 \leq \hat{\mu} \leq \mu, \tag{5.7}$$

where $\mathcal{L}_\mu(d|\mu, \hat{\theta}_\mu)$ is the likelihood to observe data d maximised with respect to θ given a signal strength μ. The set of nuisance parameters with values that maximise $\mathcal{L}_\mu$ is denoted $\hat{\theta}_\mu$. $\mathcal{L}(d|\hat{\mu}, \hat{\theta})$ is the likelihood to observe d maximised with respect to both μ and θ where $\hat{\mu}$ and $\hat{\theta}$ are the μ and θ which maximise $\mathcal{L}$. The constraints on $\hat{\mu}$ are imposed to force the signal normalisation to be positive and to give one-sided confidence intervals.

In order to describe the constraints on the nuisance parameters in a frequentist manner, the nuisance parameter *pdfs* $\rho(\theta|\bar{\theta})$ given by Equations 4.10, 4.13 and 5.5 are reinterpreted using Bayes' theorem as posterior *pdfs* arising from hypothetical measurements of $\bar{\theta}$ [104].

Nuisance parameter	Signal	$Z \to \tau\tau$	$Z \to ee$	$t\bar{t}$	EWK	QCD
Luminosity	4.5%	4.5%	4.5%	4.5%	-	-
$Z \to ll$ cross section	-	3%	3%	-	-	-
$t\bar{t}$ cross section	-	-	-	8%	-	-
Diboson cross sections	-	-	-	-	1%	-
Trigger efficiency	1%	1%	1%	1%	-	-
Electron selection efficiency	1%	1%	1%	1%	-	-
τ_h ID efficiency	6%	6%	-	6%	-	-
e-τ_h fake rate	-	-	6%	-	-	-
R_W	-	-	-	-	6%	-
$R_{OS/SS}$	-	-	-	-	-	5%
W estimate stat. uncertainty	-	-	-	-	1%	-
QCD estimate stat. uncertainty	-	-	-	-	-	2%
Electron energy scale	2%	2%	2%	2%	2%	-
τ_h energy scale	3%	3%	3%	3%	3%	-

Table 5.4: Nuisance parameters in the inclusive channel CL_s computation. The numbers in the table constitute the uncertainties on the estimates of the background yields, and shapes in the case of the energy scale uncertainties, and correspond to the width parameters of the *pdfs* describing the constraints on the nuisance parameters. An entry '-' denotes no dependence of a process on the corresponding nuisance parameter. The EWK background is dominated by W events with an approximately 3% contribution from diboson events such that luminosity and efficiency uncertainties on the diboson contribution are neglected.

Nuisance parameter	Signal	$Z \to \tau\tau$	$Z \to ee$	$t\bar{t}$	W	QCD
Luminosity	4.5%	4.5%	4.5%	4.5%	-	-
$Z \to ll$ cross section	-	3%	3%	-	-	-
$t\bar{t}$ cross section	-	-	-	8%	-	-
Trigger efficiency	1%	1%	1%	1%	-	-
Electron selection efficiency	1%	1%	1%	1%	-	-
τ_h ID efficiency	6%	6%	-	6%	-	-
e-τ_h fake rate	-	-	6%	-	-	-
b-tagging efficiency	10%	10 %	10%	10%	-	-
JES	5%	3.5%	3.5%	13.5%	-	-
R_W	-	-	-	-	6%	-
$R_{OS/SS}$	-	-	-	-	-	5%
W estimate stat. uncertainty	-	-	-	-	15%	-
QCD estimate stat. uncertainty	-	-	-	-	-	25%
Electron energy scale	2%	2%	2%	2%	2%	-
τ_h energy scale	3%	3%	3%	3%	3%	-

Table 5.5: Nuisance parameters in the b-tagged channel CL_s computation. The numbers in the table constitute the uncertainties on the estimates of the background yields, and shapes in the case of the energy scale uncertainties, and correspond to the width parameters of the *pdfs* describing the constraints on the nuisance parameters. An entry '-' denotes no dependence of a process on the corresponding nuisance parameter.

If the distribution of the measurements is $p(\bar{\theta}|\theta)$,

$$\rho(\theta|\bar{\theta}) \sim p(\bar{\theta}|\theta) \cdot \pi(\theta) \tag{5.8}$$

where $\pi(\theta)$ are uniform prior *pdfs* for the measurements. For gamma distributed $\rho(\theta|\bar{\theta})$, $p(\bar{\theta}|\theta)$ is Poisson distributed and for normal or log-normal distributed $\rho(\theta|\bar{\theta})$, $p(\bar{\theta}|\theta)$ is normally distributed [105].

The likelihood functions in Equation 5.6 are of the form

$$\mathcal{L}(d|\mu,\theta) = \prod_n p_n(d_n|\mu,\theta) \cdot \prod_i p_i(\bar{\theta}_i|\theta_i) \cdot \prod_j p_j(\bar{\theta}_j|\theta_j) \cdot \prod_k p_k(\bar{\theta}_k|\theta_k) \tag{5.9}$$

where $p_n(d_n|\mu,\theta)$ is the Poisson probability to observe d_n data in the nth bin of the $m_{e+\tau_h}$ histogram given μ and θ. The nuisance parameters θ_i represent the statistical uncertainties on data yields in background control regions, θ_j the luminosity, cross section and selection efficiency uncertainties, and θ_k the shape morphing parameters discussed in Section 5.5.1. The corresponding constraining *pdfs*, p_i, p_j and p_k, are derived according to Equation 5.8 using the posterior *pdfs* given in Equations 4.10, 4.13 and 5.5 respectively.

d may be the observed data or pseudodata generated by MC. When d is the observed data, $q_\mu = q_\mu^{obs}$ and $\hat{\theta}_\mu^{obs}$ and $\hat{\theta}_0^{obs}$ are the sets of nuisance parameters that maximise $\mathcal{L}_\mu$ for signal plus background ($\mu \neq 0$) and background only ($\mu = 0$) hypotheses respectively. MC pseudodata for the $\mu \neq 0$ and $\mu = 0$ hypotheses is generated using the best fit $\hat{\theta}_\mu^{obs}$ and $\hat{\theta}_0^{obs}$ respectively and is used to construct *pdfs* of the test statistic for each hypothesis. Given these *pdfs*, two p-values, p_μ and p_0, are defined which quantify the probability of observing the data under the $\mu \neq 0$ and $\mu = 0$ hypotheses respectively. The ratio of the two p-values is defined as $\mathrm{CL_s}(\mu)$:

$$p_\mu = \mathrm{P}\left(q_\mu \geq q_\mu^{obs} | \mu \cdot s(\hat{\theta}_\mu^{obs}) + b(\hat{\theta}_\mu^{obs})\right), \tag{5.10}$$

$$p_0 = \mathrm{P}\left(q_\mu \geq q_\mu^{obs} | b(\hat{\theta}_\mu^{obs})\right), \tag{5.11}$$

$$\mathrm{CL_s}(\mu) = \frac{p_\mu}{p_0}. \tag{5.12}$$

The value of μ at which $\mathrm{CL_s}(\mu) = 0.05$ is denoted μ_{95}. Signal yields greater than μ_{95} are excluded at 95% CL. Expected limits are evaluated by generating 10,000 sets of pseudodata under the $\mu = 0$ hypothesis, treating them as data, and computing $\mathrm{CL_s}(\mu_{95})$ for each pseudodataset. A cumulative $\mathrm{CL_s}(\mu_{95})$ distribution is then constructed and the

expected limit is given by the median of the distribution. The $\pm 1\sigma$ and $\pm 2\sigma$ bands are given by $\mathrm{CL_s}(\mu_{95})$ at the 16% and 84%, and 2.5% and 97.5% quantiles of the distribution respectively. For each m_A point, limits on $\sigma(pp \to \Phi) \cdot \mathcal{B}(\Phi \to \tau\tau)$ are given by

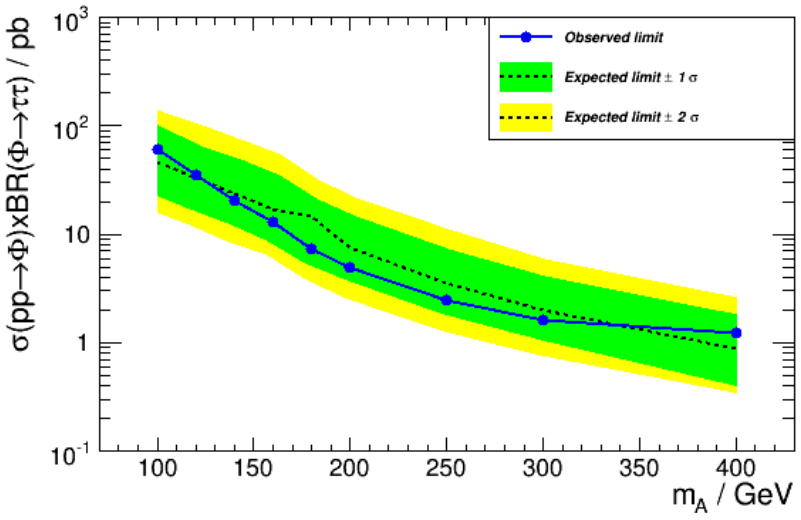
$$\sigma(pp \to \Phi) \cdot \mathcal{B}(\Phi \to \tau\tau) < \frac{\mu_{95}}{\mathcal{L}_{int} \cdot \mathcal{A} \cdot \epsilon \cdot \mathcal{B}_{e+\tau_h}}, \text{ 95\% CL.} \tag{5.13}$$

The products of signal acceptance, efficiency and branching ratio to the $e + \tau_h$ final state, $\mathcal{A} \cdot \epsilon \cdot \mathcal{B}_{e+\tau_h}$, for each channel and production mechanism are given in Table 5.2. For each mass point, the total $\mathcal{A} \cdot \epsilon \cdot \mathcal{B}_{e+\tau_h}$ is derived by addition of the $gg \to \Phi \to \tau\tau$ and $gg \to bb\Phi \to bb\tau\tau$ contributions weighted by the yields of the respective processes which are estimated using MC. The signal efficiency uncertainties are incorporated into the limit computation by the dependence of the signal yield on θ. The expected and observed upper limits on $\sigma(pp \to \Phi) \cdot \mathcal{B}(\Phi \to \tau\tau)$ in each channel are shown in Figure 5.14.

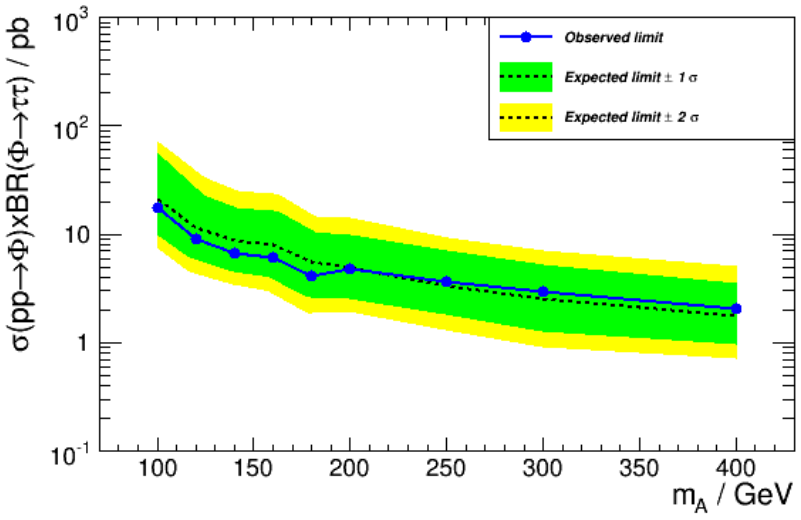
The limits are more stringent for large m_A where the expected backgrounds are small. Values of $\sigma(pp \to \Phi) \cdot \mathcal{B}(\Phi \to \tau\tau)$ greater than 18 and 1.3 pb are excluded at 95% CL for $m_A = 100$ and 400 GeV respectively. A peak in the expected limit in the inclusive channel at $m_A = 180$ GeV results from the $m_{e+\tau_h}$ distribution of the signal peaking in the same region as the $Z \to ee$ background. In the $140 \leq m_A \leq 300$ GeV region in the inclusive channel the background estimates in the corresponding $m_{e+\tau_h}$ region are in excess of the observed data resulting in a more stringent observed limit than expected. The converse is true for $m_A \leq 120$ GeV and $m_A = 400$ GeV. In the b-tagged channel the observed limit is more stringent than expected for $m_A \leq 180$ GeV and less stringent than expected for $m_A \geq 250$ GeV. The observed limits are within $\pm 1\sigma$ of the expected limits in all cases.

5.5.3 Interpretation in the MSSM

The $\Phi \to \tau\tau$ cross section is a function of $\tan\beta$ in the MSSM. The limits derived in Section 5.5.2 are therefore interpreted as upper limits on $\tan\beta$ in the m_h^{max} scenario of the MSSM, using the $\Phi \to \tau\tau$ cross sections given Section 1.5.1. These limits are shown in Figure 5.15 for the inclusive and b-tagged channels. The theoretical uncertainties on the observed limits are evaluated using the cross section uncertainties also given in Section 1.5.1. These uncertainties include renormalisation and factorisation scale, PDF and α_s uncertainties. Considering the central cross section predictions only, values of $\tan\beta$ greater than 12 and 40 are excluded at 95% CL for $m_A = 100$ and 400 GeV respectively.

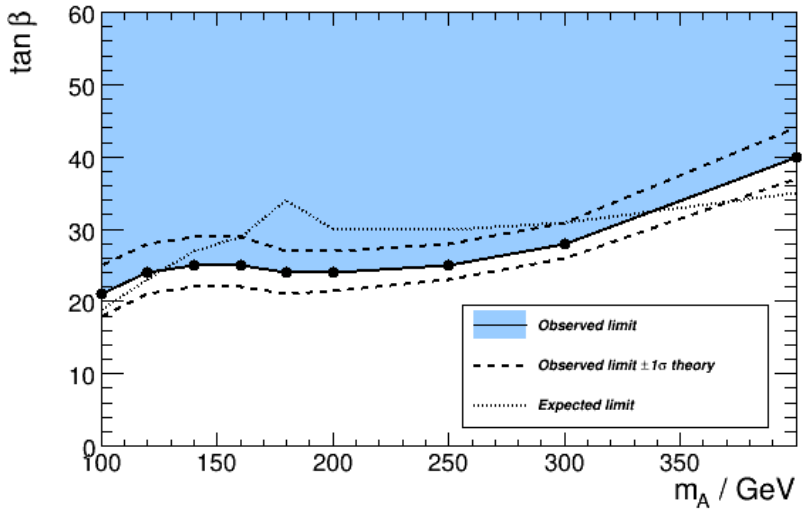


(a)

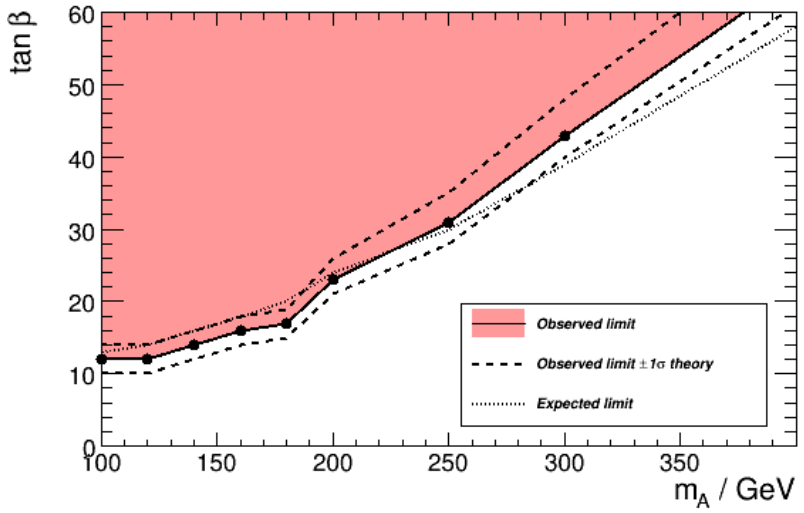


(b)

Figure 5.14: Expected and observed upper limits on $\sigma(pp \to \Phi) \cdot \mathcal{B}(\Phi \to \tau\tau)$ at 95% CL in the inclusive (a) and b-tagged (b) channels.



(a)



(b)

Figure 5.15: Expected and observed upper limits at 95% CL on tan β for varying m_A in the m_h^{max} scenario of the MSSM in the inclusive (a) and b-tagged (b) channels. The theoretical uncertainties on the observed limits include renormalisation and factorisation scale, PDF and α_s uncertainties. Considering the central $\Phi \to \tau\tau$ cross section predictions, the shaded areas of the parameter space are excluded at 95% CL.

Figure 5.16 shows a comparison of the limits on tan β between the inclusive and b-tagged channels with the central cross section predictions. At small m_A, $m_A \leq 200$ GeV, the limits in the b-tagged channel are most stringent as a result of the larger signal to background ratio in that channel in the corresponding $m_{e+\tau_h}$ region. At large m_A, backgrounds are small in the corresponding $m_{e+\tau_h}$ regions of both channels, as shown in Figures 5.7 and 5.8, and signal efficiency is greatest in the inclusive channel. As a result, the most stringent limits for $m_A \geq 250$ GeV are set in the inclusive channel.

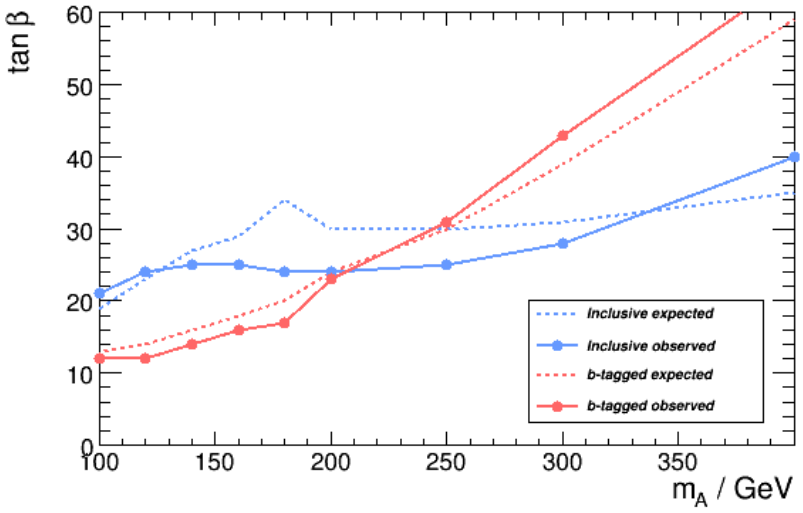


Figure 5.16: Expected and observed upper limits on tan β at 95% CL in the inclusive and b-tagged channels.

The inclusive and b-tagged channels are therefore seen to provide complementary tan β sensitivity in the $e + \tau_h$ channel. The greatest sensitivity at large and small m_A is provided by the inclusive and b-tagged channels respectively. These results may be combined with searches in the $\mu + \tau_h$ and $e + \mu$ final states, as described in Chapter 6, to improve the constraints on tan β.

Chapter 6

Conclusions

The spontaneous breaking of the electroweak symmetry in the SM predicts the existence of a massive scalar Higgs boson. The hierarchy problem arising from the quadratically divergent radiative corrections to the bare Higgs boson mass has a natural solution in supersymmetric models. The LHC facilitates exploration of physics at the TeV scale and searches for Higgs bosons and new physics such as SUSY. The neutral Higgs bosons of the MSSM have enhanced couplings to down-type fermions at large tan β relative to the SM so that $\Phi \to bb$ and $\Phi \to \tau\tau$ are the dominant decay modes of these Higgs bosons. The $\Phi \to \tau\tau$ mode may be isolated from large hadronic backgrounds making it the optimal search channel at the LHC.

MSSM neutral Higgs boson searches at the LHC may be conducted inclusive of both the $gg \to \Phi$ and $gg \to bb\Phi$ production mechanisms or exclusively in the $gg \to bb\Phi$ channel with the use of b-tagging. $\Phi \to \tau\tau$ analyses at the LHC require efficient reconstruction of τ-jets with discrimination against quark and gluon jet backgrounds. τ_h reconstruction and ID methods are calibrated by reconstructing $Z \to \tau\tau$ events and measuring the cross section $\sigma(pp \to Z + X) \cdot \mathcal{B}(Z \to \tau\tau)$.

A measurement of the $Z \to \tau\tau$ cross section in the $e + \tau_h$ final state at CMS with 2.1 fb^{-1} of integrated luminosity at $\sqrt{s} = 7$ TeV yields the result

$$\sigma\left(pp \to Z + X\right) \cdot \mathcal{B}\left(Z \to \tau\tau\right) = 1.06^{+0.05}_{-0.04}\left(\text{stat.}\right) \pm 0.07\left(\text{syst.}\right) \pm 0.05\left(\text{lumi.}\right) \text{nb} \quad (6.1)$$

which is in good agreement with the SM prediction of 0.972 ± 0.042 nb at NNLO and CMS measurements of the $Z \to ee$ and $Z \to \mu\mu$ cross sections. This result validates the CMS τ_h reconstruction algorithms and the methods used to estimate backgrounds to

$Z \to \tau\tau \to e + \tau_h$ events which result primarily from QCD multijet, $Z \to ee$ and W+jet production.

In a search for neutral Higgs bosons decaying to τ pairs in the $e + \tau_h$ final state, using the benchmark methods of the $Z \to \tau\tau$ cross section measurement and the same dataset, no significant excess of events is observed over the predicted SM backgrounds. $\Phi \to \tau\tau$ cross sections, $\sigma(pp \to \Phi + X) \times \mathcal{B}(\Phi \to \tau\tau)$, greater than 18 and 1.3 pb are excluded at 95% CL for $m_A = 100$ and $m_A = 400$ GeV respectively. The results are interpreted as upper limits on tan β in the MSSM. Values of tan β greater than 12 and 40 are excluded at 95% CL for $m_A = 100$ and 400 GeV respectively in the m_h^{max} scenario. The inclusive and and b-tagged channels are found to be complementary in the $e + \tau_h$ final state with the inclusive channel providing the greatest tan β sensitivity at large m_A and the b-tagged channel providing the greatest sensitivity at small m_A.

The CMS $\Phi \to \tau\tau$ searches in the inclusive and b-tagged channels and $e + \tau_h$, $\mu + \tau_h$ and $e + \mu$ final states are combined, accounting for correlations of systematic uncertainties between channels, to maximise overall sensitivity to Φ production. The combination using the CMS Run 2011A and Run 2011B datasets, corresponding to $\mathcal{L}_{int} = 4.6$ fb^{-1}, yields no evidence for MSSM neutral Higgs bosons and stringent constraints on tan β as shown in Figure 6.1 [93]. A large area of the m_A-tanβ parameter space of the m_h^{max} scenario is excluded at 95% CL.

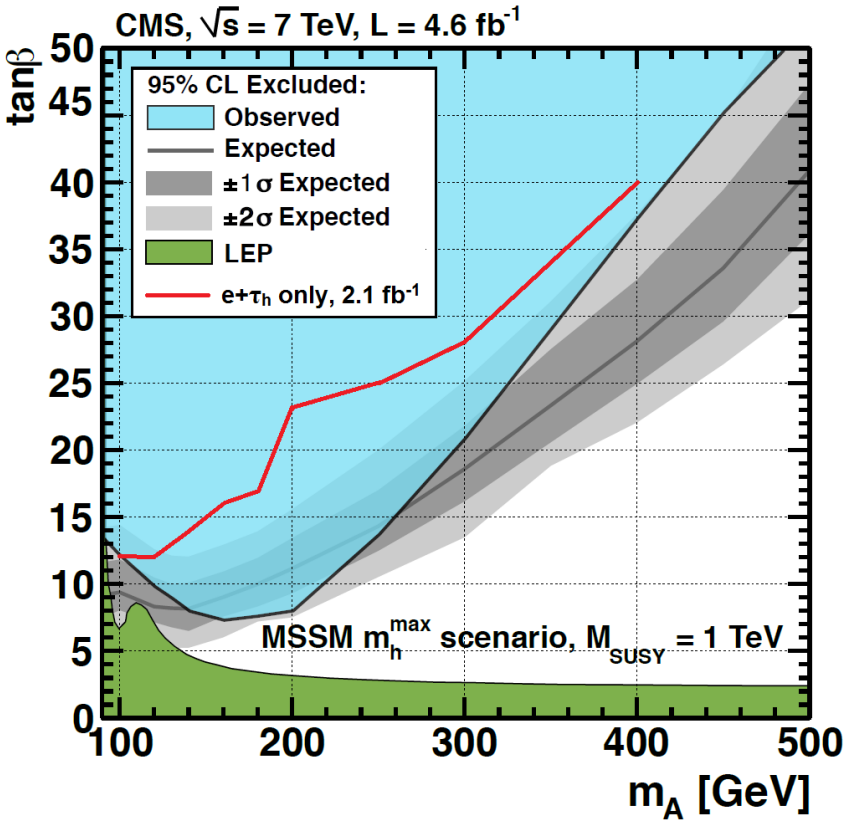


Figure 6.1: Expected and observed upper limits on tan β at 95% CL as a function of m_A in the m_h^{max} scenario of the MSSM resulting from $\Phi \to \tau\tau$ searches conducted by CMS with 4.6 fb^{-1} of integrated luminosity $\sqrt{s} = 7$ TeV [93]. The results are derived from searches in the $e + \tau_h$, $\mu + \tau_h$ and $e + \mu$ final states. The theoretical uncertainties on the Φ production cross section are incorporated into the limit computation. The observed limits obtained in the $e + \tau_h$ final state only with 2.1 fb^{-1} as derived in Chapter 5 are also shown.

Bibliography

[1] Particle Data Group, "Review of Particle Physics". *J. Phys. G* **37** (2010) 075021. doi:10.1088/0954-3899/37/7A/075021.

[2] D. Griffiths, "Introduction to Elementary Particles". 2nd edition, Wiley-VCH (2008). ISBN 978-3527406012.

[3] I. J. R. Aitchison and A. J. G. Hey, "Gauge Theories in Particle Physics, Volume I: From Relativistic Quantum Mechanics to QED". 3rd edition, Taylor and Francis (2003). ISBN 978-0750309509.

[4] I. J. R. Aitchison and A. J. G. Hey, "Gauge Theories in Particle Physics, Volume II: QCD and the Electroweak Theory". 3rd edition, Taylor and Francis (2003). ISBN 978-0750308649.

[5] S. Weinberg, "A Model of Leptons". *Phys. Rev. Lett.* **19** (1967) 1264. doi:10.1103/PhysRevLett.19.1264.

[6] A. Salam, "Volume 8 of Nobel Symposium. Elementary particle theory: relativistic groups and analyticity". John Wiley and Sons (1968), pp367. ISBN 978-0470838426.

[7] D. J. Gross and F. Wilczek, "Asymptotically Free Gauge Theories. I". *Phys. Rev. D* **8** (1973) 3633. doi:10.1103/PhysRevD.8.3633.

[8] G. 't Hooft, "Renormalizable Lagrangians for massive Yang-Mills fields". *Nucl. Phys. B* **35** (1971) 167. doi:10.1016/0550-3213(71)90139-8.

[9] A. J. Gross and F. Wilczek, "Ultraviolet Behaviour of Non-Abelian Gauge Theories". *Phys. Rev. Lett.* **30** (1973) 1343. doi:10.1103/PhysRevLett.30.1343.

[10] H. D. Politzer, "Reliable Perturbative Results for Strong Interactions". *Phys. Rev. Lett.* **30** (1973) 1346. doi:10.1103/PhysRevLett.30.1346.

[11] S. L. Glashow, "Partial-symmetries of weak interactions." *Nucl. Phys.* **22** (1961)

579. doi:10.1016/0029-5582(61)90469-2.

[12] A. Salam and J. C. Ward, "Gauge Theory of Elementary Interactions". *Phys. Rev.* **136** (1964) B763. doi:10.1103/PhysRev.136.B763.

[13] N. Cabibbo, "Unitary Symmetry and Leptonic Decays". *Phys. Rev. Lett.* **10** (1963) 531. doi:10.1103/PhysRevLett.10.531.

[14] M. Kobayashi and T. Maskawa, "CP-Violation in the Renormalizable Theory of Weak Interaction". *Prog. Theor. Phys.* **49** (1973) 652. doi:10.1143/PTP.49.652.

[15] P. W. Higgs, "Spontaneous Symmetry Breakdown without Massless Bosons". *Phys. Rev.* **145** (1966) 1156. doi:10.1103/PhysRev.145.1156.

[16] T. W. B. Kibble, "Symmetry breaking in non-Abelian gauge theories". *Phys. Rev.* **155** (1967) 1554. doi:10.1103/PhysRev.155.1554.

[17] S. P. Martin, "A Supersymmetry Primer". arXiv:hep-ph/9709356v6 (2011).

[18] WMAP Collaboration, "Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation". *Astrophys. J. Suppl.* **180** (2009) 330. doi:10.1088/0067-0049/180/2/330.

[19] LHC Study Group, "The Large Hadron Collider: conceptual design". *CERN Preprint* (1995). CERN-AC-95-05-LHC.

[20] J. D. Wells, "Lectures on Higgs Boson Physics in the Standard Model and Beyond". arXiv:0909.4541v1 [hep-ph] (2009).

[21] LEP Electroweak Working Group, "The LEP Electroweak Working Group: Status of March 2012". http://lepewwg.web.cern.ch/LEPEWWG (2012).

[22] J. Wess and B. Zumino, "A lagrangian model invariant under supergauge transformations". *Phys. Lett. B* **49** (1974) 52. doi:10.1016/0370-2693(74)90578-4.

[23] O. Buchmueller *et al.*, "Supersymmetry in Light of 1/fb of LHC Data". arXiv:1110.3568v1 [hep-ph] (2011).

[24] S. Dimopoulos and H. Georgi, "Softly broken supersymmetry and SU(5)". *Nucl. Phys. B* **193** (1981) 150. doi:10.1016/0550-3213(81)90522-8.

[25] M. Carena and H. E. Haber, "Higgs Boson Theory and Phenomenology". *Prog. Part. Nucl. Phys.* **50** (2003) 63. doi:10.1016/S0146-6410(02)00177-1.

[26] L. Evans and P. Bryant, "LHC Machine", *JINST* **3** (2008) S08001. doi:10.1088/1748-0221/3/08/S08001.

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

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

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

[30] ALICE Collaboration, "The ALICE Experiment at the CERN LHC". *JINST* **3** (2008) S08002. doi:10.1088/1748-0221/3/08/S08002.

[31] CERN, "LHC Cooldown Status". http://cdsweb.cern.ch/record/1212916 (2009).

[32] CMS Collaboration, "CMS Luminosity - Public Results". https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults (2012).

[33] TOTEM Collaboration, "First measurement of the total proton-proton cross section at the LHC energy of $\sqrt{s} = 7$ TeV". *EPL* **96** (2011) 21002. doi:10.1209/0295-5075/96/21002.

[34] R. K. Ellis, W. J. Stirling and B. R. Webber, "QCD and Collider Physics". Cambridge University Press (2003). ISBN 978-0521545891.

[35] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt, "Parton distributions for the LHC". *Eur. Phys. J. C* **63** (2009) 189. doi:10.1140/epjc/s10052-009-1072-5.

[36] ALEPH, DELPHI, L3 and OPAL Collaborations, "Search for neutral MSSM Higgs bosons at LEP". *Eur. Phys. J. C* **47** (2006) 547. doi:10.1140/epjc/s2006-02569-7.

[37] CDF Collaboration, "Search for Higgs Bosons Predicted in Two-Higgs-Doublet Models via Decays to Tau Lepton Pairs in 1.96 TeV $p\bar{p}$ collisions". *Phys. Rev. Lett.* **103** (2009) 201801. doi:10.1103/PhysRevLett.103.201801.

[38] D$\emptyset$ Collaboration, "Search for Higgs bosons decaying to $\tau^+\tau^-$ pairs in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV". *Phys. Lett. B* **707** (2012) 323. doi:10.1016/j.physletb.2011.12.050.

[39] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs cross sections: 1. Inclusive observables". arXiv:1101.0593v3 [hep-ph] (2011).

[40] LHC Higgs Cross Section Working Group, "Handbook of LHC Higgs Cross Sections: 2. Differential Distributions". arXiv:1201.3084v1 [hep-ph] (2012).

[41] S. Heinemeyer, W. Hollik and G. Weiglein, "FeynHiggs: a program for the calculation of the masses of the neutral CP-even Higgs bosons in the MSSM". *Comput. Phys. Commun.* **124** (2000) 76. doi:10.1016/S0010-4655(99)00364-1.

[42] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt, "Heavy-quark mass dependence in global PDF analyses and 3- and 4- flavour parton distributions". *Eur. Phys. J. C* **70** (2010) 51. doi:10.1140/epjc/s10052-010-1462-8.

[43] M. Spira, "HIGLU: A Program for the Calculation of the Total Higgs Production Cross Section at Hadron Colliders via Gluon Fusion including QCD Corrections". arXiv:hep-ph/9510347v1 (1995).

[44] R. V. Harlander and W. B. Kilgore, "Next-to-Next-to-Leading Order Higgs Production at Hadron Colliders". *Phys. Rev. Lett.* **88** (2002) 201801. doi:10.1103/PhysRevLett.88.201801.

[45] R. V. Harlander and W. B. Kilgore, "Production of a pseudo-scalar Higgs boson at hadron colliders at next-to-next-to leading order". *JHEP* **10** (2002) 017. doi:10.1088/1126-6708/2002/10/017.

[46] S. Dittmaier, M. Krämer and M. Spira, "Higgs Radiation off Bottom Quarks at the Tevatron and the LHC". *Phys. Rev. D* **70** (2004) 074010. doi:10.1103/PhysRevD.70.074010.

[47] S. Dawson, C. B. Jackson, L. Reina and D. Wackeroth, "Exclusive Higgs Boson Production with bottom quarks at Hadron Colliders". *Phys. Rev. D* **69** (2004) 074027. doi:10.1103/PhysRevD.69.074027.

[48] R. V. Harlander and W. B. Kilgore, "Higgs boson production in bottom quark fusion at next-to-next-to-leading order". *Phys. Rev. D* 68 (2003) 013001. doi:10.1103/PhysRevD.68.013001.

[49] ATLAS Collaboration, "ATLAS detector and physics performance: Technical Design Report, 2". *CERN Scientific Committee Paper* (1999). CERN-LHCC-99-015.

[50] CMS Collaboration, "CMS Physics Technical Design Report, Volume II: Physics

Performance". *J. Phys. G* **34** (2007) 995. doi:10.1088/0954-3899/34/6/S01.

[51] ATLAS Collaboration, "Search for neutral MSSM Higgs bosons decaying to $\tau^+\tau^-$ pairs in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector". *ATLAS Note* (2011). ATLAS-CONF-2011-132.

[52] CMS Collaboration, "Search for Neutral Higgs Bosons Decaying to Tau Pairs in pp Collisions at $\sqrt{s} = 7$ TeV". *CMS Physics Analysis Summary* (2011). CMS-PAS-HIG-11-009.

[53] S. D. Drell and T. M. Yan, "Massive Lepton-Pair Production in Hadron-Hadron Collisions at High Energies". *Phys. Rev. Lett.* **25** (1970) 316. doi:10.1103/PhysRevLett.25.316.

[54] R. Gavin, Y. Li, F. Petriello and S. Quackenbush, "FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order". *Comput. Phys. Commun.* **182** (2011) 2388. doi:10.1016/j.cpc.2011.06.008.

[55] CMS Collaboration, "Measurement of the inclusive Z cross section via decays to tau pairs in pp collisions at $\sqrt{s} = 7$ TeV". *JHEP* **08** (2011) 117. doi:0.1007/JHEP08(2011)117.

[56] CMS Collaboration, "CMS Physics Technical Design Report Volume I: Detector Performance and Software". *CERN Scientific Committee Paper* (2006). CERN-LHCC-2006-001.

[57] CMS Collaboration, "CMS Tracking Performance Results from Early LHC Operation". *Eur. Phys. J. C* **70** (2010) 1165. doi:10.1140/epjc/s10052-010-1491-3.

[58] R. E. Kalman, "A new approach to linear filtering and prediction problems". *J. Basic Eng.* **82** (1960) 35. doi:10.1109/ICASSP.1982.1171734.

[59] CMS Collaboration, "Tracking and Primary Vertex Results in First 7 TeV Collisions". *CMS Physics Analysis Summary* (2010). CMS-PAS-TRK-10-005.

[60] CMS Collaboration, "Absolute Calibration of Luminosity Measurement at CMS: Summer 2011 Update". *CMS Physics Analysis Summary* (2011). CMS-PAS-EWK-11-001.

[61] E. Chabanat and N. Estre, "Deterministic Annealing for Vertex Finding at CMS". *CHEP 2004 Conference Note* (2004). CERN-2005-002.

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

[63] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus and E_T^{miss}". *CMS Physics Analysis Summary* (2009). CMS-PAS-PFT-09-001.

[64] CMS Collaboration, "Jet Performance in pp Collisions at $\sqrt{s} = 7$ TeV". *CMS Physics Analysis Summary* (2010). CMS-PAS-JME-10-003.

[65] CMS Collaboration, "Missing Transverse Energy Performance in Minimum-Bias and Jet Events from Proton-Proton Collisions at $\sqrt{s} = 7$ TeV". *CMS Physics Analysis Summary* (2010). CMS-PAS-JME-10-004.

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

[67] W. Adam, R. Frühwirth, A. Strandlie and T. Todorov, "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC". *J. Phys. G* **31** (2005) N9. doi:10.1088/0954-3899/31/9/N01.

[68] CMS Collaboration, "Commissioning of the Particle-Flow Reconstruction in Minimum-Bias and Jet Events from pp Collisions at $\sqrt{s} = 7$ TeV". *CMS Physics Analysis Summary* (2010). CMS-PAS-PFT-10-002.

[69] CMS Collaboration, "Commissioning of the particle-flow event reconstruction with leptons from J/Psi and W decays at 7 TeV". *CMS Physics Analysis Summary* (2010). CMS-PAS-PFT-10-003.

[70] M. Cacciari, G. P. Salam and G. Soyez, "The anti-k_t jet clustering algorithm". *JHEP* **04** (2008) 063. doi:10.1088/1126-6708/2008/04/063.

[71] CMS Collaboration, "Determination of the Jet Energy Scale in CMS with pp Collisions at $\sqrt{s} = 7$ TeV". *CMS Physics Analysis Summary* (2010). CMS-PAS-JME-10-010.

[72] UA2 Collaboration, "Measurement of production and properties of jets at the CERN anti-pp collider". *Z. Phys. C* **20** (1983) 117. doi:10.1007/BF01573214.

[73] DØ Collaboration, "Determination of the absolute jet energy scale in the DØ calorimeters". *Nucl. Inst. Meth. A* **424** (1999) 352. doi:10.1016/S0168-9002(98)01368-0.

[74] CMS Collaboration, "Electromagnetic calorimeter calibration with 7 TeV data". *CMS Physics Analysis Summary* (2010). CMS-PAS-EGM-10-003.

[75] CMS Collaboration, "Electron reconstruction and identification at $\sqrt{s} = 7$ TeV". *CMS Physics Analysis Summary* (2010). CMS-PAS-EGM-10-004.

[76] S. Baffioni *et al.*, "Electron Reconstruction in CMS". *CMS Note* (2006). CMS-NOTE-2006-040.

[77] J. Marriner, "Secondary Vertex Fit with Mass and Pointing Constraints (CTVMFT)". *CDF Note* (1993). CDF/DOC/SEC_VTX/PUBLIC/1996.

[78] CMS Collaboration, "Measurement of the inclusive W and Z production cross sections in pp collisions at $\sqrt{s} = 7$ TeV with the CMS experiment". *JHEP* **10** (2011) 132. doi:10.1007/JHEP10(2011)132.

[79] CMS Collaboration, "Performance of muon identification in pp collisions at $\sqrt{s} = 7$ TeV". *CMS Physics Analysis Summary* (2010). CMS-PAS-MUO-10-002.

[80] CMS Collaboration, "Performance of τ-lepton reconstruction and identification in CMS". *JINST* **7** (2012) P01001. doi:10.1088/1748-0221/7/01/P01001.

[81] CMS Collaboration, "Performance of b-jet identification in CMS". *CMS Physics Analysis Summary* (2011). CMS-PAS-BTV-11-001.

[82] CMS Collaboration, "Status of b-tagging tools for 2011 data analysis". *CMS Physics Analysis Summary* (2011). CMS-PAS-BTV-11-002.

[83] CMS Collaboration, "Observation of Z+b, Z$\to$ *ee*, $\mu\mu$ with CMS at $\sqrt{s} = 7$ TeV". *CMS Physics Analysis Summary* (2011). CMS-PAS-EWK-10-015.

[84] CMS Collaboration, "Missing transverse energy performance of the CMS detector". *JINST* **6** (2011) P09001. doi:10.1088/1748-0221/6/09/P09001.

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

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

[87] S. Alioli, P. Nason, C. Oleari and E. Re, "NLO vector-boson production matched with shower in POWHEG", *JHEP* **07** (2008) 060. arXiv:1002.2581 [hep-ph].

[88] P. Nason and B. Webber, "Next-to-Leading-Order Event Generators". arXiv:1202.1251v1 [hep-ph] (2012).

[89] J. Alwall *et al.*, "MadGraph/MadEvent v4: The New Web Generation". *JHEP* **09** (2007) 028. doi:10.1088/1126-6708/2007/09/028.

[90] T. Sjöstrand, S. Mrenna and P. Z. Skands, "PYTHIA 6.4 physics and manual". *JHEP* **05** (2006) 026. arXiv:hep-ph/0603175.

[91] S. Jadach, Z. Was, R. Decker and J. H. Kuhn, "The tau decay library TAUOLA: version 2.4". *Comput. Phys. Commun* **76** (1993) 361. doi:10.1016/0010-4655(93)90061-G.

[92] J. Allison *et al.*, "Geant4 developments and applications". *IEEE Trans. Nucl. Sci.* **53** (2006) 270. doi:10.1109/TNS.2006.869826.

[93] CMS Collaboration, "Search for neutral Higgs bosons decaying to tau pairs in pp collisions at $\sqrt{s} = 7$ TeV". arXiv:1202.4083v1 [hep-ex] (2012).

[94] C. Cuenca Almenar, "Search for the neutral MSSM Higgs bosons in the ditau decay channels at CDF Run II". *Fermilab Thesis* (2008). FERMILAB-THESIS-2008-86.

[95] CMS Collaboration, "Search for Neutral MSSM Higgs Bosons Decaying to Tau Pairs in pp Collisions at $\sqrt{s} = 7$ TeV". *Phys. Rev. Lett.* **106** (2011) 231801. doi:10.1103/PhysRevLett.106.231801.

[96] CMS Collaboration, "Combination of top pair production cross section measurements". *CMS Physics Analysis Summary* (2011). CMS-PAS-TOP-11-024.

[97] CMS Collaboration, "Measurement of WW and observation of WZ and ZZ in leptonic modes". *CMS Physics Analysis Summary* (2011). CMS-PAS-EWK-11-010.

[98] R. J. Barlow, "Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences". Reprint edition, Wiley-Blackwell (1989). ISBN 978-0471922957.

[99] C. Forbes, M. Evans, N. Hastings and B. Peacock, "Statistical Distributions". 4$^{\text{th}}$ edition, Wiley-Blackwell (2011). ISBN 978-0470390634.

[100] CMS Collaboration, "Search for Neutral Higgs Bosons Decaying to Tau Pairs in pp Collisions at $\sqrt{s} = 7$ TeV". *CMS Physics Analysis Summary* (2011). CMS-PAS-HIG-11-020.

[101] J. Alwall, Q. Li and F. Maltoni, "Matched predictions for Higgs production via

heavy-quark loops in the SM and beyond". arXiv:1110.1728v1 [hep-ph] (2011).

[102] A. L. Read, "Modified frequentist analysis of search results (the CL_s method)". *CERN Article* (2000). CERN-OPEN-2000-205.

[103] ATLAS Collaboration, CMS Collaboration and LHC Higgs Combination Group, "Procedure for the LHC Higgs boson search combination in Summer 2011". *ATLAS and CMS Note* (2011). ATL-PHYS-PUB-2011-11. CMS-NOTE-2011-005.

[104] T. Bayes, "An Essay towards solving a Problem in the Doctrine of Chances". *Phil. Trans.* **53** (1763) 370. doi:10.1098/rstl.1763.0053.

[105] CMS Collaboration, "SM Higgs Combination". *CMS Physics Analysis Summary* (2011). CMS-PAS-HIG-11-011.

Appendix: List of acronyms

ALICE A Large Ion Collider Experiment 33

ATLAS A Toroidal Large Hadron Collider Apparatus 33

AVF Adaptive Vertex Finder 60

CKM Cabibbo-Kobayashi-Maskawa 24

CL Confidence Level 27

CMS Compact Muon Solenoid 33

CSC Cathode Strip Chamber 56

DT Drift Tube 56

EB Electromagnetic Calorimeter Barrel 52

ECAL Electromagnetic Calorimeter 47

EE Electromagnetic Calorimeter Endcaps 52

EWK Electroweak . . . 112

GSF Gaussian Sum Filter . . . 62

HB Hadron Calorimeter Barrel . . . 54

HCAL Hadron Calorimeter . . . 53

HE Hadron Calorimeter Endcaps . . . 54

HF Forward Calorimeter . . . 54

HLT High Level Trigger . . . 47

HPS Hadron Plus Strips . . . 77

HO Hadron Calorimeter Outer . . . 54

ID Identification . . . 40

JES Jet Energy Scale . . . 63

L1 Level-1 . . . 47

LEP Large Electron-Positron Collider . . . 35

LHC Large Hadron Collider . . . 25

LHCb Large Hadron Collider beauty 33

LO Leading Order 24

MC Monte Carlo 62

MPF Missing transverse energy Projection Fraction 66

MSSM Minimal Supersymmetric Standard Model 28

MSTW08 Martin-Stirling-Thorne-Watt-2008 36

NLO Next-to-Leading Order 24

NNLO Next-to-Next-to-Leading Order 24

OS Opposite Sign 71

PDF Parton Distribution Function 35

pdf probability density function 113

PF Particle Flow 59

PU Pile-Up 35

PV Primary Vertex 59

QCD Quantum Chromodynamics 23

RPC Resistive Plate Chamber 56

SM Standard Model 23

SS Same Sign 104

SSV Simple Secondary Vertex 85

SSVHE Simple Secondary Vertex High Efficiency 85

SSVHP Simple Secondary Vertex High Purity 85

SUSY Supersymmetry 28

SV Secondary Vertex 84

TEC Tracker Endcaps 50

TIB Tracker Inner Barrel 50

TID Tracker Inner Disks 50

TOB Tracker Outer Barrel 50

VEV Vacuum Expectation Value 26