The search for the Higgs boson in tauon pairs at the ATLAS experiment

Qi Tao Shao

Submitted in total fulfilment of the requirements

of the degree of Doctor of Philosophy

April 2013

School of Physics

The University of Melbourne

Abstract

The Higgs boson is a particle that's predicted to exist by spontaneous electroweak symmetry breaking. Electroweak symmetry breaking is an essential part of the Standard Model of particle physics, as it generates masses for the electroweak gauge bosons. Finding the Higgs boson is integral to our understandings of the fundamental particles and their interactions.

Searches for the Higgs boson are conducted by the ATLAS experiment using proton-proton collisions at the Large Hadron Collider. One of these searches is performed using the $H \rightarrow \tau\tau$ decays, which has a clean detection signature and, with $H \rightarrow b\bar{b}$, is one of the only two viable fermonic search channels. Using collision the 4.7 fb^{-1} of data collected at $\sqrt{s}$ = 7 TeV, the $H \rightarrow \tau\tau$ analysis excludes the Higgs boson at approximately 3 times the expected cross section for 100 < m_H < 120 GeV and 5 to 12 times the expected cross section for 130 < m_H < 150 GeV. The $H \rightarrow \tau\tau$ search results are combined with those from the other channels to achieve better sensitivities.

The combined results have excluded most Higgs masses between 110 and 500 GeV. The only region that is not excluded is at m_H = 126 GeV, where an excess above the background expectations is observed in multiple bosonic channels. This excess has a combined local significance of 5.9 σ. ATLAS claims this observed excess as a discovery of a new bosonic particle, whose properties have thus far been measured to be consistent with that of the Higgs boson.

Declaration

This is to certify that

i. the thesis comprises only my original work towards the PhD except where indicated in the Preface,

ii. due acknowledgement has been made in the text to all other material used,

iii. the thesis is fewer than 100 000 words in length, exclusive of tables, maps, bibliographies and appendices.

Qi Tao Shao

Preface

The ATLAS collaboration consists of approximately 3000 scientists from 164 institutions worldwide. The experimental work is manpower intensive and as such, analyses are usually performed in groups. The standard practice is for individuals to join analysis groups, where they are are assigned specific roles that contribute to the group's overall research goals. The collaborative nature of this setup also allows results to be cross checked.

The work in chapter 5 was performed by the author with the data-driven cross checks provided by the other collaborators of the $Z \to \tau\tau \to ll + 4\nu$ sub-group. The cross section measurements presented in chapter 6 was performed by the author in collaboration with the same group. The author had direct contributions to all the work presented, except for the data-driven estimates of the $W \to l\nu$ and $t\bar{t}$ cross sections and the estimation of the $e\mu$ multijet background, which were performed by the other group members.

The $H \to \tau\tau \to ll + 4\nu$ analysis, presented in chapter 8, was performed by the author in collaboration with the $H \to \tau\tau \to ll + 4\nu$ sub-group. All work presented had direct contributions from the author, with exception to the $Z \to \tau\tau$ and $\gamma* / Z \to ee, \mu\mu$ background estimates, which were developed and performed by the other members of the group. The combined $H \to \tau\tau$ search results, presented in chapter 9, includes results from the hadronic $H \to \tau\tau$ sub-channels. These analyses were performed by the collaborators of the hadronic $H \to \tau\tau$ sub-groups. Finally, the results from the non-$H \to \tau\tau$ search channels, presented in chapter 10, were performed by the collaborators of the other Higgs analysis groups.

Acknowledgements

I would first like to thank my primary supervisor, Elisabetta Barberio, who has been a great mentor and teacher over the years. She has always been there to help me whenever I needed it. A lot of what I have achieved during my Ph.D can be directly attributed to her efforts on my behalf. Thanks also goes out to Geoffrey Taylor, who's my second supervisor and also the head of Experimental Particle Physics (EPP) group in Melbourne. His management of the group has been outstanding over the years, ensuring that the group always had the resources it needed. Thank you also to Andrew Melatos, my tertiary supervisor, who has always been there to check how I was going and to provide necessary advice.

Completing a Ph.D in EPP has been very enjoyable mainly because of the people in the group. I would like to thank all the EPP post doctorates, past and present, who have always been there to help me when I needed it. So thank you to Ulrich Felzmann, Martin White, Aldo Saavedra, Guilherme Hanninger, Takashi Kubota and in particular Antonio Limosani, who has been always been there for me both in Australia and abroad. I would also like to thank all the EPP students who I've had the pleasure to work and live with. These include Anna Phan, Phillip Urquijo, Nadia Davidson, Will Davey, Nik Patel, T'Mir Danger Julius, Vivien Lee, Anthony Morley, Stefanie Elbracht-Leong, Anja Schubert, Dave Peak, David Jennens, KG Tan, Nick Rodd, Pere Rados and many others past and present.

To all my friends in the physics department, you've all been great to hang and procastinate with. So thank you to Mark Bennett, Alex Malone, Francis Dillon, Jayne Thompson, Letizia Sammut, Steve Petrie, Brad Greig, Masum Rab, Ahmad Galea, Marko Milisevic, Maxim Priymak, Andrea Ruff, Christina Magoulas, Laurence Spiller, Catherine de Burgh-Day, Loren Bruns Jr and many others. Thank you also to all my friends at CERN, who have made all the months I spent in the town of St. Genis so pleasant. So to Maximilian Micheler, James Mylroie-Smith, Dermot Moran, Stephan Farry, Seamus O'Geibheannaigh, Sarah (the) Baker, John Almond, Gareth Brown, Vikash Chavda, Vicki Moeller, Abdi Noor, John Alison, Dominic Olivito, John Penwell, Jamie Saxon, and so many others, I thank you all.

Last but not least, I would like to thank my parents for supporting me in my Ph.D efforts. They bought me an apartment just one minute walk away from my office at the start of my Ph.D. Life's just convenient when you don't have to worry about commuting everyday.

Contents

List of Tables

List of Figures

Introduction

The Higgs boson is a particle that's hypothesised to exist as a consequence of the spontaneous electroweak symmetry breaking. Electroweak symmetry breaking was first proposed as a way of generating masses for the electroweak gauge bosons. Ever since its prediction, many experiments have searched for this particle.

At the time at which the analyses in this thesis were performed, the ATLAS experiment searched for the Higgs using proton-proton collisions at the Large Hadron Collider (LHC). ATLAS is general purpose detector that's designed to measure and search for a wide range of phenomena, which includes searching for the Higgs boson. The main analyses presented in this thesis used collision data collected at a centre-of-mass energy of $\sqrt{s}$ = 7 TeV.

Precision electroweak measurements have so far preferred a low mass Higgs boson ($\lesssim$ 140 GeV/c^2). For these masses, $H \to \tau\tau$, $H \to b\bar{b}$, $H \to \gamma\gamma$, $H \to WW$ and $H \to ZZ$ are all viable search channels. Analyses are performed in each of these channels by the ATLAS experiment and their results are combined to achieve the best statistical significance. Observations of the Higgs boson in multiple decay modes is also required to confirm any discovery. This thesis will focus on the $H \to \tau\tau$ analyses and in particular the $H \to \tau\tau \to ll + 4\nu$ sub-channel. The $H \to \tau\tau$ search channel is important because it has a high branching ratio, a clean signature and provides an essential test for the Yukawa Higgs couplings to the fermions.

For Higgs masses below 130 GeV, the most dominant background to the $H \to \tau\tau$ analyses comes from the $Z \to \tau\tau$ process. This background is irreducible, due to the similar topology and kinematics of its final state particles. To ensure that this background is well modelled, the $Z \to \tau\tau$ production cross section is measured using the $Z \to \tau\tau \to ll + 4\nu$ and $Z \to \tau\tau \to l\tau_h + 3\nu$ decay modes. This thesis will focus on the $Z \to \tau\tau \to ll + 4\nu$ analysis, where the cross section is measured using the $e\mu$ and $\mu\mu$ sub-channels. The ee channel is not used for this measurement because the signal to background ratio is too low.

Multijet processes are significant backgrounds for both the $H \to \tau\tau \to ll + 4\nu$ and $Z \to \tau\tau \to ll + 4\nu$ analyses, due to its high production cross section. The contributions from this background are measured using data-driven methods in both analyses, where some assumptions are made about the composition and efficiencies of these events. To validate

these assumptions, the composition of the multijet background is measured using simulated multijet samples.

Chapter 1 describes the theoretical motivations behind existence of the Higgs and Chapter 2 describes the expected production mechanisms of the Higgs and its backgrounds at the LHC. This is followed by descriptions of the LHC accelerator complex and the ATLAS detector in Chapter 3. The simulated and data samples used in this thesis along with any applied corrections are described in Chapter 4. Following this, Chapter 5 details the study of the multijet composition. The $Z \rightarrow \tau\tau \rightarrow ll + 4\nu$ analysis is detailed in Chapter 6 and Chapter 7 briefly describes the $Z \rightarrow \tau\tau \rightarrow l\tau_h + 3\nu$ analyses and the combined $Z \rightarrow \tau\tau$ cross section measurement. The $H \rightarrow \tau\tau \rightarrow ll + 4\nu$ analysis is detailed in Chapter 8 and the combined results of all the $H \rightarrow \tau\tau$ search channels are presented in Chapter 9. Finally, the combined results from all search channels in ATLAS are presented in Chapter 10.

1

Theoretical motivation

Particle physics is a study of fundamental particles and their interactions. Particle interactions are described by quantum field theory (QFT), where particles are represented by quantised fields and their interactions are mediated by the exchange of spin-1 gauge fields. The Standard Model (SM) of particle physics is the best current theory for describing fundamental particles and their interactions, excluding gravity. This chapter provides an overview of the SM and will introduce the Higgs mechanism as a way of generating the masses for the weak gauge bosons.

1.1 Standard Model

The SM is a local gauge invariant QFT that combines the strong, electromagnetic and weak interactions into a single framework that's invariant under $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$. The $SU(3)_C$ group describes the strong interaction and the $SU(2)_L \otimes U(1)_Y$ group describes the electroweak interactions. The properties of the gauge bosons are given in Table 1.1.

Gauge boson	Interaction	Electric charge	Mass
g	Strong	0	0
γ	Electromagnetic	0	0
$W^{\pm}$	Weak	± 1	80.385 ± 0.015 GeV
Z	Weak	0	91.188 ± 0.002 GeV

Table 1.1: The properties of the SM gauge bosons [1].

Matter particles consists of twelve spin-$\frac{1}{2}$ fermions, which are uniquely distinguished by an internal quantum number, *flavour*. There are main two types of fermions and they are *quarks*, which interact with all the SM gauge fields, and *leptons*, which only interact with the electroweak fields. There are six flavours of quarks and leptons, which are each arranged into three sets of doublets or *generations*. These generations are arranged in ascending order based on the fermion masses. The basic properties of the quarks and leptons are given in Table 1.2.

Leptons			**Quarks**		
Flavour	Electric Charge	Mass	Flavour	Electric Charge	Mass
e	-1	0.511 MeV	u	+2/3	$2.3^{+0.7}_{-0.5}$ MeV
ν_e	0	< 0.3 eV	d	-1/3	$4.8^{+0.7}_{-0.3}$ MeV
μ	-1	106 MeV	c	+2/3	1.28 ± 0.03 GeV
ν_μ	0	< 0.3 eV	s	-1/3	0.95 ± 0.05 GeV
τ	-1	1.78 GeV	t	+2/3	174 ± 1 GeV
ν_τ	0	< 0.3 eV	b	-1/3	4.18 ± 0.03 GeV

Table 1.2: The properties of the SM quarks and leptons [1]. Uncertainties on the charged lepton masses are greatly smaller than the quoted values and are not shown.

1.1.1 The electromagnetic interaction

Quantum Electrodynamics (QED) is a local gauge invariant QFT that describes electromagnetic interactions. QED is invariant under $U(1)_Q$ and has the following Lagrangian:

$$\mathcal{L}_{QED} = \bar{\psi_q}(i\gamma^\mu \not{D}_{QED,\mu} - m_q)\psi_q - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{1.1}$$

where ψ_q and $\bar{\psi_q}$ are the fermion and anti-fermion fields, respectively, with electric charge q and mass m_q. $F_{\mu\nu}$ is the electromagnetic field tensor, which is given by:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \tag{1.2}$$

where A_μ is the photon gauge field introduced to maintain local U(1) gauge invariance.

The covariant derivative of QED that describes the fermion interactions with the photon mediator and is given by:

$$\not{D}_{QED,\mu} = \partial_\mu + iqA_\mu. \tag{1.3}$$

1.1.2 The electroweak interaction

In the SM, the electromagnetic and weak interactions are simultaneously described by the electroweak interaction. The electroweak interaction is invariant under $SU(2)_L \otimes U(1)_Y$ gauge transformations and has the following Lagrangian:

$$\mathcal{L}_{EW} = \sum_f \bar{\psi}_f i\gamma^\mu \not{D}_{EW,\mu} \psi_f - \frac{1}{4} W^a_{\mu\nu} W_a^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \tag{1.4}$$

where ψ_f represents the fermion fields. The fermion fields of the electroweak interaction are given by:

$$\ell^n_L = \begin{pmatrix} \nu^n_L \\ e^n_L \end{pmatrix}, e^n_R, \quad Q^n_L = \begin{pmatrix} u^n_L \\ d^n_L \end{pmatrix}, u^n_R, d^n_R, \tag{1.5}$$

where ℓ denotes the lepton fields with e and ν representing the charged lepton and neutrino fields, respectively, with generation n and chirality L or R. Similarly, Q represents the quark fields with u denoting the up-type quarks and d denoting the down-type quarks. Equation 1.5 contains no right-chiral neutrinos terms, ν_R, because right-chiral neutrinos have never been observed.

For the theory to remain invariant under $SU(2)_L \otimes U(1)_Y$ gauge transformations, the covariant derivative $\not{D}_{EW,\mu}$ requires the introduction of the B_μ and W^i_μ gauge fields:

$$\not{D}_{EW,\mu} = \partial_\mu + ig\frac{Y}{2} B_\mu + ig' \frac{\tau^i_L}{2} W^i_\mu, \tag{1.6}$$

where g and g' are the weak coupling constants, Y is the weak hypercharge and τ^i_L are the generators of SU(2). The weak hypercharge is defined to be $Y = 2(Q - I_3)$, where Q denotes the electric charge and I_3 is the third component of weak isospin. For the fermion fields that are doublets in $SU(2)_L$, the ν^n_L and u^n_L components have I_3 = +1/2 and e^n_L and d^n_L have I_3 = -1/2. For the remaining right-chiral fermion fields, which are singlets in $SU(2)_L$, I_3 = 0. The third component of weak isospin is a conserved quantity in all SM interactions.

The W^i_μ and B_μ fields are not the gauge boson fields of the electroweak interaction. Instead, the electroweak gauge boson fields, $W^\pm_\mu$, Z_μ and A_μ, are linear combinations of B_μ and W^i_μ:

$$W^\pm_\mu = \frac{1}{\sqrt{2}}(W^1_\mu \mp i W^2_\mu), \tag{1.7}$$

$$Z_\mu = cos(\theta_W) W^3_\mu - sin(\theta_W) B_\mu, \tag{1.8}$$

$$A_\mu = sin(\theta_W) W^3_\mu + cos(\theta_W) B_\mu, \tag{1.9}$$

where $\theta_W = cot(\frac{g'}{g})$ is the Weinberg angle which has been experimentally been measured to be $cos(\theta_w) = 0.88173 \pm 0.00032$ at Q = 91.2 GeV[1].

1.1.3 Higgs mechanism

The Lagrangian of the electroweak interaction describes interactions for massless gauge bosons. The weak gauge bosons are known to have mass so a complete theory must reflect this. The Higgs mechanism offers a way of generating mass terms for the weak gauge bosons in a way that preserves unitarity and renormalisation [2]. This is achieved by introducing a complex scalar field into the Lagrangian:

$$\mathcal{L}_{Higgs} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2, \tag{1.10}$$

where ϕ is a SU(2) doublet:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{1.11}$$

The last two terms of equation 1.10 is the potential of the Higgs field and is denoted by V. For V to be minimised requires

$$\frac{\partial V}{\partial \phi} = \phi(\mu^2 + \lambda \phi^2) = 0. \tag{1.12}$$

Equation 1.12 has two sets of solutions, which depends on μ^2. If we consider $\mu^2 < 0$, this has the solution

$$\phi^2 = \frac{-\mu^2}{\lambda} \equiv v^2, \tag{1.13}$$

where v is defined to be the value of ϕ at which the potential is minimised.

Equation 1.11 contains 4 real fields ϕ_1, ϕ_2, ϕ_3 and ϕ_4. The simplest choice is to assign $\phi_3 = v$ and $\phi_1 = \phi_2 = \phi_4 = 0$, which gives the vacuum, ϕ_0, a non-zero expectation value

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{1.14}$$

To obtain Higgs interaction terms, one must consider expanding around the minimum potential as follows:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \tag{1.15}$$

where $H(x)$ is the Higgs field, which can also be interpreted as excitations of the vacuum. By choosing $\phi_3 = v$, the symmetry of the potential is broken. This mechanism is known as *spontaneous symmetry breaking* and its consequences are described by the Goldstone theorem [3]. For any Lagrangian with global or local symmetries, spontaneous symmetry breaking will generate terms for a scalar particle with mass

$$m_H^2 = -2\mu^2 = \frac{2v^2}{\lambda}. \tag{1.16}$$

This massive scalar is the Higgs boson and its existence is a prediction of spontaneous symmetry breaking.

Spontaneous symmetry breaking can be applied to the Lagrangian of the electroweak interaction by replacing ∂_μ in equation 1.10 with $\not{D}_{EW,\mu}$ and ϕ with $\phi(x)$. The $SU(2)_L \otimes U(1)_Y$ symmetries of the electroweak interaction contains three local symmetries. For Lagrangians with local symmetries, spontaneous symmetry breaking generates one gauge boson mass term for each broken local symmetry. Therefore, three mass terms are generated through the electroweak symmetry breaking, which are required for the three weak gauge bosons. The mass terms of the weak gauge bosons can be written as a function of v and the weak coupling constants g and g':

$$m_{W^\pm} = \frac{1}{2} v g, \quad m_Z = \frac{v}{2}\sqrt{g^2 + g'^2}. \tag{1.17}$$

The scalar field ϕ contains four degrees of freedom. After symmetry breaking, three of those degrees of freedom became mass terms for the weak gauge bosons and the last remaining degree of freedom became the Higgs boson. From equation 1.16, the mass of the Higgs boson depends on v and λ, which are free parameters that can only be experimentally measured. Therefore to find the Higgs boson, one must search for it across a wide range of possible masses.

After symmetry breaking $SU(2)_L$ and $U(1)_Y$ are broken symmetries. However by construction, the $U(1)_Q$ symmetry is preserved and the photon remains massless. For this reason, the electromagnetic interaction can still be accurately described by QED. The Higgs couplings to the weak gauge fields are given by:

$$\begin{aligned}
\mathcal{L}_{HB} &= \frac{vg^2}{2} W_\mu^+ W^{-\,\mu} H + \frac{g^2}{4} W_\mu^+ W^{-\,\mu} H^2 + \frac{vg^2}{4cos^2\theta_W} Z_\mu Z^\mu H + \frac{g^2}{8cos^2\theta_W} Z_\mu Z^\mu H^2, \\
\mathcal{L}_{HB} &= g m_W W_\mu^+ W^{-\,\mu} H + \frac{g^2}{4} W_\mu^+ W^{-\,\mu} H^2 + \frac{g m_W}{2cos^2\theta_W} Z_\mu Z^\mu H + \frac{g^2}{8cos^2\theta_W} Z_\mu Z^\mu H^2,
\end{aligned} \tag{1.18}$$

where the first and third terms describe the three-point interactions between the weak gauge bosons and the Higgs field and the second and fourth terms describe the four-point interactions.

For the fermions to attain mass terms that are invariant under SU(2) transformations, the following Yukawa couplings are added to the Lagrangian:

$$\mathcal{L}_{Yukawa} = -g_e^{i,j} \bar{L_e^i} \phi e_R^j - g_d^{i,j} \bar{Q_L^i} \phi d_R^j - g_u^{i,j} \bar{Q_L^i} \phi u_R^j + \text{Hermitian conjugate}, \tag{1.19}$$

where i and j are the generation indices of the fermions and $g_{e,u,d}^{i,j}$ are 3×3 matrices containing the fermion coupling constants. Since no mass mixing has ever been observed in the charged lepton and quark sectors, $g_{e,u,d}$ are diagonal matrices.

The right-chiral neutrino does not exist in the SM Lagrangian and therefore neutrinos do not attain mass via Yukawa couplings to the Higgs field. In the SM, neutrinos are modelled as massless particles, since their masses are known to be small and are negligible in most calculations. The most recent experiments place an upper limit on the combined masses of the three neutrino flavours to be less than 0.3 eV [4].

By replacing ϕ equation 1.19 with $\phi(x)$, the following fermion terms are generated:

$$\begin{aligned}
\mathcal{L}_{Yukawa} &= \sum_i^3 \frac{g_e^i v}{\sqrt{2}} \bar{e_L^i} e_R^i + \frac{g_e^i}{\sqrt{2}} \bar{e_L^i} e_R^i H, \\
&+ \sum_j^3 \frac{g_u^j v}{\sqrt{2}} \bar{u_L^j} u_R^j + \frac{g_u^j}{\sqrt{2}} \bar{u_L^j} u_R^j H \\
&+ \sum_k^3 \frac{g_d^k v}{\sqrt{2}} \bar{d_L^k} d_R^k + \sum_k \frac{g_d^k}{\sqrt{2}} \bar{d_L^k} d_R^k H + \text{Hermitian conjugate}.
\end{aligned} \tag{1.20}$$

The first term of each line in equation 1.20 are the fermion masses and the second terms are the fermion couplings to the Higgs field. Equation 1.20 can be rewritten in terms of the fermion masses, m_f, to become:

$$\mathcal{L}_{Yukawa} = \sum_f m_f \bar{f}_L f_R + \frac{m_f}{v} \bar{f}_L f_R H, \tag{1.21}$$

$$\mathcal{L}_{Yukawa} = \sum_f m_f \bar{f}_L f_R + \frac{g m_f}{2 m_W} \bar{f}_L f_R H. \tag{1.22}$$

Equation 1.22 shows a linear relationship between the fermions' masses and their coupling strengths to the Higgs field. Therefore more massive fermions will have proportionally stronger couplings to the Higgs field. The coupling strengths of the Higgs directly affects its branching ratio, which is described in the next section.

1.2 Higgs decay modes

The branching ratios of the Higgs boson decays will depend on its mass and its relative coupling strengths to other particles. The coupling strengths of the Higgs are shown in equation 1.18 for the gauge bosons and in equation 1.22 for the fermions. A plot of the Higgs branching ratios as a function of its mass, m_H, is shown in Figure 1.1. Although the Higgs boson does not couple directly to gluons or photons, the $H \rightarrow gg$ and $H \rightarrow \gamma\gamma$ decay modes can occur through higher order loop diagrams.

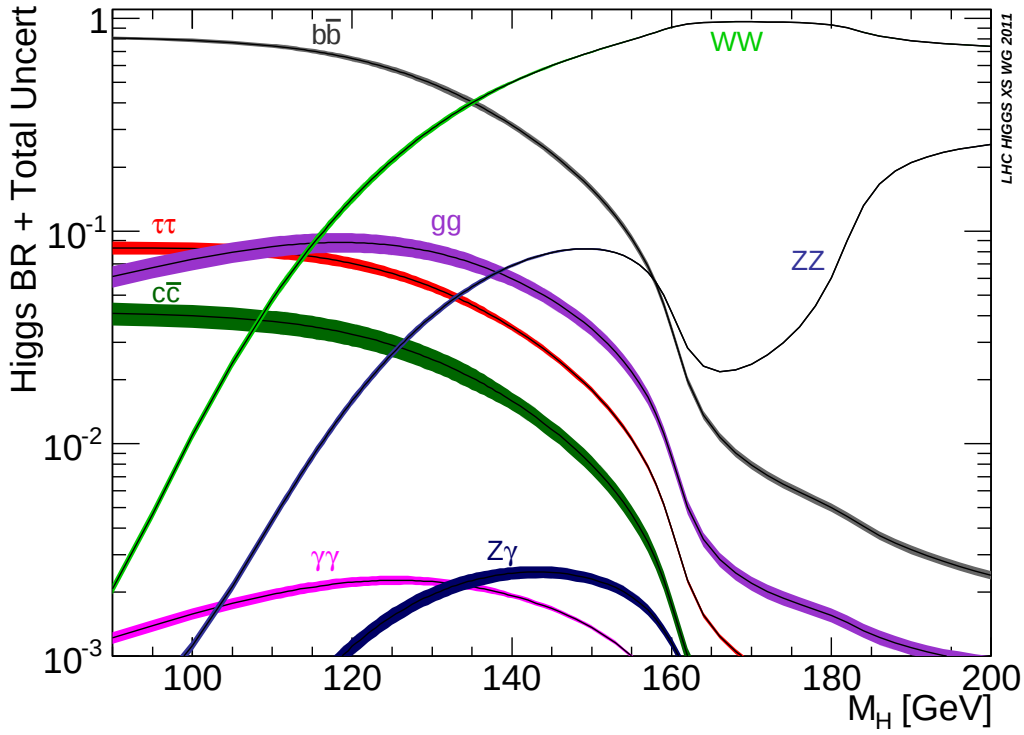


Figure 1.1: Branching ratios for the Higgs boson as a function of m_H [5].

1.3 Direct searches and best fit masses

Before the analyses that are presented in this thesis were conducted, there was no experimental evidence for the Higgs boson. Prior to the LHC, the experiments at LEP and Tevatron searched for the Higgs over a range of possible masses. These searches have provided limits on the Higgs masses, which are shown in Figure 1.2. This plot shows the masses that are excluded at 95% confidence [6] [7].

Precision electroweak measurements have favoured a Higgs with $m_H \lesssim 140$ GeV. Combining the precision measurements with direct searches from LEP and Tevatron, the most likely Higgs mass is found to be 120^{+12}_{-5} GeV, which is also shown in Figure 1.2. In this mass region, $H \to \tau\tau$, $H \to b\bar{b}$, $H \to \gamma\gamma$, $H \to WW$ and $H \to ZZ$ are all viable search channels. However, $H \to \tau\tau$ and $H \to b\bar{b}$ are the only fermionic search channels. Measurements of the fermionic decay modes are essential for testing the Higgs' Yukawa couplings in order to confirm any observed resonance as the Higgs boson.

The $H \to \tau\tau$ search channel has some advantages over $H \to b\bar{b}$. The $H \to b\bar{b}$ channel is dominated by the $b\bar{b}$ background. The $H \to \tau\tau$ channel has a cleaner signature and less backgrounds, so a better signal to noise ratio can be achieved.

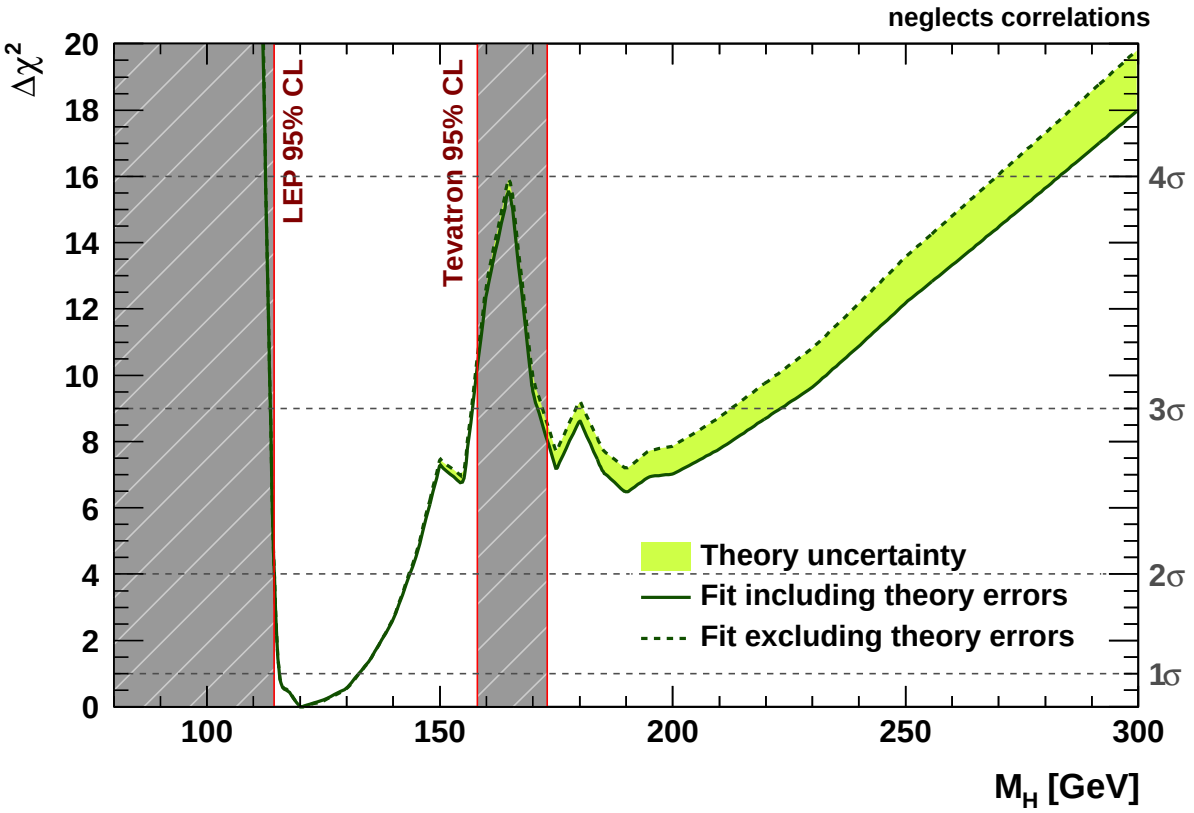


Figure 1.2: Indirect determination of the Higgs mass showing the $\Delta\chi^2$ of the fit as a function of m_H [8]. Excluded masses are represented by the hatched grey areas.

1.4 $H \rightarrow \tau\tau$ decay modes

Tauons are the heaviest charged leptons and can decay hadronically or leptonically. Leptonically decaying tauons will produce a tau-neutrino and either an electron with an anti-electron-neutrino or a muon with an anti-muon-neutrino. The branching fractions the tauon's leptonic decay modes are given Table 1.3.

For reasons already discussed, this thesis will focus on searching for the Higgs boson in the $H \rightarrow \tau\tau$ channel and more specifically, the sub-channels where both the tauons decay leptonically. Leptonically decaying tauons have cleaner signatures than their hadronic counterparts. So despite having lower branching ratios, the sensitivities of the di-leptonic sub-channels are comparable to the sub-channels with hadronically decaying tauons.

There are three distinct $H \rightarrow \tau\tau \rightarrow\ ll + 4\nu$ final states and these are referred to as the ee, $e\mu$ and $\mu\mu$ channels. All three channels are used for the $H \rightarrow \tau\tau \rightarrow\ ll + 4\nu$ analysis.

Decay mode	Branching ratio (%)
$\tau \rightarrow e\bar{\nu}_e\nu_\tau$	17.85 ± 0.05
$\tau \rightarrow \mu\bar{\nu}_\mu\nu_\tau$	17.36 ± 0.05
$\tau \rightarrow \tau_h\nu_\tau$	63.86 ± 0.15

Table 1.3: The leptonic decay modes of τ leptons and their branching ratios [1].

1.5 Summary

This chapter has described the importance of the Higgs boson to electroweak symmetry breaking as part of the SM. The theoretical interactions of the Higgs boson with other SM particles has also been described with particular focus on the $H \rightarrow \tau\tau$ decay modes. Chapter 2 will use the Lagrangians introduced in this chapter to describe the expected production mechanisms of the Higgs boson and its backgrounds at the LHC.

2

Higgs production and backgrounds

This chapter will describe Higgs boson production mechanisms at the LHC and the topology of the $H \to \tau\tau$ events. Backgrounds to the $H \to \tau\tau$ events are also described with particular focus on the irreducible $Z \to \tau\tau$ background.

2.1 Higgs production at the LHC

There are several different ways of producing the Higgs boson. This section will focus on the production mechanisms that are most relevant for proton-proton collisions at the LHC. (The proton structure function is further discussed in appendix A.)

At the LHC, there are three main Higgs production mechanisms: gluon-gluon fusion (ggH); vector boson fusion (VBF); and Higgsstrahlung (VH). Feynman diagrams for each of these processes are shown in figures 2.1, 2.2 and 2.3, respectively. The ggH production has the highest cross section followed by the VBF and the VH. The expected production cross sections are shown in Figure 2.4, which are calculated using the PDF4LHC prescriptions [9].

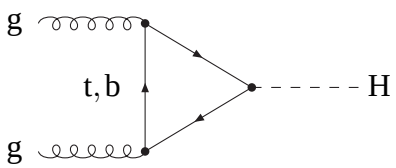


Figure 2.1: Feynman diagram of the gluon-gluon fusion Higgs production. Any fermion can be placed in the triangle loop, however leading contributions come from the heavier top and bottom quarks.

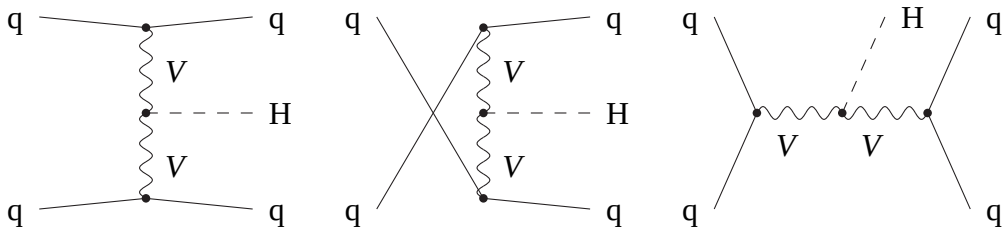


Figure 2.2: Leading order Feynman diagrams for the Higgs production via VBF. q denotes any quark or antiquark and *V* stands for W and Z boson.

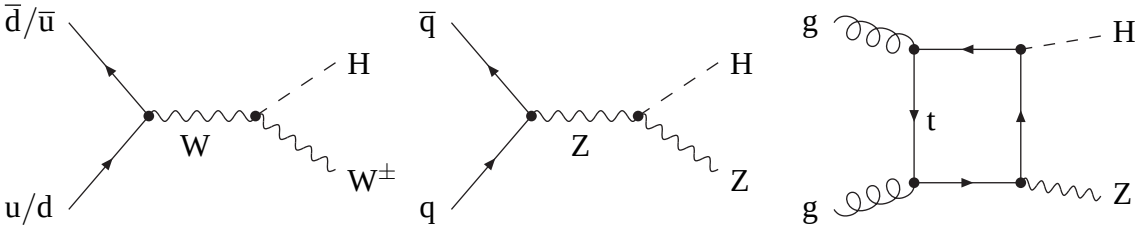


Figure 2.3: Leading order Feynman diagrams for the VH processes.

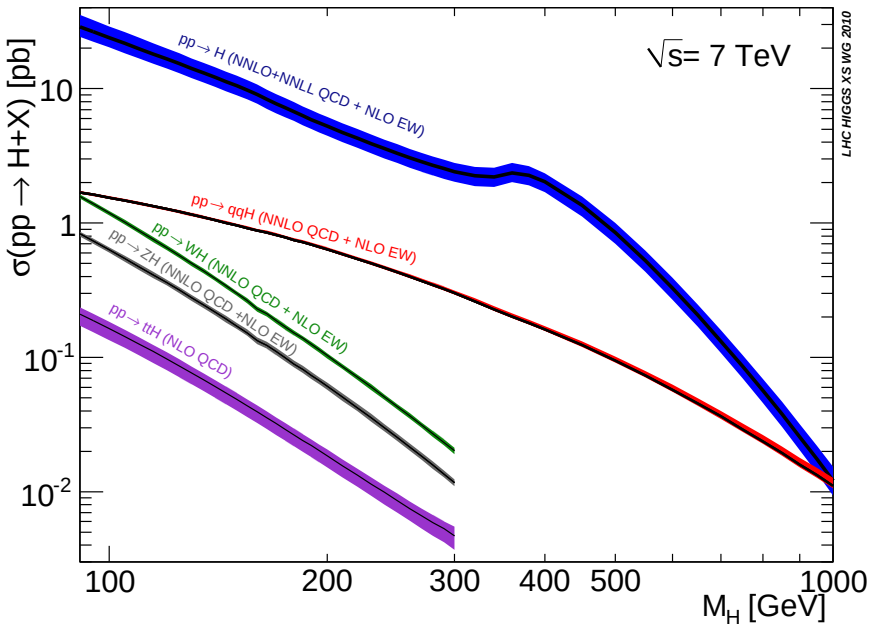


Figure 2.4: Expected production cross sections of the Higgs boson [9]. The gluon-gluon fusion process is shown in blue, the VBF is shown in red and the *WH* (*ZH*) Higgsstrahlung is shown in green (grey).

2.2 $H \rightarrow \tau\tau \rightarrow ll + 4\nu$ search topology

The $H \rightarrow \tau\tau \rightarrow ll + 4\nu$ events are characterised by two oppositely charged leptons with E_T^{miss} from the four neutrinos. The large mass difference between the Higgs and the tauon provides the daughter tauons with significant boosts. The charged leptons and neutrinos from the tauon decays are focused by this boost. If the Higgs is produced at rest in the transverse plane, then the tauons are mostly back-to-back in the ϕ-direction and the E_T^{miss} is soft. However, if the Higgs is produced with some radial momenta, then the tauons will be focused along this direction, producing higher p_T leptons and stronger E_T^{miss}.

Higgs bosons produced via the VBF mechanism are further characterised by the presence of two additional jets formed by the two spectator quarks. Since these quarks do not take part in the hard scattering process, they are likely to retain most of their initial longitudinal momenta. Jets formed by these quarks travel close to the beam line and are referred to as "forward" jets.

The VH production mechanisms creates a W or Z boson in conjunction with a Higgs. VH events with hadronically decaying bosons are additionally characterised by the presence of two jets with an invariant mass approximately equal to the masses of the vector bosons.

The gluon-gluon fusion Higgs production mechanism does not have any additional features that further distinguishes it from background processes. As such, the ggH events are expected to have a lot of background.

2.3 $Z \rightarrow \tau\tau$ background

The topology of the $Z \rightarrow \tau\tau \rightarrow ll + 4\nu$ events is extremely similar to that of the $H \rightarrow \tau\tau \rightarrow ll + 4\nu$ process. For m_H < 140 GeV, this background is irreducible. For VBF and VH events, the presence of forward jets and jets from a vector boson decay are strong discriminants against this background. However, for the ggH production, only a slight difference in the kinematics of the leptons and the amount of E_T^{miss} can be used as discriminants.

The main production mechanism for Z bosons at the LHC is via quark anti-quark annihilation. A Feynman diagram of this process is shown in Figure 2.5. Since the final state particles of the $Z \rightarrow \tau\tau \rightarrow ll + 4\nu$ and $H \rightarrow \tau\tau \rightarrow ll + 4\nu$ events are the identical, this background affects all dilepton channels equally.

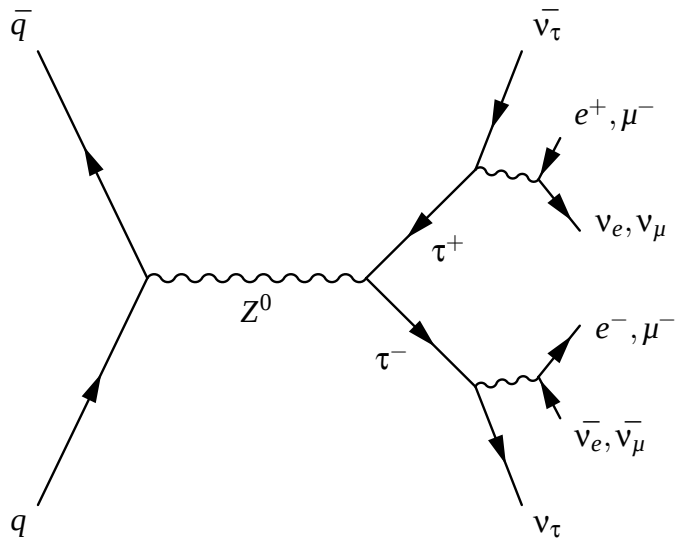


Figure 2.5: A feynman diagram of the $Z \rightarrow \tau\tau$ lepton decay modes [10].

2.4 Other backgrounds

All SM processes that can produce two oppositely charged leptons can be considered as backgrounds to the $H \rightarrow \tau\tau \rightarrow ll + 4\nu$ analysis. When measuring the cross section of the $Z \rightarrow \tau\tau \rightarrow ll + 4\nu$ process, these processes are also backgrounds to that analysis. The expected production cross sections of the signal and the main backgrounds are shown in Figure 2.6. This plot highlights the orders of magnitude differences in the production rates of the signal and various background processes. Brief descriptions for each of the major backgrounds are given below:

- **Multijets:** Multijets refer to physical processes that produce quarks and gluons via the strong interaction. These processes can produce real leptons from heavy flavoured quark decays or charged hadrons that fake leptons. Both mechanisms will produce events with two oppositely charged leptons. Lepton candidates from multijets are expected to have low selection efficiencies. However, the large production cross section of this background allows it to have significant contributions in all dilepton channels.

- $\gamma^*/\mathbf{Z} \rightarrow \mathbf{ee}, \mu\mu$**+ jets:** The leptonic decay modes of γ^*/Z processes produce two oppositely charged leptons of the same flavour. This process is expected to be the dominant background in the ee and $\mu\mu$ channels. Although the $e\mu$ channel is less affected, contributions from this background can occur if one of the leptons is not detected and additional jets fake a lepton.

- $\mathbf{W} \rightarrow \mathbf{e}\nu_\mathbf{e}, \mu\nu_\mu$ **+ jets:** Leptonically decaying W bosons that are produced with additional jets can mimic the signal if any of the additional jets fake a lepton. Since

jets are more likely to fake electrons than muons, contributions from this background are more prominent in the ee and $e\mu$ channels.

- **$W \rightarrow \tau\nu_\tau$ + jets:** Tauons produced from W boson decays can decay either leptonically or hadronically. Hadronically decaying tauons produce mainly charged kaons or pions, which can fake electrons or muons. Leptonically decaying tauons directly produce an electron or muon. For two oppositely charged leptons to be present, additional jets are required to produce a second lepton. Jets are more likely to fake electrons than muons, so this background is expected to affect the ee and $e\mu$ channels more so than the $\mu\mu$ channel. This background will also produce more E_T^{miss} than the $W \rightarrow e\nu_e$ and $W \rightarrow \mu\nu_\mu$ backgrounds.

- **$t\bar{t}$**: Top anti-top pairs decays can produce either zero, one or two leptons. The two lepton decay modes affects all three dilepton channels. Contributions from the zero and one lepton decay modes are more likely to affect the ee and $e\mu$ channels.

- **Dibosons (WW, WZ, ZZ):** W^+W^- pairs where both bosons decay leptonically produce two leptons of opposite charge. For the WZ and ZZ processes, there are various combinations of decay modes that produce two or more charged leptons. For the decay modes that produce more than two charged leptons, the signal can be mimicked if the additional leptons escape detection.

- **Single top:** Two oppositely charged leptons can be produced from single top events if the top quark decays leptonically and additional jets fake a lepton. Contributions from this background are found to be negligible for the $Z \rightarrow \tau\tau$ analysis. For the $H \rightarrow \tau\tau$ analysis, contributions from the single top t-channel, s-channel and Wt production mechanisms are considered.

2.5 Summary

This chapter has described the main production mechanisms and topologies of the $H \rightarrow \tau\tau$ process and its backgrounds. These will be become important later in this thesis when discussing the event selections of the $H \rightarrow \tau\tau \rightarrow ll + 4\nu$ and $Z \rightarrow \tau\tau \rightarrow ll + 4\nu$ analyses.

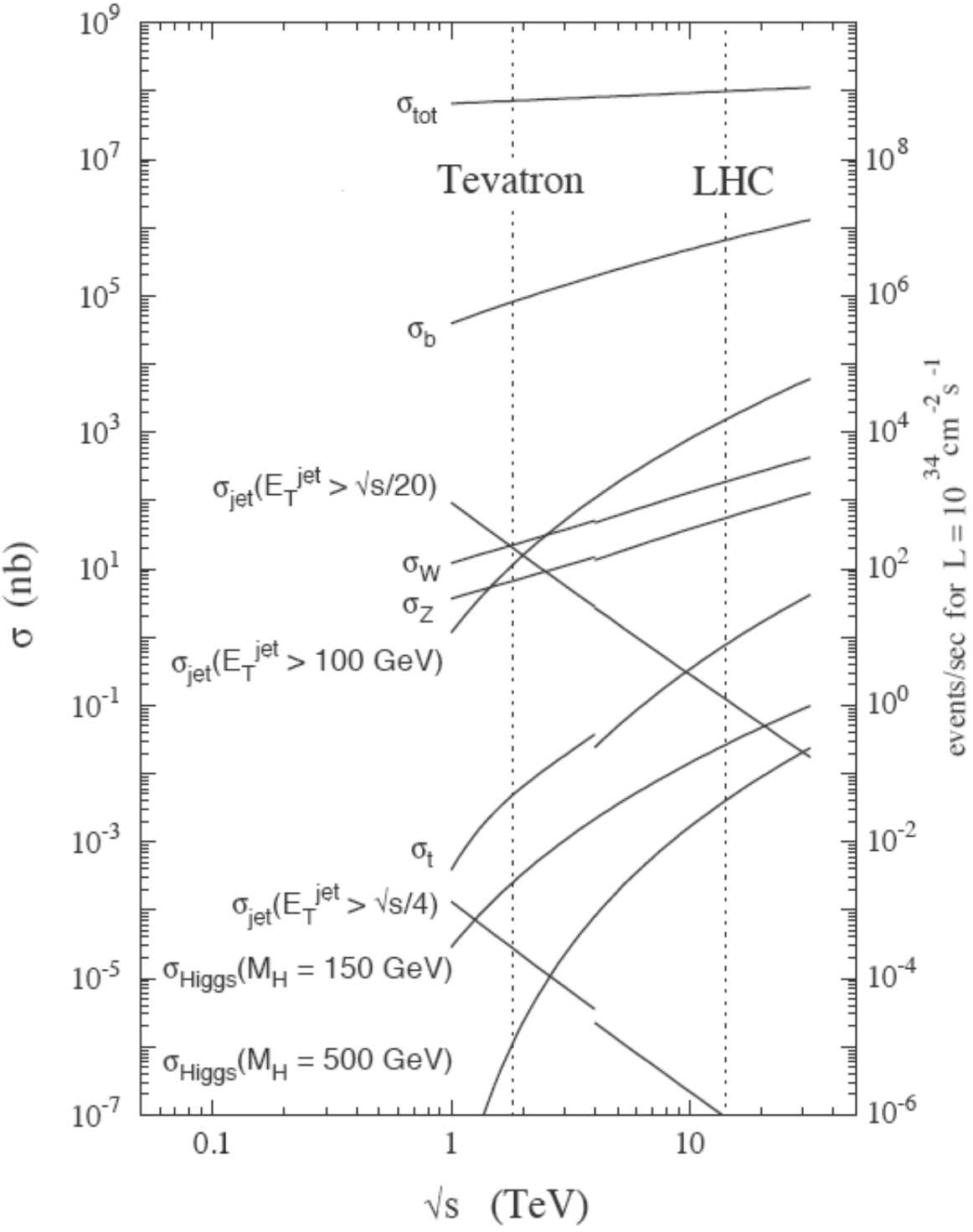


Figure 2.6: The cross sections of the main SM processes shown as a function of $\sqrt{s}$ [11].

3

The ATLAS Experiment

The ATLAS experiment is one of four major experiments on the Large Hadron Collider (LHC) ring. It is operated by a collaboration of approximately 3000 scientists from 164 institutions and 38 countries. The LHC is located at CERN, Switzerland. The main goal of the LHC experiments is to search for the Higgs boson and find new physics beyond the SM. This chapter will provide an overview of the LHC accelerator, the ATLAS detector and the software used by the ATLAS collaboration for particle reconstruction.

3.1 The Large Hadron Collider

The LHC is currently the world's most energetic particle collider. It is designed to accelerate two separate beams of protons to energies of up to 7 TeV per beam. In 2010 and 2011, the LHC operated at 3.5 TeV per beam with a peak luminosity of $3.9 \times 10^{33} cm^{-2} s^{-1}$ [12] [13]. In 2012, the operating energies were higher at 4.0 TeV per beam with a peak luminosity of $7.7 \times 10^{33} cm^{-2} s^{-1}$.

There are four interaction points along the LHC ring where collisions can occur. At each of these points there is a detector that utilises these collisions for physics studies, as shown in Figure 3.1. These four main LHC experiments are:

- ATLAS and CMS, which are general purpose detectors [14] [15],
- ALICE, which is designed to study the physics of heavy ion collisions [16] and
- LHCb, which is designed for studying heavy flavour physics through precision measurements of CP violation and rare decays [17].

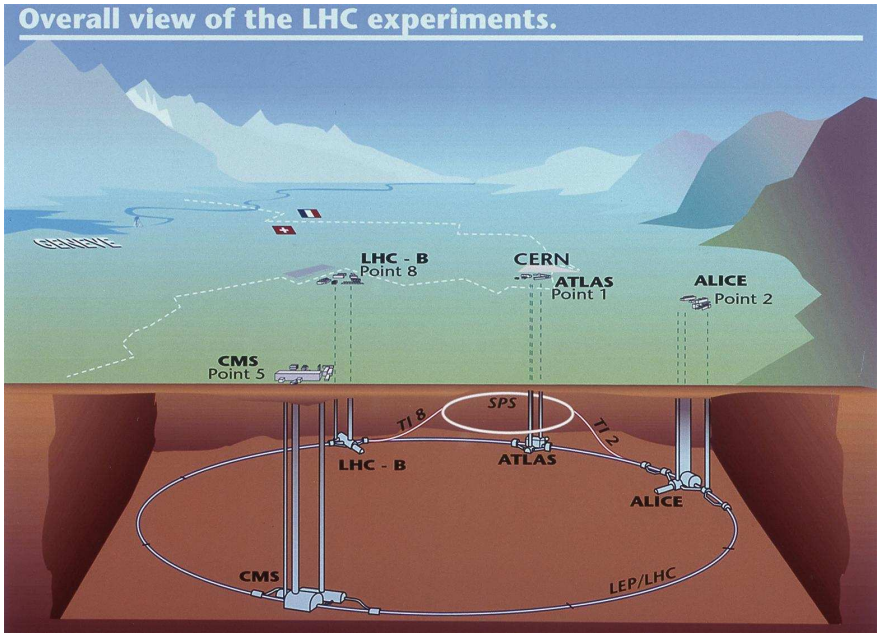


Figure 3.1: Layout of the four main experiments on the LHC ring [18].

The LHC uses a multistage acceleration process to boost protons to the desired collision energies. The process begins by ionizing hydrogen atoms to produce protons and accelerating them to a momentum of 50 MeV using a linear accelerator, LINAC2. They are then passed into the Proton Synchrotron Booster where they are accelerated to a momentum of 1.4 GeV. Following this, they are then accelerated through the Proton Synchrotron where they leave with a momentum of 25 GeV and go into the Super Proton Synchrotron where they are combined into bunches and accelerated to 450 GeV. The protons enter the main LHC ring from the SPS via one of the two 2.5 km injection tunnels, depending on the desired fill direction. Once there, the protons are accelerated to their final desired collision energies using the main LHC magnets. A detailed schematic of the LHC acceleration processes is shown in Figure 3.2.

There are 1232 superconducting dipole magnets along the LHC ring for bending and accelerating the beam. The coils of these magnets are made of niobium-titanium (NbTi), which is superconducting at a temperature of 1.9 K. The dipole magnets are "twin bore" in design, which allows them to accelerate two separate beams in opposite directions. For the twin bore magnet system, the energies of the two beams will always be entwined. In addition to the dipole magnets, the LHC ring also contains 392 quadrupole magnets for focusing and correcting the beam as well as many sextupole, octupole, decapole and dodecapole magnets for higher-order beam corrections.

A key aspect of the LHC is the ability to deliver a high luminosity for physics studies. The luminosity of a collider is the measure of the number of interactions it can provide per unit time. It is related to the number of generated events as follows:

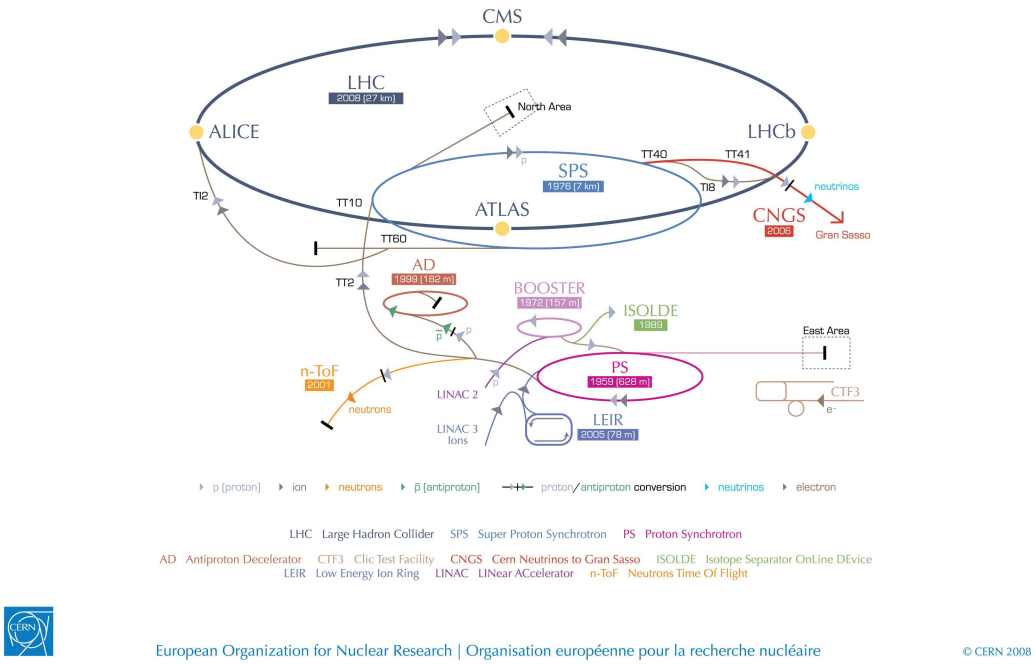


Figure 3.2: The LHC ring complex showing the chain of particle accelerators used to boost the particles to the required energies [19].

$$N_{event} = L\sigma_{event}, \tag{3.1}$$

where σ_{event} is the cross section of the given process measured in units of area and L is the machine luminosity measured in units of inverse area and inverse time. For a Gaussian beam the luminosity can be approximately be given by:

$$L = \frac{N_p^2 n_b f_{rev} \gamma_r}{4\pi\varepsilon_n \beta^*} F, \tag{3.2}$$

where N_p is the number of particles per bunch, n_b is the number of bunches, f_{rev} is the revolution frequency of the particles around the ring, γ_r is the relativistic gamma factor, ε_n is the normalized transverse beam emittance, β^* is the beta function at the collision point and F is the geometric luminosity reduction factor due to the crossing angle at the interaction point (IP):

$$F = \left(1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)^2\right)^{-1/2}, \tag{3.3}$$

where θ_c is the full crossing angle at the IP, σ_z is the RMS of the bunch length and σ^* is the transverse RMS of the beam size at the IP.

At design, a maximum of 2808 bunches circulates the main ring with collisions occurring every 25 ns. Each bunch contains approximately 1.15×10^{11} protons, which gives an expected number of 19 interactions per bunch crossing. The majority of these interactions are proton-proton scatterings that don't create particles with high p_T. These soft interactions, although uninteresting, do affect detector responses. Interactions that occur within the bunch collisions on top of the hard scattering process is known as the *underlying event*. Detector responses are also affected by previous bunch crossings since most detector elements have readout latencies that are higher than 25 ns. The combination of multiple interactions and readout latencies is referred to as *pileup*. The effects of pileup are significant on physics analyses and must be correctly accounted for to accurately describe the observed interactions. A summary of the LHC design specifications is shown in Table 3.1.

LHC machine parameters	
Circumference	26.7 km
Dipole magnetic field	8.33 T
Dipole magnet temperature	1.9 K
Quadrupole magnetic field	6.85 T
Quadrupole magnet temperature	1.9 K
Number of main dipole magnets	1232
Number of main quadrupole magnets	392
Number of other magnets	7970
Proton beam parameters	
Proton energy	7 TeV
Number of protons per bunch	1.15×10^{11}
Number of bunches	2808
Time between bunch crossings	25 ns
Circulating beam current	0.582 A
Stored energy per beam	362 MJ
Peak luminosity parameters	
RMS bunch length (σ_z)	7.55 cm
RMS bunch size (σ^*)	16.7 μm
Beta function (β^*)	0.55 m
Relativistic gamma factor (γ_r)	7461
Normalized transverse emittance ε_n	3.75 μm
Full crossing angle (θ_c)	285 μrad
Geometric luminosity reduction factor (F)	0.836
Peak luminosity	$10^{34} cm^{-2} s^{-1}$
Interaction parameters	
Inelastic cross section	60 mb
Total cross section	100 mb
Average interactions per bunch crossing	19.02

Table 3.1: Summary of the main LHC specifications for proton-proton collisions running at design energy and luminosity at the ATLAS and CMS IPs [12]

3.2 The ATLAS detector

ATLAS is a general purpose detector designed to accommodate a broad scope of physics studies using collisions provided by the LHC. To accomplish this task, ATLAS was designed to have the following features [14]:

- High pseudorapidity and full azimuthal coverage. The physics processes of interest have low production cross sections so a large detector acceptance maximises the number of reconstructed events. A full azimuthal coverage is also essential for measuring missing transverse energy, which is a key component in many SM and beyond the SM analyses.
- Detector elements must have fast readouts and be radiation-hard to cope with the high collision rates of the LHC. A high detector granularity is also required to handle the high particle fluxes and to reduce the effects of pileup.
- An inner detector tracking system with good momentum resolution and reconstruction efficiency. Tracking elements must also be present close the interaction point to reconstruct secondary vertices from τ-leptons and b-jets.
- Good electromagnetic calorimetry for electron and photon identification and energy measurements. Electron and photon channels have clean signatures and are expected to play key roles in SM analyses and searches for new physics.
- Good hadronic calorimetry for accurate reconstructions of jet and missing transverse energy. Jets are ever-present in the LHC collision environment and must be well reconstructed for any physics analysis.
- Good muon identification, charge determination and momentum resolution over a wide range of momenta. Muons have the cleanest signatures amongst all the common final state particles and are crucial components in physics analyses.
- An efficient triggering on low p_T objects with high background rejection rates to retain the maximum number of interesting physics events.

The key performance aspects of the ATLAS detector are quantified and summarised in Table 3.2 and a layout of the detector is shown in Figure 3.3. This chapter provides a brief overview on the main sub-systems of the ATLAS detector, which have been designed to fulfil the described performance requirements. A more comprehensive description of the detector can be found in references [14] and [20] and the expected performance of the detector can be found in reference [21].

Detector component	Required resolution	η coverage	
		Measurement	Trigger
Tracking	$\sigma_{P_T}/P_T = 0.05\% \oplus 1\%$	$\pm$ 2.5	-
EM calorimetry	$\sigma_E/E = 10\% / \sqrt{E} \oplus 0.7\%$	$\pm$ 3.2	$\pm$ 2.5
Hadronic calorimetry			
barrel and endcap	$\sigma_E/E = 50\ \% / \sqrt{E} \oplus 3\%$	$\pm$3.2	$\pm$ 3.2
forward	$\sigma_E/E = 100\ \% / \sqrt{E} \oplus 10\%$	3.1 < |η| < 4.9	3.1 < |η| < 4.9
Muon spectrometer	$\sigma_{P_T}/P_T = 10\ \%$ at $p_T = 1$ TeV	$\pm$ 2.7	$\pm$ 2.4

Table 3.2: The general performance goals of the ATLAS detector. For high-p_T muons, the muon spectrometer performance is independent of the tracking system.

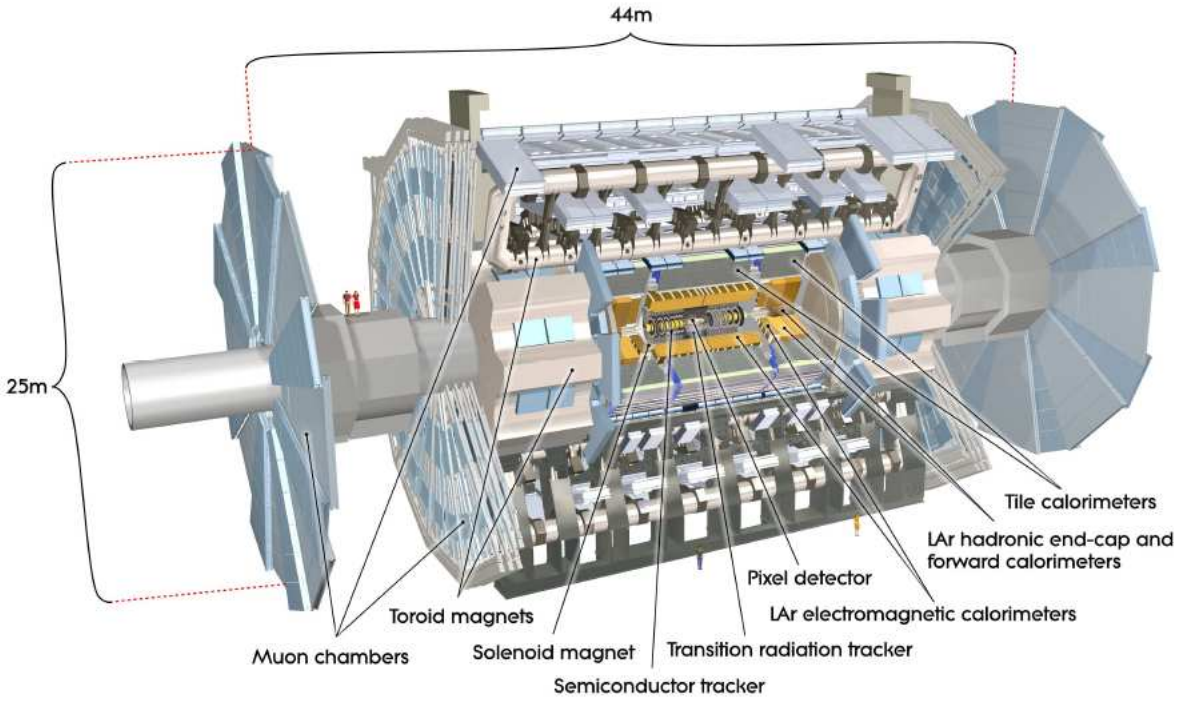


Figure 3.3: A cut away view of the ATLAS detector.

3.2.1 Coordinate system

The ATLAS collaboration uses a standard right-handed Cartesian coordinate system to describe the detector and the particles created during collisions. The z-axis points along the beam line with the positive direction anti-clockwise along the LHC ring when viewed from above. The x and y axes are perpendicular to the beam line with the positive y-direction pointing upwards and the positive x-direction pointing towards the centre of the main ring. The ATLAS detector is divided into two sides, which is split directly down the middle of the detector across the z-axis. The side with $z > 0$ is called *Side A* and the side with $z < 0$ is called *Side C*.

It is convenient to describe the ATLAS detector in terms of a cylindrical coordinate system (R, ϕ, z), where R is the transverse distance from the beam line and $\phi \in [0, 2\pi)$ is the azimuthal angle with 0 starting at the x-axis. Spherical coordinates (R, ϕ, θ) can also be used with $\theta \in [0, \pi)$ being the polar angle with 0 starting at the beam line.

At hadron colliders, particle vectors are often described by the parameters (p_T, ϕ, η), where p_T is the momentum in the transverse direction and η is the pseudorapidity defined by $\eta = -ln[tan(\theta/2)]$. Pseudorapidity is often used in collider experiments because the cross sections of physical processes are approximately constant in η. It is related to rapidity, which is given by $y = \frac{1}{2} ln[(E + p_z)/(E - p_z)]$.

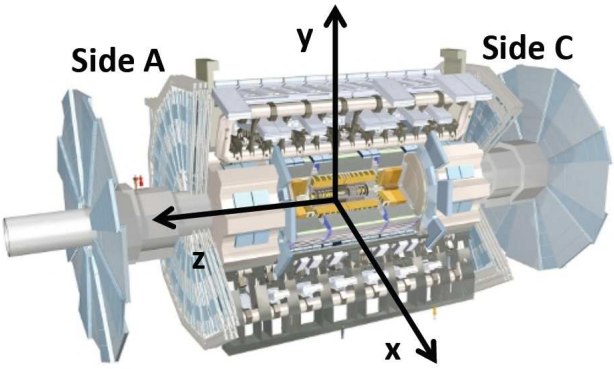


Figure 3.4: The Cartesian coordinate system used by the ATLAS collaboration.

3.2.2 Magnet system

The ATLAS detector uses a superconducting magnet system that consists of a standard central solenoid surrounded by a three air-core toroidal system. Two toroids are in the end-caps and one is in the barrel. The magnetic field of the inner detector is supplied by the central solenoid, while the toroid system provides the magnetic field for the muon spectrometer. The layout of the magnet system is depicted in Figure 3.5.

The central solenoid is positioned just in front of the electromagnetic calorimeter and provides a 2 T magnetic field. The toroid magnets have peak magnetic fields of 3.9 and 4.1 T for the barrel and end-cap toroids respectively. Each toroid consists of eight coils placed radially symmetric around the beam line.

The superconducting magnets of the ATLAS detector are made of NbTi. To achieve superconductivity, the magnets need to be cooled to a temperature of 4.5 K. The main cooling is provided by flowing 4.5 K helium gas through tubes welded onto the casings of the coil windings. To provide additional cooling, the central solenoid magnet is also coupled to a refrigerator while the toroid magnets have extra helium pumps.

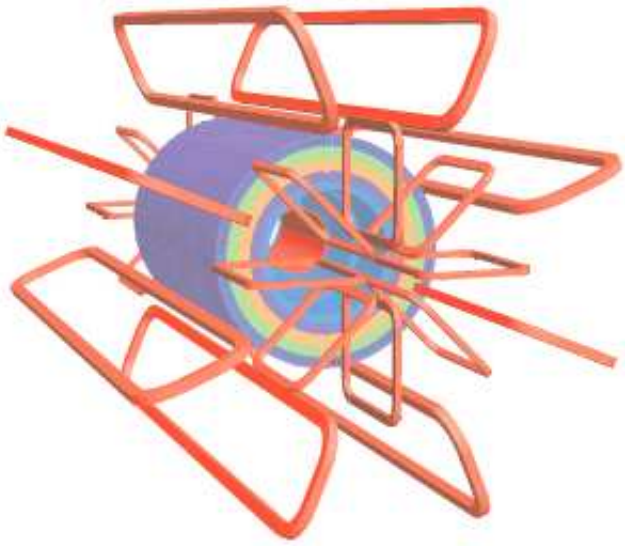


Figure 3.5: Geometry of the toroid magnets shown in red surrounding the central solenoid shown in blue. The coloured layers represent the calorimeter layers that enclose the central solenoid [14].

3.2.3 Inner detector

The inner detector is the closest system to the interaction point and provides momentum, direction, vertex and charge measurements for particles within the pseudorapidity range $|\eta| < 2.5$. It is contained within a cylinder of length 5.5 m and radius 1.15 m and is immersed in an almost homogeneous 2 T magnetic field that points parallel to the beam line. During LHC operations, approximately 1000 particles will emerge from the interaction point within the coverage of the inner detector per proton-proton collision. To handle such high particle densities and to meet the performance goals of ATLAS, the inner detector has been designed to be radiation-hard and attain high-resolution measurements with fine detector granularity. There are three sub-detector systems for the inner detector and they are: the pixel detectors, the Silicon Microstrip (SCT) detectors and the Transition Radiation Tracker (TRT). A diagram of the inner detector is shown in Figure 3.6.

The pixel detectors are located closest to the interaction point and use semiconducting silicon detectors. Each module provides a 2-dimensional read out tangential to the plane of the module. There are 1744 identical pixel modules in the inner detector providing an overall coverage of up to $|\eta| < 2.5$. Track density is higher for the pixel detectors than for any other system and as such, they have been designed to be the most intrinsically accurate system with a resolution of $10\mu m$ in the azimuthal direction and $115\mu m$ in the z-direction.

The SCT detectors lie between the pixel detectors and the TRT and use silicon strips similar to the pixel detectors. The 15912 SCT modules of the inner detector spread across the

barrel and endcaps provides tracking coverage in the range $|\eta| < 2.5$. Unlike the pixel modules, each SCT wafer can only provide a 1-dimensional position read out. To achieve a 2-dimensional position measurement two wafers are glued back-to-back at a relative angle of 40 mrad. The SCT modules have an intrinsic accuracy of $17\mu m$ in the azimuthal direction and $580\mu m$ in the z direction.

The TRT is the furthest inner detector sub-system from the interaction point and is positioned in front of the calorimetry system. It covers the range $|\eta| < 2.2$ and uses drift (straw) tube technology to detect charged particles. Each straw is 4 mm in diameter and is filled with a gas mixture of 70% Xe, 27% CO_2 and 3% O_2. At the centre of each straw there is a $31\mu m$ diameter tungsten wire coated with gold, which acts as the anode. The straw walls are coated with a thin layer of Al, which acts as the cathode. Charged particles traversing through the TRT straws will ionize the gas contained within them and these ions will drift towards the central wire. The TRT can achieve an accuracy of 30 μm. One of the main purposes of this technology is to provide discrimination between electrons and pions. Transition radiation is emitted by charged particles when they cross the boundaries between media with different dielectric constants. Transition radiation photons produced from highly relativistic particles have frequencies in the X-ray bands. These photons further ionises the TRT gas mixture, which are measured as high threshold hits. The intensity of the transition radiation is directly proportional to the particle's Lorentz factor, so electrons produce more high threshold hits than pions.

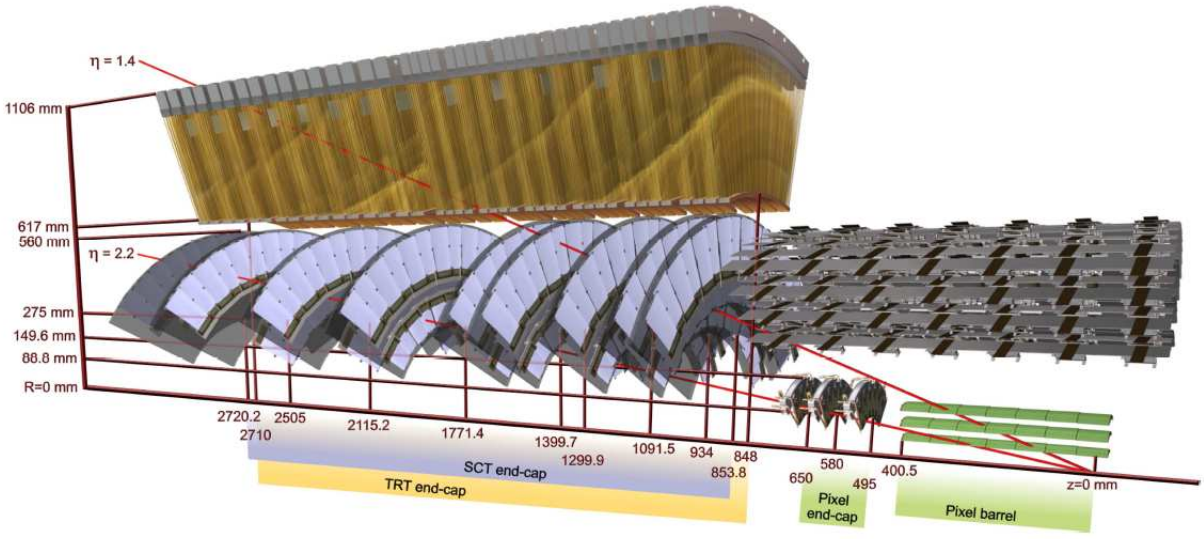


Figure 3.6: An illustration of the ATLAS Inner detector [14].

3.2.4 Calorimetry

The ATLAS calorimetry system surrounds the inner detector and provides energy measurements for electromagnetic and hadronic particles. The calorimeters consist of a number of sampling layers with full ϕ coverage around the beam axis and a pseudorapidity coverage of up to $|\eta| < 4.9$. There are three cryostats that house the calorimeters, one for the barrel and two for the end-caps. The barrel cryostat contains the electromagnetic barrel calorimeter while the end-cap cryostats contain the electromagnetic end-cap calorimeters, the hadronic end-cap calorimeters and the forward calorimeters. These calorimeters all use liquid argon (LAr) as the active detector medium and thus require cryostats for active cooling. The hadronic calorimetry is provided in the barrel and extended barrel regions by the Tile calorimeter, which uses polystyrene scintillator tiles that operate at room temperature. A cut-away view of the ATLAS calorimetry system is shown in Figure 3.7.

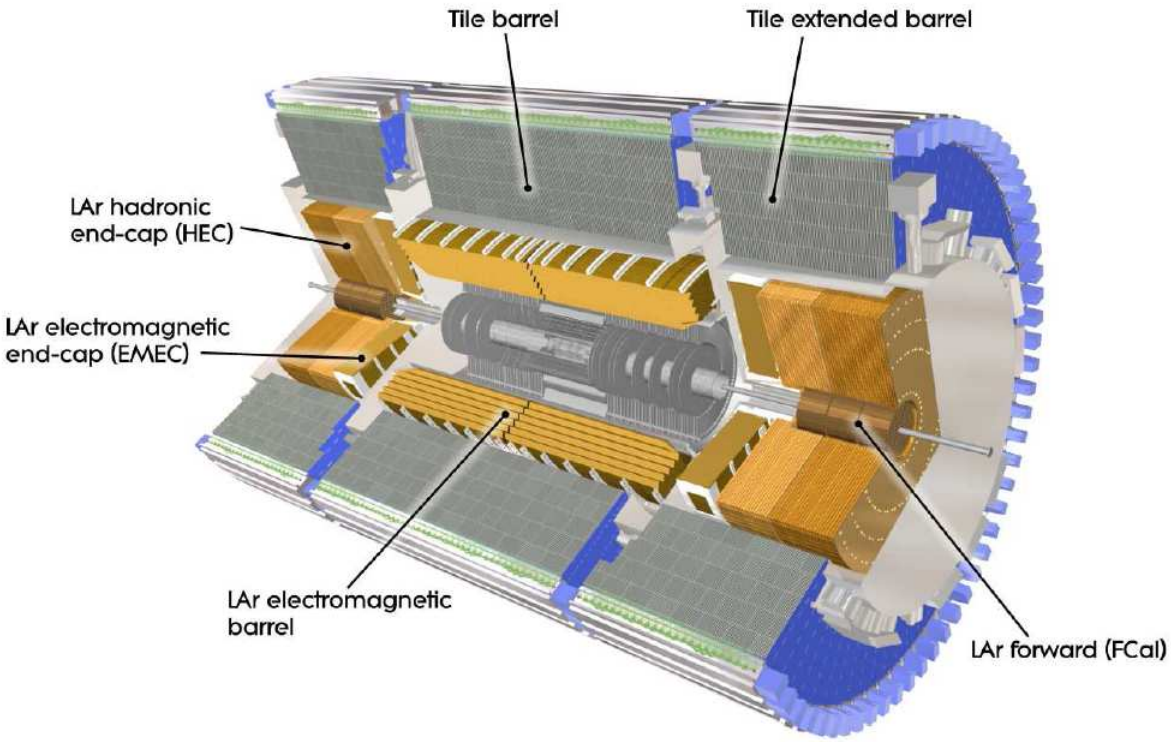


Figure 3.7: A cut-away view of the ATLAS calorimetry system [14].

The electromagnetic calorimeter uses lead in combination with liquid argon (LAr) to maximise electromagnetic showering. The barrel section covers the range $|\eta| < 1.475$ and is divided into two identical half-barrels separated by a small gap of 4 mm at z = 0. The end-cap sections are mechanically divided into two coaxial wheels with the outer wheels covering $1.375 < |\eta| < 2.5$ and the inner wheel covering $2.5 < |\eta| < 3.2$. The detector pieces are arranged in an accordion geometry to ensure complete azimuthal coverage. The thickness

of the lead plates varies as a function of η and has been designed for optimal performance in energy resolution. The EM calorimeter is segmented into multiple layers, three for the region $|\eta| < 2.5$ and two for the higher-η regions, $2.5 < |\eta| < 3.2$. In the region $|\eta| < 1.8$, an additional presampler detector is placed in front of the first EM calorimetry layer, which is used to correct for energy losses caused by material interactions. The presampler provides an extra sampling and consists of an active LAr layer of thickness 1.1 cm in the barrel and 0.5 cm thickness in the endcaps. The total depth of the EM calorimeter is approximately 22 to 30 radiation lengths (X_0) in the barrel and 24 to 33 X_0 in the endcaps, which is sufficient to contain the majority of EM showers at ATLAS.

The hadronic calorimetry system consists of three different detector types: the Tile calorimeter, the LAr hadronic endcap calorimeter and the LAr forward calorimeter. The Tile calorimeter is placed directly behind the barrel LAr calorimeter and covers the pseudorapidity rangess $|\eta| < 1.7$. The LAr hadronic endcap calorimeter is positioned directly behind the EM endcap calorimeter and covers the pseudorapidity range $1.5 < |\eta| < 3.2$. The higher-η regions, $3.1 < |\eta| < 4.9$, are covered by the LAr forward calorimeter, which surround the beam pipe starting from approximately 4.7 m away the interaction point. The LAr hadronic endcap calorimeter slightly overlaps with both the Tile and forward calorimeters to ensure minimal leakage in the transition regions.

The Tile calorimeter uses steel as the absorber material and is divided azimuthally into 64 modules. They are segmented into three layers with approximately 1.5, 4.1 and 1.8 interaction lengths (λ) for the barrel and 1.5, 2.6 and 3.3 λ for the extended barrel. The total thickness of the entire detector at the outer edge of the tile calorimetry system is approximately 9.7 λ at $\eta = 0$.

The LAr hadronic endcap calorimeter consists of two independent wheels per endcap, each assembled from 32 identical wedge-shaped modules. Each wheel is segmented into two layers with the first layer containing 24 parallel copper plates, each 25 mm thick, and the second layer containing 16 copper plates, each 50 mm thick. The total thickness of the LAr hadronic endcap calorimeter is approximately 14 to 18 λ.

The LAr forward calorimeter consists of three 45 cm thick modules. The first module is used to measure electromagnetic particles and mainly uses copper as the absorber material. The second and third modules are used to measure hadronic particles and mainly uses tungsten as the absorber material. The interaction lengths of each module in order from the first to the third are 2.66 λ, 3.68 λ and 3.60 λ.

3.2.5 Muon spectrometer

The muon spectrometer is the furthest sub-detector from the interaction point and encompasses the calorimetry system. Muons interact weakly with matter and can pass through the detector systems with minimal energy losses. The muon spectrometer provides measurements of muon tracks, which is used to determine their charge and momentum. There are four main complementary sub-detectors that comprise the muon spectrometer: the Monitored Drift Tubes (MDTs), the Cathode Strip Chambers (CSCs), the Resistive Plate Chambers (RPCs) and the Thin Gap Chambers (TGCs). A cut-away diagram of the ATLAS muon spectrometer system is shown in Figure 3.8.

The MDTs provide precision tracking measurements with coverages of up to $|\eta| < 2.0$ for the inner most layer and $|\eta| < 2.7$ for the remaining layers. This sub-detector uses drift tube technology, which consists of bundles of pressurised tubes filled with a gas mixture of 93% Ar and 7% CO_2 at a pressure of 3 bar. Each tube is 29.970 mm in diameter and contains a central tungsten-rhenium wire with a width of 50 μm. The tubes are made of aluminium which acts as the cathode while the central wire acts as the anode with a wire potential of 3080 V. The maximal drift time from the wall to the wire is about 700 ns. The MDTs are aligned along the ϕ direction and have an intrinsic resolution of approximately 35 μm in the z-direction.

The RPCs' primary function is to provide triggering for the barrel regions within $|\eta| < 1.05$. It also provides the azimuthal coordinate measurement for tracks within its coverage. The RPC detectors consist of three layers. The first two layers, placed relatively close together, are designed to provide triggering for tracks with p_T between 6 and 9 GeV. The third layer

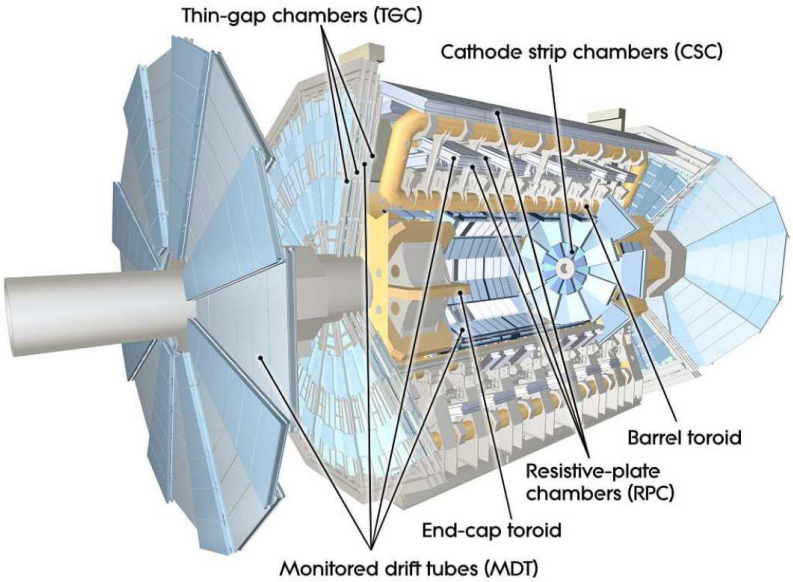


Figure 3.8: A cut away view of the ATLAS muon spectrometer [14].

is placed much further away and provides triggering for tracks with p_T greater than 9 GeV. The RPC uses gaseous parallel electrode-plate detectors that consist of two resistive plates made of phenolic-melaminic plastic laminate, which are separated by 2 mm using insulating spacers. The gas mixture in between the plates consists of 94.7% $C_2H_2F_4$, 5.0% Iso-C_4H_{10} and 0.3% SF_6. The chosen composition allows an operating voltage of around 4.9 kV/mm between the plates. The signal is read out through metallic strips placed on the outer faces of the plates. Each strip is approximately 25-35 mm wide and is separated from neighbouring strips by 2 mm with a 0.3 mm ground strip in the centre for improved readout decoupling. To provide measurements in ϕ and η, each RPC layer consists of two perpendicularly aligned RPC detectors. The RPC has a resolution of approximately 10 mm in the z and ϕ directions.

The CSCs provide precision tracking for the end cap regions within the range $2.0 < |\eta| < 2.7$. They are multi-wire chamber detectors with their anode wires orientated in the radial direction. The two cathodes either side of the wires are both segmented into strips. One set of strips is orientated perpendicular to the wires and the other parallel to them. This provides 2-dimensional measurements in ϕ and η. There are two disks of CSC detectors, one in each endcap. These disks are segmented equally into eight large chambers overlapped with eight small chambers. Each chamber contains four layers of CSC detectors resulting in four independent measurements of η and ϕ along each track. The CSC has an intrinsic resolution of approximately 40 μm in R and 5 mm in ϕ.

The TGCs provide triggering and azimuthal coordinate measurements in the regions $1.05 < |\eta| < 2.4$. They are multi-wire chamber detectors that have a wire-to-cathode distance of 1.4 mm and a wire-to-wire distance of 1.8 mm. A gas mixture of CO_2 and n-C_5H_{12} (n-pentance) is used for these detectors with a wire potential of approximately 2.9 kV. The cathodes of the TGC modules are made of graphite with one side grounded and the other side connected to readout strips for measurements in ϕ. The anode wires provide the measurements in η. The high voltage and small wire-to-wire distances have been chosen to achieve quick time resolutions, which are essential for triggering purposes. The TGC has an intrinsic resolution of approximately 2-6 mm in R and 3-7 mm in ϕ. A summary of the muon spectrometer sub-detector parameters is given in Table 3.3.

		Chamber resolution (RMS)			**Measurements/track**		**Number of**	
Type	**Function**	z/R	ϕ	time	barrel	end-cap	chambers	channels
MDT	tracking	35 μm (z)	-	-	20	20	1150	354k
CSC	tracking	40 μm (R)	5 mm	7 ns	-	4	32	30.7k
RPC	trigger	10 mm (z)	10 mm	1.5 ns	6	-	606	373k
TGC	trigger	2-6 mm (R)	3-7 mm	4 ns	-	9	3588	318k

Table 3.3: Parameters of the four sub-systems of the muon spectrometer [14]. Resolutions do not include chamber-alignment uncertainties.

3.3 Particle reconstruction and identification

The subsystems of the ATLAS detector each play key roles in reconstructing and identifying the particles created by the LHC collisions. The ATLAS detector is capable of reconstructing and identifying electrons, photons, muons, hadronically decaying tauons and jets, which originate from hadrons, gluons and quarks. For particles with no detectable signatures, their presence can be inferred through the calculation of missing transverse energy. This section briefly describes the methods used to reconstruct and identify the particles that are relevant for the analyses in this thesis.

Track reconstruction

Tracks in the inner detector are groups of sequential detector hits left by traversing charged particles. The track reconstruction process begins by taking adjacent hits found on pixel and SCT detectors and grouping them into clusters [22]. Clusters are then used to form 3-dimensional space points. For the pixel detectors, a 3-dimensional space point is determined by the centre of a cluster, while for the SCT detectors, two clusters from each side of the modules are combined to form a space point. For the TRT, the recorded drift times are converted into drift distances.

Once the space points are reconstructed, a pattern recognition algorithm begins searching for possible track candidates from the space points closest to the IP and extending radially outwards. This is referred to as the "inside-out" algorithm. Track candidates are seeded from three consecutive space points and are used to guide the search for more space points. Once the last silicon space points have been identified, a more sophisticated fit is performed, which also removes any overlapping track candidates. The silicon track is used to define the direction for a possible TRT extension. If a TRT extension is found, then the track is refitted including the TRT drift distances. The inside-out procedure can reconstruct track candidates with p_T > 100 MeV.

Following the inside-out sequence, an "outside-in" algorithm is performed on the space points and drift circles that have not yet been associated to any track. The outside-in procedure begins at the TRT and extrapolates towards the interaction point. The purpose of this algorithm is to pick up any tracks that may have originated away from the interaction point, such as photon conversions and long-lived hadron decays. The outside-in procedure can reconstruct track candidates with p_T > 300 MeV.

The status of a track at any point in the inner detector can be fully described using five independent track parameters. The five parameters used commonly in ATLAS tracking system are (d_0, z_0, ϕ, θ, q/p), where d_0 and z_0 are the radian and longitudinal impact parameters, respectively. The impact parameter is the track's closest distance of approach

to the interaction point. The two angles, ϕ and θ, describe the direction of the track. The charge and momentum of the track is denoted by q and p, respectively, which in combination describes the track curvature.

Energy reconstruction

Hadrons, electrons and photons produce a cascade of secondary particles when they interact with the calorimeters. Energy deposits are produced in groups of neighbouring calorimeter cells, which are known as clusters. Cluster reconstruction algorithms aim to include as much of the released energy as possible by maximising the signal to noise ratio. ATLAS uses two main clustering algorithms: the sliding window algorithm and the topological clustering algorithm [23].

The sliding window algorithm is best used for particles with well defined energy deposition patterns, such as electrons and photons. This algorithm searches for clusters by incrementally scanning across the electromagnetic calorimeter using a fixed sized window. A search window of 5×5 calorimeter towers is used, where a calorimeter tower is a line of cells that begins at the inner most calorimeter layer and ends at the outer most layer. Regions with a local E_T maximum greater than 3 GeV are used as precluster seeds. If two precluster seeds are found within 2×2 towers of each other, then the precluster with the higher E_T is kept. The precluster seed defines the centre of a cluster where the energies of the neighbouring cells, in a predefined cluster size, are summed. For electrons, a cluster size of $\Delta\eta \times \Delta\phi = 0.075 \times 0.175$ is used for the barrel region and $\Delta\eta \times \Delta\phi = 0.125 \times 0.125$ is used for the endcaps.

The topological clustering algorithm bases clustering on the significance of energy contents within neighbouring cells, which results in variable cluster sizes. This type of clustering is more suitable for hadronic showers where the energy response is more varied. Topological clusters are seeded by cells with a signal to noise ratio greater than 4. Neighbouring cells with a signal to ratio greater than 2 are iteratively added to the cluster. The thresholds are optimised to minimise noise and to maximise the efficiency. To prevent energy deposits from close particles being amalgamated into a single cluster, clusters are split between any local maxima measurements.

3.3.1 Electron reconstruction and identification

Electrons will leave tracks in the inner detector and deposit energy in the electromagnetic calorimeter. The tracks are required to have silicon hits with a TRT extension and the clusters are reconstructed using the sliding window algorithm. The electron reconstruction algorithm [24] begins by searching for a cluster in the electromagnetic calorimeter, which is

then matched to an inner detector track. The matching is based on the closeness of the track's extrapolated position to the cluster barycenter. For clusters with multiple track candidates, the track with the closest $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ match is chosen.

Electrons used in the analyses described in this thesis are required be within the pseudorapidity range of $|\eta|$ < 2.47 but outside the range 1.37 < $|\eta|$ < 1.52. The first requirement ensures all electrons are within the tracking acceptance of the detector and the second ensures them to be outside the transition region between the barrel and endcap calorimeters, where the efficiency and resolution of the instruments are poor. Furthermore, electrons are removed if they fall within regions of the calorimeter with known instrumentation issues, such as readout problems, dead electronics or noisy cells. The energies of electrons are measured from their clusters and their directions are measured from their tracks.

In the $Z \to \tau\tau \to ll + 4\nu$ and multijet analyses, electron candidates are required to pass the "medium" identification requirement. This requires the candidates to have cluster shapes that are consistent with electron showers, a track with at least seven silicon hits and a basic track to cluster matching. For the $H \to \tau\tau \to ll + 4\nu$ analysis, the electrons are required to pass the "tight" identification requirement. This requirement is more stringent on all the criteria used by the "medium" identification and in addition, uses the number of high threshold TRT hits as a discriminant. A tighter selection is used for the $H \to \tau\tau \to ll + 4\nu$ analysis because a stronger background suppression is required.

Electron isolation Tracking and calorimeter activities around reconstructed electrons are measured using the $p_T$cone and $E_T$cone algorithms. The $p_T$cone algorithm sums the p_T of all reconstructed tracks in a predefined cone size around the electron, while the $E_T$cone algorithm similarly sums the E_T of all energy deposits. Each algorithm operates using one of three standard cone sizes defined by the angular separation ΔR = 0.2, 0.3 and 0.4, which are labeled as "20", "30" and "40", respectively. For example, $p_T$cone40 uses a cone size of $\Delta R = 0.4$ and $E_T$cone30 uses a cone size of $\Delta R = 0.3$.

3.3.2 Muon reconstruction and identification

Muons will usually leave a track in the inner detector, small energy deposits in the calorimeters and a track in the muon spectrometer. Muon reconstruction begins by reconstructing a track in the muon spectrometer. For muons described in this thesis, this reconstruction is performed using the STACO algorithm [25].

Track reconstruction in the muon spectrometer is a multi-stage process. The first stage is to identify regions where at least one TGC or RPC hit has occurred in both the η and ϕ

coordinates. For the tracking stations within these regions, straight track segments are reconstructed locally within each station with a loose interaction point constraint. The track segments from neighbouring stations are connected to provide a rough momentum estimation. Starting with this estimation and adding slight variations, the segments from each station are extrapolated to all other stations to provide initial track candidates. A global fit is then performed on the track hits that belonged to the best candidate. At this stage, all the hits are classified as good or bad and only the good hits are kept. Finally, the track is refitted again with material effects incorporated for a more precise measurement.

Muons can also be reconstructed using the inner detector. The inner detector and muon spectrometer tracks can be connected to form combined muon tracks. The matching procedure is performed using a χ^2 minimisation. Muon candidates described this thesis are required to be combined muons and within the muon spectrometer coverage of $|\eta| < 2.4$. The difference between the z-positions of the muon track extrapolated to the beam line and the primary vertex is required to be less than 10 mm to remove atmospheric muons. To ensure the muon track is well reconstructed and to reduce the fakes from hadrons, the inner detector segment is required to have: at least two pixel hits; at least six SCT hits; and at least 90% of the TRT hits to lie within 1σ of the best fit track.

Muon isolation The same $p_T$cone and $E_T$cone algorithms used to measure electron isolation are similarly available for muons. The expected muon energy losses in the calorimeters are subtracted from the $E_T$cone calculations.

3.3.3 Jet reconstruction

Jet reconstruction aims to recover the original four-momentum of fragmenting partons by summing up the energies of their daughter particles. All jets described in this thesis are reconstructed using the Anti-k_T algorithm.

Anti-k_T is a sequential recombination algorithm that groups particles into jets based on their energies and angular separations [26]. All clusters with a matching track are considered as protojets. Each protojet is assigned an ordering parameter as follows:

$$d_i = E_{T,i}^2, \tag{3.4}$$

where $E_{T,i}$ is the transverse energy of the i^{th} protojet. All combinations of protojet pairs are also added to the same list and their ordering parameter is given by:

$$d_{ij} = min(E_{T,i}^{2k}, E_{T,j}^{2k})[(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2]/\Delta R^2, \tag{3.5}$$

where ΔR is the distance scale ($\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$) and k is the relative power of the momentum and distance scales. The Anti-k_T algorithm uses a particular setting of $k = -1$ [27]. After calculating d for all protojets and protojet combinations, the list is put in a descending order. If the smallest item in the list is a d_{ij} object, then the protojets i and j are combined to form a new protojet and the list is updated. If the smallest item on the list is a d_i object, then the protojet i is removed from the protojet list and added to the list of jets. These steps are repeated until there are no more objects in the protojet list. The jets in this thesis are all reconstructed using a distance scale of $\Delta R = 0.4$ and are required to be within the calorimeter pseudorapidity coverage of $|\eta| < 4.5$.

In the $H \to \tau\tau \to ll + 4\nu$ analysis, the jet vertex fraction (JVF) is used as an additional requirement on jet selection. This requirement ensures that the jets are produced by the hard scattering interaction and not from the underlying event. The JVF is defined as follows:

$$JVF = \frac{\sum_{i=1}^{N_{PV,tracks}} p_{T,i}(jet)}{\sum_{j=1}^{N_{tracks}} p_{T,j}(jet)}, \tag{3.6}$$

where $N_{PV,tracks}$ is the number of jet tracks that are associated with the primary vertex and N_{tracks} is the number of jet tracks associated to any vertex.

3.3.4 Flavour tagging

The identification of jets that originate from b-flavoured quark decays is an important discriminant for many processes, such as $t\bar{t}$ events. When b quarks are produced, hadronisation quickly occurs to form b-hadrons. The lifetimes of b-hadrons are relatively long and, if significantly boosted, will decay at measurable distances away from the primary vertex. Such displacements can be used to discriminate b-flavoured jets from other jets. There are many different flavour tagging algorithms, however this section will only describe two of these: the IP3D and JetFitter algorithms [28], which are relevant to this thesis.

The IP3D algorithm uses the radial and longitudinal impact parameter significances, d_0/σ_{d_0} and z_0/σ_{z_0}, to provide a likelihood measurement of a jet being b-flavoured, where σ_{d_0} and σ_{z_0} denote the uncertainties on the radial and longitudinal impact parameters, respectively. The distributions for each jet type are obtained from MC with corrections applied to account for discrepancies with the observed data [29].

The JetFitter algorithm reconstructs the intermediate decay chains within a jet to identify b-jets. The tracks of the jet are used to find a common flight path, where the b-hadron decay and its subsequent c-hadron decays are assumed to lie [30]. With this approach, several

secondary vertices may be reconstructed for a single jet, forming a rough decay chain. This decay chain is used to discriminate between different flavoured jets based on:

- the number of vertices with at least two tracks,
- the total number of tracks at these vertices,
- the number of additional single track vertices on the flight axis,
- the invariant mass of all particles attached to the chain,
- the fraction of the jet's total energy that exists as part of the decay chain,
- the displacement significance, d/σ_d, where the position is a weighted average of all the secondary vertices.

In this thesis, b-jets are identified using a combination of the IP3D and JetFitter likelihoods. This combination is based on an Artificial Neural Network technique [31].

3.3.5 Missing transverse energy reconstruction

The reconstruction of missing energy is important for inferring the presence of particles which cannot be directly detected, such as neutrinos. For a hadron collider, the initial momentum of the system is unknown in the z-direction, as the colliding partons each carry a varied fraction of the proton's total momentum. However, in the transverse direction, the momenta of the partons are negligible and assumed to be zero. Therefore the missing energy can be measured in the radial direction by applying this initial condition and the conservation of energy and momentum [32].

In the $Z \rightarrow \tau\tau \rightarrow ll + 4\nu$ analysis, the missing transverse energy (E_T^{miss}) is calculated by combining the energy deposits in the calorimeters with the momenta of the muon tracks. The missing transverse energy terms are given by:

$$E_{x(y)}^{miss} = \sum_{i=1}^{N_{cell}} E_i^{calo} sin\theta_i cos\phi_i + \sum_{j=1}^{N_\mu} [p_{x,j}^{\mu} - E_{x,j}^{\mu,loss}], \quad (3.7)$$

$$E_y^{miss} = \sum_{i=1}^{N_{cell}} E_i^{calo} sin\theta_i sin\phi_i + \sum_{j=1}^{N_\mu} [p_{y,j}^{\mu} - E_{y,j}^{\mu,loss}], \quad (3.8)$$

$$E_T^{miss} = \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2}, \quad (3.9)$$

where E^{calo} is the energy of measured in the calorimeter cells, p^{μ} is the momentum of the muon tracks and $E^{\mu,loss}$ is a correction term that accounts for the muon energy losses within the calorimeters.

In the $H \rightarrow \tau\tau \rightarrow ll + 4\nu$ analysis, the E_T^{miss} is calculated using a more precise algorithm, which was not available at the time when the $Z \rightarrow \tau\tau \rightarrow ll + 4\nu$ analysis was performed. This algorithm calculates the E_T^{miss} by combining the contributions from all reconstructed electrons, photons, muons, hadronic tauons, jets and energy deposits that are not associated with any reconstructed particles. The x and y components of the E_T^{miss} are given by:

$$E_{x(y)}^{\text{miss}} = E_{x(y)}^{\text{miss,e}} + E_{x(y)}^{\text{miss},\gamma} + E_{x(y)}^{\text{miss},\tau} + E_{x(y)}^{\text{miss,jets}} + E_{x(y)}^{\text{miss,softjets}} + E_{x(y)}^{\text{miss,calo},\mu} + E_{x(y)}^{\text{miss,cellout}}, \tag{3.10}$$

where

- $E^{\text{miss,e}}$ contains the contributions from reconstructed electrons,
- $E^{\text{miss},\gamma}$ contains the contributions from reconstructed photons,
- $E^{\text{miss},\tau}$ contains the contributions from hadronically decaying tauons,
- $E^{\text{miss,jets}}$ contains the contributions from jets,
- $E^{\text{miss,softjets}}$ contains the contributions from low p_T jets,
- $E^{\text{miss,calo},\mu}$ contains the energy losses of muons crossing the calorimeters and
- $E^{\text{miss,cellout}}$ contains the contributions that are not associated with any of the terms mentioned above.

3.4 Trigger and data acquisition system

At the LHC design luminosity, the rate of proton-proton collisions (40 MHz) far exceeds the rate at which data can be recorded (200 Hz). However, most collisions produce only inelastic proton scatterings, which are uninteresting for physics studies. To select only the events that are relevant for the desired analyses, ATLAS uses a specialised three-level trigger and data acquisition system (TDAQ). Further details of the level one (L1) and Higher-level trigger systems (HLT) can be found in references [33] and [34], respectively.

The first level trigger, L1, searches events for high p_T muons; electrons and photons; jets; and hadronically decaying tauons, as well as high amounts of missing transverse energy. Only the detector systems with fast response times are used by the L1 trigger. Muon p_T

is calculated using only the dedicated RPC and TGC trigger systems, while for the other particles, only the calorimeter clusters are used to measure their E_T. The L1 trigger assesses an event within 2.5 μs, reducing the event rate to a maximum of 75 kHz.

The L1 trigger also identifies areas of the detector where significant activity has occurred, known as Region of Interests (ROIs). To increase the processing speed, the level two trigger (L2) only applies its assessments on the ROIs. The L2 trigger reduces the event rate to around 3 kHz and has an average processing time of 40 ms. For muons, this reduction mainly comes from the refinement of muon's p_T using additional tracking information from the inner detector and the precision muon chambers. For electrons and tauons, a high p_T matching track is required, which reduces the number of possible candidates. In the case of photons, jets and the missing transverse energy, the rejection power at the L2 trigger is limited, since the tracking systems do not provide further discrimination. A dedicated system of programmable processors are used to process the L2 reconstructions within the desired time frame.

The final stage of the trigger system is performed by the Event Filter (EF). Here, offline algorithms and methods adapted for the online environment are used to reconstruct the particle candidates. The most recent calibrations, alignments and magnetic field maps are also incorporated by the EF. The more stringent selections reduce the event rate to the targeted 200 Hz, with each event having an average processing time of about 4s. To complete the reconstruction within the required time frame, a large farm of CPUs is dedicated to the EF. Events that satisfy the EF triggers are moved to permanent storage by the TDAQ system. At capacity, an output data rate of approximately 100 MB/s is recorded.

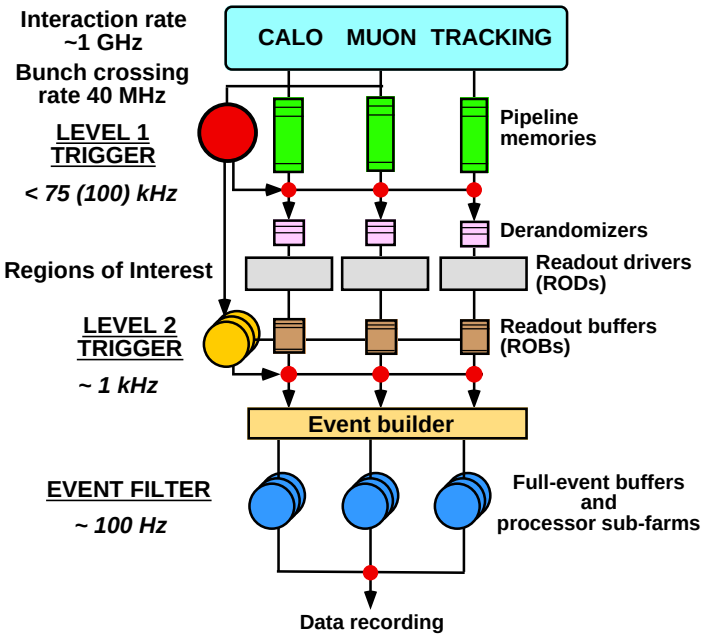


Figure 3.9: Schematic view of the ATLAS TDAQ system [20].

4

Simulation and data samples

This chapter will describe the data and simulated samples used for the multijet composition, $Z \rightarrow \tau\tau \rightarrow ll + 4\nu$ and $H \rightarrow \tau\tau \rightarrow ll + 4\nu$ analyses. All collision data used in the analyses described in this thesis were collected at $\sqrt{s}$ = 7 TeV during LHC operations in 2010 and 2011 and all simulated samples are generated at $\sqrt{s}$ = 7 TeV to match the data. The corrections applied to the simulated samples and the methods used to derive them are also described in this chapter.

4.1 Simulation

4.1.1 Event generation

Monte Carlo (MC) simulations of collision events are used extensively throughout the analyses described in this thesis. Event simulation is a multistep process that begins with generating the initial state partons. The Q^2 transfer of the hard-scattering parton interaction varies from collision to collision and for the simulations used by the analyses described in this thesis, this variation is modelled by either CTEQ [35] or MRST [36].

The next step of the simulation involves generating the final state partons. Final state partons are created from hard-scattering interactions and spectator quarks. The application of this step varies from generator to generator. In the analyses described in this thesis, a number of different generators are used, which include PYTHIA [37], HERWIG[38]/JIMMY[39], ALPGEN [40], POWHEG [41] [42], AcerMC [43] and MC@NLO [44]. All these programs use leading order matrix elements, except for MC@NLO and POWHEG, which use next-to-leading order matrix elements.

Once the final state partons have been generated, parton radiation and showering effects are simulated. This process consists of adding initial and final state radiation partons onto

existing partons. The generation of parton showers are determined probabilistically and continues until the energies of all partons are below a predetermined threshold.

After showering, all final state partons will undergo hadronisation. ALPGEN and MC@NLO use HERWIG/JIMMY to simulate this process, while AcerMC and POWHEG use PYTHIA. HERWIG is used in conjunction with JIMMY in order simulate the effects of multiple hard-scattering interactions. For all simulated samples, τ-lepton decays are additionally simulated using the TAUOLA package [45] and all generators are interfaced with PHOTOS [46] to simulate final state QED radiation.

Pileup effects from multiple proton collisions per bunch crossing are also considered at the generator level. This is modelled by overlaying simulated underlying events over the original hard-scattering process.

4.1.2 Detector simulation

To simulate detector responses and material effects, all events are passed through the ATLAS detector simulation [47], which is based on GEANT4 [48]. Relevant records of material interactions, such as charge and energy depositions, are converted into simulated detector responses. These responses are tuned to mimic the real detector as closely as possible using results gathered from laboratory tests, test beam data and cosmic ray runs. The detector simulation is also updated with the latest detector conditions, which are provided by on-going data studies.

All simulated datasets contain two types of event information: *truth* and *reconstructed*. *Truth* data refers to the information that describes events at the generator level. The exact kinematics, interactions and decay chains of all generated particles are recorded in truth data. *Reconstructed* data refers to events that are reconstructed from simulated detector responses. Only measurements that can be made with the real detector can be observed in reconstructed data.

4.2 $Z \to \tau\tau \to ll + 4\nu$ and multijet analysis samples

The simulated samples used in the $Z \to \tau\tau \to ll + 4\nu$ analysis are all generated using PYTHIA, except for the $t\bar{t}$ sample, which used MC@NLO and the diboson samples, which used HERWIG/JIMMY. All simulated samples are generated with a varied number of collision interactions to simulate different pileup conditions [10].

The cross sections of the W and Z production are calculated at NNLO using FEWZ [49]. For the Drell-Yan production, additional scale corrections are taken from reference [50] to obtain the NNLO cross section. The cross sections of the $t\bar{t}$ and diboson processes are

both calculated using MC@NLO with additional scale corrections applied to the $t\bar{t}$ sample to match the NNLO order predictions given in references [51] and [52].

Simulated samples of multijet processes are used to estimate the composition of the multijet background (see chapter 5) and for comparison purposes in the $Z \to \tau\tau \to ll + 4\nu$ analysis. The multijet samples, Mulitjet(μ), are filtered at the truth level to contain at least one prompt muon with p_T > 8 GeV and $|\eta|$ < 3. These samples are sliced into mutually exclusive sub-samples that cover different p_T ranges of the leading jet. The filtered samples offer an increase in the MC statistics but have the drawback of omitting contributions where both leptons are fake. The electron filtered Multijet(e) samples are similarly available and are filtered and sliced in the same way.

The main results of the multijet composition study are measured using a simulated sample of Minimum Bias events that's filtered on the presence of at least one 6 GeV final state particle. This sample contains 40 million events of all known QCD processes combined in the best known proportions. The advantage of using this dataset is that it is unbiased against events with fake leptons. A summary of the simulated samples used the $Z \to \tau\tau$ analyses and the multijet composition study is shown in Table 4.1.

Process	**Generator**	$\sigma\times$ **BR** $\times\varepsilon_{\mathbf{F}}$ **[nb]**
$Z \to ee\ (m_{ll} > 60GeV)$	PYTHIA	990
$Z \to \mu\mu\ (m_{ll} > 60GeV)$	PYTIHA	990
$Z \to \tau\tau\ (m_{ll} > 60GeV)$	PYTHIA	990
$W \to e\nu$	PYTHIA	1046
$W \to \mu\nu$	PYTHIA	1046
$W \to \tau\nu$	PYTHIA	1046
$t\bar{t}$	MC@NLO	91.50
$\gamma^*/Z \to ee\ (15GeV < m_{ll} < 60GeV)$	PYTHIA	146.2
$\gamma^*/Z \to \mu\mu\ (15GeV < m_{ll} < 60GeV)$	PYTHIA	146.2
$\gamma^*/Z \to \tau\tau\ (10GeV < m_{ll} < 60GeV)$	PYTHIA	396.7
WW	HERWIG	11.3
WZ	HERWIG	3.5
ZZ	HERWIG	1.0
$Multijet(\mu)$ 8-17 GeV	PYTHIA	8.48×10^5
$Multijet(\mu)$ 17-35 GeV	PYTHIA	8.14×10^5
$Multijet(\mu)$ 35-70 GeV	PYTHIA	2.21×10^5
$Multijet(\mu)$ 70-140 GeV	PYTHIA	2.85×10^5
$Multijet(e)$ 17-35 GeV	PYTHIA	9.11×10^5
$Multijet(e)$ 35-70 GeV	PYTHIA	2.50×10^5
$Multijet(e)$ 70-140 GeV	PYTHIA	3.57×10^5
Minimum Bias 6 GeV	PYTHIA	2.54×10^7

Table 4.1: Summary of the MC samples used in the $Z \to \tau\tau \to ll + 4\nu$ analysis. The generator of each sample is shown along with the expected theoretical cross sections multiplied by the branching ratio (BR) and filter efficiencies (ε_F).

4.3 $H \to \tau\tau \to ll + 4\nu$ simulated samples

The $H \to \tau\tau \to ll + 4\nu$ analysis was performed after the $Z \to \tau\tau \to ll + 4\nu$ analysis where more accurate simulations were available.

The ggH, VBF and VH signal samples are generated using POWHEG. For the $\gamma^*/Z \to ll$ ($l = e, \mu, \tau$) simulated samples, the ALPGEN generator is used because it provides a better description of the associated jet production than the other generators. The MC@NLO generator is used for all the top related samples and HERWIG is used to generate the diboson samples. The pileup setup of the $H \to \tau\tau \to ll + 4\nu$ simulated samples have smaller bunch crossing spacings to match the increased luminosities of the 2011 data.

The cross sections used for the Higgs processes was described earlier in Section 2.1. For the background processes, the cross sections are calculated using the same methods as the $Z \to \tau\tau \to ll + 4\nu$ analysis. A summary of the MC samples used for the $H \to \tau\tau \to ll + 4\nu$ analysis is shown in Table 4.2.

Process	**Generator**	**$\sigma\times$ BR [pb]**
$gg \to H \to \tau\tau$, m_H = 120 GeV	POWHEG	1.18
$VBF \to H \to \tau\tau$, m_H = 120 GeV	POWHEG	9.0×10^{-2}
$WH \to H \to \tau\tau$, m_H = 120 GeV	POWHEG	4.7×10^{-2}
$ZH \to H \to \tau\tau$, m_H = 120 GeV	POWHEG	2.6×10^{-2}
$\gamma^*/Z \to (m_{ll} > 10$ GeV, $l = e, \mu, \tau)$	ALPGEN	31.5×10^3
$t\bar{t}$	MC@NLO	15.0×10^3
Single-top t-channel	MC@NLO	64.6
Single-top s-channel	MC@NLO	4.6
Single-top Wt	MC@NLO	15.7
WW	HERWIG	44.9
WZ	HERWIG	18.0
ZZ	HERWIG	5.6

Table 4.2: Summary of the MC samples used in the $H \to \tau\tau \to ll + 4\nu$ analysis. The generator of each sample is shown along with the expected theoretical cross sections multiplied by the branching ratio (BR) and filter efficiencies (ε_F).

4.4 Data samples

Data collected during LHC operations are required to satisfy basic quality requirements. These requirements include the functionality of all ATLAS sub-detector systems relevant for the analyses presented in this thesis [53].

The $Z \rightarrow \tau\tau \rightarrow ll + 4\nu$ analysis is performed on most of the data collected during the 2010 LHC runs, which was almost all the data that was available at the time. The data has an integrated luminosity of 35.5 pb^{-1} with an uncertainty of 3.4%. The specific data periods used in the analysis and their corresponding integrated luminosities are shown in Table 4.3.

2010 data period	**Integrated luminosity (pb^{-1})**
E	0.514
F	1.743
G	5.531
H	6.984
I	20.735
Total	35.507

Table 4.3: The data periods used for the $Z \rightarrow \tau\tau \rightarrow ll + 4\nu$ analysis [54].

The $H \rightarrow \tau\tau \rightarrow ll + 4\nu$ analysis is performed on the data collected in 2011. This data has an integrated luminosity of 4.7 fb^{-1}[55] with an uncertainty of 3.9%. The data collected in 2010 is not used because they were collected under different pileup conditions, which are not replicated in the simulations used by this analysis. The data periods used for this analysis shown in Table 4.4.

2011 data period	**Integrated luminosity (pb^{-1})**
B	11.7
D	161.9
E	48.8
F	136.1
G	537.6
H	259.5
I	324.8
J	226.4
K	561.7
L	1387.7
M	1005.1
Total	4661.3

Table 4.4: The data periods used for the $H \rightarrow \tau\tau \rightarrow ll + 4\nu$ analysis [56].

4.4.1 Primary vertex selection

Primary vertices in the $Z \rightarrow \tau\tau \rightarrow ll + 4\nu$ and $H \rightarrow \tau\tau \rightarrow ll + 4\nu$ analyses are required to have a minimum of three associated tracks. This ensures the vertices come from collisions rather than from background sources such as cosmics. All events are required to have at least one primary vertex. Vertices used for pileup corrections and other selections must also satisfy these conditions.

4.4.2 Jet quality

Events that are identified to have poorly reconstructed jets or jets arising from non-collision effects are rejected to maintain a high quality of E_T^{miss} reconstruction. Jets that pass their analyses' selections are required to satisfy an additional set of quality selections based on the distribution of calorimeter energy deposits, cluster shapes and jet timings.

In the $Z \to \tau\tau \to ll + 4\nu$ analysis, the quality selections are further described in references [10] and [57]. For the $H \to \tau\tau \to ll + 4\nu$ analysis, these are further described in references [58] and [59].

4.5 Monte Carlo corrections

In order for the simulations to better describe the observed collisions, corrections are applied to the relevant distributions of the simulations to account for any observed differences it has with the data.

In the $Z \to \tau\tau \to ll + 4\nu$ and $H \to \tau\tau \to ll + 4\nu$ analyses, corrections are applied to the trigger, particle identification and isolation efficiencies. In addition, extra resolution smearing and scaling corrections on are applied to the p_T or E_T of reconstructed muons, electrons and jets. In each case, either control regions or separate studies are used to obtain the corrections. Corrections are also applied to the pileup conditions of the simulations to replicate those of the data.

4.5.1 $Z \to \tau\tau \to ll + 4\nu$ corrections

Pileup reweighting

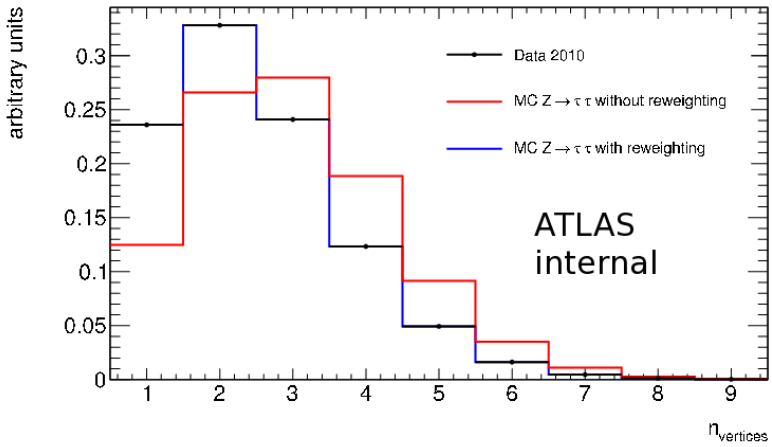
The collision data used in this analysis was collected during periods where the number of interactions per event varied. The number of reconstructed primary vertices is used to estimate the number of interactions per event. Simulated events are reweighted by matching the primary vertex distribution of the signal sample to that of the data [10].

The pileup weights are calculated after the data quality, trigger, vertex and jet quality selections and are shown in Table 4.5. Since all simulated samples are generated with the same pileup conditions, these weights are applied to all the samples. The vertex multiplicities of the data and the signal MC before and after the reweighting procedure are shown in Figures 4.1(a) and 4.1(b) for the $e\mu$ and $\mu\mu$ channels, respectively.

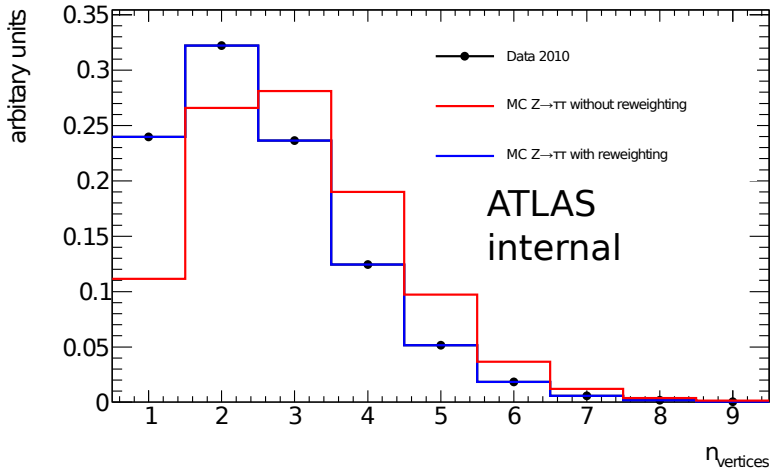
The systematic uncertainty of the pileup reweighting procedure is estimated by varying the reweighting factors given in Table 4.5 coherently up and down by one standard error. These variations change the estimated signal and background yields by approximately 0.6%.

N (vertices)	$e\mu$ channel	$\mu\mu$ channel
1	1.891 ± 0.013	2.149 ± 0.060
2	1.234 ± 0.006	1.211 ± 0.022
3	0.861 ± 0.004	0.840 ± 0.015
4	0.654 ± 0.004	0.654 ± 0.016
5	0.539 ± 0.004	0.530 ± 0.025
6	0.465 ± 0.006	0.500 ± 0.039
7	0.415 ± 0.010	0.459 ± 0.069
8	0.448 ± 0.021	0.447 ± 0.069
9	0.407 ± 0.038	0.282 ± 0.069
≥ 10	0.582 ± 0.106	0.573 ± 0.143

Table 4.5: The MC vertex weights used for the $e\mu$ and $\mu\mu$ channels.



(a) $e\mu$ channel



(b) $\mu\mu$ channel

Figure 4.1: Vertex multiplicities for the $e\mu$ and $\mu\mu$ channels between the data and signal MC before and after the pileup reweighting [10].

Trigger efficiencies

The efficiency of the 15 GeV "medium" electron trigger is measured using a tag-and-probe method [10]. Tag-and-probe is a method that uses resonant decays with strong detection signatures to measure the efficiencies of one of its daughter particles, which in this case are the leptons. Stringent selections are used to obtain a pure sample of the signature events, which is referred to as "tagging." The daughter particle, whose efficiency is then measured, is referred to as the "probe." To minimize selection biases, the "tag" selections usually does not use any kinematic information from the "probe."

For electrons with 16 < p_T < 20 GeV, the efficiencies are derived from $W \rightarrow e\nu$ events and for electrons with p_T > 20 GeV, $Z \rightarrow ee$ events are used. Discrepancies between the data and MC are corrected for by rescaling the simulated efficiencies, which are given in Table 4.6 [10].

The trigger efficiency corrections for the 15 GeV "medium" electron trigger has an uncertainty ranging from 2.5% for $p_T(e) < 20$ GeV to 0.5% for $p_T(e)$ > 20 GeV. For the $\mu\mu$ triggers, the uncertainties on the scale factors vary between 1-3% for the 13 GeV and 13 GeV "tight" muon triggers; and between 2-7% for the 10 GeV muon trigger. The tight muon trigger differs from the other triggers because it is seeded by a 10 GeV L1 trigger. The non-tight muon triggers require only two trigger hits and hence, has no p_T requirement. The 10 GeV muon trigger collected less data than the other two triggers and therefore has a larger statistical uncertainty.

The statistical and systematic uncertainties on the trigger efficiency corrections are combined in quadrature and the total uncertainty is varied up and down by $\pm\sigma$ to measure the trigger systematic uncertainties.

p_T range	Correction factor
16 < E_T < 18 GeV	1.004 $\pm$ 0.025
18 < E_T < 20 GeV	0.987 $\pm$ 0.023
E_T > 20 GeV	0.995 $\pm$ 0.005

Table 4.6: Correction factors applied to the MC for the 15 GeV "medium" electron trigger.

The efficiency corrections for the single muon triggers, described in Section 6.1, is measured using a tag-and-probe method on $Z \rightarrow \mu\mu$ events [54]. The derived corrections vary mostly between $\pm$ 5%. In the $\mu\mu$ channel, the single muon triggers can be fired by either one of the two muons, so the overall efficiency of the triggers is given by:

$$\varepsilon = 1 - (1 - \varepsilon(\mu_1))(1 - \varepsilon(\mu_2)), \tag{4.1}$$

where $\varepsilon(\mu_1)$ and $\varepsilon(\mu_2)$ are the efficiencies of the two muons passing the triggers individually. For events with two muons, the overall corrections to the trigger efficiencies is found to vary mostly between $\pm$ 1%.

Lepton energy scale and resolution

The p_T (or E_T) scale and resolution of leptons in MC are corrected to match the observations in data using $Z \to ee$ events for electrons; and $Z \to \mu\mu$ and $W \to \mu\nu$ events for muons.

The energy resolution of simulated electrons is 1% narrower in the barrel and 2 - 4% narrower in the endcaps. Additional resolution smearing is applied to simulated electrons as a function of η across 50 bins [60].

The momentum resolution of muons in data and MC are studied individually for the inner detector and muon spectrometer track segments [61]. The resolutions in simulations is 1 - 3% narrower for the inner detector tracks and 4 - 9% narrower for the muon spectrometer tracks. Corrections to the resolutions are applied separately for the two track segments and then combined for the full track.

The corrections to the muons' inner detector and muon spectrometer momentum resolutions; and the corrections to the electrons' energy resolutions have statistical and systematic uncertainties. The four components of uncertainties for the muons are combined in quadrature and the total is varied by $\pm\sigma$. Similarly for the electrons, the two components are comined in quadrature and the total is varied. The variations are propagated through the analysis to measure its systematic effects. The variations on the p_T of the leptons are also propagated into the calculation of the E_T^{miss} for each event.

Lepton efficiencies

Lepton identification and isolation efficiencies are measured for electrons and muons using tag-and-probe methods. For muons, $Z \to \mu\mu$ events are used to measure the identification and isolation efficiencies. For electrons, a combined study using $Z \to ee$ and $W \to e\nu$ events are used to measure the identification efficiency, while the isolation efficiency is measured using only $Z \to ee$ events.

The efficiency of the muon identification is measured to be up to 2% lower in data than in MC and a set of η dependent corrections are applied to account for this discrepancy [62] [63]. The isolation efficiencies of muons are mostly consistent between data and MC with the largest discrepancy of 2% observed for muons with p_T < 25 GeV [54]. Correction factors are applied for the muon isolation efficiencies as a function of p_T.

Electron identification efficiency is slightly lower in data than MC, with the largest discrepancy of about 5% observed for electrons with p_T < 25 GeV. Corrections are applied to simulated

electrons over eight bins of η and seven bins of p_T [64]. The isolation efficiencies of electrons are consistent within the uncertainties between the data and MC and as such, no corrections are applied.

The uncertainties on the scale factors to correct lepton identification, reconstruction and isolation efficiencies are considered as sources of systematic uncertainty. For electrons, an uncertainty of 1.5% on the reconstruction efficiency is used. This uncertainty along with the uncertainties of the identification and isolation efficiencies are combined in quadrature to give an overall uncertainty. Similarly for muons, the uncertainties on the identification and isolation are combined in quadrature. For both lepton flavours, the combined uncertainties are varied by $\pm\sigma$ and propagated through the analysis.

4.5.2 $H \to \tau\tau \to ll + 4\nu$ corrections

Pileup corrections

The collision data used in the $H \to \tau\tau \to ll + 4\nu$ analysis have higher pileup conditions than the data used in the $Z \to \tau\tau \to ll + 4\nu$ analysis. As such, a more accurate reweighting of the simulated pileup conditions is required. In this analysis, the average number of interactions per bunch crossing is reweighted instead of the primary vertex multiplicity. In this way, any discrepancies in the vertex reconstruction efficiencies are mitigated. However, since this efficiency is expected to be similar between data and MC, the distributions of the vertex multiplicities are compared after the reweighting procedure as a check. This comparison is shown in Figure 4.2, where a rough agreement is observed between the data and the corrected MC.

The systematic uncertainties associated with the simulated pileup conditions are considered by smearing the jet energy as a function of η and p_T. The extra smearing is performed to mimic the effects of calorimeter noise from the underlying event.

Trigger efficiencies

The efficiencies of the triggers used in the $H \to \tau\tau \to ll + 4\nu$ analysis are measured from data using $Z \to ee$ and $Z \to \mu\mu$ events. For the triggers that require both an electron and a muon, the efficiencies of the two leptons are assumed to be independent and are measured separately. The corrections for these triggers are also applied to the electron and muon separately.

For the single electron triggers, the efficiencies between the data and MC mostly agree within 2% with the largest differences measured to be around 5% [58]. For the 12 GeV di-electron trigger, the efficiencies agree within the uncertainties between the data and MC

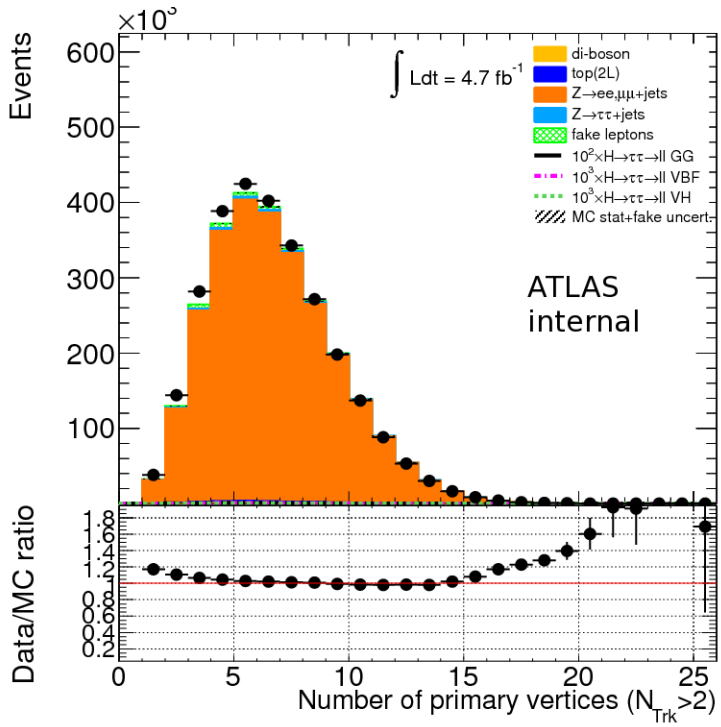


Figure 4.2: Comparison between the number of primary vertices in data and MC after applying pileup corrections [58].

and no corrections are applied. For the single muon triggers, corrections of up to 5% are applied for the muons in the barrel and up to 2% for the muons in the endcaps.

Systematic uncertainties of the trigger efficiency corrections are measured similarly to the $Z \to \tau\tau \to ll + 4\nu$ analysis.

Lepton scale and resolution

The corrections for the lepton energy scale and resolution are derived using the same tag-and-probe methods, as described for the $Z \to \tau\tau \to ll + 4\nu$ analysis. Corrections for the inner detector and muon spectrometer track segments are applied separately for muons. Due to the higher pileup conditions, the corrections to the electron energy scale additionally considers the number of interactions, as the underlying event affects the responses of the calorimeters [58].

The corrections for muon inner detector and muon spectrometer track segments have statistical and systematic uncertainties. These components are combined in quadrature and varied by $\pm\,\sigma$ independently for the two track segments. The two uncertainties derived from each track segment are then combined in quadrature. For electrons, the statistical and systematic uncertainties of the corrections are combined in quadrature and varied by $\pm\,\sigma$.

Lepton efficiencies

The corrections on the lepton identification, reconstruction and isolation efficiencies are derived similarly to the $Z \to \tau\tau \to ll + 4\nu$ corrections. However, the electron isolation corrections additionally take into account the effects of pileup by factoring in the number of reconstructed vertices [58]. Corrections are applied to all these efficiencies, except for the muon isolation efficiency, which is found to be consistent between the data and simulation.

The lepton trigger, reconstruction and identification scale factors have statistical and systematic uncertainties. These uncertainties are combined in quadrature and varied by $\pm\sigma$ separately for each correction. Each variation is propagated separately through the analysis to calculate their systematic effects.

4.6 Summary

This chapter has described the data and simulated samples used by the multijet composition, $Z \to \tau\tau \to ll + 4\nu$ and $H \to \tau\tau \to ll + 4\nu$ analyses, which will be described later in chapters 5, 6 and 8, respectively. The important efficiencies and distributions of the simulations are checked against the data for consistency and any observed differences are corrected for. The $H \to \tau\tau \to ll + 4\nu$ analysis was performed after the other two analyses and thus uses some simulated samples and corrections that are more accurate. Therefore the systematic uncertainties associated with the simulations are smaller in the $H \to \tau\tau \to ll + 4\nu$ analysis.

5

Multijet composition

Multijet processes are significant backgrounds for both the $Z \to \tau\tau \to ll + 4\nu$ and $H \to \tau\tau \to ll + 4\nu$ analyses. This background consists of a variety of different QCD processes and this chapter describes a method that measures their relative contributions to the dilepton analyses. Detailed knowledge of the multijet composition is essential for understanding the underlying assumptions that are used when estimating this background using data-driven methods.

5.1 Lepton selection

Reconstructed leptons are required to satisfy the requirements described in sections 3.3.1 and 3.3.2. Electrons are required to have p_T > 15 GeV and muons are required to have p_T > 10 GeV. In addition, the muon inner detector and muon spectrometer track segments must be matched with a χ^2 < 150; and the momentum of the muon spectrometer segment must be greater than 60% of the inner detector segment. The first requirement ensures the two track segments are well matched and the second ensures the track segments come from the same muon, which is expected to lose some of its energy whilst traversing the calorimeters.

Isolation selection Isolation requirements in the form of $p_T$cone40/p_T and $E_T$cone30/p_T are strong discriminants against multijet lepton candidates. This study uses the following isolation selections:

- Electrons: $p_T$cone40/p_T < 0.06 and $E_T$cone30/p_T < 0.10,
- Muons: $p_T$cone40/p_T < 0.06 and $E_T$cone40/p_T < 0.06.

The selections of this study were chosen before both the $Z \rightarrow \tau\tau \rightarrow ll + 4\nu$ and $H \rightarrow \tau\tau \rightarrow ll + 4\nu$ analyses were conducted with the intention of determining the multijet composition for these analyses. The lepton selections were therefore chosen to be as close as possible to those expected to be used. As such, the described selections are similar but not identical to those used in the two dilepton analyses. However, the results achieved by this study are expected to be similar enough to the composition of the two dilepton analyses to be useful.

5.2 Monte Carlo multijet study

The production of multijet events at the LHC is so large that generating the sufficient number of MC events is not feasible. To overcome this problem, a "factorisation" method is used to calculate the selection efficiencies of this background.

5.2.1 Estimating the number of two lepton events

The simulated Minimum Bias sample used for this study (see Table 4.1) contains very few events with two or more reconstructed leptons. As such, the expected number of two lepton events is estimated using factorisation. This method assumes the reconstruction efficiencies of the two leptons are independent. This assumption is tested later in the systematic uncertainties section (Section 5.3). The efficiency of the two lepton selection is given by:

$$\varepsilon_{\ell 1,\ell 2} = \varepsilon_{\ell 1} \times \varepsilon_{\ell 2}, \tag{5.1}$$

where ε_ℓ is the reconstruction efficiency of lepton ℓ, ℓ_1 is the leading lepton and ℓ_2 is the sub-leading lepton. ε_ℓ is measured separately for different multijet processes and is uniquely defined by the lepton flavour (e or μ); whether its real or fake; and the quark flavour of its mother particle (b, c or light flavoured). Real leptons here only refer to prompt leptons that are produced by the heaviest flavoured hadrons in each event. All non-prompt or secondary leptons are considered as fakes. These definitions provide the best two category grouping for leptons with similar selection efficiencies. ε_ℓ is calculated using the following equation:

$$\varepsilon_\ell = \frac{N_\ell}{2 \times Br(q \rightarrow \ell) \times N(q\bar{q})} \tag{5.2}$$

where

- q is the flavour of the multijet event,

- ℓ is the type of lepton,
- N_ℓ is the number of events with at least one reconstructed lepton originating from a $q \to \ell$ process,
- $N(q\bar{q})$ is the total number of $q\bar{q}$ events in the simulated sample, and
- $Br(q \to \ell)$ is the branching fraction of $q \to \ell$. For this study the branching fraction for b quarks to electrons (or muons) is assumed to be 11% (11%) and c quarks to electrons (or muon) is assumed to be 10% (10%).

All fake leptons are assumed to come from jets. The probability of producing a fake lepton is expected to be uncorrelated with the decay modes of the heavy flavoured quarks and so a branching fraction of 1 is assumed for fake leptons. The measured lepton reconstruction efficiencies are shown in Table 5.1.

Lepton type	$\mathbf{b\bar{b}}$ (10^{-3})	$\mathbf{c\bar{c}}$ (10^{-3})	$\mathbf{q\bar{q}}$ (10^{-3})
Real electron	5.33	0.26	-
Fake electron	0.69	0.47	0.36
Real muon	12.80	1.09	-
Fake muon	0.51	0.10	0.07

Table 5.1: Lepton reconstruction efficiencies for $b\bar{b}$, $c\bar{c}$ and light-flavoured multijet events [65].

A total of nine channels considered for this study are they are listed in Table 5.2. Heavy flavoured quarks (b and c) produce both real and fake leptons and as such, all four dilepton permutations are considered. Light flavoured multijet events are dominated by fakes (> 99%) and to simplify the study, all leptons produced by these events are assumed to be fake.

Quark flavour	**Lepton 1**	**Lepton 2**
$b\bar{b}$	real	real
$b\bar{b}$	real	fake
$b\bar{b}$	fake	real
$b\bar{b}$	fake	fake
$c\bar{c}$	real	real
$c\bar{c}$	real	fake
$c\bar{c}$	fake	real
$c\bar{c}$	fake	fake
$q\bar{q}$ (light flavoured)	fake	fake

Table 5.2: A list of the considered multijet channels.

For all channels, the efficiency of the two lepton selection is calculated using the following equation:

$$\varepsilon_{2\ell} = \varepsilon_{\ell 1,\ell 2} \times \varepsilon_{OS} \times A_{comb}, \tag{5.3}$$

where A_{comb} is the number of combinatorial permutations, which can either be 1 or 2, and ε_{OS} is the efficiency of the opposite sign requirement.

The efficiency of the opposite sign requirement is measured using the Multijet(e, μ) 17-35 GeV filtered sample, as the Minimum Bias sample doesn't contain enough two lepton events. Since the filtering of the Multijet samples are performed at the truth level, ε_{OS} can only be measured in the channels with at least one real lepton. The measured values are presented in Table 5.3.

QCD channel	OS events	SS events	ε_{OS}
$b\bar{b}$ - Real electron, real muon	366	127	0.74 ± 0.10
$b\bar{b}$ - Fake electron, real muon	206	219	0.49 ± 0.06
$c\bar{c}$ - Real electron, real muon	12	2	0.86 ± 0.48
$c\bar{c}$ - Fake electron, real muon	84	80	0.51 ± 0.10
$b\bar{b}$ - Real muon, real muon	495	136	0.78 ± 0.10
$b\bar{b}$ - Real muon, fake muon	66	97	0.40 ± 0.08
$c\bar{c}$ - Real muon, real muon	55	7	0.89 ± 0.23
$c\bar{c}$ - Real muon, fake muon	11	17	0.39 ± 0.19

Table 5.3: The number of OS and SS events and ε_{OS} for events where both leptons have $p_T > 10$ GeV. The electron p_T requirement is relaxed from 15 GeV because some channels have insufficient statistics [65].

In the $b\bar{b}$ channels with two real leptons, ε_{OS} slightly deviates away from unity. In the $c\bar{c}$ channels with two real leptons, ε_{OS} is still consistent with unity at 1σ. For all channels with one fake lepton, ε_{OS} is close to 50%, which indicates that the charge of the fake lepton is mostly uncorrelated with the charge of the real lepton. For the multijet channels with two fake leptons, ε_{OS} cannot be directly measured and is assumed to be 50% following the results of the channels with one fake lepton. To simplify the observed measurements, the multijet channels are assigned the following ε_{OS} values:

- $b\bar{b}$ - Real, real: $\varepsilon_{OS} = 0.75$,
- $c\bar{c}$ - Real, real: $\varepsilon_{OS} = 1$,
- All other channels: $\varepsilon_{OS} = 0.5$.

Using the measured efficiencies and equation 5.3, the number of two lepton events for each channel is calculated and shown in Table 5.4. The results are calculated for an integrated luminosity of 1.549 $\times 10^{-2}$ pb^{-1}, which is the approximate size of the Minimum Bias sample.

Multijet channel	Expected events $\pm\sigma_{sys} \pm \sigma_{stat}$		
	ee	$e\mu$	$\mu\mu$
$b\bar{b}$ - real real	$0.045 \pm 0.004 \pm 0.004$	$0.62 \pm 0.06 \pm 0.04$	$2.10 \pm 0.21 \pm 0.07$
$b\bar{b}$ - real fake	$0.079 \pm 0.008 \pm 0.007$	$0.15 \pm 0.02 \pm 0.01$	$1.01 \pm 0.10 \pm 0.03$
$b\bar{b}$ - fake real	-	$0.54 \pm 0.05 \pm 0.04$	-
$b\bar{b}$ - fake fake	$0.052 \pm 0.005 \pm 0.005$	$0.19 \pm 0.02 \pm 0.02$	$0.18 \pm 0.02 \pm 0.01$
$c\bar{c}$ - real real	$0.001 \pm 0.001 \pm 0.001$	$0.019 \pm 0.002 \pm 0.002$	$0.104 \pm 0.010 \pm 0.005$
$c\bar{c}$ - real fake	$0.016 \pm 0.002 \pm 0.002$	$0.009 \pm 0.001 \pm 0.001$	$0.096 \pm 0.010 \pm 0.005$
$c\bar{c}$ - fake real	-	$0.18 \pm 0.02 \pm 0.01$	-
$c\bar{c}$ - fake fake	$0.15 \pm 0.02 \pm 0.01$	$0.16 \pm 0.02 \pm 0.01$	$0.045 \pm 0.005 \pm 0.003$
$q\bar{q}$ - fake fake	$0.28 \pm 0.03 \pm 0.01$	$0.27 \pm 0.03 \pm 0.01$	$0.06 \pm 0.001 \pm 0.003$

Table 5.4: Expected two lepton events for all dilepton channels for an integrated luminosity of 1.524E-2 pb^{-1} [65].

Two lepton estimate cross check

The estimates shown in Table 5.4 can be tested in the channels with at least one real muon by comparing the factorised results to the number of events directly measured in the Multijet(μ) 17-35 GeV sample. The Multijet(μ) 17-35 GeV sample is chosen to perform this cross check because it covers a p_T spectrum that's similar to the Minimum Bias sample, where a comparison is shown in Figure B.1. This cross check is only performed in the channels with a sufficient number of two lepton events. When directly measuring the number of two lepton events, one must ensure that the two leptons are not produced by the same jet. In the $Z \rightarrow \tau\tau \rightarrow ll + 4\nu$ and $H \rightarrow \tau\tau \rightarrow ll + 4\nu$ analyses, the invariant mass of the two leptons is required to be above 25 GeV and above 30 GeV, respectively. As the number of multijet events drops off quickly with increasing p_T, only lepton pairs that are relatively back-to-back will pass these selections. To ensure that the two leptons are relatively back-to-back, a ΔR > 2.0 requirement is applied when measuring the number of two lepton events in the Multijet(μ) sample.

The Multijet(μ) 17-35 GeV sample has an integrated luminosity of 2.21 pb^{-1}, which is approximately 137 times the size of the Minimum Bias sample. The factorised estimates are scaled up to match the integrated luminosity of the Multijet(μ) sample and the results are shown in Table 5.5, where only the channels with a sufficient number of two lepton events are shown.

The factorised estimates agree with the direct measurements within the given uncertainties. This suggests that the factorisation methodology is generally valid for the tested channels and is assumed to be equally as valid for the channels that could not be tested.

Multijet channel	Factorised estimate ($\pm\sigma_{stat} \pm \sigma_{sys}$)	Observed in $J1(\mu)$
$b\bar{b}$ - Real electron, real muon	$84 \pm 5 \pm 8$	86
$b\bar{b}$ - Fake electron, real muon	$74 \pm 5 \pm 7$	73
$c\bar{c}$ - Real electron, real muon	$2.5 \pm 0.3 \pm 0.3$	2
$c\bar{c}$ - Fake electron, real muon	$24 \pm 1 \pm 2$	26

Table 5.5: Comparison table for the factorised estimates of the MinBias6 sample and the direct measurement of the $J1(\mu)$ sample [65].

5.2.2 Isolation efficiency

Once the number of two lepton events is calculated, the isolation selections efficiencies are similarly calculated for each lepton type and then factorised. To keep the results of this study as general as possible, the efficiencies of additional selections are not calculated. The results of this study are therefore most valid after the dilepton and isolation requirements, but can be used as an upper bound estimate when considering further selections.

The isolation efficiencies are measured for each type of multijet lepton. These are calculated from events with at least one reconstructed lepton. For events with more than one reconstructed lepton, each lepton contributes independently to its own type. The measured isolation efficiencies are given in Table 5.6.

Lepton type	$b\bar{b}$	$c\bar{c}$	$q\bar{q}$
Real electron	0.048 ± 0.009	0.094 ± 0.024	-
Fake electron	0.105 ± 0.017	0.099 ± 0.008	0.128 ± 0.006
Real muon	0.038 ± 0.004	0.044 ± 0.005	-
Fake muon	0.028 ± 0.006	0.065 ± 0.009	0.119 ± 0.007

Table 5.6: Lepton isolation efficiencies [65].

5.2.3 Estimated multijet events

Assuming the lepton reconstruction efficiencies are independent of the isolation efficiencies, the expected number of multijet events with two isolated leptons is given by:

$$N_{multijet,i} = Scale_{\mathcal{L}} \times N_{ll,i} \times \varepsilon_{isolation,1} \times \varepsilon_{isolation,2}, \tag{5.4}$$

where $N_{multijet,i}$ is the expected multijet background for a given channel i, $Scale_{\mathcal{L}}$ is the luminosity scale, $N_{ll,i}$ is the expected number of two lepton events and $\varepsilon_{isolation}$ is the lepton isolation efficiency. For the integrated luminosity of the $Z \rightarrow \tau\tau \rightarrow ll + 4\nu$ analysis (35.5 pb^{-1}), the number of expected events in each channel is given Table 5.7, neglecting any

preselection efficiencies such as vertex or trigger requirements. The relative contributions of the multijet events are presented in Table 5.8.

Multijet channel	$N_{multijet}(\pm\sigma_{stat}\pm\sigma_{sys})$		
	ee	$e\mu$	$\mu\mu$
$b\bar{b}$ - real real	$0.24 \pm 0.09 \pm 0.02$	$2.60 \pm 0.11 \pm 0.26$	$7.11 \pm 1.13 \pm 0.71$
$b\bar{b}$ - real fake	$0.91 \pm 0.21 \pm 0.09$	$0.45 \pm 0.03 \pm 0.05$	$2.48 \pm 0.58 \pm 0.25$
$b\bar{b}$ - fake real	-	$4.95 \pm 0.30 \pm 0.50$	-
$b\bar{b}$ - fake fake	$1.29 \pm 0.38 \pm 0.13$	$1.29 \pm 0.08 \pm 0.13$	$0.32 \pm 0.14 \pm 0.03$
$c\bar{c}$ - real real	$0.017 \pm 0.008 \pm 0.002$	$0.17 \pm 0.01 \pm 0.02$	$0.45 \pm 0.10 \pm 0.05$
$c\bar{c}$ - real fake	$0.35 \pm 0.09 \pm 0.04$	$0.12 \pm 0.01 \pm 0.01$	$0.93 \pm 0.15 \pm 0.09$
$c\bar{c}$ - fake real	-	$1.74 \pm 0.07 \pm 0.17$	
$c\bar{c}$ - fake fake	$3.37 \pm 0.50 \pm 0.34$	$2.40 \pm 0.10 \pm 0.24$	$0.42 \pm 0.11 \pm 0.04$
$q\bar{q}$ - fake fake	$10.79 \pm 0.98 \pm 1.08$	$9.46 \pm 0.24 \pm 0.95$	$2.08 \pm 0.25 \pm 0.21$
Total	$16.97 \pm 2.27 \pm 1.70$	$23.20 \pm 0.96 \pm 2.32$	$13.80 \pm 2.47 \pm 1.38$

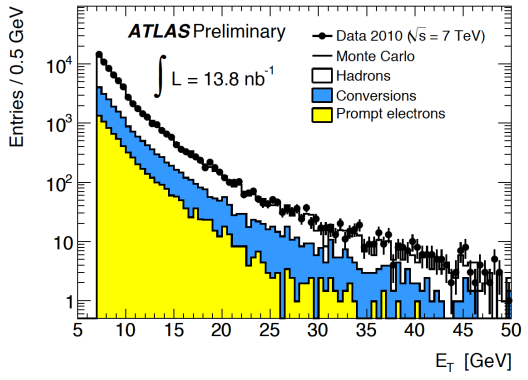
Table 5.7: Expected multijet events with two isolated leptons for 35.5 pb^{-1} [65].

Multijet channel	Percentage of the total ($\pm\sigma_{stat}\pm\sigma_{sys}$)		
	ee	$e\mu$	$\mu\mu$
$b\bar{b}$ - real real	1.43%±0.53%±0.14%	11.22%±0.48%±1.12%	51.54%±8.19%±5.15%
$b\bar{b}$ - real fake	5.35%±1.25%±0.53%	1.96%±0.13%±0.20%	17.98%±4.42%±1.80%
$b\bar{b}$ - fake real	-	21.35%±1.30%±2.13%	-
$b\bar{b}$ - fake fake	7.61%±2.26%±0.76%	5.57%±0.35%±0.56%	2.34%±1.02%±0.23%
$c\bar{c}$ - real real	0.10%±0.05%±0.01%	0.74%±0.04%±0.07%	3.29%±0.73%±0.33%
$c\bar{c}$ - real fake	2.08%±0.53%±0.21%	0.52%±0.04%±0.05%	6.73%±1.10%±0.67%
$c\bar{c}$ - fake real	-	7.52%±0.30%±0.75%	-
$c\bar{c}$ - fake fake	19.85%±2.97%±1.98%	10.35%±0.43%±1.03%	3.07%±0.80%±0.31%
$q\bar{q}$ - fake fake	63.59%±5.76%±6.36%	40.78%±1.04%±4.08%	15.06%±1.83%±1.51%

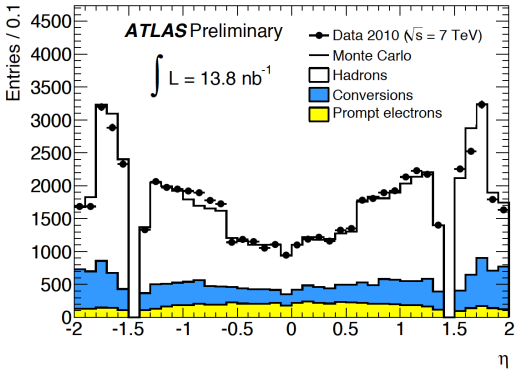
Table 5.8: Contributions from each multijet channel given as a percentage of the total background [65].

Cross section scale corrections

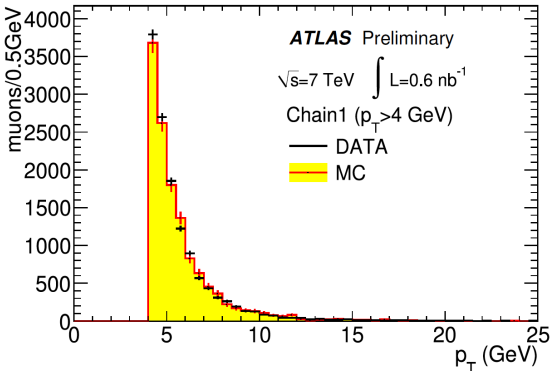
The predicted number of multijet events exceeds the amount observed in data. This discrepancy is attributed to a mismodelling of the multijet cross sections, which is corrected by rescaling the MC to match the data. The kinematic distributions of the processes are assumed to be well modelled, which is supported by studies of multijet events that produce electrons and muons [66] [67]. Comparisons of the lepton p_T and η distributions are shown in Figure 5.1, where the MC distribution is generated using the same Minimum Bias sample as the one used for this study and the MC scale is adjusted to match the number of observed events in data.



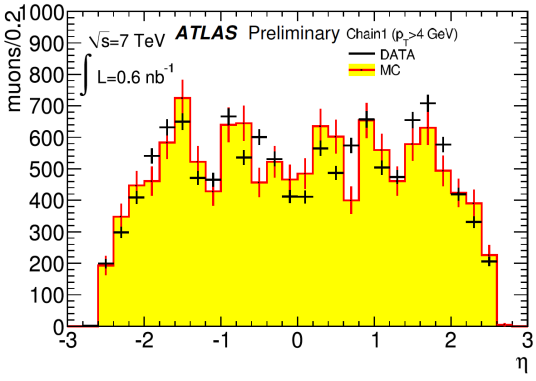
(a) Electron E_T



(b) Electron η



(c) Muon p_T



(d) Muon η

Figure 5.1: The electron and muon p_T (E_T) and η distribution comparisons [66] [67].

The multijet cross section correction factor, k, is calculated after the two lepton requirement using:

$$k = N^{MC}_{multijet}/N^{Data}_{multijet}, \tag{5.5}$$

where $N^{Data}_{multijet}$ and $N^{MC}_{multijet}$ are the number of mulitjet events observed in data and MC, respectively. For $N^{Data}_{multijet}$, the expected contamination from non-mulitjet processes is estimated using simulations and subtracted. For $N^{MC}_{multijet}$, the number of events is estimated using the sum of the Multijet(μ) samples. The k factor is calculated only from the $\mu\mu$ channel where the contributions from two fake leptons is the smallest. A scale factor of $k = 0.664 \pm 0.057 \pm 0.100$ is found [65], where a 15% systematic uncertainty assigned for the expected contributions from light-flavoured fake events, which are not included in the filtered samples.

Mulitjet estimate cross check

The factorised multijet estimate in the $e\mu$ channel is cross checked against a data-driven estimate to validate the methodologies described in this chapter. The data-driven estimate uses the same lepton selections as the factorised estimate and the method is described in Section 6.4.1. The 10 GeV electron and 6 GeV muon dilepton trigger was used to collect the data. For a fair comparison, the factorised estimate is scaled by efficiency of this trigger, which is measured to be $\varepsilon_{trigger,e\mu} = 0.818 \pm 0.023$ (stat.) [65]. The factorised estimate in the $e\mu$ channel is:

$$N^{multijet}_{MC,e\mu} = k \times \varepsilon_{trigger,e\mu} \times N_{e\mu}, \quad (5.6)$$

$$= 12.60 \pm 1.47 \text{ (stat.)} \pm 3.06 \text{(sys.)}. \quad (5.7)$$

The equivalent data-driven method estimates the mulitjet contributions to be [65]:

$$N^{multijet}_{data,e\mu} = 16.37 \pm 8.57 \text{ (stat.)}. \quad (5.8)$$

Systematic uncertainties are not estimated for the data-driven cross check. The two results are consistent within the given uncertainties, which gives great confidence to the factorisation methods used in this chapter.

5.3 Systematic uncertainties

The correlation between the reconstruction efficiencies of the two leptons is considered as a systematic uncertainty. The extent of the correlation is tested in heavy flavoured quark events by plotting the p_T of one quark against the probability that its anti-quark has a p_T > 10 GeV (or p_T > 15 GeV), which is shown in Figure 5.2. The correlation tests are conducted on the mother particles instead of the leptons because there are not enough two lepton events in the Minimum Bias sample. The p_T of prompt leptons is strongly associated with the p_T of its mother particle and therefore the correlation between the mother particles is a good estimate of the lepton correlation.

The p_T of the $b\bar{b}$ and $c\bar{c}$ quark pairs are noticeably correlated. If the p_T of the first quark is above 10 or 15 GeV, then the probability of its partner having a p_T> 10 GeV increases by about 0.1 for both flavours. If the p_T correlation between the quarks is 100%, then the plot in Figure 5.2 should be step-function that rises at $p_T = 10$ GeV. If there is no p_T correlation between the quarks, then the plot should have zero gradient. Figure 5.2 shows a slight gradient , which suggests that the p_T of the quark pairs are closer to being uncorrelated than

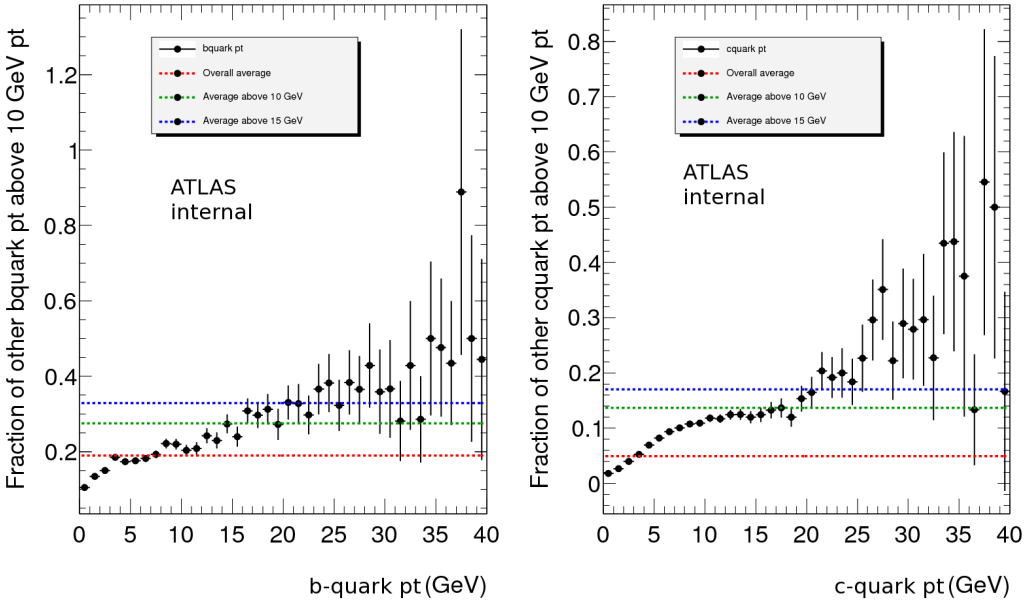


Figure 5.2: The probability of the opposite quark having a $p_T > 10$ GeV as a function of the first quark's p_T. The red line represents the global average, the green line represents the average for events with the first quark's $p_T > 10$ GeV and the blue line for events with the first quark's $p_T > 15$ GeV [65].

fully correlated. The correlation is estimated to be 10% since the blue and green lines are about 10% of the distance above the red line tending towards unity. Assuming the quarks' momentum is correlation is directly proportional to the lepton reconstruction efficiencies, a systematic uncertainty of 10% is assigned for the predicted yields in the $b\bar{b}$ and $c\bar{c}$ channels.

The p_T correlations between light-flavoured quarks are not studied because their mother particles are not distinct. Leptons produced in light quark events are generally non-prompt, so the p_T correlation is expected to be less than that of the heavy flavoured quarks. Conservatively, a systematic uncertainty of 10% is also assigned for the estimates in the light-flavoured channels.

5.4 Summary

The results of this study show that the three dilepton channels are affected differently by the various multijet processes. Overall, these studies have shown that multijet processes are most likely to produce real muons and fake electrons.

In the following chapters, the $Z \to \tau\tau \to ll + 4\nu$ analysis uses data-driven methods to estimate the multijet background contribution (see 6.4). In the $e\mu$ channel, the method assumes that the charges of the two multijet leptons are independent. The results of this study supports this assumption, where approximately 88% of the $e\mu$ multijet events have at least one fake

lepton. In the $\mu\mu$ channel, the same assumption cannot be made, as over 50% of the multijet events come from real $b\bar{b}$ events. As such, the control regions in the $\mu\mu$ channel are defined using muons that fail the isolation selection instead.

In the $H \rightarrow \tau\tau \rightarrow ll + 4\nu$ analysis, the 'fake' background is estimated by modelling the fake distribution in a control region and fitting that shape in the signal region (see Section 8.4.4). This method requires the fake composition to be approximately the same in the signal and control regions and assumes the two leptons' charges are independent. To ensure similar fake compositions, only one lepton is allowed to fail the isolation requirement to minimize the composition bias towards heavy-flavoured quark decays. This selection is motivated by the results of this study, which shows heavy-flavoured quark decays fail the isolation requirements more so than fake leptons. The assumption on the two leptons' charge in dependence is supported by this study in the ee and $e\mu$ channels, but less so in the $\mu\mu$ channel. In the $H \rightarrow \tau\tau \rightarrow ll + 4\nu$ fake estimation, a larger discrepancy is observed in the $\mu\mu$ channel, which can be explained by the results presented.

6

$Z \rightarrow \tau\tau \rightarrow ll + 4\nu$ cross section

The measurement of the $Z \rightarrow \tau\tau$ production cross section is an important test for electroweak predictions at the LHC and for the $H \rightarrow \tau\tau$ analyses, where $Z \rightarrow \tau\tau$ events are an irreducible background. This chapter will describe the methods used to measure the $Z \rightarrow \tau\tau$ cross section using the $Z \rightarrow \tau\tau \rightarrow e\mu + 4\nu$ and $Z \rightarrow \tau\tau \rightarrow \mu\mu + 4\nu$ decay modes.

6.1 Trigger selection

The 15 GeV "medium" quality single electron trigger was used to collect the data for the $e\mu$ channel because it has the lowest p_T threshold of the unprescaled electron triggers.

The $\mu\mu$ channel uses a combination of 10 GeV, 13 GeV and 13 GeV "tight" single muon triggers, where the lowest threshold unprescaled trigger for each period of data taking was chosen. The integrated luminosities and the fraction of the collected data from each trigger are given in Table 6.1. In MC, events are selected with these triggers such that the fractions between the triggers matches those of the data.

Trigger	Lumi (pb^{-1})	Lumi (%)
μ, p_T > 10 GeV	3.02	8.2
μ, p_T > 13 GeV	15.3	42.9
μ, p_T > 13 GeV "tight"	17.3	48.8
Total	L = 35.5 pb^{-1}	

Table 6.1: Triggers used in the $\mu\mu$ analysis with the corresponding luminosities and fraction of total luminosity [10].

6.2 Particle selection

Selecting $Z \rightarrow \tau\tau \rightarrow ll + 4\nu$ events requires well reconstructed electrons, muons and missing transverse energy. These particles must be within the detector acceptance and are required to satisfy basic quality criteria.

Electron candidates are required to have an E_T > 16 GeV to ensure the candidate electrons are above the trigger threshold. The electron energy scale for data events is calibrated using $Z \rightarrow ee$ events. The calibrations are dependent on the calorimeter region with adjustments ranging between $\pm$ 5%. [60].

In the $\mu\mu$ channel, the leading muon is required to have p_T > 15 GeV to be above the trigger thresholds. The p_T threshold of the sub-leading muon is 10 GeV to maintain a high signal efficiency. For the $e\mu$ channel, a muon with p_T > 10 GeV is required for the same reason.

Jets are required to have E_T > 20 GeV. Descriptions of the calibrations used to obtain the jet energies are given in reference [68].

6.3 Event selection

This section describes the selections used for the $Z \rightarrow \tau\tau \rightarrow ll + 4\nu$ analysis.

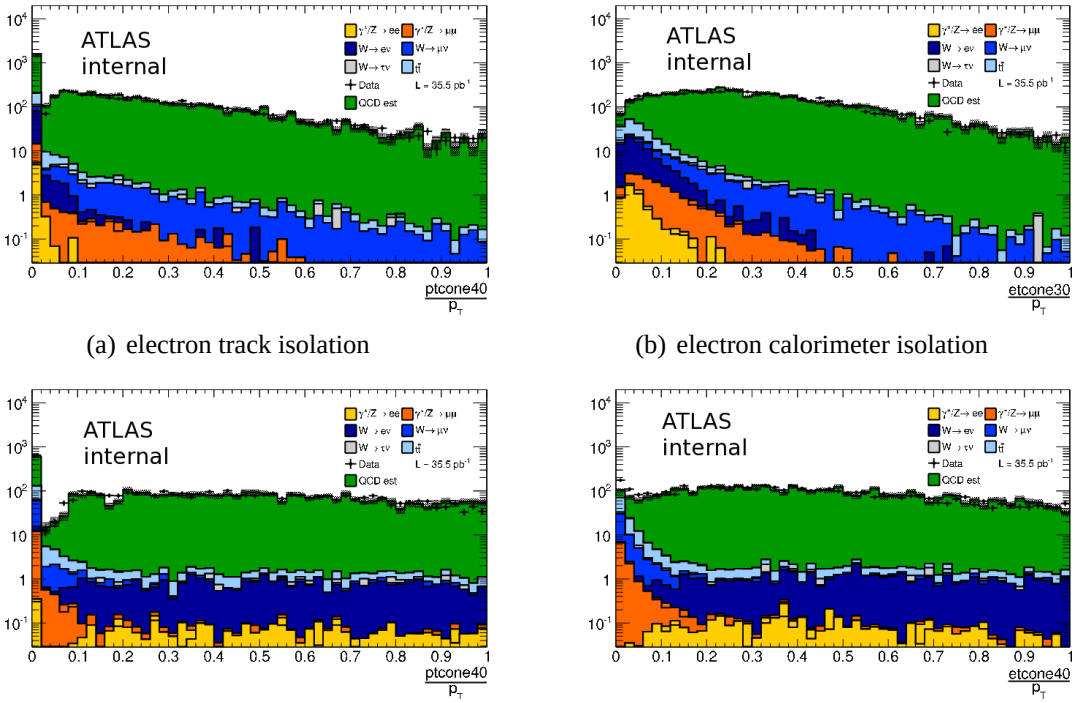
6.3.1 $Z \rightarrow \tau\tau \rightarrow e\mu + 4\nu$ selection

$Z \rightarrow \tau\tau \rightarrow e\mu + 4\nu$ candidate events are selected by requiring exactly one electron and one muon of opposite charge. Since the flavours of the two leptons are different, the contributions from $\gamma^*/Z \rightarrow ee$ and $\gamma^*/Z \rightarrow \mu\mu$ events are greatly reduced. However, contributions from multijet, $W \rightarrow l\nu$ and $t\bar{t}$ events are still significant.

Lepton isolation

After the $e\mu$ selection, there are approximately 20 times more multijet events than the $Z \rightarrow \tau\tau \rightarrow e\mu + 4\nu$ events. Leptons produced from multijet events tend to be not isolated. The measured tracking and calorimeter activities in a cone around the lepton candidates are effective discriminants against those produced from multijets. The following lepton isolation requirements are applied [10] and the distributions are shown in Figure 6.1:

- Muons: $p_T$cone40/p_T < 0.06 and $E_T$cone40/p_T < 0.06,
- Electrons: $p_T$cone40/p_T < 0.06 and $E_T$cone30/p_T < 0.10.



(a) electron track isolation

(b) electron calorimeter isolation

(c) muon track isolation

(d) muon calorimeter isolation

Figure 6.1: Data and MC comparison of the isolation variables after the $e\mu$ selection. The hatched areas represent the uncertainties of the MC and multijet estimations. Multijet contributions (QCD est.) are estimated from data (Section 6.4) [10].

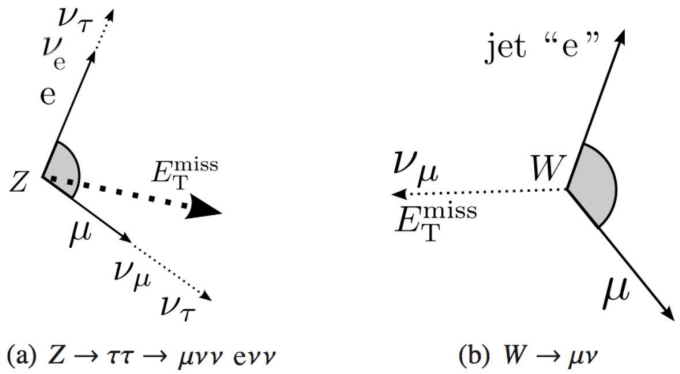
W + jets suppression

The angular separations between the two final state leptons and the E_T^{miss} direction are good discriminants against $W \to e\nu$ and $W \to \mu\nu$ events. Figure 6.2 shows the typical topologies of a $Z \to \tau\tau \to e\mu + 4\nu$ event compared to a $W \to l\nu$ event. In the case of the signal events, the two tauons produced from the Z decay are significantly boosted, which focuses the directions of the tauon's daughter particles. The E_T^{miss} of the signal is dominated by the transverse momenta of the four neutrinos and as such, the direction of the E_T^{miss} should lie mostly between the opening angles of the two leptons or at worst be roughly parallel with one of them. For the $W \to l\nu$ events, the second lepton is usually produced from additional jets. The transverse momenta of the jets, the charged lepton and the neutrino should roughly balance in the transverse plane. Since the neutrino dominates the E_T^{miss} for $W \to l\nu$ events, the E_T^{miss} direction usually lies outside the opening angle of the two charged leptons.

The difference in topology between the signal and $W \to l\nu$ events is exploited by using the variable

$$\sum cos\Delta\phi = cos(\phi(e) - \phi(E_T^{miss})) + cos(\phi(\mu) - \phi(E_T^{miss})). \quad (6.1)$$

Signal events have $\sum cos\Delta\phi$ values that peak at zero with a tail in the positive direction. Positive values of $\sum cos\Delta\phi$ are produced by events where the Z boson is boosted in the transverse direction. However, the majority of Z bosons are created almost at rest in the transverse direction, in which case the tauons and its decay products are nearly back-to-back. For these events, $\sum cos\Delta\phi$ is almost zero. For the $W \rightarrow l\nu$ events, $\sum cos\Delta\phi$ is usually negative because the average ϕ separation between the E_T^{miss} and the two leptons is > $\pi/2$ ($cos\Delta\phi$ < 0). Events are required to have $\sum cos\Delta\phi$ > -0.15, which selects the majority of signal events whether at rest or boosted and rejects most of the $W \rightarrow l\nu$ background.



(a) $Z \rightarrow \tau\tau \rightarrow \mu\nu\nu\ e\nu\nu$ (b) $W \rightarrow \mu\nu$

Figure 6.2: Topologies of a $Z \rightarrow \tau\tau \rightarrow e\mu + 4\nu$ event and a $W \rightarrow l\nu$ event [10].

Top suppression

$t\bar{t}$ events are characterised by the presence of a large number of high p_T jets, two leptons and large E_T^{miss}. The total energy in $t\bar{t}$ events are on average higher than $Z \rightarrow \tau\tau \rightarrow e\mu + 4\nu$ events. To reject this background the following variable is defined:

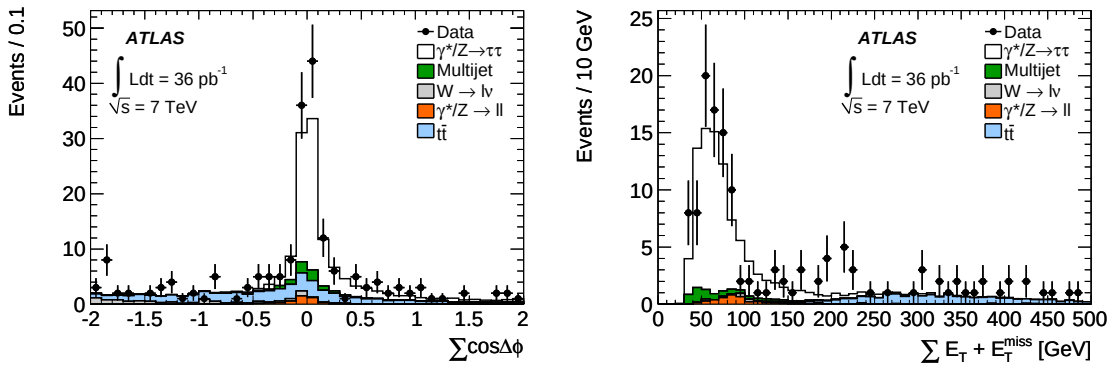
$$\sum E_T + E_T^{miss} = E_T(e) + E_T(\mu) + E(jets) + E_T^{miss}, \tag{6.2}$$

where the E_T summation includes all reconstructed muons, electrons and jets. Events are selected if they have $\sum E_T + E_T^{miss}$ < 150 GeV. This selection rejects over 98% of $t\bar{t}$ events while retaining over 90% of the signal [10].

Invariant Mass

The invariant mass of the electron and muon is required to be within 25 GeV < $m_{e\mu}$ < 80 GeV. This selection is chosen to include the majority of the signal events, while rejecting any events outside the signal spectrum.

The $\sum cos\Delta\phi$ after the $e\mu$ and isolation selections and the $\sum E_T + E_T^{miss}$ distributions after the $\sum cos\Delta\phi$ selection are shown in Figure 6.3. The $e\mu$ invariant mass distribution after the $\sum E_T + E_T^{miss}$ selection is shown in Figure 6.4.



(a) (b)

Figure 6.3: $\sum cos\Delta\phi$ distributions after the $e\mu$ and isolation selections and $\sum E_T + E_T^{miss}$ distributions after the $\sum cos\Delta\phi$ selection. Multijet contributions (QCD est.) are estimated from data (Section 6.4) [69].

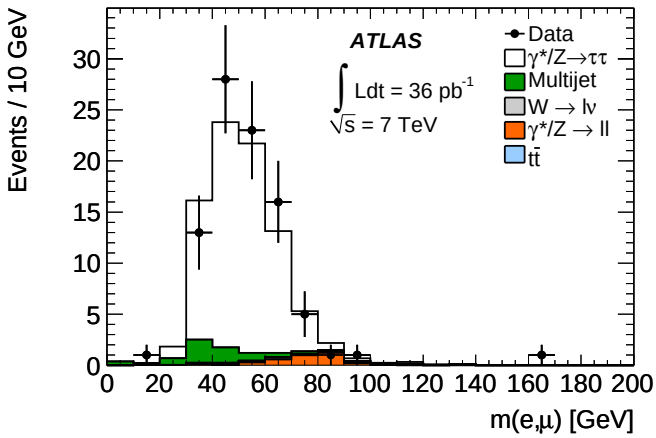


Figure 6.4: $m_{e\mu}$ distributions after all selections except for the invariant mass selection. Multijet contributions (QCD est.) are estimated from data (Section 6.4) [69].

6.3.2 $Z \to \tau\tau \to \mu\mu + 4\nu$ selection

The $Z \to \tau\tau \to \mu\mu + 4\nu$ candidate events are required to have exactly two muons of opposite charge.

Lepton isolation

The two selected muons are required to pass the following isolation requirements to reduce background contributions from multijet processes:

- $E_T$cone40/p_T < 0.20,
- $p_T$cone40/p_T < 0.15.

The isolation requirements are lower than the $e\mu$ channel because the multijet contributions are smaller in this channel. A rejection factor of approximately 100 is achieved against multijet events. The distributions of the isolation variables are shown in Figure C.3 (Appendix C) for events with exactly two muons of opposite charge.

Invariant mass selection

After the isolation criteria, the di-muon spectrum is dominated by $\gamma^*/Z \to \mu\mu$ and $\Upsilon \to \mu\mu$ events. To reduce these backgrounds, an invariant mass requirement of 25 < $m_{\mu\mu}$ < 65 GeV is applied. The upper and lower bounds are chosen to be away from the Z and Υ resonances. The invariant mass distribution of $\mu\mu$ events after isolation is shown in Figure 6.5.

$\gamma^*/Z \to \mu\mu$ suppression

After applying the invariant mass requirement, there are still approximately 200 times more $\gamma^*/Z \to \mu\mu$ events than $Z \to \tau\tau \to \mu\mu + 4\nu$ events. Muons from $\gamma^*/Z \to \mu\mu$ events have similar kinematics to those produced by the signal, making this background particularly challenging to remove. Two independent methods are used to reduce this background. The first uses a series of standard selections optimised to discriminate the signal from $\gamma^*/Z \to \mu\mu$ events. This technique, although robust, does not achieve the best possible signal to noise ratio. To improve this, the second method uses a multivariate technique, in the form of a Boosted Decision Tree (BDT), to distinguish between the two processes.

Standard selection method

In the standard method, the $\gamma^*/Z \to \mu\mu$ background is suppressed using the following requirements after the invariant mass selection:

- $\Delta\phi(\mu_1, \mu_2)$ > 2.7, to ensure the two muons are back-to-back in the transverse plane.

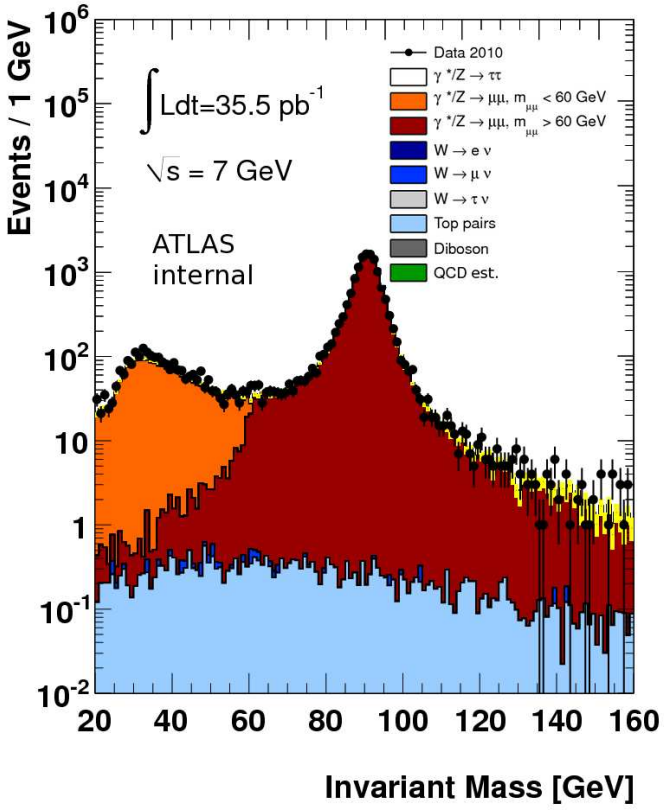


Figure 6.5: The invariant mass distribution for $\mu\mu$ events that pass the isolation selections. The yellow hatches indicate the uncertainty on the MC estimates. Multijet contributions (QCD est.) are estimated from data (Section 6.4) [10].

- $\Delta\phi(\mu_1, MET)$ > 2.7 or $\Delta\phi(\mu_1, MET) < 0.3$. The $\Delta\phi$ separation between the leading lepton and the direction of the E_T^{miss} is a good discriminant against $\gamma^*/Z \rightarrow \mu\mu$ events. Neutrinos from signal events travel approximately collinearly with the muons, so the E_T^{miss} direction is most likely to be either aligned or anti-aligned with the leading muon. For $\gamma^*/Z \rightarrow \mu\mu$ events, the E_T^{miss} is mostly fake and is expected to be orientated randomly with respect to the direction of the leading muon.

- $p_T(\mu_1) - p_T(\mu_2)$ > 5 GeV. After the two previous selections, only $\gamma^*/Z \rightarrow \mu\mu$ events with little or no boost in the transverse direction are left. For these events, the p_T of the two muons are similar. This differs from the signal events where the 3-body decay of the tauons allow the muons to have a larger p_T spread.

- $|d_0(\mu_1)|$ + $|d_0(\mu_2)|$ > 0.04 mm, where d_0 denotes the radial impact parameter of a muon with respect to the primary vertex. d_0 is larger for the signal muons because of the long tauon lifetime. The sum of the two muons' impact parameters provides a better discrimination against the background than either one alone.

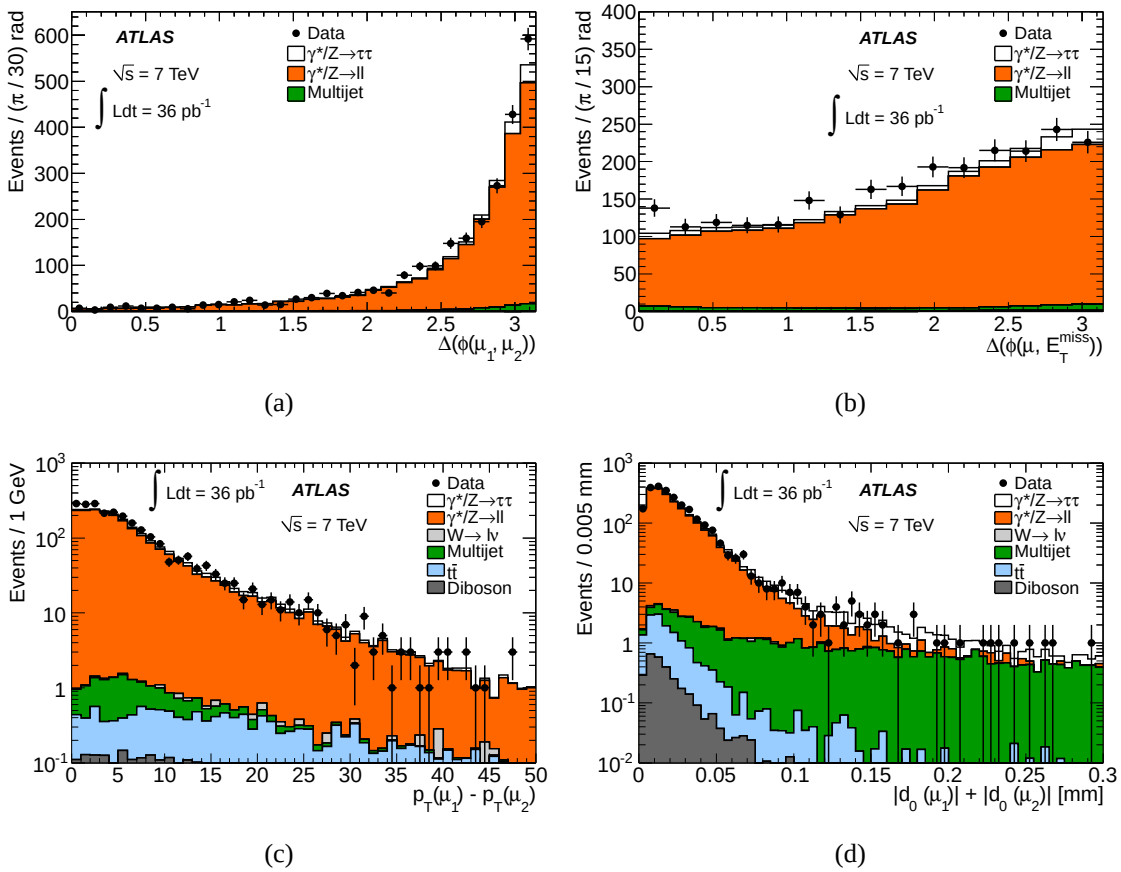


Figure 6.6: Data and MC comparisons of the $\gamma^*/Z \to \mu\mu$ suppression variables shown after the $\mu\mu$, isolation and invariant mass selections. Multijet contributions are estimated from data (Section 6.4)[69].

Distributions of the $\gamma^*/Z \to \mu\mu$ suppression variables are shown in Figure 6.6. The $\Delta\phi(\mu_1, \mu_2)$ and $|d_0(\mu_1)|$ + $|d_0(\mu_2)|$ distributions for the signal and $\gamma^*/Z \to \mu\mu$ events are shown specifically in Figure C.1 and the correlations of $\Delta\phi(\mu_1, \mu_2)$ with respect to the $\Delta\phi(\mu, E_T^{miss})$ and $P_T(\mu_1) - P_T(\mu_2)$ variables are shown in Figure C.2.

Boosted decision trees

Boosted Decision Trees (BDTs) are a multivariate analysis technique that uses a number of given discriminating variables to optimally separate signal and background processes [31]. Training events are separated into two sub-samples or *nodes* by applying a selection on the variable with the most discriminating power. This selection is optimised over 20 iterations. At each subsequent node, the same process occurs, creating more sub-nodes. A depiction of this process is shown in Figure 6.7, where a tree-like structure is formed. Node splitting continues until either the maximum number of levels has been reached; or there are insufficient number of events to create more nodes. In this analysis, nodes must contain at least 50 events and the trees have a maximum depth level of 3.

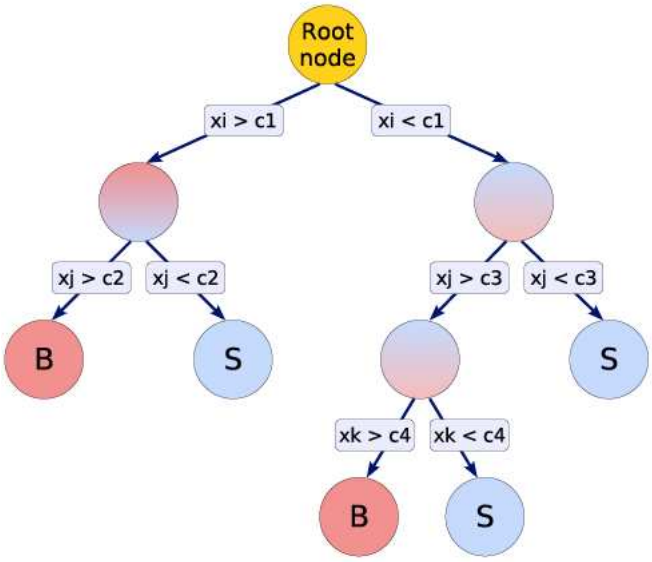


Figure 6.7: An illustration of a decision tree [31].

The nodes at the end of the trees which are referred to as "leaf nodes" and are used to classify events. The classification value is known as the "BDT output" or the "BDT response" and has a range [-1,1], where -1 indicates a purely background-like event and 1 a purely signal-like event. The classifications are determined by the number of signal and background events found within the leaf nodes during training.

Signal and background events from the training sample that are misclassified are given larger weights to help correctly classify them. The reweighted events are used to develop a new decision tree. This process is known as *boosting* and is typically applied several hundred times creating a "forest" of trees. In this analysis, 400 trees are used. Boosting is a key feature of BDTs that improves the signal and background separation and prevents bias classifications caused by statistical fluctuations. The Boosting for the tree used in this analysis is performed using the Adaptive Boost algorithm, which derives its boost weight using:

$$\alpha = \frac{1 - err}{err}, \tag{6.3}$$

where err is the misclassification rate of the previous tree. After applying the boost weights, the weights of the entire sample are renormalised such that the sum of the weights remains constant. Applying the Boost weights across all trees, the overall BDT classification of an event is given by:

$$y_{Boost}(x) = \frac{1}{N_{trees}} \cdot \sum_{i}^{N_{trees}} ln(\alpha_i) \cdot h_i(x), \tag{6.4}$$

where x denotes the vector of input variables, $h(x)$ is the result of an individual classifier and N_{trees} is the number of trees in the BDT.

Multivariate selection method

In the multivariate analysis method, a BDT is trained after the invariant mass selection to distinguish $Z \to \tau\tau \to \mu\mu + 4\nu$ events from $\gamma^*/Z \to \mu\mu$ events.

Training uses the $Z \to \tau\tau$ simulations as the signal and the $\gamma^*/Z \to \mu\mu$ simulations as the backgrounds. During the training and testing procedures, the isolation requirement is removed to increase statistics. Other background samples are not used in the training, as their inclusion dilutes the discriminating power of the BDT against $\gamma^*/Z \to \mu\mu$ events. The same four variables that are used in the standard method, along with the p_T of the leading muon are used to train the BDT:

- $\Delta\phi(\mu_1, \mu_2)$,
- $\Delta\phi(\mu_1, MET)$,
- $p_T(\mu_1) - p_T(\mu_2)$,
- $|d_0(\mu_1)| + |d_0(\mu_2)|$,
- $p_T(\mu_1)$.

About 50% of the signal and 30% of the background events are used in the training. This is about 4000 events for the signal and 10000 events for the backgrounds. The remainder of the MC events are used for testing the consistency of the BDT responses against the training sample. This check is conducted to ensure the BDT has not been over-trained. Figure 6.9 shows an overlay of the training and test samples for the signal and background, where a consistency between the training and test samples is observed. This is a strong indication that the BDT has not over-trained.

The purities and efficiencies of signal and background are shown in Figure 6.8 for the trained BDT. The $S/\sqrt{S+B}$ ratio is also shown as a function of the BDT output, where a maximum is observed at approximately 0.07. Events with a BDT output below this threshold are rejected.

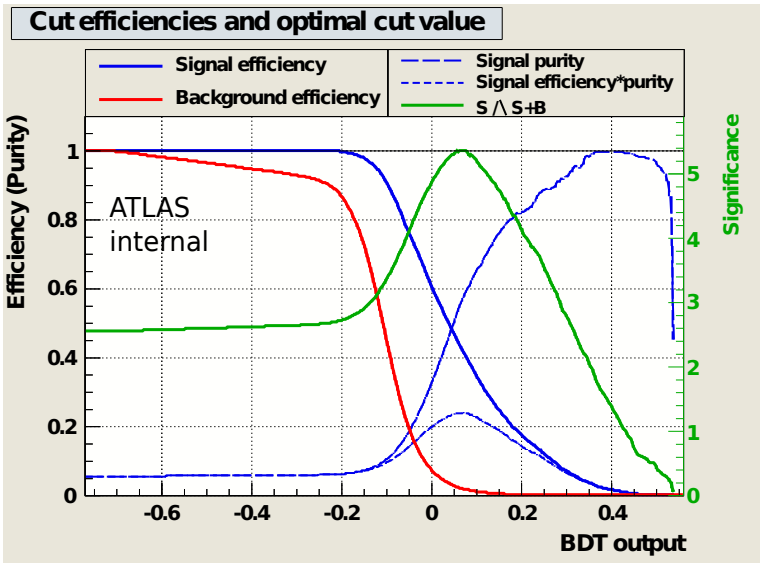


Figure 6.8: The efficiencies of the signal and background processes for the trained BDT [10].

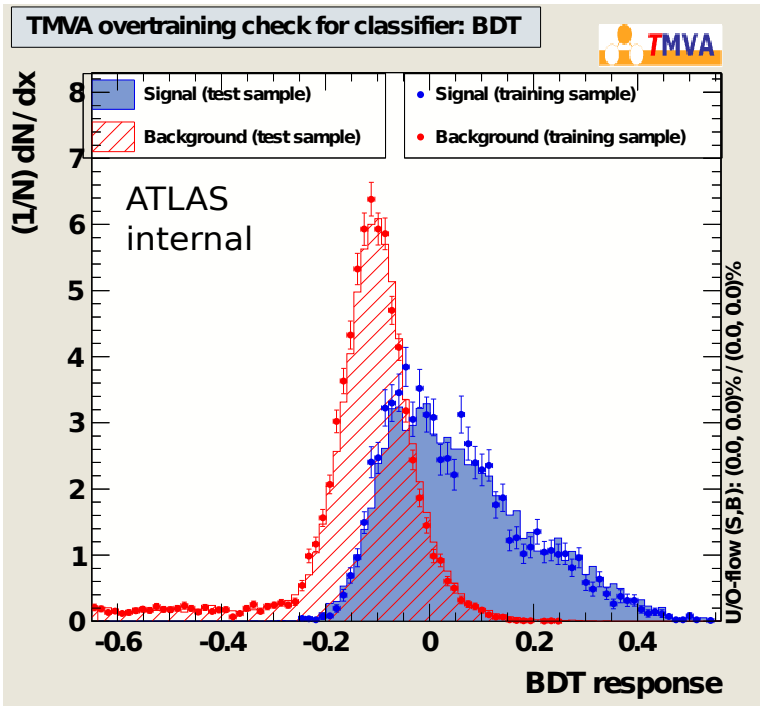


Figure 6.9: An overlay of the BDT responses between for the training and test samples [10].

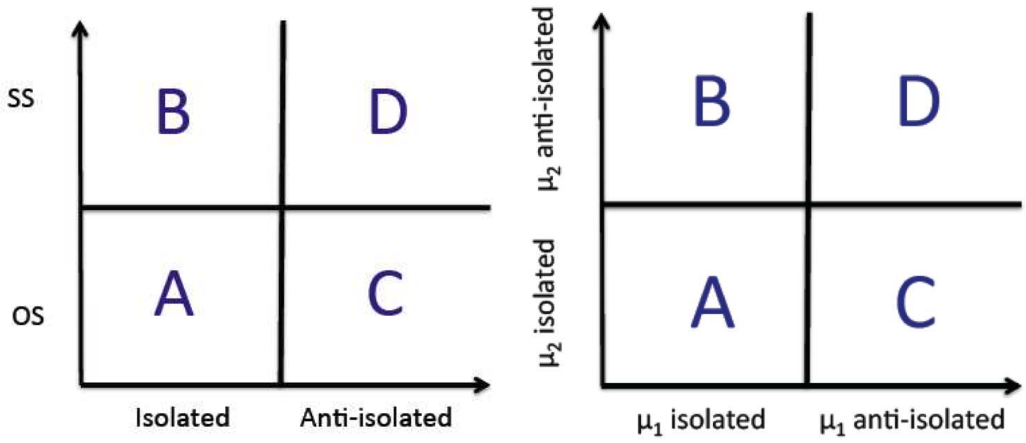
6.4 Estimating multijet contributions

The multijet background contribution in the $e\mu$ and $\mu\mu$ channels is estimated using data-driven methods. In both cases, three exclusive control regions with enriched multijet purities are defined to estimate the contributions in the signal region. The multijet purity can be enriched by choosing events with same signed leptons or leptons that fail the isolation selections. Leptons fitting the latter description are referred to as being "anti-isolated". The three control regions "B","C" and "D" are used to estimate the multijet contribution region "A". This technique is colloquially known as the "ABCD" method. A diagram of these regions is shown in Figure 6.10 for the $e\mu$ and $\mu\mu$ channels.

The contamination of electroweak and $t\bar{t}$ events in each control region is estimated using MC and then subtracted from the sample. The remaining events are assumed to be all multijet events.

$$N^i = N^i_{data} - N^i_{Z\to ll} - N^i_{t\bar{t}} - N^i_{W\to l\nu} - N^i_{diboson}, \tag{6.5}$$

where $l = e, \mu, \tau$.



(a) $e\mu$ regions (b) $\mu\mu$ regions

Figure 6.10: The four regions used to estimate the multijet background for the (a) $e\mu$ and (b) $\mu\mu$ channels [10].

6.4.1 ABCD method for $Z \to \tau\tau \to e\mu + 4\nu$

For the $e\mu$ channel, the four regions are defined after the dilepton selection using the following definitions:

- **Region A** is the region that contains events with an isolated electron and an isolated muon with opposite charges.
- **Region B** is the control region containing events with an isolated electron and an isolated muon with the same charges.
- **Region C** is the control region containing events with an anti-isolated electron and an anti-isolated muon with opposite charges.
- **Region D** is the control region containing events with an anti-isolated electron and an anti-isolated muon with the same charges.

A diagram of the four regions for the $e\mu$ channel is shown in Figure 6.10(a). The multijet isolation efficiencies are estimated using the ratio of events between region "D" and region "B". This efficiency is applied to region "C" to estimate the number of events in region "A". The selection efficiencies are assumed to be the same for same-sign and opposite-sign events. The estimated number of events is given by:

$$N^A = N^B \frac{N^C}{N^D} = N^B R_{OS/SS}, \tag{6.6}$$

where $R_{OS/SS}$ is the ratio of opposite sign and same sign events, which is measured to be 1.55 ± 0.04.

N^A is an estimate of the contributions after the $e\mu$ and isolation selections. Due to the limited statistics in region "B", the multijet contribution after the full selection is estimated by applying an efficiency factor to N^A. This efficiency is derived by passing events in region "D" through the analysis selections of the $e\mu$ channel. Assuming the events in region "A" and "D" have the same selection efficiencies, the number of multijet events in the signal region is estimated by:

$$N^{multijet}_{signal} = N^A \times \varepsilon^D_{signal}. \tag{6.7}$$

The validity of this assumption is checked by comparing the distributions of the selection variables in the two regions. However, the number of events in region "A" is statistically limited, so this assumption cannot be checked directly. Instead two additional regions are defined where only one of the two leptons are required to be isolated. These intermediate

Region C:	Anti-isolated OS	RegionD:	Anti-Isolated SS
$\gamma^*/Z(ee)$	0.25 ± 0.10	$\gamma^*/Z(ee)$	0.08 ± 0.03
$\gamma^*/Z(\mu\mu)$	0.28 ± 0.04	$\gamma^*/Z(\mu\mu)$	0.26 ± 0.04
$W(e\nu)$	1.72 ± 0.33	$W(e\nu)$	0.40 ± 0.15
$W(\mu\nu)$	4.07 ± 0.43	$W(\mu\nu)$	3.48 ± 0.39
$W(\tau\nu)$	0.75 ± 0.37	$W(\tau\nu)$	0.72 ± 0.36
$t\bar{t}$	4.57 ± 0.15	$t\bar{t}$	3.15 ± 0.12
Diboson	0.11 ± 0.01	Disboson	0.05 ± 0.01
$\gamma^*/Z(\tau\tau)$	0.61 ± 0.09	$\gamma^*/Z(\tau\tau)$	0.51 ± 0.14
Sum MC	12.36 ± 0.69	Sum MC	8.65 ± 0.58
Data	3771	Data	2432
Multijet	3758.64	Multijet	2423.35
Region A:	Isolated OS	Region B:	Isolated SS
$\gamma^*/Z(ee)$	0.06 ± 0.02	$\gamma^*/Z(ee)$	0.05 ± 0.03
$\gamma^*/Z(\mu\mu)$	5.47 ± 0.21	$\gamma^*/Z(\mu\mu)$	2.81 ± 0.15
$W(e\nu)$	0.22 ± 0.11	$W(e\nu)$	0.11 ± 0.07
$W(\mu\nu)$	9.16 ± 0.64	$W(\mu\nu)$	9.55 ± 0.66
$W(\tau\nu)$	0.98 ± 0.44	$W(\tau\nu)$	0.52 ± 0.26
$t\bar{t}$	48.67 ± 0.50	$t\bar{t}$	0.38 ± 0.04
Diboson	9.11 ± 0.13	Diboson	0.55 ± 0.02
$\gamma^*/Z(\tau\tau)$	86.93 ± 1.17	$\gamma^*/Z(\tau\tau)$	1.44 ± 0.17
Sum MC	160.60 ± 1.54	Sum MC	15.41 ± 0.75
Data	190	Data	22
Multijet	29.4	Multijet	6.59

Table 6.2: The number of data and MC events for the four regions of the multijet background estimation after the $e\mu$ selection [10].

regions have higher statistics, so the independence of the isolation and charge with respect to the other selection variables can be tested.

The distributions of the event selection variables for the intermediate regions and the control regions "C" and "D" are shown in Figure 6.11. The shapes of these distributions are roughly consistent across the four shown regions, given the uncertainties. This indicates that the selections of the $e\mu$ channel are approximately independent of the charge and isolation. Thus, the selection efficiencies derived from region "D" can be used estimate the multijet contributions in the signal region. Any observed shape differences between the control regions are accounted for with systematic uncertainties and is discussed in Section 6.7.4. Region "D" is also used to model the shapes of the multijet distributions in all plots.

The number of events in each region is given in Table 6.2. The estimated number of multijet events after the $e\mu$ and isolation selections is

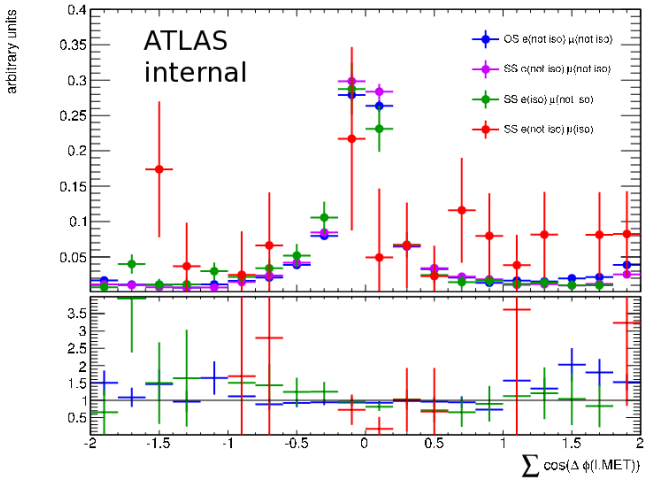
$$N^A = 10.2 \pm 7.4(stat.) \pm 2.5(syst.). \tag{6.8}$$

The efficiency of the remaining selections is measured to be

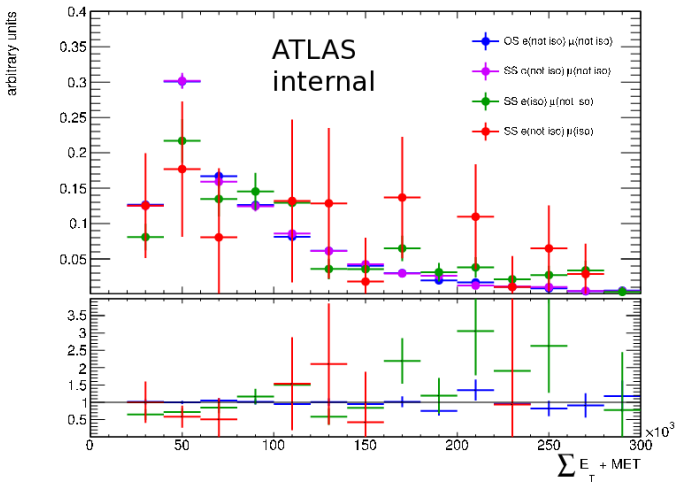
$$\varepsilon^D_{signal} = \frac{N^D_{signal}}{N^D_{dilepton}} = 0.56 \pm 0.01(stat.) \pm 0.01(syst.). \tag{6.9}$$

Thus, the expected number of multijet events in the $e\mu$ signal region is estimated to be

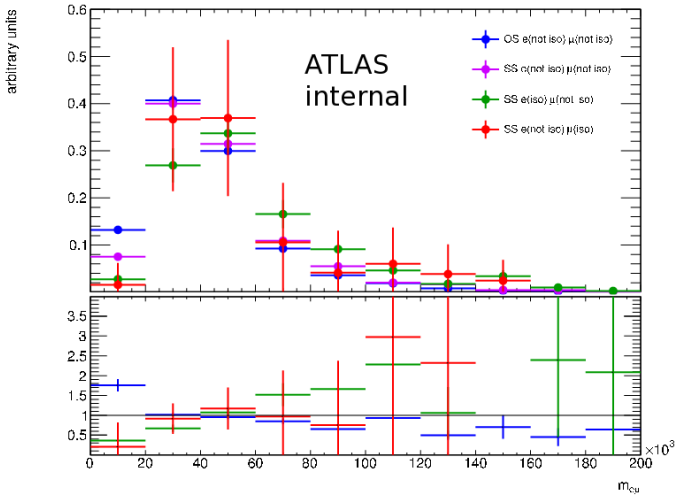
$$N^A_{signal} = 5.7 \pm 4.1(stat.) \pm 0.9(syst.). \tag{6.10}$$



(a)



(b)



(c)

Figure 6.11: The shape distributions for the variables used in the $e\mu$ event selections. The intermediate regions withs region "C" and "D" are shown for each variable [10].

6.4.2 ABCD method for $Z \rightarrow \tau\tau \rightarrow \mu\mu + 4\nu$

The ABCD method is used in the $\mu\mu$ channel at two different points of the event selections: the first after the dilepton and isolation criteria; and the second after the full selection. The first estimate is used to check the normalisation of the $\gamma^*/Z \rightarrow \mu\mu$ background (Section 6.5) and the second is used to estimate the multijet contributions in the signal region. The four regions for the $\mu\mu$ channel is defined after the dilepton selection using the following definitions:

- **Region A** is the region that contains events where both muons are isolated.
- **Region B** is the control region where the leading muon is isolated and the sub-leading muon is anti-isolated.
- **Region C** is the control region where the leading muon is anti-isolated and the sub-leading muon is isolated.
- **Region D** is the control region where both muons are anti-isolated.

A diagram of the ABCD regions for the $\mu\mu$ channel is shown in Figure 6.10(b). For the estimate after the full selections, regions "A" and "C" are modified to additionally include the full event selections.

Regions "B" and "D" are used to measure the isolation efficiency of the leading muon, which is applied to region "C" to estimate the number of multijet events in region "A". Any correlations found between the control regions are accounted for using a k factor. Thus, the number of multijet events in region "A" is given by:

$$N^A = N^C \left(\frac{N^B}{N^D} \right) \times k. \tag{6.11}$$

At the first multijet estimate, k corrects for the correlation between the two muon's isolation efficiencies. The $E_T$cone40/p_T and $p_T$cone40/p_T efficiencies are measured for both muons for when the other is isolated and anti-isolated. These results are shown in Table C.1 and their distributions are shown in Figure C.4. The four measured k values are averaged to obtain a single k factor. The systematic uncertainty on k is taken as the largest difference between any of the measured k values.

The k factors are similarly calculated for the multijet estimation in the $\mu\mu$ signal region for the standard and multivariate analysis methods. The $E_T$cone40/p_T and $p_T$cone40/p_T efficiencies of the leading muon are measured for events that pass the full event selections. The two k values measured for the signal region are averaged to obtain to a single k factor with a systematic uncertainty that's equal to the difference between the two. The measured

values are shown in Table 6.3 and the $E_T$cone40/p_T and $p_T$cone40/p_T distributions for these events are shown in Figure 6.12.

The estimated multijet contributions in signal regions are given in Table 6.4 for the standard and multivariate analysis methods. The number of events in each of the four ABCD regions after the $\mu\mu$ selection are shown in Table 6.5 and their distributions are shown in Figure C.5. For the multijet estimate in the signal regions, the number of events in the modified region "C" is presented in Table 6.6.

Isolation Variable	Pass standard selections and >1 iso μ	μ_2 not isolated	k factor
$\varepsilon(\mu_1)$ E_TCone40/p_T	0.279 ± 0.072	0.180 ± 0.006	1.55 ± 0.41
$\varepsilon(\mu_1)$ p_TCone40/p_T	0.200 ± 0.060	0.136 ± 0.004	1.47 ± 0.45
Average k factor = 1.51 ± 0.61 (stat) ± 0.08 (sys)			
Isolation Variable	Pass BDT selections and >1 iso μ	μ_2 not isolated	k factor
$\varepsilon(\mu_1)$ E_TCone40/p_T	0.167 ± 0.021	0.180 ± 0.006	0.93 ± 0.15
$\varepsilon(\mu_1)$ p_TCone40/p_T	0.127 ± 0.019	0.136 ± 0.004	0.93 ± 0.16
Average k factor = 0.93 ± 0.22 (stat) ± 0.01 (sys)			

Table 6.3: The isolation selection efficiencies for the leading muon for events that pass all other $\mu\mu$ event selections for the standard (top) and multivariate (bottom) analysis methods. The ratio of these efficiencies are given as k factors [10].

After selection	Multijet estimate ($\pm\sigma_{stat}$ $\pm\sigma_{sys}$)
Invariant mass $m_{\mu\mu}$	$69.3 \pm 11.7 \pm 10.7$
Standard selection	$1.8 \pm 0.8 \pm 0.1$
BDT selection	$10.2 \pm 2.5 \pm 0.7$

Table 6.4: Estimated number of multijet events in the $\mu\mu$ channel after the selection on the invariant mass selection and after the full event selections for the standard and multivariate analysis methods [10].

Region B:	μ_1 not isolated, μ_2 isolated	Region D:	μ_2 not isolated, μ_1 not isolated
Data	766	Data	6550
$Z \to \tau\tau$	$9.2 \pm 0.4 \pm 0.4$	$Z \to \tau\tau$	$0.4 \pm 0.1 \pm 0.0$
$W \to \mu\nu$	$43.9 \pm 1.5 \pm 3.0$	$W \to \mu\nu$	$2.3 \pm 0.4 \pm 0.3$
$t\bar{t}$	$22.7 \pm 0.6 \pm 1.8$	$t\bar{t}$	$4.3 \pm 0.2 \pm 0.4$
$\gamma^*/Z \to \mu\mu$	$231.4 \pm 3.3 \pm 15.0$	$\gamma^*/Z \to \mu\mu$	$9.5 \pm 0.8 \pm 0.7$
$W \to \tau\nu$	$2.7 \pm 0.7 \pm 0.1$	$W \to \tau\nu$	$1.2 \pm 0.5 \pm 0.1$
$\gamma^*/Z \to ee$	0.0	$\gamma^*/Z \to ee$	$0.2 \pm 0.1 \pm 0.0$
$W \to e\nu$	0.0	$W \to e\nu$	0.0
Multijet	$456.1 \pm 21.7 \pm 15.4$	Multijet	$6532.1 \pm 80.8 \pm 0.9$
Region A:	μ_1 & μ_2 isolated	Region C:	μ_1 isolated, μ_2 not isolated
Data	2491	Data	366
$Z \to \tau\tau$	$115.1 \pm 1.4 \pm 8.0$	$Z \to \tau\tau$	$3.2 \pm 0.2 \pm 0.2$
$W \to \mu\nu$	$1.5 \pm 0.3 \pm 0.2$	$W \to \mu\nu$	$4.0 \pm 0.4 \pm 0.8$
$t\bar{t}$	$12.4 \pm 0.4 \pm 1.1$	$t\bar{t}$	$4.9 \pm 0.3 \pm 1.0$
$\gamma* /Z \to \mu\mu$	$2137.2 \pm 10.0 \pm 148.7$	$\gamma* /Z \to \mu\mu$	$49.0 \pm 1.7 \pm 6.3$
$W \to \tau\nu$	0.0	$W \to \tau\nu$	$0.2 \pm 0.2 \pm 0.0$
$\gamma* /Z \to ee$	0.0	$\gamma* /Z \to ee$	0.0
$W \to e\nu$	0.0	$W \to e\nu$	0.0
Multijet	$224.8 \pm 51.1 \pm 149.0$	Multijet	$304.7 \pm 17.5 \pm 6.4$

Table 6.5: The number of events in each region for the $\mu\mu$ multijet estimation method [10].

	Region C: Standard selection	Region C: BDT selection
Data	18	108
$Z \to \tau\tau$	$0.5 \pm 0.1 \pm 0.0$	$1.1 \pm 0.1 \pm 0.2$
$W \to \mu\nu$	$0.1 \pm 0.1 \pm 0.1$	$0.3 \pm 0.1 \pm 0.2$
$t\bar{t}$	$0.1 \pm 0.03 \pm 0.0$	$1.4 \pm 0.1 \pm 0.2$
$\gamma* /Z \to \mu\mu$	$0.5 \pm 0.2 \pm 0.1$	$0.7 \pm 0.2 \pm 0.2$
$W \to \tau\nu$	0.0	0.0
$\gamma* /Z \to ee$	0.0	0.0
$W \to e\nu$	0.0	0.0
Multijet	$16.8 \pm 4.1 \pm 0.2$	$156.5 \pm 12.5 \pm 0.4$

Table 6.6: The number of events for the $\mu\mu$ channel in region "C" of the the multijet estimation method, for events that pass the full $\mu\mu$ event selections barring the isolation of one muon [10].

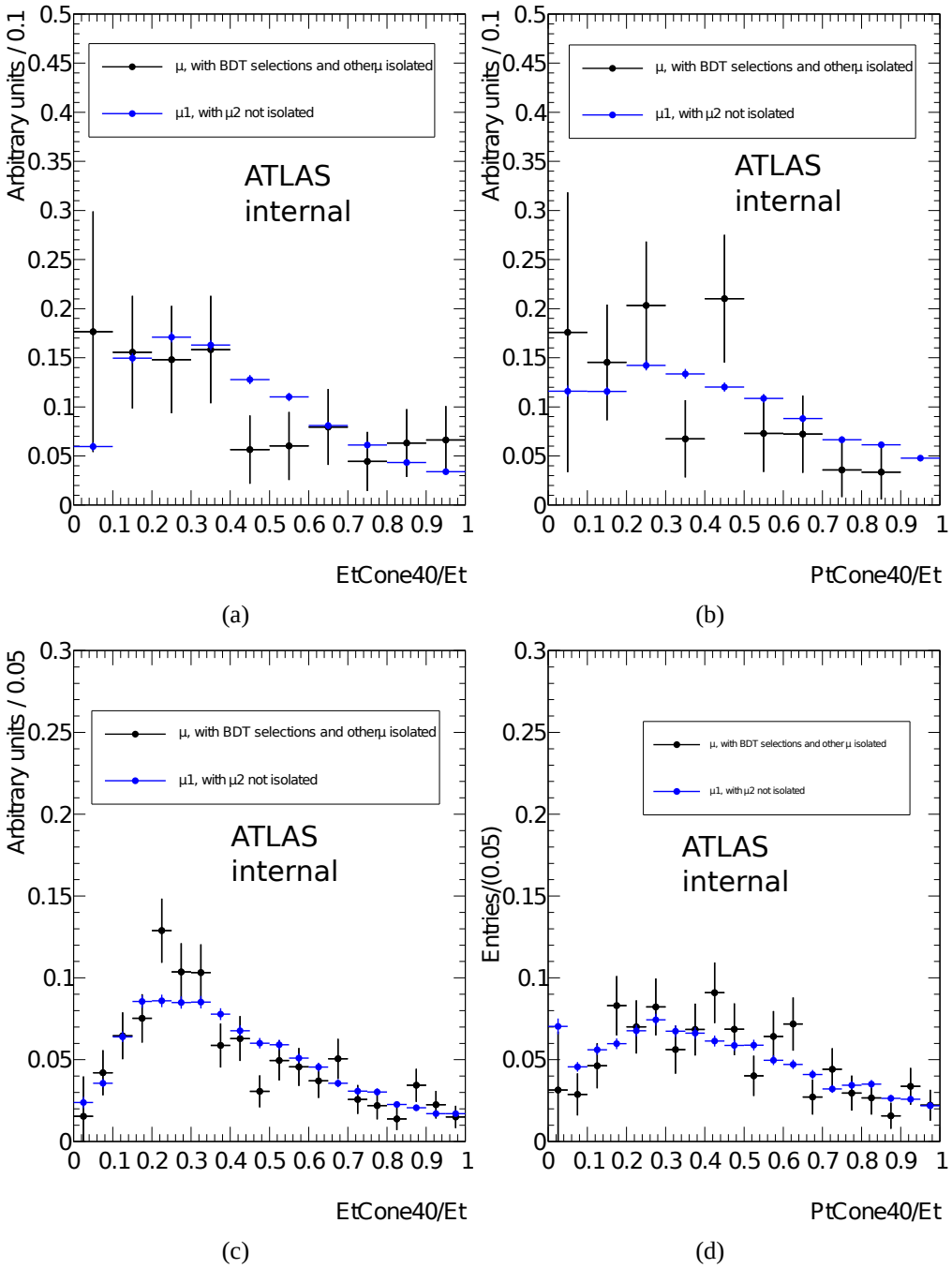


Figure 6.12: Distributions of the $\mu\mu$ multijet events in data for the E_TCone40/p_T and p_TCone40/p_T in the different ABCD regions after all selections in both the standard and multivariate analysis methods. The distributions have been normalised to unity for viewing purposes [10].

6.5 Estimation of electroweak backgrounds

The background contribution from W, γ^*/Z and $t\bar{t}$ events is estimated using simulations. The production cross sections used for each sample is checked in control regions using data.

6.5.1 W normalisation

The expected cross section of $W \to \mu\nu$ and $W \to e\nu$ events are checked using control regions defined by the following selections:

- Exactly one muon and one electron of opposite charge,
- The muon ($W \to \mu\nu$ region) or the electron ($W \to e\nu$ region) satisfies the isolation criteria,
- Isolation requirements for the other lepton are not applied,
- $\sum E_T + E_T^{miss}$ < 150 GeV,
- 60 < M_T < 100 GeV,

where M_T is the transverse mass:

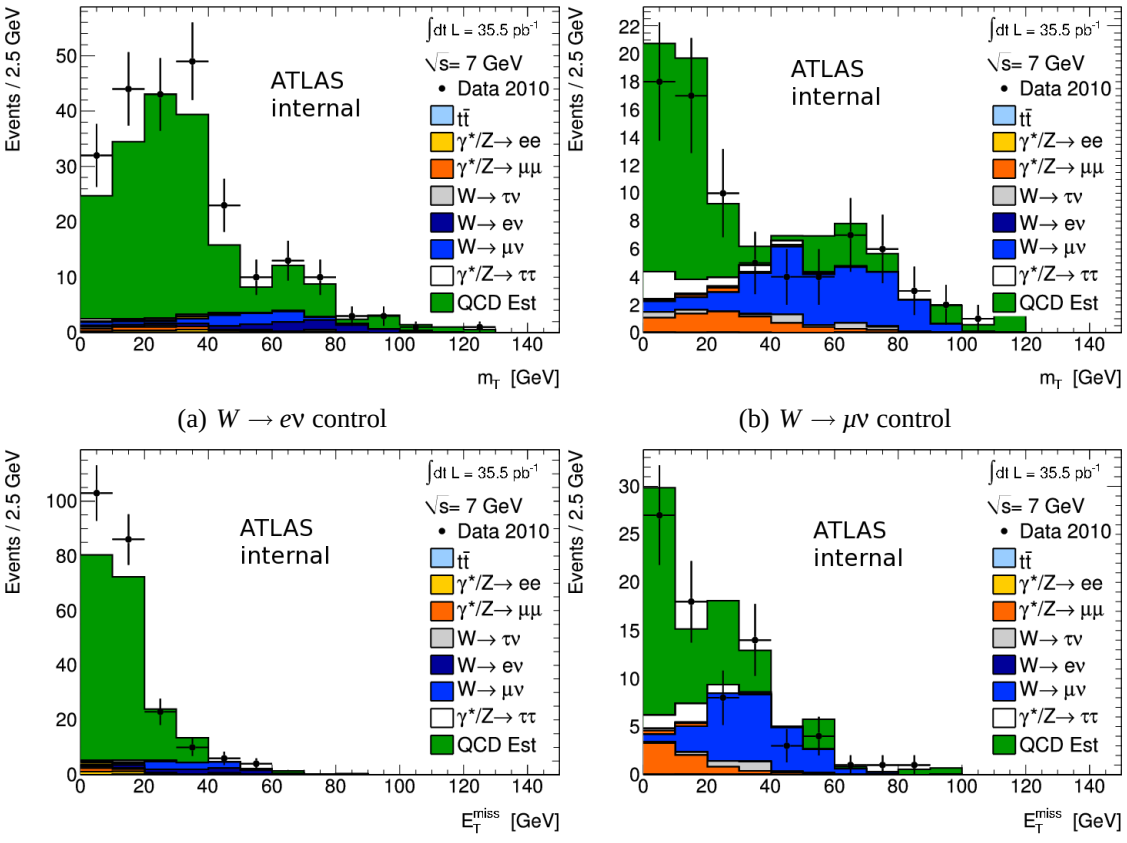
$$M_T = \sqrt{2p_{T,l}E_T^{miss}(1 - cos\Delta(l, E_T^{miss}))}. \quad (6.12)$$

$W \to l\nu$ + jets events can fake the signal if one of the associated jets is misidentified as a lepton. By removing the isolation requirements for one of the lepton candidates, the signal contamination is reduced and the amount of $W \to l\nu$ events is increased. The $\sum E_T + E_T^{miss}$ requirement reduces the number of $t\bar{t}$ events within the control regions. The M_T distribution of $W \to l\nu$ events peaks around the mass of the $W^{\pm}$ bosons and by applying a selection on M_T around this mass, the purity $W \to l\nu$ events is further increased.

Table 6.7 shows the number of events within the $W \to e\nu$ and $W \to \mu\nu$ control regions, respectively. The number of multijet events within these regions is estimated using the ABCD method. The distributions of the M_T and E_T^{miss} variables for these regions are shown in Figure 6.13. The shapes of the distributions and the selection efficiencies of the multijet events are taken from region "D". Within the uncertainties, the $W \to l\nu$ estimates from simulations are consistent with the number of observed events in data.

	$W \to e\nu$ control region		$W \to \mu\nu$ control region	
	N_{TOT}	$N_{60<M_T<100GeV}$	N_{TOT}	$N_{60<M_T<100GeV}$
$Z \to ee$	2.39 ± 0.16	0.32 ± 0.06	0.04 ± 0.02	0.01 ± 0.01
$Z \to \mu\mu$	2.75 ± 0.14	0.10 ± 0.02	6.75 ± 0.22	0.59 ± 0.06
$W \to e\nu$	8.51 ± 0.77	5.24 ± 0.62	0.11 ± 0.07	0.05 ± 0.05
$W \to \mu\nu$	9.15 ± 0.68	2.81 ± 0.39	25.28 ± 1.19	10.69 ± 0.78
$W \to \tau\nu$	1.25 ± 0.53	0.40 ± 0.32	2.23 ± 0.64	0.62 ± 0.35
$t\bar{t}$	0.37 ± 0.05	0.12 ± 0.03	0.06 ± 0.02	0.02 ± 0.01
$\gamma* \to ee$	1.05 ± 0.24	0.02 ± 0.02	0.02 ± 0.02	0.00 ± 0.00
$\gamma* \to \mu\mu$	0.14 ± 0.07	0.00 ± 0.00	0.81 ± 0.21	0.00 ± 0.00
$\gamma* \to \tau\tau$	0.09 ± 0.09	0.00 ± 0.00	0.32 ± 0.17	0.00 ± 0.00
$Z \to \tau\tau$	2.12 ± 0.20	0.18 ± 0.06	4.63 ± 0.30	0.11 ± 0.05
Mulitjet	167.40 ± 14.03	17.32 ± 1.45	48.55 ± 9.79	5.17 ± 1.04
Data	232	29	77	18
Total	195.2 ± 14.08	26.1 ± 1.66	88.78 ± 9.89	17.28 ± 1.35

Table 6.7: Number of events in the $W \to l\nu$ control regions for data and MC [10].



(a) $W \to e\nu$ control

(b) $W \to \mu\nu$ control

(c) $W \to e\nu$ control

(d) $W \to \mu\nu$ control

Figure 6.13: Distributions of M_T and E_T^{miss} variables for the $W \to l\nu$ control regions. Multijet contributions are shown in green (QCD est) [10].

6.5.2 Top normalisation

The expected $t\bar{t}$ cross section is checked with a control region defined as follows:

- One electron and one muon of opposite charge,
- Both leptons pass isolation,
- $\sum cos\Delta\phi$ < -0.15,
- $\sum E_T + E_T^{miss}$ > 150 GeV.

The $\sum cos\Delta\phi$ requirement reduces the signal contamination in the control region. Since $t\bar{t}$ events contain more energy than the signal and other backgrounds, the $\sum E_T + E_T^{miss}$ > 150 GeV requirement is used to improve the $t\bar{t}$ purity. The multijet contribution in the $t\bar{t}$ control region is estimated using the same methods as described previously.

The number of events in the $t\bar{t}$ control region is shown in Table 6.8 for data and MC. The observed events agree with the MC estimates within the uncertainties.

	N_{TOT}
$\gamma*/Z \rightarrow ll\ (l = e, \mu)$	0.19 ± 0.04
$W \rightarrow e\nu$	0.08 ± 0.07
$W \rightarrow \mu\nu$	1.29 ± 0.26
$W \rightarrow \tau\nu$	0.33 ± 0.25
$t\bar{t}$	26.96 ± 0.38
$\gamma*/Z \rightarrow \tau\tau$	0.31 ± 0.07
Multijet	2.30 ± 2.75
Data	31
Sum MC	29.18 ± 2.80

Table 6.8: Number of events in the $t\bar{t}$ control region for data and MC [10].

6.5.3 γ^*/Z normalisation

The expected $\gamma^*/Z \rightarrow \mu\mu$ production is checked using events that pass the $\mu\mu$, isolation and 25 < $m_{\mu\mu}$ < 65 GeV selections. In this region, over 90% of the events are expected to come from the $\gamma^*/Z \rightarrow \mu\mu$ process (see Figure 6.5), which makes it suitable for checking the scale of this background.

The multijet contribution in this region is estimated as described previously (Section 6.4). Other electroweak and $t\bar{t}$ contributions are estimated from simulations. The number of events measured in this region is shown in Table 6.9 for the data and MC. A slight excess of about 7% is observed in the data. However, this excess is consistent within one standard deviation of the statistical and systematic uncertainties of the $\gamma^*/Z \to \mu\mu$ estimate, which is approximately 9%.

Process	Number of events $\pm\sigma_{stat} \pm \sigma_{sys}$
$Z \to \tau\tau$	$115.1 \pm 1.4 \pm 8.0$
$W \to \mu\nu$	$1.5 \pm 0.3 \pm 0.2$
$t\bar{t}$	$12.4 \pm 0.4 \pm 1.1$
Multijet	$69.3 \pm 11.7 \pm 10.7$
Data	2491
γ^*/Z MC	$2137.2 \pm 11.1 \pm 148.7$
γ^*/Z data	$2292.7 \pm 51.3 \pm 6.1$
Ratio data/MC	$1.07 \pm 0.02 \pm 0.08$

Table 6.9: Comparison between data and predicted MC yields for $\gamma^*/Z \to \mu\mu$ in the $\mu^+\mu^-$ channel [10].

6.6 Cross section methodology

The measurement of the $Z \to \tau\tau \to ll + 4\nu$ cross section is made in both the fiducial and total phase space for events satisfying 66 < $m_{\tau\tau}$ < 116 GeV. The cross sections are calculated as follows:

$$\sigma^{tot.}_{\gamma^*/Z} \times BR(\gamma^*/Z \to \tau\tau) = \frac{N_{obs} - N_{bkg}}{A_Z \cdot C_Z \cdot \mathcal{L}}, \tag{6.13}$$

$$\sigma^{fid.}_{\gamma^*/Z} \times BR(\gamma^*/Z \to \tau\tau) = \frac{N_{obs} - N_{bkg}}{C_Z \cdot \mathcal{L}}, \tag{6.14}$$

where

- N_{obs} is the number of observed candidate events in data
- N_{bkg} is the number of estimated background events

- A_Z is the kinematic and geometric acceptance for the signal process. It is determined using truth information from signal MC events and is defined as:

$$A_Z = \frac{N_{dressed}^{gen\ kin}}{N_{Born}^{gen\ m_{inv}}}, \tag{6.15}$$

  where $N_{Born}^{gen\ m_{inv}}$ is the number of truth events that are generated at the Born level with an invariant mass, $m_{\tau\tau}$, between 66 - 116 GeV. $N_{dressed}^{gen\ kin}$ denotes the number of $N_{Born}^{gen\ m_{inv}}$ events that also fall within the defined fiducial regions (see below).

  The impact of QED final state radiation on the acceptance of the τ leptons is investigated using a dedicated sample where the PHOTOS package was switched off. The effect of QED final state radiation on the acceptance is measured to be -1.2% in the $e\mu$ channel and -1.3% in the $\mu\mu$ channel [10].

  The central values for A_Z are measured from the default PYTHIA $\gamma^*/Z \to \tau\tau$ ($m_{\tau\tau}$ < 60 GeV and $m_{\tau\tau}$ > 60 GeV) samples [10] and their values are shown in Table 6.10.

- C_Z is the correction factor that accounts for the acceptance and reconstruction efficiencies of the detector. It is defined as:

$$C_Z = \frac{N^{reco\ pass}}{N_{dressed}^{gen\ kin}}, \tag{6.16}$$

  where $N^{reco\ pass}$ is the number of MC signal events that pass the event selections of the analysis. Events that pass the selections but have an invariant mass $m_{\tau\tau}$ outside the defined window are excluded.

- $\mathcal{L}$ is the integrated luminosity of the data used in the analysis.

For a fiducial cross section measurement, A_Z is set to unity because an extrapolation of the full $Z \to \tau\tau$ phase space is not required. Fiducial cross sections are often used because they are less affected by theoretical uncertainties, which are usually considered as systematic uncertainties on A_Z.

The fiducial regions for the $e\mu$ channel are:

- Electron: $E_T > 16$ GeV, $|\eta| < 2.47$, excluding $1.37 < |\eta| < 1.52$,
- Muon: $p_T > 15$ GeV, $|\eta| < 2.4$,
- Event: $\sum \cos\Delta\phi > -0.15$,
- $25 < m_{e\mu} < 80$ GeV.

The fiducial regions are defined separately for the two analysis methods in the $\mu\mu$. For the standard analysis they are:

- $p_T(\mu_1) > 15GeV$, $|\eta| < 2.4$,
- $p_T(\mu_2) > 10GeV$, $|\eta| < 2.4$,
- $25GeV < m_{\mu_1,\mu_2} < 65GeV$,
- $\Delta\phi(\mu_1,\mu_2) > 2.7$,
- $\Delta\phi(\mu_1,MET) > 2.7$,
- $p_T(\mu_1) - p_T(\mu_2) > 5GeV$.

For the multivariate analysis they are:

- $p_T(\mu_1) > 15GeV$, $|\eta| < 2.4$,
- $p_T(\mu_2) > 10GeV$, $|\eta| < 2.4$,
- $25GeV < m_{\mu_1,\mu_2} < 65GeV$.

	$e\mu$ channel	$\mu\mu$ channel (Standard)	$\mu\mu$ channel (BDT)
A_Z	0.1139	0.0488	0.1557
C_Z	0.2857	0.3607	0.2661

Table 6.10: Central values of A_Z and C_Z for the $e\mu$ and $\mu\mu$ channels [10].

6.7 Systematic uncertainties

This section details the effects of systematic uncertainties on the measurement of the $\gamma^*/Z \rightarrow \tau\tau \rightarrow ll + 4\nu$ cross section. The effects considered include the uncertainties of A_Z and C_Z for the signal; and any theoretical or experimental uncertainties that may affect the estimation of the backgrounds.

6.7.1 Theoretical cross section uncertainties

The theoretical uncertainties on the W and γ^*/Z cross sections are estimated at NNLO by varying the factorisation and renormalisation scale; the uncertainty eigenvector of the parton distribution function; and the strong coupling constant [70]. An uncertainty of $\pm$ 5% is assigned for the production cross sections of the W and γ^*/Z processes.

For the $t\bar{t}$ production cross section, a theoretical uncertainty of $\pm$ 7% is assigned. This uncertainty is a combination of the uncertainties in the $t\bar{t}$ cross section measurement [71]; the parton density function; and the strong coupling constant.

6.7.2 Uncertainty for electrons in the problematic calorimeter regions

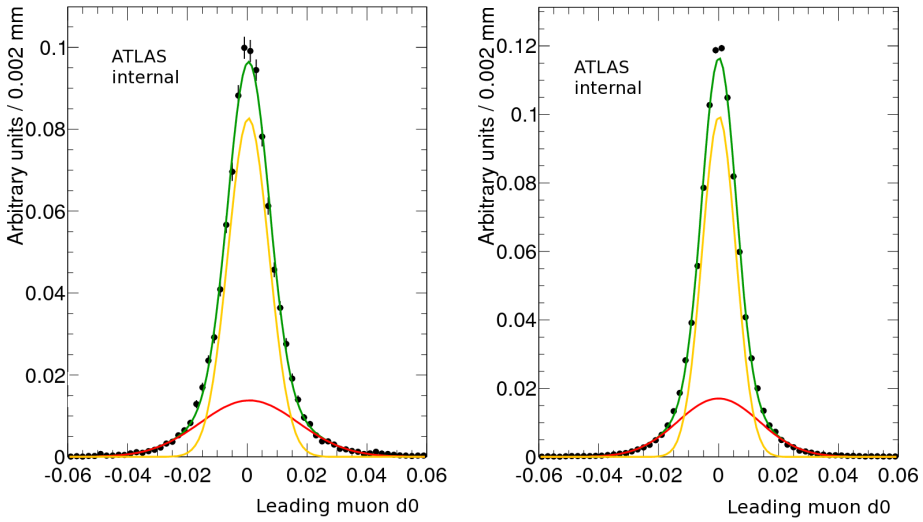
During data taking periods, some regions of the calorimeter were identified to be problematic. The map that determines whether or not clusters are in a problematic region of the calorimeter only corresponds to approximately 52% of the data used in this analysis. The systematic uncertainty related for using only one map for the whole dataset is estimated to be 0.4% [10]. This only affects electrons and therefore is only considered in the $e\mu$ channel.

6.7.3 Muon d_0

The radial muon impact parameter, d_0 is used to discriminate the $Z \to \tau\tau \to \mu\mu + 4\nu$ events from $\gamma^*/Z \to \mu\mu$ events in the $\mu\mu$ channel. The resolution of this variable is compared between the data and MC using the leading muon from on-shell $Z \to \mu\mu$ events. These events are selected by requiring exactly two isolated muons of opposite charge with an invariant mass between 75 and 105 GeV. The d_0 resolution is narrower in the MC and to correct for this discrepancy, additional smearing is applied to the simulations to match data.

The muon d_0 distribution can be represented as a sum of two Gaussian functions in both the data and MC. The fits for both distributions are shown in Figure 6.14. The measured amplitudes, means and widths of these Gaussians are shown in Table 6.11. The narrower of the two Gaussians is referred to as the "inner" Gaussian and the wider one is referred to as the "outer" Gaussian.

The smearing technique applied to the MC events needs to preserve the double Gaussian nature of the muon d_0 distributions. To achieve this, the inner and outer components of the Gaussians are each assigned a smearing term, $G(\mu, \sigma_i)$ and $G(\mu, \sigma_o)$, where G is a function that returns a random number from a normal distribution with mean μ and width σ. For each muon in the MC, the d_0 is smeared using the following equation:



(a) Data fit (b) MC fit

Figure 6.14: Double Gaussian fits for the data and MC events [10].

Gaussian Component	Amplitude	Mean	Width	χ^2/NDF
Data inner	$82.7 \pm 1.6 \mu m$	$0.4 \pm 0.1 \mu m$	$6.83 \pm 0.2 \mu m$	1.64
Data outer	$13.7 \pm 1.6 \mu m$	$0.7 \pm 0.3 \mu m$	$16.3 \pm 0.6 \mu m$	1.64
MC inner	$99.4 \pm 0.6 \mu m$	$0.1 \pm 0.02 \mu m$	$5.63 \pm 0.04 \mu m$	12.4
MC outer	$17.1 \pm 0.6 \mu m$	$0.0 \pm 0.07 \mu m$	$13.6 \pm 0.1 \mu m$	12.4

Table 6.11: The parameters for the double Gaussian fits on the data and MC d_0 distributions [10].

$$d_0(MC)_{new} = d_0(MC)_{original} + G(\mu, \sigma). \tag{6.17}$$

The μ term translates the mean value of the two MC Gaussians to match the mean of the two data Gaussians. The μ term is calculated using a weighted average:

$$\mu = (w_{i,data} \times \mu_{i,data} + w_{o,data} \times \mu_{o,data}) - (w_{i,mc} \times \mu_{i,mc} + w_{o,mc} \times \mu_{o,mc}), \tag{6.18}$$

where w_i and w_o is the weight given to each of the inner and outer Gaussians, respectively. The weight of $w_{i,data}$ is calculated using

$$w_{i,data} = \frac{f(0)_{i,data}}{f(0)_{i,data} + f(0)_{o,data}}, \tag{6.19}$$

where $f(0)$ is the value of the Gaussians evaluated at zero. The values for $w_{o,data}$, $w_{i,mc}$ and $w_{o,mc}$ are calculated similarly.

The σ term changes the width of the original Gaussians into the smeared Gaussians and is calculated using the following formula for the inner (or outer)Gaussians:

$$\sigma_{i(o)} = \sqrt{\sigma^2_{i(o),data} - \sigma^2_{i(o),mc}}. \tag{6.20}$$

The μ and σ values required to achieve the desired smearing are shown in Table 6.12.

Gaussian Component	Data σ (μm)	MC σ (μm)	G σ (μm)
Inner	6.84 ± 0.19	5.63 ± 0.14	3.88 ± 0.26
Outer	16.3 ± 0.7	13.6 ± 4.7	8.89 ± 0.95

Table 6.12: Measured parameters of the muons D0 distributions for data and MC [10].

When smearing the original distribution, only one of the smearing factors, $G(\mu, \sigma_i)$ or $G(\mu, \sigma_o)$, is applied for each event. The term that is used is determined probabilistically and is dependent on the value of d_0. The probability of using $G(\mu, \sigma_i)$ is given by:

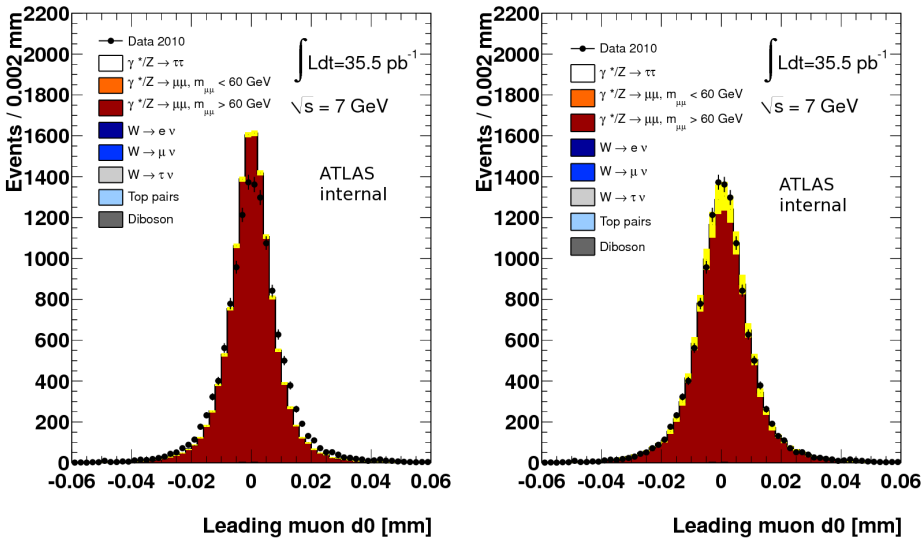
$$p = \frac{f(d_0)_{i,mc}}{f(d_0)_{i,mc} + f(d_0)_{o,mc}}, \tag{6.21}$$

where $f(d_0)$ is the value of the Gaussians evaluated at the given d_0 value. To determine whether or not to use this value, a random number, y, is generated from uniform distribution between 0 to 1. If $p > y$, then $G(\mu, \sigma_i)$ is used for the smearing, otherwise $G(\mu, \sigma_o)$ is used. The MC and data distributions before and after the smearing are shown in Figure 6.15.

Three systematic uncertainties are considered for the smearing of the muon d_0 and they are: the uncertainties in the widths of the smearing functions, $G(\mu, \sigma)$; the uncertainty in using a double Gaussian smearing technique; and the uncertainty on the scale in the tails of the d_0 distributions. These three systematic uncertainties are evaluated independently and combined in quadrature.

Muon d_0 width The widths of the smearing Gaussians, $G(\mu, \sigma)$, are varied up and down by the uncertainties that are quoted in Table 6.12. The uncertainties on these Gaussians are calculated with the equation $\delta = \sigma\sqrt{\chi^2/NDF}$, using the measured values from Table 6.11. The inner and outer Gaussians are varied coherently and the measured deviations away from the central values are given as the $\pm$ σ uncertainties.

Muon d_0 shape The robustness of using a double Gaussian smearing technique is evaluated by comparing it to the results of a single Gaussian smearing technique. The smearing is performed in the same way using equation 6.17, but only using one $G(\mu, \sigma)$



(a) (b)

Figure 6.15: Distributions of muon d_0 in data and MC with (right plot) and without (left plot) the d_0 smearing applied. The yellow hatches indicate the MC errors, which includes the systematics uncertainties of the muon corrections, vertex reweighting and d_0 smearing [10].

term. The width of $G(\mu,\sigma)$ calculated from a weighted average of the inner and outer Gaussians given by:

$$\sigma = \sqrt{w_{i,mc}\sigma^2_{i,mc} + w_{o,mc}\sigma^2_{o,mc}}, \tag{6.22}$$

where w is the weight of the inner and outer Gaussians with w_i calculated by

$$w_i = \frac{f(d0)^2_i}{f(d0)^2_i + f(d0)^2_o}, \tag{6.23}$$

$$w_o = \frac{f(d0)^2_o}{f(d0)^2_i + f(d0)^2_o}, \tag{6.24}$$

where $f(d0)$ is the value of the Gaussians evaluated at d_0. The systematic uncertainty of the d_0 smearing shape is taken as the difference between the values obtained when smearing with a single Gaussian and the nominal value.

Muon d_0 scale The number of events with $|d_0| > 0.04$ mm is checked for consistency between the data and MC after applying the smearing. The $\mu\mu$ analysis channel selects events with high $|d_0|$ values, so agreement within these regions is important. The number of events in data and MC with $|d_0| > 0.04$ mm is shown in Table 6.13.

The data and MC differ by 2% for events with $|d_0| > 0.04$ mm. This discrepancy is assigned as the systematic uncertainty for the muon d_0 scale.

Region	Data events	MC events	Ratio: data/mc
All d_0	13756	13623	1.01
$\|d_0\| > 0.04$ mm	574	565	1.02

Table 6.13: Data and MC scale comparison for the selected Z events [10].

6.7.4 Multijet systematics

The multijet estimation in this analysis assumes that the variables used to define the ABCD control regions are uncorrelated. In the $e\mu$ channel this refers to the correlation between the charge distribution of the leptons (OS or SS) and the isolation variables ($p_T$cone and $E_T$cone). For the $\mu\mu$ channel, this refers to the correlation between the leading and sub-leading muon's isolation efficiencies. Systematic uncertainties on these correlations and on $\varepsilon_{multijet}$ are considered.

The correlations between the isolation selections in the $\mu\mu$ channel has already been discussed in Section 6.4.2 using the k factors. The measured uncertainties on the k factors are propagated through the ABCD method to obtain the systematic uncertainties on the multijet estimate.

For the $e\mu$ channel, $R_{OS/SS}$ is checked against the isolation variables for correlations using $e\mu$ events with no isolation requirements. The lack of isolation selections enriches the multijet purity and the expected electroweak contamination is subtracted. For these events, $R_{OS/SS}$ is plotted against $E_T$cone30 and $p_T$cone40 for electrons; and $E_T$cone40 and $p_T$cone40 for muons, which are shown in Figure 6.16. Linear fits are made to these distributions to measure any deviation away from a zero gradient. The first bin is excluded from the fit because it contains too much signal contamination and is shown only for visual purposes.

To calculate the systematic uncertainty, the fitted function is extrapolated to the centre of the first bin and is denoted as R_{ex}. The last three bins of each distribution are merged to increase statistics and is denoted as $R'_{OS/SS}$. The results of the linear fits with their uncertainties are given in Table 6.14. The relative systematic uncertainty on $R_{OS/SS}$ is given by

$$\Delta R_{OS/SS} = |\frac{R_{ex}}{R'_{OS/SS}} - 1|. \tag{6.25}$$

The linear fits for the electrons measure gradients that are consistent with zero. However, for the muons, the gradients are measured to deviate away from zero. As a conservative

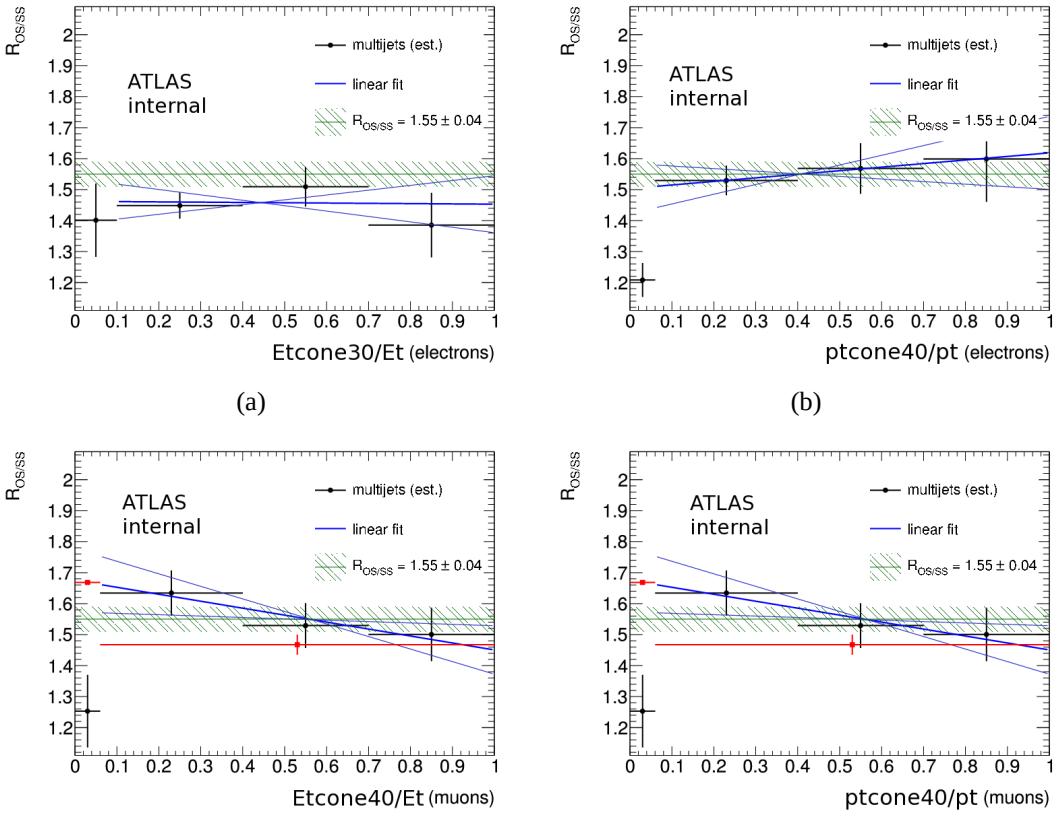


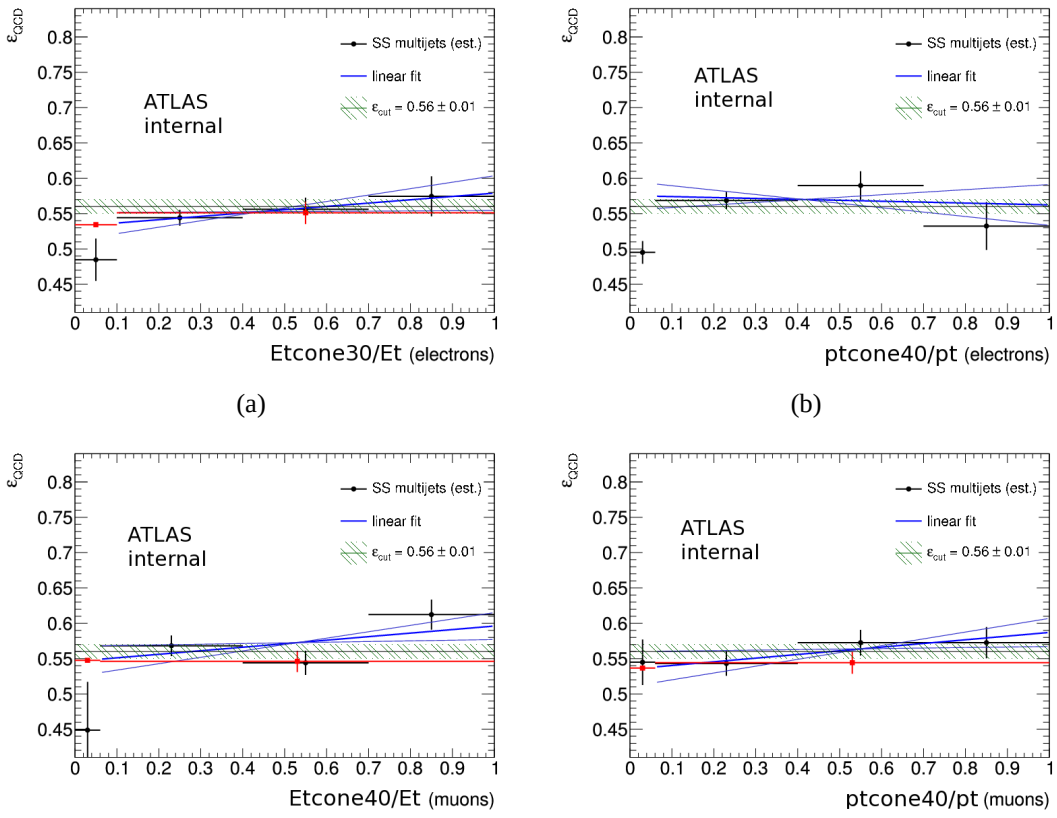
Figure 6.16: $R_{OS/SS}$ plotted as a function $E_T$cone (left) and $p_T$cone (right) for electrons (top) and muons (bottom). The blue lines represent the $\pm 1\sigma_{stat.}$ variations of the linear fit and the red boxes denote the R_{ex} and $R'_{OS/SS}$, respectively [10].

	a	b	R_{ex}	$R'_{OS/SS}$	$\Delta R_{syst.}$
$E_T$cone30/E_T (electrons)	1.46±0.07	-0.01±0.16	–	–	0%
$p_T$cone40/p_T (electrons)	1.50±0.08	0.12±0.20	–	–	0%
$E_T$cone40/E_T (muons)	1.63±0.08	-0.31±0.15	1.62	1.44	12.6%
$p_T$cone40/p_T (muons)	1.68±0.10	-0.22±0.18	1.67	1.47	13.7%

Table 6.14: Results of the linear fit $y = a + b \cdot x$ on $R_{OS/SS}$ [10].

estimate, the muons' uncertainty is used for both lepton flavours. The $R_{OS/SS}$ uncertainties are averaged and an overall uncertainty of 13.1% is considered [10].

The systematic uncertainty on $\varepsilon_{multijet}$ is calculated using a similar method. Again, $e\mu$ events are selected with no isolation requirement with the expected electroweak contributions subtracted. The efficiency, $\varepsilon_{multijet}$, is then measured as a function of the $p_T$cone and $E_T$cone for electrons and muons, shown in Figure 6.17. The results of the linear fits are given in Table 6.15. The fits with deviations larger than 1σ from zero are averaged to obtain a systematic uncertainty on $\varepsilon_{multijet}$, which is measured to be 2.4%.



(a) (b)

(c) (d)

Figure 6.17: Plots of $\varepsilon_{multijet}$ as a function of $E_T$cone (left) and $p_T$cone (right) for electrons (top) and muons (bottom). The blue lines represent the $\pm 1\sigma_{stat}$ variations of the linear fit and the red boxes denote the $\varepsilon^{ex}_{multijet}$ and $\varepsilon'_{multijet}$ respectively [10].

	a	b	$\varepsilon^{ex}_{multijet}$	$\varepsilon'_{multijet}$	$\Delta\varepsilon_{multijet,syst.}$
$E_T$cone30/E_T (electrons)	0.53±0.02	0.047±0.044	0.534	0.564	3.0%
$p_T$cone40/p_T (electrons)	0.57±0.02	-0.013±0.049	–	–	0%
$E_T$cone40/E_T (muons)	0.55±0.02	0.050±0.041	0.548	0.546	0.3%
$p_T$cone40/p_T (muons)	0.54±0.02	0.052±0.045	0.537	0.544	1.4%

Table 6.15: Results of the linear fit $y = a + b \cdot x$ on $\varepsilon_{multijet}$ [10].

6.7.5 A_Z uncertainty

There are three systematic effects that are considered for the acceptance factor A_Z. These are: the uncertainties on the CTEQ6.6 NLO PDF set used for generating the signal MC; the deviations between different PDF sets; and the uncertainty of the parton shower modelling.

The first effect is determined using the 22 PDF error eigenvectors of the CTEQ6.6 NLO PDF set [72] by combining them in quadrature and varying the total up and down one standard error. The second effect is measured by using the HERAPDF1.0 PDF set to measure A_Z and taking the difference between that and the nominal value. Finally, the MC@NLO signal sample is used to measure A_Z instead of the default PYTHIA sample to measure the uncertainty in the parton showering modelling. Combining these uncertainties in quadrature, the relative systematic uncertainty on A_Z is 2.9% for the $e\mu$ channel and 7% for the $\mu\mu$ channel, which is shown in Table 6.16.

Source of uncertainty	$\delta A_Z/A_Z$ $e\mu$ (%)	$\delta A_Z/A_Z$ $\mu\mu$ Standard (%)	$\delta A_Z/A_Z$ $\mu\mu$ BDT (%)
CTEQ 6.6 NLO PDF	1.3	1.3	1.2
Different PDF sets	1.8	2.6	2.0
Model dependence	1.8	1.2	6.5
Total uncertainty	2.9	3.1	6.9

Table 6.16: Relative sources of uncertainties of the A_Z [10].

6.7.6 C_Z uncertainty

The systematic uncertainties described previously in this section that affect the signal acceptance are all considered systematic uncertainties on C_Z. The list of these uncertainties and their relative contributions is given in Table 6.17 for the $e\mu$ channel and Table 6.18 for the $\mu\mu$ channel.

Source of Uncertainty	$\delta C_Z/C_Z(\%)$
Electron ID, reco, isolation, trigger efficiency	6.1
μ ID and isolation efficiency	2.6
Energy scale electron, jet and clusters	1.7
μ Energy scale and resolution	0.1
Electron resolution	0.1
Problematic regions in the calorimeter	0.4
Electron charge misidentification	0.3
Pileup re-weighting	0.6
Jet cleaning	0.4

Table 6.17: Relative systematic uncertainties for C_Z in the $e\mu$ channel [10].

Systematic uncertainty	Standard $\delta C_Z/C_Z(\%)$	BDT $\delta C_Z/C_Z(\%)$
Muon ID, reco, scale, trigger and isolation efficiency	4.4	4.7
Muon energy scale	0.3	0.5
E_T^{miss} smearing	0.0	0.0
Muon d_0 width	0.5	0.2
Muon d_0 scale	2.0	2.0
Muon d_0 shape	1.1	0.9
Pile-up re-weighting	0.0	0.0
Jet cleaning	0.0	0.2

Table 6.18: Relative systematic uncertainties for C_Z for the $\mu\mu$ channel for the standard and BDT selections [10].

6.7.7 Systematic uncertainty summary

The systematic uncertainties that affect the background estimations of the $Z \to \tau\tau$ cross section measurement are provided in Table 6.20 for the $e\mu$ channel and Table 6.19 for the $\mu\mu$ channel. All uncertainties are combined in quadrature to obtain the total uncertainty.

Source	Standard selection uncertainty (%)			
	$\gamma^*/Z \to ee, \mu\mu$	W + jets	$t\bar{t}$	Multijet
Muon scale factor	4.4	0.0	0	1.8
Muon p_T	0.2	0.0	0	0.7
Missing E_T	0.0	0.0	0	0.0
Muon d_0 width	3.8	0.0	0	0.0
Muon d_0 scale	2.0	2.0	0	0.0
Muon d_0 shape	0.9	0.0	0	0.0
Vertex Weight	0.0	0.0	0	0.8
Cross Section	5.0	5.0	6.0	2.3
Jet Cleaning	0.0	0.0	0	0.0
Luminosity	3.4	3.4	3.4	3.6
Multijet	–	–	–	8.2
Total Systematic	9.9	6.4	6.9	9.5
Source	BDT selection uncertainty (%)			
	$\gamma^*/Z \to ee, \mu\mu$	W + jets	$t\bar{t}$	Multijet
Muon scale factor	4.5	25	6.3	0.3
Muon p_T	0.8	0.0	6.3	0.1
Missing E_T	0.1	0.0	6.3	0.0
Muon d_0 width	5.1	0.0	6.3	0.1
Muon d_0 scale	2.0	2.0	2.0	0.0
Muon d_0 shape	0.6	0.0	6.3	0.2
Vertex Weight	0.3	0.0	0.0	0.1
Cross Section	5.0	5.0	6.0	0.3
Jet Cleaning	0.3	0.0	0.0	0.0
Luminosity	3.4	3.4	3.4	3.4
Multijet	–	–	–	6.2
Total Systematic	9.6	25.8	15.7	7.1

Table 6.19: A summary of the systematic uncertainties for the $\gamma^*/Z \to ee, \mu\mu$, W + jets, $t\bar{t}$ and multijet backgrounds in the $\mu\mu$ channel [10].

Source	Uncertainty (%)			
	$\gamma^*/Z \to ee, \mu\mu$	W + jets	$t\bar{t}$	Multijet
Muon p_T	0.12	0.00	0.42	0.05
Muon scale factor	1.23	1.82	1.72	4.89
Electron scale factor	5.38	6.37	5.92	4.66
Elecron/Cluster/Jet E	7.48	11.86	9.27	2.59
Vertex Weight	0.55	0.55	0.58	0.64
Luminosity	3.40	3.40	3.40	3.79
Cross Section	5.00	5.00	7.00	5.64
Problematic calorimeter regions	0.37	0.37	0.40	0.30
Electron Charge	0.28	0.28	0.28	0.31
Jet cleaning	2.10	0.79	16.24	0.40
Multijet estimation	–	–	–	13.00
Total Systematic	11.31	14.91	21.18	16.38

Table 6.20: A summary of the uncertainties for the $\gamma^*/Z \to ee, \mu\mu$, W + jets, $t\bar{t}$ and multijet backgrounds in the $e\mu$ channel [10].

6.8 Cross section measurement

The number of $Z \to \tau\tau \to ll + 4\nu$ candidates in data and the expected background contributions are shown in Table 6.21 along with their corresponding A_Z, C_Z and integrated luminosity.

	$e\mu$ channel	$\mu\mu$ channel standard	$\mu\mu$ channel BDT
N_{obs}	85	45	90
N_{bkg}	$8.94 \pm 4.12 \pm 0.84$	$24.4 \pm 1.3 \pm 2.0$	$47.5 \pm 9.9 \pm 3.5$
A_Z	$0.1139 \pm 0.0004 \pm 0.0033$	$0.0488 \pm 0.0004 \pm 0.0015$	$0.156 \pm 0.001 \pm 0.011$
C_Z	$0.2887 \pm 0.005 \pm 0.020$	$0.3607 \pm 0.0127 \pm 0.0219$	$0.266 \pm 0.006 \pm 0.0167$
$\mathcal{L}$	35.51 ± 1.21		

Table 6.21: A summary of the number of selected events, background contributions, A_Z, C_Z and integrated luminosity for the $e\mu$ and $\mu\mu$ channels. The first error is statistical and the second systematic [10].

6.8.1 $Z \to \tau\tau \to e\mu + 4\nu$ cross section

The product of the fiducial cross section and the branching ratio is measured to be:

$$\sigma^{\text{fid.}}_{\gamma^*/Z} \times \text{BR}(\gamma^*/Z \to \tau\tau) = 7.50 \pm 1.00\ (\text{stat.}) \pm 0.50\ (\text{sys.}) \pm 0.26\ (\text{lumi.})\ \text{pb}. \quad (6.26)$$

The total cross section is measured to be:

$$\sigma^{\text{tot.}}_{\gamma^*/Z}(66 < m_{\tau\tau} < 116GeV) = 1062.6 \pm 142.3\ (\text{stat.}) \pm 77.8\ (\text{sys.}) \pm 36.17\ (\text{lumi.})\ \text{pb}. \tag{6.27}$$

6.8.2 $Z \to \tau\tau \to \mu\mu + 4\nu$ cross section

The product of the fiducial cross section and the branching ratio for the standard and multivariate selections are measured to be:

- Standard selection:

  $\sigma^{\text{fid.}}_{\gamma^*/Z} \times \text{BR}(Z/\gamma* \to \tau\tau) = 1.64 \pm 0.54\ (\text{stat.}) \pm 0.21\ (\text{sys.}) \pm 0.06\ (\text{lumi.})\ \text{pb}.$

- BDT selection:

  $\sigma^{\text{fid.}}_{\gamma^*/Z} \times \text{BR}(Z/\gamma* \to \tau\tau) = 4.50 \pm 1.05\ (\text{stat.}) \pm 0.59\ (\text{sys.}) \pm 0.15\ (\text{lumi.})\ \text{pb}.$

The total cross section for both methods are:

- Standard selection:

  $\sigma^{\text{tot.}}_{\gamma^*/Z}(66 < m_{\tau\tau} < 116GeV) = 1113 \pm 365\ (\text{stat.}) \pm 145\ (\text{sys.}) \pm 38\ (\text{lumi.})\ \text{pb}.$

- BDT selection:

  $\sigma^{\text{tot.}}_{\gamma^*/Z}(66 < m_{\tau\tau} < 116GeV) = 959 \pm 225\ (\text{stat.}) \pm 130\ (\text{sys.}) \pm 33\ (\text{lumi.})\ \text{pb}.$

The measured total cross sections in both channels are in agreement with each other and the theoretical prediction of

$$\sigma^{\text{tot,theory}}_{\gamma^*/Z}(66 < m_{\tau\tau} < 116GeV) = 960 \pm 49.5\ \text{pb}. \tag{6.28}$$

Assuming lepton universality, the measurements also agree with the combined cross section measurement of the $Z \to ee$ and $Z \to \mu\mu$ processes, which is measured by ATLAS to be

$$\sigma^{\text{tot.}}_{\gamma^*/Z}(66 < m_{ll} < 116GeV) = 945 \pm 6\ (\text{stat.}) \pm 40\ (\text{syst.}) \pm 32\ (\text{lumi.})\ \text{pb}. \tag{6.29}$$

6.9 Summary

This chapter has described the analysis used to measure the $Z \to \tau\tau$ cross section using its leptonic decay modes. These results will be combined with the results obtained using the $Z \to \tau\tau \to l\tau_h + 3\nu$ decay modes to achieve the best possible precision.

7

$Z \to \tau\tau$ combined cross section

The cross section of the $Z \to \tau\tau$ process is also measured using its semi-leptonic decay modes: $Z \to \tau\tau \to \mu\tau_h + 3\nu$ and $Z \to \tau\tau \to e\tau_h + 3\nu$. These decay channels are referred to as the $e\tau_h$ and $\mu\tau_h$, respectively and as the $l\tau_h$ channels collectively. This chapter will briefly summarise the measurements of the $Z \to \tau\tau$ cross sections in these channels; and the combination of these measurements with the $Z \to \tau\tau \to ll + 4\nu$ results. A more detailed description of the $e\tau_h$ and $\mu\tau_h$ analyses channels can be found in references [54] and [69].

7.1 $Z \to \tau\tau \to l\tau_h + 3\nu$ selections

7.1.1 Trigger selection

Events in the $e\tau_h$ channel are collected using the same trigger as the $e\mu$ channel. For the $\mu\tau_h$ channel, the triggers are the same as those used for the $\mu\mu$ channel.

7.1.2 Particle selections

The $l\tau_h$ channels require exactly one electron or muon with exactly one τ_h candidate of opposite charge. The lepton selections and the definition of the E_T^{miss} are the same between the $Z \to \tau\tau \to l\tau_h + 3\nu$ and $Z \to \tau\tau \to ll + 4\nu$ analyses, with the exception of the electron identification requirement where a "tight" selection is applied instead. The tighter identification requirement reduces the contributions from multijets, which is expected to be larger background in the $l\tau_h$ channels. The same isolation requirements used in the $e\mu$ channel are applied for the electrons and muons in the $l\tau_h$ channels.

The τ_h reconstruction algorithms are optimised to distinguish between τ_h and jets produced from multijet processes [73]. The τ_h identification selections uses three discriminating

variables, which are based on the energy-weighted cluster width of the jet; the momentum-weighted track width of the jet; and the fraction of the candidate's p_T carried by the leading track. For τ_h jets, the widths of the jets are usually narrower than jets produced from multijet processes and the p_T fraction of the leading track is usually higher. The τ_h identification requirements used for the $l\tau_h$ channels have an efficiency of $\sim$ 40% for τ_h jets with one charged pion and $\sim$ 30% for τ_h jets with three charged pions.

7.1.3 Event selections

The $W \rightarrow l\nu$ + jet process is a major electroweak background for the $l\tau_h$ channels since these events contain an isolated lepton from the W boson and the associated jet can fake a τ_h candidate. To reduce this background, two selections are applied. The first requirement, $\sum cos\Delta\phi$ > -0.15, is the same as the one used for the $e\mu$ channel and the second requires the transverse mass to be below 50 GeV. Distributions of $\sum cos\Delta\phi$ and the transverse mass, m_T, are shown in Figure 7.1 and Figure 7.2, respectively.

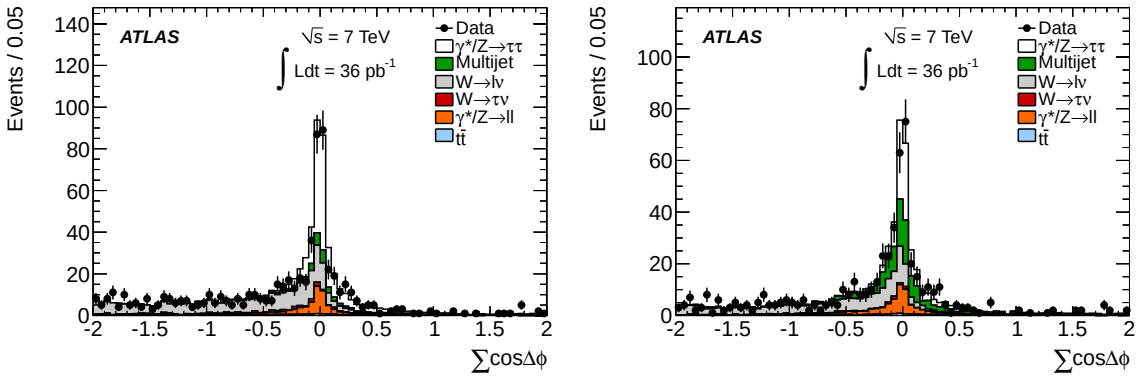
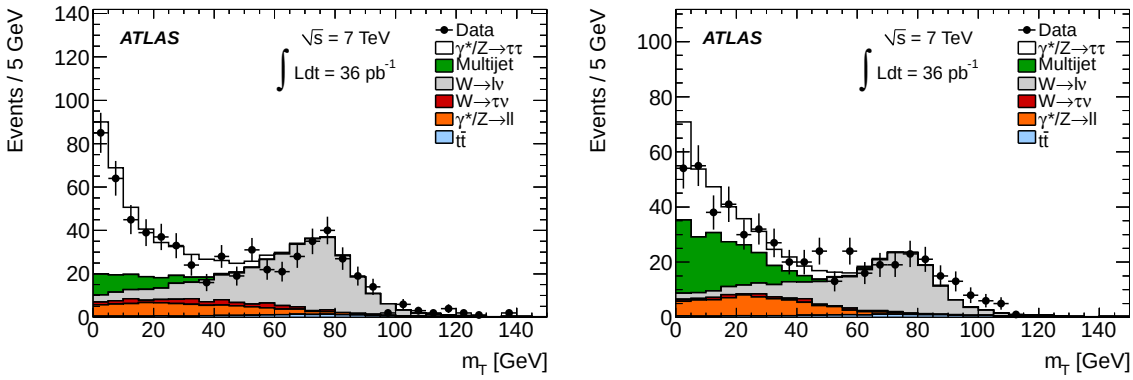


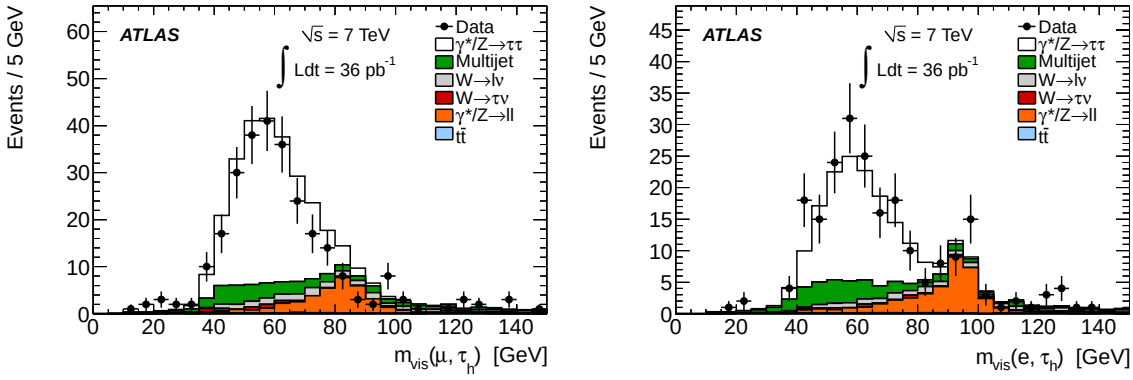
Figure 7.1: The $\sum cos\Delta\phi$ distributions shown in the $e\tau_h$ (a) and $\mu\tau_h$ (b) channels after selecting one lepton and τ_h of opposite charge.



(a) (b)

Figure 7.2: The m_T distributions shown in the $e\tau_h$ (a) $\mu\tau_h$ (b) channels after selecting one lepton and τ_h of opposite charge.

The final selection in the $l\tau_h$ channels requires the invariant mass of the τ_h and the lepton to be within the range 35 < m_{l,τ_h} < 75 GeV. This selection window encompasses the majority of the signal spectrum and rejects contributions from on-shell $Z \rightarrow ll$ events where one of the leptons fakes a τ_h. The m_{l,τ_h} distributions after applying all other event selections are shown in Figure 7.3.



(a) (b)

Figure 7.3: The distributions of the m_{l,τ_h} shown after the $\sum cos\Delta\phi$ and m_T selections in the $e\tau_h$ (a) and $\mu\tau_h$ (b) channels.

7.1.4 Candidates and acceptance

The number of $Z \rightarrow \tau\tau \rightarrow l\tau_h + 3\nu$ candidates observed after all event selections are shown in Table 7.1 along with the expected signal and background contributions. All background contributions are estimated using simulation, except for the multijet background, which is estimated from data. A variation of the ABCD method is used to estimate the multijet

background, which is similar to the methods described in Section 6.4. A more detailed description of the background estimations in the $l\tau_h$ channels can be found in reference [54].

Channel	$e\tau_h$	$\mu\tau_h$
$\gamma^*/Z \to ll$ $(l = e, \mu)$	6.9 ± 0.4	11.1 ± 0.5
$W \to l\nu$ $(l = e, \mu)$	4.8 ± 0.4	9.3 ± 0.7
$W \to \tau\nu$	1.5 ± 0.4	3.6 ± 0.8
$t\bar{t}$	1.02 ± 0.08	1.3 ± 0.1
Diboson	0.18 ± 0.01	0.28 ± 0.02
Multijet	23 ± 6	24 ± 6
$\gamma^*/Z \to \tau\tau$	98 ± 1	186 ± 2
Total expected events	135 ± 6	235 ± 6
N_{obs}	151	213

Table 7.1: The $Z \to \tau\tau \to l\tau_h + 3\nu$ candidates with the expected signal and background contributions. Only the statistical uncertainties are shown.

The fiducial regions of the $e\tau_h$ and $\mu\tau_h$ channels are defined as follows:

- μ p_T > 15 GeV and $|\eta|$ < 2.4 ($\mu\tau_h$ only),
- e p_T > 16 GeV and 0 < $|\eta|$ < 1.37 or 1.52 < $|\eta|$ < 2.47 ($e\tau_h$ only),
- τ_h p_T > 20 Gev and 0 < $|\eta|$ < 1.37 or 1.52 < $|\eta|$ < 2.47,
- $\sum cos\Delta\phi$ > -0.15,
- m_T < 50 GeV,
- 35 < $m_{l\tau_h}$ < 75 GeV.

The correction factors A_Z and C_Z of the $l\tau_h$ channels are defined similarly to those in the $e\mu$ and $\mu\mu$ channels (see Section 6.6). These are estimated from the $\gamma^*/Z \to \tau\tau$ MC samples. The uncertainties on A_Z and C_Z are also calculated using the methods described for the $e\mu$ and $\mu\mu$ channels. Table 7.2 shows the measured values of A_Z and C_Z; the observed data candidates minus the estimated backgrounds ($N_{obs} - N_{bkg}$); the branching ratio of the final states (BR); and the integrated luminosity ($\mathcal{L}$) of the data used in the cross section measurement.

Channel	$e\tau_h$	$\mu\tau_h$
$N_{obs} - N_{bkg}$	$114 \pm 14 \pm 3$	$164 \pm 16 \pm 4$
A_Z	0.101 ± 0.003	0.117 ± 0.004
C_Z	0.12 ± 0.02	0.20 ± 0.03
BR	0.2313 ± 0.0009	0.2250 ± 0.0009
$\mathcal{L}$	$35.5 \pm 1.2\ pb^{-1}$	$35.5 \pm 1.2\ pb^{-1}$

Table 7.2: The components used for the $Z \rightarrow \tau\tau$ cross section calculations in the $e\tau_h$ and $\mu\tau_h$ analysis channels. For $N_{obs} - N_{bkg}$, the first uncertainty is the statistical component and the second is the systematic. For all other values, the quoted uncertainties contains all components combined in quadrature.

7.2 Combined cross section measurement

The $Z \rightarrow \tau\tau$ cross section is measured in the $l\tau_h$ channels using the same method as described in Section 6.6. The fiducial and total cross sections measured in the four analysis channels are presented in Table 7.3.

Channel	Fiducial cross section (pb)	Total cross section ([66,116] GeV) (nb)
$e\tau_h$	$27 \pm 3 \pm 5 \pm 1$	$1.14 \pm 0.14 \pm 0.20 \pm 0.04$
$\mu\tau_h$	$23 \pm 2 \pm 3 \pm 1$	$0.86 \pm 0.08 \pm 0.12 \pm 0.03$
$e\mu$	$7.5 \pm 1.0 \pm 0.5 \pm 0.3$	$1.06 \pm 0.14 \pm 0.08 \pm 0.04$
$\mu\mu$	$4.5 \pm 1.1 \pm 0.6 \pm 0.2$	$0.96 \pm 0.22 \pm 0.12 \pm 0.03$
Combined total $Z \rightarrow \tau\tau$ cross section		$\mathbf{0.97 \pm 0.07 \pm 0.06 \pm 0.03}$

Table 7.3: Summary of the fiducial and total cross sections of the $Z \rightarrow \tau\tau$ process measured in the four analysis channels. The quoted uncertainties contain the statistical, systematic and luminosity components, respectively [69].

The total cross section measurements are combined using the Best Linear Unbiased Estimate (BLUE) method [74] [75]. The BLUE method provides an estimate of the $Z \rightarrow \tau\tau$ total cross section by linearly combining the individual measurements and their uncertainties. The correlations between the systematic uncertainties of each channel are considered, which are assumed to be either fully correlated or fully uncorrelated. The corrections to the particles' reconstruction and isolation efficiencies are assumed to be fully correlated for the same particle types, but fully uncorrelated between different particle types. For the trigger systematics, the uncertainties are assumed to be correlated if the channels used the same triggers, but otherwise they are assumed to be uncorrelated. For the electron and tauon energy scale, their uncertainties are conservatively taken to be fully correlated. The multijet background is estimated using similar methods in the $e\mu$, $e\tau_h$ and

$\mu\tau_h$ channels and their uncertainties are conservatively assumed to be fully correlated. Finally, the uncertainties of the luminosity; the normalisation of the electroweak and $t\bar{t}$ backgrounds; and A_Z are considered to be fully correlated.

The measurements of the total $Z \to \tau\tau$ cross section agree with the theoretical predictions individually and combined. Assuming lepton universality, the results also agree with the measured total cross section of the $Z \to ll$ ($l = e, \mu$) processes. A plot of these measurements for each channel and their combination is shown in Figure 7.4. The combined result is also consistent with the total cross section measured by the CMS experiment of $\sigma^{\text{tot.}}_{\gamma^*/Z}(66 < m_{\tau\tau} < 116\ GeV) = 1.00 \pm 0.04\ (\text{stat.}) \pm 0.08\ (\text{sys.}) \pm 0.04\ (\text{lumi.})$ nb [76].

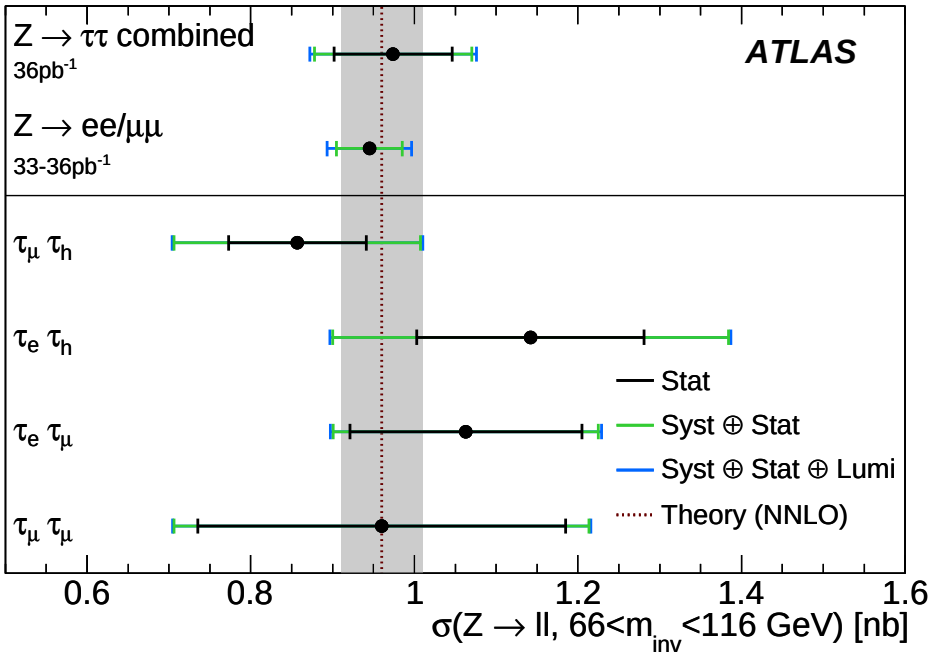


Figure 7.4: The total $Z \to \tau\tau$ cross section measured in the four analysis channels. The greyed area indicates the theoretically predicted value and its uncertainties [69].

8

$H \rightarrow \tau\tau \rightarrow ll + 4\nu$ search

This chapter will detail the analysis used to search for the Higgs boson using the $H \rightarrow \tau\tau \rightarrow ll + 4\nu$ decay modes. The event selections are optimised for a Higgs mass of about 120 GeV, but hypothesised Higgs masses between 100 - 150 GeV are also considered.

8.1 Trigger selection

The p_T of the signal leptons are expected to be soft. To maintain a high signal yield, only the unprescaled triggers with the lowest p_T thresholds are used.

To maximise the efficiency and acceptance of signal events, multiple triggers are used in conjunction to collect the data for the $e\mu$ channel. Events are selected if they pass any of the following triggers:

- muon p_T > 6 GeV and "medium" electron E_T > 10 GeV,
- muon p_T > 18 GeV,
- "medium" electron E_T > 22 GeV (20 GeV for some data).

The di-muon and single muon triggers are similarly used in conjunction for the $\mu\mu$ channel. For the di-muon trigger, the leading muon is required to have p_T > 15 GeV and the sub-leading muon is required to have p_T > 10 GeV. A p_T threshold of 18 GeV is required for the single muon trigger.

Data for the ee channel was collected using the di-electron trigger. Both electrons are required to have E_T > 12 GeV and pass the "medium" identification criteria.

8.2 Particle selection

The topology of $H \to \tau\tau \to ll + 4\nu$ events requires well reconstructed electrons, muons, jets, and missing transverse energy.

Electrons are required to have E_T > 15 GeV. If the single electron trigger is used, then the electron is also required to be matched to that trigger. The matching electron is required to have E_T > 23 GeV to be above the trigger threshold where the efficiencies are well understood.

To reduce the background contributions from multijets, the following isolation requirements are applied to all electrons: $E_T$cone20/E_T < 0.08 and $p_T$cone40/p_T < 0.06.

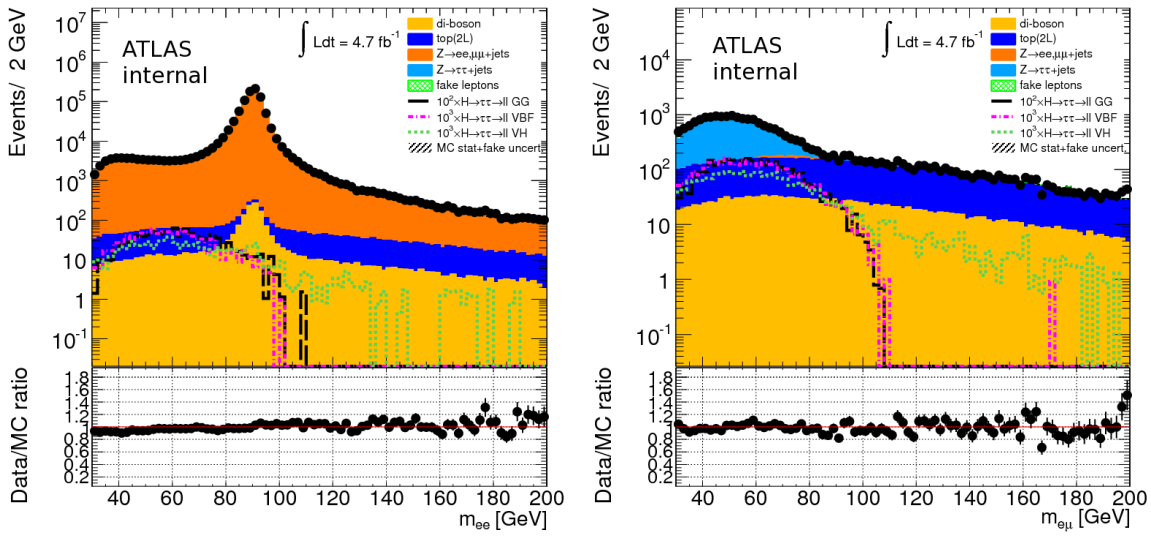
Muons are required to have p_T > 10 GeV and if the single muon trigger is used, then a trigger matching muon with p_T > 17 GeV is required. The reason for this requirement is the same as that for the electrons. Similarly, to reduce the multijet background, muons are required to satisfy the following isolation selections: $E_T$cone20/p_T < 0.02 and $p_T$cone40/p_T < 0.04.

All reconstructed jets must have an E_T > 20 GeV and satisfy the quality selections described earlier in this thesis.

8.3 Event selection

Events selected by this analysis are divided into four exclusive categories: the 2-jet vector boson fusion (VBF) category; the 2-jet Higgsstrahlung (VH) category; the 1-jet category; and the 0-jet category. The 2-jet categories focus on selecting events that are produced via the VBF and VH production mechanisms. The 1-jet category is a generic search channel that selects the VBF and VH events that don't pass the stringent requirements of the 2-jet categories. In addition, Higgs events produced via the gluon-gluon fusion mechanism with associated jets may also be selected in the 1-jet category. Lastly, the 0-jet category focuses mainly on selecting events produced via the gluon-gluon fusion mechanism. This category is mainly an inclusive search and has the highest background contributions.

The first selection for all categories requires exactly two leptons of opposite charge. All two lepton flavour combinations are accepted, except in the 0-jet channel where only the $e\mu$ channel is used to reduce the $\gamma^*/Z \to ee, \mu\mu$ background. The distributions of the three channels after applying the two lepton selection are shown in Figure 8.1. The fake background contribution in all the plots and tables presented in this chapter is estimated using a data-driven method, which is discussed later in Section 8.4.4. The $\gamma* /Z \to \tau\tau$ background is also estimated from data and is described in Section 8.4.1. All other main backgrounds are estimated using simulations and are either rescaled or cross checked using data-driven methods.



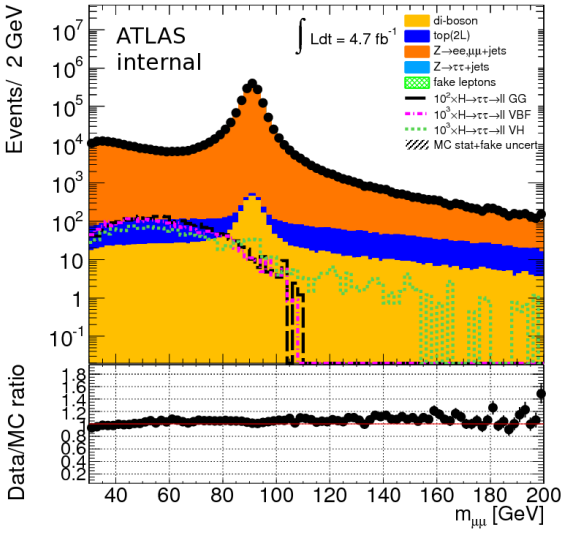


Figure 8.1: The m_{ll} distributions for the three channels after the dilepton selection [58].

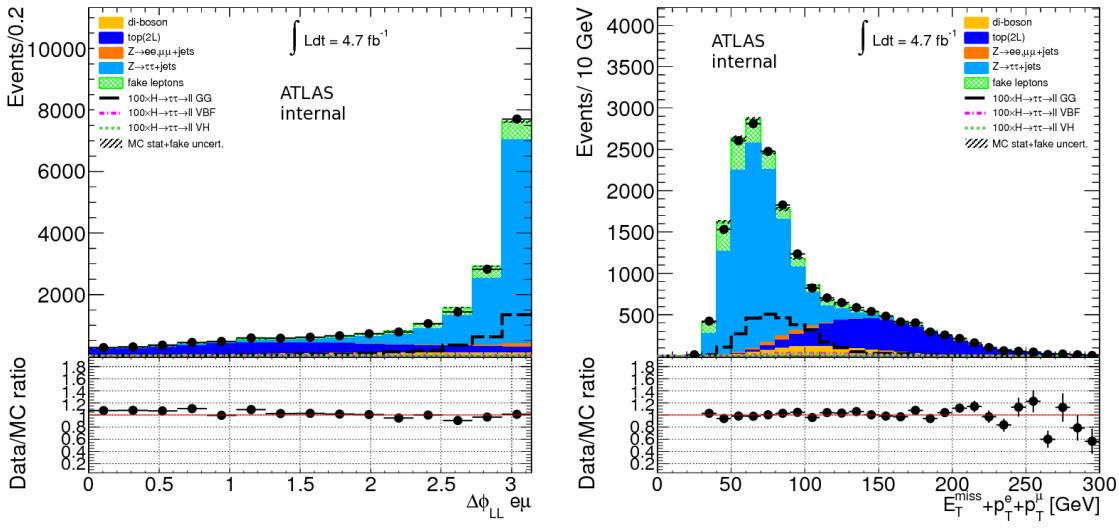
0-jet category selections The 0-jet category has the following event selections:

- 30 GeV < $m_{e\mu}$ < 100 GeV. This invariant mass window selects the majority of the signal events whilst rejecting any background contributions outside the signal spectrum. The $e\mu$ invariant mass distribution is shown in Figure 8.1(b).

- $\Delta\phi_{e\mu}$ > 2.5. Higgs bosons created via the gluon-gluon production mechanism usually have minimal boost in the transverse direction. The final state leptons are therefore back-to-back in the azimuthal plane. This selection has a high rejection against $t\bar{t}$, multijet and W+jet events. The distribution for this variable is shown Figure 8.2(a).

- H_T^{Lep} = p_T(e) + $p_T(\mu)$ + E_T^{miss} < 120 GeV. The centre-of-mass energy of $t\bar{t}$ events are

usually higher than that of the signal. Therefore the sum of the lepton p_T and E_T^{miss} is higher for this background, which is rejected effectively by this criteria. This distribution is shown in Figure 8.2(b).



(a) (b)

Figure 8.2: The $\Delta\phi_{e\mu}$ and H_T^{Lep} distributions shown after the invariant mass selection [58].

1-jet and 2-jet category common selections

- 30 GeV < $m_{ee,\mu\mu}$ < 75 GeV and 30 GeV < $m_{e\mu}$ < 100 GeV. For the ee and $\mu\mu$ channels the invariant mass of the two leptons is chosen to reject background contributions from $Z \to ee$ and $Z \to \mu\mu$ events. Since this background is minimal in the $e\mu$ channel, the mass window is increased to maximise signal efficiency. These distributions are shown in Figure 8.1.

- $N_{jet} \geq 1$ with p_T > 40 GeV (JVF > 0.75 if $|\eta|$ < 2.4, section 3.3.3). The p_T threshold of the leading jet is chosen to be as high as possible with minimal signal losses from the VBF and VH processes. The distributions for the jet multiplicities and jet p_T are shown in Figure 8.3.

- E_T^{miss} > 40 GeV for ee and $\mu\mu$ channels and E_T^{miss} > 20 GeV for $e\mu$ channel. This selection has a high rejection against multijet events. In the same flavoured channels, a higher requirement is used to further reject $\gamma^*/Z \to ee, \mu\mu$ events. The distributions of the E_T^{miss} after applying the previous criteria are shown in Figure 8.4.

- $0 < x_1, x_2 < 1$, where x_1 and x_2 are the fractions of the tauon's momenta in the form of neutrinos estimated by the collinear approximation [77]. The collinear approximation is a common mass reconstruction method used for τ-lepton decays. This method assumes the daughter particles of the tauons all travel approximately in the same direction. The transverse momentum of the charged leptons and missing transverse energy can be written as:

$$\vec{\tau}_{1,T} + \vec{\tau}_{2,T} = \vec{l}_{1,T} + \vec{l}_{2,T} + E_T^{miss} \equiv \frac{\vec{l}_{1,T}}{x_1} + \frac{\vec{l}_{2,T}}{x_2}, \tag{8.1}$$

where l_T denotes the transverse momentum of the charged leptons; and x_1 and x_2 represent the fractions of neutrino momenta. Equation 8.1 can be rewritten in terms of two coupled equations to solve for the unknowns, x_1 and x_2:

$$E_x^{miss} = (\frac{1}{x_1} - 1)\vec{l}_{1,x} + (\frac{1}{x_2} - 1)\vec{l}_{2,x}, \tag{8.2}$$

$$E_y^{miss} = (\frac{1}{x_1} - 1)\vec{l}_{1,y} + (\frac{1}{x_2} - 1)\vec{l}_{2,y}. \tag{8.3}$$

The invariant mass of the tauon pair can be reconstructed using the collinear mass:

$$m_{\tau\tau} = \frac{m_{ll}}{\sqrt{x_1 x_2}}. \tag{8.4}$$

The $0 < x_1, x_2 < 1$ selection ensures the two fractions give a real $m_{\tau\tau}$ mass term. For background processes, the assumptions of the collinear approximation may not be valid and can often yield imaginary masses, which are unphysical. Therefore this selection also provides rejection for events that don't have topologies similar to the $H \to \tau\tau \to ll + 4\nu$ events. The distributions of x_1 and x_2 after the previous selections are shown in Figure 8.5.

- $0.5 < \Delta\phi_{ll} < 2.5$. Higgs bosons created by the VBF and VH mechanisms generally have some boost in the transverse direction from recoiling off their associated jets. The boost focuses the angular separations of the leptons in the azimuthal plane. This selection rejects background contributions from $\gamma^*/Z \to ll$ ($l = e, \mu, \tau$) and $t\bar{t}$ events, which are usually created with small transverse momenta. The upper bound also ensures that the 1-jet and 2-jet categories are exclusive to the 0-jet category. The distribution for this variable is shown in Figure 8.6 after applying all previous requirements.

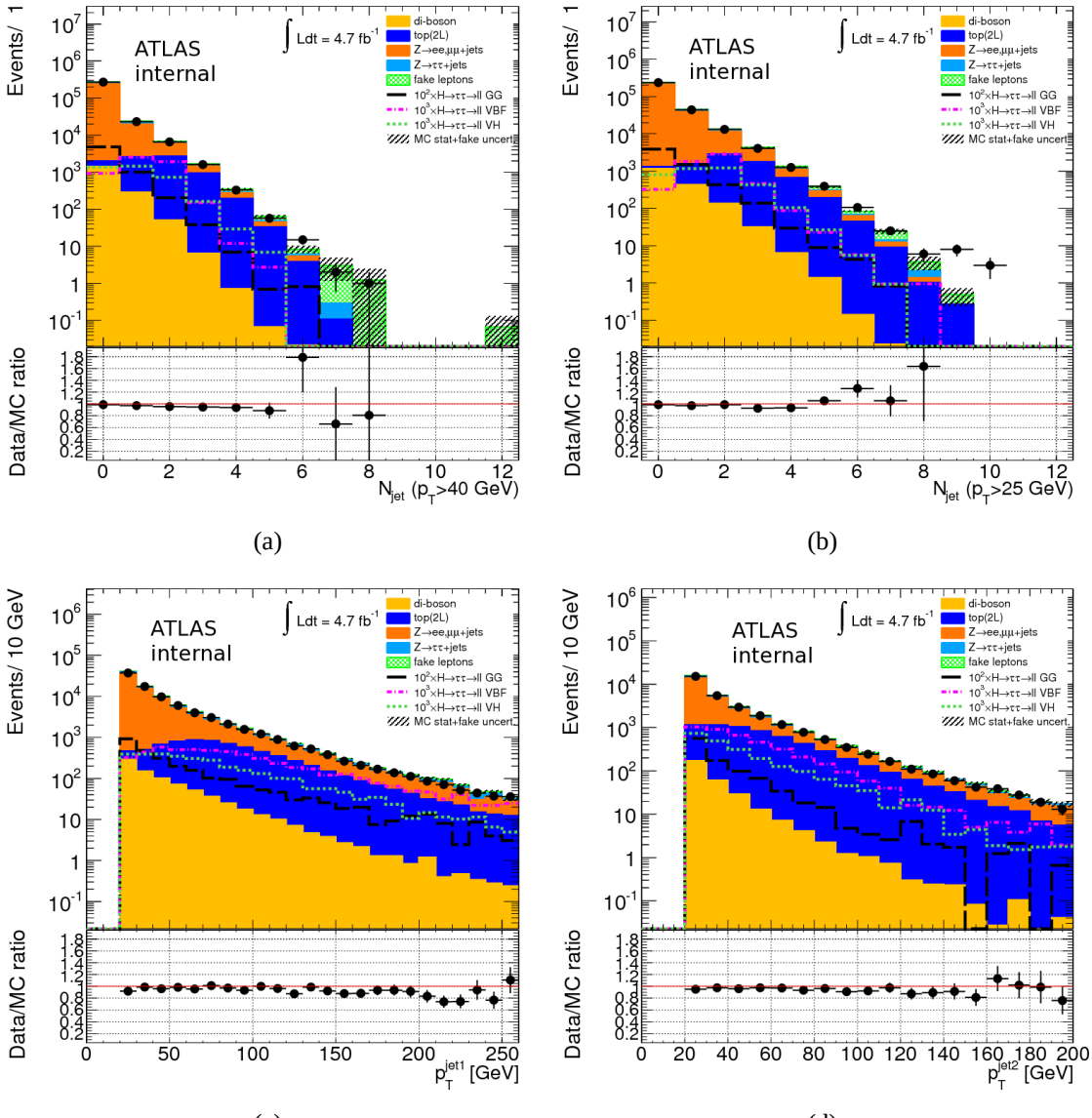
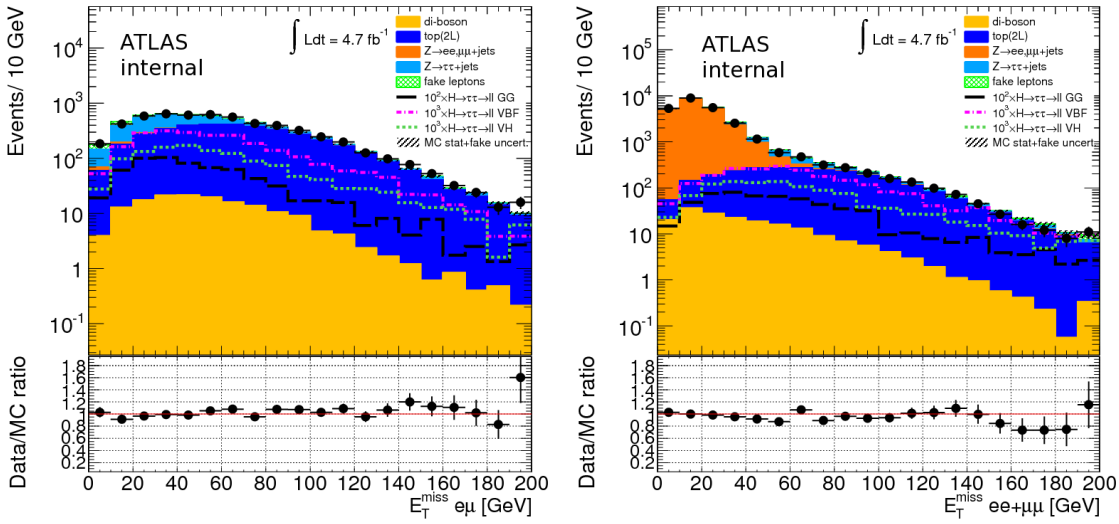
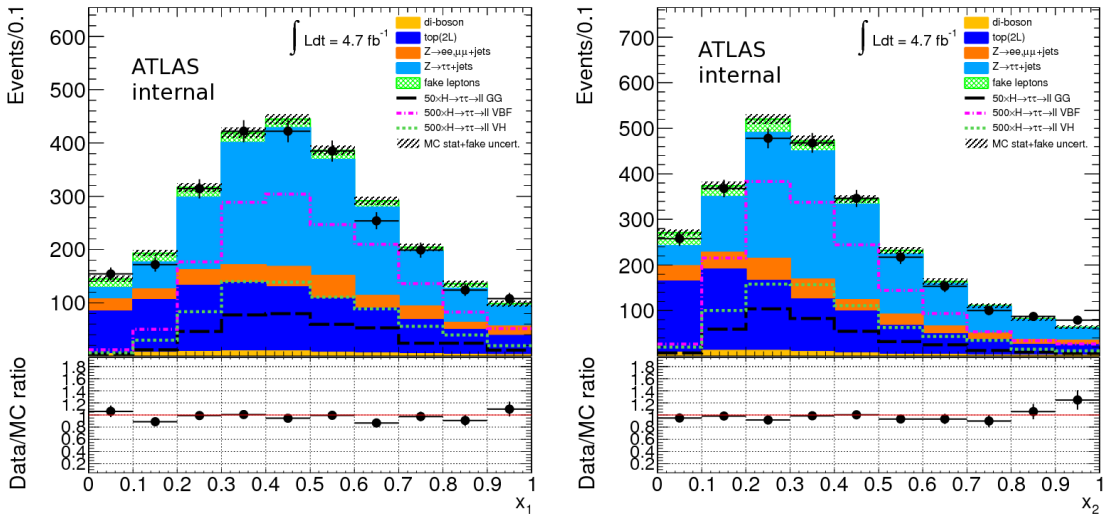


Figure 8.3: The distributions for the number of jets with $p_T > 40$ GeV and $p_T > 25$ GeV after the invariant mass selection are shown in (a) and (b), respectively. The p_T of the leading jet for events with at least one jet is shown in (c) and the p_T of the sub-leading jet for events which have at least 2 jets is shown in figure 8.3(d) [58].



(a) (b)

Figure 8.4: The distributions of the E_T^{miss} after the invariant mass and jet p_T > 40 GeV selections for the $e\mu$ (left) and same flavoured channels (right) [58].



(a) (b)

Figure 8.5: The distributions of x_1 and x_2 for all channels after the invariant mass, jet p_T and E_T^{miss} selections [58].

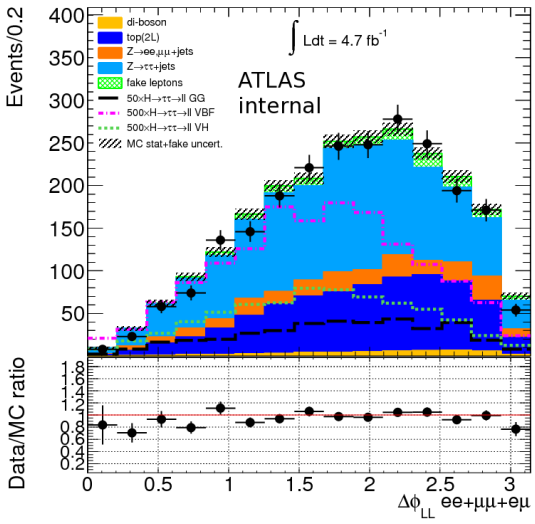
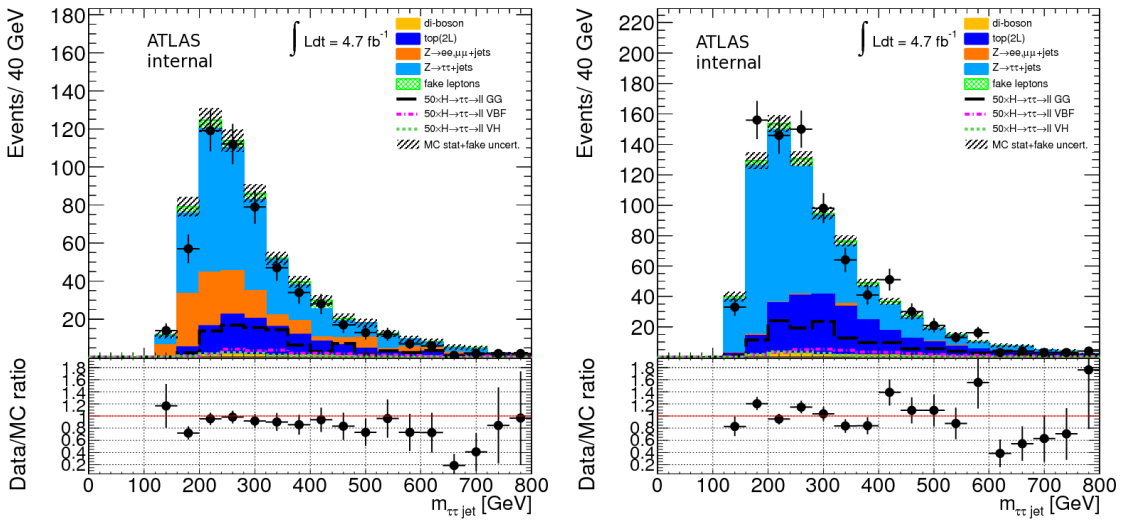


Figure 8.6: The $\Delta\phi_{ll}$ distributions for all channels after the invariant mass, jet p_T, E_T^{miss} and $x_{1,2}$ selections [58].

1-jet selections

- Excludes events that pass the 2-jet VBF or 2-jet VH selections.
- $m_{\tau\tau j}$ > 225 GeV. The invariant mass of the two tauons and the leading jet must be greater than 225 GeV, where $m_{\tau\tau}$ is calculated using the collinear approximation. This selection reduces background contributions from $\gamma^*/Z \to ll$ ($l = e, \mu, \tau$) processes. The $m_{\tau\tau j}$ distributions are shown in Figure 8.7 for events in this category.
- b-jet veto. Jets identified by the combination of the IP3D and JetFitter flavour tagging algorithms (section 3.3.4) are considered as b-jets. Events are rejected if they have any identified b-jets. This selection provides great discrimination against the $t\bar{t}$ and single-top backgrounds.



(a) $ee + \mu\mu$ (b) $e\mu$

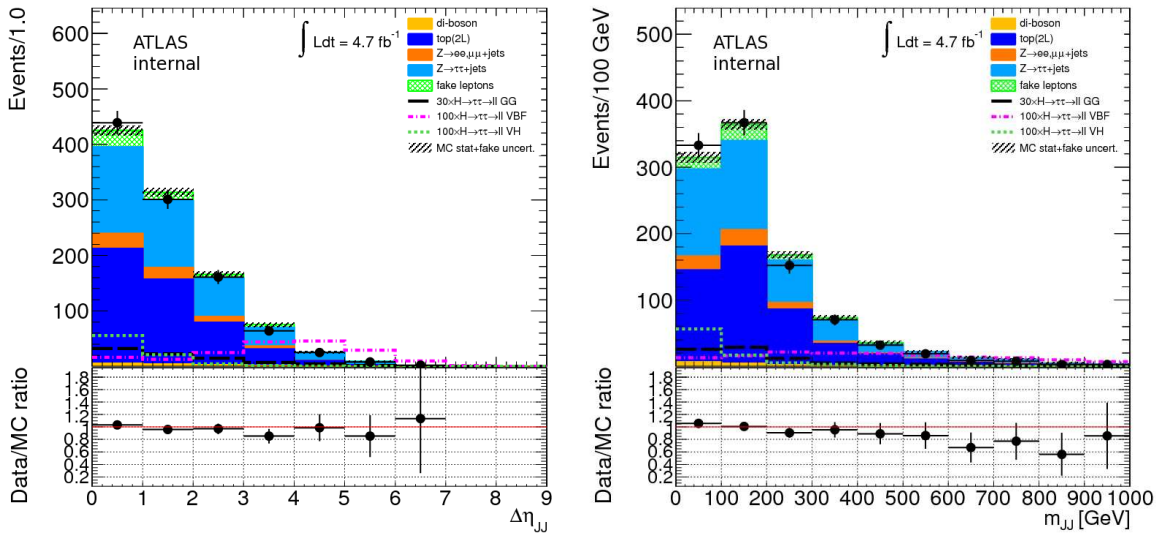
Figure 8.7: The $m_{\tau\tau j}$ distributions for events in the 1-jet category [58].

2-jet VBF selections

- A second jet with p_T > 25 GeV (JVF > 0.75 if $|\eta|$ < 2.4)
- $\Delta\eta_{jj}$ > 3.0. The pseudorapidity difference between the two leading jets must be greater than 3.0. This requirements favours the selection of forward jets produced by the VBF production mechanism and rejects $t\bar{t}$ and $Z \to ll$+jet ($l = e, \mu, \tau$) events where the jets are more central. The $\Delta\eta_{jj}$ distributions are shown in Figure 8.8(a).
- m_{jj} > 350 GeV. The invariant mass of the two jets must be above 350 GeV. Jets produced by the VBF mechanism are on average more energetic than those from $t\bar{t}$ and $Z \to ll$+jet ($l = e, \mu, \tau$) events and thus, this selection provides a good discrimination against these backgrounds. The m_{jj} distributions are shown in Figure 8.8(b).
- b-jet veto, as described in the 1-jet category.
- Central jet veto. No additional jets with p_T > 25 GeV can be found within the pseudorapidity of the two leading jets. This selection is mainly used to reject $t\bar{t}$ events. The jet η distributions for the leading and sub-leading jets are shown in Figure 8.9. A discrepancy is observed in the η distribution of the leading jet, which is attributed to a mismodelling of the parton PDF in the $Z \to ll$ ($l = (e, \mu)$) ALPGEN samples [78]. The systematic uncertainties considered for the parton PDF and the cross section of the $Z \to ll$ processes is expected to be sufficient to cover this discrepancy.

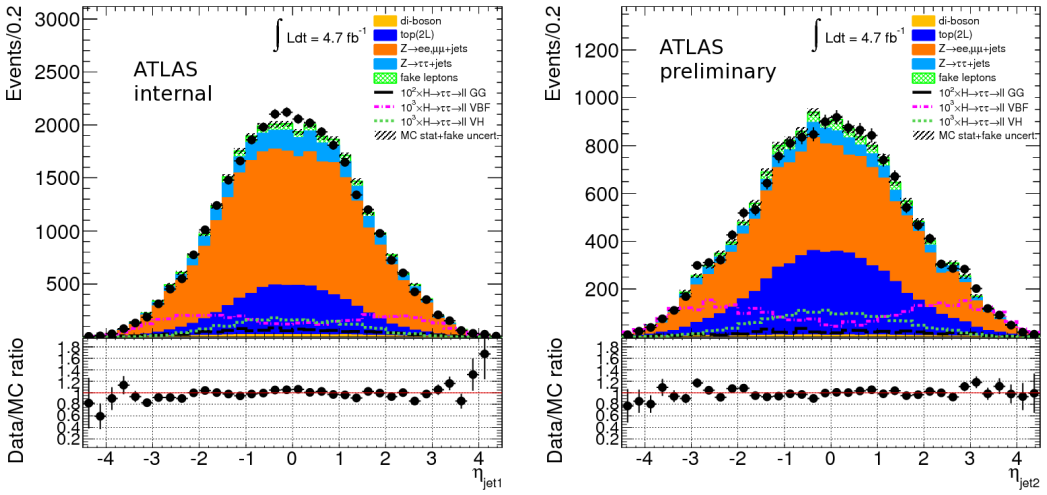
2-jet VH selections

- A second jet, as described in the VBF category.
- $\Delta\eta_{jj}$ < 2.0. For the VH production mechanism, the vector boson is boosted back-to-back in the transverse plane with respect to the Higgs. The jets produced by vector boson decays are therefore focused in the direction of this boost. This selection ensures the jets are within the same hemisphere, which favours the VH production topology. The distribution of this variable is shown in Figure 8.8(a).
- 50 < m_{jj} < 120 GeV. The invariant mass of the two jets is required to be around the invariant masses of the $W^{\pm}$ and Z bosons. This requirement distinguishes jet pairs originating from a vector boson decay from those produced by other processes, such as $t\bar{t}$ events. The distribution of this variable is shown in Figure 8.8(b).
- b-jet veto, as described in the 1-jet category.



(a) $\Delta\eta jj$ (b) m_{jj}

Figure 8.8: The $\Delta\eta_{jj}$ and m_{jj} distributions for events that satisfy the common 1-jet and 2-jet common selections [58].



(a) Leading jet (b) Sub-leading jet

Figure 8.9: The η distributions of the leading jet after after the invariant mass and jet p_T > 40 GeV selections. The η distributions for the sub-leading jet are shown for events with at least a second jet with p_T > 25 GeV. Events from all channels are shown together [58].

After applying all selections, the distributions of the four Higgs categories are shown in Figure 8.10. The collinear mass is used for the signal mass reconstruction in the 2-jet and 1-jet categories, while the effective mass is used for the 0-jet category. The effective mass is defined as follows:

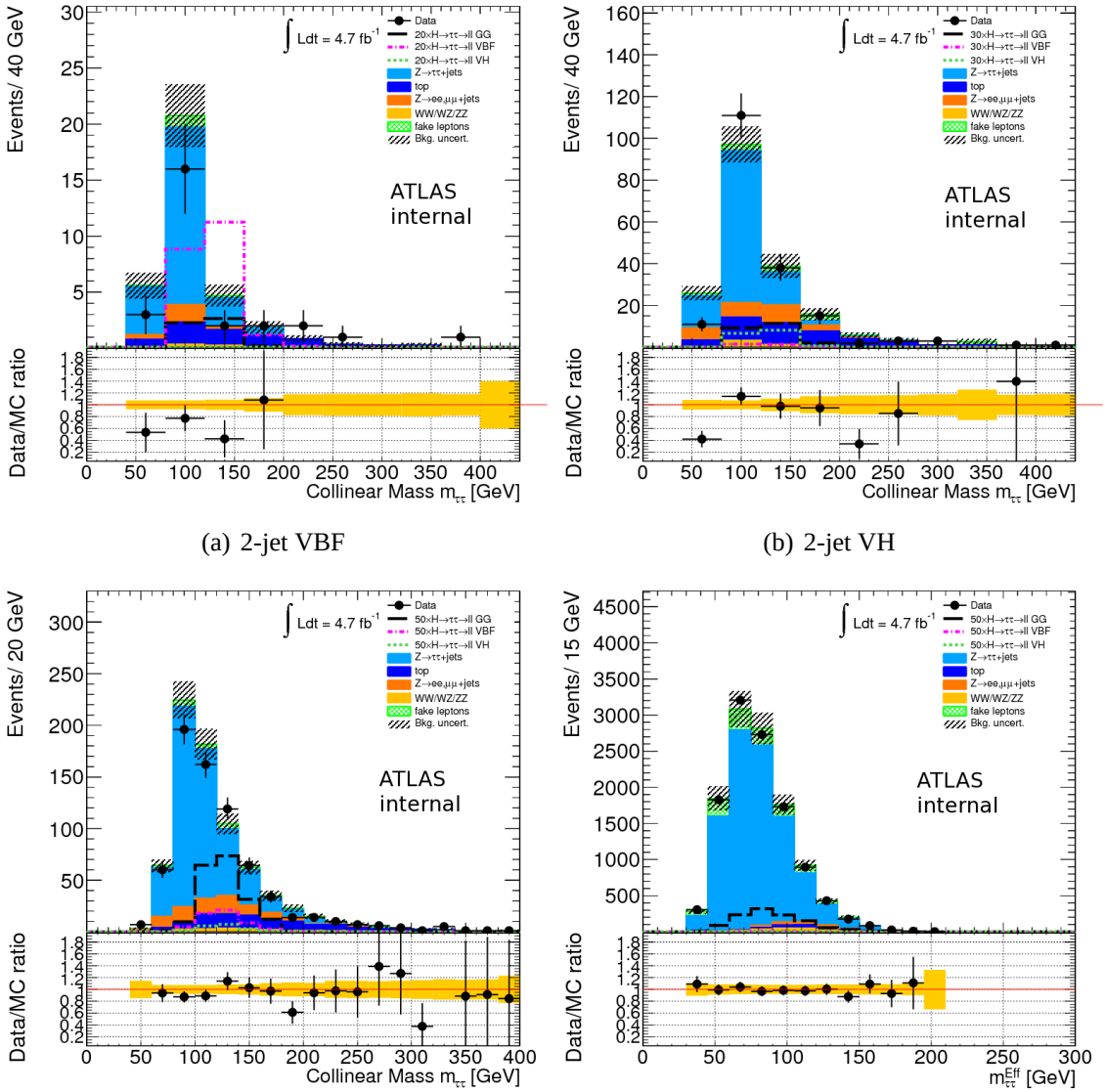
$$m_{\tau\tau}^{effective} = \sqrt{(p_e + p_\mu + E_T^{miss})\alpha(p_e + p_\mu + E_T^{miss})_\alpha}, \tag{8.5}$$

where the four-momentum of the E_T^{miss} component is defined to be:

$$(E_T^{miss})^{\alpha} = (|E_T^{miss}|, E_x^{miss}, E_y^{miss}, 0). \tag{8.6}$$

The mass reconstruction techniques used for each category are chosen because they provide different shapes for the signal and background processes. This increases the Higgs sensitivities when setting an exclusion limit, which is discussed in section 8.6.

The number of observed signal candidates and the estimated yields in the signal region are shown in Table 8.1. The expected number of signal events for a range of possible Higgs masses are shown in Table 8.2.



(a) 2-jet VBF

(b) 2-jet VH

(c) 1-jet

(d) 0-jet

Figure 8.10: The distributions of the four $H \to \tau\tau \to ll + 4\nu$ signal regions [58].

	$ee + \mu\mu + e\mu$			
	VBF category	VH category	1-jet category	0-jet category
$gg \to H$ ($m_H = 120\ GeV$)	0.26±0.06±0.10	0.8±0.1±0.2	3.9±0.2±1.0	23±1±3
VBF ($m_H = 120\ GeV$)	1.08±0.03±0.11	0.10±0.01±0.01	1.15±0.03±0.01	0.75±0.03±0.06
VH ($m_H = 120\ GeV$)	0.01±0.01±0.01	0.53±0.02±0.07	0.40±0.02±0.03	0.52±0.02±0.04
$\gamma^*/Z \to \tau\tau$	$24 \pm 3 \pm 2$	$107 \pm 12 \pm 9$	$516 \pm 11 \pm 41$	9676±50±68
$\gamma^*/Z \to ee, \mu\mu$	$2 \pm 1 \pm 1$	$25 \pm 4 \pm 9$	$83 \pm 7 \pm 27$	185±11±14
Top	$7 \pm 1 \pm 2$	$42 \pm 2 \pm 6$	$98 \pm 3 \pm 12$	169±4±14
Di-boson	$0.9 \pm 0.3 \pm 0.3$	$6 \pm 1 \pm 1.0$	$21 \pm 1 \pm 3$	221±3±18
Fake backgrounds	$1.3 \pm 0.8 \pm 0.6$	$13 \pm 2 \pm 5$	$30 \pm 4 \pm 12$	1183±13±473
Total background	$35 \pm 3 \pm 4$	193±7 ± 20	748±14 ± 52	11437±53±478
Observed data	27	185	702	11420

Table 8.1: The number of signal and background events for each of the four Higgs categories after applying all selections. The quoted uncertainties denote the statistical and systematic components, respectively [58].

Higgs boson (VBF+VH+$gg \to H$) and Background expectations				
Sample	$m_H = 100GeV$	$m_H = 110GeV$	$m_H = 115GeV$	$m_H = 120GeV$
VBF category	1.5 ± 0.1	1.4 ± 0.1	1.4 ± 0.1	1.3 ± 0.1
VH category	1.8 ± 0.2	1.9 ± 0.2	1.8 ± 0.2	1.4 ± 0.2
1-jet category	6 ± 1	7 ± 1	6 ± 1	5 ± 1
0-jet category	28 ± 3	27 ± 3	27 ± 3	24 ± 3
Sample	$m_H = 130GeV$	$m_H = 140GeV$	$m_H = 150GeV$	Bkgs
VBF category	1.0 ± 0.1	0.5 ± 0.1	0.3 ± 0.1	35 ± 5
VH category	1.0 ± 0.2	0.6 ± 0.2	0.3 ± 0.1	193 ± 19
1-jet category	5 ± 1	3 ± 1	1.3 ± 0.4	$(7.5 \pm 0.4) \cdot 10^2$
0-jet category	18 ± 2	11 ± 2	4.7 ± 0.5	$(1.14 \pm 0.05) \cdot 10^4$

Table 8.2: The expected number of signal events for a various Higgs mass hypotheses [58].

8.4 Estimation of background contributions

The main backgrounds to this analysis are estimated either using data-driven methods or simulations that have been cross checked with data-driven methods.

8.4.1 $\gamma^*/Z \to \tau\tau$ background

The $Z \to \tau\tau$ process is the main background to this analysis because it has a topology that's similar to the signal's. Ideally, this background would be estimated from data but a pure sample of $Z \to \tau\tau$ events is difficult to obtain. Furthermore, any pure sample of $Z \to \tau\tau$ events is likely to contain signal contamination.

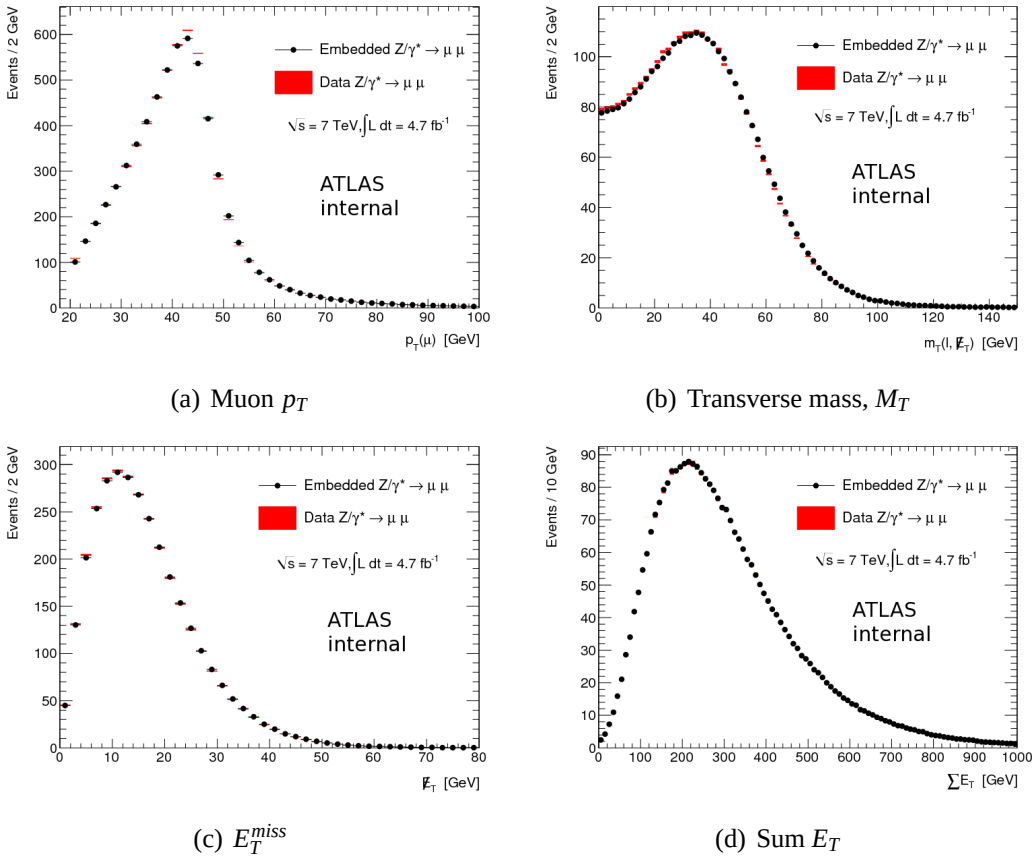
To provide the best possible estimate of this background, a pure sample of $Z \to \mu\mu$ events is selected, from which the muons are replaced with simulated tauons [58]. The $Z \to \mu\mu$ process is chosen for this task because it is easy to isolate and contains minimal signal contamination. This method is known as "embedding" and has the advantage of fully replicating the kinematics of real collision events. Assuming lepton universality, the kinematics of $Z \to \mu\mu$ and $Z \to \tau\tau$ events are expected to be almost identical, with a slight difference caused by muons and tauons having different masses. The resulting hybrid dataset is referred to as the τ-embedded $Z \to \mu\mu$ sample.

The $Z \to \mu\mu$ events are selected from data using the following requirements:

- Two oppositely charged muons with p_T > 20 GeV,
- Both muons have $p_T$cone20/p_T < 0.2,
- Both muons have a common primary vertex,
- Invariant mass $m_{\mu\mu}$ > 55 GeV.

Once selected, all detector hits associated with the daughter muons are removed. The four-vectors of these muons are then replaced by the four-vectors of the simulated tauons with the mass term and 3-momentum slightly adjusted to account for the mass difference between muons and tauons. The simulated tauons are processed by the TAUOLA and PHOTOS packages before being passed through the full ATLAS detector simulation and reconstruction algorithms.

The validation of the embedding method is performed by replacing the muons in the data events with simulated muons instead of tauons. This makes it possible to check the systematic effects of the method itself without any interference from the $Z \to \tau\tau$ modelling.



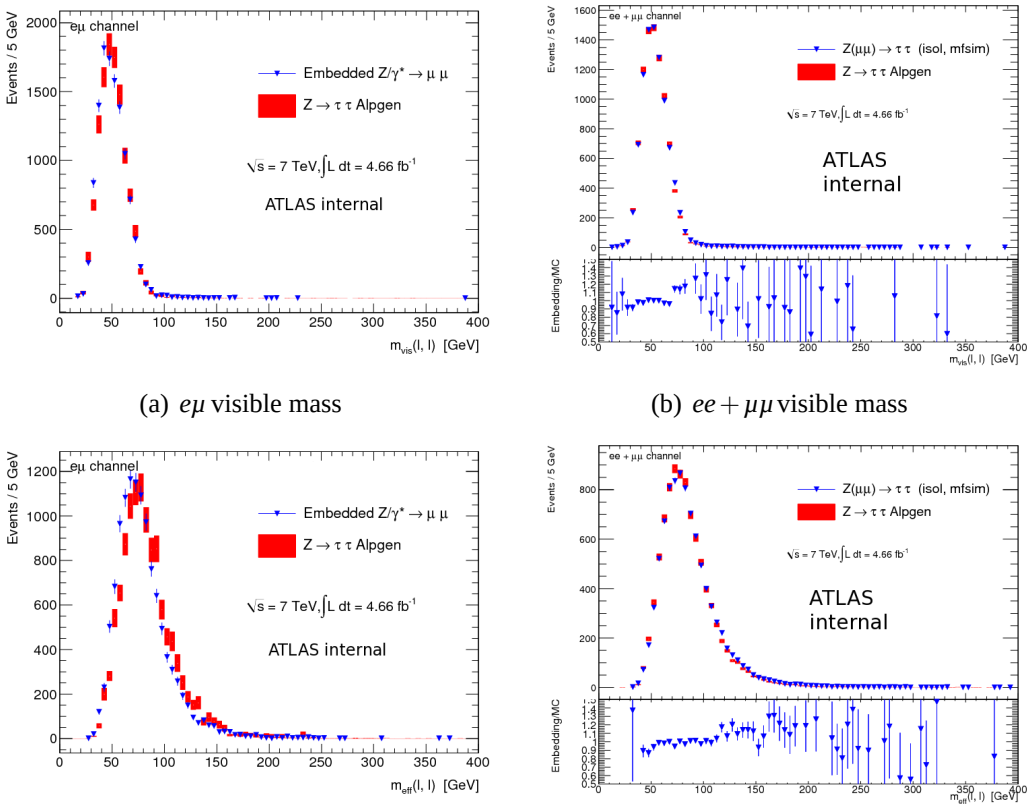
(a) Muon p_T (b) Transverse mass, M_T

(c) E_T^{miss} (d) Sum E_T

Figure 8.11: The comparisons between the data $Z \rightarrow \mu\mu$ events and the μ-embedded validation sample [58].

The key kinematic distributions of the $Z \rightarrow \mu\mu$ data events and the μ-embedded sample are shown in Figure 8.11, where a good agreement is observed.

The τ-embedded $Z \rightarrow \mu\mu$ sample is used in this analysis to model the shapes of all relevant distributions and the selection efficiencies after the triggers. However, the scale of the embedded sample cannot be easily obtained due to its biases with the trigger and dilepton selections. Instead, the ALPGEN $Z \rightarrow \tau\tau$ sample is used to provide the scale of the embedded sample after the trigger and dilepton selections. To ensure the kinematics of the ALPGEN sample are similar enough to the embedded sample, a comparison is made using the invariant mass and effective mass distributions of the two samples. The plots are shown in Figure 8.12, where a rough agreement is observed. The peak values in the mass distributions are slightly different between the embedded and ALPGEN samples. This discrepancy does not have a strong impact on the analysis since the ALPGEN sample is only used to measure the trigger and dilepton selection efficiencies. The systematic uncertainties considered for the production cross section of the Z boson is expected to be enough to cover this discrepancy.



(a) $e\mu$ visible mass

(b) $ee + \mu\mu$ visible mass

(c) $e\mu$ effective mass

(d) $ee + \mu\mu$ effective mass

Figure 8.12: The visible mass and effective mass distributions of the ALPGEN $Z \rightarrow \tau\tau$ sample and the τ-embedded $Z \rightarrow \mu\mu$ sample after the dilepton selection for the $e\mu$ and $ee + \mu\mu$ channels [58].

8.4.2 $\gamma^*/Z \rightarrow ee, \mu\mu$ background

The shapes of the kinematic distributions and the selection efficiencies of the $\gamma^*/Z \rightarrow ee, \mu\mu$ backgrounds are derived using MC. However, the scale of these backgrounds are determined using a data-driven method to reduce the uncertainties on the mismodelling of the E_T^{miss}.

Two control regions are defined after the ee and $\mu\mu$ selections for events with an invariant mass between 80 and 100 GeV. One of these regions requires events to have E_T^{miss} > 40 GeV and the other requires E_T^{miss} < 40 GeV. The chosen invariant mass window ensures a high purity of $\gamma^*/Z \rightarrow ee, \mu\mu$ events. An illustration of these regions is shown in Figure 8.13, where the control regions are labelled as "B" and "D". Region "C" is used to calculate systematic uncertainties of this method.

Any discrepancies in the tails of the E_T^{miss} distributions between the data and MC are corrected for by rescaling the MC estimate. Assuming that the E_T^{miss} is independent of the m_{ll}, the corrected estimate in the signal region, "A", is given by:

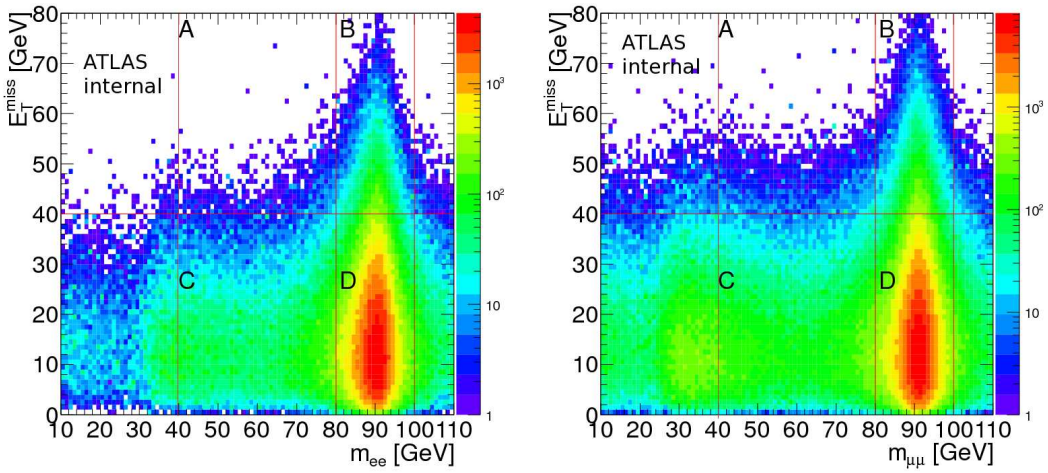


Figure 8.13: The MC distributions of the $Z \to ee$ (left) and $Z \to \mu\mu$ (right) events plotted as a function of m_{ll} and E_T^{miss} [58].

$$A_{MC}^{corrected} = A_{MC} \times \frac{B_{data}}{B_{data} + D_{data}} \frac{B_{MC} + D_{MC}}{B_{MC}}. \tag{8.7}$$

The event selections of the four categories are applied to regions "B" and "D" to measure the rescaling factors. In all control regions, the expected contributions from other electroweak processes are estimated using MC and subtracted from the data. The number of $\gamma^*/Z \to ee, \mu\mu$ events is estimated in B_{MC} and D_{MC} using the ALPGEN $\gamma^*/Z \to ee, \mu\mu$ samples. This method assumes the m_{ll} and E_T^{miss} distributions are well modelled by the MC. This is checked by comparing the ratio of events in regions "C" and "D" between data and MC. The measured differences are considered as systematic uncertainties [58]. The measured rescaling factors for each category and its systematic uncertainties are shown in Table 8.3.

Estimated γ^*/Z+jets correction	
Estimate	Rescaling factor
ee+0j	$0.91 \pm 0.01(stat) \pm 0.04(syst)$
$\mu\mu$+0j	$0.93 \pm 0.01(stat) \pm 0.02(syst)$
ee+1j	$0.86 \pm 0.02(stat) \pm 0.09(syst)$
$\mu\mu$+1j	$0.98 \pm 0.02(stat) \pm 0.01(syst)$
ee VBF	$0.87 \pm 0.05(stat) \pm 0.02(syst)$
$\mu\mu$ VBF	$1.08 \pm 0.04(stat) \pm 0.15(syst)$
ee VH	$0.95 \pm 0.04(stat) \pm 0.07(syst)$
$\mu\mu$ VH	$1.00 \pm 0.03(stat) \pm 0.08(syst)$

Table 8.3: The rescaling factors for the $\gamma^*/Z \to ee, \mu\mu$ processes [58].

8.4.3 Top backgrounds

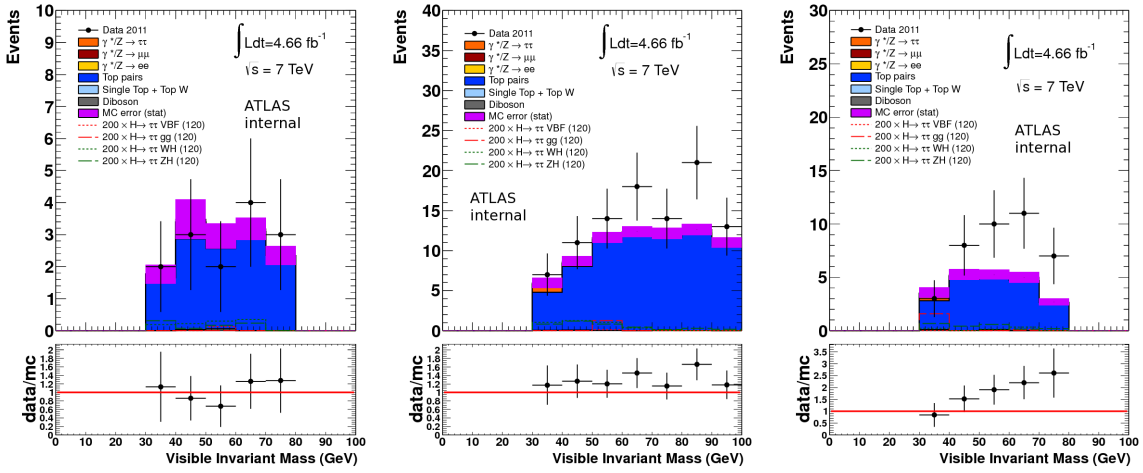
The selection efficiencies and the kinematic distributions of the top related backgrounds are estimated using MC. The predicted cross sections are checked with data-driven estimates in control regions where the purity of top events is enhanced. In the 2-jet and 1-jet categories, the control regions use the same selections as the normal analysis except with the b-tag veto selection inverted. This is the same in the 0-jet category except the H_T^{Lep} selection is inverted instead. The purity of top events is above 90% in the VBF and 1-jet categories and above 75% in the VH and 0-jet categories. The number of events in each top control region is shown in Table 8.4 for the data and MC. The m_{ll} distributions for these regions are shown in Figures 8.14, 8.15, 8.16(a) and 8.16(b) for the VH, 1-jet, VBF and 0-jet categories, respectively.

The scale of the top backgrounds are determined using the following equation:

$$R_{\mathrm{MC}}^{\mathrm{scale}} = \frac{N_{\mathrm{data}}^{CR} - N_{\mathrm{Other\ Bkgs}}^{CR}}{N_{\mathrm{top\ Bkg}}^{CR}}, \tag{8.8}$$

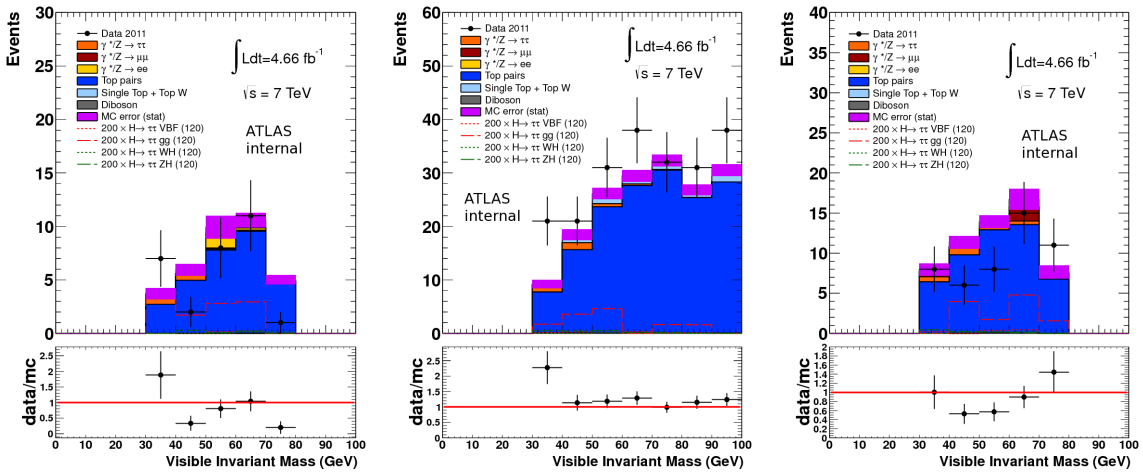
	Top Backgrounds	Other Backgrounds	Data
VBF category	6 ± 1	1.0 ± 0.5	5
VH category	106 ± 2	4 ± 1	151
1-jet category	252 ± 3	13 ± 2	289
0-jet category	474 ± 4	166 ± 6	619

Table 8.4: The number of events in the top background control regions. Only statistical errors are shown [58].



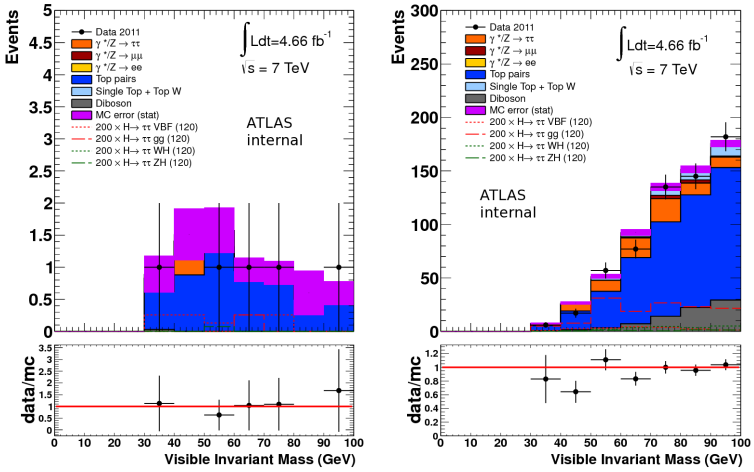
(a) ee (b) $e\mu$ (c) $\mu\mu$

Figure 8.14: Comparisons of the m_{ll} distributions for events in the VH top control regions [58].



(a) ee (b) $e\mu$ (c) $\mu\mu$

Figure 8.15: Comparisons of the m_{ll} distributions for events in the 1-jet top control regions [58].



(a) VBF all channels (b) 0-jet

Figure 8.16: Comparisons of the m_{ll} distributions for events in the VBF and 0-jet top control regions [58].

where, N^{CR}_{data} is the number of data events in the control region, $N^{CR}_{\text{Other Bkgs}}$ is the estimated contribution from non-top related processes and $N^{CR}_{\text{top Bkg}}$ is the estimated contribution from top related events. $N^{CR}_{\text{Other Bkgs}}$ and $N^{CR}_{\text{top Bkg}}$ are estimated from simulations. The measured scales are shown in Table 8.5. The systematic uncertainties are calculated by propagating all considered effects through the MC estimates. All scaling factors are consistent with unity within 2σ. These results validate the predicted cross sections of the top backgrounds and as such, they are used for the estimation of top events in the signal region.

	Top MC estimate	Scale
VBF category	6 ± 1	$0.82 \pm 0.46 \pm 0.13$
VH category	41 ± 1	$1.39 \pm 0.12 \pm 0.12$
1-jet category	105 ± 2	$1.09 \pm 0.07 \pm 0.10$
0-jet category	171 ± 3	$0.96 \pm 0.05 \pm 0.07$

Table 8.5: Expected number of top background events in the signal regions and the relative scale factors. The first uncertainty is the statistical component and the second is the systematic. For the top MC estimate, only the statistical errors are shown [58].

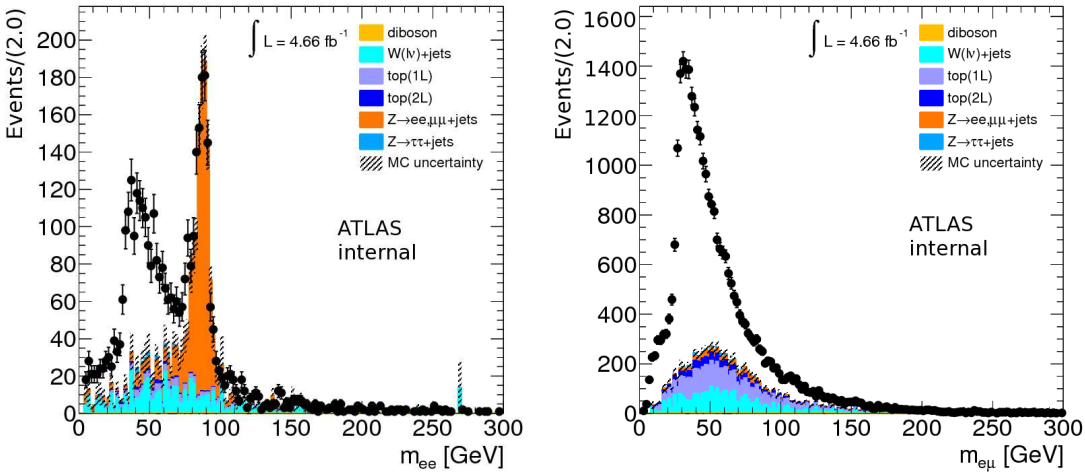
8.4.4 Fake background

The fake background includes all processes with at least one reconstructed lepton not originating from a tauon or vector boson decay. This background mainly consists of multijet, W+jets and $t\bar{t}$ events. To avoid double counting, the zero and one lepton $t\bar{t}$ decays are removed from the top simulated samples. The scale, shapes and selection efficiencies of the fakes are estimated from data.

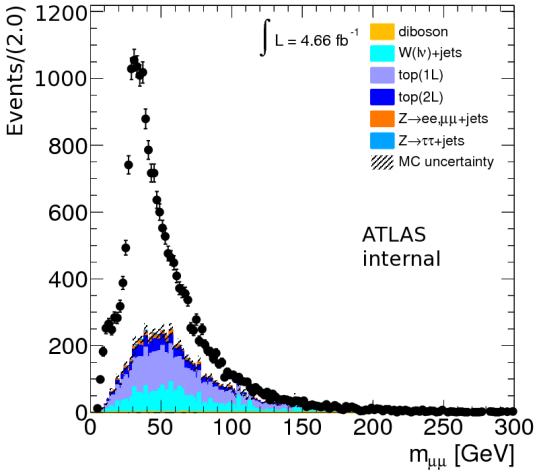
The shapes of the fake distributions are derived using control regions where the purity of these events are enhanced. The selections used for these regions are the same as nominal analysis but with the following changes:

- Exactly one of the leptons fails the $p_T$cone requirement,
- Leptons are required to have the same charge (No charge requirement for the *ee* channel),
- Electron identification requirement reduced to "medium".

The requirements of the fake control regions are chosen to select fake distributions with compositions that resemble those expected in the signal region. By requiring exactly one of the leptons to fail the track isolation requirement, contributions from heavy-flavoured multijets are reduced. Heavy-flavoured multijets are not expected to have significant contributions because they are likely to fail the isolation requirements. The $E_T$cone selection is kept nominal in all channels to reduce contributions from real leptons that undergo bremsstrahlung in the inner detector. For the W+jets and light-flavoured di-jet events, the charge of the lepton candidates are expected to be uncorrelated and therefore the same-sign requirement favour these events and reduces signal contamination. In the *ee* channel, opposite-signed events are also used to increase statistics. All electroweak contaminations are estimated from simulations and subtracted from these regions. The purity of the fake control regions are shown in figure 8.17 for events that satisfy the dilepton selections.



(a) ee (b) $e\mu$



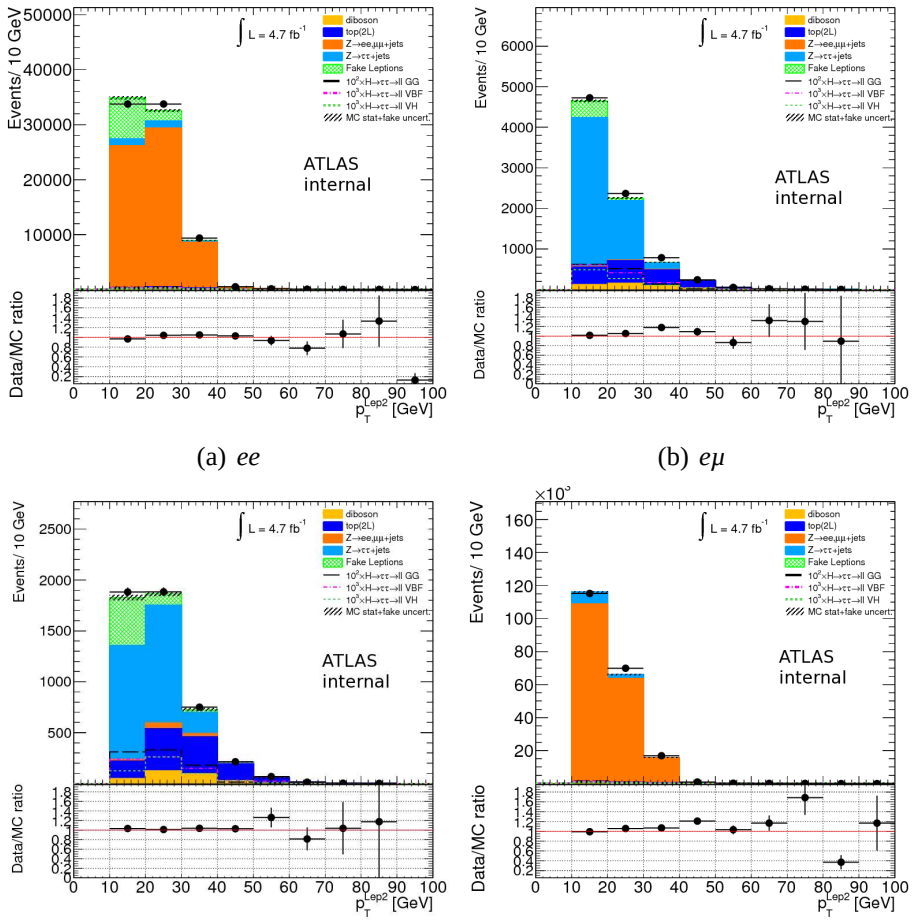
(c) $\mu\mu$

Figure 8.17: The m_{ll} distributions of fake control regions after applying the dilepton selection [58].

Given the tight requirements of the lepton isolation and the E_T^{miss}, the fakes are expected to have a small contributions in the signal region. Thus, the scale of this background is calculated after the dilepton and invariant mass selections using the "template" method [79]. The template method takes a particular variable distribution in data and subtracts the expected contributions from all non-fake processes. The remaining events are assumed to be composed entirely of fakes and using these shapes, the scale which best fits the data distribution is used.

In this analysis, the p_T of the sub-leading lepton is used to determine the fake scale in all channels. In addition, the $e\mu$ channel is split further into two sub-channels determined by whether the leading lepton is a muon (μe) or an electron ($e\mu$). In this way, the fake

background is more accurately described, since its composition may differ between μe and $e\mu$ events. The p_T of the sub-leading lepton is chosen because its distribution shape is different for fakes and other electroweak events. The distributions of the fake events are compared between the control and signal regions, which is shown in Figure D.1 as a function of the sub-leading lepton p_T. The distributions agree within the uncertainties, which indicates that the fake compositions are similar between the signal and control regions. The expected fake contributions are shown in Figure 8.18 as a function of the sub-leading lepton p_T.



(a) ee (b) $e\mu$

(c) μe (d) $\mu\mu$

Figure 8.18: The p_T distributions of the sub-leading lepton including the fake contributions estimated using the template method [58].

The selection efficiencies of the fake background are calculated by passing the events in the control regions through the selections of the analysis. Assuming the scale is independent of the selection efficiencies, the fake contribution at each selection stage is estimated by multiplying the efficiencies with the template fitted scale.

8.5 Systematic uncertainties

This section details the systematic uncertainties considered for this analysis. These include the resolution and scale factors of reconstructed particles, the pileup conditions, the cross section of the signal and backgrounds and the integrated luminosity.

8.5.1 Jet energy scale and resolution

The corrections for jet energy scale contain a number of different effects, which are considered as systematic uncertainties. These are detailed in references [80] and [81]. All uncertainties associated with the jet energy scale are combined in quadrature to obtain an overall uncertainty. The energy scale varies between 2-7% depending on the p_T and η of the jets. The $\pm\sigma$ variations are propagated through this analysis to determine its effect on the signal and background estimates.

The methods used to determine the uncertainties in the jet energy resolution are described in reference [82]. The systematic uncertainty of the jet energy resolution is determined by increasing the resolution width by 1σ and propagating the variation through all simulated estimates.

8.5.2 E_T^{miss} reconstruction

All systematic uncertainties pertaining to the scale and resolution of reconstructed particles are also propagated to the E_T^{miss}. Systematic uncertainties specific to the E_T^{miss} include the uncertainties on the low p_T jets and the energy deposits that are not associated with any reconstructed particle. These effects are considered by coherently varying the soft jet and cell out terms by $\pm$10% and propagating the variations through all simulated estimates.

8.5.3 Fake background estimation

The uncertainties of the fake background estimation are derived by varying the p_T range of the sub-leading lepton from which the template fit is performed. The maximal deviation from the nominal value is taken as the systematic uncertainty, which is measured to be 40%.

8.5.4 SM cross sections

For the background samples, an uncertainty of 4% is considered for the cross section of γ^*/Z production. For γ^*/Z events with at least one additional jet, an additional uncertainty

of 24% is considered. Different predictions between the ALPGEN and PYTHIA MC are also considered and is measured to be about 2%. For $t\bar{t}$ and single-top events, an uncertainty of 8% is used [83] and an uncertainty of 4% is used for all diboson processes.

For the signal events, the uncertainties on their cross sections depend on m_H. Uncertainties of 15 - 25% are used for the gluon-gluon fusion production and 2.5 - 3% is considered for the VBF and VH productions [84].

8.5.5 $\gamma^*/Z \rightarrow \tau\tau$ embedding

The systematic uncertainty of the embedding method is obtained by using a sample of $\gamma^*/Z \rightarrow \mu\mu$ events where the track isolation requirement is removed. The difference between the nominal $\gamma^*/Z \rightarrow \mu\mu$ and alternate sample is taken as the uncertainty.

8.5.6 Parton distribution functions

The uncertainty of the proton PDF results is an additional source of uncertainty for the signal and background cross sections. An uncertainty of 3% is used for all background processes. For the signal, 7.8% is considered for the gluon-gluon fusion process and 2.3% is considered for the VBF and VH processes [84].

The systematic uncertainties measured for this analysis are shown in Tables D.1, D.2, D.3 and D.4 for the VBF, VH, 1-jet and 0-jet categories, respectively.

8.6 Higgs limit setting

The exclusion limits on the production of the Higgs boson are determined as a function of the Higgs mass. The procedure used to compute the limits is based on the modified frequentist method known as CLs [85]. This method is used to determine the consistency of the observed events with a background-only or a signal+background hypothesis. These are measured in terms of likelihoods, which are calculated from the signal region distributions shown in Figure 8.10.

The uncertainties on the background and signal+background estimates are assumed to have Gaussian distributions. For both hypotheses, each source of uncertainty is randomly varied about their central values, over thousands of iterations, to measure the average variance. This is then used to measure the likelihoods of each hypothesis. For the uncertainties that affects the shapes of the mass distributions, the variations are performed on a bin-by-bin basis to achieve a more precise estimate. In this analysis, only the jet energy scale variations are performed in this manner. For all other systematics, only the scale is varied.

Correlations between the systematic uncertainties are also taken into account in the CLs method, where the uncertainties are assumed to be either fully correlated or fully uncorrelated. For uncertainties that are fully correlated, their variations are performed coherently. In this analysis, the uncertainties on the luminosity, energy scale and acceptance are assumed to be fully correlated and all other uncertainties are assumed to be uncorrelated.

The likelihood function used to determine the limits in the four signal categories, denoted by j, is given by:

$$L(\mu,\theta) = \prod_{j=1}^{4} \text{Poisson}[N_j|\mu(s_j^{gg\to H} + s_j^{VBF} + s_j^{VH}) + b_j] \prod_{\theta} \text{Gaussian}(\theta|0,1) \tag{8.9}$$

where μ is the signal strength, θ are the sources of uncertainty, N is the number of observed events, s is the number of expected signal events and b is the number of expected background events.

The test statistic, q, measures how well each hypothesis fits the observed distributions by comparing them to the best fit scenario. They are defined as follows:

$$q_0 = -2\, ln \frac{L(0,\hat{\theta}_0^{obs})}{L(\hat{\mu},\hat{\theta})} \text{ (background-only)}, \tag{8.10}$$

$$q_\mu = -2\, ln \frac{L(\mu,\hat{\theta}_\mu^{obs})}{L(\hat{\mu},\hat{\theta})} \text{ (signal+background)}, \tag{8.11}$$

where $L(\hat{\mu},\hat{\theta})$ denotes the maximum likelihood fit of the data points for the given θ uncertainties with the constraint $0 \leq \hat{\mu} \leq \mu$. The lower bound on $\hat{\mu}$ ensures that the signal yield is positive and the upper bound ensures that an excess of observed events is described by the signal+background hypothesis. $L(0,\hat{\theta}_0^{obs})$ and $L(\mu,\hat{\theta}_\mu^{obs})$ denote the maximum likelihood fit of the data assuming a background-only or a signal+background hypothesis, respectively.

The probability distribution functions of the two hypotheses are constructed from the test statistics, q_0 and q_μ, which are denoted by $f(q_\mu|\mu,\hat{\theta}_\mu^{obs})$ and $f(q_0|0,\hat{\theta}_0^{obs})$, respectively. The probabilities of observing an excess above the given data points is determined by integrating these functions from the measured q values to all possible higher q values (assuming different values of μ). The probabilities are measured in terms of p-values for the two hypotheses:

$$p_0 = \int_{q_0}^{\infty} f(q_0|0,\hat{\theta}_0^{obs})dq_0 \text{ (background-only)}, \tag{8.12}$$

$$p_\mu = \int_{q_\mu}^{\infty} f(q_\mu|\mu,\hat{\theta}_\mu^{obs})dq_\mu \text{ (signal+background)}. \tag{8.13}$$

The confidence limit of the signal, $CLs(\mu)$, is calculated as a ratio of the two p-values:

$$CLs(\mu) = \frac{p_\mu}{p_0}. \tag{8.14}$$

For $CLs = \alpha$, the inverse, $(1-\alpha)$, is often quoted as the confidence level. The exclusion limits are expressed relative to the SM Higgs production cross section. Figure 8.19 shows the exclusion limits as a function of the Higgs mass. The most sensitive Higgs masses are between 100 - 130 GeV, where an exclusion of 5-6 times the SM cross-section is observed. For the higher mass range, the exclusion limit decreases to 7-15 times the SM predictions.

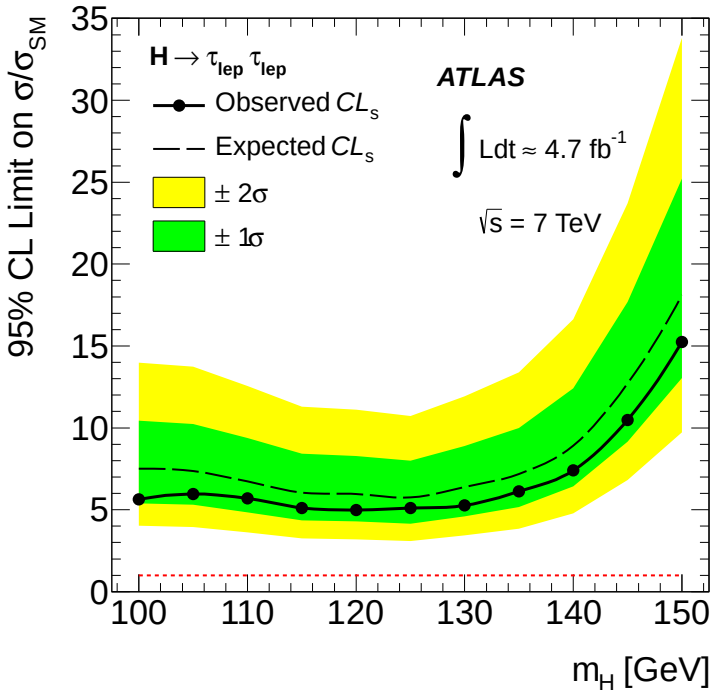


Figure 8.19: Exclusion limits of the $H \to \tau\tau \to ll + 4\nu$ analysis. The expected and observed 95% confidence-level limits are shown as solid and dashed lines, respectively [86].

9

$H \to \tau\tau$ combined limit

The search for the Higgs boson is also conducted in the $H \to \tau\tau \to l + \tau_h + 3\nu$ and $H \to \tau\tau \to \tau_h\tau_h + 2\nu$ final states. This chapter briefly describes the analyses used for these searches and the exclusion limits they achieve both individually and in combination with all the $H \to \tau\tau$ search channels.

9.1 Event selections

Triggers

For the $l\tau_h$ search channels, a single "medium" electron trigger with a p_T > 20 GeV requirement is used for the $e\tau_h$ channel and the a muon trigger with a p_T > 18 GeV requirement is used for the $\mu\tau_h$ channel.

For the $\tau_h\tau_h$ search channel, a di-τ_h trigger is used, which has p_T thresholds of 29 GeV and 20 GeV for the leading and sub-leading hadronic tauons, respectively.

$H \to \tau\tau \to l + \tau_h + 3\nu$ selections

Events in this channel are required to have exactly one isolated electron (or muon) with one hadronic tauon of opposite charge. The isolation requirements on the electron or muon are the same as the ones used in $H \to \tau\tau \to ll + 4\nu$ analysis. The electron or muon have p_T > 25 GeV and p_T > 20 GeV, respectively. For the τ_h candidate, p_T > 20 GeV is required.

To reduce contributions from W + jets and $t\bar{t}$ processes, the transverse mass is required to be above 30 GeV. Events that satisfy these selections are split into seven exclusive categories based on their jet properties and their E_T^{miss} value. Separating events into different categories

achieves a better signal to noise ratio. This analysis has one H+2-jet VBF category, two H+1-jet categories and four H+0-jet categories, which require events to have at least two jets, at least one jet and exactly zero jets, respectively.

The H+2-jet VBF category requires events to have a E_T^{miss} > 20 GeV and at least two jets with p_T > 25 GeV. The two leading jets are required to be in opposite hemispheres ($\eta_{j1} \cdot \eta_{j2}$ < 0), have a $\Delta\eta_{jj}$ > 3 and have an invariant mass (m_{jj}) greater than 300 GeV. The lepton and τ_h candidate must also be within the pseudorapidity range of the two leading jets. The $e\tau_h$ and $\mu\tau_h$ events are combined in this category due to the limited number of events.

The two H+1-jet categories requires events to have E_T^{miss} > 20 GeV and at least one jet with p_T > 25 GeV. Only the events that fail the VBF selections are included. The $e\tau_h$ and $\mu\tau_h$ events are analysed separately.

The four H+0-jet categories includes all events that have no jets with p_T > 25 GeV. The events are categorised on whether they are $e\tau_h$ or $\mu\tau_h$ final states and whether they have E_T^{miss} > 20 GeV or E_T^{miss} < 20 GeV.

For each category, the mass of the $\tau\tau$ pair is reconstructed using a method that exploits the relative orientations of the neutrinos and leptons that are consistent with the kinematics of a τ-lepton decay. This method is referred to as the Missing Mass Calculator (MMC) and has a reconstruction efficiency of 99% and 13 - 20% resolution on the mass [87].

$H \rightarrow \tau\tau \rightarrow \tau_h\tau_h + 2\nu$ selections

In the $H \rightarrow \tau\tau \rightarrow \tau_h\tau_h + 2\nu$ analysis, events are required to have exactly two oppositely charged τ_h candidates with p_T > 35 GeV and p_T > 25 GeV to be above the di-τ_h trigger thresholds. Events are rejected if they contain any isolated leptons to reduce the background contributions from electroweak processes.

Only a single H+1-jet category is considered in this search channel. The backgrounds in this channel mainly come from multijets that fake τ_h candidates and $Z \rightarrow \tau\tau$ events where both tauons decay hadronically. After the two τ_h selection, a 0 < x_1,x_2 < 1 selection is applied, where x_1 and x_2 are the collinear approximation momentum fractions (section 8.3). The collinear approximation is used for the reconstruction of $m_{\tau\tau}$ in this channel and this selection ensures that the solutions are always real. The events are also required to have E_T^{miss} > 20 GeV, at least one jet with p_T > 40 GeV, an angular separation between the two τ_h candidates that satisfies $\Delta R(\tau,\tau)$ < 2.8 and the invariant mass of the $\tau\tau$ pair with the leading jet ($m_{\tau\tau j}$) to be above 225 GeV.

9.2 Background estimations

$H \to \tau\tau \to l + \tau_h + 3\nu$ estimation

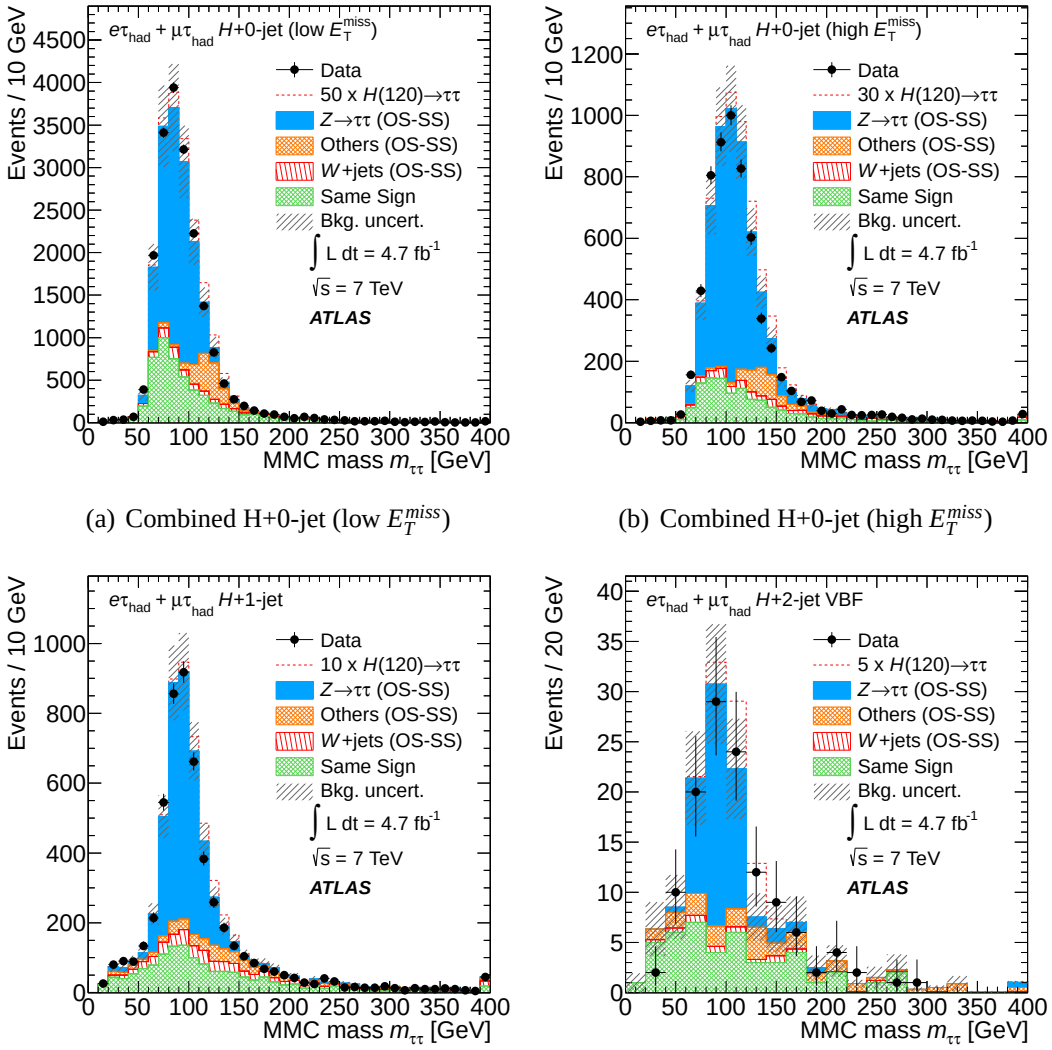
The background contribution is estimated using a control region with the same event selections as the signal but with the opposite charge (OS) requirement changed to a same charge (SS) requirement. The number of background events (n_{OS}^{bkg}) can be expressed as:

$$n_{OS}^{bkg} = n_{SS}^{all} + n_{OS-SS}^{W+jets} + n_{OS-SS}^{Z\to\tau\tau} + n_{OS-SS}^{other} \tag{9.1}$$

where n_{SS}^{all} is the number of SS events observed in data control region and the remaining terms are the differences between the number of OS and SS events for the considered backgrounds. For the multijet background, $n_{OS-SS}^{multijet} = 0$ is assumed, since the charge of the fake τ_h candidate is expected to be uncorrelated with the charge of the lepton. This assumption is validated in a multijet enriched control sample where a value of $r_{OS/SS}^{multijet} = 1.10 \pm 0.01(stat.) \pm 0.09(syst.)$ is observed [86]. The main advantage of using this method is that the fake distributions from all processes are derived from data. This reduces any uncertainties related to the mismodelling of fakes in the electroweak backgrounds and provides a data-driven estimate of the multijet background.

The $\gamma^*/Z \to \tau\tau$ contribution is estimated using the τ-embedded $Z \to \mu\mu$ samples (8.4.1). For the W+jets background, the shapes of the kinematic distributions are obtained from MC and the scale is calculated from data using a W enriched control region. For the remaining backgrounds, MC estimates are used.

The MMC is used to define the mass spectrum of this analysis, as it provides a better discrimination between the signal and the non-$\gamma^*/Z \to \tau\tau$ backgrounds [87]. The MMC distributions for data with the expected signal and background contributions are shown in Figure 9.1.



(a) Combined H+0-jet (low E_T^{miss})

(b) Combined H+0-jet (high E_T^{miss})

(c) Combined H+1-jet

(d) H+2-jet VBF

Figure 9.1: The MMC distributions in the signal regions for the $H \rightarrow \tau\tau \rightarrow l + \tau_h + 3\nu$ analysis categories [86].

$H \to \tau\tau \to \tau_h\tau_h + 2\nu$ estimation

The $\gamma^*/Z \to \tau\tau$ and multijet processes are the dominant backgrounds in this analysis and are estimated using data-driven methods. All other backgrounds are estimated using simulations.

The kinematic distribution of the $\gamma^*/Z \to \tau\tau$ background is obtained using the embedded sample. The scale of this background is calculated from a control region by fitting the track multiplicity in a cone size of ΔR = 0.6 around the reconstructed τ_h candidates. The $\gamma^*/Z \to \tau\tau$ shapes are modelled from the simulation and the multijet shapes are derived from data using SS events. All other processes are modelled from simulations. The track multiplicity of multijet events is expected to be larger than the $\gamma^*/Z \to \tau\tau$ events. The two distributions are expected to be separated well enough for a clean fit to be made. The scale of the multijet background is estimated using the same fitting method as described for the $\gamma^*/Z \to \tau\tau$ background using the track multiplicity. However, this fit is performed in the signal region after all event selections.

The mass distribution of the signal events is shown in figure 9.2.

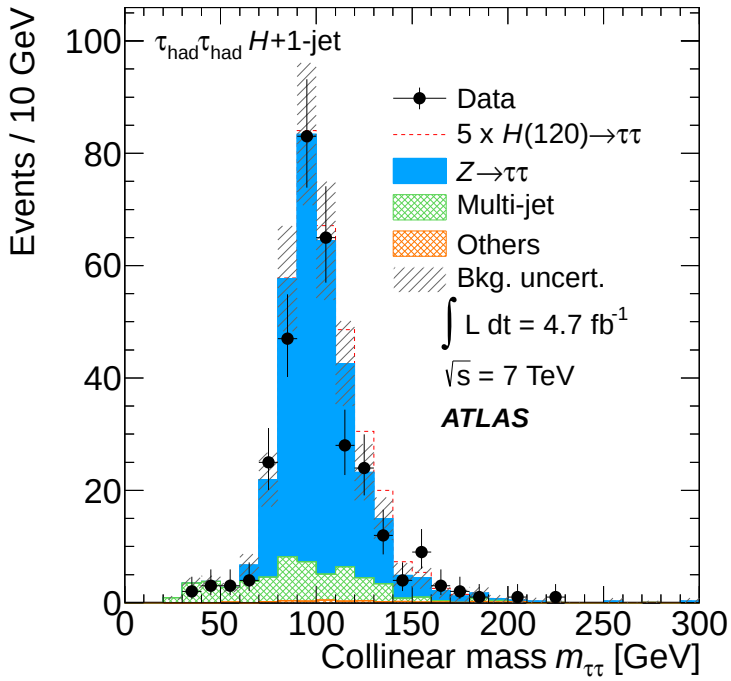


Figure 9.2: The signal region of the $H \to \tau\tau \to \tau_h\tau_h + 2\nu$ search channel [86].

9.3 Systematic uncertainties

The systematic uncertainties on to the production cross sections, the integrated luminosity, the jets, the lepton efficiencies and the E_T^{miss} are evaluated using the same methods as the $H \rightarrow \tau\tau \rightarrow ll + 4\nu$ analysis. The only systematics that are unique to the hadronic $H \rightarrow \tau\tau$ search channels are the uncertainties on the τ_h identification and energy scale.

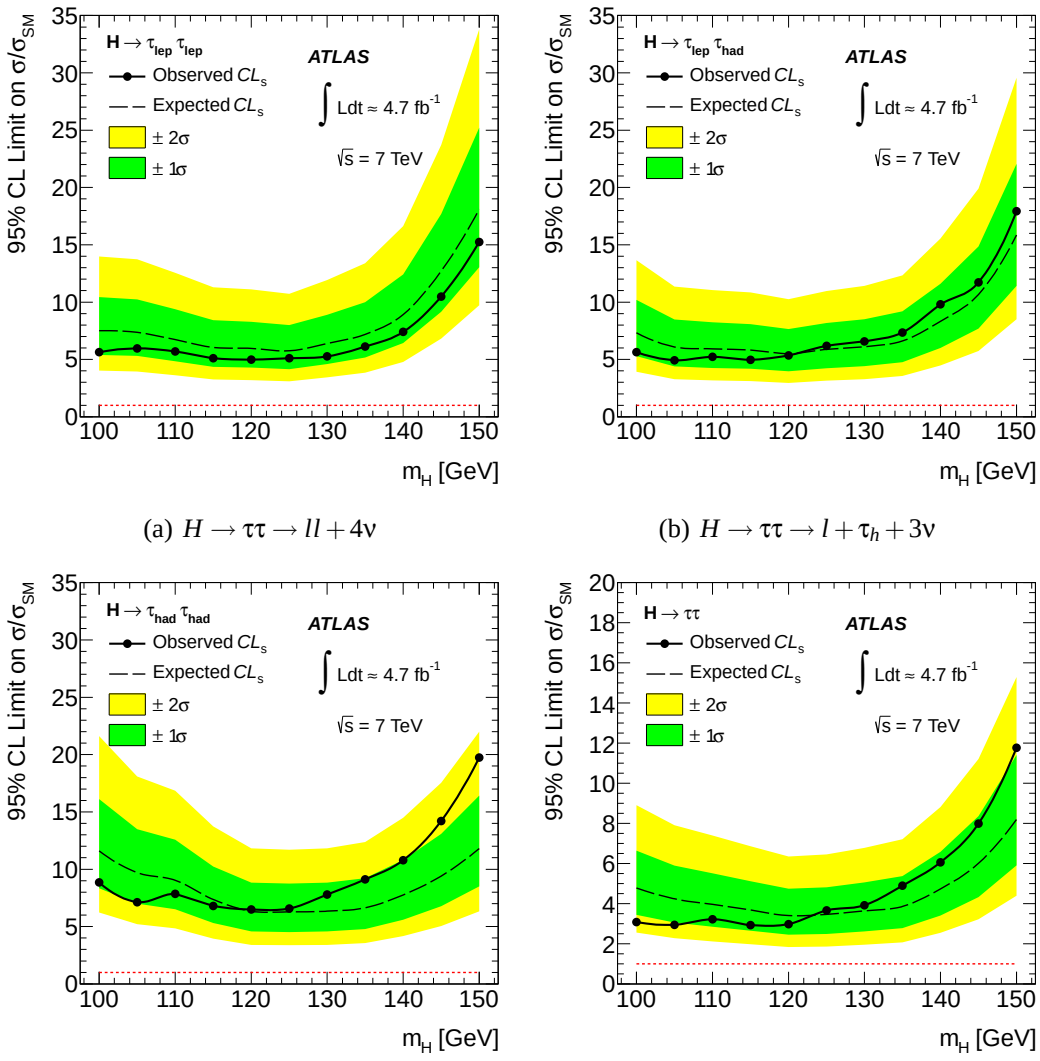
The difference in the τ_h identification efficiency between data and MC is calculated by comparing the number of $W \rightarrow \tau\nu$ and $\gamma^*/Z \rightarrow \tau\tau$ events in control regions where the purity of these processes are enriched. The efficiency difference is found to be less than 4%.

The τ_h energy scale is calibrated by comparing the single hadron responses in the calorimeters between the collision data and the 2004 test beam data [88]. The resulting calibrations are then validated in an enriched data sample of $\gamma^*/Z \rightarrow \tau\tau$ events. The uncertainties in the scale are found to be between 2 - 5%.

9.4 Exclusion limits

The combined exclusion limits are calculated from the twelve $H \rightarrow \tau\tau$ search categories. Systematic uncertainties that are common between the different search channels are considered to be fully correlated and all other systematic uncertainties are considered to be uncorrelated. The limits set using the $H \rightarrow \tau\tau$ analyses are shown in Figure 9.3, for the three sub-channels individually and in combination.

The combined results are most sensitive in the mass region between 100 and 120 GeV, where the observed limit is approximately 3 times the SM prediction. For the higher mass regions, the exclusion power of these analyses decrease to approximately 6 times the SM prediction at m_H = 140 GeV and 12 times at m_H = 150 GeV.



(a) $H \rightarrow \tau\tau \rightarrow ll + 4\nu$

(b) $H \rightarrow \tau\tau \rightarrow l + \tau_h + 3\nu$

(c) $H \rightarrow \tau\tau \rightarrow \tau_h\tau_h + 2\nu$

(d) $H \rightarrow \tau\tau$ combined

Figure 9.3: The exclusion limits set by the $H \rightarrow \tau\tau$ analysis channels shown individually and in combination. The excluded production cross sections are shown as a ratio over the expectations of the Standard Model (σ/σ_{SM}) as a function of the Higgs mass (m_H) [86].

10

ATLAS Higgs combination

The ATLAS experiment searches for the Higgs boson across many final state channels. Besides the $H \to \tau\tau$ channel, which has already been described, searches have also been conducted in the $H \to b\bar{b}$ [89], $H \to \gamma\gamma$ [90], $H \to ZZ \to 4l$ [91] and $H \to WW \to l\nu l\nu$ [92][93] search channels.

10.1 7 TeV combination

The results from the $H \to \tau\tau$, $H \to b\bar{b}$, $H \to \gamma\gamma$, $H \to WW$ and $H \to ZZ$ search channels are combined to achieve results with greater statistical significance. The correlations between the systematic uncertainties are taken into account and are assumed to be either fully correlated or fully uncorrelated. The correlated uncertainties mainly include: the integrated luminosity; the efficiency of the electron and photon trigger identification; the electron and photon energy scales; the jet energy scale and E_T^{miss}; and the theoretical cross sections of the signal and background processes [94]. All other systematic uncertainties are considered to be fully uncorrelated.

The exclusion limits achieved by each search channel and their combined limits are shown in Figure 10.1. The majority of the mass points between 110 and 150 GeV are excluded at 95% confidence. A slight excess of events is observed above the background expectations in the $H \to \gamma\gamma$ and $H \to ZZ$ channels at a hypothesised Higgs mass of 126 GeV. These are quantified in terms of local p_0 values, which are shown in Figure 10.2. The excess has a combined significance of 2.9 σ.

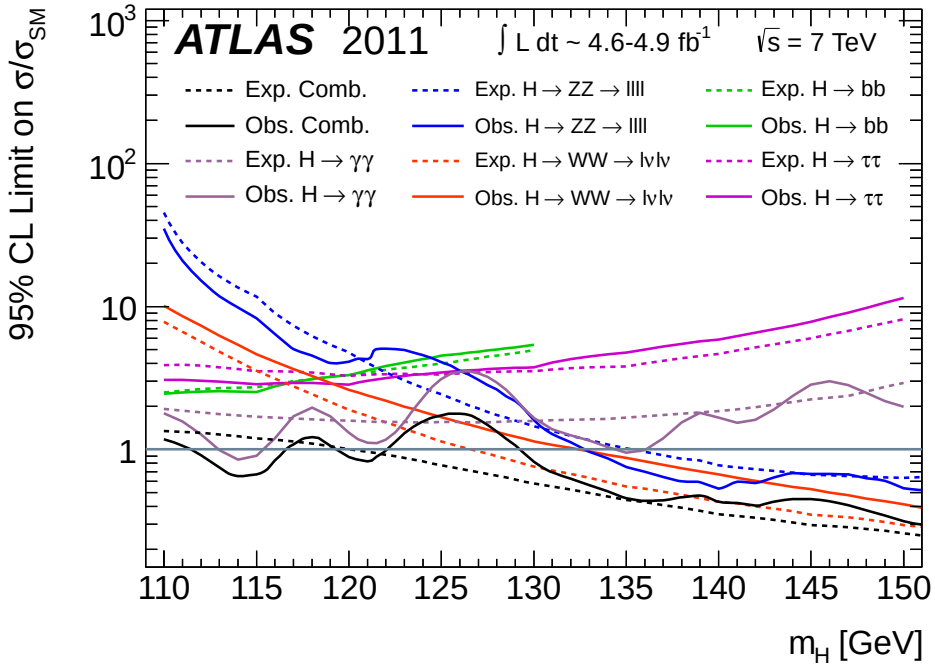


Figure 10.1: The exclusion limits achieved by Higgs search channels using the $\sqrt{s}$ = 7 TeV dataset [95].

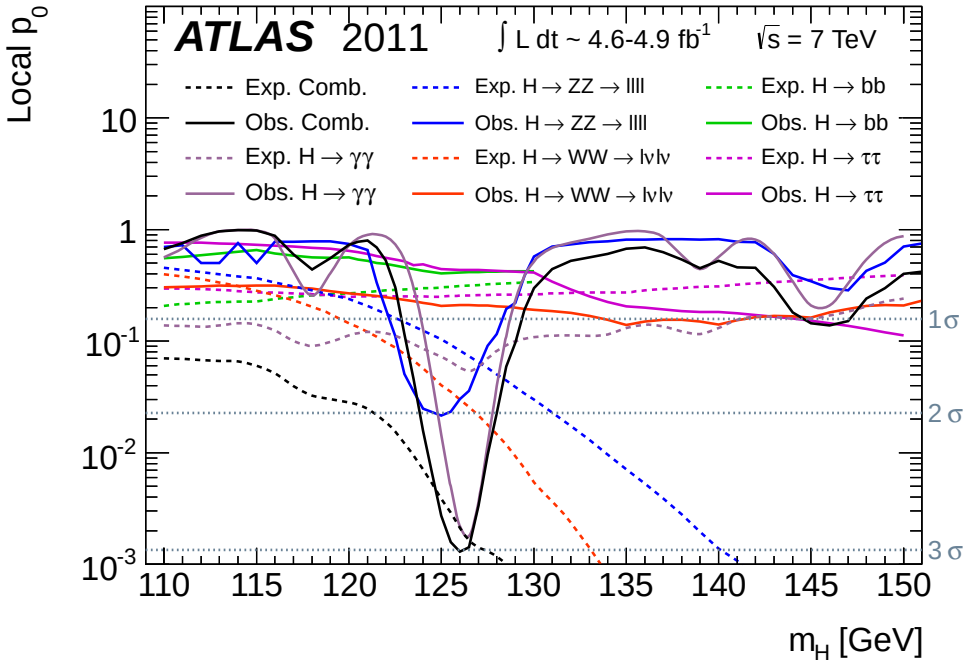


Figure 10.2: The local p_0 values for the Higgs search channels using the $\sqrt{s}$ = 7 TeV dataset [95].

10.2 7 TeV and 8 TeV combination

The $H \to \gamma\gamma$, $H \to ZZ$ and $H \to WW$ analyses are also performed on the $\sqrt{s}$ = 8 TeV dataset. The combined limits of the $\sqrt{s}$ = 7 TeV and $\sqrt{s}$ = 8 TeV datasets are shown in Figure 10.3, where most mass points below 500 GeV are excluded at 95% confidence. The only region that isn't excluded is around m_H = 126 GeV, where an excess is observed in the $H \to \gamma\gamma$, $H \to ZZ$ and $H \to WW$ search channels. The Higgs mass at which the excess is observed is consistent between the $\sqrt{s}$ = 7 TeV and $\sqrt{s}$ = 8 TeV datasets. Figure 10.4 shows the combined local p_0 values, where a significance of 5.9 σ is measured.

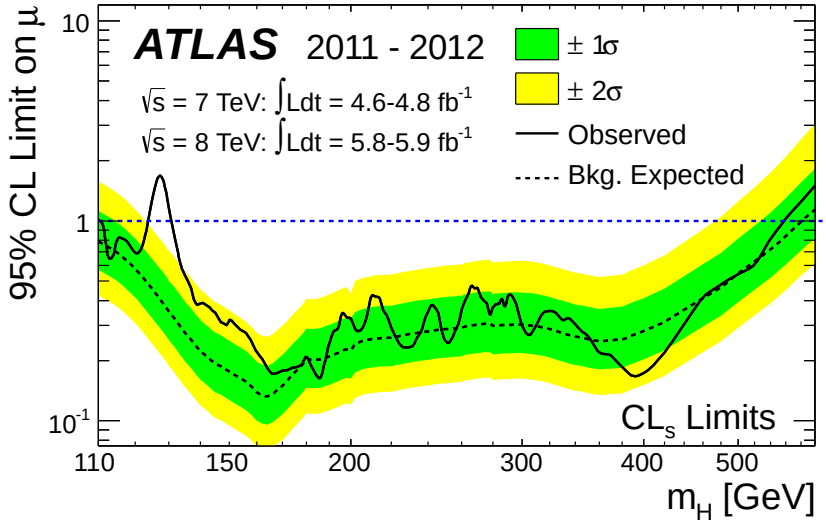


Figure 10.3: The combined exclusion limits of the ATLAS Higgs search channels.

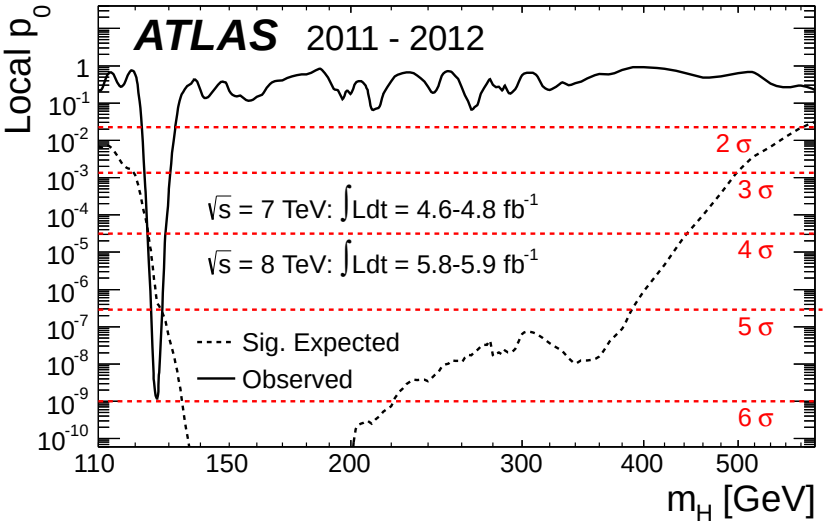


Figure 10.4: The combined local p_0 values of the ATLAS Higgs search channels.

The best fit signal strengths of the observed excess are shown in Figure 10.5 for the various search channels and their combination. A signal strength of $\mu = 1.4 \pm 0.3$ is found to best fit the data using a Higgs mass hypothesis of m_H = 126 GeV.

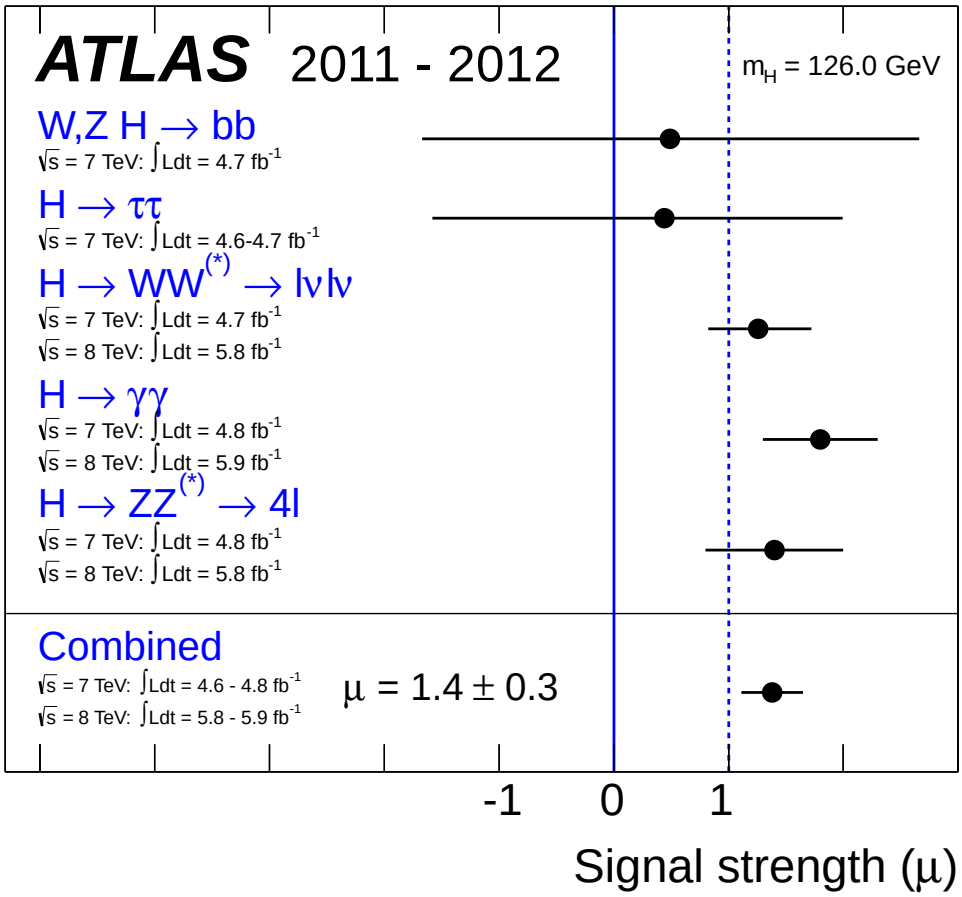


Figure 10.5: The best fit signal strengths for the various Higgs search channels and their combination.

The local p_0 value measured at m_H = 126 GeV exceeds the 5σ limit required to claim discovery. ATLAS claims the discovery of a new particle with a best fit mass of 126 $\pm$ 0.4 (stat.) $\pm$ 0.4 (sys.) GeV. This particle has been observed to decay into a pair of gauge bosons, which means it must be a boson. The measured excess is also consistent with the expected yields of the SM Higgs. Thus far, all measured properties of this new particle are consistent within the uncertainties of a SM Higgs.

11

Conclusions and outlook

This thesis began by introducing spontaneous electroweak symmetry breaking as way of generating masses for the electroweak gauge bosons. This mechanism predicts the existence of a massive scalar particle, the Higgs boson. The Higgs interaction terms are obtained by expanding around the minimum potential and from these terms, the Higgs decay modes and expected cross sections were calculated and shown.

To achieve the most from the $H \rightarrow \tau\tau \rightarrow ll + 4\nu$ analysis, the multijet and $Z \rightarrow \tau\tau$ backgrounds were studied in further detail to ensure these processes are well modelled. The multijet composition study showed that the contributions in the ee and $e\mu$ channels are dominated by light-flavoured jets faking leptons, while in the $\mu\mu$ channel, they're dominated by leptonically decaying $b\bar{b}$ pairs. The production cross section of the $Z \rightarrow \tau\tau$ process was measured in the $e\mu$ and $\mu\mu$ decay channels and combined with the measurements from the $e\tau_h$ and μ_h channels for better precision. A combined total cross section of 0.97 $\pm$ 0.07 (stat.) $\pm$ 0.06 (syst.) $\pm$ 0.03 (lumi.) was measured, which is in agreement with the NNLO predictions and the CMS measurements.

The $H \rightarrow \tau\tau \rightarrow ll + 4\nu$ analysis excluded the Higgs at 5 - 7 times the SM cross section for 100 < m_H < 130 GeV and 7 - 15 times the SM cross section for 130 < m_H < 150 GeV. Combining all the $H \rightarrow \tau\tau$ search results, an exclusion of 3 - 5 times the SM cross section is achieved for 100 < m_H < 130 GeV and 5 - 12 times the SM cross section is achieved for 130 < m_H < 150 GeV. These results are obtained only from the data collected by ATLAS in 2011 at $\sqrt{s}$ = 7 TeV, which has an integrated luminosity of 4.7 fb^{-1}.

A new boson was discovered with a significance of 5.9 σ in the $H \rightarrow \gamma\gamma$, $H \rightarrow WW \rightarrow l\nu l\nu$ and $H \rightarrow ZZ \rightarrow 4l$ search channels using the 4.8 fb^{-1} of $\sqrt{s}$ = 7 TeV data collected in 2011 and the 5.9 $fb-1$ of $\sqrt{s}$ = 8 TeV data collected in 2012. This particle has a best fit mass of 126 $\pm$ 0.4 (stat.) $\pm$ 0.4 (syst.) and a best fit signal strength of $\mu = 1.4 \pm 0.3$. These properties are consistent, within the uncertainties, to that of the SM Higgs boson.

The newly discovered particle has so far only been observed to decay into gauge bosons. To confirm this discovery as the Higgs, its fermionic decay modes must also be observed, which can be achieved in the $H \to \tau\tau$ and $H \to b\bar{b}$ search channels with more data. The spin of this particle also needs to be measured to test whether or not it is a scalar. This measurement is most easily performed in the $H \to \tau\tau$ channel, where the branching ratio is high and the angular distribution of the final state particles are easy to measure. At the time at which this thesis was written, the LHC continues to produce data at $\sqrt{s}$ = 8 TeV, which can be used to make the necessary measurements to confirm this discovery. Therefore the future goals of the $H \to \tau\tau$ analyses are to firstly, observe the signal and secondly, to measure the spin of this new particle if it does decay into tauons.

A

Parton distribution function

Protons are composite particles that consists of quarks and gluons. They contain three valence quarks that are two up quarks and one down quark to give protons a +1 net electric charge ($2 \times u(+2/3) + d(-1/3) = +1$). In the most simplistic view, the three valence quarks are held together in a triplet by gluon interactions. However, high-energy experiments that have probed the proton structure reveal a more complicated picture. One of these more recent experiments has been conducted at the Hadron Electron Ring Anlage (HERA), located at DESY in Hamburg, Germany. HERA is a proton-electron collider with two detectors that measure the proton's parton distribution function (PDF): H1 and ZEUS. A detailed description of HERA and the H1 and ZEUS detectors can be found in references [96] and [97].

The most recent combined measurement of the proton PDF from H1 and ZEUS is presented in Figure A.1. This plot shows the probabilities of valence quarks, sea quarks and gluons carrying a fraction of the proton's total momentum at centre-of-mass energy of $Q^2 = 10000$ GeV^2. These results show that sea quarks and gluons dominate the proton PDF for low values of x and the valence quarks peak at around the $x = 10^{-1}$. The measurements of the proton's PDFs is important for calculating expected cross sections of physics processes at hadron colliders.

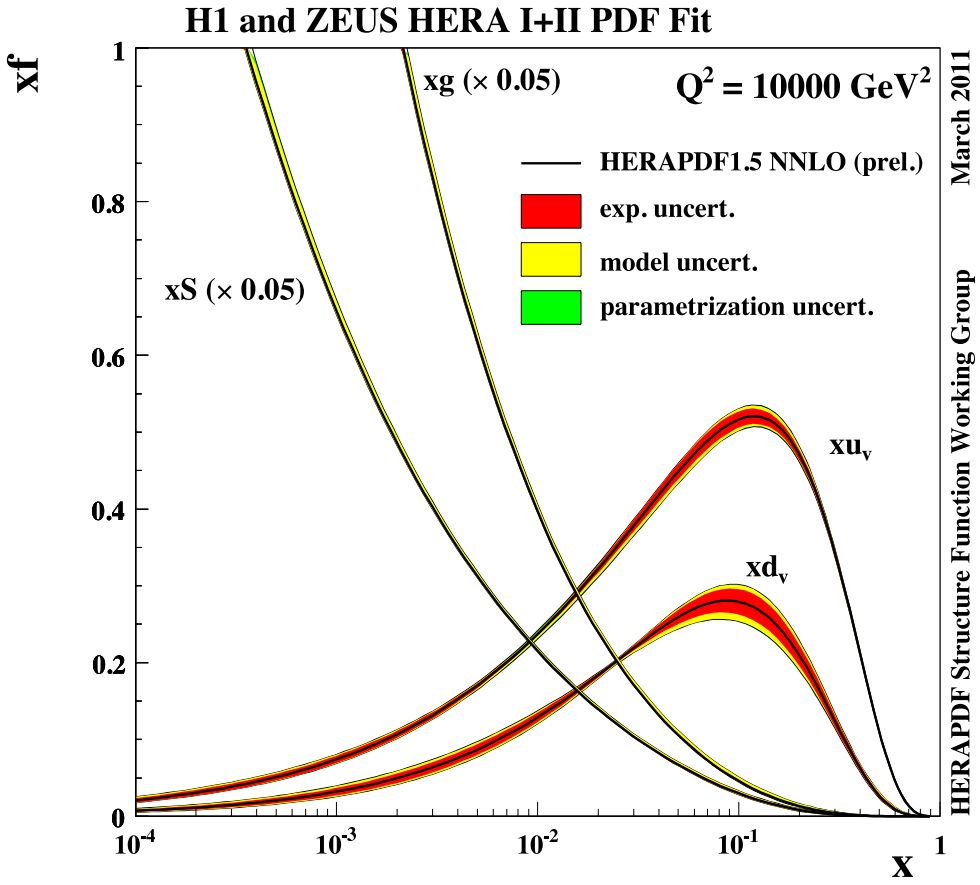
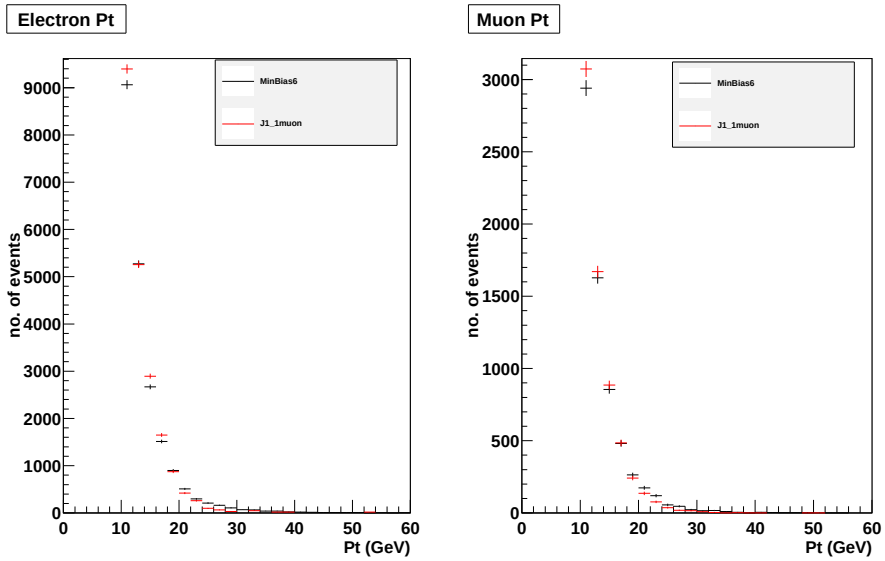


Figure A.1: Combined measurement of the proton PDF by the H1 and ZEUS experiments. f is the probability of finding a parton with a fraction of the proton's total momentum, x. xu_v and xd_v represent the up and down valence quarks respectively and xS and xg represent the sea quarks and the gluons respectively. Both the sea quark and gluon distributions have been down scaled by a factor of 20 for viewing purposes.

B

Multijet composition plots

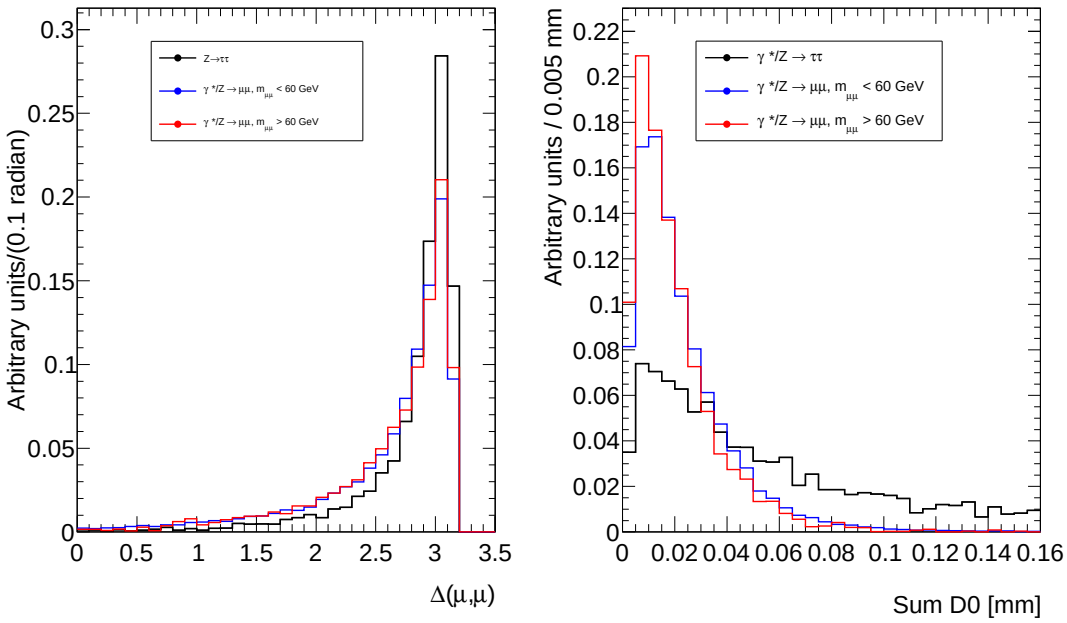


(a) Electron Pt (b) Muon Pt

Figure B.1: p_T distribution for the Minimum Bias and the Multijet(μ) 17-35 GeV samples [65] [65].

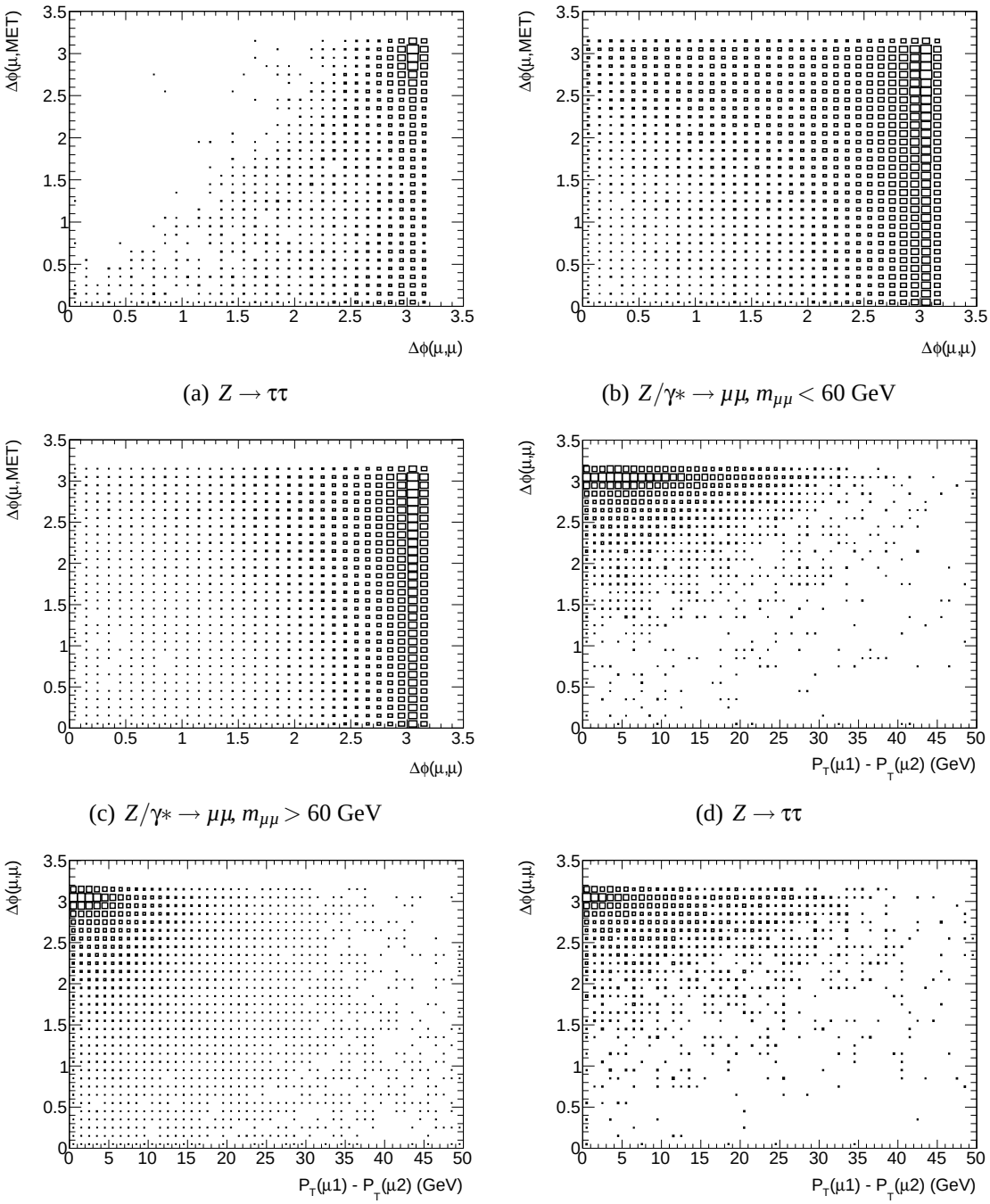
C

$Z \rightarrow \tau\tau \rightarrow ll + 4\nu$ plots and tables



(a) $\Delta\phi(\mu_1, \mu_2)$ (b) $|d_0(\mu_1)| + |d_0(\mu_2)|$

Figure C.1: $\Delta\phi(\mu_1, \mu_2)$ and $|d_0(\mu_1)| + |d_0(\mu_2)|$ distributions for the $Z \rightarrow \tau\tau \rightarrow \mu\mu + 4\nu$ and $Z/\gamma* \rightarrow \mu\mu$ events normalised to unity for viewing purposes. The distributions are shown for events that satisfy the $\mu\mu$, isolation and invariant mass selections [10].



(a) $Z \rightarrow \tau\tau$

(b) $Z/\gamma* \rightarrow \mu\mu$, $m_{\mu\mu} < 60$ GeV

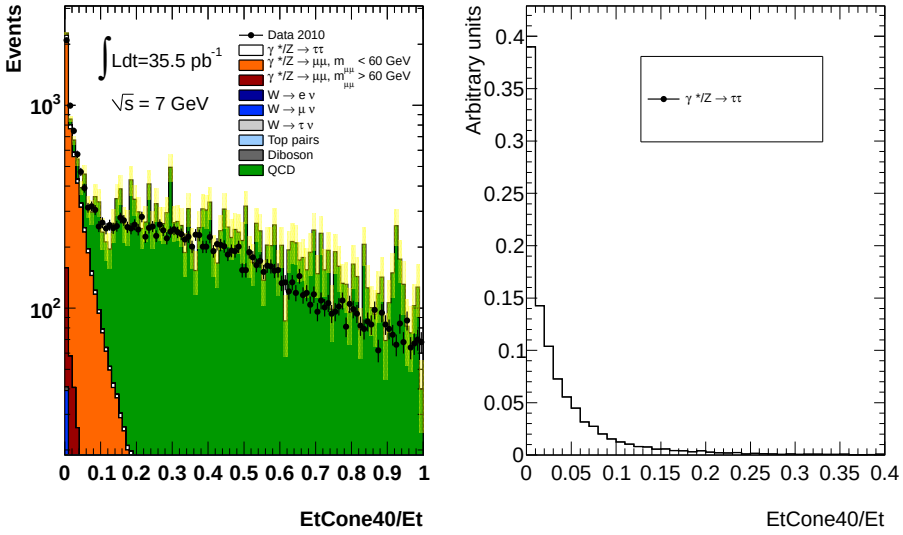
(c) $Z/\gamma* \rightarrow \mu\mu$, $m_{\mu\mu} > 60$ GeV

(d) $Z \rightarrow \tau\tau$

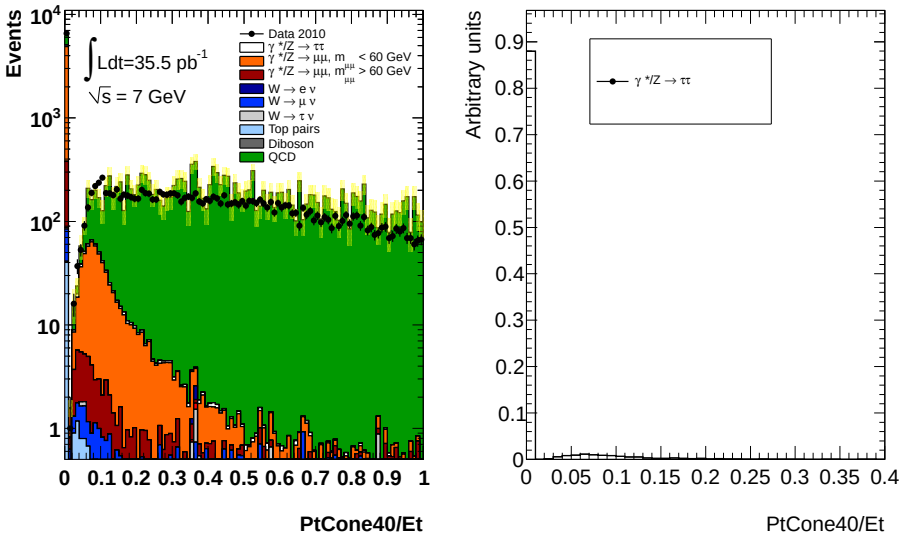
(e) $Z/\gamma* \rightarrow \mu\mu$, $m_{\mu\mu} < 60$ GeV

(f) $Z/\gamma* \rightarrow \mu\mu$, $m_{\mu\mu} > 60$ GeV

Figure C.2: 2D histograms of $\Delta\phi(\mu_1,\mu_2)$ versus $\Delta\phi(\mu, E_T^{miss})$ shown in (a), (b) and (c). 2D histograms of $\Delta\phi(\mu_1,\mu_2)$ versus $P_T(\mu_1) - P_T(\mu_2)$ shown in (d), (e) and (f). Distributions are shown after the $\mu\mu$, isolation and invariant mass selections. The size of the squares is proportional to the number of events in each bin [10].

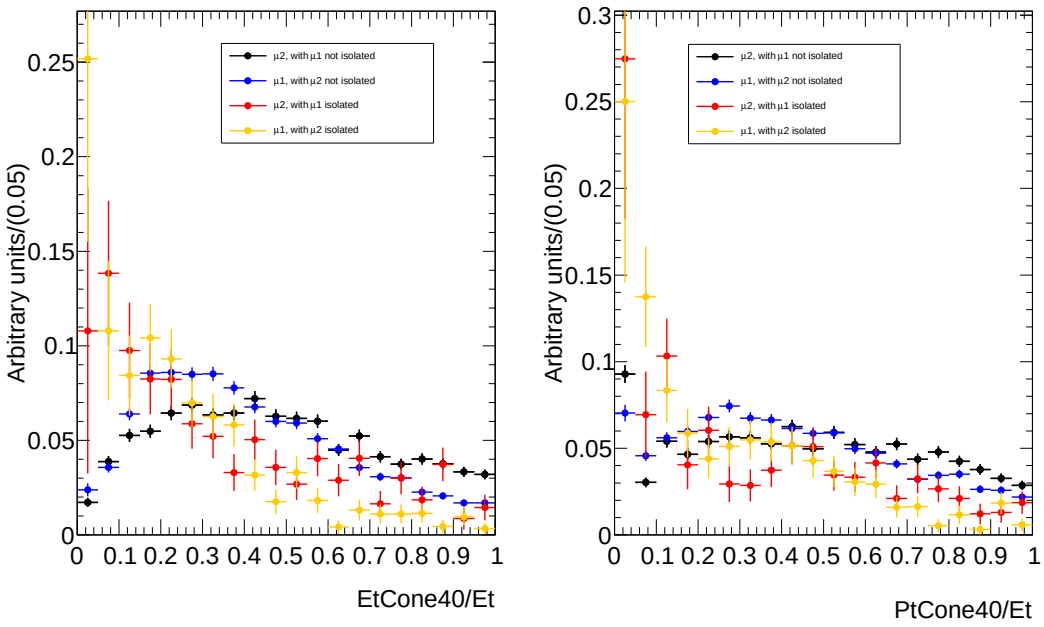


(a) Muon $E_T$cone40/p_T (b) Muon $E_T$cone40/p_T signal



(c) Muon $p_T$cone40/p_T (d) Muon $p_T$cone40/p_T signal

Figure C.3: Distributions of the isolation variables for events with exactly two muons of opposite charge. A data and MC comparison is shown in (a) and (c), where the yellow hatches indicate the uncertainty on the MC. The multijet contribution is estimated using Multijet(μ) samples, which is scaled to match the number of observed events in data. The shape of the signal distribution is shown explicitly in (b) and (d) for viewing purposes [10].

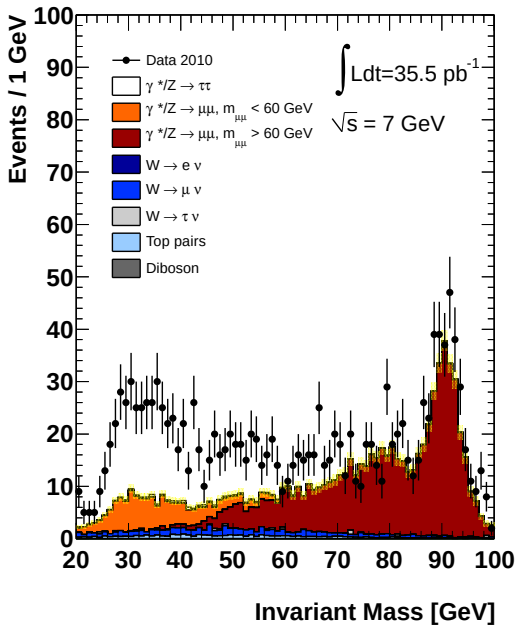


(a) (b)

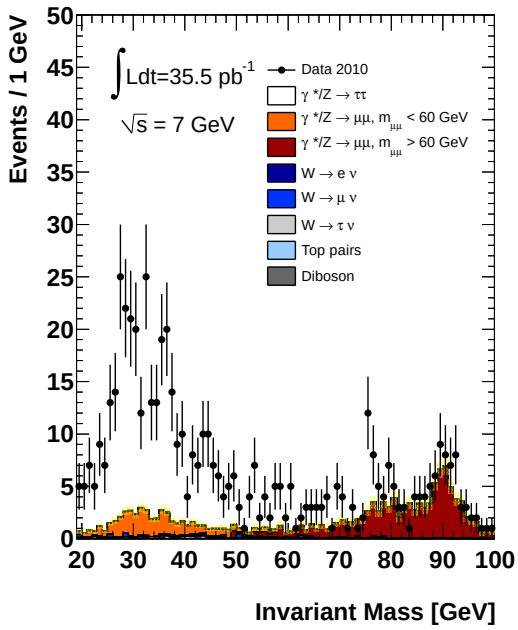
Figure C.4: Isolation distributions of $\mu\mu$ multijet events in data in the different ABCD regions after the visible mass selection. The distributions have been normalised to unity for viewing purposes [10].

Isolation Variable	μ_1 isolated	μ_1 not isolated	k factor
$\varepsilon(\mu_2)$ E_TCone40/P_T	0.378 ± 0.028	0.110 ± 0.004	3.44 ± 0.28
$\varepsilon(\mu_2)$ P_TCone40/P_T	0.369 ± 0.028	0.107 ± 0.004	3.45 ± 0.29
Isolation Variable	μ_2 isolated	μ_2 not isolated	k factor
$\varepsilon(\mu_1)$ E_TCone40/P_T	0.528 ± 0.040	0.180 ± 0.006	2.93 ± 0.24
$\varepsilon(\mu_1)$ P_TCone40/P_T	0.437 ± 0.035	0.136 ± 0.004	3.21 ± 0.27
Average k factor = 3.26 ± 0.55 (stat) ± 0.50 (sys)			

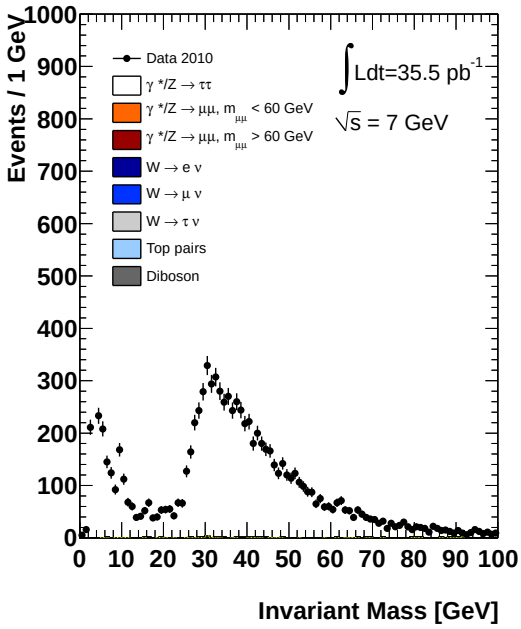
Table C.1: The isolation selection efficiencies of the sub-leading muon are shown for when the leading muon passes the isolation selections and when it does not (top). Conversely, the efficiencies of the leading muon are shown for when the sub-leading muon passes isolation and when it does not (bottom). The ratio of these efficiencies are given as k factors [10].



(a) Region B



(b) Region C

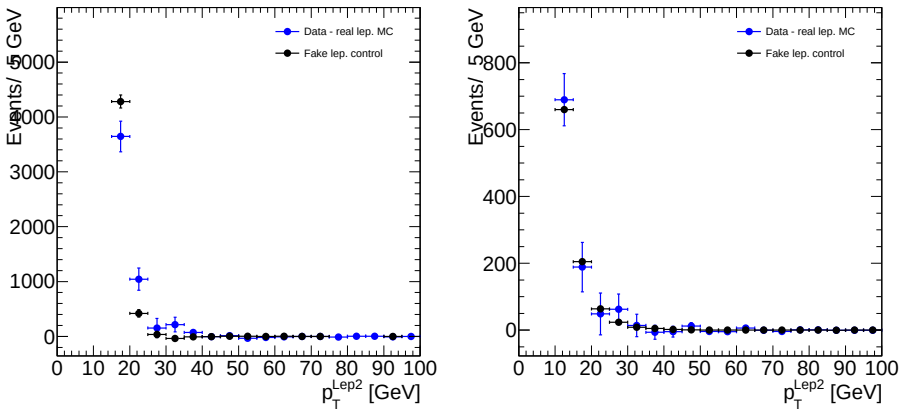


(c) Region D

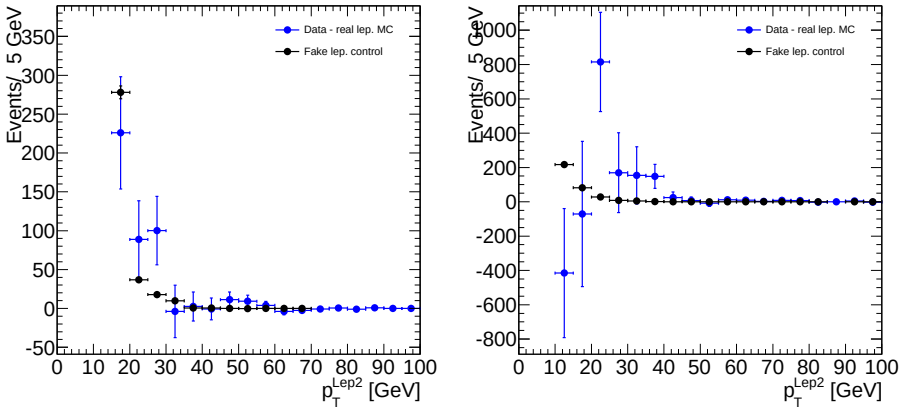
Figure C.5: The $\mu\mu$ invariant mass distributions of the multijet control regions B, C and D for events passing the dilepton selection [10].

D

$H \longrightarrow \tau\tau \longrightarrow ll + 4\nu$ plots and tables



(a) ee (b) $e\mu$



(c) μe (d) $\mu\mu$

Figure D.1: The comparisons between the nominal and fake control regions for the p_T of the sub-leading lepton after the dilepton and invariant mass selections. The distributions are normalised to the same integral [58].

Table D.1: Systematic uncertainties for the signal (top) and backgrounds (bottom) in the 2-jet VBF category [58].

2-jet VBF Category Relative Uncertainty (%)			
Uncertainty Source	$gg \to H$ (m_H=120 GeV)	VBF (m_H=120 GeV)	VH (m_H=120 GeV)
	Reconstruction		
Trigger scale factors (%)	+0.5/-1.7	+0.5/-1.7	+0.5/-1.7
Electron scale factors (%)	+1.8/-1.9	+1.8/-1.9	+1.8/-1.9
Electron energy scale (%)	+0.3/-0.3	+0.3/-0.3	+0.3/-0.3
Electron energy resolution (%)	+0.2/-0.5	+0.2/-0.5	+0.2/-0.5
Muon scale factors (%)	+2.2/-2.2	+2.2/-2.2	+2.2/-2.2
Muon momentum scale (%)	+0.3/-0.3	+0.3/-0.3	+0.3/-0.3
Muon momentum resolution (%)	+0.2/-0.5	+0.2/-0.5	+0.2/-0.5
Jet energy scale (%)	+3.2/-4.9	+0.8/-7.5	+0.3/-0.1
Jet energy resolution (%)	$\pm$2.7	$\pm$0.3	0.0
Pile-up (%)	0.0	$\pm$0.8	$\pm$18.0
E_T^{miss} reconstruction (%)	0.0	0.0	$\pm$0.1
b-tagging (%)	$\pm$0.2	$\pm$0.1	0.0
	Monte Carlo modelling		
PDF (gg) (%)	$\pm$8.0	-	-
PDF ($q\bar{q}$) (%)	-	$\pm$4.0	$\pm$4.0
	Process rate		
$gg \to H$ (%)	$\pm$25	-	-
VBF/VH H (%)	-	$\pm$1.0	$\pm$1.0
Luminosity (%)	$\pm$3.9	$\pm$3.9	$\pm$3.9
MC statistics (%)	$\pm$25.9	$\pm$4.7	$\pm$30.0

Uncertainty Source	$Z \to \tau\tau$	Fakes	Other Backgrounds
	Reconstruction		
Trigger scale factors (%)	+1.7/-3.9	-	+1.7/-3.9
Electron scale factors (%)	+0.5/-1.5	-	+0.5/-1.5
Electron energy scale (%)	+0.8/-0.8	-	+0.8/-0.8
Electron energy resolution (%)	+0.3/+2.4	-	+0.3/+2.4
Muon scale factors (%)	+0.5/-1.8	-	+0.5/-1.8
Muon momentum scale (%)	+0.8/-0.8	-	+0.8/-0.8
Muon momentum resolution (%)	+0.3/-2.8	-	+0.3/-2.8
Jet energy scale (%)	-	-	+20.3/-16.7
Jet energy resolution (%)	-	-	$\pm$3.3
Pile-up (%)	-	-	$\pm$0.9
E_T^{miss} reconstruction (%)	-	-	$\pm$0.8
b-tagging (%)	-	-	$\pm$1.8
	Monte Carlo modelling		
PDF (gg) (%)	-	-	$\pm$4.4
PDF ($q\bar{q}$) (%)	$\pm$4	-	$\pm$1.8
	Process rate		
Fake background normalisation (%)	-	$\pm$40	-
$Z \to \tau\tau$ embedding (%)	$\pm$1	-	
Cross-section Z+jets (%)	$\pm$2.5	-	$\pm$0.5
Cross-section di-boson (%)	-	-	$\pm$0.5
Cross-section $t\bar{t}$ (%)	-	-	$\pm$3.3
Luminosity (%)	$\pm$3.9	-	$\pm$3.9
MC statistics (%)	$\pm$11.1	$\pm$35.1	$\pm$2.3

Table D.2: Systematic uncertainties for the signal (top) and backgrounds (bottom) in the 2-jet VH category [58].

2-jet VH Category Relative Uncertainty (%)			
Uncertainty Source	$gg \to H$ (m_H=120 GeV)	VBF (m_H=120 GeV)	VH (m_H=120 GeV)
	Reconstruction		
Trigger scale factors (%)	+0.5/-1.7	+0.5/-1.7	+0.5/-1.7
Electron scale factors (%)	+1.8/-1.9	+1.8/-1.9	+1.8/-1.9
Electron energy scale (%)	+0.3/-0.3	+0.3/-0.3	+0.3/-0.3
Electron energy resolution (%)	+0.2/-0.5	+0.2/-0.5	+0.2/-0.5
Muon scale factors (%)	+2.2/-2.2	+2.2/-2.2	+2.2/-2.2
Muon momentum scale (%)	+0.3/-0.3	+0.3/-0.3	+0.3/-0.3
Muon momentum resolution (%)	+0.2/-0.5	+0.2/-0.5	+0.2/-0.5
Jet energy scale (%)	+2.2/-17.0	+2.1/-9.8	+3.7/-8.9
Jet energy resolution (%)	$\pm$0.1	$\pm$0.2	$\pm$1.3
Pile-up (%)	$\pm$1.2	0.0	$\pm$0.2
E_T^{miss} reconstruction (%)	$\pm$1.1	0.0	$\pm$0.3
b-tagging (%)	$\pm$0.2	$\pm$0.1	$\pm$0.2
	Monte Carlo modelling		
PDF (gg) (%)	$\pm$8.0	-	-
PDF ($q\bar{q}$) (%)	-	$\pm$4.0	$\pm$4.0
	Process rate		
$gg \to H$ (%)	$\pm$25	-	-
VBF/VH H (%)	-	$\pm$1.0	$\pm$1.0
Luminosity (%)	$\pm$3.9	$\pm$3.9	$\pm$3.9
MC statistics (%)	$\pm$14.3	$\pm$12.5	$\pm$6.9

Uncertainty Source	$Z \to \tau\tau$	Fakes	Other Backgrounds
	Reconstruction		
Trigger scale factors (%)	+1.7/-3.9	-	+1.7/-3.9
Electron scale factors (%)	+0.5/-1.5	-	+0.5/-1.5
Electron energy scale (%)	+0.8/-0.8	-	+0.8/-0.8
Electron energy resolution (%)	+0.3/-2.4	-	+0.3/-2.4
Muon scale factors (%)	+0.5/-1.8	-	+0.5/-1.8
Muon momentum scale (%)	+0.8/-0.8	-	+0.8/-0.8
Muon momentum resolution (%)	+0.3/-2.8	-	+0.3/-2.8
Jet energy scale (%)	-	-	+18.5/-23.6
Jet energy resolution (%)	-	-	$\pm$1.2
Pile-up (%)	-	-	$\pm$0.3
E_T^{miss} reconstruction (%)	-	-	$\pm$0.3
b-tagging (%)	-	-	$\pm$1.5
	Monte Carlo modelling		
PDF (gg) (%)	-	-	$\pm$4.6
PDF ($q\bar{q}$) (%)	$\pm$4	-	$\pm$1.7
	Process rate		
Fake leptons normalisation (%)	-	$\pm$40	-
$Z \to \tau\tau$ embedding (%)	$\pm$1	-	
Cross-section Z+jets (%)	$\pm$2.5	-	$\pm$0.8
Cross-section di-boson (%)	-	-	$\pm$0.5
Cross-section $t\bar{t}$ (%)	-	-	$\pm$3.4
Luminosity (%)	$\pm$3.9	-	$\pm$3.9
MC statistics (%)	$\pm$5.0	$\pm$17.3	$\pm$6.5

Table D.3: Systematic uncertainties for signal (top) and backgrounds (bottom) in the 1-jet category [58].

1-jet Category Relative Uncertainty (%)			
Uncertainty Source	$gg \to H$ (m_H=120 GeV)	VBF H (m_H=120 GeV)	VH (m_H=120 GeV)
	Reconstruction		
Trigger scale factors (%)	+0.5/-1.7	+0.5/-1.7	+0.5/-1.7
Electron scale factors (%)	+1.8/-1.9	+1.8/-1.9	+1.8/-1.9
Electron energy scale (%)	+0.3/-0.3	+0.3/-0.3	+0.3/-0.3
Electron energy resolution (%)	+0.2/-0.5	+0.2/-0.5	+0.2/-0.5
Muon scale factors (%)	±2.2	±2.2	±2.2
Muon momentum scale (%)	+0.3/-0.3	+0.3/-0.3	+0.3/-0.3
Muon momentum resolution (%)	+0.2/-0.5	+0.2/-0.5	+0.2/-0.5
Jet energy scale (%)	+0/-1.7	+0.9/-0.6	+0.3/-4.7
Jet energy resolution (%)	±0.7	±1.9	±2.0
Pile-up (%)	±0.7	0.0	±2.0
E_T^{miss} reconstruction (%)	±0.3	±0.1	±0.2
b-tagging (%)	±0.1	±0.1	±0.1
	Monte Carlo modelling		
PDF (gg) (%)	±8.0	-	-
PDF ($q\bar{q}$) (%)	-	±4.0	±4.0
	Process rate		
$gg \to H$ (%)	±20	-	-
VBF/VH H (%)	-	±1.0	±1.0
Luminosity (%)	±3.9	±3.9	±3.9
MC statistics (%)	±6.5	±4.4	±5.0

Uncertainty Source	$Z \to \tau\tau$	Fakes	Other Backgrounds
	Reconstruction		
Trigger scale factors (%)	+1.7/-3.9	-	+1.7/-3.9
Electron scale factors (%)	+0.5/-1.5	-	+0.5/-1.5
Electron energy scale (%)	+0.8/-0.8	-	+0.8/-0.8
Electron energy resolution (%)	+0.3/+2.4	-	+0.3/+2.4
Muon scale factors (%)	+0.5/-1.8	-	+0.5/-1.8
Muon momentum scale (%)	+0.8/-0.8	-	+0.8/-0.8
Muon momentum resolution (%)	+0.3/-2.8	-	+0.3/-2.8
Jet energy scale (%)	-	-	+13.8/-11.8
Jet energy resolution (%)	-	-	±0.2
Pile-up (%)	-	-	±0.4
E_T^{miss} reconstruction (%)	-	-	±0.1
b-tagging (%)	-	-	±2.8
	Monte Carlo modelling		
PDF (gg) (%)	-	-	±3.8
PDF ($q\bar{q}$) (%)	±4	-	±2.1
	Process rate		
Fake leptons normalisation (%)	-	±40	-
$Z \to \tau\tau$ embedding (%)	±1	-	
Cross-section Z+jets (%)	±2.3	-	±0.9
Cross-section di-boson (%)	-	-	±0.5
Cross-section $t\bar{t}$ (%)	-	-	±2.8
Luminosity (%)	±3.9	-	±3.9
MC statistics (%)	±2.3	±12.7	±4.0

Table D.4: Systematic uncertainties for the signal (top) and backgrounds (bottom) in the 0-jet category [58].

0-jet Category Relative Uncertainty (%)			
Uncertainty Source	$gg \to H$ (m_H=120 GeV)	VBF H (m_H=120 GeV)	VH (m_H=120 GeV)
	Reconstruction		
Trigger scale factors (%)	+0.5/-1.7	+0.5/-1.7	+0.5/-1.7
Electron scale factors (%)	+2.0/-2.1	+2.0/-2.1	+2.0/-2.1
Electron energy scale (%)	+0.3/-0.3	+0.3/-0.3	+0.3/-0.3
Electron energy resolution (%)	+0.2/-0.5	+0.2/-0.5	+0.2/-0.5
Muon scale factors (%)	+2.0/-2.1	+2.0/-2.1	+2.0/-2.1
Muon momentum scale (%)	+0.3/-0.3	+0.3/-0.3	+0.3/-0.3
Muon momentum resolution (%)	+0.2/-0.5	+0.2/-0.5	+0.2/-0.5
Jet energy scale (%)	+0.2/-0.2	+1.9/-1.7	+2.4/-2.3
Jet energy resolution (%)	0.0	$\pm$0.7	$\pm$0.6
Pile-up (%)	$\pm$0.1	0.3	$\pm$0.3
E_T^{miss} reconstruction (%)	$\pm$0.2	$\pm$0.3	$\pm$0.4
b-tagging (%)	-	-	-
	Monte Carlo modelling		
PDF (gg) (%)	$\pm$8.0	-	-
PDF ($q\bar{q}$) (%)	-	$\pm$4.0	$\pm$4.0
	Process rate		
$gg \to H$ (%)	+12/-7.0	-	-
VBF/VH H (%)	-	$\pm$1.0	$\pm$1.0
Luminosity (%)	$\pm$3.9	$\pm$3.9	$\pm$3.9
MC statistics (%)	$\pm$4.0	$\pm$5.6	$\pm$6.0

Uncertainty Source	$Z \to \tau\tau$	Fakes	Other Backgrounds
	Reconstruction		
Trigger scale factors (%)	+1.7/-3.9	-	+1.7/-3.9
Electron scale factors (%)	+0.5/-1.6	-	+0.5/-1.6
Electron energy scale (%)	+0.8/-0.8	-	+0.8/-0.8
Electron energy resolution (%)	+0.3/+2.6	-	+0.3/+2.6
Muon scale factors (%)	+0.5/-1.6	-	+0.5/-1.6
Muon momentum scale (%)	+0.8/-0.8	-	+0.8/-0.8
Muon momentum resolution (%)	+0.3/-2.6	-	+0.3/-2.6
Jet energy scale (%)	-	-	+2.7/-2.6
Jet energy resolution (%)	-	-	$\pm$0.1
Pile-up (%)	-	-	$\pm$0.1
E_T^{miss} reconstruction (%)	-	-	$\pm$0.1
b-tagging (%)	-	-	-
	Monte Carlo modelling		
PDF (gg) (%)	-	-	$\pm$2.5
PDF ($q\bar{q}$) (%)	$\pm$4	-	$\pm$2.7
	Process rate		
Fake leptons normalisation (%)	-	$\pm$40	-
$Z \to \tau\tau$ embedding (%)	$\pm$1	-	
Cross-section Z+jets (%)	$\pm$2.2	-	$\pm$2.2
Cross-section di-boson (%)	-	-	$\pm$2.1
Cross-section $t\bar{t}$ (%)	-	-	+0.7/-1.3
Luminosity (%)	$\pm$3.9	-	$\pm$3.9
MC statistics (%)	$\pm$0.5	$\pm$1.1	$\pm$2.3

Bibliography

[1] Particle Data Group, K. Nakamura *et al.*, "Review of particle physics," *J. Phys.* **G37** (2010) 075021.

[2] Ta-Pei Cheng and Ling-Fong Li, *Gauge theory of elementary particle physics*. Oxford University Press, 1984.

[3] G. L. Kane, *Modern Elementary Particle Physics*. Westview Press, 1993.

[4] S. A. Thomas, F. B. Abdalla, and O. Lahav, "Upper Bound of 0.28 eV on Neutrino Masses from the Largest Photometric Redshift Survey," *Phys. Rev. Lett.* **105** (Jul, 2010) 031301. http://link.aps.org/doi/10.1103/PhysRevLett.105.031301.

[5] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, and M. Spira, "Standard Model Higgs-Boson Branching Ratios with Uncertainties," *Eur.Phys.J.* **C71** (2011) 1753, arXiv:1107.5909 [hep-ph].

[6] LEP, ALEPH, DELPHI, L3 and OPAL Collaborations , R. Barate *et al.*, "Search for the standard model Higgs boson at LEP," arXiv:hep-ex/0306033 [hep-ex].

[7] CDF and D. Collaborations, "Standard Model Higgs Boson Combination at the Tevatron," arXiv:1201.5107 [hep-ex].

[8] Baak, M. and Goebel, M. and Haller, J. and Hoecker, A. and Ludwig, D. and others, "Updated Status of the Global Electroweak Fit and Constraints on New Physics," arXiv:1107.0975 [hep-ph].

[9] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.), "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables," *CERN-2011-002* (CERN, Geneva, 2011) , arXiv:1101.0593 [hep-ph].

[10] E. Barberio *et al.*, "Measurement of the cross section $\sigma \times BR(Z \to \tau\tau)$ in the dilepton channel with the ATLAS detector: Supporting Note," Tech. Rep. ATL-COM-PHYS-2011-418, CERN, Geneva, Apr, 2011.

[11] P. Jenni, M. Nessi, M. Nordberg, and K. Smith, *ATLAS high-level trigger, data-acquisition and controls: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 2003.

[12] M. Benedikt *et al.*, "LHC Design Report," *CERN-2004-003* (2004) .

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

[14] ATLAS Collaboration , G. Aad *et al.*, "The ATLAS Experiment at the CERN Large Hadron Collider," *JINST* **3** (2008) S08003.

[15] CMS Collaboration , S. Chatrchyan *et al.*, "The CMS experiment at the CERN Large Hadron Collider," *JINST* **3** (2008) S08004.

[16] ALICE Collaboration , K. Aamodt *et al.*, "The ALICE experiment at the CERN Large Hadron Collider," *JINST* **3** (2008) S08002.

[17] LHCb Collaboration , A. Augusto Alves Jr *et al.*, "The LHCb Detector at the LHC," *JINST* **3** (2008) S08005.

[18] CERN AC Term, "The four main LHC experiments," *CERN-AC-9906026* (Jun, 1999) .

[19] C. Lefèvre, "The CERN accelerator complex. Complexe des accélérateurs du CERN," *CERN-DI-0812015* (Dec, 2008) .

[20] ATLAS Collaboration , A. Airapetian *et al.*, "ATLAS Detector and Physics Performance Technical design report," *CERN-LHCC-99-14* (1999) .

[21] ATLAS Collaboration , G. Aad *et al.*, "Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics," `arXiv:0901.0512 [hep-ex]`.

[22] T. Cornelissen *et al.*, "Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)," *ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002* (2007) .

[23] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, S. Menke, S. Rajagopalan, D. Rousseau, S. Snyder, and G. Unal, "Calorimeter Clustering Algorithms: Description and Performance," Tech. Rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva, Apr, 2008.

[24] V. Gallo, "Identification and reconstruction of electrons and photons with the ATLAS detector at the LHC.," Tech. Rep. ATL-PHYS-PROC-2012-044, CERN, Geneva, Feb, 2012.

[25] B. Resende, "Muon identification algorithms in ATLAS," Tech. Rep. ATL-PHYS-PROC-2009-113, CERN, Geneva, Sep, 2009.

[26] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions," *Phys.Rev.* **D48** (1993) 3160–3166, arXiv:hep-ph/9305266 [hep-ph].

[27] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-k(t) jet clustering algorithm," *JHEP* **0804** (2008) 063, arXiv:0802.1189 [hep-ph].

[28] ATLAS Collaboration , G. Aad *et al.*, "Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data," Tech. Rep. ATLAS-CONF-2011-102, CERN, Geneva, Jul, 2011.

[29] ATLAS Collaboration , G. Aad *et al.*, "Performance of Impact Parameter-Based b-tagging Algorithms with the ATLAS Detector using Proton-Proton Collisions at $\sqrt{s}$ = 7 TeV," Tech. Rep. ATLAS-CONF-2010-091, CERN, Geneva, Oct, 2010.

[30] G. Piacquadio and C. Weiser, "A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS," *J.Phys.Conf.Ser.* **119** (2008) 032032.

[31] A. Hocker *et al.*, "TMVA: Toolkit for multivariate data analysis," *PoS* **ACAT** (2007) 040, arXiv:physics/0703039.

[32] ATLAS Collaboration , G. Aad *et al.*, "Performance of Missing Transverse Momentum Reconstruction in Proton-Proton Collisions at $\sqrt{s}$ = 7 TeV with ATLAS," *Eur. Phys. J. C* **72** no. arXiv:1108.5602. CERN-PH-EP-2011-114, (Sep, 2011) 1844. 33 p.

[33] ATLAS Collaboration , P. Bright-Thomas *et al.*, "ATLAS First-Level Trigger Technical Design Report," *CERN-LHCC-98-14* (1998) .

[34] ATLAS Collaboration , B. Caron *et al.*, "ATLAS High-Level Trigger, Data Acquisition and Controls Technical Design Report," *CERN-LHCC-2003-022* (2003) .

[35] W.-K. Tung, H. L. Lai, J. Pumplin, P. M. Nadolsky, and C. P. Yuan, "Global QCD Analysis and Collider Phenomenology - CTEQ," arXiv:0707.0275 [hep-ph].

[36] A. Sherstnev and R. Thorne, "Different PDF approximations useful for LO Monte Carlo generators," arXiv:0807.2132 [hep-ph].

[37] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual," *JHEP* **0605** (2006) 026, arXiv:hep-ph/0603175 [hep-ph].

[38] G. Corcella *et al.*, "HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes)," *JHEP* **01** (2001) 010, arXiv:hep-ph/0011363.

[39] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, "Multiparton interactions in photoproduction at HERA," *Z. Phys.* **C72** (1996) 637–646, arXiv:hep-ph/9601371.

[40] M. L. Mangano *et al.*, "ALPGEN, a generator for hard multiparton processes in hadronic collisions," *JHEP* **07** (2003) 001, arXiv:hep-ph/0206293.

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

[42] S. Alioli, P. Nason, C. Oleari, and E. Re, "NLO Higgs boson production via gluon fusion matched with shower in POWHEG," *JHEP* **04** (2009) 002, arXiv:0812.0578 [hep-ph].

[43] B. P. Kersevan and E. Richter-Was, "The Monte Carlo event generator AcerMC version 2.0 with interfaces to PYTHIA 6.2 and HERWIG 6.5," arXiv:hep-ph/0405247.

[44] S. Frixione and B. R. Webber, "The MC@NLO 3.3 Event Generator," arXiv:hep-ph/0612272 [hep-ph].

[45] S. Jadach, Z. Was, R. Decker, and J. Kühn, "The τ decay library TAUOLA, version 2.4," *Computer Physics Communications* **76** no. 3, (1993) 361 – 380. http://www.sciencedirect.com/science/article/pii/001046559390061G.

[46] P. Golonka and Z. Was, "PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays," *Eur.Phys.J.* **C45** (2006) 97–107, arXiv:hep-ph/0506026 [hep-ph].

[47] ATLAS Collaboration , G. Aad *et al.*, "The ATLAS Simulation Infrastructure," *Eur. Phys. J. C* **70** no. arXiv:1005.4568. CERN-PH-EP-2010-044, (May, 2010) 823–874. 53 p.

[48] S. Agostinelli *et al.*, "GEANT4: A Simulation toolkit," *Nucl. Instrum. Meth.* **A506** (2003) 250–303.

[49] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, "High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO," *Phys.Rev.* **D69** (2004) 094008, arXiv:hep-ph/0312266 [hep-ph].

[50] S. Catani, L. Cieri, G. Ferrera, D. de Florian, and M. Grazzini, "Vector boson production at hadron colliders: A Fully exclusive QCD calculation at NNLO," *Phys.Rev.Lett.* **103** (2009) 082001, arXiv:0903.2120 [hep-ph].

[51] S. Moch and P. Uwer, "Theoretical status and prospects for top-quark pair production at hadron colliders," *Phys. Rev.* **D78** (2008) 034003, arXiv:0804.1476 [hep-ph].

[52] U. Langenfeld, S. Moch, and P. Uwer, "New results for $t\bar{t}$ production at hadron colliders," arXiv:0907.2527 [hep-ph].

[53] M. Baak *et al.*, “Data Quality Status Flags and Good Run Lists for Physics Analysis in ATLAS,” Tech. Rep. ATL-COM-GEN-2009-015, CERN, Geneva, Mar, 2009.

[54] C. Boddy *et al.*, “Measurement of Z->tautau production cross-section in proton-proton collisions at sqrts=7 TeV with the ATLAS detector - Support Note for lep-had channels,” Tech. Rep. ATL-COM-PHYS-2011-416, CERN, Geneva, Apr, 2010.

[55] ATLAS Collaboration , G. Aad *et al.*, “Updated Luminosity Determination in pp Collisions at root(s)=7 TeV using the ATLAS Detector,” Tech. Rep. ATLAS-CONF-2011-011, CERN, Geneva, Mar, 2011.

[56] C. Anders *et al.*, “Search for SM H $\rightarrow \tau^+\tau^- \rightarrow l\tau_h$ with the ATLAS Detector in 7 TeV Proton-Proton Collisions,” Tech. Rep. ATL-COM-PHYS-2012-037, CERN, Geneva, Jan, 2012.

[57] T. A. J. performance group, “Recommandations from jet-MET group for jet cleaning in 2010 data,”. `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/HowToCleanJets`.

[58] E. Barberio *et al.*, “Search for the Standard Model Higgs Boson in the Decay Mode H $\rightarrow \tau^+\tau^- \rightarrow$ ll + 4 Neutrinos in Proton-Proton Collisions at sqrt(s)=7 TeV with the ATLAS Detector,” Tech. Rep. ATL-COM-PHYS-2011-1611, CERN, Geneva, Nov, 2011.

[59] T. A. J. performance group, “Recommandations from jet-MET group for jet cleaning in 2011 data,”. `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/HowToCleanJets2011`.

[60] T. A. egamma performance group, “Egamma recommendation on energy scale and energy resolution for release 16 data,”. `https://twiki.cern.ch/twiki/bin/view/AtlasProtected/EnergyScaleResolutionRecommendations#2010dataRelease16`.

[61] F. Cerutti *et al.*, “Muon Momentum Resolution in First Pass Reconstruction of pp Collision Data Recorded by ATLAS in 2010,” Tech. Rep. ATLAS-COM-CONF-2011-003, CERN, Geneva, Jan, 2011.

[62] ATLAS Collaboration , G. Aad *et al.*, “Muon Reconstruction Performance,” Tech. Rep. ATLAS-CONF-2010-064, CERN, Geneva, Jul, 2010.

[63] ATLAS Collaboration , G. Aad *et al.*, “Determination of the muon reconstruction efficiency in ATLAS at the Z resonance in proton-proton collisons at sqrt(s)=7 TeV,” Tech. Rep. ATLAS-CONF-2011-008, CERN, Geneva, Feb, 2011.

[64] O. Arnaez and D. Froidevaux, “Recommendations for electron efficiencies,”. https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/EfficiencyMeasurements.

[65] E. Barberio *et al.*, “First observation of the process Z ->tautau->emu + 4nu with the ATLAS detector: Supporting Note,” Tech. Rep. ATL-COM-PHYS-2011-078, CERN, Geneva, Jan, 2011.

[66] ATLAS Collaboration , G. Aad *et al.*, “Observation of prompt inclusive electrons in the ATLAS experiment at sqrt(s) = 7 TeV.,” Tech. Rep. ATL-COM-PHYS-2010-280, CERN, Geneva, May, 2010.

[67] ATLAS Collaboration , G. Aad *et al.*, “Muon Performance in Minimum Bias pp Collision Data at $\sqrt{s}$ = 7 TeV with ATLAS,”.

[68] ATLAS Collaboration , G. Aad *et al.*, “Jet energy scale and its systematic uncertainty for jets produced in proton-proton collisions at sqrt(s) = 7 TeV and measured with the ATLAS detector,” Tech. Rep. ATLAS-CONF-2010-056, CERN, Geneva, Jul, 2010.

[69] ATLAS , G. Aad *et al.*, “Measurement of the Z to tau tau Cross Section with the ATLAS Detector,” *Phys.Rev.* **D84** (2011) 112006, arXiv:1108.2016 [hep-ex].

[70] ATLAS Collaboration , G. Aad *et al.*, “A measurement of the total $W^{\pm}$ and Z/γ^{*} cross sections in the e and μ decay channels and of their ratios in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector,” Tech. Rep. ATLAS-CONF-2011-041, CERN, Geneva, Mar, 2011.

[71] ATLAS Collaboration , G. Aad *et al.*, “A combined measurement of the top quark pair production cross-section using dilepton and single-lepton final states,” Tech. Rep. ATLAS-CONF-2011-040, CERN, Geneva, Mar, 2011.

[72] P. M. Nadolsky *et al.*, “Implications of CTEQ global analysis for collider observables,” *Phys.Rev.* **D78** (2008) 013004, arXiv:0802.0007 [hep-ph].

[73] ATLAS Collaboration , G. Aad *et al.*, “Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons,” Tech. Rep. ATLAS-CONF-2011-077, CERN, Geneva, May, 2011.

[74] D. G. Louis Lyons and P. Clifford, “How to combine correlated estimates of a single physical quantity,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **270** no. 1, (1988) 110 – 117. http://www.sciencedirect.com/science/article/pii/0168900288900186.

[75] A. Valassi, "Combining correlated measurements of several different physical quantities," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **500** no. A, (2003) 391 – 405. http://www.sciencedirect.com/science/article/pii/S0168900203003292.

[76] CMS , S. Chatrchyan *et al.*, "Measurement of the Inclusive Z Cross Section via Decays to Tau Pairs in pp Collisions at $\sqrt{s} = 7$ TeV," *JHEP* **1108** (2011) 117, arXiv:1104.1617 [hep-ex].

[77] R.K. Ellis, I. Hinchliffe, M. Soldate and J.J. Van der Bij,, "Higgs Decay to $\tau^+ \tau^-$: A Possible Signature of Intermediate Mass Higgs Bosons at the SSC," *Nucl. Phys.* **B297** (1988) 221.

[78] K. Bierwagen, U. Blumenschein, J. F. Grivaz, T. Kanno, S. Meeham, M. Sandhoff, J. Sauvan, and A. Quadt, "Measurement of the cross section for jets produced in association with Z bosons," Tech. Rep. ATL-COM-PHYS-2012-735, CERN, Geneva, Jun, 2012.

[79] ATLAS Collaboration , G. Aad *et al.*, "Data-driven estimation of the background to charged Higgs boson searches using hadronically-decaying tau final states in ATLAS,".

[80] T. Barillari, "Jet energy scale uncertainties in ATLAS," Tech. Rep. ATL-COM-PHYS-2012-676, CERN, Geneva, May, 2012.

[81] ATLAS , G. Aad *et al.*, "Jet energy measurement with the ATLAS detector in proton-proton collisions at sqrt(s) = 7 TeV," arXiv:1112.6426 [hep-ex].

[82] ATLAS Collaboration , G. Aad *et al.*, "Jet energy resolution and selection efficiency relative to track jets from in-situ techniques with the ATLAS Detector Using Proton-Proton Collisions at a Center of Mass Energy $\sqrt{s}$ = 7 TeV," Tech. Rep. ATLAS-CONF-2010-054, CERN, Geneva, Jul, 2010.

[83] ATLAS Collaboration , G. Aad *et al.*, "Measurement of the top quark pair production cross-section based on a statistical combination of measurements of dilepton and single-lepton final states at $\sqrt{s}$ = 7 TeV with the ATLAS detector," Tech. Rep. ATLAS-CONF-2011-108, CERN, Geneva, Aug, 2011.

[84] ATLAS Collaboration , G. Aad *et al.*, "Combined Standard Model Higgs boson searches with up to 2.3 fb^{-1} of pp collisions at $\sqrt{s}$ = 7 TeV at the LHC," Tech. Rep. ATLAS-CONF-2011-157, CERN, Geneva, Nov, 2011.

[85] A. L. Read, "Presentation of search results: the CLs technique," *J. Phys.* **G28** (2002) 2693.

[86] ATLAS , G. Aad *et al.*, "Search for the Standard Model Higgs boson in the H to tau+ tau- decay mode in sqrt(s) = 7 TeV pp collisions with ATLAS," arXiv:1206.5971 [hep-ex].

[87] A. Elagin, P. Murat, A. Pranko, and A. Safonov, "A New Mass Reconstruction Technique for Resonances Decaying to di-tau," *Nucl.Instrum.Meth.* **A654** (2011) 481–489, arXiv:1012.4686 [hep-ex].

[88] ATLAS Collaboration , E. Abat, "Response and Shower Topology of 2 to 180 GeV Pions Measured with the ATLAS Barrel Calorimeter at the CERN Test-beam and Comparison to Monte Carlo Simulations," Tech. Rep. ATL-CAL-PUB-2010-001, CERN, Geneva, May, 2010.

[89] ATLAS , G. Aad *et al.*, "Search for the Standard Model Higgs boson produced in association with a vector boson and decaying to a b-quark pair with the ATLAS detector," arXiv:1207.0210 [hep-ex].

[90] ATLAS Collaboration , G. Aad *et al.*, "Observation of an excess of events in the search for the Standard Model Higgs boson in the gamma-gamma channel with the ATLAS detector," Tech. Rep. ATLAS-CONF-2012-091, CERN, Geneva, Jul, 2012.

[91] ATLAS Collaboration , G. Aad *et al.*, "Observation of an excess of events in the search for the Standard Model Higgs boson in the $H \to ZZ^{(*)} \to 4\ell$ channel with the ATLAS detector.," Tech. Rep. ATLAS-CONF-2012-092, CERN, Geneva, Jul, 2012.

[92] ATLAS Collaboration , G. Aad *et al.*, "Observation of an Excess of Events in the Search for the Standard Model Higgs Boson in the $H \to WW^{(*)} \to \ell\nu\ell\nu$ Channel with the ATLAS Detector," Tech. Rep. ATLAS-CONF-2012-098, CERN, Geneva, Jul, 2012.

[93] ATLAS , G. Aad *et al.*, "Search for the Standard Model Higgs boson in the H $\to$ WW(*) $\to$ l ν l ν decay mode with 4.7 fb^{-1} of ATLAS data at sqrt(s) = 7 TeV," arXiv:1206.0756 [hep-ex].

[94] ATLAS , G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys.Lett.B* (2012) , arXiv:1207.7214 [hep-ex].

[95] ATLAS Collaboration , G. Aad, , *et al.*, "Combined search for the Standard Model Higgs boson in pp collisions at $\sqrt{s}$=7 TeV with the ATLAS detector," *Phys. Rev. D* **86** (Aug, 2012) 032003. http://link.aps.org/doi/10.1103/PhysRevD.86.032003.

[96] H1 collaboration , "The H1 detector at HERA," *DESY-H1-96-0* (1993) .

[97] ZEUS collaboration , “The ZEUS detector,”. http://www-zeus.desy.de/bluebook/bluebook.html.