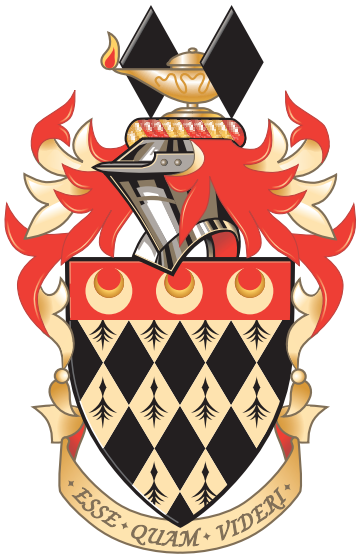
A search for the Standard Model Higgs boson produced in association with a pair of top quarks and decaying to a pair of bottom quarks.

Joshuha Thomas-Wilsker

Department of Physics

Royal Holloway, University of London



A thesis submitted to the University of London for the

Degree of Doctor of Philosophy

June 13, 2016

DECLARATION

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the document.

Joshuha Thomas-Wilsker June 13, 2016

Abstract

A search for the Standard Model Higgs boson produced in association with a pair of top-quarks and decaying to a pair of b-quarks ($t\bar{t}H(H\rightarrow b\bar{b})$) is presented. The analysis uses 20.3 fb^{-1} of data taken from proton-proton collision at centre of mass energy $\sqrt{s} = 8$ TeV, collected using the ATLAS detector. This thesis focuses on the search for dilepton events where the two leptons can be either an electron or a muon.

A study is presented comparing different techniques ("pairing methods") for correctly identifying which two b-jets in an event come from the decay of the Higgs boson. The best alternative to the default pairing method used in the Run I analysis, improves the matching efficiency by more than 25%. Individual neural networks are then trained to separate signal from background using input variables based on each pairing method.

No significant improvement on the separation obtained with the Run I neural network is achieved when using an alternative pairing method. The neural network performances are then compared with an alternative multi-variate classifier using the same pairing method, namely boosted decision trees. Comparing the two classifiers shows the best performance of the neural network is better than the best performance for the boosted decision tree. Similarly, for each individual pairing method, the neural network outperforms the boosted decision tree trained using the same input variables and pairing method.

The inclusion of additional input variables, which provide information on the topology of the event, improves the performance of the alternative classifier (boosted decision tree) by on average $\approx 5\%$. The best performing BDT, goes from a separation power of 0.1915 ± 0.003 (stat.) to 0.2012 ± 0.003 (stat.). The best performing NN goes from a separation power of 0.1925 ± 0.002 (stat.) to 0.1946 ± 0.003 (stat.). For the best performing classifier, the alternative classifier is seen to improve on the separation achieved by the neural network by 3.4% of the neural networks separation (1.5% stat. error on neural network separation).

A calibration of the MV1 b-tagging algorithm using 14.34 fb^{-1} of data collected at 8 TeV using the ATLAS detector is also presented. It is performed on a sample of jets selected from $t\bar{t}$ dilepton events, which ensures a very high-purity, inclusive sample of b-jets. b-tagging scale factors are then derived from two separate samples of jets: those which contain a muon and those which

do not. A ratio of the scale factors derived using each jet category is used as a measure of the potential bias. No significant bias is seen in either of the jet p_{T} bins, however due to the size of the uncertainties, the measurement is not sensitive to a bias of less than $\approx 8\%$.

Acknowledgements

Firstly, I'd like to thank Professor Pedro Teixeira-Dias whose supervision over the last four years has helped my development as physicist. I'd also like to thank my family for all of their encouragement and support. Without you, none of this would have been possible. I must also express my appreciation to all the people I have worked with, in particular Ian Connelly and Ricardo Gonçalo with whom I worked closely for many years. Finally, a big thank you everyone who has made the last 4 years so enjoyable. You are now scattered across the world, but you should all know how grateful I am.

Shout-out to Wychwood Machining for providing financial support over the final months.

Dedication

To my family Alison, David and Reuben.

Contents

List of Figures

List of Tables

Preface

> *Facts are meaningless. You can use facts to prove anything that's even remotely true!*

Homer Simpson, *The Simpsons*

In July 2012, the ATLAS and CMS collaborations showed strong evidence supporting the discovery of a new particle with mass 125 GeV in the search for the Standard Model Higgs boson. This new particle is referred to throughout this thesis as the Higgs boson. All of the work I performed and present in this thesis was carried out using data taken by the ATLAS detector at the LHC.

In this thesis I present a measurement of a potential bias in b-tagging scale factors.These scale factors are used to correct for the discrepancy in a b-tagging algorithms ability to b-tag b-jets in data and simulation. The potential bias is thought to be introduced when a calibration is performed on a sample of b-jets selected using the muon that comes from the decay of the b hadron. Several calibration methods used by the ATLAS collaborations during Run I are performed on such a subset of b-jets. The calibration analysis was performed on two separate categories of b-jets: b-jets where the b hadron has decayed to a muon and a neutrino, and b-jets that have decayed another way. These b-jets are selected from an enriched sample of b-jets originating from $t\bar{t}$ events. b-tagging scale factors are derived separately using each category of jet. They are then compared by taking the ratio of the scale factor derived using b-jets in which the b hadron decayed to a muon, and the scale factor derived using b-jets that decay another way. This work is presented in Chapter 4 which is derived from an ATLAS internal note [1], written by myself. The work I performed on behalf of the ATLAS b-tagging group also contributed to another ATLAS internal note [2].

After the b-tagging bias measurement, I present the $t\bar{t}H(H\rightarrow b\bar{b})$ analysis. The Higgs boson is the most recent addition to the family of fundamental particles in the Standard Model of particle physics. However we still have a relatively limited understanding of many of the particle's attributes. One of the most important attributes is its coupling to fermions. For any given Higgs boson mass, these can be predicted by the Standard Model. Values of the Higgs boson couplings that deviate away from the values predicted by the Standard Model would suggest new physics. A measurement of the coupling strength of the Higgs boson to fermions and vector bosons has been performed by the ATLAS and CMS experiments separately by combining the results from multiple search channels such as $H \rightarrow ZZ \rightarrow 4l$, $H \rightarrow WW \rightarrow l\nu l\nu$ and $H \rightarrow \gamma\gamma$. However, to better understand the nature of this particle we must also study the Higgs boson couplings to quarks. In this thesis I also present a search for the Standard Model Higgs boson produced in association with two top quarks and decaying to a pair of b-quarks, namely the $t\bar{t}H(H\rightarrow b\bar{b})$ analysis. In particular, searching for events in which both top-quarks decay to W bosons which each produce a lepton and a neutrino. This channel is of particular importance as it enables the direct study of the top Yukawa coupling along with the Higgs boson coupling to b-quarks.

Events in the $t\bar{t}H(H\rightarrow b\bar{b})$ dilepton analysis are divided into analysis regions according to their jet and b-jet multiplicity. In the three most signal significant regions, neural networks are trained to separate signal and background. These neural networks use a set of input variables that have good signal/background discrimination. Some of these variables require reconstructing a Higgs boson candidate for the event. Reconstructing the Higgs boson candidate means assigning two b-jets to the b hadrons that decayed from the Higgs boson. The method used in the analysis performed on the data taken during the first data taking period at the LHC (Run I), correctly assigned the b-jets in only 31% of the true $t\bar{t}H(H\rightarrow b\bar{b})$ dilepton events. In this thesis I present a study of alternative methods of assigning these b-jets which I have called "pairing methods". The aim of this study is to improve the efficiency of the pairing method used in the analysis in order to be able to construct neural network input variables with better discrimination power.

I also present an in depth study of the multivariate techniques used in the Run I analysis. I compared the performance of a boosted decision tree with the original neural network. Both classifiers are trained on the same events, variables and pairing method so that a like-for-like comparison can be made. I show the impact that alternative pairing methods have on the performance of each multivariate classifiers and how the performance of each classifier is affected by the use of additional

input variables. This study is documented in Chapter 5.

Subsequently, an evaluation of the parton shower systematic uncertainty in the $t\bar{t}H(H\rightarrow b\bar{b})$ analysis is presented. This uncertainty is evaluated by rerunning the analysis using an alternative signal simulation parton shower model. The systematic sample was originally generated with the HERWIG parton shower model, however this sample had particularly low statistics which meant it could not be used to get a good estimate of the systematic uncertainty. Instead a kinematic reweighting is derived and used to reweight the high-statistics nominal signal sample. The reweighted signal sample is then used in place of the systematic sample to evaluate the uncertainty. The procedure for evaluating this uncertainty is documented in the ATLAS internal note [3]. It is also included in the $t\bar{t}H$ paper published in the European Physics Journal C [4].

The structure of this thesis is as follows:

- In Chapter 1, a description of the Standard Model of particle physics is presented. A summary of its components is given along with an outline of its mathematical construction. The Higgs field is described in terms of the mass mechanism and spontaneous symmetry breaking. The Higgs boson is then associated with the relevant terms in the Standard Model Lagrangian. This is followed by a description of the properties of the Higgs boson along with some key measurements made and their consistency with theory.

- In Chapter 2, an overview of the LHC and a description of the relevant components of the ATLAS detector are given. This includes a discussion of the basic principles in accelerator physics and provides a description of the physical concepts behind the relevant components of the detector.

- In Chapter 3, an overview of how the ATLAS collaboration simulates events is given along with a description of how the physics objects are reconstructed. I make reference to the object definitions in the two analyses presented, namely the $t\bar{t}H(H\rightarrow b\bar{b})$ analysis and the b-tagging analysis, highlighting any relevant differences between the two.

- In Chapter 4, the b-tagging calibration analysis is presented. In this chapter I provide an introduction to b-tagging along with the motivation behind this calibration, the calibration methods used and finally I summarise with the results and conclusions.

- In Chapter 5, I present the $t\bar{t}H(H\rightarrow b\bar{b})$ analysis, with particular focus on the dilepton analysis. I present the motivation for the analysis, along with the analysis methods followed by the

final results and conclusions. In Section 5.11.3 I present the motivation, method and results of the study into possible event reconstruction methods and their efficiencies. In Section 5.13 I present the motivation, method and results of a study of the multivariate analysis methods. The evaluation of the signal hadronisation model uncertainty is provided in Section 5.14.

In Chapters 4 and 5 I present my own work whereas Chapters 1, 2 and 3 are a literature review of the information necessary to perform both of the analyses. Citations are provided for all of the information that is not my own work.

Throughout this thesis, natural units are used as shown:

$$c = \hbar = 1. \tag{1}$$

Energy, momentum and mass are therefore given in electron-volts (eV).

Chapter 1

Theory and Motivation

> *Light thinks it travels faster than anything but it is wrong. No matter how fast light travels, it finds that darkness has always got there first and is waiting for it.*
>
> Terry Pratchett, *Reaper Man*

1.1 The Standard Model of Particle Physics

1.1.1 The Components of the Standard Model

The Standard Model of Particle Physics (SM) describes the properties and interactions of the fundamental particles of nature in a single over-arching theory. The known elementary particles are divided into two families according to the value of their spin quantum number. This is an intrinsic property of a particle and refers to the spin angular momentum of the particle. Particles with half-integer spin are called fermions and particles with integer spin are called bosons. The fermion family can be further subdivided into quarks and leptons. There are six types of quark that form three 'generations' all of which posses colour charge. As one progresses from the first to the third generation, quarks get progressively heavier as shown in Table 1.1. The six quarks in order of mass from lightest to heaviest are up, down, strange, charm, bottom and top and have fractional electric charges of either $+\frac{2}{3}$ or $-\frac{1}{3}$. The lepton family is made up of electrons, muons, taus and their associated neutrinos. The electron, muon and tau all have an electric charge of -1 and are massive

particles whereas the neutrinos are electrically neutral and have an incredibly small mass relative to all other particles. For every quark and lepton there also exists an anti-particle with equal and opposite electric charge.

The SM also describes three fundamental forces: the strong force, the weak force and the electromagnetic force. Each of these forces derives from a fundamental field whose quantisations lead to the existence of spin-1 gauge bosons. These gauge bosons are the particles that mediate the force in particle interactions. The photon is a massless and electrically neutral gauge boson that is a quantisation of the electromagnetic field. Gluons are gauge bosons that mediate the strong force; they are also massless and electrically neutral but carry colour charge. The weak force is mediated by three gauge bosons, namely two electrically charged bosons called W^+ and W^- and an electrically neutral Z boson. The fifth boson in the SM is the Higgs boson which is not associated with any fundamental force but is associated with the Higgs field, the field that provides the mechanism which gives particles mass.

The nature of the strong force coupling is such that it becomes asymptotically stronger at smaller energy scales. This means that at high energy scales, quarks and gluons interact weakly and therefore behave like quasi-free particles. Furthermore, at very low energies, the force between quarks and gluons gets stronger. This effect is constantly seen in particle detectors. After a collision the quarks involved separate, getting further and further apart. As they do so, the energy in the gluon field increases until there is enough energy to create a quark anti-quark pair out of the vacuum. These particles then also separate creating more particles and the process repeats. The result is a 'jet' of particles travelling in a similar direction.

Eventually, the particles in these jets will not have enough energy to separate further and no more particle pairs are created. At this point, the strong force holding the quarks together cannot be overcome, and they are bound together inside what are known as mesons or baryons.

A meson is a particle made of one quark and one anti-quark bound together by the strong force. mesons have integer spin and are relatively unstable particles so typically live $O(10^{-8})$ s. baryons are formed from 3 quarks bound together by the strong force. They have half integer spin and are typically heavier than mesons. The lifetimes of baryons varies dramatically depending on the type of baryon e.g. Λ baryon lifetime $O(10^{-10})$s, whereas proton decay has not been observed [5].

	Quarks			
	Up (u)	**Charm** (c)	**Top** (t)	**F**
Mass:	$2.3\,\mathrm{MeV}/c^2$	$1.275\,\mathrm{GeV}/c^2$	$173.07\,\mathrm{GeV}/c^2$	
Charge:	$+\frac{2}{3}e$	$+\frac{2}{3}e$	$+\frac{2}{3}e$	
Spin:	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	**E**
	Down (d)	**Strange** (s)	**Bottom** (b)	
Mass:	$4.8\,\mathrm{MeV}/c^2$	$95\,\mathrm{MeV}/c^2$	$4.18\,\mathrm{GeV}/c^2$	**R**
Charge:	$-\frac{1}{3}e$	$-\frac{1}{3}e$	$-\frac{1}{3}e$	
Spin:	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	
				M
	Leptons			
	Electron (e)	**Muon** (μ)	**Tau** (τ)	**I**
Mass:	$0.51\,\mathrm{MeV}/c^2$	$105.66\,\mathrm{MeV}/c^2$	$1776.82\,\mathrm{MeV}/c^2$	
Charge:	$\pm 1\,e$	$\pm 1\,e$	$\pm 1\,e$	
Spin:	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	**O**
	Electron Neutrino (ν_e)	**Muon Neutrino** (ν_μ)	**tau Neutrino** (ν_τ)	
Mass:	$<2\,\mathrm{eV}/c^2$	$<0.19\,\mathrm{MeV}/c^2$	$<18.2\,\mathrm{MeV}/c^2$	**N**
Charge:	$0\,e$	$0\,e$	$0\,e$	
Spin:	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	
				S
	Photon (γ)	**Gluon** (g)		**B**
Mass:	$0\,\mathrm{eV}/c^2$	$0\,\mathrm{eV}/c^2$		
Charge:	$0\,e$	$0\,e$		**O**
Spin:	1	1		
	W-Boson ($W^\pm$)	**Z-Boson**(Z^0)		**S**
Mass:	$80.385\,\mathrm{GeV}/c^2$	$91.1876\,\mathrm{GeV}/c^2$		
Charge:	$\pm 1\,e$	$0\,e$		**O**
Spin:	1	1		
	Higgs boson (H)			**N**
Mass:	$125.9\,\mathrm{GeV}/c^2$			
Charge:	$0\,e$			**S**
Spin:	0			

Table 1.1: Fundamental particles in the SM of particle physics. All quantities are quoted from the particle data group 2014 review [5]. All particles in the SM are accompanied by an anti-particle (not shown in this table) which possess equal and opposite charge to the particle. All other quantities are the same. e is the absolute electric charge of the electron.

1.1.2 Constructing the Standard Model

All processes in the SM are described by a renormalisable quantum field theory in which the SM Lagrangian is invariant under local gauge transformations. This invariance is required to make

the theory renormalisable and is introduced to the SM through the local gauge symmetry $U(1)_Y \times SU(2)_L \times SU(3)_c$. Each of these terms represents a group of transformations acting on a set of vectors or matrices and are associated with one or more fields. It is these symmetries that give rise to the fundamental fields in the standard model that are associated with the spin-1 gauge bosons.

Weak hypercharge is the generator of the gauge group $U(1)_Y$, a gauge symmetry of the SM. In order to make the SM Lagrangian invariant under a local $U(1)_Y$ gauge transformation, an additional vector field B_μ is introduced to the SM Lagrangian along with a covariant derivative. The $U(1)_Y$ gauge symmetry is a component of the electroweak gauge group SU(2)×U(1) and when its associated fundamental field B_μ mixes with the fundamental electroweak fields, the Z boson and photon are generated as quantisations of the resulting physical gauge fields as shown for the photon in Equation 1.1 and the Z boson in Equation 1.3:

$$A_\mu = \sin\theta_W W_\mu^{(3)} - \cos\theta_W B_\mu \tag{1.1}$$

In both equations θ_W is the Weinberg angle (weak mixing angle).

The Standard Electroweak Model is based on the more complex SU(2)×U(1) symmetry. This symmetry introduces a singlet and a triplet with weak isospin as its generator. The triplet of additional fundamental fields are labelled $W_\mu^{(1)}$, $W_\mu^{(2)}$ and $W_\mu^{(3)}$ and are associated with the SU(2) symmetry whereas the singlet B_μ is associated with the U(1) symmetry. The combination of fields $W_\mu^{(1)}$ and $W_\mu^{(2)}$ is interpreted as two W bosons as shown in Equation 1.2:

$$W^\pm = \frac{1}{\sqrt{2}}(W_\mu^{(1)} \mp i W_\mu^{(2)}) \tag{1.2}$$

The $W_\mu^{(3)}$ field is associated with an electrically neutral boson which is part of the Z boson as shown in Equation 1.3 and B_μ remains associated with weak hypercharge:

$$Z_\mu = \cos\theta_W W_\mu^{(3)} - \sin\theta_W B_\mu \tag{1.3}$$

The third field in the standard model is the colour field which is associated with the $SU(3)_c$ group. Once again, by enforcing local gauge invariance the theory predicts the existence of a set of associated gauge bosons called gluons.

The wave-function of a particle in the Standard Model is represented as a plane wave multiplied by a four component spinor $u(E,\vec{p})$, where

$$\Psi = u(E,\vec{p})e^{i(\vec{p}\vec{r}\cdot Et)}. \tag{1.4}$$

The spinor $u(E,\vec{p})$ has both a left- and a right-handed chiral component. However, in Standard Model charged-current weak interactions, the W bosons couple to matter in vector-axial interactions. This means the associated coupling terms include the projection operators,

$$\begin{aligned} P_L &= \frac{1}{2}(1-\gamma^5) \\ P_R &= \frac{1}{2}(1+\gamma^5) \end{aligned} \tag{1.5}$$

where,

$$\gamma^5 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{1.6}$$

and P_L and P_R are the chirality operators. P_L and P_R operate on particle and anti-particle spinors, leaving just the left or right-handed components of the spinor respectively. Hence, only fermions (anti-fermions) with a left-handed (right-handed) chirality interact via the charged weak interaction.

In the Standard Model, all fermions form weak isospin doublets. Weak isospin is a quantum number which is associated with the weak interaction and is typically denoted with the letter 'T'. The third component of weak isospin namely 'T_3', is a conserved quantity and is used to group left-handed fermions (with negative chirality) into doublets. These are called left-handed chiral doublets and are formed from a pair of particles which have the same weak isospin value, T= $\frac{1}{2}$ but have equal and opposite values of the third component of weak isospin, $T_3 = \pm\frac{1}{2}$. The 6 left-handed chiral doublets are shown in Equation 1.7. Similarly, the anti-fermions in the Standard Model form their own doublets which have T= $\frac{1}{2}$ but opposite T_3 values to their fermion counterpart. Fermions with positive chirality form right-handed chiral singlets. The particles in these singlets have values of T= 0 and $T_3 = 0$ and are shown in Equation 1.8. Similarly, anti-fermions with negative chirality form right-handed chiral singlets.

$$\begin{aligned}\psi_L^{Leptons} &= \begin{pmatrix}\nu_e\\ e^-\end{pmatrix}_L, \begin{pmatrix}\nu_\mu\\ \mu^-\end{pmatrix}_L, \begin{pmatrix}\nu_\tau\\ \tau^-\end{pmatrix}_L \\ \psi_L^{Quarks} &= \begin{pmatrix}u\\ d'\end{pmatrix}_L, \begin{pmatrix}c\\ s'\end{pmatrix}_L, \begin{pmatrix}t\\ b'\end{pmatrix}_L\end{aligned} \tag{1.7}$$

$$\begin{aligned}\psi_R^{Leptons} &= e_R^-,\ \mu_R^-,\ \tau_R^- \\ \psi_R^{Quarks} &= u_R,\ d_R,\ c_R,\ s_R,\ t_R,\ b_R\end{aligned} \tag{1.8}$$

The strength of flavour changing weak interactions between quarks is characterised in the SM by the Cabibbo-Kobayashi-Maskawa (CKM) matrix. The CKM matrix is a unitary matrix in the Standard Model that contains information on the strength of flavour-changing weak decays. Each element of the matrix is proportional to the probability that a given quark of flavour i will decay to another quark of flavour j and denoted $|V_{i,j}|$. The CKM matrix containing the current best measurements of the couplings of each quark to another via the charge-current weak interaction is shown in Equation 1.9 [5]:

$$\begin{aligned}M_{CKM} &= \begin{pmatrix}|V_{u,d}| & |V_{u,s}| & |V_{u,b}|\\ |V_{c,d}| & |V_{c,s}| & |V_{c,b}|\\ |V_{t,d}| & |V_{t,s}| & |V_{t,b}|\end{pmatrix} \\ &= \begin{pmatrix}0.97425\pm0.00022 & 0.2253\pm0.0008 & 0.00413\pm0.00049\\ 0.225\pm0.008 & 0.986\pm0.016 & 0.041\pm0.0013\\ 0.0084\pm0.0006 & 0.04\pm0.0027 & 1.021\pm0.032\end{pmatrix}.\end{aligned} \tag{1.9}$$

One thing to note about this matrix, is that the off-diagonal elements which are proportional to the probabilities of decays between quarks from different generations are smaller.

To describe the dynamics and processes associated with SM processes, it is convenient to use a Lagrangian. The SM Lagrangian can be broken down into five parts [6]:

$$\mathcal{L} = \mathcal{L}_{Fermions} + \mathcal{L}_{Gauge} + \mathcal{L}_{Int.} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa} \tag{1.10}$$

In Equation 1.10, $\mathcal{L}_{Fermions}$ represents all Fermion kinetic terms, $\mathcal{L}_{Gauge}$ represents all the gauge field kinetic terms, $\mathcal{L}_{Int.}$ represents the gauge couplings with the Higgs field, fermion fields and the self-interaction terms, $\mathcal{L}_{Higgs}$ represents the scalar Higgs field kinetic energies, $\mathcal{L}_{Yukawa}$ represents the Yukawa interaction terms which mathematically represents the coupling between the Higgs

field and massless fermion fields. The inclusion of the Higgs field terms will be discussed in the next section.

1.2 The Higgs Boson and the Origins of Mass

Renormalisation ensures the cancellation of divergences in higher order Feynmann diagrams by the explicit addition of counter-terms. Gauge invariance plays a key role in the renormalisation especially in the unified electroweak theory. Requiring local gauge invariance results in the addition of gauge field terms to the SM Lagrangian which cancels out many divergent integrals and renders the theory renormalisable. However, this gauge invariance also requires all gauge bosons to be massless. This contradicts experimental evidence where both the W and Z bosons have been found to be massive particles. Therefore, the SM requires a separate mechanism to provide W and Z gauge bosons with mass. The introduction of the Higgs field provides the mechanism for which the massive particles in the SM can acquire mass without breaking gauge symmetries, whilst retaining the renormalisability of the theory. The addition of the Higgs field simultaneously introduces an associated massive gauge boson called the Higgs boson.

1.2.1 Spontaneous Symmetry Breaking

Part of the mass mechanism through which massive fermions and bosons obtain mass in the SM is called spontaneous symmetry breaking. When constructing the Lagrangian for a field, a perturbation procedure can be applied which expresses the field in terms of fluctuations from the ground state. This is done by finding the minima of the field potential and choosing a ground state.

Choosing an arbitrary ground state conceals the original symmetry of the Lagrangian in a process known as spontaneous symmetry breaking. This happens because the vacuum state (the field configuration that gives the chosen ground state) does not share the same symmetry as the original Lagrangian. In general, the new non-symmetric Lagrangian yields one massive field and at least one massless scalar field [7].

1.2.2 The Mass Mechanism

The Standard Electroweak Model is based on the $SU(2)_L \times U(1)_Y$ symmetry. Through the enforcement of local gauge invariance, the Standard Electroweak Lagrangian yields 4 massless gauge

bosons. In the Electroweak model, the mass mechanism is applied to the SU(2)×U(1) gauge symmetry of the Electroweak force to give mass to the $W^{\pm}$ and Z bosons. The simplest Higgs model in this case forms the Higgs field out of complex scalar fields placed in the weak isospin doublet:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.11}$$

The Lagrangian density for this doublet is defined in Equation 1.12:

$$\mathcal{L} = (\delta_\mu \phi)^\dagger (\delta^\mu \phi) - V(\phi) \tag{1.12}$$

where V(ϕ) is the Higgs potential shown in Equation 1.13:

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \tag{1.13}$$

and $\delta_\mu \phi$ as defined in Equation 1.14 is the derivative of ϕ with respect to each of the 4 space-time coordinates.

$$\delta_\mu \phi = \frac{\delta \phi}{\delta x^\mu} \tag{1.14}$$

Here, λ represents the Higgs self-coupling parameter which is a free parameter in the SM and μ^2 represents the minimum for the potential of the field ϕ.

The shape of the potential depends on the sign of μ^2. If $\mu^2 < 0$, the potential has an infinite set of minima, satisfying Equation 1.15:

$$\phi^\dagger \phi = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = \frac{v^2}{2} = -\frac{\mu^2}{2\lambda} \tag{1.15}$$

where 'v' is the value if the field ϕ, when ϕ is in the physical vacuum state. The shape of the Higgs field potential is shown in Figure 1.1.

These minima must correspond to the non-zero expectation value of a neutral field. This non-zero expectation value results in a massive field which therefore cannot be associated with the neutral photon field and must correspond to the neutral scalar field ϕ^0:

$$\langle 0 | \phi | 0 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{1.16}$$

Choosing the vacuum state to be in the real direction, the fields can then be expressed as an expan-

sion about the minimum:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1(x) + i\phi_2(x) \\ v + \eta(x) + i\phi_4(x) \end{pmatrix}. \tag{1.17}$$

After spontaneous symmetry breaking there is a massive scalar field '$\eta(x)$', and 3 massless scalar fields (associated with the Goldstone bosons). By choosing the appropriate gauge, the massless fields provide the longitudinal degrees of freedom for the $W^{\pm}$ and Z bosons and are thus eliminated from the Lagrangian.

In this gauge $\phi(x)$ becomes:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \tag{1.18}$$

Here $h(x)$ (i.e. the Higgs field) represents $\eta(x)$. The new Lagrangian written in terms of $\phi(x)$ is no longer symmetric.

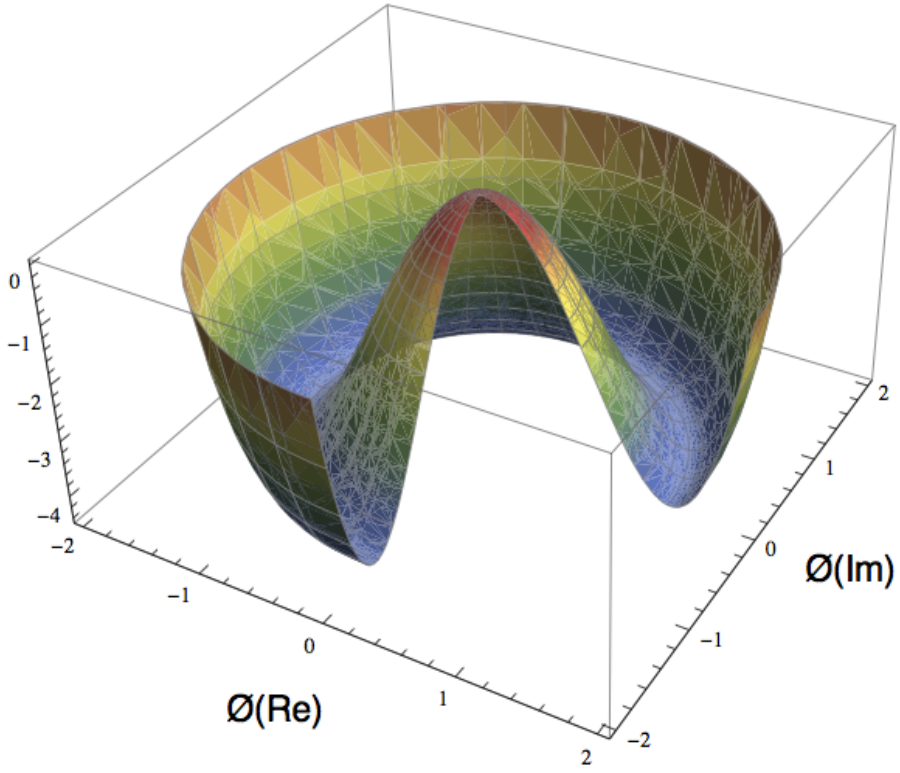


Figure 1.1: Plot of the Higgs field potential $V(\phi)$ for $\mu^2 < 0$.

The SM requires this new Lagrangian to be locally gauge invariant under infinitesimal local SU(2) gauge transformations. This is ensured by the introduction of a covariant derivative and the appropriate transformations of all fields. Once local $SU(2)_L \times U(1)_Y$ gauge invariance is implemented, mass terms can be identified through the quadratic terms in the new non-symmetric Lagrangian.

The fundamental W^1 and W^2 fields combine to generate a physical gauge field. The mass of this physical field (the W boson mass) is shown in Equation 1.19. It has dependencies on the weak coupling constant 'g_W' of the $SU(2)_L$ gauge interactions and the vacuum expectation value of the Higgs field:

$$m_W = \frac{1}{2} g_W v. \tag{1.19}$$

The Lagrangian terms that are quadratic in the neutral W_3 and A_μ fields mix into two physical

particles, with the masses shown in Equation 1.20:

$$\begin{aligned} m_A &= 0 \\ m_Z &= \frac{1}{2} v \sqrt{g_W^2 + g'^2} \end{aligned} \tag{1.20}$$

where g_W and g‘ are the couplings associated with the $SU(2)_L$ and $U(1)_\gamma$ gauge symmetries respectively.

The Glashow-Salam-Weinberg model predicts that the mass of the W and Z bosons are related by Equation 1.21:

$$\frac{m_W}{m_Z} = \cos\theta_W . \tag{1.21}$$

The physical fields are generated through the mixing of massless underlying fields associated with the $SU(2)_L \times U(1)_Y$ local gauge symmetries. The Z boson which is associated with the neutral Goldstone boson, acquires mass through the Higgs mechanism whereas the photon remains neutral. The relationship between the physical and underlying fundamental fields is shown in Equation 1.22 where θ_W is the weak mixing angle:

$$\begin{aligned} A_\mu &= \cos\theta_W B_\mu + \sin\theta_W W_\mu^{(3)} \\ Z_\mu &= -\sin\theta_W B_\mu + \cos\theta_W W_\mu^{(3)} . \end{aligned} \tag{1.22}$$

The ratio of the couplings, $\frac{g'}{g_W}$, can also be related using the weak mixing angle:

$$\frac{g'}{g_W} = \tan\theta_W \tag{1.23}$$

The electroweak symmetry breaking model is defined by 4 parameters; the couplings associated with the two local gauge symmetries, g_W and g', the vacuum expectation value ν and λ from the Higgs potential. The Higgs potential parameter is related to the Higgs field vacuum expectation value, ν, and boson mass, m_H such that:

$$
\begin{aligned}
v^2 &= -\frac{\mu^2}{\lambda} \\
m_H^2 &= 2\lambda v^2
\end{aligned}
\tag{1.24}
$$

The values of g_W (≈ 0.18), g' (≈ 0.097) and ν can be extracted from three independent physical measurements; the Z boson mass, $\sin\theta_W$ and the Fermi constant G_F are the most accurately measured and are therefore used for this purpose. The measurements of the Z boson mass and $\sin\theta_W$ are determined from the Z line shape scan at LEP whereas the Fermi constant G_F ($1.1663787 \times 10^{-5}\text{GeV}^{-2}$) is defined using the muon lifetime formula and the fine structure constant α ($\approx 1/137$), which is currently dominated by the $e^{\pm}$ anomalous magnetic moment measurement. The value of λ depends on the Higgs boson mass and can be obtained from measurements made at the LHC. The current best measurement of the Higgs vacuum expectation value ν is 246 GeV [5].

1.2.3 The Higgs Yukawa Coupling

The Lagrangian $\mathcal{L}_D$, for a relativistic spin-half field (Dirac field) represented by a four-component spinor Ψ, that satisfies the free-particle Dirac equation is:

$$
\mathcal{L}_D = i\bar{\Psi}\gamma^\mu\delta_\mu\Psi - m\bar{\Psi}\Psi. \tag{1.25}
$$

Here, Ψ is composed of left and right-chiral components, $\Psi = \Psi_L + \Psi_R$. Due to the different transformation properties of left- and right-handed-chirality states, the mass terms for fermions in the Dirac Lagrangian are not gauge invariant under the $\text{SU}(2)_L \times \text{U}(1)_Y$ symmetry. In the SM, the mass mechanism that generates the W and Z boson masses can also be relied upon to generate the fermion masses. The Higgs fields coupling to fermions is:

$$
g_f = \sqrt{2}\frac{m_f}{v} \tag{1.26}
$$

Couplings between a Dirac field and a scalar field are known as Yukawa couplings. As is shown in Equation 1.26 the coupling of the Higgs field to fermions is proportional to the mass of the particle it is coupling to (for bosons, the coupling is proportional to the square of the mass). For the top quark with a mass of 173.21 ± 0.51 GeV/c^2 the Yukawa coupling is close to unity [8].

1.2.4 Production Processes & Branching Ratios of the Higgs

The four main processes for Higgs boson production in proton-proton collisions at the LHC are: gluon-gluon fusion (ggF), vector boson fusion (VBF), associated vector boson production (VH) and associated top production ($t\bar{t}H$). The cross-section for each of these processes at the LHC is shown in a Table 1.2. The exact cross-section of each process for a Standard Model Higgs boson of mass $m_H = 125$ GeV/c^2 is shown in Table 1.2.

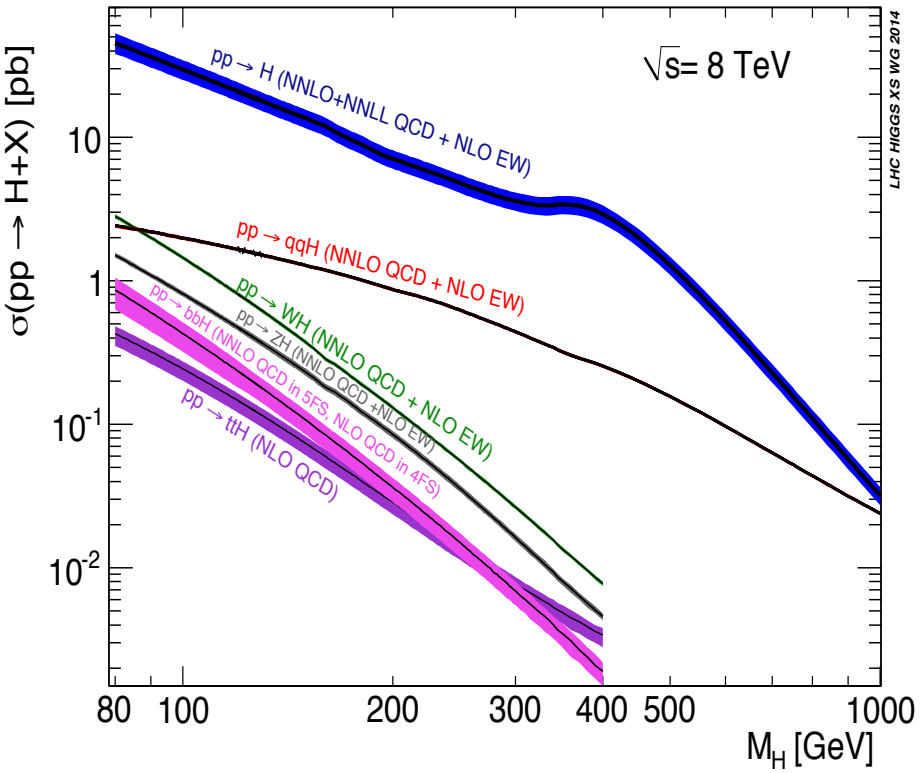


Figure 1.2: Cross-sections [pb] for the main Higgs boson production processes at the Large Hadron Collider for the Higgs boson mass range 100-500 GeV/c^2 [9]

.

Due to the large centre of mass energy in proton-proton collisions at the LHC, gluon-gluon fusion has the largest cross-section for producing Standard Model Higgs bosons [10]. This process is a factor of ten larger than all other production cross sections [11]. Seeing as the Higgs boson couples to mass, this process must be mediated by massive particles. Although the calculation of the cross-section for a gg$\rightarrow$ H process involves loops of all massive coloured fermions (quark loops), it is dominated by the top-quark loop amplitude. The next-heaviest fermion loop contributes

no more than 10% of the cross-section for this process at leading order calculations [12]. This process is subject to extremely large QCD radiative corrections [13].

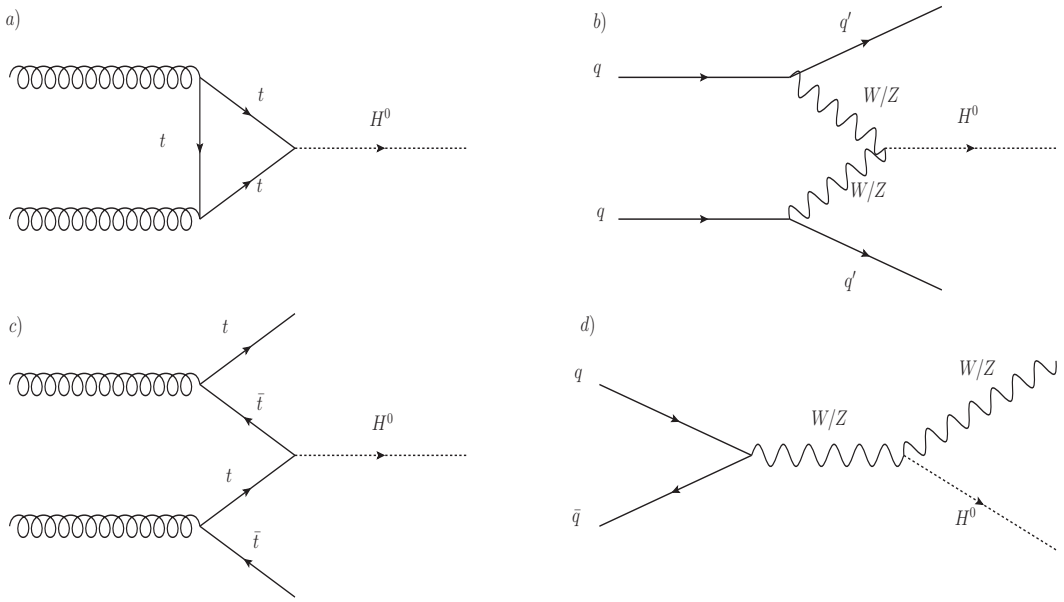


Figure 1.3: The four main Higgs boson production processes at the LHC: a) Gluon-gluon fusion (ggF) b) Vector boson fusion (VBF) c) Associated top quark production ($t\bar{t}H$) d) Associated vector boson production (VH)

Another process by which the Higgs boson is produced is by VBF. This is the process by whereby two vector bosons annihilate to form the Higgs boson (see Figure 1.3). The VBF cross-section at the LHC is an order of magnitude lower than that of gluon-gluon fusion. In this process, the Higgs boson is produced in association with two jets that are scattered in the forward direction as a result of the recoiling quarks [14]. There is not a lot of hadronic activity in the rapidity region between these jets because they are not colour connected. This makes them excellent signatures when trying to identify VBF events. The Higgs boson decay products are much more central which means they are separated from the forward part of the event and can be well reconstructed. Gluon-gluon fusion events produced with two additional jets provide the dominant background to VBF events.

Higgs boson production cross-sections (pb)					
ggF	VBF	WH	ZH	$t\bar{t}H$	Total
$19.3^{+15\%}_{-15\%}$	$1.58^{+3\%}_{-2\%}$	$0.70^{+4\%}_{-5\%}$	$0.41^{+6\%}_{-6\%}$	$0.13^{+12\%}_{-18\%}$	22.1

Table 1.2: Higgs boson production cross-sections in proton-proton collisions with center of mass energy $\sqrt{s} = 8$ TeVat the LHC for $m_H = 125$ GeV/ c^2 [5]

The VH processes have a smaller cross-section than ggF and VBF processes. However, it remains an important production channel for observing the Higgs boson as the small cross-section is compensated for by techniques that utilise easy-to-distinguish signatures such as b-jets and leptons to minimise background processes [15].

The $t\bar{t}H$ channel is a particularly important production process as it offers the possibility of measuring the top Yukawa coupling. It is difficult to observe the Higgs boson in this channel as $t\bar{t}H$ final states will suffer from large $t\bar{t}$ background. As a result, extremely accurate background and detector descriptions are required for observations of and analysis of this process.

1.2.5 Higgs Boson Branching Ratios

The coupling between the Higgs field and any fermion (boson) is proportional to the particle mass (square of the mass). This, among other factors, determines the frequency with which the Higgs boson decays into a particle and is predicted by the Standard Model.

The largest branching ratio of the 125 GeV/c^2 Standard Model Higgs boson is to a pair of b-quarks which is BR($H \rightarrow b\bar{b}$)= $57.7\%^{+3.3\%}_{-3.2\%}$ [5]. The Higgs boson can also decay via gluon pairs (gg) and photon pairs ($\gamma\gamma$). This process is mediated by loops of massive particles as the Higgs boson couples to mass. These loops provide additional indirect information on the Higgs boson couplings to heavy particles such as WW and $t\bar{t}$.

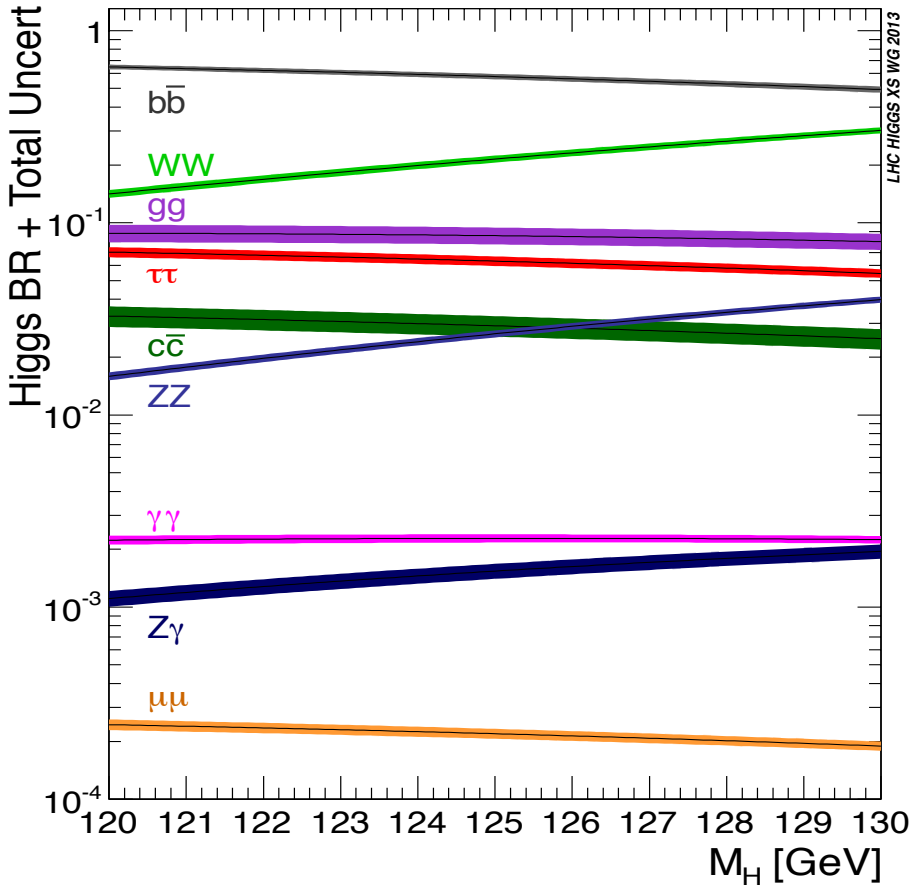


Figure 1.4: The branching ratios for the main decays of the Standard Model Higgs boson with mass near $m_H = 125$ GeV/c^2. The theoretical uncertainties indicated by the relative size of the bands [5].

1.2.6 Measurements of the Standard Model Higgs Boson

The search for the Higgs boson is performed by looking at the signature decay products of Higgs boson candidate events and minimising the possibility that they could have come from any other type of event through a series of experimental techniques. Significant evidence supporting the existence of a SM-like Higgs boson was obtained by the ATLAS and CMS experiments in 2012. The search for the Higgs boson was carried out by analysis teams from each collaboration performing separate searches in individual decay channels before combining them to calculate a total significance. The final states that contributed to the discovery of the Higgs boson were $H \to \gamma\gamma$, $H \to ZZ \to 4l$ and $H \to WW \to l\nu l\nu$ where 'l' denotes a charged lepton. The signal strength in each of the channels is shown in Figure 1.5. The two most sensitive searches are those for $H \to \gamma\gamma$

and H $\rightarrow$ ZZ $\rightarrow$ 4l. In both of these channels, the Higgs boson candidate is reconstructed on an event-by-event basis from the invariant mass of the decay products. These two channels are particularly sensitive because, despite relatively small signals, the background processes can be reduced significantly.

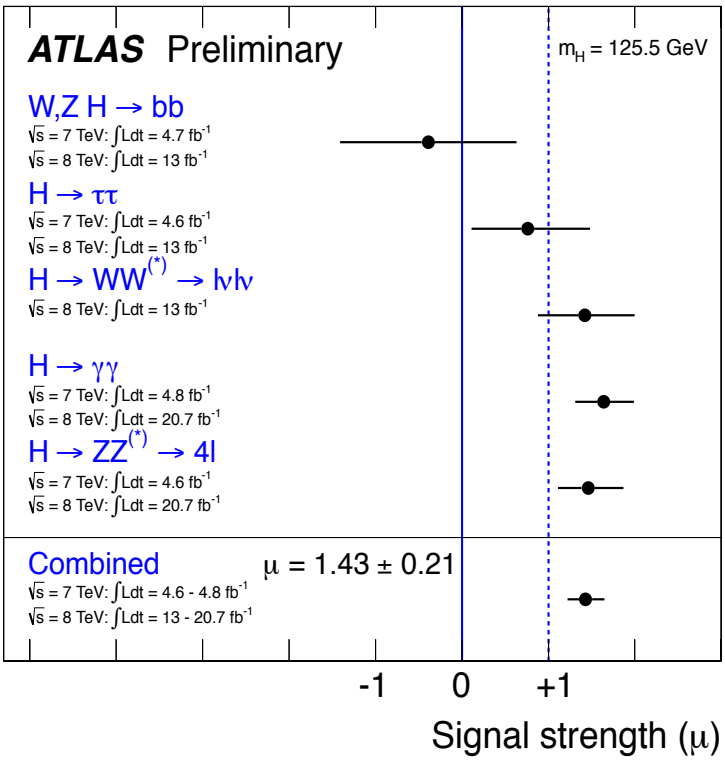


Figure 1.5: Measurements of the signal strength parameter μ for m_H= 125.5 GeV/c^2 in individual channels and their combination using the ATLAS detector [16]

.

The current best measurement of the Higgs boson mass performed by the ATLAS collaboration [16] using 4.8 fb^{-1}at $\sqrt{s}$ = 7 TeVand 20.7 fb^{-1}at $\sqrt{s}$ = 8 TeV, and combining all five decay channels is:

$$m_H = 125.5 \pm 0.4(stat) \pm 0.4(sys) \text{ GeV}/c^2 \tag{1.27}$$

The best-fit signal strength for individual channels and all channels combined is also measured. The signal strength parameter μ is measured and expressed in units of the SM Higgs boson cross-

section. The ATLAS current best measurement for the combined signal strength for a Higgs boson mass of 125.5 GeV/c^2 is:

$$\mu = 1.43 \pm 0.16(stat) \ \pm 0.14(sys). \tag{1.28}$$

The same measurement performed by the CMS experiment with an integrated luminosity of 5.1 fb^{-1}of 7 TeV data and 19.6 fb^{-1} of 8 TeV data taken from proton-proton collisions using the LHC [17] also using all five decay channels is:

$$m_H = 125.7 \pm 0.3(stat) \ \pm 0.3(sys) \ \text{GeV}/c^2 \tag{1.29}$$

The CMS current best measurement for the combined signal strength for a Higgs boson of mass 125.7 GeV/c^2 is:

$$\mu = 0.8 \pm 0.14. \tag{1.30}$$

The consistency of the measurement of the couplings of the observed resonance with respect to those predicted by the SM for the Higgs boson was tested by both experiments and no significant deviations from the SM Higgs boson couplings were found.

The observed $m_{\gamma\gamma}$ spectrum from the di-photon analysis is shown in Figure 1.6. The solid red curve show the fitted signal-plus-background model where the Higgs boson mass has been fixed at 125.4 GeV. The background component of the fit is shown by the blue dotted line. The solid black curve close to the x-axis shows the signal component of the fit. The plot beneath the main figure shows the ratio between the data and and the background component of the fitted model.

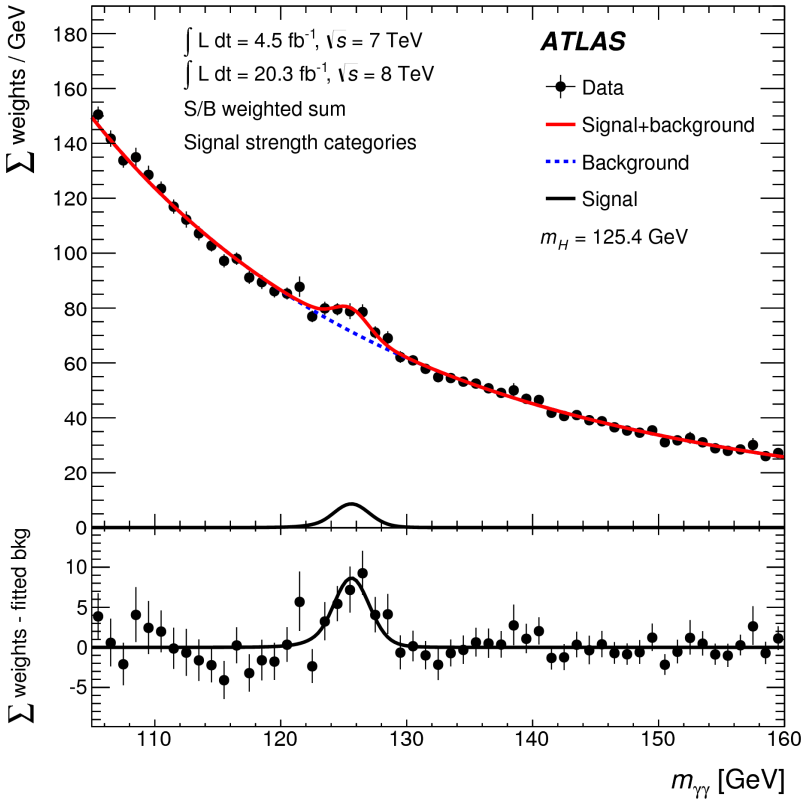


Figure 1.6: The observed $m_{\gamma\gamma}$ spectrum from the diphoton Higgs boson search [18].

Chapter 2

The Large Hadron Collider and the ATLAS Detector

> *Doomsday device? Ah, now the balls in Farnsworths court. I suppose I could part with one and still be feared.*
>
> Professor Hubert J. Farnsworth, *Futurama*

2.1 The LHC

The Large Hadron Collider (LHC) is a high energy particle collider located in Geneva, Switzerland. The main ring of the collider is 27 km in circumference located underground at the border between France and Switzerland. The collider was designed with the intention of discovering new physics at high energies using a range of experiments based on the CERN (Conseil Européen pour la Recherche Nucléaire) main site.

One of the main functions of the LHC is to collide protons together. In order to create the protons, hydrogen atoms are stripped of their electrons readying the protons for injection into the first stage of acceleration. This is a linear accelerator called Linac 2 which brings the protons energy up to 50 MeV. The protons are then passed into the Proton Synchrotron Booster which consists of 4 rings. Each ring is surrounded by powerful magnets to bend the protons trajectory round the circle and a cyclical electromagnetic field accelerates the particles to 1.4 GeV. Subsequently the protons are accelerated by the Proton Synchrotron (PS) to 25 GeV, and then by the Super-Proton

Synchrotron (SPS) up to 450 GeV. The protons are then injected into the LHC main ring where they were accelerated to a maximum energy of 3.5 TeV in 2011, 4 TeV in 2012 and 6.5 TeV in 2015.

Two separate beams of protons are injected into the LHC in opposite directions. Metallic chambers containing electromagnetic fields oscillating at radio-frequency (400 MHz) are used to accelerate the particles around the LHC main ring. These are called radio-frequency (RF) cavities. A RF powered generator is used to set up a an resonant electromagnetic field inside the cavity. The electromagnetic field transfers energy to the charged particles passing through the cavity pushing them forwards along the accelerator. An ideally timed particle (synchronised with the RF frequency) will see no acceleration whereas particles arriving slightly later/earlier will see an acceleration/deceleration respectively. This separates beams into swarms of protons called bunches which contain roughly 10^{11} protons per bunch. There can be up to 1400 bunches present in each beam.

When in circulation protons are travelling close to the speed of light. Dipole magnets are used to steer the protons round the LHC main ring. Because protons are electrically charged, proton beams diverge so quadrupole magnets are used to control the width and the height (focus) of the beams. A very large electromagnetic force is required to steer the protons and maintain the focus of the beam. This requires a large magnetic field and electric current which are achieved by using superconducting magnets. These magnets need to be cooled to temperatures close to absolute zero to obtain their superconducting state.

Protons are injected into the LHC main ring by the SPS and accelerated to the full potential energy. Once fully accelerated, the circulating beams begin colliding. There are four predetermined collision points at each of the major experiments around the LHC ring. Multipole magnets are used to focus the beams at the collision points producing around 27 collisions per bunch crossing. The number of proton-proton collisions per bunch crossing is known as pileup. Not all of the collisions in a bunch crossing will be kept for further analysis as they do not all contain the signatures of an interesting physics collision. However, they can interfere with the recorded information from events that do. The effect of pileup is taken into account when performing physics analyses.

Proton-proton collisions reduce the number of protons left in the circulating beams and causes the instantaneous luminosity to decrease. When the instantaneous luminosity of the decreases below a threshold value, the beams are rapidly extracted from their orbit and fired into large graphite

absorbers to end their circulation.

Around the LHC main ring are 4 collision points at which the proton-proton beams are focused. At each of these collision points, a detector is located. The ATLAS detector is one of 4 main detectors around the main circuit at the LHC. It is a multi-purpose particle physics detector designed to be able to measure all possible final states we expect to observe from proton-proton (Pb ion) collisions.

2.2 The ATLAS Coordinate System

To describe the position of particles within the detector, ATLAS uses a cylindrical coordinate system. All points are described relative the orthogonal $x-$, $y-$, and $z-$axes with the positive $z-$direction pointing along the beam axis around the LHC circuit. The transverse plane is then described by $x-$ and $y-$coordinates. The $x-$axis points towards the centre of the LHC circuit, whereas the $y-$axis points towards ground level. The azimuthal angle ϕ is in the transverse $x-y$ plane and is defined from the positive $x-$axis. The polar angle θ is defined from the positive $z-$axis in the $z-y$ plane.

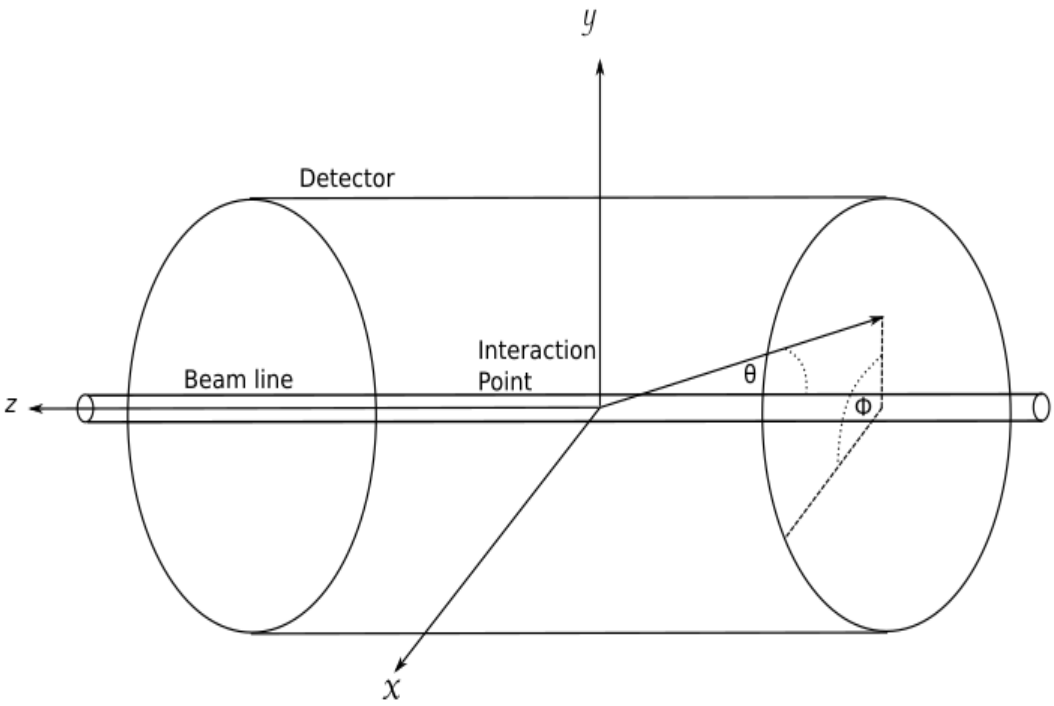


Figure 2.1: The ATLAS coordinate system with the $x-$axis pointing towards the centre of the LHC ring, the $y-$axis pointing upwards toward ground level and the $z-$axis pointing along the beam direction.

As the protons involved in collisions at the LHC don't have equal and opposite momentum, their centre of mass frame can be boosted along the z-direction with respect to the lab frame. Subsequently, the value of $\Delta\theta$ between two particles changes depending on the size of this boost. Lorentz invariant quantities are therefore used which enables the comparison of kinematic quantities without distorting results with these effect.

Rapidity is defined in the following way:

$$y = \ln \sqrt{\frac{E + p_z}{E - p_z}}. \tag{2.1}$$

In this equation y represents rapidity, E represents the Energy of the particle and p_z represents the $z-$component of the momentum. This variable can be used in place of θ because it is Lorentz-invariant and is dependant on p_z, where the size of the $z-$component of the particles momentum is proportional to the angle θ. For particles with most of their momentum down the beam line, and therefore at low θ, the rapidity will tend to infinity. Particles with most of their momentum in the

transverse plane and therefore large θ, rapidity will tend to 0.

Rapidity calculations require both the total energy and p_z of the particle to be known. These values are difficult to obtain for particles which take a trajectory close to the beam line due to design limitations and so rapidity is not commonly used in hadron collider physics. Instead, An approximation of the rapidity variable called pseudorapidity is used which is defined by equation 2.2. For highly relativistic particles, pseudorapidity and rapidity are approximately equal.

$$\eta = -\ln\tan\frac{\theta}{2} \tag{2.2}$$

2.3 The ATLAS Detector

The ATLAS detector has a cylindrical geometry consisting of several sub-detector systems which are organised concentrically around the beam pipe. The inner most layer is called the 'inner detector'. It is situated inside the central solenoid magnet, immersing it in a 2 Tesla axial magnetic field. This causes the trajectories of charged particles traversing the inner detector to be curved leaving an arced track. The track creates a segment of an incomplete circle in the detector, the radius of which can be used in equation 2.3 to calculate the momentum/charge ratio of charged particles (e.g. electrons, muons, hadrons).

$$\frac{mv^2}{r} = Bqv. \tag{2.3}$$

In this equation m is the particles mass, v is its velocity, r is the radius of the circle, B is the magnetic field strength and q is the charge of the particle.

Outside the inner detector are two calorimeters; the electromagnetic calorimeter (EM calorimeter) and the tile calorimeter (or hadronic calorimeter). Each designed to absorb the energy deposited by different types of particles, either via electromagnetic or hadronic interactions respectively, which result from the proton-proton collision.

The muon spectrometer is located in the around the calorimeters. This is because the muons are the only particle aside from the neutrinos that consistently reach the outermost part of the detector. Like the inner detector, the muon spectrometer is kept inside a magnetic field to give it tracking capabilities. The magnetic field in this case is created using toroidal magnets which are described in more detail in section 2.3.8. The detector's closed volume allows one to measure

neutrino's transverse momentum by calculating the missing transverse energy. The details of these subsystems will be discussed further in this chapter.

2.3.1 The Inner Detector

The ATLAS inner detector (ID) is comprised of three subsystems. These are the pixel detector, the silicon micro-strip semiconductor tracker (SCT) and the transition radiation tracker (TRT). The three subsystems are used together to improve tracking measurements.

The innermost subsystem is the pixel detector, followed by the SCT and then the TRT. The inner detector consists of three units, the barrel section and two end caps to complete the cylindrical geometry of the detector. In the barrel region, the sub-detectors are arranged in concentric cylinders around the z-axis while the end caps are circular disks placed perpendicular to the z-axis in the x-y plane allowing tracking coverage up to $|\eta| \leq 2.5$. With an average of 36 hits per track, the ATLAS inner detector provides continuous tracking and improved momentum resolution over $|\eta| < 2.0$ along with an electron identification capability that is complementary to that provided by the calorimeters.

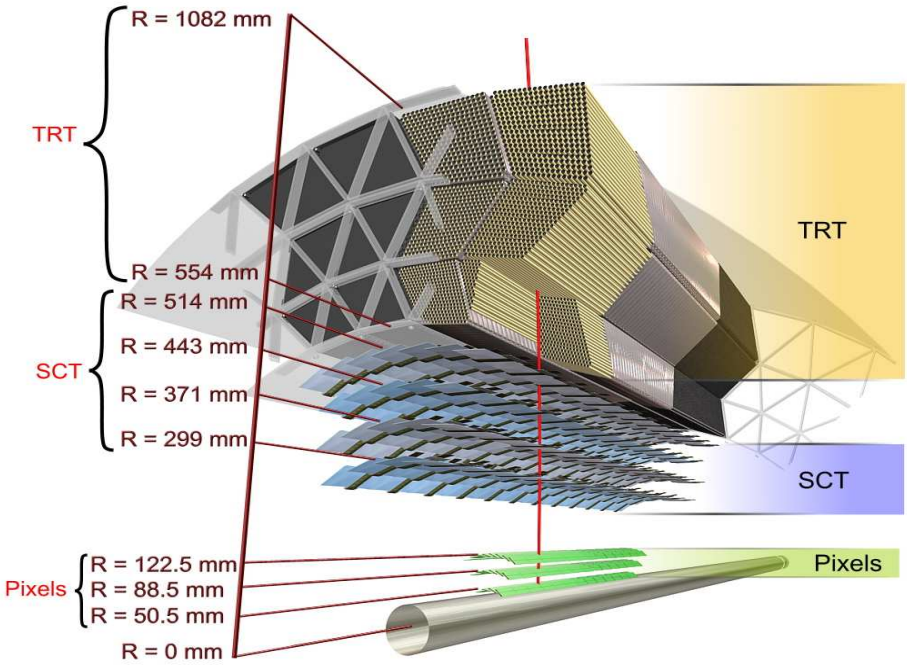


Figure 2.2: Schematic of the ATLAS inner detector highlighting the separated subsections: transition radiation tracker, semi-conductor tracker and pixel detector. [19]

Both the pixel detector and the SCT use semiconductor tracking technology. An electric field is induced across the semiconducting material in the trackers. As the charged particles traverse the semiconductor, an electron-hole pair is created which, due to the electric field, drift towards specially doped regions of the material where the charge builds up. This subsequently produces a current that flows towards the electronic readout system. As the charge passes through the detector, it creates these electron-hole pairs in multiple layers of detector (see Sections 2.3.2 and 2.3.3 for more detail). These measurements are combined to make up a complete track tracing the particles trajectory through the detector. This is what is then used to calculate the charge-momentum ratio of the particle. The technology used in the TRT is described in detail in section 2.3.4.

2.3.2 The Pixel Detector

The Pixel detector consists of concentric layers of silicon pixels around the the beam axis. These are inside the central solenoid which creates a nominal magnetic field of 2 T. The silicon layers in the barrel are situated at radii of 50.5 mm, 88.5 mm and 122.5 mm from the z-axis. The end caps are made up of 3 circular disks located at each end of the detector. The 3 layers are situated at ± 495 mm, ± 580 mm and ± 650 mm along the z-axis from the nominal interaction point. This gives it a coverage of $|\eta| < 2.5$ for position measurements up to 11 cm from the nominal collision point.

2.3.3 The Silicon Micro-strip Semiconductor Tracker (SCT)

The silicon micro-strip semiconductor tracker is composed of 4 layers of silicon wafers situated around the pixel detector. The aim of the SCT is to provide a complementary measurement of a charged particle's trajectory, perpendicular to the beam axis. Its micro-strip technology allows for a more practical coverage of a larger area when compared with the pixel detector. Each silicon wafer is 6.3 by 6.4 cm^2 with 768 active strips. Individual modules consist of four detectors, two on each side back to back, at a 20 mrad angle. The nominal spatial resolution of the SCT is 17 μm in r-ϕ and 580 μm in z-x. As with the pixel detector, the SCT is composed from three units, the barrel and two end caps, with coverage up to $|\eta| \leq 2.5$.

This tracker is used to take 4 high precision measurements per track in r-ϕ-z coordinates at an intermediate radial range and uses small angled stereo to obtain a z measurement. The SCT has reduced granularity when compared to the pixel detector but due to being positioned further

away from the interaction point it has a lower occupancy rate. The SCT is used to contribute towards transverse momentum calculations but also vertex locating and measurements of the impact parameters from vertices of interest.

2.3.4 The Transition Radiation Tracker (TRT)

The transition radiation tracker is a straw detector located around the SCT at the radial range 554-1082 mm. Each straw has a diameter of 4 mm. The straws contain a Xenon-based gas mixture as the active material and a gold-plated Tungsten-Rhenium readout wire. When a particle traverses the $XeCO_2$ gas it ionises the Xenon atoms liberating electrons. The resulting free electrons drift towards the anode wire where they are amplified and read out.

The straws are interleaved with layers of polypropylene fibres (barrel region) and foils (end caps) which act as radiators [20]. The dielectric medium changes drastically between a radiator and a straw which instigates the production of transition radiation photons by the highly relativistic particles that traverse the medium. The resulting radiation (5-30 keV) is subsequently absorbed by the Xe atoms, leading to the release of more electrons and significantly increasing the read out signal. The amount of transition radiation is proportional to how relativistic the particle is which means that typically, lighter particles produce more transition radiation than heavier particles. Subsequently, this information can be used to distinguish between electrons and charged pions.

The TRT is used alongside the SCT and Pixel detector, helping with pattern recognition when reconstructing the tracks, by adding around 36 additional hits per track. It also improves the momentum measurement in the inner detector by contributing a precision that would correspond to a single point measurement of 50 μm.

2.3.5 The Electromagnetic Calorimeter

The electromagnetic calorimeter is designed to make precision measurements of the energies of electrons (positrons) and photons above a few MeV and also has the capability to measure the energy deposited in the calorimeter by jets via electromagnetic processes. It is a sampling calorimeter which consists of layers of material which force electromagnetically interacting particles form electromagnetic showers.

The ATLAS electromagnetic calorimeter uses lead absorption layers to force the shower. When

a particle interacts with the lead absorption layer they produce a shower of low energy electromagnetically interacting particles, for example electrons, positrons and photons (electromagnetic shower). The main processes involved in creating this shower are pair production from photons and Bremsstrahlung radiation by energetic electrons. Absorption layers are interleaved with layers of liquid argon which is used as the active element to measure the energy deposits. Liquid argon was chosen as it is a radiation hard noble gas with a high atomic number (Z) (and subsequently a short radiation length). The charged component of the shower ionises the LAr in the gaps inducing more negatively charged electrons and positively charged ions. The negative charge drifts towards the copper electrodes where the signal is summed and read out. By measuring the charge deposited on the electrode it is possible to calculate the original energy of particle when it entered calorimeter.

The electromagnetic calorimeter is located around the inner detector in the radial range 1150-2220 mm. It is divided along the z-axis into three sections: the barrel region and the two end caps, one either side of the barrel region. Because all three regions use liquid argon as the active material, they can share the same cryostat. Each section is a located inside an aluminium cryostat which keeps the temperature of the liquid argon at a constant 87 K. The signal response depends on the temperature of the liquid argon (varies 2%/K) so the uniformity of the temperature inside the detector is of vital importance. The barrel covers the pseudorapidity range $0 \leq |\eta| \leq 1.52$ and the end caps cover the range $1.375 \leq |\eta| \leq 3.2$. The barrel contains 16 3.2 m long modules which stretch radially outward. They consist of 64 accordion-shaped absorbers interleaved with readout electrodes immersed in liquid argon as shown in figure 2.3. The accordion structure minimises dead-zones in the ϕ-direction.

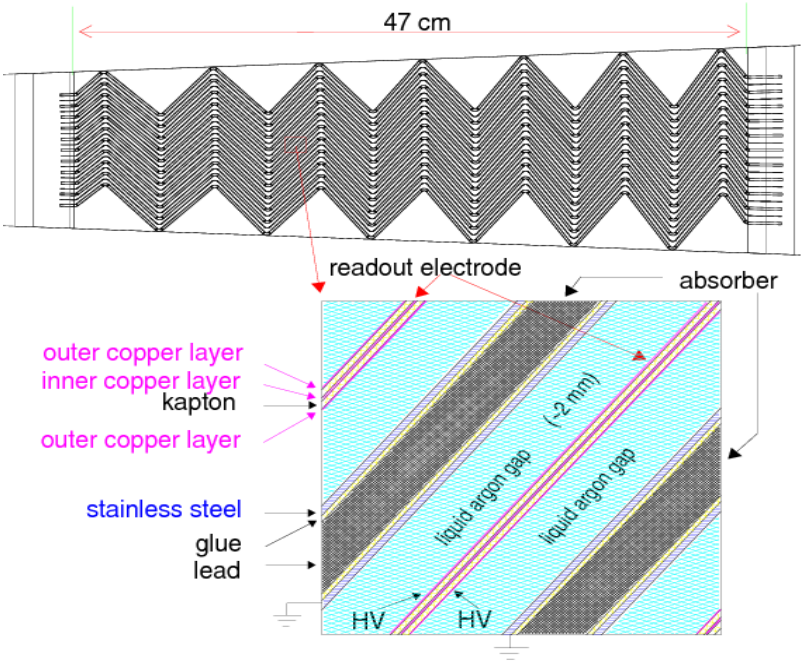


Figure 2.3: Schematic of the ATLAS electromagnetic calorimeter. Image shows the accordion structure of the electrodes that run through the calorimeter and the interleaved layers of lead absorbers and electrode read-out wires immersed in liquid argon [21],[22].

The calorimeter is composed of three concentric sampling layers, each with a different granularity. At the innermost edge of the electromagnetic calorimeter is the pre-sampler which is used to correct for the energy lost in the material between the beam axis and the calorimeter. Outside this there is the 1^{st} sampling layer, as show in figure 2.4 which has the finest granularity of $\Delta\eta \times \Delta\Phi = 0.0031 \times 0.098$. Due to its ultra-fine strip cells, this layer is used to distinguish between photons and π^0 hadrons using shower-shape variables and looking for substructure. There are then two more sampling layers with increasing cell sizes. The square cells in the second sampling layer are $\Delta\eta \times \Delta\Phi = 0.025 \times 0.0245$ and in the third layer they are $\Delta\eta \times \Delta\Phi = 0.05 \times 0.0245$.

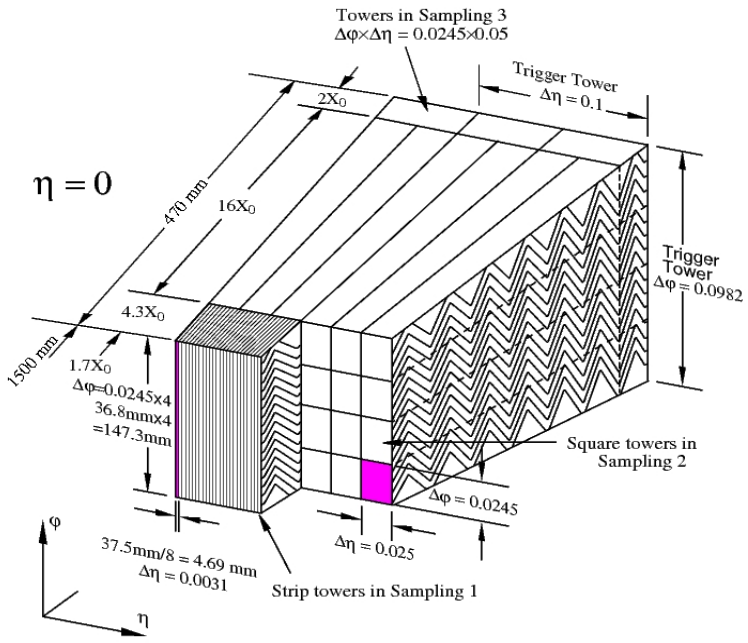


Figure 2.4: Schematic of the ATLAS electromagnetic calorimeter. Image shows the layout of a the three sampling layers in the calorimeter [23]

2.3.6 The Hadronic Tile Calorimeter

The Hadronic Tile Calorimeter (HCAL) is a large hadronic sampling calorimeter. It is used to measure the energy of hadronic showers produced by the interactions of the mesons and baryons from a collision with the atomic nuclei of the detector material. The tile calorimeter is composed of steel sheets interleaved with layers of scintillating tiles. Along the edges of each scintillating sheet are wavelength shifting (WLS) fibres which transport the photons generated in the scintillator to the photomultiplier tubes (PMTs).

When a particle passes through the steel absorber, it creates a shower of lower momentum particles (hadronic shower). When these particles reach the scintillating material, they create many electron-hole pairs. The scintillating material is doped with impurities. Points at which the impurity atoms exist in the scintillators lattice are called activator sites. Electron-hole pairs drift to the activator site as the ionisation energy is lower. The electron de-excites by transitioning through energy levels emitting visible/UV light. The light produced in the scintillating material is collected by WLS fibres which carry the photons to the PMT. These then convert the light into an electric current. The size of the signals read out is then used to calculate the energy of the primary hadrons

that entered the hadronic calorimeter. The nuclear interaction length (defined for a material as the average distance a hadron has to travel inside that medium before a nuclear interaction occurs) for hadronic showers is longer due to the smaller cross-section of the nuclear interactions and thus, the tile calorimeter is radially larger than the EM calorimeter.

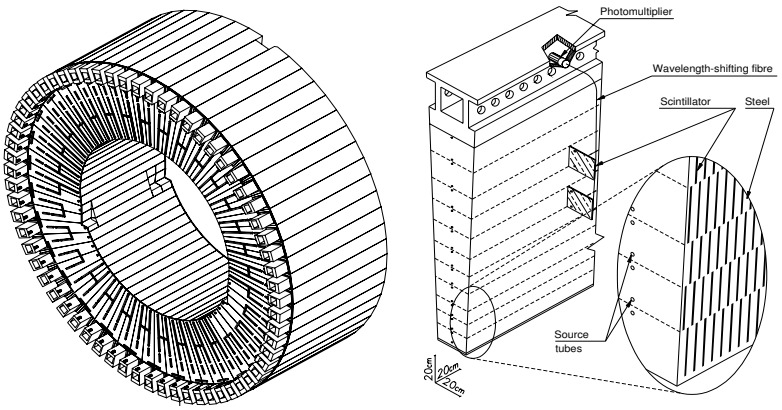


Figure 2.5: Schematic of the ATLAS tile calorimeter. Image shows the interleaved layers of scintillator and steel with WLS fibres running along sides of scintillator sheets toward PMTs [20].

This calorimeter comes in the form of a cylindrical shell, situated around the EM calorimeter with an inner edge at a radial depth of 2280 mm and an outer radial depth of 4230 mm from the beam axis. The calorimeter is divided into three regions: the barrel and two extended barrel regions either side. The central barrel region covers $|\eta| < 1$ and the extended barrel covers $0.8 < |\eta| < 1.7$. It is also subdivided into three layers radially, each with a radiation depth (which describes the rate of energy loss by a particle traversing the material) of 1.4, 3.9, 1.8 nuclear interaction lengths from the innermost to the outermost layer, respectively.

2.3.7 Energy Resolution

Good calorimeter energy resolution is essential for many analyses. For example, good mass resolution is required from the EM calorimeter in search for the Higgs boson decaying to a pair of photons as the analysis is designed to search for a narrow di-photon resonance above a large irreducible $\gamma\gamma$ background. Furthermore, good jet energy resolution is required by many analyses such as SUSY searches which often look for di-jet resonances and for example the b-tagging calibration

presented where jet energy resolution is a large systematic uncertainty for low p_{T} jets.

The energy resolution of calorimeters can be parametrised as [24],

$$\frac{\sigma}{E} = \frac{\alpha}{\sqrt{E}} \oplus \frac{\beta}{E} \oplus \gamma \tag{2.4}$$

where α is the sampling term coefficient which denotes the statistical shower developement, β is the noise term (e.g. electronic noise) and γ is a constant with contributions from instrumental effects such as leakage effects (both longitudinal and lateral) and loss of energy in dead (non-detecting) material. To meet the mass resolution required for Higgs searches, the sampling term is required to be 10% for the EM calorimeter, 50% for the hadronic calorimeter and 100% for the forward calorimeters (EM end caps and extended barrel regions of the HCAL).

2.3.8 The Muon Spectrometer

The Muon Spectrometer (MS) is the largest sub-detector in ATLAS. It uses a strong magnetic field to bend the trajectory of charged particles. The curvature of the trajectory is then used to derive the momentum of the particle. The ATLAS MS is designed to measure this independently of the inner detector and can therefore be used in a complementary way.

The MS barrel region is composed from three concentric layers around the beam pipe at radial depths of 5, 7.5, 10 m from the beam axis. The layers are referred to as the inner, middle and outer layers with the inner layer closest to the beam pipe located just outside the Hadronic Tile Calorimeter. The MS also consists of 4 end-cap disks situated at $\pm$ 7, 10.8, 14, and 21.5 m along the z-axis from the nominal interaction point. The MS is capable of momentum measurements in the pseudorapidity range $|\eta| < 2.7$ and can trigger on particles in the region $|\eta| < 2.4$ using resistive plate chambers (RPCs). The detector layout is shown in figure 2.6.

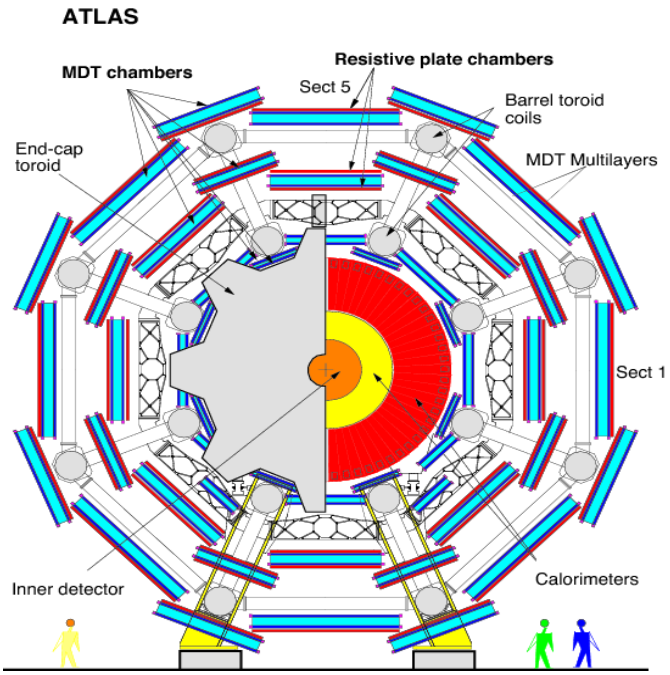


Figure 2.6: Schematic of the ATLAS Muon Spectrometer [25]

A precision measurement of the track coordinates in the principal bending direction of the magnetic field is provided by three layers of Monitored Drift Tube chambers (MDTs). In addition to the MDTs, the forward region of $2.0 < |\eta| < 2.7$ Cathode-Strip Chambers are used in the inner layer to give a better time resolution and higher rate capability.

The muon spectrometer momentum resolution is parameterised according to the following equation:

$$\frac{\sigma(p)}{p} = \frac{p_0^{MS}}{p_T} \oplus p_1^{MS} \oplus p_2^{MS} \cdot p_T \tag{2.5}$$

where p_0^{MS}, p_1^{MS} and p_2^{MS} are coefficients related to the energy loss in the calorimeter material, multiple scattering and the intrinsic resolution terms respectively. The spectrometer is capable of providing a momentum measurement with a resolution better than 3% over a wide p_{T} range and 10% at $p_{\mathrm{T}} = 1$ TeV, where the p_{T} is the momentum of the muon in the plane transverse to the beam pipe [26].

The magnet system of the MS consists of 8 large air-core superconducting barrel toroids and two end-cap air toroidal magnets. The magnetic field provides the deflection of the particle trajectories which is orthogonal to the the trajectory itself and allows for a standalone measurement. The

MS can provide a stand-alone muon momentum measurement with a momentum resolution of 10% for a 1 TeV track. A 3-point measurement of a muon track is made in the *r*-*z* projection parallel to the bending direction of the magnetic field, which is accurate to within 50μm. Muons traversing the MDTs ionise the Ar/CO_2 gas liberating electrons which then drift toward the Tungsten-Rhenium wire to be read out.

2.3.9 The ATLAS Trigger System

The LHC design luminosity is around $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. Bunch crossings occur at 50 ns intervals which corresponds to a 20 MHz crossing rate with an average of 27 proton-proton interactions per bunch crossing. The ATLAS experiment data output rate in 2012 was ≈ 400 MHz which is far too many events to record. Hence, the number of events must be reduced and this must be done in a way that we ensure we are not disregarding potentially interesting events. This is done is using 'triggers' which are a chain of algorithms that make a very fast decision as to whether the event comes from an interesting physical process or not. If an event is interesting, the event information is kept and if not, it is thrown away. These types of decisions are performed in three stages.

The Level-1 (L1) trigger is a hardware based trigger which makes the first decision using just the muon spectrometer and the calorimeters. This trigger is capable of identifying 'regions-of-interest' (ROI), which are cones in η-ϕ space that contain muon candidate tracks or clusters of energy deposits in the calorimeter. Cones in η-ϕ space are identified as ROIs if the these objects exceed a threshold value. The L1 trigger takes roughly 2.5 μs to make its decision. If no ROI is found for a given collision, the event is disregarded, otherwise it will be buffered and the next trigger in line (the Level-2 trigger) is required to make a decision. The L1 trigger reduces the data rate to around 100 kHz.

The software-based Level-2 (L2) trigger looks at ROI's in the detector that have been identified by the L1 trigger. The L2 trigger utilises ID tracking information as well as input from the muon spectrometer and calorimeters but only from the ROI. This is used to reduce the data output rate to around 3 kHz. For each event, the L2 trigger makes a decision in 40 ms, after which the event information is passed to the Event Filter.

The Event Filter (EF) is the final trigger level in the ATLAS triggering system. If an event reaches the EF trigger and fulfils all the criteria warranted to keep an event, it will be kept in permanent storage. The EF is a software based trigger which uses offline analysis procedures on

fully built events to reduce the dataset to around 400 Hz, making its per event decision in a few seconds. L2 and the EF form what is known as the High Level Trigger (HLT) and both use all of the detector subsystems and their full granularity to make their decisions.

The data acquisition system is what steers the event information until the point where it is placed in permanent storage. It receives event data at the L1 trigger rate which it transmits to the L2 trigger upon the L2 triggers request. Events passing the L2 requirements are built and then passed to the EF by the data acquisition system and moved to permanent storage.

Data can be streamed off into various output files based on the trigger decision made for an event. Stream names represent the kind of trigger signatures the events will contain. For example if an events passes the single-electron or photon trigger it will end up in the so-called "Egamma" stream. ATLAS uses an inclusive streaming model which means that if an event activates more than one trigger it can be stored in multiple streams.

2.3.10 Data Quality

Data collected and stored by the ATLAS is required to undergo a series of quality checks. To define a good dataset, data quality information from a series of luminosity blocks (grouped into a "run") is analysed by a group dedicated to this task. A luminosity block is a unit of time for data taken and they last typically around 2 minutes. Recommended good runs lists (GRLs) are compiled by the data quality group. For both analyses the "AllGood" GRL was used which means all detector components were functioning normally for the duration of each run in the list.

Chapter 3

Event Simulation and Object Reconstruction

> *Seeing, contrary to popular wisdom, isn't believing. It's where belief stops, because it isn't needed any more.*
>
> Terry Pratchett, *Pyramids*

3.1 Introduction

The studies presented make reference to two analyses, namely the 'b-tagging calibration analysis' referred to in Chapter 4 and the $t\bar{t}H(H\rightarrow b\bar{b})$ analysis referred to in Chapter 5. A description of the event simulation and object reconstruction used in each of the analyses is outlined in this chapter along with any corrections that are applied. However, a discussion of the simulated signal and background processes is saved until the relevant chapters.

The $t\bar{t}H(H\rightarrow b\bar{b})$ analysis is concerned with the decay of the Higgs boson produced in association with a pair of top-quarks and decaying to a pair of b-quarks. Three $t\bar{t}H(H\rightarrow b\bar{b})$ analyses were performed before the LHC shutdown in 2013. The analyses ensured orthogonality between their signal events by targeting events according to the decay of the top quarks and rejecting events that the other analyses selected as signal events. The dilepton analysis looks for events where both the W bosons from the top quarks decay to a lepton and a neutrino, the single lepton looks for events where one decays leptonically and the other hadronically and the all-hadronic analysis looks for

events where both W bosons decay hadronically. This thesis focuses on the dilepton analysis. The high jet (and b-jet) multiplicity final states of this analysis mean that the focus of this chapter is on the reconstruction of jets (and b-jets). Furthermore, a description of how leptons are reconstructed is necessary as the focus is on the dilepton analysis.

Particular attention is paid to muons due to their importance in the b-tagging calibration analysis. A comprehensive understanding of this analysis also means that the reconstruction of tracks and vertices inside the ATLAS inner detector is necessary. However, an extensive discussion of the b-tagging procedure is left until Chapter 4.

The b-tagging calibration analysis is performed using $t\bar{t}$ dilepton events. This is the main background in the $t\bar{t}H(H\rightarrow b\bar{b})$ analysis and so a repeated description of the reconstruction of many of the same object types is unnecessary. However, any discrepancies in the way objects are reconstructed have been highlighted, otherwise one should assume the same definition is used in both analyses.

3.2 Event Simulation

Protons are composite particles consisting of a large number of quarks and gluons. The proton's total momentum is distributed amongst the constituent particles according to its parton distribution function (PDF). When two protons collide it is the interaction of a gluon (or quark) in each proton that determines the nature of the collision. Simulations of processes in proton-proton collisions at the LHC are based on phenomenological models which use these PDFs.

Simulated collisions are described in two parts according to the energy scale the process occurs on: the hard scatter and the underlying event. The hard scatter describes processes which occur at an energy scale of more than a few GeV whereas the underlying event describes the softer processes. Monte Carlo generators often use lowest order matrix element (ME) calculations to simulate a given hard process. The particles generated in the matrix element calculation are not stable particles.

Final state matrix element particles may not be stable particles, which either decay into or radiate more particles. In the case of quarks or gluons (partons), subsequent partons also radiate further partons developing into a parton shower. Partons produced at the end of the parton shower cannot exist as free particles due to colour confinement and therefore need to be formed into hadrons; this process is called hadronisation. A given energy scale, dictated by the parton shower program,

is predefined as the point at which this happens. Parton shower evolution and hadronisation are modelled with reference to experimental data.

The underlying event includes all partons not involved in the hard scattering process. This includes particles that come from the break up of the protons after a parton is knocked out of each of the two protons, initial and final state radiation (ISR/FSR) such as the addition of soft radiation and loop processes to the final state matrix element particles and multiple parton interactions.

Event simulations also model how the generated particles in the parton shower interact with the detector. This models both the particle trajectories and the detector's response. A complete simulation of detector material and geometry is available for the ATLAS detector using the GEANT program [27].

3.3 Object Reconstruction

The methods used to reconstruct physics objects in simulation and data are the same and are outlined in this section. The reconstruction methods used for most of the physics objects are the same in both the b-tagging calibration and the $t\bar{t}H(H\rightarrow b\bar{b})$ analysis presented. However, each analysis presented in this thesis was performed at different points during Run I and suggestions for object definitions from the ATLAS combined performance groups changed during this time. Any relevant differences between object reconstruction methods for the two analyses are clearly highlighted in this section.

3.3.1 Jets

High energy quarks and gluons are not observable as free particles. Instead, they create a highly collimated shower of hadrons produced along a similar trajectory as the original parton. These showers are produced by the Quantum Chromodynamics (QCD) phenomenon of fragmentation. The shower of hadrons is observed as a group of topologically related energy deposits in the electromagnetic and hadronic calorimeters which are clustered together to create topological clusters (clusters formed from neighbouring cells with significant signal compared to noise) [28]. These are then used to create an object, called a jet, that is measurable in a detector. On average, roughly 60% of the energy in a jet is in the form of charged particles ($\pi^{\pm}$), 30% is in the form of photons ($\pi^0 \rightarrow \gamma\gamma$) and 10% is in the form of neutral hadrons (e.g. neutrons).

The ATLAS calorimeter is non-compensating which means the detector response for electrons

(EM scale) and hadrons (hadronic scale) is different. Corrections are derived at the EM scale by looking a fully simulated sample of jets and comparing the reconstructed jets with the truth information. These corrections are them applied to the reconstructed jets in data. The response of the detector to jets at the EM scale is defined in Equation 3.1:

$$\mathcal{R}_{EM}^{jet} = E_{EM}^{jet} / E_{truth}^{jet} \tag{3.1}$$

where E_{EM}^{jet} is the energy of the jet at the EM scale. E_{truth}^{jet} is the energy of the truth jet which has been built from stable particles defined to have lifetimes longer than 10 ps excluding muons and neutrinos. Reconstruction inefficiencies and energy deposits outside the calorimeters lower the response of the detector to both EM and hadronic particles in different ways. Calibration schemes based on cell energy density improve the jet energy resolution by weighting energy deposits from EM and hadronic showers differently.

Local Cluster Weighting (LCW)

Local cluster weighting calibration is a jet energy scale correction that locally corrects the 3-dimensional topological clusters (topo-clusters) in the calorimeter before any jets are formed. The calibration procedure is determined by looking at fully simulated MC samples of jets and comparing the reconstructed jets in this sample with the truth information. The corrections derived from this comparison are applied to reconstructed data from collisions.

The first stage in the calibration uses cluster shape variables to characterise clusters of energy deposits as electromagnetic or hadronic. The cluster properties used are the energy density in the cells forming the cluster, the fraction of their energy deposited in the different calorimeter layers, the cluster isolation and its depth in the calorimeter. A more detailed description of the procedure can be found in Reference [29].

All corrections are derived from simulated samples of single charged and neutral pions. A flat distribution in the logarithm of pion energies from 200 MeV up to 2 TeV is used. The corrections are derived with respect to the true deposited energy in the calorimeter. True energy deposits are classified into three categories:

1. **Visible energy**: For example, the energy deposited via ionisation;

2. **Invisible energy**: For example, the energy absorbed in nuclear reactions;

3. **Missing energy**: Energy missing from the jet; for example, the energy carried away by neutrinos.

The local cluster weighting calibration is performed in 4 steps:

1. **Cluster classification**: The expected population in logarithmic bins of topo-cluster energy, the cluster depth in the calorimeter and the average energy density in the cells forming the cluster are used to calculate classification weights. The weights are calculated for small $|\eta|$-regions ($|\eta|$ bins of width 0.2) on a sample created by mixing single neutral and charged pions under the assumption that $2/3$ of pions are charged. This is used to classify clusters as mostly hadronic or mostly electromagnetic where the weight denotes the probability the cluster originated from a hadronic interaction (p_{had}).

2. **Hadronic weighting**: Topological clusters receive calorimeter cell correction weights derived from detailed MC simulations of charged pions. Each cell is weighted according to the topo-cluster energy and the calorimeter cell energy density. The hadronic energy correction weights are calculated with respect to the true energy deposits in the MC simulation. This is then multiplied by the probability p_{had} that the cluster is hadronic. The applied weight is:

$$w_{Hadronic} \cdot p_{had} + w_{EM} \cdot (1 - p_{had}) \tag{3.2}$$

   where $w_{Hadronic}$ is the hadronic energy correction weight and $w_{EM} = 1$. The result is a look-up-table for each calorimeter layer in bins of width 0.2 in $|\eta|$. Tables are binned in logarithmically in topological cluster energy and cell energy density (E/V).

3. **Dead material corrections**: Corrects for energy deposits in materials in regions of the detector outside the calorimeters (e.g. upstream materials like the inner wall of the cryostat). These corrections are η-dependant.

4. **Out-of-cluster corrections**: Corrects for the signal losses from the calorimeter clustering and jet reconstruction. This is for isolated energy deposits inside the calorimeter but outside the topological clusters. These are energy deposits that do not pass the noise thresholds applied during clustering. Out-of-cluster corrections are η-dependant.

Additional corrections are also applied to the LCW calibrated jets. These are derived in 3 steps:

- **Pileup correction**: The average additional energy due to additional proton-proton collisions is subtracted from the jet energy measured in the calorimeter. Correction constants are obtained from *in situ* measurements [29].

- **Vertex correction**: The direction of each jet is corrected so that the pointing direction is toward the primary interaction vertex (as defined in Section 3.3.6) and not the geometric centre of the detector (as is the case when jets are reconstructed).

- **Jet energy and direction correction**: Restores the reconstructed jet energy (the partial measurement of the energy deposited by hadrons due to calorimeter non-compensation to the true energy of the corresponding jet of stable particles (MC truth jet) [29]). The LCW-scale jet energy response is defined as in Equation 3.3:

$$\mathcal{R}_{LCW}^{jet} = E_{LCW}^{jet} / E_{truth}^{jet}. \tag{3.3}$$

A jet response calibration function $F_{LCW+JES}$ is derived as a function of the detector pseudo-rapidity and jet energy from a fit parametrised by the jet energy response and the measured jet energy at the LCW scale after the pileup and vertex corrections, E_{LCW}^{jet}. The final jet energy scale correction is:

$$E_{LCW+JES}^{jet} = E_{LCW}^{jet} / F_{LCW+JES}(E_{LCW}^{jet}) \tag{3.4}$$

After the jet origin and energy corrections, the origin-corrected jet η is further corrected for a bias due to poorly instrumented regions of the calorimeter. This is needed as topological clusters in poorly-instrumented regions of the calorimeter are reconstructed with lower energies with respect to the better-instrument regions. Therefore jet directions are biased towards the better instrumented regions of the detector.

The η-correction is derived from simulation as the average difference $\Delta\eta = \eta_{truth} - \eta_{origin}$, where η_{origin} is the origin-corrected pseudorapidity and η_{truth} is the true pseudorapidity of the jet. An η-correction function is derived in bins of $(E^{truth}, \eta_{detector})$. The corrections described above are then be applied to reconstructed jets in data and MC. Jets that have been corrected using the LCW correction along with the additional JES calibration described here are referred to as LCW+JES jets or just LC jets.

Recombination Scheme

A recombination scheme provides a meaningful single-jet mass using these calibrated topological clusters [30]. These clusters can be considered as Lorentz four-momentum vectors that are summed together to calculate the four-momentum of the final jet object.

Recombination schemes are used iteratively in cone algorithms such as the anti-k_t jet algorithm [31]. When a candidate cone is identified, its momentum vector is calculated using a recombination scheme and the cone is re-defined around the new centre. This is then repeated until it converges on a stable cone (adding soft radiation does not change the cone definition).

The jets used in the b-tagging calibration and the $t\bar{t}H(H\rightarrow b\bar{b})$ analysis presented are built using the anti-k_t algorithm with a radius parameter $R = 0.4$. Jets reconstructed with the anti-k_t algorithm start from energy deposits in the EM and hadronic calorimeters. The anti-k_t method clusters energy deposits from soft particles last, favouring those from high-p_T particles to seed jets, which leads to a more regular cone-like geometry for jets. This is one of the main advantages of the anti-k_t algorithm as it make jet energy calibration and pileup/underlying event subtraction less complicated. Algorithms such as the original k_t [32] and Cambridge-Aachen methods [33] cluster soft collinear energy deposits first and only at the end do they cluster the high-$p_T$particles which can lead to an irregular shape. The response of the calibrated jets at the LCW scale is shown in Figure 3.1.

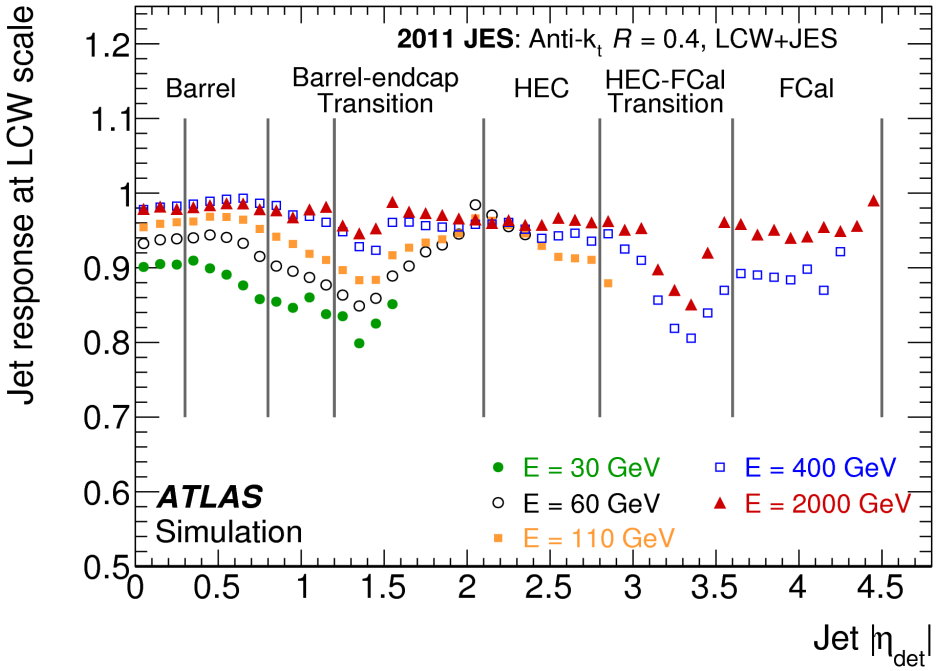


Figure 3.1: Average response of simulated anti-k$_t$ jets formed from topological clusters with a cone of radius 0.4. The response is defined by $\mathcal{R}^{jet}_{LCW} = E^{jet}_{LCW}/E^{jet}_{truth}$. A response of 1 means the truth-jet energy is the same as the energy of the jet after the LCW correction has been applied. Different calorimeter regions are shown, indicated by the vertical lines [29].

Jet and electron candidates are both reconstructed from energy deposits in the calorimeters and during jet reconstruction there is no distinction between the two. It is therefore possible to double-count an energy deposit as both a jet and an electron. To remedy this, if there are any jets that lie within a $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.2$ from an already identified electron the single closest jet is discarded; after this, electrons that have a $\Delta R < 0.4$ are removed. All jets are required to have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$. Those which have $p_{\mathrm{T}} < 50\,\mathrm{GeV}$ and $|\eta| < 2.4$ are required to have a jet vertex fraction (JVF) of above 0.5. The JVF is the fraction of tracks associated with the jet that originate from the primary vertex. If a jet has a JVF>0.5 the jet is considered to have come from the primary vertex and not generated by pileup interactions.

Selection	Jets
Type	anti-k_t LC Jets
p_{T}	$> 25\,\mathrm{GeV}$
$\lvert\eta\rvert$	< 2.5
Isolation	$\Delta R_{jet,e} > 0.2$
Specifics	JVF > 0.5 if jet $p_{\mathrm{T}} < 50\,\mathrm{GeV}$ and $\lvert\eta\rvert < 2.4$

Table 3.1: Standard selection cuts for signal jets used in both analyses.

3.3.2 B-jets

Jets are identified as coming from the hadronisation of a b-quark using the MV1 b-tagging algorithm (see Section 4.1.2) which combines discriminating quantities such as the impact parameters of displaced tracks and the topological properties of secondary and tertiary decay vertices within a jet. The working point used for the analyses presented in this thesis corresponds to a 70% efficiency of correctly tagging a jet that originated from a b-quark. However, the working point used also has a $\approx 1\%$ efficiency of incorrectly tagging a light-jet and a $\approx 20\%$ efficiency of incorrectly tagging a charm-jet, as determined for b-tagged jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $\lvert\eta\rvert < 2.5$ in simulated $\mathrm{t\bar{t}}$ events using the combinatorial likelihood calibration method [34]. In the $\mathrm{t\bar{t}H(H\rightarrow b\bar{b})}$analysis, b-tagging efficiencies in simulation samples are corrected for data by using scale factors derived by the $\mathrm{t\bar{t}}$ based combinatorial likelihood calibration.

3.3.3 Muons

Muons can traverse the entire ATLAS detector without being stopped by any of the intermediate material. For this reason the muon spectrometer is the outermost layer of the detector. Muons are also charged particles and therefore leave a track in the inner detector. Hence, muons can be precisely reconstructed from track segments in the muon spectrometer which are matched with tracks found in the inner detector [35]. The final muon candidates are refitted using complete track information from both parts of the detector. Muons reconstructed this way are know as combined MUID muons.

Isolation requirements are used to separate prompt muons coming from the hard interaction, from non-prompt muons or fake muons caused by e.g. a decaying heavy flavoured particle, charged

hadrons penetrating the MS and from decay in flight of pions/kaons. This also serves to prevent any overlap with the soft muons used in the semileptonic selection described later in Chapter 4. Muon candidates are required to be in the range $|\eta| < 2.5$ and separated from selected jets by $\Delta R > 0.4$. They must also satisfy a p_{T}-dependent track-based isolation requirement. This makes sure the scalar sum (labelled $p_{\mathrm{T}}^{cone\ var}$) of the p_{T} of any tracks in a cone of variable $\Delta R = 10\,\mathrm{GeV}/p_{\mathrm{T}}^{\mu}$ around the muons axis, excluding the muon track itself, is less than 5% of the muon p_{T}. The longitudinal impact parameter ($|z_0|$), which is the point along the beam axis at which the longitudinal distance of the track of the particle from the primary vertex is smallest, is required to be less than 2 mm. Muons are also required to have a hit pattern in the inner detector that is consistent with a well constructed track [36].

Object Selection	Muons
Type	combined MUID muons
p_{T}	> 25 GeV
$\|\eta\|$	< 2.5
Isolation	$p_{\mathrm{T}}^{cone\ var}/p_{\mathrm{T}}^{\mu} < 5\%$ $\Delta R_{\mu,jet} > 0.4$
Specifics	$\|z_0\| < 2\,\mathrm{mm}$

Table 3.2: Standard selection cuts for signal muons in both analyses.

3.3.4 Electrons

Electrons are reconstructed from energy clusters in the electromagnetic calorimeter which are matched to reconstructed tracks in the inner detector. During jet reconstruction, no distinction is made between identified electrons and jets. If any jets have a $\Delta R < 0.2$ from an already identified electron, the single closest jet is discarded, after which electrons that have a $\Delta R < 0.4$ are removed.

Prompt electrons (the electron with the highest p_{T} in the event) are required to have $p_{\mathrm{T}} \geq 25\,\mathrm{GeV}$ (where the trigger is fully efficient) however, in the $\mathrm{t\bar{t}H(H\to b\bar{b})}$ analysis the sub-leading lepton is only required to have $p_{\mathrm{T}} > 15\,\mathrm{GeV}$. Electrons must also have pseudorapidity in the range $0 < |\eta| < 1.37$ or $1.52 < |\eta| < 2.47$. Electrons found in the region $1.37 < |\eta| < 1.52$ are removed as they fall into the calorimetry transition region.

Not all electrons built by the electron reconstruction algorithms are signal electrons. Many

background sources such as jets, in-jet hadron decay and photon induced pair-production can also produce electrons. To reject these while keeping as many signal electrons as possible, a list of selection cuts on variables that discriminate signal electrons from background sources is applied. In the b-tagging calibration analysis, the tight++ cut-based algorithm is used [37].

Further isolation requirements are used to reduce the number of electrons originating from hadron decays. The sum p_{T} of the tracks within a cone of $\Delta R = 0.3$ (p_{T}^{cone30}), not including the electron track itself, must satisfy $p_{\mathrm{T}}^{cone30}/p_{\mathrm{T}}^{e} \leq 4\%$. Furthermore, the sum of the energy that does not come from the electron itself but is deposited inside the calorimeter within a cone of radius $\Delta R = 0.2$ from the electron direction (E_{T}^{cone20}), must satisfy $E_{\mathrm{T}}^{cone20}/\ E_{\mathrm{T}}^{e} \leq 7\%$.

In the $\mathrm{t\bar{t}H(H\rightarrow b\bar{b})}$ analysis, the selection of electrons was optimised to increase the acceptance since the fake rate is quite small. Instead of the tight++ cut based identification, the medium working point of the likelihood-based identification is used. Relaxed isolation requirements mean the ratio of the sum of the p_{T} of the tracks excluding the electron track, with respect to the p_{T} of the electron track itself must satisfy $p_{\mathrm{T}}^{cone30}/p_{\mathrm{T}}^{e} < 0.12$.

An additional quality requirement applied in both analyses means that an electron's longitudinal impact parameter (z_0) must be < 2 mm. This is used to help distinguish electrons from converted photons.

Standard Electron Selections		
Selection	B-tagging Calibration Cuts	$\mathrm{t\bar{t}H(H\rightarrow b\bar{b})}$ Analysis Cuts
Identification	cut-based tight++	Likelihood-based medium
p_{T}	> 25 GeV	> 25 GeV (subleading > 15 GeV)
$\lvert\eta\rvert$	$0 < \lvert\eta\rvert < 1.37$ or $1.52 < \lvert\eta\rvert < 2.47$	$0 < \lvert\eta\rvert < 1.37$ or $1.52 < \lvert\eta\rvert < 2.47$
Isolation	p_{T}^{cone30}Isolation efficiency = 90% $p_{\mathrm{T}}^{cone30}/\ p_{\mathrm{T}}^{e} \leq 4\%$ E_{T}^{cone30}isolation efficiency = 90% $E_{\mathrm{T}}^{cone20}/\ E_{\mathrm{T}}^{e} \leq 7\%$, $\Delta R_{e,jet} > 0.4$	$p_{\mathrm{T}}^{cone30}/p_{\mathrm{T}}^{e} < 0.12$ $\Delta R_{e,jet} > 0.4$
Trigger p_{T} threshold [GeV]	24 or 60	24 or 60
Impact Parameters	$z_0 < 2$ mm	$z_0 < 2$ mm

Table 3.3: Standard selection cuts for final state electrons used in calibration.

3.3.5 Photons

Photons are neutral, electromagnetically-interacting particles. They leave no track in the inner detector and create a shower of electromagnetic particles in the EM calorimeter. Clusters where no track is matched are classified as unconverted photons. Photons can also convert to an electron-positron pair before they reach the EM calorimeter through interactions with the nuclei in the detector material. These are known as converted photons and are characterised by the presence of at least one track matching an EM cluster originating from a conversion vertex inside the tracker. This results in an ambiguity between electrons and converted photons as both candidates have a track matched with an EM calorimeter cluster. However, if a cluster in the second sampling layer of the EM calorimeter is matched to a track that does not originate from a vertex found in the beam interaction region but originates from a photon conversion vertex, the candidate is reconstructed as a converted photon, otherwise it is reconstructed as an electron.

In-jet processes such as $\pi^0 \rightarrow \gamma\gamma$ result in the deposition of a substantial fraction of their energy in the EM calorimeter and can mimic the signature of a prompt photon. The key features used for distinguishing between jets and prompt photons are: hadronic leakage, lateral showering and shower sub-structure. Hadronic leakage is a measure of the fraction of energy deposited in the hadronic calorimeter compared to the EM calorimeter. This fraction is higher for hadronic objects such as jets when compared with electromagnetic objects such as prompt photons.

Lateral showering is a measure of the shower width in η-ϕ space. The reason for measuring this is that a prompt photon should produce a narrower cluster compared to a jet. The calorimeter can resolve the lateral showering differences between prompt photons and jets, providing a powerful discriminating variable.

Substructure in showers is used to distinguish prompt photons from jets in which a neutral meson has decayed to two photons. Because the meson decay produces two photons that are highly collimated, the showers can appear to have originated from a single photon. However, ultra-fine strips in the 1^{st} layer of the EM calorimeter are used to obtain a finer granularity and thus can resolve each photon individually.

3.3.6 Track and Vertex Reconstruction

Track reconstruction is the procedure in which the trajectory of a particle inside the detector is reconstructed. It is based on several algorithms and uses measurements made by the inner detec-

tor. Points in $\eta - \phi - z$ space are formed from clusters of neighbouring silicon pixels in the ID. Seeds which make the track candidates are formed by combining three points. These seeds are then extrapolated into tracks by including measurements from the subsequent layers of the ID and refitting its trajectory. Track candidates are extended out to the TRT which reduces the error on the trajectory. Tracks which don't meet certain criteria (e.g. track candidates with $p_{\mathrm{T}} < 400\,\mathrm{MeV}$) are rejected [38].

Various vertex topologies can be produced when colliding two bunches of protons in the LHC. A typical bunch-crossing will result in several primary vertices along the beam axis each corresponding to a separate proton-proton collision, vertices from the decay of heavy long lived particles, vertices in jets and vertices from photon conversions. Each type of vertex has its own distinguishing features which can be used to aid their reconstruction.

Identifying the primary vertex created by the hard scattering process amongst many other proton-proton collisions is particularly important in the environment created in collisions at the LHC. Interaction vertices are reconstructed by first finding a selection of tracks that are likely to come from the interaction region. A single primary vertex is then fitted combining all of the selected tracks. Tracks that are considered to be outliers are disassociated from this vertex but are used to seed a new competing vertex. As the number of candidates grows, candidates compete with one another to gain more tracks and the method is iterated. At the end of the procedure, the vertex with the highest summed track p_{T} is considered to be the primary vertex corresponding to the hard scatter.

The ATLAS detector can also distinguish b-jets from jets originating from the decays of light and/or charm hadrons. This requires the reconstruction of the vertex where the b-hadron decayed (the secondary vertex). The high track density in the detector means a dedicated vertex finding algorithm is used to find secondary vertices. Typically, the decay vertex of a b-hadron is significantly displaced from the primary vertex. Displaced tracks are selected and used to construct an single inclusive secondary vertex much in the same way the primary vertex is reconstructed. A similar procedure is used to remove outlier tracks and reconstruct the secondary vertex. Specific b-tagging algorithms are discussed at greater length in Chapter 4.

Chapter 4

A measurement of the performance of the ATLAS multivariate b-tagging algorithm on b-jets with muons using top-quark pair events

> *Not everything that counts can be counted, and not everything that can be counted counts.*
>
> Albert Einstein

The number of true b-jets an algorithm tags with respect to the total number of true b-jets in the sample is know as the algorithms tagging efficiency. Similarly, the algorithms mistag efficiency is the number of non b-jets that are tagged with respect to the total number of non b-jets in the sample. The tagging efficiencies and mistag efficiencies for a given working point (see Section 4.1.2) of a b-tagging algorithm can be different in data and simulation.

Working points must be calibrated which means calculating a scale factor that corrects for a possible discrepancy in efficiency between data and simulation. The calibration of any b-tagging method can be performed in several ways. The calibration method presented here is the $t\bar{t}$ dilepton kinematic selection method (Section 4.7) which calibrates working points using an enriched b-jet sample by selecting $t\bar{t}$ dilepton events by their kinematic properties. The working point calibrated

here corresponds to an efficiency of 70%.

The $t\bar{t}$ dilepton kinematic selection method performs a calibration on a high purity, inclusive sample of b-jets. However other calibration procedures such as p_{T}^{Rel} (see Section 4.5) use a subset of b-jets, where the b-jet decays to a muon plus neutrino. Low momentum in-jet muons (soft muons) are used in many b-tagging calibrations because they have a relatively high reconstruction efficiency (compared to for instance, electrons) and provide a distinguishable signature. However, muons deposit very little energy in the calorimeters and neutrinos traverse the detector without depositing any energy. Unlike jets decaying semileptonically to an electron and neutrino and hadronically decaying jets, where the electrons/hadrons will deposit most of their energy in the calorimeters, jets which decay to a muon can have a largely different p_{T} scale, which affects the efficiency of b-tagging algorithms. Hence, it is possible that a calibration performed on b-jets decaying to a muon plus neutrino could yield different results.

The aim of this analysis is to compare two identical calibrations on two orthogonal subsets of b-jets to see if such a bias is introduced. The initial sample of b-jets is obtained using the $t\bar{t}$ dilepton kinematic selection method. Subsequently, the initial sample is split into two subsamples. A subsample of b-jets with an associated soft in-jet muon are selected using the p_{T}^{Rel} calibration selection method (described in Section 4.5). These jets are labelled "SL jets". The remaining jets subsample is labelled "NSL jets". B-jets that decay to an electron plus neutrino are included in the NSL subset.

4.1 Introduction to b-tagging

The identification of b-jets (b-tagging) is essential to many physics analyses. It has proven to be especially useful in the search for the Higgs boson and the top-quark sector as it provides a way of selecting signal events containing b-jets and reducing background processes which often have light and charm quarks in their final state.

Several properties distinguish b-jets from other jets. Firstly, when a b-quark fragments, the b-jet retains a large fraction of the b-jets original momentum (on average $\approx 70\%$). This means the transverse momentum of the b-jet can often be distinguished from those of lighter jets, which create a softer p_{T} spectrum. Secondly, the mass of the b-hadron is quite large ($> 5\,\mathrm{GeV}$) which means its decay products tend to have a large transverse energy with respect to the jet axis.

Furthermore, the $|\mathrm{V}_{c,b}|$ and $|\mathrm{V}_{u,b}|$ vertices in the CKM matrix are off-diagonal elements. Their

small values mean that these quark transitions are very suppressed. When a B meson decays, the amplitude calculation for the process will include one of these vertex terms which prolongs the average lifetime of b-mesons to around $O(10^{-12})$ s [5]. This means they can travel a distance of $c\tau = O(mm)$ from the point at which the protons in the beam collide (the primary vertex). This results in a decay vertex (secondary vertex) inside the beam pipe, which can be reconstructed. Special algorithms are used to reconstruct the secondary vertex and associate tracks with it. Many variables used to distinguish b-jets use characteristics of the secondary vertex and the tracks it is associated with. A schematic of a b-hadron decay is illustrated in Figure 4.1.

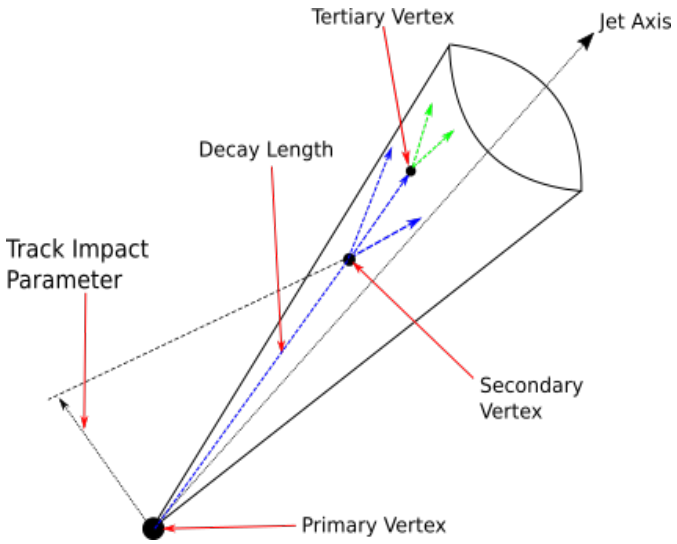


Figure 4.1: An example of the topology of a decay within a jet. In this figure primary and secondary vertices are shown along with the impact parameter, jet axis and decay length. Two impact parameter variables can be constructed: the transverse and the longitudinal parameters.

One can look for a lepton inside the jet which comes from the semileptonic decay of the b-hadron. Due to the hard fragmentation and high mass of the b-hadron, these leptons have a large transverse momentum in the detector and a large transverse momentum relative to the jet axis. The B mesons exclusive semileptonic branching ratio to leptons (electrons and muons) is $\approx 10\%$. A breakdown of the B meson branching fraction to each of the leptons is shown in Table 4.1. The calibration method described in Section 4.5 only uses b-jets in which the B meson has decayed muonically. b-jet in which the B mesons has decayed to a tau lepton or electron are not considered in the semileptonic sample of jets.

Branching ratios	
$B \to X + e + \nu_e$	$10.86 \pm 0.35\%$
$B \to X + \mu + \nu_\mu$	$10.95^{+0.29\%}_{-0.25\%}$
$B \to X + \tau + \nu_\tau$	$2.48 \pm 0.26\%$

Table 4.1: Semileptonic branching ratios of B-hadrons [5]

The two impact parameters including the one shown in Figure 4.1 are key in many b-tagging algorithms. The transverse impact parameter, d_0, is defined as the distance of closest approach between the extrapolated track and the primary vertex in the *x*-*y* plane. This impact parameter is signed according to the charge of the particle and its angular momentum about the beam line. The longitudinal impact parameter, z_0, is the longitudinal distance at the point on the extrapolated track which defines the value of d_0. The sign of this impact parameter depends on the direction of z_0 relative to the primary vertex. Impact parameters tend to be relatively large for track stemming from displaced vertices, whereas track that originate from the primary vertex have much smaller impact parameters.

4.1.1 B-tagging Algorithms

In this section three algorithms with three different approaches to b-tagging are discussed: SV1, IP3D and JetFitter. All three of these algorithms were calibrated on samples of di-jet events. There are alternative algorithms used by the ATLAS collaboration but only these three are discussed here as their outputs are used as inputs to the multivariate classifier MV1 which is used for this study [39].

The SV1 algorithm reconstructs secondary vertices using the tracks left in the inner detector by the decay products of the b-hadron and the subsequent c-hadron decay [40]. A common first step for all b-tagging algorithms is to reject tracks associated with V^0 decay vertices because they mimic tracks formed by b-hadrons. V^0 is the commonly used name for two-pronged vertices coming from the decay of long-lived particles (which in this case does not include b-hadrons). Some of the more common causes for V^0 vertices are: $\Lambda^0 \to p^+\pi^-$, $K_S \to \pi^+\pi^-$, $K_L \to \pi^+\pi^-\pi^-$. Preselection cuts [40] (e.g. on impact parameter variables) are used to remove V^0 vertices along with vertices from other long-lived particles and secondary interactions from non b-hadrons. Once all tracks from V^0 candidates are removed, all two track pairs that form a good vertex are built, using only tracks associated with the jet and are far enough away from the primary vertex. All

tracks are then combined into an inclusive vertex as shown in Figure 4.2. The worst-matching tracks are removed one-by-one until a good vertex fit is found.

Once the secondary vertex has been reconstructed, the SV1 algorithm uses many of its features to obtain the likelihood the jet is a b-jet. The algorithm produces a likelihood distribution of jets. The discriminating variables used by SV1 are:

1. the significance of the decay length projected onto the jet axis;
2. the invariant mass of all tracks associated with the secondary vertex;
3. the ratio sum of the energies of tracks associated to the vertex with respect to all tracks in the jet;
4. the number of two-track vertices;
5. the ΔR between the jet axis and the line joining the secondary and primary vertices (see Figure 4.1).

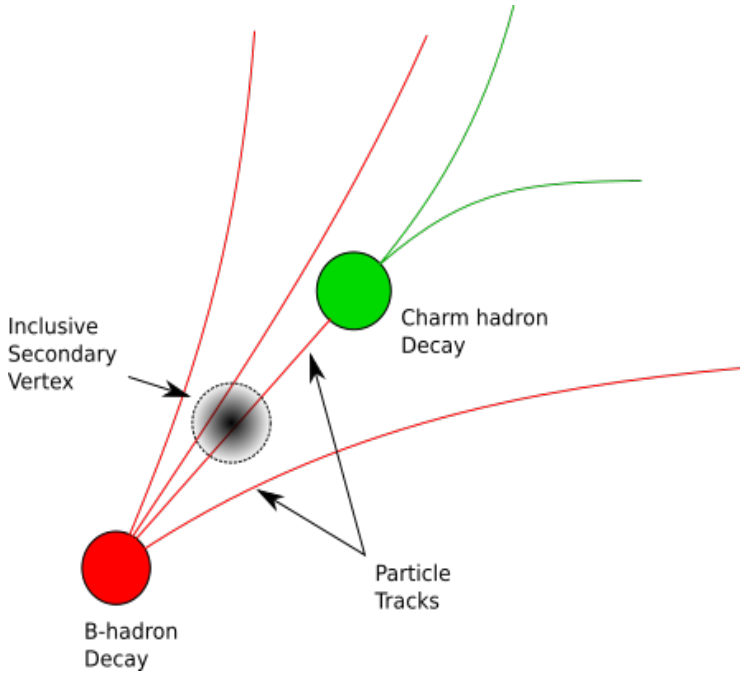


Figure 4.2: A schematic of the method used by the SV1 algorithm for reconstructing a secondary vertex. All tracks from the b-hadron and (typically) c-hadron decay vertices are fit to a single inclusive vertex before removing the outliers.

The IP3D algorithm uses a likelihood ratio technique in which impact parameter variables from candidate tracks are compared to pre-defined distributions for b- and light-jets. It then combines the log-likelihood of the track impact parameters for a jet to obtain the likelihood that the jet is a b-jet [41].

The JetFitter algorithm has an improved method of reconstructing the secondary vertex because the underlying hypothesis of a single geometrical (inclusive) vertex used by the other algorithms is flawed. When the distance between the b- and c-hadron decay vertices is significant compared to the vertex resolution in the direction of flight, the tracks of one of the two vertices can be lost in the fit. The JetFitter algorithm exploits the topology of weak decays of b- and c-hadrons inside jets. The algorithm works on the assumption that the b- and c-hadron decay vertices lie on the same line as the b-hadron flight path. This is a reasonable assumption as the hard fragmentation of the b-hadron means the lateral displacement of the c-hadrons decay vertex from the b-hadron flight path is small enough not to violate the basic assumption within the typical resolution of the tracking detector. This information is used to approximate the decay length of the b-hadron. A Kalman filter is used to find the line that intersects the primary, secondary and tertiary vertices, where the tertiary vertex is from the subsequent decay of a c-hadron. The position of each vertex along this line is also found. The discrimination between b-, c- and light-jets is then done by constructing a combined likelihood from variables similar to those used by the SV1 algorithm, along with decay length significances. A high-performance b-tagging algorithm JetFitterCombNN uses the output from the JetFitter algorithm and combines it with the output from the IP3D algorithm by means of a neural network [39, 41]. For a single jet, both JetFitter and JetFitterCombNN output the probability of the a jet being a b-jet.

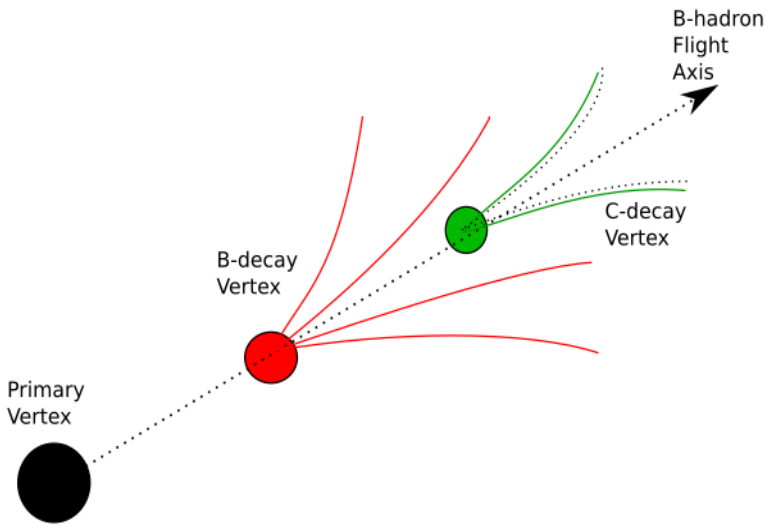


Figure 4.3: A schematic of the JetFitter multi-vertex fit using the b-hadron flight direction constraint.

4.1.2 ATLAS MV1 b-tagging algorithm

The ATLAS MV1 b-tagging algorithm uses information from several other b-tagging algorithms as inputs to a neural network (NN). In this case the inputs are the output values of the SV1, IP3D and JetFitterCombNN algorithms. The NN is trained using simulated samples of b- and light-jets from a mixture of $t\bar{t}$ and dijet events. The network is trained using b-jets as signal and light-jets as background. The algorithm can then be used to compute the likelihood (tag weight) any jet is a b-jet. Once trained, the NN is used on simulation and data to produce two tag-weight distributions. The MV1 tag weight distribution for simulated events is shown in Figure 4.4. A cut is then applied to the tag weight distributions. This is the value that defines whether or not a jet is tagged as a b-jet. Fixed cuts or 'working points', are tuned to yield a specified b-jet efficiency. The efficiency is defined as the ratio of correctly b-tagged jets with respect to the total number of true b-jets in the sample. A specific cut on the MV1 weight distribution will also incorrectly assign light- and charm-jets as b-jets (mistag) with a particular efficiency. This mistag efficiency is calculated as the number of true non b-jets tagged as b-jets, with respect to the total number of true non b-jets in the sample.

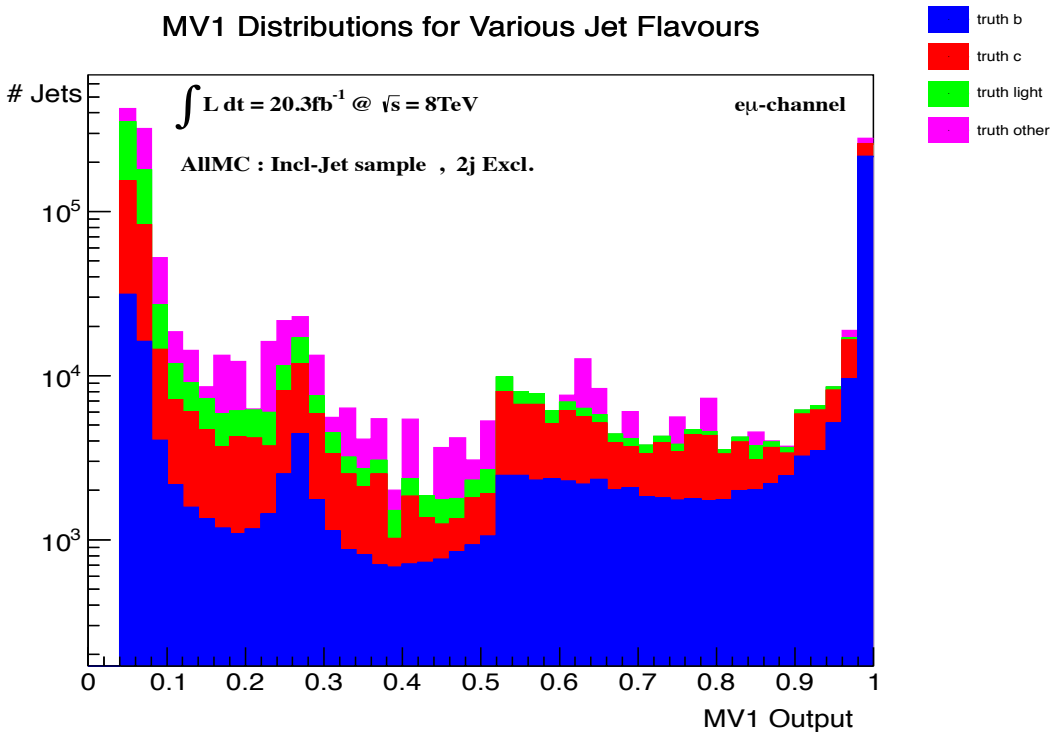


Figure 4.4: Distribution of MV1 weights for light jets, charm jets, b jets and all other jets (e.g. τ-jets) for an inclusive collection of leading and subleading jets for events in all simulated samples before any selection has been applied. The events selected are dilepton events where one lepton is a muon and the other is an electron, namely eμ events.

4.2 Simulation Samples

The main backgrounds for the $t\bar{t}$ dilepton process are $Z+$jets, diboson, single top Wt-production and fake dilepton events. Simulated samples have been generated for each of the major background processes. The hard process for the $t\bar{t}$ signal sample is generated using POWHEG [42] which uses cross-sections calculated at next-to-leading-order (NLO) accuracy to normalise the number of predicted events. A Feynman diagram for the $t\bar{t}$ hard interaction is shown in Figure 4.5. For these samples, the parton shower was modelled using PYTHIA [43] which evolves the simulation from the NLO calculation of the hard process, to a complicated multi-hadronic final state.

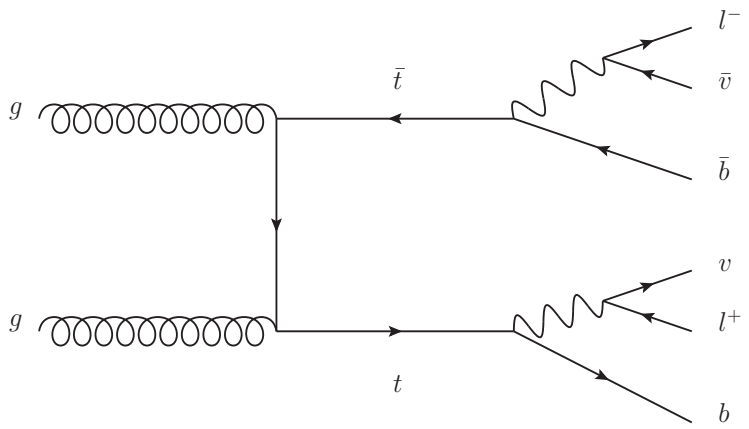


Figure 4.5: A Feynman diagram of a $t\bar{t}$ dilepton event. As a fraction of the number of times a top-quark decays via a Wq process, Wb decays make up 97% of the interactions $\left(\text{R=B}(t \to Wb)/\text{B}(t \to Wq) = 0.97 + 0.09 - 0.08\ [44]\right)$.

One of the most important background processes comes from Z + jets events. For this process, ALPGEN [45] was used to simulate the hard interaction in which the cross-sections are calculated at leading order (LO) accuracy. A Feynman diagram of the Z + jets process is shown in Figure 4.6. The evolution and hadronisation of the parton shower is performed by the HERWIG package [46].

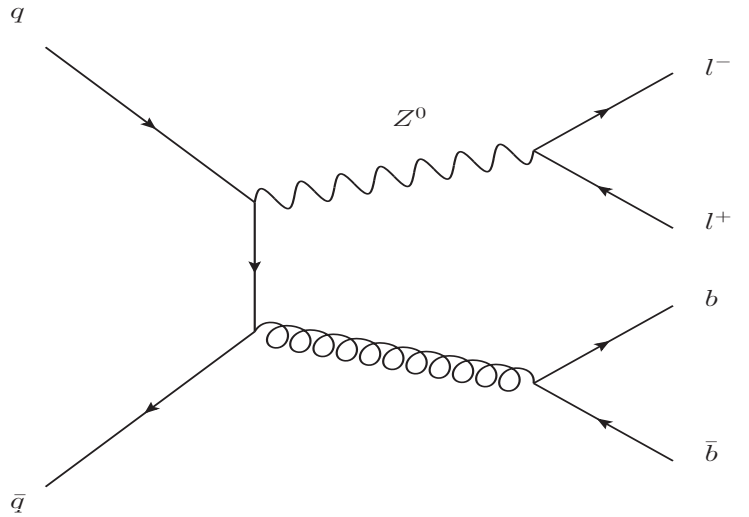


Figure 4.6: Feynman diagram of the Z + jets background process.

A smaller contribution to the background comes from diboson events which were generated for ZZ, WZ and WW events separately. Feynman diagrams of the diboson processes simulated

are shown in Figure 4.7. The ZZ samples simulate the inclusive decay of one of the Z bosons with the other decaying to a pair of leptons ($ZZ \rightarrow \text{incl.} + \ell\ell$). Similarly, the WZ samples were generated with the W boson decaying inclusively and the Z boson decaying to a pair of leptons ($WZ \rightarrow \text{incl.} + \ell\ell$). The WW samples were generated for decays in which one W decays semileptonically and the other decays hadronically ($WW \rightarrow \ell\nu + qq'$) or both decay semileptonically ($WW \rightarrow \ell\nu + \ell\nu$). This is the dominant diboson process in the eμ channel and for these events to be categorised as dilepton events, it requires two jets to be produced in the hadronisation stage of the event generation. For this reason all diboson events were simulated using ALPGEN as it is a multi-leg leading order generator which enables the calculation of the matrix element with additional hard partons. HERWIG was used to model the parton shower in this sample.

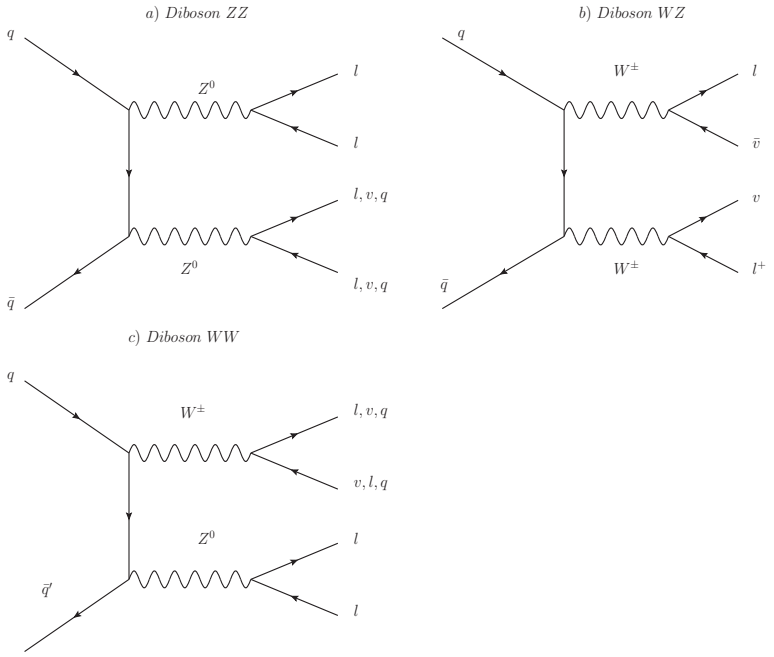


Figure 4.7: Feynman diagrams for the WW, WZ, and ZZ diboson background processes.

Single top events were generated in three channels separately: s-channel, t-channel and the Wt process. All three channels are depicted in Figure 4.8. The t-channel has the largest cross-section contributing around 75% of the inclusive LO cross-section, with Wt-channel representing the next highest fraction and the $s-$channel representing the smallest fraction. However the Wt-channel is the process that provides the largest background when selecting $\bar{\text{tt}}$ events in this measurement.

The t-channel events were simulated using ACERMC [47] to calculate the matrix element of

the hard interaction at NLO accuracy. PYTHIA was used to model the subsequent parton shower appropriately. The *s*-channel and Wt process were generated using the POWHEG package to calculate the matrix element at NLO accuracy using PYTHIA to model the parton shower. Nominal samples were generated using the diagram removal technique (DR) [48] which removes $t\bar{t}$ diagrams that interfere with the calculation of the Wt cross-section above LO. An alternative technique called diagram subtraction (DS) [48] was implemented when simulating additional single top samples. These samples are not used in the nominal analysis but are used to estimate an additional systematic uncertainty associated with the chosen method for removal of interference diagrams.

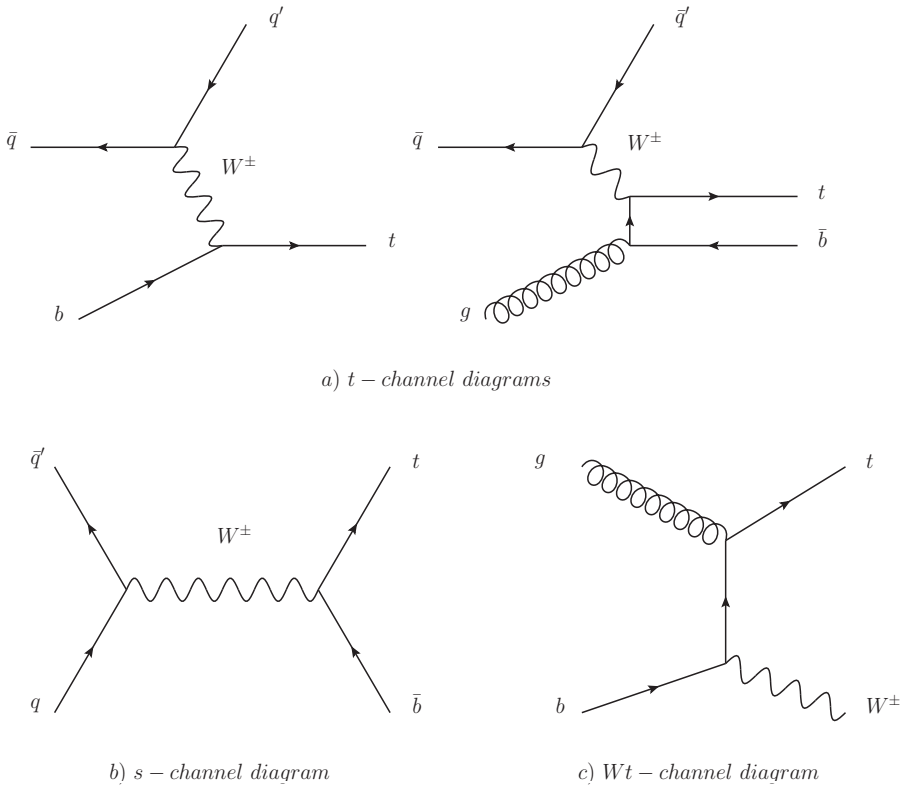


Figure 4.8: Feynman diagrams of the single top background processes: a) t-channel, b) s-channel and c) Wt-channel.

Although the $t\bar{t}$ dilepton signature is realtively clean, fake events still contaminate the channel. Additional events that will be selected in data are found to come from events which contain either an electron-like jet or a non-prompt lepton (decayed from a heavy-flavour hadron or pion/Kaon) which are reconstructed as prompt electrons or muons (prompt meaning charged isolated leptons coming from a W or Z boson decay). These are known as fake events and they primarily come

from processes such as W+jets, single lepton $t\bar{t}$ events or single top s- or t-channels.

The true prompt leptons from the W boson decays in a $t\bar{t}$ event are oppositely charged, which motivates the selection of events with two oppositely charged leptons. However, fake events that have oppositely charged leptons can pass the calibration selection. Some of the processes volunteering such events can be diffcult or are inefficient to model with simulated data and so data-driven techniques are used.

A same-sign fakes estimate was used in this calibration which works on the assumption that events with oppositely charged fake leptons should have the same distributions as the events with same charge fake leptons (same-sign events). These events are selected from data. Truth matching of leptons is done in the simulated samples which ensures events are opposite sign and therefore prevent any double-counting of fake events from simulation and data.

Simulated samples used in this study are generated using cross-sections calculated at LO or NLO accuracy. The best available total cross-section calculations at the time of this analysis were calculated at next-to-next-to-leading-order (NNLO) accuracy. A ratio of cross-sections (k-factor) for NNLO with respect to NLO was used to scale the weight of the event. All simulated samples are scaled using the best available theoretical cross-sections and k-factors accordingly. The cross-sections and k-factors used in this analysis can be found in Tables A.1 and A.2. The simulated samples were normalised to the integrated luminosity of the data used for this analysis.

4.3 Data Samples

This analysis was performed on 14.34 $\text{fb}^{-1}$of data collected using the ATLAS detector from proton-proton collisions at 8 TeV centre-of-mass energy. Data for this analysis was taken from the "MuonsEgammaMerged" stream which means that the events have either passed the single electron trigger, the single muon trigger chain or both requirements. Both of these trigger chains are fully efficient for leptons with $p_{\text{T}} \geq 25$GeV. In the case where an event activates both triggers, the analysis technique used is to take the event from the Egamma stream and disregard it from the Muons stream to avoid double counting eμ events.

4.4 Event Selection

The dilepton kinematic selection method obtains a sample of jets enriched in b-jets on which to perform the calibration. The selection method used here provides an inclusive sample of b-jets which is particularly important as this prevents any bias being introduced as a result of using a subsample of b-jets, as is the case in some calibrations. The selection looks for dilepton $t\bar{t}$ events as depicted in Figure 4.5 to enrich the jet sample with b-jets due to the top-quarks large b-quark branching ratio, $R = B(t \to Wb)/B(t \to Wq) = 0.97$ [44]. To obtain the highest b-jet purity possible, the event selection implements a series of cuts on all events which reduces the number of events coming from background processes and therefore avoids contaminating the calibration jets sample with large numbers of non b-jets. The event selection implemented in this analysis is shown in Table 4.2.

The event level selection outlined in Table 4.2 ensures a high purity of good top-pair dilepton events are selected. A truth record selection cut is applied to all simulated samples. Part of the truth information stored is the type of particle that was simulated which can be used to ensure only true dilepton events are used. This prevents double-counting fake events when using the data-driven fake estimation.

Unbalanced energy deposits in the detector can be caused by cosmic ray events which occur when extra-terrestrial high energy particles collide with nuclei in the upper atmosphere producing pions. These decay to muons which pass through the detector and can be reconstructed as jets. Such events are removed by the selection.

In real data, if an event passed both trigger chains it could theoretically exist as two identical events in the "EgammaMuonsMerged" data stream. To avoid double counting these events, data is taken from "Egamma" stream unless the single electron trigger was not fired. In this case, if the muon trigger was fired, the event will be taken from the "Muons" stream or else it is discarded. ATLAS triggers are not 100% efficient, so each event is reweighted to correct for the detectors response. Weights are derived from data-driven techniques to correct for the trigger inefficiency. Because leptonic triggers were used in this calibration, at least one lepton is required to be trigger-matched. This means the lepton reconstructed offline and used in the calibration is the trigger object that fired the leptonic trigger.

The H_T variable is the scalar sum of the transverse momentum of all isolated leptons and jets in the final state. Selecting eμ events with $H_T \geq 130$ GeV rejects $Z \to \tau\tau$+jets background events

which are expected to have a lower H_{T} . In the ee and $\mu\mu$ channels the H_{T} requirement is not implemented but instead the missing energy is required to be greater than 60 GeV to reduce the amount of Z + jets background. Events are required to have at least 2 jets and exactly 2 oppositely charged leptons, where the lepton requirement removes the contribution of same-sign fake events. Requiring a dilepton invariant mass in excess of 15 GeV removes low dilepton invariant mass processes such as $\mathrm{J}/\Psi \to \mathrm{ee}/\mu\mu$. Furthermore, for ee or $\mu\mu$ dilepton events, the Z + jets background is suppressed by removing events where the dilepton invariant mass lies within ± 10 GeV of the mass of a Z-boson, $M_z = 91$ GeV.

Jet cleaning rejects events with jets that typically have a problem with the calorimeter measurement, otherwise known as "bad jets". This often manifests itself as a large EM energy fraction or large energy deposit in a single calorimeter layer. Such behaviour is caused by large non-physical energy spikes or cosmic muon interactions in the detector.

A clear discrepancy between data and simulation was seen in the dilepton invariant mass spectrum of the ee- and $\mu\mu$-channels (see figures 4.9 and 4.12). An excess of predicted events is seen in the Z-peak where the Z + jets sample is dominant. The ee- and $\mu\mu$-channels would also usually require a cut on missing transverse energy ($E_T^{miss} \geq 60$ GeV), which introduces a non-zero slope in the bin-by-bin data/MC ratio of the dilepton invariant mass distribution. A reweighting was derived to correct the Z p_{T} spectrum to data before the E_T^{miss} cut was implemented but this did not remove the mismodelling of the Z + jets E_T^{miss} spectrum. The dilepton invariant mass distributions after applying the E_T^{miss} cut and the Z p_{T} reweighting are shown in Figures 4.11 and 4.14. As a result, the final calibration is performed using events only from the eμ-channel as the Z + jets sample does not have a large contribution in this channel (Z boson does not decay to $e\mu$) and is therefore less sensitive to any such mismodellings. This is shown by the distributions in Figure 4.15 and Figure 4.18. This does not have a large impact on the calibration as the eμ-channel has higher statistics and a higher b-jet purity as there is less background contamination.

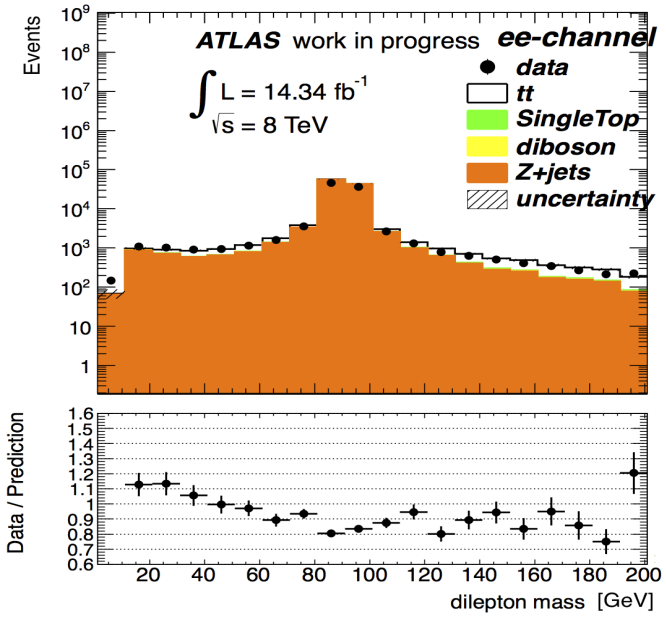


Figure 4.9: The dilepton invariant mass distribution for the ee-channel before E_T^{miss} cut and before the Z p_{T} reweighting. All simulated samples are normalised to an integrated luminosity of 14.34 fb^{-1}.

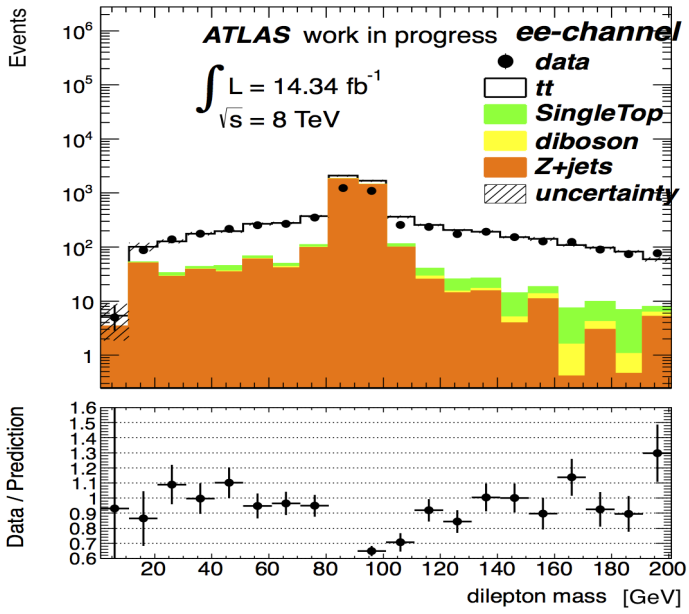


Figure 4.10: The dilepton invariant mass distribution for the ee-channel after the E_T^{miss} cut and before the Z p_{T} reweighting. All simulated samples are normalised to an integrated luminosity of 14.34 fb^{-1}.

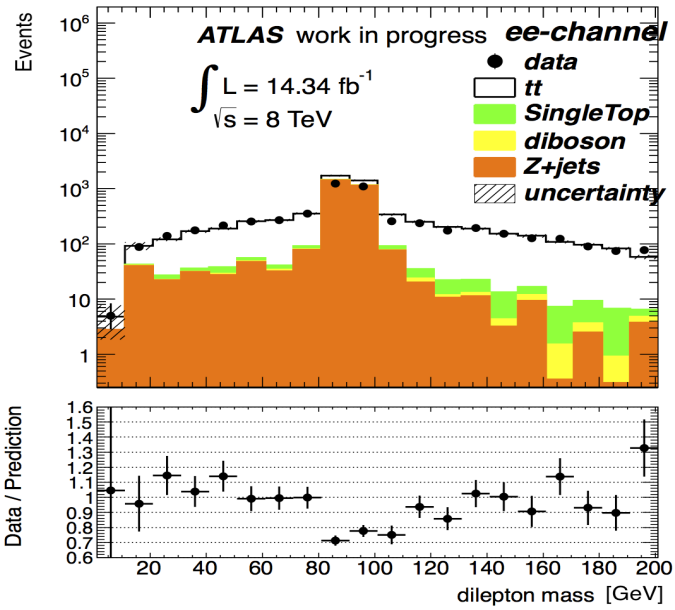


Figure 4.11: The dilepton invariant mass distribution for the ee-channel after the E_T^{miss} cut and after the Z p_T reweighting. All simulated samples are normalised to an integrated luminosity of 14.34 fb^{-1}.

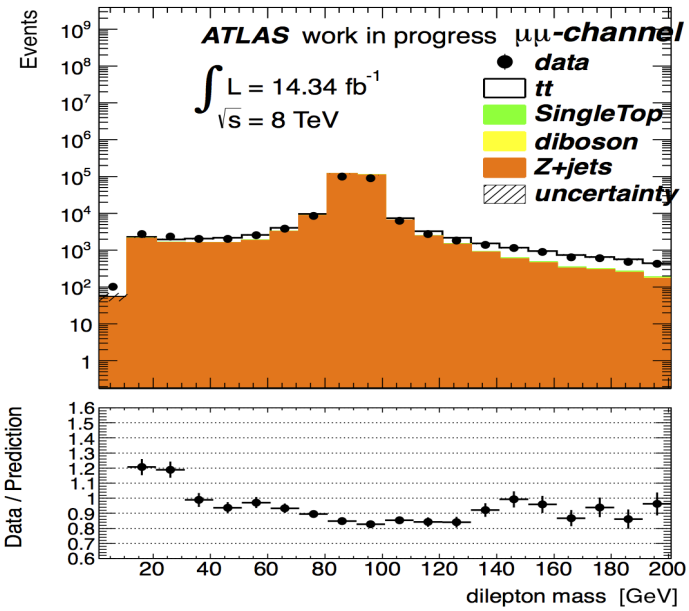


Figure 4.12: The dilepton invariant mass distribution for the $\mu\mu$-channel before E_T^{miss} cut and before the Z p_T reweighting. All simulated samples are normalised to an integrated luminosity of 14.34 fb^{-1}.

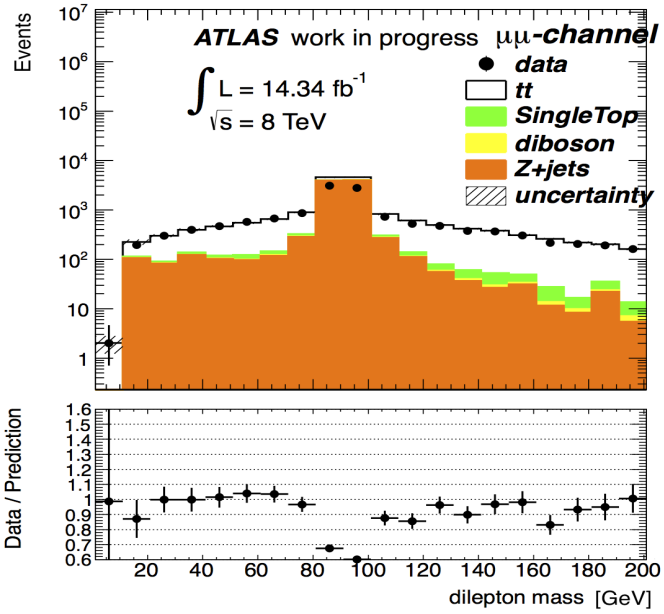


Figure 4.13: The dilepton invariant mass distribution for the $\mu\mu$-channel after the E_T^{miss} cut and before the Z p_{T} reweighting. All simulated samples are normalised to an integrated luminosity of 14.34 fb^{-1}.

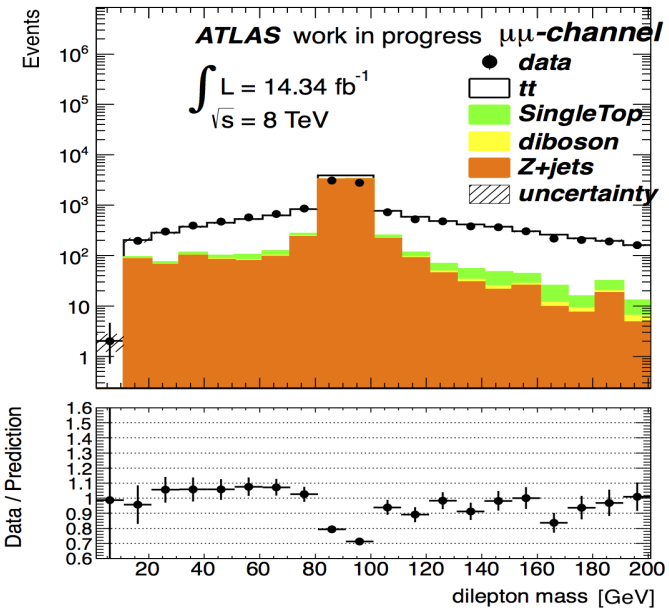
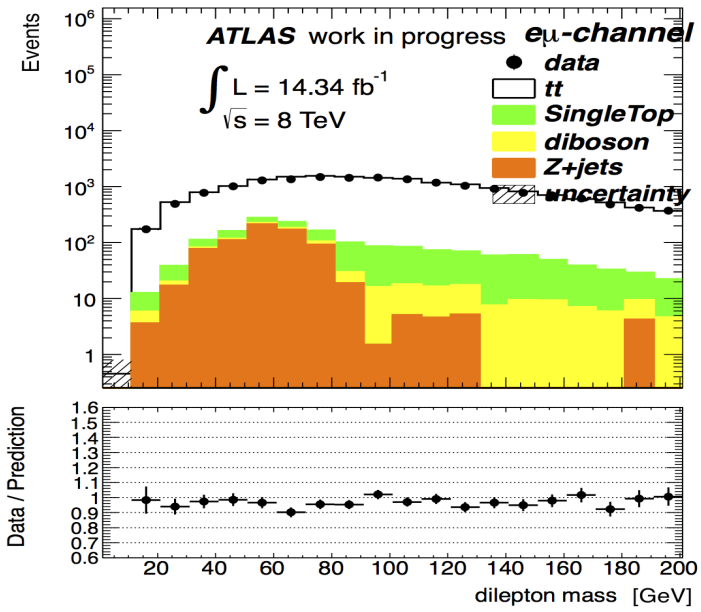
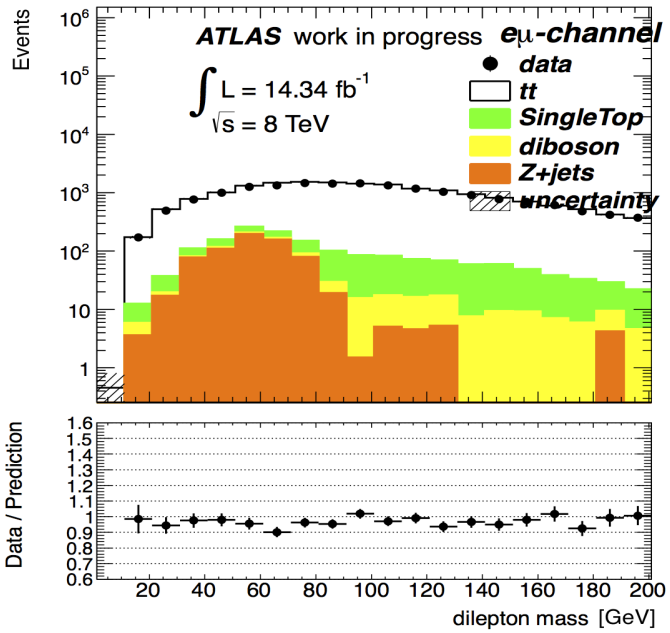


Figure 4.14: The dilepton invariant mass distribution for the $\mu\mu$-channel after the E_T^{miss} cut and after the Z p_{T} reweighting have been applied. All simulated samples are normalised to an integrated luminosity of 14.34 fb^{-1}.



(a)



(b)

Figure 4.15: The dilepton invariant mass distribution for the eμ-channel (a) before and (b) after the Z p_{T} reweighting have been applied. All simulated samples are normalised to an integrated luminosity of 14.34 fb^{-1}.

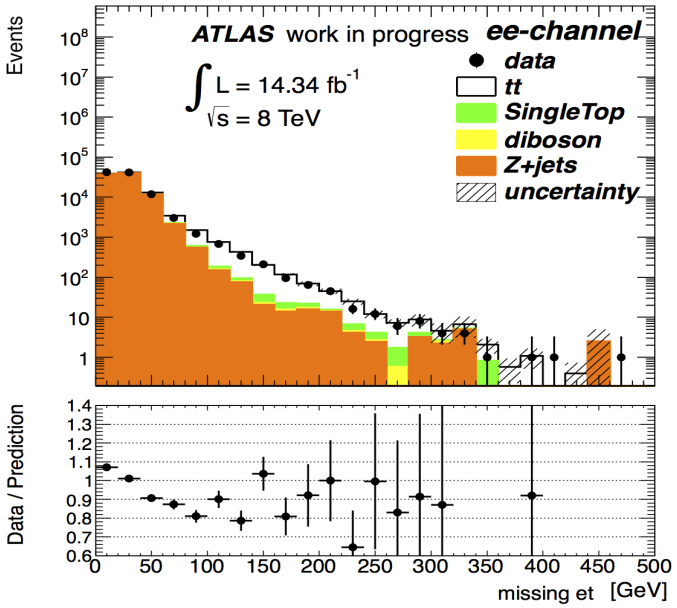


Figure 4.16: The E_T^{miss} distribution for the ee-channel. No E_T^{miss} cut, no Z-mass veto and no Z p_T reweighting have been applied to these ditributions. All simulated samples are normalised to an integrated luminosity of $14.34\,\text{fb}^{-1}$.

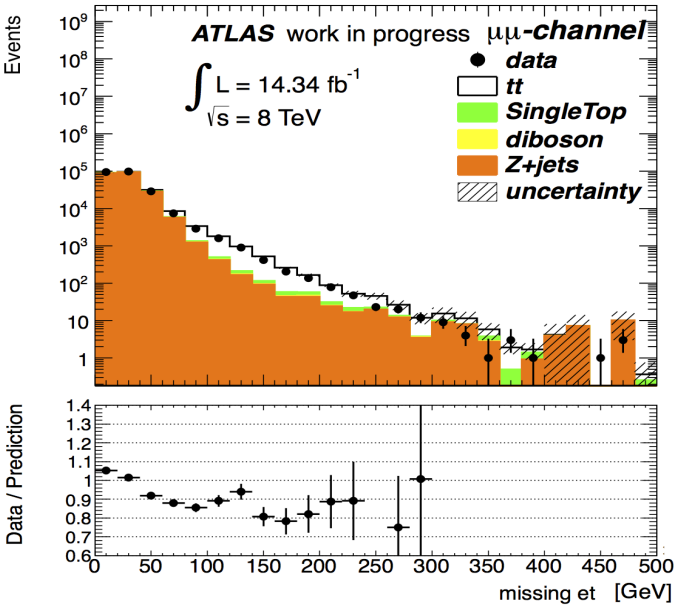


Figure 4.17: The E_T^{miss} distribution for the $\mu\mu$-channel. No E_T^{miss} cut, no Z-mass veto and no Z p_T reweighting have been applied to these ditributions. All simulated samples are normalised to an integrated luminosity of $14.34\,\text{fb}^{-1}$.

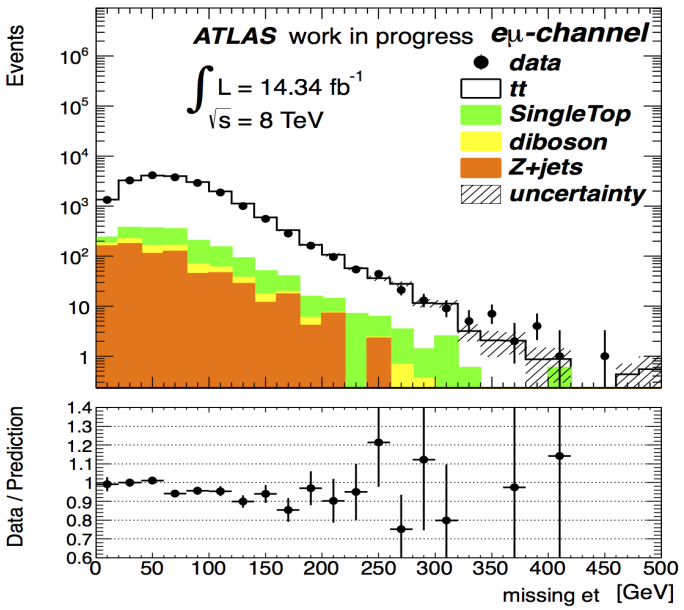


Figure 4.18: The E_T^{miss} distribution for the eμ-channel. No E_T^{miss} cut, no Z-mass veto and no Z p_{T} reweighting have been applied to these ditributions. All simulated samples are normalised to an integrated luminosity of $14.34\,\mathrm{fb}^{-1}$.

Event Selection				
Cut #	Requirement	Data	$t\bar{t}$	Single Top
0	Initial number of events	509436383	11587578	1776505
1	True dilepton event (simulation) or event in GRL (data)	478057981	1349946	1776505
2	Trigger	304840407	1045801	458782
3	No. of tracks associated with primary vertex > 4	300334438	1044862	458108
4	Cosmics	300334435	1044862	458108
5	≥ 2 leptons ($p_{\mathrm{T}} > 25$ GeV)	7687469	296914	27869
6	Single electron or single muon trigger matched	7687356	296837	27859
7	No eμ overlap found in event	7687073	296781	27846
8	Reject event with bad jets	7679603	296356	27807
9	$H_{\mathrm{T}} \geq 130$ GeV(eμ-channel) or MET ee/$\mu\mu$-channels)	122997	225282	18977
10	≥ 2 good jets ($p_{\mathrm{T}} > 20$ GeV)	39629	184346	8553
11	Exactly 2 good leptons ($p_{\mathrm{T}} > 25$ GeV)	39513	184329	8553
12	Two oppositely charged leptons	39010	183471	8496
13	Dilepton invariant mass ≥ 15 GeV	38922	183078	8480
14	Z-Veto	30735	172485	8053
15	Leptons found in truth record (simulation)	30735	171693	8021
16	Veto event with LAr error (remove corrupted events)	30735	171693	8021

Table 4.2: Number of events in the data, nominal simulated $t\bar{t}$and single top samples passing each stage of the $t\bar{t}$ dilepton event selection. Here, "good jets" refers to jets that were not removed during jet cleaning (removal of bad jets). Cosmic events are also removed. Such events are caused by high energy particles originating mainly from outside the solar system which collide with atmospheric molecules. This results in a shower of pions and leptons that can interact with the detector.

All object selection, calibration, background estimations and MC samples for 2012 data follow the guidelines set out by the top analysis group [49].

4.5 p_{T}^{Rel} Calibration Method

One method of calibrating a b-tagging algorithm is to obtain the number of b-jets before and after tagging. If one uses a subset of jets, namely those which decay semileptonically to a muon plus a muon neutrino, it is possible to calculate the p_{T}^{Rel} variable for each jet. This is constructed from the transverse momentum of an in-jet muon with respect to the combined muon plus jet axis (see Figure 4.19).

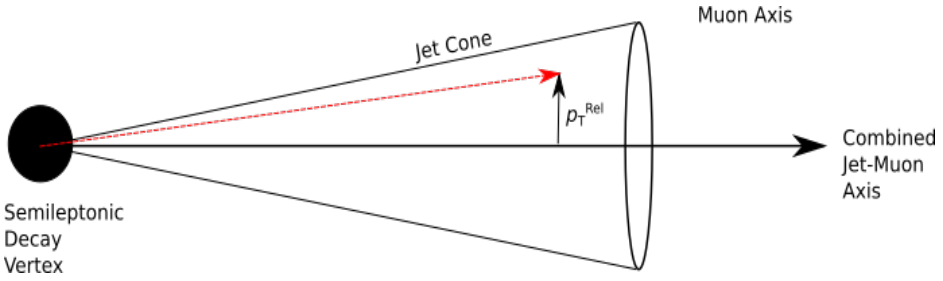


Figure 4.19: Schematic diagram of the p_{T}^{Rel} variable defined as the transverse momentum of the muon with respect to the combined jet-muon direction.

Muons originating from b-hadron decays have a harder p_{T}^{Rel} spectrum than those coming from c- and light-hadrons. By constructing p_{T}^{Rel} spectrum templates of muons from b-, c- and light-hadrons and fitting them to the p_{T}^{Rel} spectrum of muons from jets in data, jet flavour fractions before and after requiring a b-tag are obtained. The efficiency is then calculated using the fraction of b-jets in the pre-tagged (f_b) and tagged (f_b^{tagged}) samples of jets, the total number of jets in those samples (N and N^{tagged} respectively), and the factor C which is a correction factor for biases introduced through differences between the modelling of the b-hadron direction and the heavy flavour contamination of the p_{T}^{Rel} templates for light flavour jets in data with respect to simulation [50]. The equation used to calculate the efficiency is shown in Equation 4.1.

$$\varepsilon_b^{data} = \frac{f_b^{tagged} N^{tagged}}{f_b N} \cdot C. \tag{4.1}$$

The ratio of the b-tagging efficiency on data with respect to simulation is used to calculate a calibration scale factor. Because this method does not use an inclusive sample of b-jets to perform the calibration, it could in principle be biased.

The p_{T}^{Rel} calibration method described in this section is not used in the analysis but a basic knowledge of the technique and selection of semileptonic jets is required, as the same jet selection is used in this study.

4.6 Semileptonic and Non-Semileptonic Jets

Jets selected using the $t\bar{t}$ dilepton kinematic selection procedure are divided up into two orthogonal samples according to the decay of the b-hadron: semileptonic (SL) and nonsemileptonic (NSL). SL-jets are labelled as such if the b-hadron inside the jet decays to a muon and a neutrino. Decays to an electron and a neutrino are not considered SL jets for the sake of this calibration analysis but are included in the NSL jet sample. The relative fraction of SL with respect to NSL jets as a subset of inclusive jets are shown in Table 4.3.

Jet Subset	**# Jets**
Inclusive	1798878
Semileptonic	145135
Non-semileptonic	1653743

Table 4.3: Total numbers of inclusive, semileptonic and non-semileptonic jets for complete set of simulation samples.

To produce the independent subsets the first step is to select a collection of soft muons. After the selection procedure described in section 3.3.3, the soft muon selection procedure shown in Table 4.4 is implemented. Any possibility of overlap with muons from the hard interaction is removed by inverting the isolation requirement.

Selection for soft in-jet muons	
1	Combined "MUID" muon
2	$\Delta R \leq 0.3$

Table 4.4: Selection criteria for soft muons. The selection criteria used in this analysis are the same as those used for the p_{T}^{Rel} analysis. Here, ΔR is measured between the direction of the jet and the muon. Combined "MUID" muons are the muons described in section 3.

If the muon passes the selection criteria, an attempt is made to match it with a b-jet. The ΔR between the soft muon and every selected jet is calculated. The soft muon must be within $\Delta R = 0.3$ of a b-jet if it is to be considered to have originated from the decay of the b-hadron. Any permutation that results in a ΔR greater than 0.3, is rejected. The soft muon is paired with the jet with which it has the smallest ΔR.

The energy of a jet in which the b-hadron decays semileptonically to a muon plus neutrino is often underestimated because a large portion of the muon's energy not being deposited in the calorimeters and the neutrino escaping the detector without depositing any energy. The inclusive rate at which a semileptonic b-decay produces a muon is large ($\approx 20\%$ which includes both direct decay and indirect decay via charm/τ decays) so this can have a large effect. The correction is performed in two stages reflecting the two sources of the energy underestimate. The jet energy is first corrected to include the energy of any nearby muons $\left(\Delta R(j,\mu) < 0.3\right)$. Then, a scaling from a look-up-table is applied on a jet-by-jet basis to correct for the neutrino energy. The correction is p_{T}- and η-dependent and is applied to all SL jets. Corrections are applied to the jet energy which brings it closer to the original b-quark momentum. The full jet momentum is changed so both angles and p_{T} change, which can cause jets to migrate between p_{T} bins in the analysis.

If the jet is matched with a soft muon, a dynamic p_{T} requirement must be fulfilled if the jet is to be labelled an SL jet. This is used to enhance the b-jet purity of the semileptonic jet collection as muons from b-hadrons have a harder p_{T} spectrum than those from charm- or light-hadrons. The cut enforces a minimum p_{T} requirement for soft muons which increases with the p_{T} of the associated jet as seen in Table 4.5. Only jets with associated muons passing this requirement are considered semileptonic jets. This is implemented using the corrected b-jet energy. NSL jets within $\Delta R(j,\mu) < 0.3$ but whose muon does not exceed the dynamic p_{T} threshold still have their energy corrected but are not labelled SL for this calibration.

Dynamic p_{T}-cut for Soft Muons in B-jets	
Jet p_{T} range (GeV)	p_{T}^{μ} cut (GeV)
20-60	$p_{\mathrm{T}}^{\mu} > 4$
60-90	$p_{\mathrm{T}}^{\mu} > 6$
90-300	$p_{\mathrm{T}}^{\mu} > 8$

Table 4.5: Dynamic p_{T}-cut imposed on in-jet muons for varying jet p_{T}. If the muon passes the selection cut, the jet is assigned to the SL jet sample, otherwise it is assigned to the NSL jet sample.

If all these requirements are met, the jet is labelled as an SL jet, otherwise it is labelled as an NSL jet.

4.7 $t\bar{t}$ Dilepton Kinematic Selection Calibration Method

The efficiency of tagging a true b-jet (ε_b) is the number of correctly tagged true b-jets with respect to the number of true b-jets in the sample. Non b-jets are separated into 4 categories: light, charm (c), τ and fake. Here light jets refers to any jet stemming from a *u*-, *d*- or *s*- quark or a gluon. The efficiency of incorrectly tagging light c- or τ-jets (ε_l, ε_c, ε_τ) is calculated as the ratio between the number of true light, c- or τ-jets that have been incorrectly b-tagged (mistagged) with respect to the total amount in the combined simulated samples. The mistag efficiencies (ε_c, ε_l and ε_τ) have additional correction factors before they are applied data [41]. The efficiency with which jets from fake dilepton events are incorrectly b-tagged (ε_{fake}) is taken directly from the data-driven same-sign dilepton fakes sample as defined in Section 4.2. ε_{fake} is calculated as the number of jets from events in the fake sample that are being tagged, divided by the total number of jets in these events.

The number of jets b-tagged in data (N^{data}_{tagged}) by any b-tagging algorithm is equal to the sum of the number of true b-jets that are correctly tagged multiplied by the tagging efficiency of that algorithm, plus the number of non b-jets that are mistagged multiplied by the mistag efficiency of that algorithm. The number of b-tagged jets in data is:

$$N^{data}_{tagged} = N_b\varepsilon_b + N_c\varepsilon_c + N_l\varepsilon_l + N_\tau\varepsilon_\tau + N_{fake}\varepsilon_{fake} \tag{4.2}$$

where N_b, N_c, N_l and N_τ are the number of true b-, charm, light and τ-jets in the data sample, estimated from simulation and N_{fake} is the number of fake jets in data.

This assumes the relative and total number of jets in each sample is modelled correctly in simulation. Rearranging Equation 4.2 and replacing the absolute numbers of jets with relative flavour fractions (by dividing by the total number of jets in the data/prediction), one obtains the equation:

$$\varepsilon^{data}_b = \frac{f^{data}_{tagged} - \left(f^{MC}_c\varepsilon^{MC}_c + f^{MC}_l\varepsilon^{MC}_l + f^{MC}_\tau\varepsilon^{MC}_\tau + f^{data}_{fake}\varepsilon^{data}_{fake}\right)}{f^{MC}_b} \tag{4.3}$$

where f_b^{MC}, f_c^{MC}, f_l^{MC} and f_τ^{MC} are the flavour fraction of b-, charm, light and τ-jets taken from simulation, f_{fake}^{data} is the flavour fraction of fake jets obtained from the data-driven sample (as described in Section 4.2) and where we have additionally indicated with a superscript "MC", that the mistag efficiencies (ε_l^{MC}, ε_c^{MC}, ε_τ^{MC}) are also obtained from MC. The values are shown in Figure 4.20.

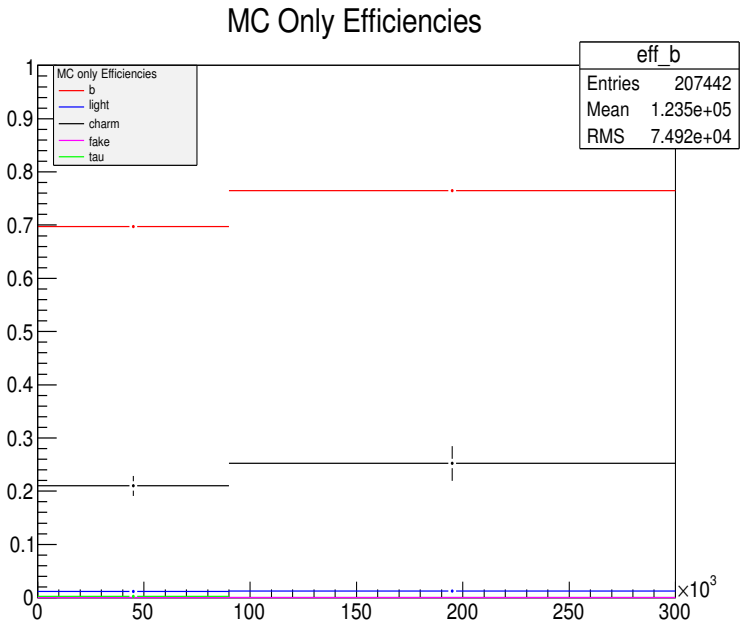


Figure 4.20: The tag/mistag efficiencies of the MV1 b-tagging algorithm for b-, c-, light, fake and τ-jets as a function of jet p_T for the 70% working point (i.e. ε_l^{MC}, ε_c^{MC}, ε_τ^{MC} and $\varepsilon_{fake}^{data}$). The b-tag efficiency is calculated as the number of b-tagged jets divided by the number of true b-jets in the $t\bar{t}$ sample ($\varepsilon_b^{t\bar{t}}$ as used in equation 4.4).

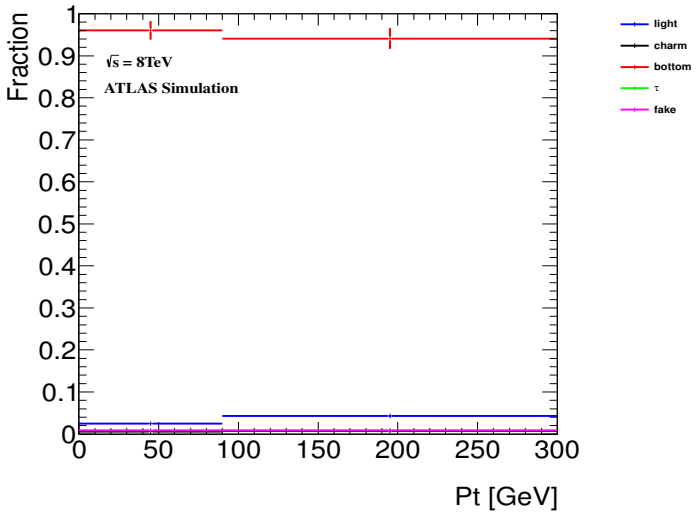
In this equation, ε_b is the b-tagging efficiency for b-jets in data - the parameter we wish to measure. One of the benefits of the $t\bar{t}$ dilepton kinematic selection method is that it derives the efficiency from an enriched sample of b-jets without requiring any information about the decay of the b-hadron, which means the sample of b-jets is inclusive. This helps prevent any calibration bias that could be introduced due to the different kinematic properties of particular subsets of b-jets for example, those which decay to a muon and neutrino.

The flavour of a jet is determined in simulation using information in the truth record which signifies the flavour of the quark from which the jet originated. This procedure is common to most flavour tagging analyses. The flavour fractions for an SL jet sample can be seen in Figure 4.21(a)

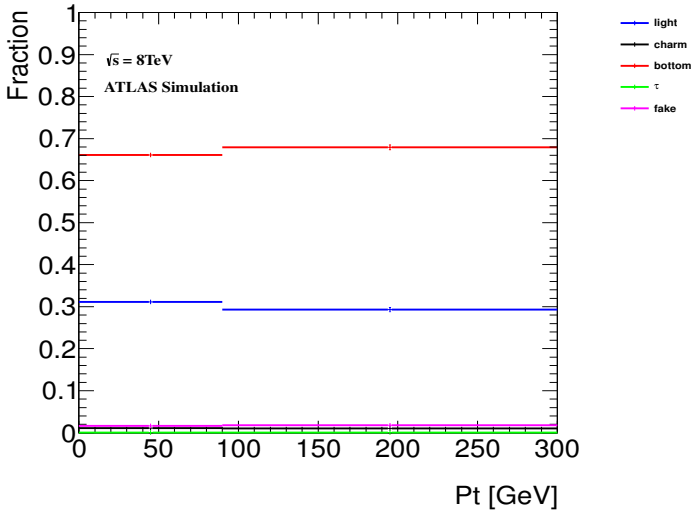
and the same plots for NSL jets are shown in Figure 4.21(b). The calibration scale factor is the ratio of the b-tagging efficiency in data with respect to the b-tagging efficiency in simulation:

$$\kappa_b = \frac{\varepsilon_b^{data}}{\varepsilon_b^{t\bar{t}}} \tag{4.4}$$

where the b-tagging efficiency on MC, $\varepsilon_b^{t\bar{t}}$, extracted from a sample of truth-tagged b-jets, is the baseline efficiency. It is calculated as the fraction of true b-jets that are b-tagged using true b-jets from the $t\bar{t}$ sample. In principle, any sample of true b-jets can be used to calculate a baseline simulation efficiency, however $t\bar{t}$ events are used given that the kinematic dilepton calibration method targets $t\bar{t}$ events in data. This ensures the kinematic properties of the b-jets in data and the baseline sample are the same and no bias has been introduced into the scale factor, which is important as we know efficiencies are affected by the kinematic properties of the jets.



(a)



(b)

Figure 4.21: Flavour fractions for a (a) semileptonic sample of jets and (b) non-semileptonic sample of jets. Plots include all simulation samples with data driven fake estimate.

4.8 Semileptonic Calibration Bias

The calibration procedure described in Section 4.7 is used to calculate b-tagging efficiencies (ε_b) and scale factors (κ_b) for each subset of jets. Efficiencies and scale factors calculated using SL and NSL jets are denoted ε_b^{SL}, κ_b^{SL}, ε_b^{NSL} and κ_b^{NSL} respectively. The final bias is defined as the ratio of scale factors calculated using the two subsets:

$$\beta^{SL,NSL} = \frac{\kappa_b^{SL}}{\kappa_b^{NSL}}. \tag{4.5}$$

The inclusive jet sample is just the sum of the SL and NSL sub-sets and therefore the b-tagging efficiency on an inclusive sample (ε_b^{Incl}) is the sum of the efficiencies on the sub-samples multiplied by their relevant fractions:

$$\begin{aligned} \varepsilon_b^{Incl,data} &= f_b^{SL,data}\varepsilon_b^{SL,data} + f_b^{NSL,data}\varepsilon_b^{NSL,data}. \\ \varepsilon_b^{Incl,MC} &= f_b^{SL,MC}\varepsilon_b^{SL,MC} + f_b^{NSL,MC}\varepsilon_b^{NSL,MC}. \end{aligned} \tag{4.6}$$

4.9 Systematics

There are various sources of systematic uncertainty that must be considered in this calibration. Most of these manifest themselves as changing event weights which are related to the efficiencies of identification and reconstruction or changes in object selection (which can come from e.g. uncertainties in energy scales). There are also systematic uncertainties which represent the effect of using different physics models (e.g. a different Monte Carlo generator), along with theoretical errors such as imperfect cross-section calculations.

The effect of a systematic uncertainty is evaluated by implementing a systematic variation and re-running the calibration. The systematic uncertainty on the b-tagging efficiency is calculated from the difference between the value using the systematic variation and the nominal value:

$$\Delta\varepsilon_b = \varepsilon_{b,syst}^{data} - \varepsilon_{b,nominal}^{data}. \tag{4.7}$$

The systematic uncertainty is propagated to the scale factor:

$$\Delta\kappa_b = \kappa_{b,syst} - \kappa_{b,nominal} = \frac{\varepsilon_{b,syst}^{data}}{\varepsilon_{b,syst}^{MC}} - \frac{\varepsilon_{b,nominal}^{data}}{\varepsilon_{b,nominal}^{MC}}. \tag{4.8}$$

The systematic uncertainties on b-tagging efficiencies and scale factors are evaluated for SL and NSL jets separately. This is then propagated to the final bias of the SL calibration as follows:

$$\Delta\beta^{SL,NSL} = \Delta\left(\frac{\kappa_b^{SL}}{\kappa_b^{NSL}}\right) = \frac{\kappa_b^{SL,syst}}{\kappa_b^{NSL,syst}} - \frac{\kappa_b^{SL,nominal}}{\kappa_b^{NSL,nominal}}. \tag{4.9}$$

Individual errors are symmetrised by taking the difference between the maximum (Δ_i^{up}) and minimum (Δ_i^{down}) systematic variations and dividing by 2:

$$\Delta_i = \frac{\Delta_i^{up} - \Delta_i^{down}}{2} \tag{4.10}$$

The total systematic error is calculated by summing the full set of independent symmetrised errors in quadrature $\left(\sum_i \Delta_i^2\right)$. Total uncertainties are calculated as the sum of the statistical and systematic uncertainties. The systematic uncertainties on the final SL/NSL bias are shown in Appendix B in Table B.1.

The most important systematic uncertainties are associated with the choice of MC generator, the parton shower model uncertainty, the uncertainty associated with the amount of initial state/final state radiation and the statistical uncertainty from the number of semileptonic jets. The meaning and importance of individual systematics are described in more detail beneath. A more detailed description can be found in [51].

$t\bar{t}$ Monte Carlo Generator Uncertainty

This uncertainty quantifies the effect of flavour fraction variations between different $t\bar{t}$ Monte Carlo generators. The nominal $t\bar{t}$ sample used in the calibration was generated using POWHEG+PYTHIA which uses the full ATLAS detector simulation. The available systematic samples for Monte Carlo generator comparison were ALPGEN+HERWIG and MC@NLO+HERWIG. To remove any hadronisation effects the "nominal" sample used to evaluate this systematic uncertainty was POWHEG+HERWIG. Both systematic samples were generated using a fast simulation of the ATLAS detector [52] whereas the nominal sample used a full simulation. Comparing with a fast simulation sample should not affect the b-tagging efficiencies as the inner detector simulation is the same as in the full simulation. The uncertainty is then evaluated by comparing the b-tagging efficiency in each jet p_{T} bin using samples from each generator. The largest variation between nominal and any of the systematic samples is used as the uncertainty.

$t\bar{t}$ Parton Shower Modelling Uncertainty

Different parton shower models result in different event kinematics. When comparing parton shower models it is important to use samples with the same Monte Carlo generator to ensure systematic variations are independent of one another. The comparison performed for this calibration is between the nominal POWHEG+PYTHIA sample and the systematic POWHEG+HERWIG sample. Both samples were simulated using a fast detector simulation. The systematic uncertainty is evaluated by comparing the b-tagging efficiency for each jet p_{T} bin between the nominal and systematic samples generated using the various parton shower models.

$t\bar{t}$ Intial/Final State Radiation Modelling Uncertainty

To evaluate the uncertainty of the modelling of the initial and final state radiation, two simulation samples were generated with increased and decreased amounts of initial and final state radiation. The two samples are ACERMC POSITIVE and ACERMC NEGATIVE respectively and are both simulated using fast detector simulations. The samples are used independently to calculate a full set of results. The systematic error is taken as half the difference between the results from the positive and negative variations.

Jet Energy Resolution (JER)

The JER was well measured using data from 2012 using the bisector technique in dijet events to measure *in situ*, the fractional jet p_{T} resolution $\sigma(p_{\mathrm{T}})/p_{\mathrm{T}}$ [53]. The JER in the MC is comparable to the resolution observed in data but small differences are still observed in the p_{T} dependant distributions of JER. The procedure recommended at the time of this calibration was to not apply any smearing on the nominal measurement and compare with a systematic result evaluated after a smearing factor has been applied as a scale factor to MC jets which increases the JER to be consistent with the JER measured in data plus its error. The difference between the nominal and smeared MC taken as the JER systematic uncertainty which covers the uncertainty on the data measurement and the small disagreement between data and simulation.

Jet Reconstruction Efficiency (JEff)

Calorimeter jets are reconstructed with an efficiency derived relative to jets built from charged tracks reconstructed in the inner detector using a tag and probe technique. The reconstruction efficiency was defined as the ratio of probe track-jets matched to calorimeter jets with respect to the number of probe track-jets. The observed difference between data and Monte Carlo was applied to Monte Carlo by discarding a fraction of jets taken at random within the inefficiency range $p_{\mathrm{T}} < 30$ GeV to evaluate the systematic variation.

Jet Vertex Fraction (JVF)

The JVF variable exploits the fraction of tracks coming from the primary vertex that are associated with the jet and estimates the contribution of multiple interactions. If the jet has a sufficient number of tracks coming from the primary vertex it is considered not to have been generated by pileup interactions. A systematic uncertainty relating to the JVF is calculated by varying the value of this cut depending on whether the jet originates from the hard interaction or a pile-up interaction. If the jet is considered a pileup jet, the JVF cut is varied from the nominal value up to 1.1 (removing all pileup jets) and down to 0.1. If the jet is not a pileup jet, the JVF cut value is varied up to 0.53 and down to 0.47 from its nominal value of 0.5. This can cause events to migrate between jet multiplicities which varies the final results of the calibration.

Jet Energy Scale Uncertainty

As described in Section 3.3, a series of corrections are used to correct the energies of simulated jets. Similarly, corrections are derived to jet energies to account for the response of the detector and *in situ* corrections are applied using $Z+$jets and $\gamma+$jets events where the p_T of the system is assumed to be balanced. All corrections have a systematic uncertainty which must be propagated through to the calibration results.

1. Baseline *in situ* uncertainty:

   The total jet energy scale calibration includes a combination of many *in situ* measurements, therefore it is preferable to reduce the number of uncertainty components. This is done using the total covariance matrix of the JES correction factors derived from the individual components of the statistical and systematic uncertainties. This matrix is decomposed into a part with strong correlations and a part with weak correlations. Diagonalising this matrix produces eigenvectors and eigenvalues which when multiplied together produce a reduced set of effective uncertainties. The result is 5 effective systematics (a subset of the full set of effective uncertainties with the largest eigenvalues) which represent the terms that are strongly correlated and a residual effective systematic to account for the remaining difference between the approximation and the total uncertainty. These are then propagated through to the b-tagging efficiency and scale factor calculations.

2. η-intercalibration

   The response of the ATLAS detector to a jet is dependent on the jet's direction, due to changing calorimeter technology and varying amounts of dead material. A calibration is therefore needed to ensure uniform calorimeter response. This is done using correction factors derived from Monte Carlo simulation using the *in situ* techniques mentioned earlier. The major sources of systematic uncertainty associated with this are due to the modelling of jets in the forward regions of the detector and the statistical methods used. Seeing as these correction factors are used for jets in this b-tagging calibration, the systematic uncertainty associated must also be taken into consideration.

3. Flavour Composition uncertainty

The jet energy response is not the same for quark-initiated jets and gluon-initiated jets. Gluon jets have a lower response than quark jets because they fragment into more and softer particles than quark jets. This means they are less likely to reach the calorimeters or be included in the topological clusters. The *in situ* jet energy scale calibrations are derived using samples with jet flavour fractions which might not reflect the composition of jets in the physics analysis. Varying the fraction of gluon jets used in the *in situ* calibrations can cause shifts in the JES and therefore a systematic uncertainty is assigned to compensate for these possible variations of flavour compositions.

4. Flavour response

   The LCW jet energy calibration scheme restores the average energy of the calorimeter jets reconstructed at the LCW scale, to the jet energy scale of jets reconstructed from stable simulated particles (truth particles), for an inclusive sample of jets. However , sub-samples are not perfectly calibrated. The flavour dependence of the calorimeters' jet energy response is typically larger in Monte Carlo simulation than in data. This occurs if the flavour content of in the data sample is not well described by the MC. Scale factors are applied to jets in simulation which resolve these differences. The derivation of these scale factors has a systematic uncertainty which propagates through to the results of this calibration.

5. B-Jet JES uncertainty

   The detector response for b-jets is different from that for the light- and gluon-jets. A separate *in situ* JES calibration is derived which has its own systematic uncertainty. The b-jet energy scale uncertainty is only applied to true b-jets instead of the flavour composition and response uncertainty.

6. Pileup corrections

   The number of reconstructed primary vertices can affect the measurement of jet energies. A number of corrections are applied which are dependent on the number of reconstructed primary vertices. These corrections also consider the active area of jets in calorimeter. Pileup corrections have a systematic uncertainty which is evaluated in any analysis by varying the

parameters of these corrections and re-evaluating the results.

7. High p_{T} single particle

   A high p_{T} hadron in a high p_{T} jet can potentially be mismodelled which can affect the JES measurement. A correction is introduced to account for this mismodelling. The derivation of this correction has a systematic uncertainty which affects the overall JES corrections for this b-tagging calibration and can cause a change in jet acceptances.

Luminosity

The luminosity calculation for the datasets in the "AllGood" good runs list described in section 4.3 has an estimated uncertainty of 2.8%. This is propagated to the final results by varying the luminosity up and down by this percentage and is performed consistently for all simulations. Seeing as Equation 4.3 uses relative flavour fractions instead of absolute numbers this uncertainty does not have an effect.

Mistag Rate

In the derivation of the b-tagging scale factors in this analysis, mistag rates for non b-jets are taken from simulation and multiplied by a scale factor before they can be applied to data. The mistag rates on data have an additional systematic uncertainty that comes from the methods used to derive the scale factors applied to the MC mistag rates. To evaluate this systematic uncertainty, the mistag efficiencies are shifted by 12-50% [50] when deriving the b-tagging scale factors.

Lepton Systematic Uncertainties

Electrons and muons both require scale factors to correct for specific aspects of data /MC agreement. In particular weights are provided to correct for trigger, reconstruction, isolation and identification efficiencies which are applied multiplicatively to events. Each and every scale factor applied has an uncertainty which must be propagated to the final result.

Electron reconstruction and isolation efficiencies from $\mathrm{Z} \rightarrow \mathrm{ee}$ and $\mathrm{W} \rightarrow \mathrm{e}\nu$ events in data and simulation are used to derive scale factors. Energy scale and resolution are compared between data and simulation using $\mathrm{Z} \rightarrow \mathrm{ee}$ and $\mathrm{J}/\psi \rightarrow \mathrm{ee}$. The values of these scale factors and their uncertainty which is propagated through this calibration can be found in Reference [37].

Measurements of the muon reconstruction efficiencies, momentum scale and resolution based on $Z \to \mu\mu$, J/ψ and $\Upsilon \to \mu\mu$ decays in data are compared to MC simulations. Derived corrections to the simulation, are used in physics analyses which are applied to both the inner detector and muon spectrometer tracks. This introduces additional systematic uncertainties resulting from the methods used to derive these corrections. The uncertainty on the efficiency is 0.001%, whereas the uncertainty on the resolution ranges from 1.7-4% and the momentum scale is known to an uncertainty between 0.05-0.2% depending on the muons rapidity. The uncertainty on the momentum scale and resolution are used to smear the p_{T} of the reconstructed muons in the analysis which affects object reconstruction and subsequently the derivation of calibration scale factors. No measurements of isolation scale factors were available at the time of this calibration. The values for all scale factors and uncertainties can be found in Reference [36].

Semileptonic Correction Uncertainty

The energy of a jet in which the b-hadron decays semileptonically can be underestimated. Corrections are applied to the jet energy that bring it closer to the original b-quark momentum. The uncertainty on this correction is close to 2% and affect the acceptance of semileptonic b-jets.

Theoretical Cross Section Uncertainty

Theoretical cross-sections of simulated backgrounds are subject to the uncertainties shown in Table A.1. The overall normalisation of individual simulated background samples are scaled up and down using the upper and lower limits of the uncertainty on a cross-section which models the different possible sample contributions. The uncertainty on this calibration is then evaluated by comparing the results acquired using the samples with systematic variations of cross-sections and the nominal samples.

Theoretical Cross-Section Uncertainties (%)	
$t\bar{t}$ cross-section	+9.34%, - 10.17%
Single Top (Wt)	±6.85%
Diboson cross-section	± 5% (inclusive normalisation) and 24% additional uncertainty per additional jet to be added in quadrature for the WW samples [54]
Z + jets cross-section	± 4% (inclusive normalisation) and 24% additional uncertainty per additional jet to be added in quadrature [55]
Fakes cross-section	± 50%

Table 4.6: The theoretical uncertainties of signal and background processes. The upper and lower limits are used to scale the theoretical cross-sections and thus the normalisations. The inclusive normalisation uncertainty for both Diboson and Z + jets samples is the default baseline uncertainty. The 24% uncertainty per additional jet is needed for samples where the additional jet required by the selection has had to come from some softer process. The extra jet comes from the additional matrix element parton simulated by ALPGEN which has a 24% uncertainty associated with it (see Table A.2 for all cross-sections).

4.10 Statistical Uncertainties

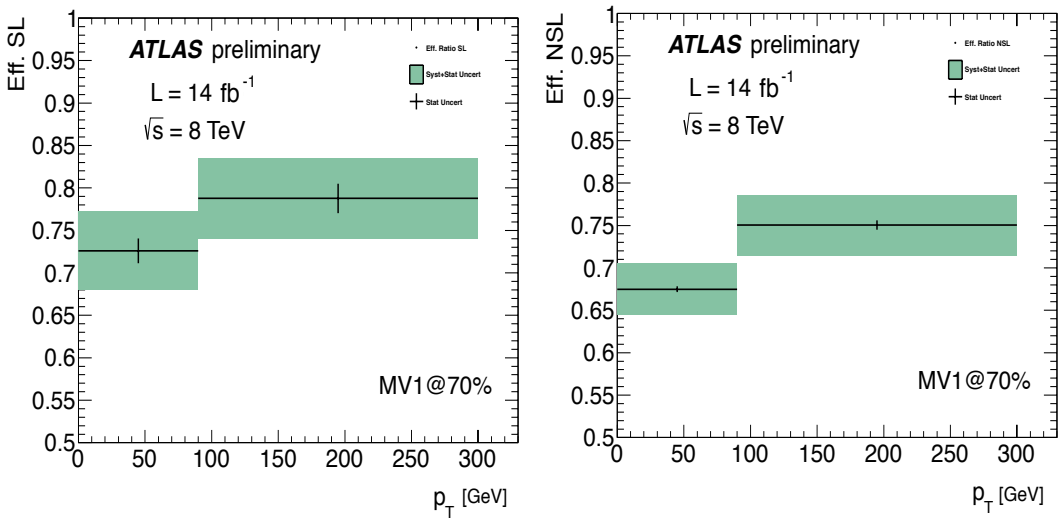
The statistical uncertainty is evaluated on the efficiency and scale factor histograms in individual jet p_{T} bins. A histogram can be interpreted as a series of independent Poisson distributions for the number of entries in each bin. For given values of some random variable n, one can calculate the Poisson probability distribution about the mean value μ of a discrete random variable N being equal to any value of n:

$$f(n;\mu) = Pr(N = n) = \frac{\mu^n}{n!}e^{-\mu} \qquad n = 0,1,2 \text{ etc.} \tag{4.11}$$

Where N follows a Poisson distribution with parameter μ. Here μ is the parameter of the distribution which represents the mean number of events in a given interval. The standard deviation of a probability distribution is defined as the square root of the variance ($\sqrt{\sigma^2} = \sqrt{\mu}$), which for one bin in a histogram, is the square root of the summed entries. Errors along the x-axis are equal to the bin width.

4.11 Results

In this section results from the calibration of the ATLAS MV1 b-tagging algorithm using the $t\bar{t}$ dilepton kinematic selection method are shown for the 70% efficiency working point. The efficiencies obtained from the calibrations using a semileptonic jet sample and non-semileptonic jet sample are shown in 4.22.



(a) (b)

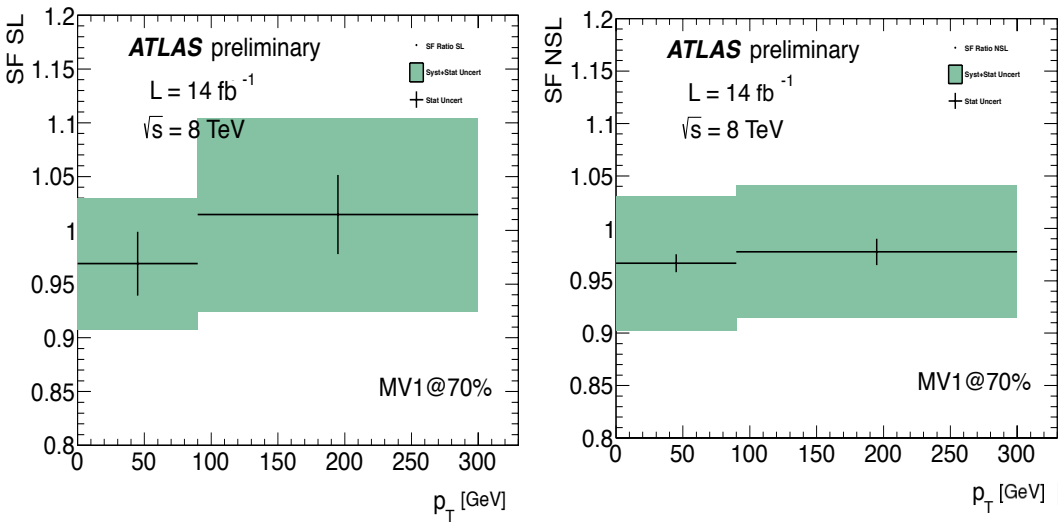
Figure 4.22: The efficiency of the MV1 b-tagging algorithm using (a) semileptonic jets and (b) non-semileptonic jets from data using the 70% MV1 working point.

The individual p_{T} bin efficiencies for semileptonic jets are in agreement with the relevant p_{T} bin efficiencies for non-semileptonic jets within uncertainties.

Jet p_{T} bin	SL Nominal Efficiency	NSL Nominal Efficiency
0-90 GeV	$0.72\ ^{+5\%}_{-5\%}$ (stat+syst)$\pm$2% (stat)	$0.67\ ^{+2\%}_{-2\%}$ (stat+syst) $\pm$1% (stat)
90-300 GeV	$0.78\ ^{+6\%}_{-6\%}$ (stat+syst)$\pm$2% (stat)	$0.74\ ^{+3\%}_{-3\%}$ (stat+syst) $\pm$1%(stat)

Table 4.7: Table of the final efficiencies in each jet p_{T} bin for the nominal calibrations.

The data/simulation scale factors for the 70% working point are shown in Figure 4.23(a) for semileptonic jets and Figure 4.23(b) for non-semileptonic jets.



(a) (b)

Figure 4.23: The p_{T} dependent data/simulation scale factor as evaluated using (a) semileptonic jets and (b) non-semileptonic jets from data and simulation using the 70% MV1 working point.

Jet p_{T} bin	SL Nominal Scale Factor	NSL Nominal Scale Factor
0-90 GeV	$0.97\ ^{+5\%}_{-6\%}$(stat+syst) $\pm 3\%$ (stat)	$0.97\ ^{+6\%}_{-7\%}$ (stat+syst) $\pm 1\%$ (stat)
90-300 GeV	$1.01\ ^{+9\%}_{-8\%}$ (stat+syst) $\pm 4\%$ (stat)	$0.98\ ^{+6\%}_{-7\%}$ (stat+syst) $\pm 1\%$ (stat)

Table 4.8: The final efficiency scale factors in each jet p_{T} bin for the nominal calibrations.

To evaluate any potential bias in a calibration using semileptonic b-jets, the ratio of the semileptonic scale factor divided by the non-semileptonic scale factor is presented in Figure 4.24.

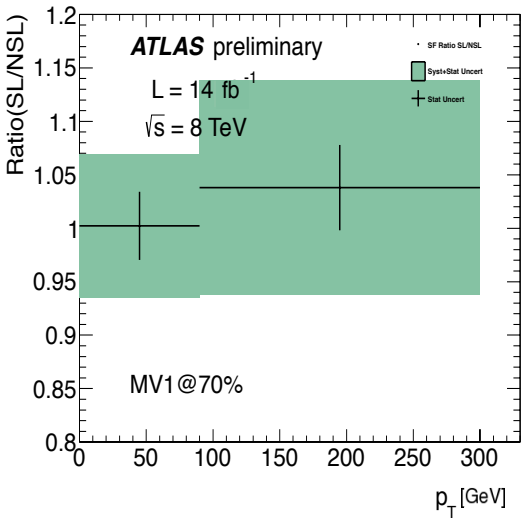


Figure 4.24: The bias, as defined in Equation 4.5 binned in jet p_{T}. The bias is defined as the scale factor derived using semileptonic jets divided by the scale factor derived using non-semileptonic jets.

A full breakdown of the systematic uncertainties is shown in Table B.1. Table 4.9 shows a summary of the full systematic and statistical uncertainty along with the nominal values for the bias. The small amount of data in the semileptonic jet sub-sample contributes approximately 3-4% of the statistical uncertainty to the final bias value in both p_{T} bins. Other large contributions to the overall uncertainty come from uncertainties such as those due to the choice of monte carlo generator, parton shower and Initial/Final state radiation to which the bias evaluation is more sensitive to in the 90-300 GeV jet p_{T} bin.

Semileptonic Bias (SL/NSL) and Associated Uncertainties		
Jet p_{T} bin	Nominal Bias	Total Uncertainty
0-90 GeV	1.00	6.75%
90-300 GeV	1.04	9.86%

Table 4.9: Table of the final bias ($\kappa_b^{SL}/\kappa_b^{NSL}$) in each jet p_{T} bin for the nominal calibrations along with their respective total uncertainties.

4.12 Conclusions

The $t\bar{t}$ dilepton kinematic selection method was used to select an enriched inclusive sample of b-jets in data exclusively from the eμ-channel. These were then further divided into two sub-samples:

semileptonic b-jets with a soft muon and non-semileptonic b-jets. The efficiency of the MV1 b-tagging algorithm was calibrated for each of these subsets of jets using the 70% working point. This was done using data and simulation alike and a b-tagging scale factor was calculated as the ratio of the efficiency on data with respect to simulation. This was done for both semiletponic and non-semileptonic b-jets. These two scale factors were then compared in a ratio (semileptonic/non-semileptonic) to evaluated whether or not a bias is caused when performing b-tagging calibrations on semileptonic b-jets.

The ratio of scale factors evaluated using jets with p_{T} in the range 0-90 GeV was found to be $1.00 \pm 3.18\%$ (Stat) $\pm 3.5\%$(Syst) and the ratio evaluated using jets with p_{T} in the range 90-300 GeV was calculated to be $1.04 \pm 3.84\%$ (Stat) $\pm 6.01\%$(Syst). These results show that the ratios of scale factors (the biases) are consistent with unity within uncertainties, therefore no bias is seen. If a bias exists, it must be less than a few percent. Improving the precision of the analysis could rule such a bias out.

Chapter 5

A search for the Standard Model Higgs boson produced in association with top quarks and decaying to a $b\bar{b}$ pair

> *Have you ever heard about the Higgs Boson Blues?*
> *Im going down to Geneva baby,*
> *Gonna teach it to you*
>
> Nick Cave and the Bad Seeds, *Higgs Boson Blues*

5.1 Introduction

The discovery of a new particle in the search for the Standard Model Higgs boson was reported by the ATLAS [56] and CMS [57] collaborations in 2012. Both collaborations showed clear evidence for a particle with a mass of 125 GeV in the decay channels $H \to \gamma\gamma$, $H \to ZZ \to 4l$ and $H \to WW \to l\nu l\nu$. Within the framework of the SM, the Higgs boson couplings are predicted for any hypothesised mass. Thus, in order to understand if the observed resonance with a mass of 125 GeV is the SM Higgs boson, its couplings must be measured and compared to those predicted for a SM Higgs boson with that mass. The largest Higgs boson coupling is the so-called top Yukawa coupling which is the coupling between the top quark and the Higgs boson. As the Higgs boson cannot decay to top quarks, this coupling must be measured via a Higgs boson production process.

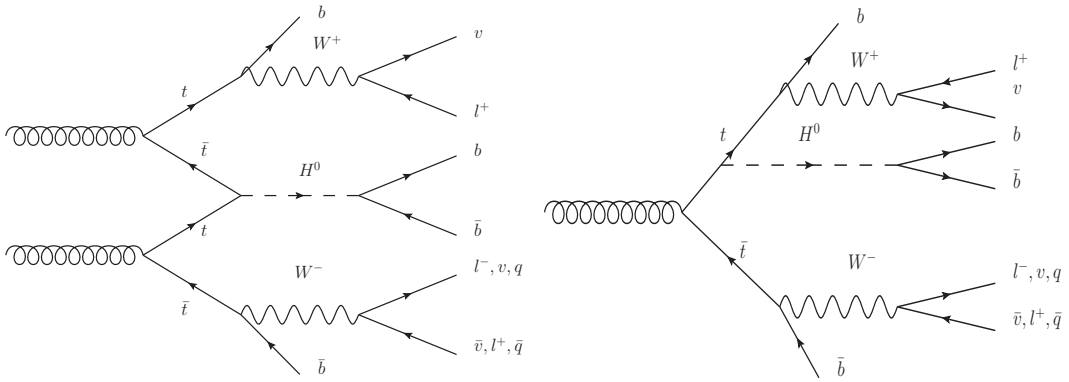


Figure 5.1: Tree-level Feynman diagrams for the production of a Higgs boson in association with a pair of top quarks and decaying to a pair of b-quarks.

The analysis presented here is a search for the Higgs boson produced in association with a pair of top quarks ($t\bar{t}H$), which aims to measure the top Yukawa coupling. The analysis targets $t\bar{t}H$ events where the Higgs boson decays to a pair of b-quarks because the $H \to b\bar{b}$ branching ratio is the largest of the 125 GeV Higgs boson. However, other decays are accepted as signal (such as $H \to WW \to l\nu l\nu$). Figure 5.1 shows the two tree-level Feynman diagrams for the $t\bar{t}H(H\to b\bar{b})$ signal process. An added bonus to searching for this Higgs boson decay signature is that it contributes to the measurement of the $H \to b\bar{b}$ coupling.

5.2 Analysis Strategy

The $t\bar{t}H(H\to b\bar{b})$ channel benefits from a high jet/b-jet multiplicity and additional leptons in its final state signature, which are used to reduce many background processes making it a favourable search channel. However, it comes with some very challenging backgrounds. The main background in this analysis comes from $t\bar{t}$ events which have at least two additional jets ($t\bar{t}+\text{jets}$). These jets most commonly come from a gluon that originated from the initial/final state. The $t\bar{t}+\text{jets}$ background is categorised according to the flavour of the additional jets produced. Events with additional b-jets ($t\bar{t}+b\bar{b}$) make up an irreducible background as in Figure 5.2, which means they have the same final state signature as the signal. The rest of the $t\bar{t}+\text{jets}$ category is composed of events where the additional jets are light jets (u,d,s), gluon jets or c-jets ($t\bar{t}+\text{light}$ and $t\bar{t}+c\bar{c}$). These events can also produce the same final state signature as signal if the jets are misidentified as b-jets and thus this analysis is very dependent on the performance of the b-tagging algorithm.

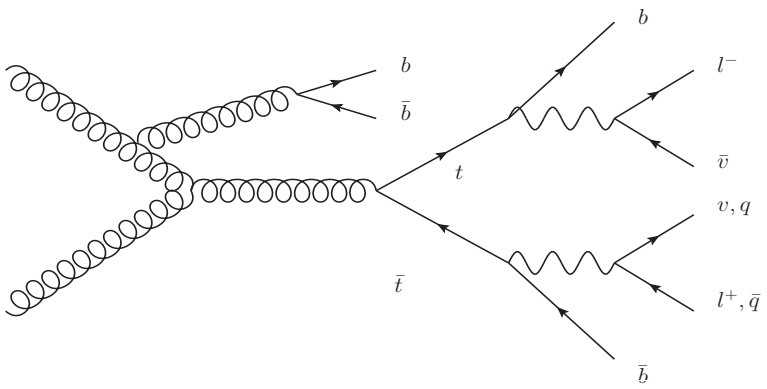


Figure 5.2: Tree-level Feynman diagram of the dominant $t\bar{t}$+$b\bar{b}$ background process for the production of the Higgs boson in association with a top pair and subsequent decay to a pair of b-quarks. The diagram shown is for the single lepton and the dilepton channels.

Separate analyses are performed according to the number of leptons found to have come from the decay of the $t\bar{t}$ system. The two analyses that are combined in Reference [58] are the so-called single lepton (or lepton+jets) and dilepton analyses. A Feynman diagram for the signal process in each analysis is shown in Figure 5.1. Using a carefully designed selection, the single lepton analysis targets events where one W boson coming from the decay of the top/anti-top quark has decayed leptonically and the other W boson has decayed hadronically. The dilepton analysis targets events where both of the W bosons have decayed leptonically.

Each analysis has its own advantages and disadvantages. The single lepton analysis has higher statistics than the dilepton analysis due to the high branching fraction of the W bosons into a charged leptons and jets, as is demonstrated in Figure 5.3. However, given that the final state signature contains more jets, the single lepton channel suffers from a larger background. The dilepton signature on the other hand has two high-p_{T} leptons which makes the signature much clearer and allows for good signal-to-background separation but suffers from a smaller branching ratio.

The two analyses are designed to be orthogonal to one another, thus allowing them to be combined. It can be seen from Figures 5.1 and 5.2 that the single lepton analysis has a larger expected significance ($S/\sqrt{B}$) than the dilepton analysis. Here '*S*' is used to denote the expected number of signal events in that region for a SM Higgs boson of mass 125 GeV and '*B*' denotes the expected number of background events. However, it can also be seen that the dilepton analysis is generally purer in signal, averaging a higher S/B across analysis regions (a concept described in more detail in in more detail later in this section) and thus contributes significant additional

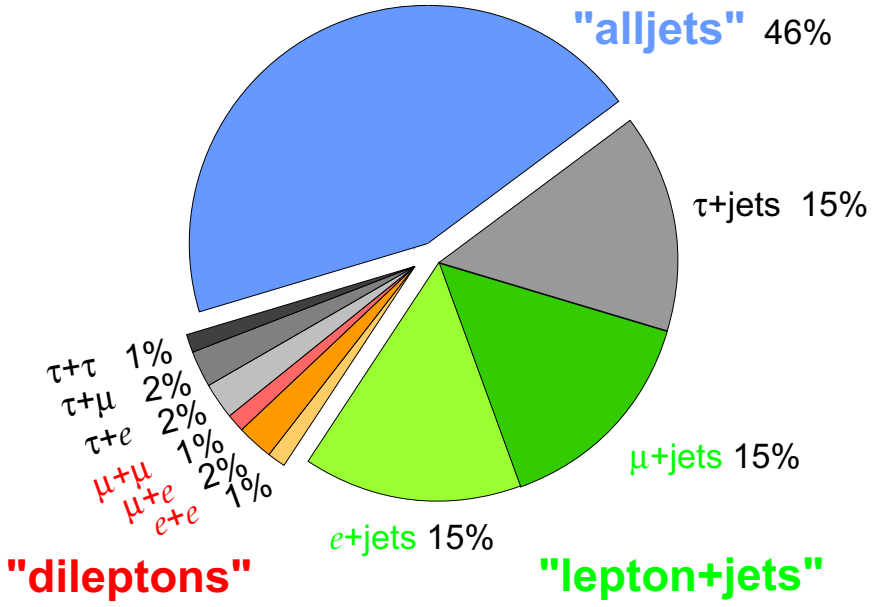


Figure 5.3: Branching fractions of the top pair system [59].

sensitivity to the search.

Moreover, combining the two channels can reduce the size of the systematic uncertainties compared to any one channel alone. For example, the (≥ 4j, ≥ 4b) dilepton analysis region is much more pure in $t\bar{t}$+$b\bar{b}$ than any of the single lepton analysis regions (as shown in Figures 5.4 and 5.5). It can therefore be used to provide improved shape and normalisation information on the $t\bar{t}$+$b\bar{b}$ background in the single lepton analysis regions via the likelihood fit (described in more detail in Section 5.15).

The work presented in this chapter focuses on the dilepton analysis which targets events with exactly 2 high-p_{T} isolated leptons (either of which can be an electron or a muon) and at least 2 jets of which at least 2 are b-tagged. Channels that include τ-particles are targeted in another analysis. The full event selection is given in Section 5.8.

Events passing the analysis selection are divided amongst analysis regions according to their jet and b-tag multiplicity. In the dilepton analysis, 6 analysis regions are used: 4 "signal-depleted" control regions and 2 "signal-rich" regions. The signal-rich regions are defined as such because they have an $S/B > 1\%$ and $S/\sqrt{B} > 0.3$. All analysis regions that do not pass these requirements are said to be signal-depleted. Signal-rich regions provide most of the signal sensitivity whereas the control regions are used to model the background contributions and constrain systematic un-

	4 j, 2 b	4 j, 3 b	4 j, 4 b
$t\bar{t}H$ (125)	31 ± 3	13 ± 2	2.0 ± 0.3
$t\bar{t}$+ light	77 000 ± 7500	6200 ± 750	53 ± 12
$t\bar{t} + c\bar{c}$	4900 ± 3000	680 ± 390	21 ± 12
$t\bar{t} + b\bar{b}$	1800 ± 1100	680 ± 380	44 ± 25
W+jets	5100 ± 3000	220 ± 130	5.5 ± 3.3
Z+jets	1100 ± 600	50 ± 27	0.9 ± 0.6
Single top	4900 ± 640	340 ± 60	6.8 ± 1.6
Diboson	220 ± 71	11 ± 4.1	0.2 ± 0.1
$t\bar{t} + V$	120 ± 40	15 ± 5.1	0.9 ± 0.3
Lepton misID	1600 ± 620	100 ± 37	3.5 ± 1.3
Total	96 000 ± 9500	8300 ± 1100	140 ± 34
Data	98 049	8752	161
S/B	< 0.001	0.002	0.014
S/$\sqrt{\text{B}}$	0.099	0.141	0.167

	5 j, 2 b	5 j, 3 b	5 j, ≥ 4 b
$t\bar{t}H$ (125)	41 ± 2	23 ± 2	6.2 ± 0.8
$t\bar{t}$+ light	38 000 ± 5500	3500 ± 520	61 ± 15
$t\bar{t} + c\bar{c}$	4300 ± 2400	810 ± 460	43 ± 25
$t\bar{t} + b\bar{b}$	1700 ± 880	890 ± 480	110 ± 63
W+jets	1900 ± 1200	140 ± 87	5.9 ± 3.9
Z+jets	410 ± 240	29 ± 17	1.5 ± 0.9
Single top	1900 ± 360	190 ± 41	8.3 ± 1.3
Diboson	97 ± 39	8.0 ± 3.4	0.4 ± 0.2
$t\bar{t} + V$	150 ± 48	26 ± 9	3.1 ± 1.0
Lepton misID	460 ± 170	70 ± 28	8.3 ± 3.7
Total	49 000 ± 7000	5700 ± 980	250 ± 75
Data	49 699	6199	286
S/B	0.001	0.004	0.025
S/$\sqrt{\text{B}}$	0.186	0.301	0.397

	≥ 6 j, 2 b	≥ 6 j, 3 b	≥ 6 j, ≥ 4 b
$t\bar{t}H$ (125)	64 ± 5	40 ± 3	16 ± 2
$t\bar{t}$+ light	19 000 ± 4400	2000 ± 460	52 ± 17
$t\bar{t} + c\bar{c}$	3700 ± 2100	850 ± 480	79 ± 46
$t\bar{t} + b\bar{b}$	1400 ± 770	970 ± 530	250 ± 130
W+jets	910 ± 620	97 ± 66	8.6 ± 6.2
Z+jets	180 ± 120	19 ± 12	1.5 ± 1.0
Single top	840 ± 220	120 ± 35	12 ± 3.7
Diboson	50 ± 24	6.0 ± 3.0	0.5 ± 0.3
$t\bar{t} + V$	180 ± 59	45 ± 14	8.5 ± 2.8
Lepton misID	180 ± 66	21 ± 8	1.1 ± 0.5
Total	26 000 ± 5800	4200 ± 1000	430 ± 150
Data	26 185	4701	516
S/B	0.002	0.01	0.04
S/$\sqrt{\text{B}}$	0.393	0.63	0.815

Table 5.1: Pre-fit event yields in each of the single lepton analysis regions for signal, background and data. The quoted uncertainties are the sum in quadrature of the statistical and systematic uncertainties on the yields.

	2 j, 2 b	3 j, 2 b	3 j, 3 b
$t\bar{t}H$ (125)	1.5 ± 0.2	5.3 ± 0.5	2.2 ± 0.3
$t\bar{t}$+ light	$14\,000 \pm 1800$	8100 ± 880	96 ± 21
$t\bar{t} + c\bar{c}$	270 ± 170	600 ± 320	76 ± 44
$t\bar{t} + b\bar{b}$	150 ± 87	260 ± 130	120 ± 65
Z+jets	330 ± 30	190 ± 49	8.2 ± 3.1
Single top	430 ± 71	270 ± 30	7.6 ± 3.5
Diboson	6.8 ± 2.2	4.2 ± 1.5	$\leq 0.1 \pm 0.1$
$t\bar{t} + V$	8.4 ± 2.7	21 ± 6	1.9 ± 0.6
Lepton misID	21 ± 10	33 ± 17	0.8 ± 0.4
Total	$15\,000 \pm 1900$	9500 ± 1000	310 ± 85
Data	15 296	9996	374
S/B	< 0.001	0.001	0.006
$S/\sqrt{B}$	0.012	0.053	0.114

	$\geq$ 4 j, 2 b	$\geq$ 4 j, 3 b	$\geq$ 4 j, $\geq$ 4 b
$t\bar{t}H$ (125)	15 ± 1	8.6 ± 0.6	2.7 ± 0.3
$t\bar{t}$+ light	4400 ± 810	120 ± 31	1.9 ± 0.8
$t\bar{t} + c\bar{c}$	710 ± 380	130 ± 74	5.0 ± 3.0
$t\bar{t} + b\bar{b}$	290 ± 150	200 ± 100	31 ± 17
Z+jets	100 ± 39	10 ± 4	0.6 ± 0.2
Single top	140 ± 55	11 ± 5	0.8 ± 0.2
Diboson	4.0 ± 1.3	0.4 ± 0.1	$\leq 0.1 \pm 0.1$
$t\bar{t} + V$	45 ± 14	7.8 ± 2.4	1.1 ± 0.4
Lepton misID	38 ± 19	4.3 ± 2.2	0.4 ± 0.2
Total	5800 ± 1000	490 ± 140	43 ± 18
Data	6006	561	46
S/B	0.003	0.015	0.059
$S/\sqrt{B}$	0.197	0.365	0.401

Table 5.2: Pre-fit event yields in each of the dilepton analysis regions for signal, background and data. The quoted uncertainties are the sum in quadrature of the statistical and systematic uncertainties on the yields.

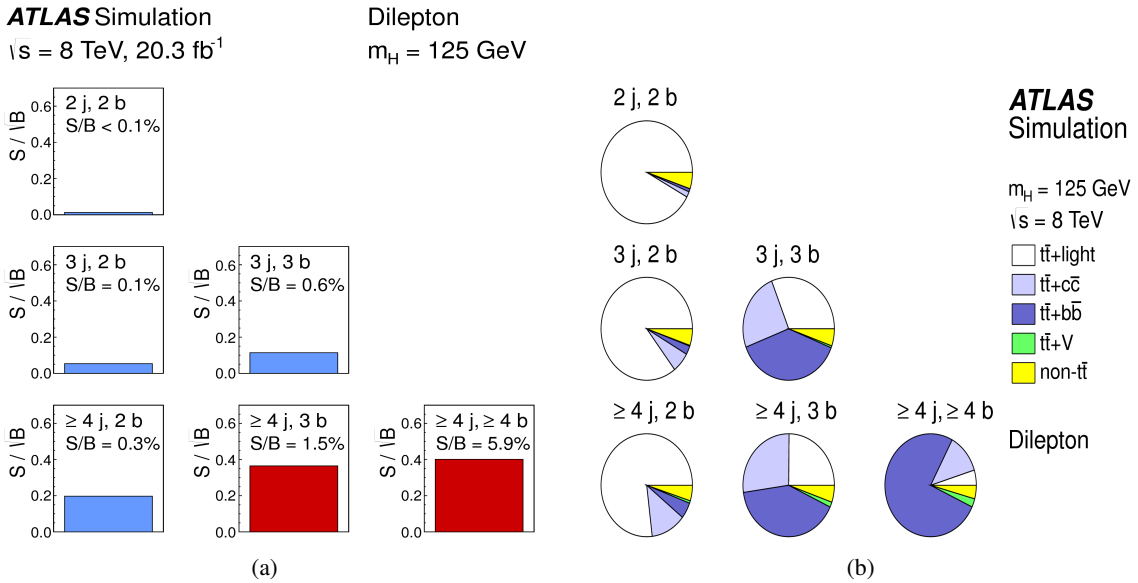


Figure 5.4: Distributions of (a) the $S/\sqrt{B}$ ratio for each of the dilepton analysis regions and (b) the relative background contributions in each dilepton analysis region [58]. Each row represents a jet multiplicity while each column represents a b-jet multiplicity. The red regions in (a) represent the signal regions while the blue regions represent the background regions.

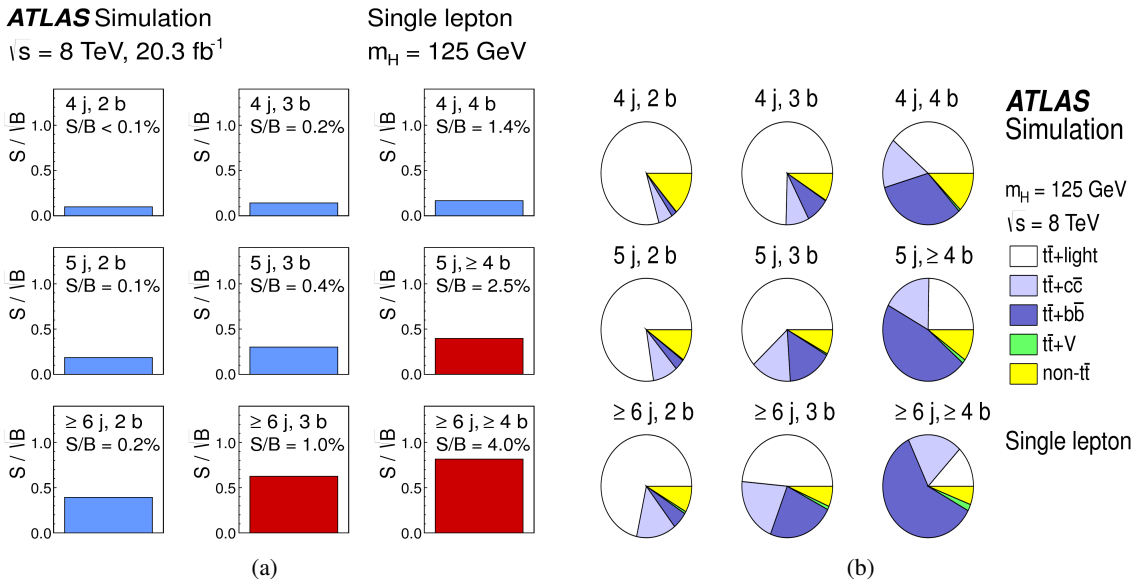


Figure 5.5: Distributions of (a) the $S/\sqrt{B}$ ratio for each of the single lepton analysis regions and (b) the relative background contributions in each single lepton region [58]. Each row represents a jet multiplicity while each column represents a b-jet multiplicity. The red regions in (a) represent the signal regions while the blue regions represent the background regions.

certainties.

A discriminant distribution is taken from each analysis region and used in the combined fit to test for the presence of a signal assuming a Higgs boson mass of 125 GeV. In the ($= 2\text{j} = 2\text{b}$), ($= 3\text{j} = 2\text{b}$) and ($\geq 4\text{j} = 2\text{b}$) regions the H_T distribution (see Section 4.4) is used whereas in the ($= 3\text{j} = 3\text{b}$), ($\geq 4\text{j} = 3\text{b}$) and ($\geq 4\text{j} \geq 4\text{b}$) regions neural network (NN) distributions are used (see Section 5.9). The neural network output distribution is used as an input to the fit. Each of these distributions has a well separated signal and background and therefore contributes to improving the sensitivity of the analysis. The 6 dilepton analysis regions and the distribution used in the likelihood fit are shown in Figure 5.6.

One of the main focuses of this chapter is a study of possible methods (pairing methods) that can be used to correctly assign which b-jets came from the Higgs boson and which came from the top quarks. This is important as the pairing can affect the shape and normalisation of some of the kinematic distributions that are used as inputs to the neural network. This will subsequently affect the ability of the neural network to separate signal from background. Several pairing methods are investigated and their reconstruction efficiencies are evaluated using the truth information in simulated samples. The effect of each pairing method on the kinematic distributions and the subsequent effect on the performance of the neural networks is also investigated. A further study

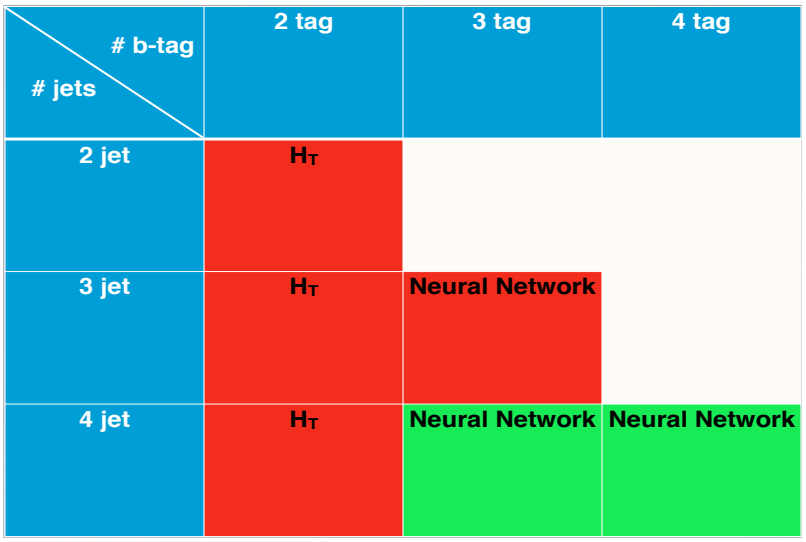


Figure 5.6: Table illustrating the 6 dilepton analysis regions along with the distribution used as input to the fit from each region. Signal regions are shown in green and background regions are shown in red.

comparing the performance of neural networks (NNs) (see Section 5.9) and boosted decision trees (BDTs) (see Section 5.12) is also presented. Because the performance of the multivariate classifier is dependent on the pairing method used, classifiers are compared when using the same pairing method to obtain a like-for-like comparison.

5.3 Statistical Analysis

The analysis presented uses a binned likelihood function constructed from the product of Poisson probabilities over all bins considered in the analysis. The likelihood function is dependent on the signal strength parameter μ, and a set of nuisance parameters θ. The signal strength parameter is defined as the ratio of the observed or expected $t\bar{t}H$ cross-section over the SM $t\bar{t}H$ cross-section. The nuisance parameters are introduced in the likelihood function as priors and incorporate the effects of the systematic uncertainties on the signal and background predictions. They act to adjust the expected signal and background predictions according to the corresponding systematic uncertainties. The resulting fitted values correspond to the amount that best fits the data as is shown by the pre-fit and post-fit event yields in Figures 5.2 and 5.3 respectively. This procedure reduces the impact of the systematic uncertainties of the analysis by taking advantage of more populated background-dominated control regions in the fit.

The likelihood is used to calculate the test statistic:

$$q_\mu = -2ln(\mathcal{L}(\mu, \hat{\hat{\theta}}_\mu)/\mathcal{L}(\hat{\mu}, \hat{\theta})) \tag{5.1}$$

where $\hat{\hat{\theta}}_\mu$ is the set of values of the nuisance parameters that maximises the likelihood function for a given value of μ, and $\hat{\mu}$ and $\hat{\theta}$ are the values of the signal strength and nuisance parameters that maximise the likelihood function with constraints $0 \leq \hat{\mu} \leq \mu$. The test statistic $q_0 = -2ln(\mathcal{L}(0, \hat{\hat{\theta}}_\mu)/\mathcal{L}(\hat{\mu}, \hat{\theta}))$ is used to evaluate the probability of obtaining a result as signal-like as this one if no signal were observed.

To obtain the final result, a simultaneous fit to data of the discriminant distributions in all analysis regions from both analysis channels is performed. Fits are performed under the signal-plus-background hypothesis where the parameter of interest, the signal strength parameter μ, is allowed to float freely in the fit. In the fit, the signal strength is constrained to be the same in all analysis regions. The normalisation of each of the backgrounds is determined from the fit at the same time as the fitted value of the signal strength parameter. Both theoretical and instrumental uncertainties are used to constrain the $t\bar{t}$, $t\bar{t}+V$, diboson, $W+$jets, $Z+$jets and single top components of the background. The relative statistical uncertainty in each bin of the discriminant distributions is also taken into account. The performance of the fit is tested by using simulated events and injecting $t\bar{t}H$ signal with a variable signal strength and comparing it to the fitted value.

5.4 Systematic Uncertainties

The systematic uncertainty in this analysis that has the largest effect on the fitted value of μ arises from the normalisation uncertainty of the irreducible $t\bar{t}$+$b\bar{b}$ background. This uncertainty is $\approx 50\%$ before the fit but is reduced by more than half of its original value by the fit. The $t\bar{t}$+$b\bar{b}$ background normalisation is pulled up by more than 40% by the fit to data, resulting in an increase in the observed $t\bar{t}$+$b\bar{b}$ yield with respect to the nominal prediction. The $t\bar{t}$+$b\bar{b}$ modelling uncertainties affecting the shape of this background, for example the renormalisation scale and parton shower recoil scheme, also affect the fitted value of μ to a large degree. These uncertainties are not correlated to other $t\bar{t}+$jets backgrounds, only to $t\bar{t}$+$b\bar{b}$. The largest of these uncertainties is given by the choice of renormalisation scale, which can drastically change the shape of the $t\bar{t}$+$b\bar{b}$ neural network distribution, making it look more signal like. The 15 most significant systematic

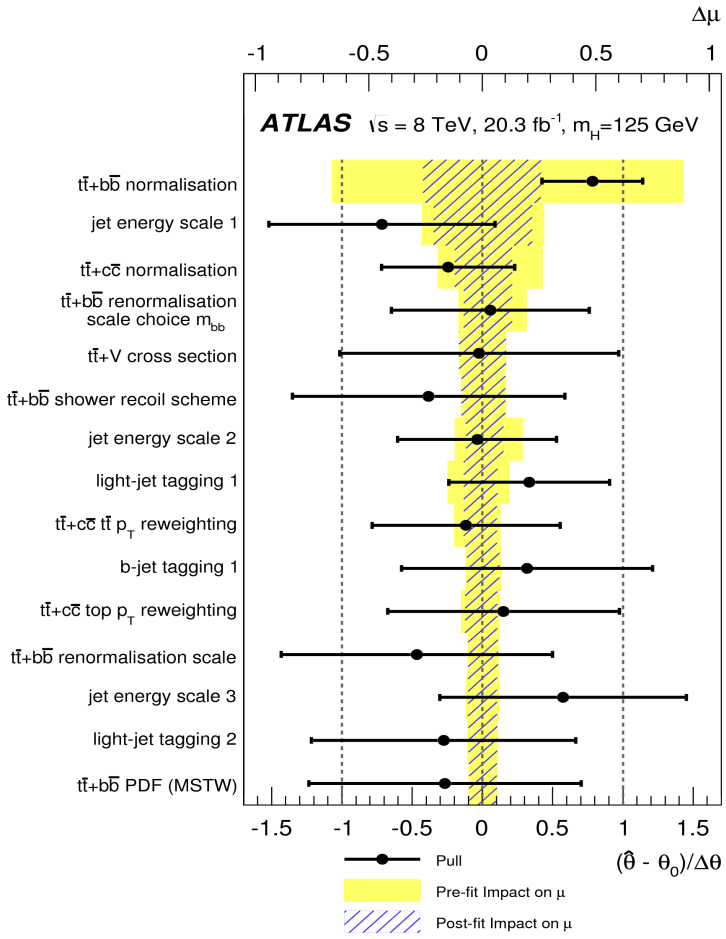


Figure 5.7: Fitted values of the nuisance parameters that have the largest impact on the parameter of interest - the measured signal strength μ_{ttH} [58]. The points represent the deviation of the fitted value of the nuisance parameter $\hat{\theta}$ from the nominal value of the nuisance parameter θ_0, in units of the pre-fit standard deviation $\Delta\theta$ and conform to the scale at the bottom of the plot. The error bars on the points show the post-fit uncertainty, σ_θ. If the data provide a reduction in the size of the uncertainty with respect to the original uncertainty, σ_θ will be less than 1. However, if the value is equal to 1, the data do not provide any further constraint on that uncertainty. The nuisance parameters are sorted such that the parameter with the largest post-fit impact on μ is at the top. The post fit impact on $\mu_{t\bar{t}H}$ is shown by the blue hatched area and conforms to the axis at the top of the plot.

uncertainties in this analysis are shown in Figure 5.7.

The calculation of an additional uncertainty associated with the choice of parton shower model for the $t\bar{t}H$ signal simulation is presented in Section 5.14 because, although it is a relatively small systematic uncertainty, it is original work that was performed entirely by the candidate. Kinematic distributions of the nominal POWHEL+PYTHIA signal sample are compared with the distributions obtained from a POWHEL+HERWIG $t\bar{t}H$ sample. An envelope is created by symmetrising the variation in the discriminant distribution between the nominal and POWHEL+HERWIG samples. The nominal is used as the reference and the systematic variation is introduced to the fit as an additional nuisance parameter which can change the normalisation and shape of the signal discriminant distributions.

The full list of systematic uncertainties considered in this analysis can be found in Appendix D. A further breakdown of the size of the effect of each systematic uncertainty in the (≥ 4j, ≥ 4b) region can be found in Table E.2. This gives a good idea of the relative importance of the uncertainty associated with the choice of parton shower model for the signal simulation (labelled 'ttH-PartonShower' in the Table).

Final results from the dilepton analysis are subsequently combined with the single lepton analysis to obtain the final "combined" $t\bar{t}H(H\rightarrow b\bar{b})$ result. Separate results from the dilepton and single lepton analyses along with the combined results are discussed at greater length in Section 5.15.

5.5 Analytical Challenges

This analysis is clearly a complicated analysis that requires a great deal of understanding of background processes. Our understanding of the shape and size of backgrounds is improved through the use of well defined control regions. An in-depth study of variables that are able to discriminate between different processes is required to fully harness the power of the combined likelihood fit. Furthermore, an improved 'pairing' method can help to understand signal and background processes. Section 5.11.2 addresses these questions in the analysis' signal region in an original study performed by the candidate.

b-tagging is also an essential part of this analysis, which allows requiring a given number of b-tagged jets in an event and subsequently reducing many background processes. However, the b-tagging of jets comes with its own associated uncertainty. This is $\approx 1\%$ per b-tag and hence the b-tagging uncertainty tends to be higher for regions in the analysis that require more b-tags. The

same b-tagging algorithm as is described in Chapter 4 is used in this analysis (the MV1 algorithm). However, the analysis uses the combinatorial likelihood b-tagging calibration method described in Reference [34] to obtain b-tagging scale-factors and efficiencies.

Furthermore, requiring high jet and high b-tag multiplicities dramatically increases the statistical uncertainty of the analysis and this is accentuated by the division of events into analysis regions. This problem is overcome by using a technique known as the 'tag rate function' or 'TRF' for short. TRF allows one to use every event in the sample before b-tagging is applied to predict the normalisation and shape of distributions after b-tagging, thus bolstering the statistics of the analysis and reducing the statistical uncertainty [60]. This is achieved by, instead of b-tagging the jets according to some b-tagging algorithm, assigning to each jet a weight associated with its probability of being b-tagged, which is parametrised as a function of p_{T} and η. For example, events with 4 jets and 2 b-tags are also used to predict the shapes and normalisations in the region with 4 jets and 4 b-tags. However, they will be assigned a lower weight than for example a true event with 4 jets and 4 b-tags and therefore have less impact in this region. Tagging probabilities are derived using an inclusive $\mathrm{t\bar{t}}$+jets sample. Since the b-tagging probability for a b-jet coming from the decay of a top quarks is slightly higher than that of a b-jet originating from another source, they are derived separately. The normalisation and shape of predictions using TRF agree well with those when using the b-tagging algorithm ('direct b-tagging'). This method is then applied consistently to all background and signal samples.

5.6 Simulation Samples

The main background processes after the event selection are $\mathrm{t\bar{t}}+\mathrm{jets}$, $\mathrm{Z}+\mathrm{jets}$, diboson, single top Wt-channel and $\mathrm{t\bar{t}}+\mathrm{V}$ (where 'V' is either a W or Z boson), all of which can produce two prompt leptons and are simulated using Monte Carlo. Events can also have non-prompt leptons or fake leptons that pass all the lepton object selection criteria. These come from $\mathrm{W}+\mathrm{jets}$, $\mathrm{t\bar{t}}$ single lepton channel or single top *s*- or *t*-channel processes. Yields for these processes are obtained from simulation and cross-checked with data-driven techniques. However, after requiring two b-tagged jets, the number of events passing the selection is negligible. All MC samples in this analysis were simulated with a top mass of 172.5 GeV and a Higgs boson mass of 125 GeV.

The main background for this search channel is $\mathrm{t\bar{t}}+\mathrm{jets}$. These events are included in the $\mathrm{t\bar{t}}$ sample which is generated using the POWHEG package [42] which is an NLO MC generator. The

showering and hadronisation of the parton shower is performed using the PYTHIA package. The sample is normalised to the theoretical cross-section calculations performed at NNLO [61].

The $t\bar{t}$ sample is inclusive in jet flavour but is divided into sub-categories that depend on the flavour of the partons that are matched to the particle jets that do not originate from the decay of the $t\bar{t}$ system. Particle jets are reconstructed by clustering stable particles, not including muons and neutrinos, using the anti-k_t algorithm with a radius parameter of 0.4 and are required to have $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ and $|\eta| < 2.5$. The matching of particle jets to partons is done by requiring $\Delta R < 0.4$ between the jet and the parton. Events with at least one b-quark not originating from the $t\bar{t}$-system, that is matched to a particle jet are labelled $t\bar{t}$+$b\bar{b}$. If an event is not labelled as such but at least one particle jet is matched to a c-quark not originating from the decay of a W boson, it is labelled $t\bar{t}+c\bar{c}$. $t\bar{t}$+$b\bar{b}$ and $t\bar{t}+c\bar{c}$ events are labelled under the umbrella term $t\bar{t}+\mathrm{HF}$ (Heavy Flavour). Any events which do not fall into the $t\bar{t}+\mathrm{HF}$ bracket are labelled $t\bar{t}+\mathrm{light}$, which includes $t\bar{t}$ events which have no additional jets.

Fully matched NLO predictions became available within SHERPA [62] plus OPENLOOPS (SHERPAOL) program [63]. The prediction from SHERPAOL is expected to model the $t\bar{t}$+$b\bar{b}$ contribution more accurately than POWHEG+PYTHIA. For the purposes of comparing the $t\bar{t}+\mathrm{jets}$ events in both generators (and propagating $t\bar{t}+\mathrm{HF}$ systematics) the $t\bar{t}+\mathrm{HF}$ events are sub-divided into two further categories. If two particle jets are both matched to b- or c-quarks not originating from the $t\bar{t}$-system, the event is labelled $t\bar{t}$+$b\bar{b}$ or $t\bar{t}+c\bar{c}$ respectively. If only one particle jet is matched to a b- or c-quark, the event is labelled $t\bar{t}+\mathrm{b}$ or $t\bar{t}+\mathrm{c}$ respectively and if a single particle jet is matched to a $b\bar{b}$ or $c\bar{c}$-pair, then the event is labelled $t\bar{t}+\mathrm{B}$ or $t\bar{t}+\mathrm{C}$ respectively.

In particular, SHERPAOL predicts a larger fraction of $t\bar{t}+\mathrm{jets}$ events coming from $t\bar{t}+\mathrm{B}$ events than POWHEG+PYTHIA. It also predicts larger fractions for any event in which an additional $b\bar{b}$-pair is produced.

The kinematic distributions of PYTHIA and SHERPAOL are similar except in the case of the very low mass and p_{T} region of the $b\bar{b}$ system. Further differences are seen in the p_{T} distributions of the $t\bar{t}$-system and top quark. Therefore a reweighting is applied to $t\bar{t}$+$b\bar{b}$ events in the POWHEG+PYTHIA sample, to correct these distributions using the best available theory correction at the time of the analysis [64]. This is used in the SHERPAOL program [62]-[63]. The reweighting is applied at generator level using distributions of the top-quark p_{T}, the p_{T} of the $t\bar{t}$-system and the ΔR and p_{T} of the dijet system not coming from the top quark decays. After the reweighting is

applied, a comparison of the relative contributions of the different categories of $t\bar{t}$+$b\bar{b}$ events in the POWHEG+PYTHIA sample is compared with the SHERPAOL. The relevant disagreements seen in the kinematic distributions are also checked.

A different reweighting is applied to $t\bar{t}+c\bar{c}$ and $t\bar{t}+$light events in POWHEG+PYTHIA which uses the ratio of the measured differential cross-section at $\sqrt{s} = 7$ TeV in data and simulation as a function of the top quark and $t\bar{t}$-system p_{T} [65]. It was verified using the simulation that the ratio derived at $\sqrt{s} = 7$ TeV is applicable at $\sqrt{s} = 8$ TeV [4].

$Z+$jets and diboson samples were produced using the ALPGEN LO generator. The parton shower evolution and hadronisation are performed by PYTHIA for $Z+$jets and HERWIG for diboson samples. The final states with hadronic Z decays of WZ samples are missing in the ALPGEN+HERWIG samples so SHERPA samples were generated for these decay modes. The $Z+$jets samples are generated with up to 5 additional partons and separated into parton flavours all of which are normalised to their NNLO theoretical cross-sections [66]. The diboson samples are generated with up to three additional partons which are normalised to the NLO theoretical cross-section calculation [67]. All event generators using HERWIG are also interfaced with JIMMY [68] to simulate the underlying event.

The single top background simulation samples were generated using the POWHEG package interfaced with PYTHIA which performs the parton shower evolution and hadronisation. Overlaps between diagrams from $t\bar{t}$ and single top final states are removed and samples are normalised to the NNLO theoretical cross section [69, 70, 71]. $t\bar{t}$+V samples were generated using the MADGRAPH LO generator using the PYTHIA package for showering and are normalised to NLO theoretical cross sections [72, 73].

The $t\bar{t}H$ signal process is modelled on NLO matrix elements using the HELAC-Oneloop package [74]. The POWHEG package is used to interface between the matrix element and the shower simulation package. Samples produced in such a way are referred to as Powhel samples. The Higgs boson decay in these samples is inclusive [75] and the $t\bar{t}H$ samples are normalised to their NLO theoretical cross-sections. The Higgs boson decay branching ratios and NLO theoretical cross sections are taken from References [76, 77].

All simulated samples are then processed through a full simulation of the ATLAS detector geometry and response using GEANT4. Simulation and data samples are processed through the same reconstruction software. Simulated events are then corrected so that object identification

efficiencies, energy scales and energy resolutions are the same as in data.

5.7 Data Samples

This analysis uses 20.3 fb^{-1} of data collected with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$. All events were collected using the single electron or single muon trigger under stable beam condition and for which all subsystems of the detector were fully functional. Both low p_{T} and a high p_{T} leptonic threshold triggers are used. The isolation requirements on the candidate lepton for the low p_{T} triggers result in inefficiencies at high p_{T} which are recovered by the high p_{T} trigger. For electrons the p_{T} thresholds are set at 24 and 60 GeV and for muons these are set at 24 and 36 GeV. The exact details of the triggers varied during the data taking period but all triggers have a similar but looser selection than the offline selection.

5.8 Event Selection

After passing the trigger selection, events are required to have at least one reconstructed vertex which has 5 associated tracks consistent with the beam collision region in the *x*-*y* plane. Where more than one vertex is found the primary vertex (PV) is taken as the vertex with the largest sum of the squared momenta of associated tracks (see JVF description in Section 4.9). Events are rejected if any jet in the event with $p_{\text{T}} > 20\,\text{GeV}$ is identified as coming from out-of-time activity (an interaction that is not consistent with the timing of the hard collision creating the PV) or calorimeter noise [78].

All jets used in the analysis are required to have $p_{\text{T}} > 25\,\text{GeV}$ and $|\eta| < 2.5$. Jets which have $p_{\text{T}} < 50\,\text{GeV}$ and $|\eta| < 2.4$ are required to have a JVF greater than 0.5 to avoid selecting jets from secondary proton-proton interactions. The leading lepton (defined as the lepton which has the highest transverse momentum) is required to have $p_{\text{T}} > 25\,\text{GeV}$ and the sub-leading lepton (defined as the lepton with the second highest transverse momentum) is required to have $p_{\text{T}} > 15\,\text{GeV}$. Events are required to have two oppositely charged leptons and at least two jets. At least one of the jets is required to be b-tagged by the MV1 b-tagging algorithm.

At this point, the type of the selected leptons is used to decide which selection cuts must be satisfied. If the event is in the eμ-channel the scalar sum of the p_{T} of the leptons and jets in the event (H_{T}) is calculated and is required to be greater than 130 GeV. This rejects $\text{Z} \to \tau\tau$+jets background

events which are expected to have a lower H_{T}. In the ee- and $\mu\mu$-channels however, the dilepton invariant mass (m_{ll}) is required to be greater than 60 GeV for events with two b-tags and greater than 15 GeV in the other regions. This is used to suppress the weak decay of J/ψ ($c\bar{c}$) and Υ ($b\bar{b}$) particles which would produce the same ee or $\mu\mu$ signature. A further dilepton Z-mass peak cut is used to reject Z + jets events. This rejects events with a dilepton invariant mass within 8 GeV of the Z boson mass peak: $|m_{ll} - m_Z| > 8$ GeV. Events with more than 2 leptons are rejected to remove possible overlap with other $t\bar{t}H$ decay channels with multilepton final states. Events which pass this selection are then divided into categories according to the number of jets and number of b-tagged jets in the event.

For each jet in an event, the MV1 algorithm will output a b-tag weight. In the direct-tag approach to b-tagging, a jet is btagged if its weight is higher than the weight that corresponds to the 70% working point. This means that the probability of correctly b-tagging a true b-jet is approximately 0.7 (70% is the average efficiency across several jet p_{T} bins). For the (4j,4b) region, 4 jets in the event must pass this cut, meaning the probability of each true b-jet being b-tagged can be multiplied together to give the total probability of a true (4j,4b) event passing the 4 b-tag selection requirement. From this one can see how by requiring events with large b-tag multiplicities can statistically limit an analysis when using the direct tag approach.

Statistical limitations cause large fluctuations in the kinematic distributions of the simulated samples. The large statistical uncertainties and subsequently unreliable systematic uncertainties, negatively effect the sensitivity of the analysis. A solution to this problem was found by using another approach to b-tagging called the tag rate function (TRF) method. Rather than requiring a particular b-tagging weight for each jet in the event and removing events with too few b-tagged jets, the probability of each jet being b-tagged is parametrised as a function of jet flavour, p_{T} and η. These weights are then combined multiplicatively into an overall event weight. This allows all events in the sample before the b-tagging requirements are implemented, to be used in predicting the normalisation and shape of the distributions after b-tagging. The shapes and normalisations of kinematic distributions obtained using the TRF method agree well with those obtained when applying the direct b-tagging method.

The expected signal and background for dilepton events in each of the analysis regions is shown in Figure 5.4 along with the relative contributions from each of the backgrounds. From these diagrams it becomes clear that the two signal regions ($\geq 4\,\mathrm{j}$, $\geq 4\,\mathrm{b}$) (4j,4b) and $\geq 4\,\mathrm{j} = 3\,\mathrm{b}$

	2 j, 2 b	3 j, 2 b	3 j, 3 b
$t\bar{t}H$ (125)	2.4 ± 1.8	8.1 ± 5.9	3.0 ± 2.2
$t\bar{t}$+ light	14 000 ± 160	8300 ± 170	84 ± 9.6
$t\bar{t}+c\bar{c}$	400 ± 110	700 ± 160	92 ± 22
$t\bar{t}+b\bar{b}$	190 ± 36	350 ± 49	140 ± 19
Z+jets	330 ± 22	200 ± 43	7.3 ± 2.4
Single top	430 ± 35	260 ± 21	7.6 ± 1.5
Diboson	6.8 ± 2.1	4.5 ± 1.4	≤ 0.1 ± 0.1
$t\bar{t}+V$	8.7 ± 2.7	21 ± 6	1.8 ± 0.6
Lepton misID	19 ± 10	30 ± 15	0.7 ± 0.4
Total	15 000 ± 120	9900 ± 82	340 ± 14
Data	15 296	9996	374

	≥ 4 j, 2 b	≥ 4 j, 3 b	≥ 4 j, ≥ 4 b
$t\bar{t}H$ (125)	22 ± 16	11 ± 8	3.1 ± 2.3
$t\bar{t}$+ light	4500 ± 150	100 ± 12	1.4 ± 0.3
$t\bar{t}+c\bar{c}$	740 ± 170	140 ± 30	4.8 ± 1.1
$t\bar{t}+b\bar{b}$	370 ± 59	230 ± 31	30 ± 4
Z+jets	100 ± 33	9.5 ± 3.1	0.4 ± 0.2
Single top	140 ± 23	11 ± 2	0.6 ± 0.1
Diboson	4.2 ± 1.3	0.3 ± 0.1	≤ 0.1 ± 0.1
$t\bar{t}+V$	43 ± 13	7.0 ± 2.1	0.9 ± 0.3
Lepton misID	34 ± 18	3.5 ± 1.8	0.2 ± 0.1
Total	5900 ± 65	520 ± 18	42 ± 4
Data	6006	561	46

Table 5.3: Post-fit event yields for signal, backgrounds and data (20.3 fb^{-1}) in each of the analysis regions. The quoted uncertainties are the sum in quadrature of the statistical and systematic uncertainties on the yields.

(4j,3b) are not signal rich, but have a relatively large fraction of events coming from the $t\bar{t}H$ signal process. Pre-fit and post-fit event yields are shown in Figures 5.2 and 5.3 respectively.

5.9 Neural Networks

Given the large number of physics objects in the $t\bar{t}H(H\to b\bar{b})$ final state, how signal-like an event looks is dependent on many variables. Therefore, using a linear cut on a single variable to separate signal and background events does not provide the best separation power. Furthermore, due to the relatively limited signal statistics ($\approx O(10)$ events in combined signal regions) it is preferable to retain as many signal events as possible when separating background and signal. Simply relying on one discriminant may disregard events that one variable will classify as background but others

may classify as signal. In such cases, multiple variables are used simultaneously to make a better informed decision. In this analysis a multivariate algorithm called a neural network is trained to separate background events from signal events in the regions where signal is expected to be most significant.

A neural network is a multivariate classifier that takes an array of several input variables and outputs a prediction as to whether the event was signal or background. Its structural make up is a series of layers of neurons interconnected by synapses where each neuron in the input layer represents a single input variable. Considering just one input variable as shown in Figure 5.8, when an event is passed to the neural network the value of the input variable a_i is passed along the connecting synapses where it is multiplied by a synapse specific weight $w_{i,j}$, and output to the hidden layer. The hidden layer neuron sums together the values from all of its input synapses z_j and applies an activation function g_j. The result a_j is then passed on to its output synapse which then applies another weight $w_{j,k}$ and passes this weighted value to the output neuron. The output neuron sums together the values from all of its input synapses z_k and applies an activation function g_k before it outputs a prediction y, as to whether the event is signal or background. In the simplified example in Figure 5.8, the predicted value is calculated as shown in Equation 5.2. The indices i, j, and k refer to the input layer, the hidden layer and the output layer respectively.

$$y(g_k) = g_k\left(\sum_j g_j\left(\sum_i a_i w_{i,j}\right) w_{j,k}\right) \tag{5.2}$$

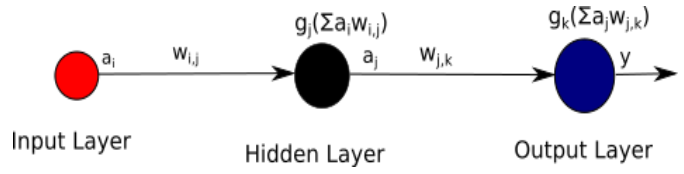


Figure 5.8: Example of a neural network with 1 input neuron, 1 neuron in the hidden layer and 1 output neuron. Here: $z_j = \sum_i a_i w_{i,j}$ with $a_j = g_j(z_j)$ and $z_k = \sum_j a_j w_{j,k}$ and finally $y = g_k(z_k)$. Each of these terms are used to construct Equation 5.2.

The networks prediction for an event y is compared with the expected classification $\hat{y}$ which will equal +1 for signal events and 0 for background events. The predicted value is scaled with respect to the maximum possible value y_{max} so that it is somewhere between 0 and +1. Considering

just a single event, the loss function is evaluated for the NN prediction. y and $\hat{y}$ are used to calculate the loss function which in this case is the "Entropy" loss function:

$$J = \frac{1}{m}\sum_{k=1}^{m}(\hat{y}_k ln(y_k) + (1-\hat{y}_k) ln(1-y_k). \tag{5.3}$$

Here, 'k' is cycling through 'm' training iterations.

The loss function provides a measure of how good the NN is at predicting the true classification. The neural network is then trained by finding the combination of weights that minimises this loss function. The number of neurons in each layer and the number of layers are hyperparameters of the network and do not change during training.

A fast way of finding the best combination of weights is called is called gradient descent. Considering a multidimensional phase space where each dimension represents the values taken by a single weight as shown in 5.9, by differentiating the loss function 'J' with respect to each of the weights an algorithm can quickly find the best direction to move in order to reach the minima quickly.

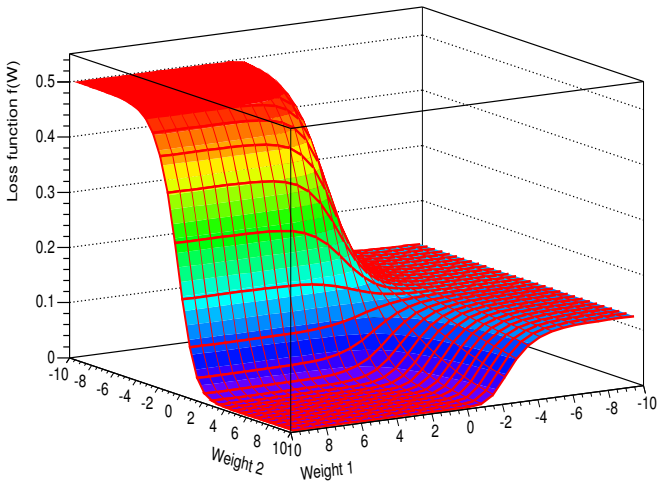


Figure 5.9: Example of the loss function surface for the simplified network in Figure 5.8. The activation functions g on the hidden layer node and output node are both sigmoid functions $1/(1+exp(-x))$ and the network is designed to return a value of 1 if given input 1. The neural network tries to find the minima in this phase space. Due to the sigmoid activation functions, large positive values of the weights provide the minimum loss.

A procedure called back propagation is used to update the network weights. The error contributed by each weight is propagated back through the neural network from the output node.

Weights that contribute more to the overall error are changed more. Each weight in the network is updated according to Equation 5.4 where weights $w_{i,j}$ represent the weights in Figure 5.8 and η is the learning rate of the network (parameter space step-size).

$$w_{i,j} = w_{i,j} - \eta(\delta J/\delta w_{i,j}) \tag{5.4}$$

The next event will then be evaluated with the new network. The procedure is repeated iteratively until the error reaches a tolerably low value, the parameters cease to update (converge) or a set number of parameter updates is reached.

An example of a network that is more representative of the NN used in this analysis is shown in Figure 5.10. Now, each layer consists of an array of neurons where every neuron in a layer has a synapse connecting it with every neuron in the next layer.

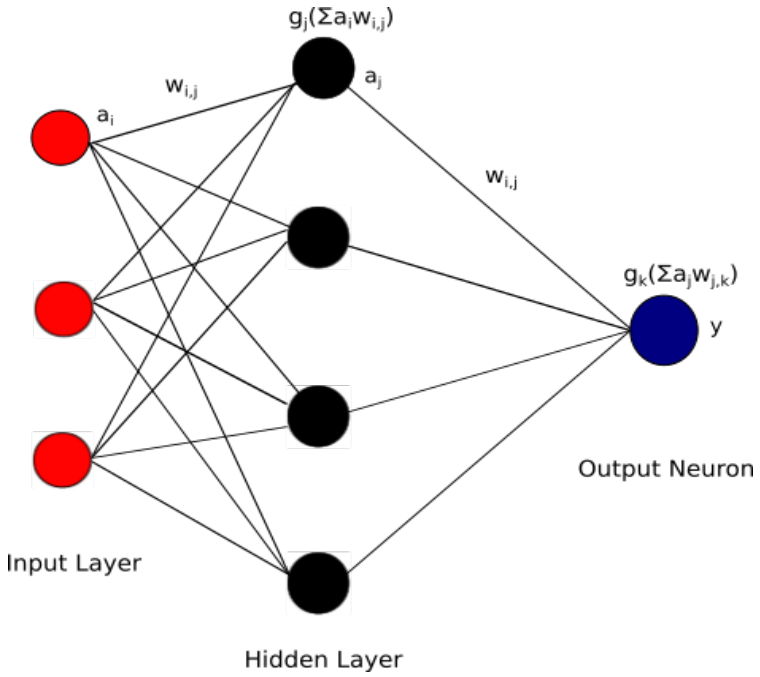


Figure 5.10: Example of a neural network with 3 input neurons, 4 neurons in the hidden layer and an output neuron.

During the training of the network, the loss function should get smaller with each iteration and will eventually plateau. To ensure the network is not overtrained, a distribution of the loss function with respect to the number of iterations is made for the networks performance on a training sample and testing sample. If the network is being overtrained, the loss function on the training sample

will become significantly lower than for the testing and the training should be stopped. Another test of overtraining is to compare the response distributions from the training and testing samples. This is shown in e.g. Figure 5.24 and is also used for the BDT (see Section 5.12) as shown in e.g. Figure 5.29. The network training used in the analysis showed no significant signs of over-training.

The final neural network is an optimised set of weights and activation functions. Once this is achieved, the neural networks configuration is frozen. The set of weights and activation functions can be used on any set of events and will give a prediction as to whether each event is signal or background. The response value of the NN is transformed according to Equation 5.5 which makes the NN distribution easier to interpret by forcing the values of y between -1 and +1. In this analysis, the frozen network is used on data to separate background and signal and the response distributions from all signal regions are combined with the discriminant distributions in the background regions to test for the presence of a 125 GeV Higgs boson signal.

$$R = 2.0/(1.0 + exp(-2.0 \cdot y(x))) - 1.0 \tag{5.5}$$

5.10 Analysis Method

Three neural networks were trained in three different analysis regions: $(\geq 4\text{j}, \geq 4\text{b})$, $\geq 4\text{j} = 3\text{b}$ and $(= 3\text{j} = 3\text{b})$(3j,3b), as each adds signal sensitivity in the analysis. All other regions in the analysis have a lower sensitivity as seen in Figure 5.4. For those regions, the scalar sum of the transverse momentum of all jets and leptons in the event (H_{T}) is used instead as the discriminant.

The NeuroBayes NN package [79] is the preferred multivariate classifier because of its robustness against over-training. The NeuroBayes package provides additional features which improve the performance of the tool. This includes an internal ranking procedure, which is used to rank the input variables by their discrimination power, and a sophisticated preprocessing of input variables.

Preprocessing input variables before training the NN is key to the success of NeuroBayes. The same preprocessing is used for all variables in this analysis. Distributions are first flattened and the mean is shifted to zero. They are then fitted using a function constructed from several polynomial functions (a spline). NeuroBayes uses the value which comes out of the fitted spline. Preprocessing essentially finds the optimal starting point for the subsequent network training.

NeuroBayes has separate training and testing phases. In the training phase an array of variables is passed to the network. The choice of variables used in the analysis is made by the ranking

procedure available in the NeuroBayes package which is based on the separation power of the input variables. An extensive list of variables based on object kinematics, global event variables, event shape variables and properties between pairs of objects is input to the neural network. NeuroBayes transforms each of the variables to a Gaussian of unit width and removes one variable at a time. When each variable is removed, the loss of discrimination power is calculated and the variable is then replaced into the list of variables. This is repeated for every variable in the list and the variable showing the smallest loss is removed. The last remaining variable will be the highest ranking variable. This process is repeated in order to rank the top 10 variables. Only the best 10 variables are used as this was deemed sufficient to achieve a signal-to-background separation close to optimal [58]. Variables which are highly correlated are removed along with variables whose distributions showed large differences between $\mathrm{t\bar{t}}$ MC generators.

A different set of variables is chosen for each region's neural network. All of the variables used in the dilepton analysis are shown in Table 5.4. In the $(\geq 4\,\mathrm{j}, \geq 4\,\mathrm{b})$ region, events are limited to a maximum of 5 jets that can be used to construct the kinematic variables (for example H_{T}). The b-tagged jets are selected first, before filling up the remaining jet vacancies with the highest p_{T} non b-tagged jets. For events with more than 5 b-tagged jets, those with the highest p_{T} are selected. For variables that are dependent on the reconstruction of the Higgs boson candidate, the b-jets originating from the top quarks in a $(\geq 4\,\mathrm{j}, \geq 4\,\mathrm{b})$ event are assigned as such by finding the two b-jets which are closest to the leptons in the event. The remaining two highest p_{T} b-jets are assigned to the Higgs boson candidate. For the other analysis regions, events with at least 4 jets but less than 4 b-tagged jets include the remaining untagged jets in the calculation of this variable. If the analysis region requires less than 4 b-tags, such variables are not used.

The procedure by which the b-jets in an event are assigned to the Higgs boson is important as it can affect the normalisation and shape of the kinematic distributions that are used as input to the neural network. This can hinder or enhance the classifiers ability to distinguish between signal and background and therefore affects the sensitivity of the analysis. Several methods for selecting the b-jet pair originating from the Higgs boson decay are investigated in Section 5.11.

Variable Name	Description	4j,4b	4j,3b	3j,3b
$\Delta\eta_{jj}^{max}$	Maximum $\Delta\eta$ between any two jets in the event.	1^{st}	1^{st}	1^{st}
$m_{bb}^{min\ \Delta R}$	Invariant mass of the two b-tagged jets with the smallest $\Delta R_{b\bar{b}}$.	2^{nd}	8^{th}	-
m_{bb}	Mass of the two b-tagged jets from the selected Higgs boson candidate system.	3^{rd}	-	-
ΔR_{hl}^{min}	ΔR between the selected Higgs boson candidate and the closest lepton.	4^{th}	5^{th}	-
N_{30}^{Higgs}	Number of b-jet pairs in an event with an invariant mass within ± 30 GeV of 125 GeV.	5^{th}	2^{nd}	5^{th}
$\Delta R_{bb}^{max\ p_{\mathrm{T}}}$	ΔR between the pair of b-tagged jets which have the largest vector sum p_{T}.	6^{th}	4^{th}	8^{th}
$Aplan_{jet}$	$1.5\lambda_2$, where λ_2 is the second eigenvalue of the momentum tensor built from all jets.	7^{th}	7^{th}	-
m_{jj}^{min}	Minimum dijet mass of any jet pair.	8^{th}	3^{rd}	2^{nd}
ΔR_{hl}^{max}	ΔR between the Higgs boson candidate and the furthest lepton.	9^{th}	-	-
$m_{jj}^{closest}$	Dijet mass of the two jets with invariant mass closest to 125 GeV	10^{th}	-	10^{th}
H_{T}	Scalar sum of the p_{T} of the jets and both leptons	-	6^{th}	3^{rd}
$\Delta R_{bb}^{max\,m}$	ΔR between two b-tagged jets with largest invariant mass	-	9^{th}	-
ΔR_{lj}^{min}	Minimum ΔR between a lepton and a jet	-	10^{th}	-
Centrality	Sum of the p_{T} divided by the sum of the E for all jets and both leptons	-	-	7^{th}
$m_{jj}^{max\,p_{\mathrm{T}}}$	Mass of the combination of the two jets with the largest p_{T} vector sum	-	-	9^{th}
$H4$	Fifth Fox-Wolfram moment computed using all jets and both leptons	-	-	4^{th}
$p_{\mathrm{T}}^{jet\,3}$	p_{T} of the third leading jet	-	-	6^{th}

Table 5.4: NeuroBayes's ranking of the top ten variables used as input to the $t\bar{t}H$ dilepton neural networks along with their descriptions.

5.11 Identifying b-jets Originating from the Higgs Boson

Constructing events in the $t\bar{t}H$ channel entails assigning b-jets to either the top quarks or the Higgs boson. The large number of final state b-jets, as well as the fact that, due to imperfect acceptance, not all of the products of the $t\bar{t}H(H\rightarrow b\bar{b})$ decay will necessarily be present, only a relatively low matching efficiency is achieved. As mentioned in Section 5.10, the current method of constructing an event is to find the two b-jets closest to the leptons and assign them to the top quarks; the two remaining b-tagged jets are assigned to the Higgs boson. However, there are many ways to construct an event, some of which have a better efficiency than the method currently used in this analysis. Here, we often use the word "pairing" to refer to an assignment of the reconstructed jets to the top quarks, or the Higgs boson. Correctly constructing events shows a better understanding of signal and background processes and may enable the construction of new discriminating variables, e.g. colour flow variables. This section is devoted to the investigation of alternative pairing methods and their impact on the power to discriminate between signal and background. Higgs boson candidate variables are only applicable to the $(\geq 4\,\mathrm{j} \geq 4\,\mathrm{b})$ analysis region as the signal signature typically contains 4 jets all of which are expected to be b-jets.

5.11.1 Truth Matching

The efficiency of a pairing method is calculated as the number of times it selects the correct Higgs boson b-jets with respect to the total number of events it was possible to obtain such a match in simulation. The baseline definition that defines whether an event is matchable or not is made using truth information e.g. flavour, parent particle, decay particles, etc. The truth matching is performed in two stages. Firstly, the truth information is used to ensure that an event is a true $t\bar{t}H(H\rightarrow b\bar{b})$ dilepton event. This means that the event is checked for two true top quarks that decay to a b-quark, a leptonically decaying W boson, and a Higgs boson that decays to a $b\bar{b}$ pair. Upon finding a true $t\bar{t}H(H\rightarrow b\bar{b})$ event, each b-quark is matched with a b-jet.

For a given set of truth b-partons and reconstructed b-jets, the ΔR between each parton-jet pair is calculated along with the sum of the ΔR values for all pairs in a given permutation. This is then repeated for every possible permutation. The permutation for which the $\sum_{i=1}^{4} \Delta R_i$ is minimised is considered the best permutation. This method is illustrated in Figure 5.11.

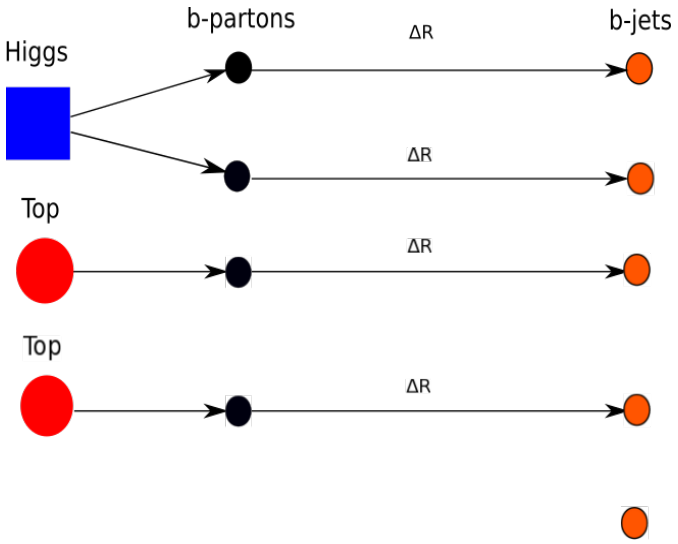


Figure 5.11: A schematic of the truth-matching procedure used to find the best permutation.

One also needs to know which are the 'correct' Higgs boson b-jets. The correct Higgs boson b-jets are the b-jets in the best permutation that can be well matched to the true Higgs boson b-partons. A b-jet is well matched to a b-parton if it is within $\Delta R < 0.4$. If either an event is not a true $t\bar{t}H(bb)$ dilepton event, or one or both of the assigned Higgs boson b-jets are not well matched, the event is discarded and not taken into account when calculating the efficiency of the pairing algorithms. Table 5.5 shows the number of events that are matchable along out of those that pass the event selection. 341 events are lost between the first and second rows in Table 5.5 because they are not true $t\bar{t}H(H\rightarrow b\bar{b})$ dilepton events. These are mostly $H \rightarrow WW$ events as shown in Figure 5.12.

Truth matching cutflow	
Events that pass full event selection	10162
True $t\bar{t}H(H\rightarrow b\bar{b})$ dilepton events	9821
True $t\bar{t}H(H\rightarrow b\bar{b})$ dilepton events with two truth matched Higgs boson candidate b-jets	8481

Table 5.5: Number of events failing at each stage of the truth matching procedure. All events that reach this stage of the selection will have passed the event selection described in Section 5.8.

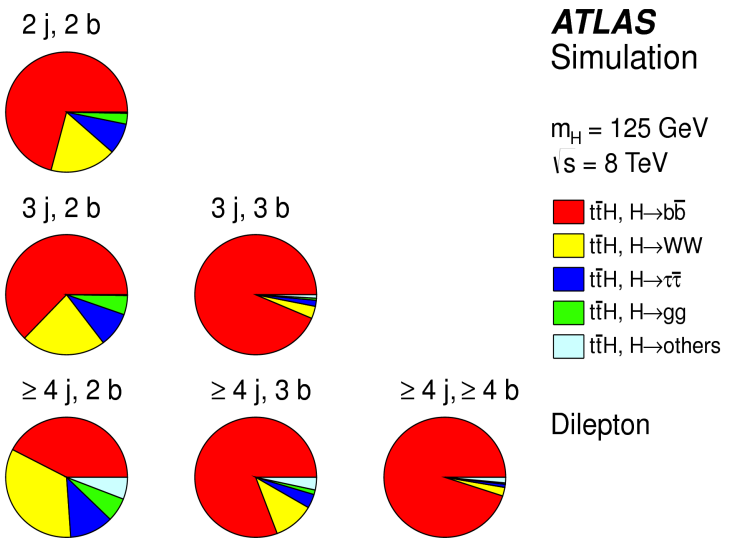


Figure 5.12: Fractional signal contributions to the total post-selection yields from events with different Higgs boson decay modes for the various analysis regions in the dilepton channel [58].

1340 events are rejected between the second and the last rows in Table 5.5. These are true $t\bar{t}H(H\rightarrow b\bar{b})$ dilepton events that have one or more poorly truth matched ($\Delta R_{(b-jet,b-parton)} > 0.4$) Higgs boson b-jets. A small fraction of these 1340 failing events are found in the high-$|\eta|$ tail of the red distribution in Figure 5.13. These events fail because one or more of the true Higgs boson b-partons is out of acceptance (large $|\eta|$). An example of such a case is if the true b-parton has gone down the beam pipe and is out of the η acceptance range of the detector but a b-jet has been reconstructed from elsewhere in the event allowing the event to pass all prior selection. When the truth matching is performed, this can cause the a large ΔR between the true parton and the reconstructed jet.

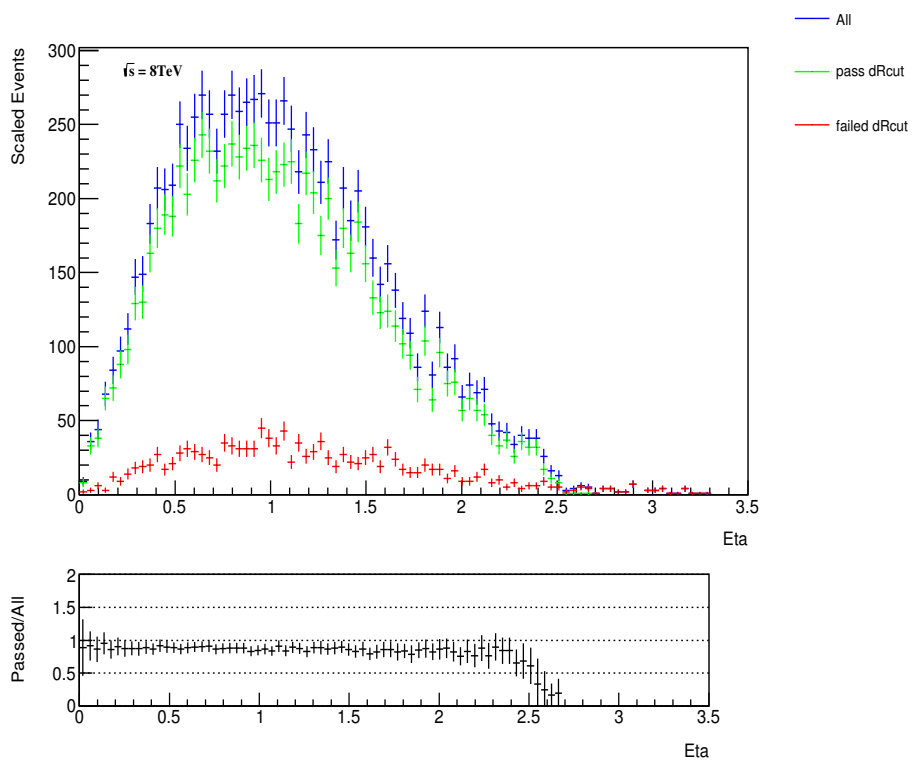


Figure 5.13: Largest $|\eta|$ of the two Higgs boson b-partons in simulated $t\bar{t}H$ dilepton sample in the ($\geq 4j \geq 4b$) region. The three distributions shown are for all true $t\bar{t}H(H\rightarrow b\bar{b})$ events passing full event selection (All), all true $t\bar{t}H(H\rightarrow b\bar{b})$ events that passed the full event selection and the truth ΔR selection (pass dRcut) and all true $t\bar{t}H(H\rightarrow b\bar{b})$ events that passed the full event selection but failed the truth ΔR selection (failed dRcut).

The reason for the remaining truth matching failures in Table 5.5 (the lower-$|\eta|$ region of the red distribution) is more subtle. The b-partons being used for the truth matching are the instances of the b-partons in the MC truth record before they have radiated gluons. However, often a b-parton will emit a gluon with a high transverse momentum and/or many softer gluons that will substantially deviate the path of the b-parton. This means that the b-parton before radiating the gluon and the reconstructed b-jet can have very different directions which can cause an event to fail the final ΔR cut.

Furthermore, the loss of energy via gluons inevitably results in a lower peak in the p_T spectrum of such b-jets. This is exemplified in Figure 5.14 where the p_T of the final instance of the b-parton in the MC truth record is plotted for events that have passed (green) and failed (red) the truth-matching. The truth-matching procedure for these events is still performed using the first instance

of the b-parton in the truth record. The average lower p_{T} of the b-partons that failed the ΔR cut is consistent with those partons having radiated gluons.

At the time of this study, information about the gluons themselves was not available in the truth record of the simulated samples available. In retrospect, a better truth matching efficiency could have been obtained by using the final instance of the b-parton in the truth record. However, the number of truth-matched events is only a bench-mark which is used to show how many events it is possible to correctly pair. Therefore, the method presented was deemed sufficient.

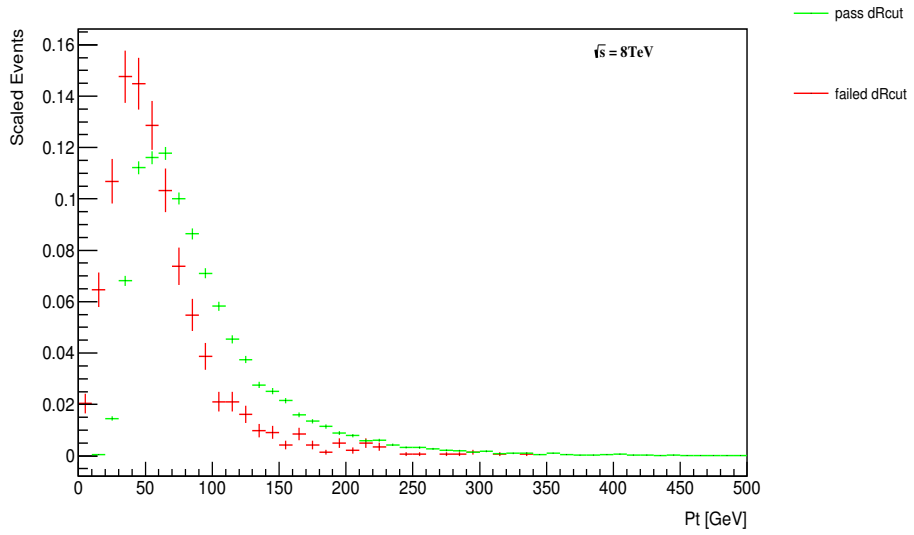


Figure 5.14: Transverse momentum of the final instance of the Higgs boson b-partons from true $\mathrm{t\bar{t}H(H\rightarrow b\bar{b})}$ events in the truth record that pass the full dilepton selection. Events that passed and failed the ΔR truth matching requirement using the first instance of the b-parton in the truth record are shown. The distributions are normalised to unity to emphasise the difference between their modal values.

An attempt to recover some of the events that failed the truth matching ΔR requirement was made by looking at the second-best permutation when the best permutation failed, instead of rejecting the event outright. The aim was to investigate whether the second-best permutation could sometimes provide two ΔR values which on average are larger and have a larger sum, but are both ≤ 0.4. In Figure 5.15 one can see the ΔR between the true Higgs boson b-parton and the b-jet it has

been matched with, for the best and second best permutations in events that failed the ΔR truth cut. The lowest two bins for the second best permutation (shown in green) show that around 100 events can be recovered by using the second-best permutation, however this is out of a total of 1340 that failed the truth matching and 9821 events that reach this cut.

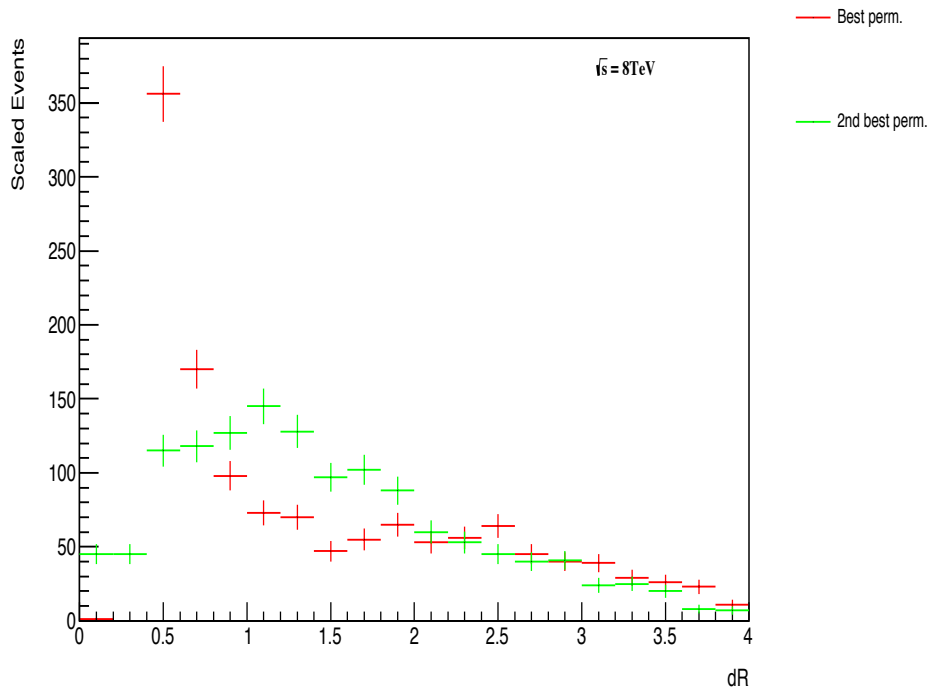


Figure 5.15: Distribution of the maximum ΔR between a Higgs boson b-parton and the associated jet for the best permutation and second-best permutation in events which failed the truth ΔR requirement.

Furthermore, when the event passes the truth requirements as a result of using the second best permutation, the large ΔR is often simply moved to the top b-jet. This can be explained by the topology in Figure 5.16, where there are two b-partons, one originating from the Higgs boson and the other from the top, are both a close ΔR match to the same b-jet (shown in green). The b-jet indicated in red, has been reconstructed at a large ΔR from the b-partons as a result of hard gluon emission. Table 5.6 shows a randomly selected collection of event outputs illustrating such behaviour. It can be seen that in such events, the 'problematic' pairing still exists, but due to the ΔR cut only being applied to the Higgs boson b-jets, the event passes the cut. It was therefore decided that such events should still be rejected.

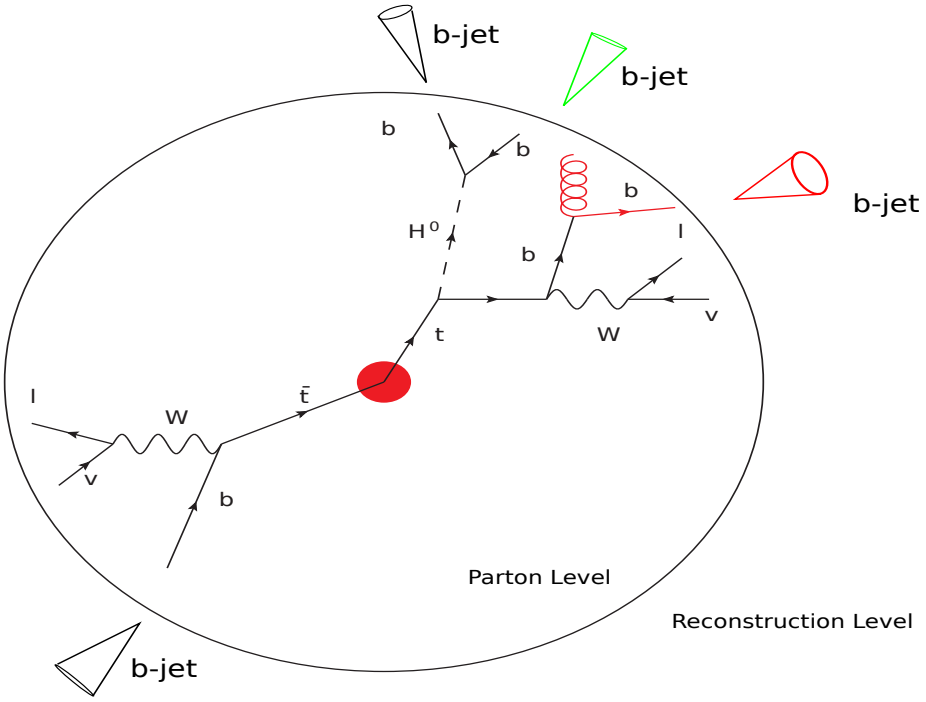


Figure 5.16: Schematic of a possible event topology that would fail the truth ΔR cut. All b-partons drawn with black lines are the b-partons that would be used in the truth matching procedure. The red b-jet is what would be reconstructed from the red b-parton after a high p_{T} gluon has been radiated. In this case, two b-partons (both shown in black, one from the Higgs boson decay and one from the top quark decay) would both have their closest matching b-jet to the green b-jet.

$\Delta R_{max}^{parton,\,jet}$	$\Delta R^{Higgsb1,\,jet}$	$\Delta R^{Higgsb2,\,jet}$	$\Delta R^{topb1,\,jet}$	$\Delta R^{topb2,\,jet}$	$\Delta R_{max}^{Higgsb,\,jet}$	$\sum \Delta R$
Best perm.	0.684	0.326	0.0300	0.031	0.684	1.072
2^{nd} best perm.	0.376	0.326	0.030	0.660	0.376	1.394
Best perm.	2.329	0.094	0.019	0.353	2.329	2.796
2^{nd} best perm.	0.12	0.094	2.441	0.353	0.120	3.010
Best perm.	0.040	2.412	0.081	0.039	2.412	2.574
2^{nd} best perm.	0.040	0.268	2.641	0.039	0.268	2.989
Best perm.	0.068	1.63	0.024	0.141	1.63	1.865
2^{nd} best perm.	1.468	3.144	0.024	0.141	3.144	4.776

Table 5.6: ΔR between the matched b-jet and b-parton for the best and second best permutations, in a few randomly selected events where the first permutation fails the $\max(\Delta R^{Higgsb1,jet}, \Delta R^{Higgsb2,jets}) < 0.4$ cut but the second permutation passes.

5.11.2 Higgs Boson b-jet Pairing Methods

In this section, different methods to determine which b-tagged jets in a $t\bar{t}H$ event originated from the Higgs decay are investigated. The methods presented here rely on simple and well-modelled information on the topology and kinematics of the events, such as the spatial separation between jets in the event (and, in the default pairing method, between jets and leptons) and jet-jet invariant masses. The methods are applicable to events in the region with four or more b-tagged jets. In this region, the jets from the signal and the remaining irreducible background are all expected to be b-jets. Therefore, the pairing algorithms only consider b-tagged jets, and any non-b-tagged jets are ignored, for pairing purposes. The five methods investigated are referred to as: the "dR method", the "dM method", the "dR dM method", "dRIso dR method" and the "dRIso dM method".

The dR method is the method that was used in the Run I analysis. It assumes that the decay products of the top are typically close together in $r - \phi$ and that the final state objects (t, $\bar{t}$, H) of a typical signal event are well separated. The two b-jets closest to the two leptons (as measured in terms of ΔR) are assigned to the b-partons originating from the top quarks and then the two remaining highest p_{T} b-jets are assigned to partons originating from the Higgs boson. This is shown schematically in Figure 5.17.

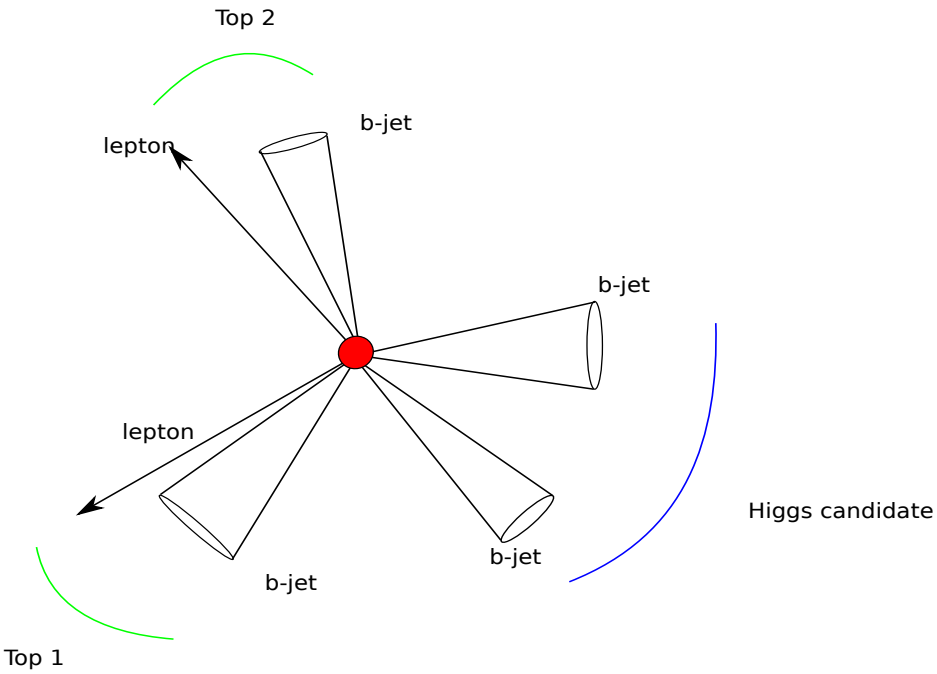


Figure 5.17: Schematic of the dR method. The b-jets paired with the leptons are assumed to have originated from the top-quarks and the remaining highest p_{T} b-jets are assumed to have originated from the Higgs boson.

The dM method is motivated by the expectation that a b-jet pair from the decay of the Higgs boson should have a dijet invariant mass that is typically closer to 125 GeV than any other pair of b-jets in the event. It was therefore hoped that finding the pair of b-jets with the closest dijet mass to 125 GeV would, more often than not, yield the b-jet pair associated with the Higgs boson. However, one possible pitfall of this approach is that one is then biased towards signal/background distributions with b-jet pairs with this dijet mass.

The dR dM method utilises the dR and dM methods sequentially. The top b-jets are assigned using the dR method, before using the dM method on the remaining b-tagged jets to assign the Higgs boson b-jets. For events with 4 b-jets this method is the same as using the dR method (as there would only be two b-tagged jets left for the dM method to iterate on), but it might be expected to produce an improvement in the small fraction of events with 5 b-tagged jets (see Figure 5.18).

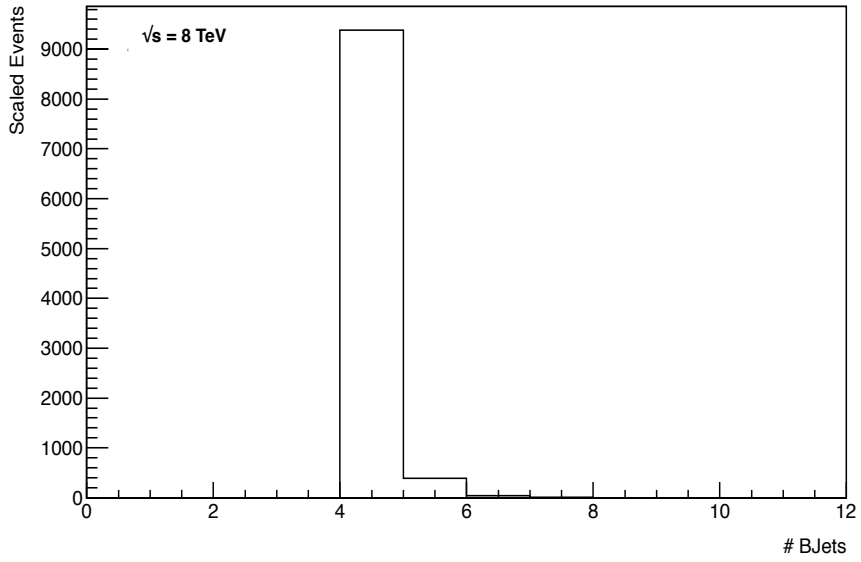


Figure 5.18: Distribution of the number of b-jets for events passing the full selection (before any truth matching cuts are implemented) in the (≥ 4j, ≥ 4b) analysis region.

The dRIso dR method is predicated on the hypothesis that the most isolated b-jet in a signal event is unlikely to have originated from the Higgs decay (seeing as the two b-jets from the Higgs decay are likely to be close to one another in $r-\phi$) - see Figure 5.17. The extreme case of this is for events with this "broomstick" topology, whereby one very well-isolated b-jet recoils against three b-jets in the opposite hemisphere of the event. In order to quantify each individual jet's isolation, the following variable, ΔR_i, is calculated for each of the four b-jets (i=1,2,3,4):

$$\Delta R_i = \sum_{j=1;j\neq i}^{4} \Delta R(i,j), \tag{5.6}$$

where the sum is over j, the index representing the other b-jets in the event (see Figure 5.19).

The b-jet which has the largest value of ΔR_i is assigned as one of the top quark b-jets and therefore removed from the list of possible Higgs boson b-jet candidates. The choice of possible Higgs boson b-jets is therefore reduced to three, upon which the dR method is then used to find the remaining top quark b-jet. This method is therefore specifically aimed at reducing the combinatorial background in events with this topology described above, by removing the possibility of the b-jet originating from the top quark on the opposite side of the event from the Higgs boson, being

assigned to the Higgs boson. For the dRIso dR method, if there are 5 b-jets in the event, out of the remaining b-tagged jets the two b-tagged jets with the highest p_{T} are assigned as the Higgs boson b-jets.

The dRIso dM method is a sequential implementation of using the ΔR_i separation in the way described above, before using the dM method to assign the Higgs boson b-jets instead of the dR method and p_{T}-ordering.

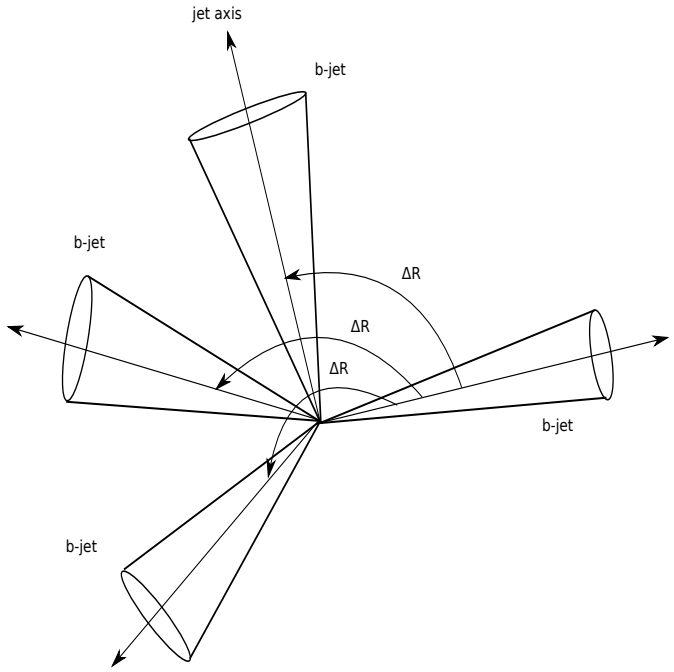


Figure 5.19: Schematic of the typical event topology targeted by the dRIso methods (see text).

5.11.3 Evaluation of the Performance of Pairing Methods

The evaluation of the performance of the pairing methods is based on the efficiency of each method. Efficiencies are calculated using Equation 5.7 where $N_{correct}$ is the number of events in which the two Higgs boson b-jets were correctly assigned to the true Higgs boson b-partons and $N_{matchable}$ is the number of events which passed the truth-matching requirements:

$$\varepsilon_{correct\ pairing} = \frac{N_{correct}}{N_{matchable}}. \tag{5.7}$$

Table 5.7 shows a summary of the performance of each of the methods.

The dR method does not perform very well, with a particularly low efficiency of just 31%.

This method makes the assumption that the lepton and b-jet originating from a top-quark are very close in $r-\phi$ and therefore a lepton can be matched to a b-jet by finding one with which it has the smallest ΔR. The blue data points in Figure 5.20 shows the angular distribution between the lepton and b-parton from the same top (the correct lepton b-parton pair). This distribution peaks at $\Delta R \approx 1$. The other distributions (incorrect lepton-b-parton pairs) show peaks at $\Delta R \approx 3$, however there are a significant number of data points from all three distributions with values less than 1, hence, choosing the smallest ΔR pair has significant contamination from the incorrect lepton-b pairings. The drop in the number of events at 0.4 in Figure 5.20, is due to the $\Delta R < 0.4$ cut in the selection.

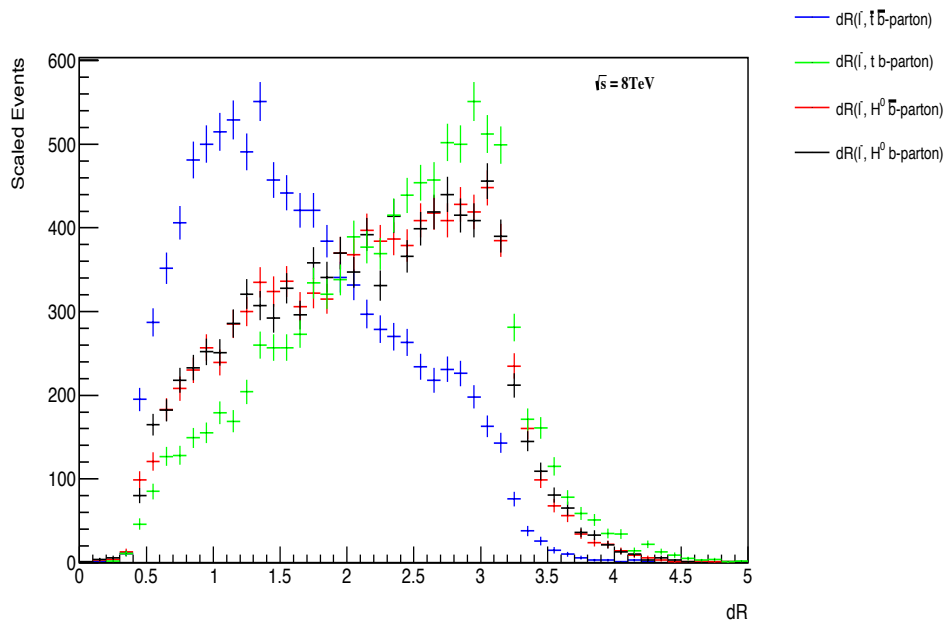


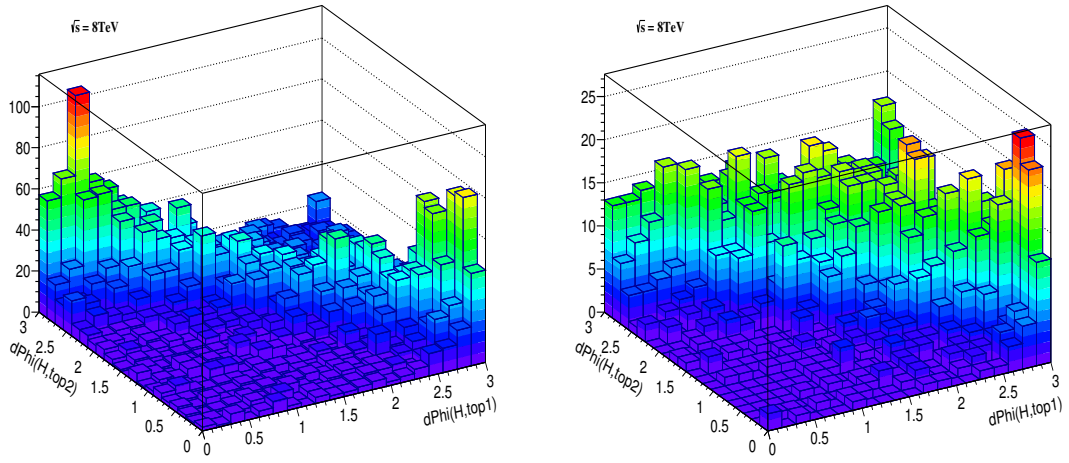
Figure 5.20: Parton level distributions of the ΔR between the lepton and each of the 4 b-partons associated with the top-quarks and the Higgs boson in simulated samples. The blue crosses refer to the correct pairing, as the ΔR is calculated using the negatively charged lepton in the event which has originated from the anti-top quark.

Particular event topologies can make the task of finding the correct pairing very difficult. The schematic shown in Figure 5.16 is of a $t\bar{t}H(H\rightarrow b\bar{b})$ topology that is particularly problematic for this technique. In such events, 3 b-jets (2 from the Higgs boson and one from the nearby top quark) are very close together, meaning the best permutation could be found by matching the wrong jets to the leptons especially if the ΔR between the lepton and the jet from which it originated is significantly

large.

This is emphasised by the distributions in Figure 5.21. These plots show the $\Delta\phi$ angle between the true Higgs boson and true top quarks in an event. Figure 5.21 (a) shows all events in which the incorrect b-jets were assigned to the Higgs boson. One can see that the two peaks of the plot are where one top has very small $\Delta\phi$(H,top) with the Higgs boson and the other has a very large $\Delta\phi$(H,top) (e.g. $\Delta\phi$(H,top1) = 0 , $\Delta\phi$(H,top2) = 3 and $\Delta\phi$(H,top1) = 3 , $\Delta\phi$(H,top2) = 0). This suggests the dR method frequently assigns the wrong b-jets to the Higgs boson, in events with similar topologies to that shown in Figure 5.19. Furthermore, one can see that events with this topology represent a significant fraction of the total number of events.

Figure 5.21 (b) shows the same plot but for events in which the dR method has assigned the correct b-jets to the Higgs boson. The level green band in the distribution shows that the dR method is equally good at assigning the correct b-jets to the Higgs boson for all topologies.



(a) (b)

Figure 5.21: Distributions of the $\Delta\phi$ between the truth Higgs boson and the leading (top1) and subleading (top2) truth top quarks for events in which the reconstructed Higgs boson b-jets were a) incorrectly and b) correctly paired by the dR method.

The dRIso dR method mildly improves upon the dR method by reducing the combinatorics of the problem. However, the poor performance is once again caused by the dR method's inability to correctly pair b-jets in the event topology shown in Figure 5.16.

The dM method's performance is significantly better than that of the dR and dRIso dR methods. However this method still only obtains the correct pairing $\approx 47\%$ of the time. The broad

reconstructed H→ $b\bar{b}$ dijet mass distribution as shown in Figure 5.22 means the true reconstructed Higgs boson mass can be up to 75 GeV from the measured Higgs boson mass. The dRIso dM method mildly improves the efficiency of the dM method by reducing the number of possible combinations of b-jets, but the improvement is $\approx 1\%$.

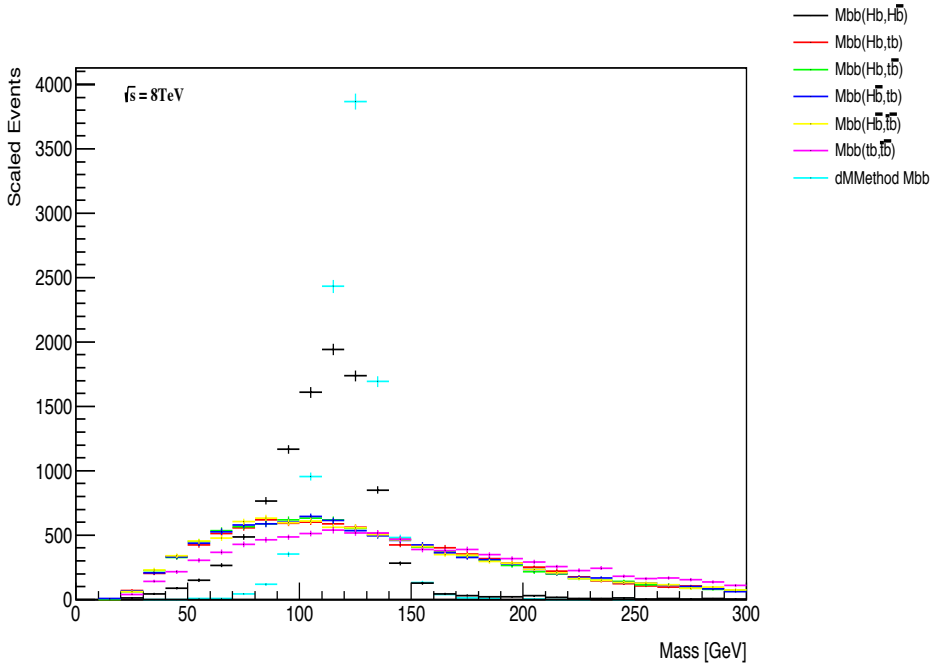


Figure 5.22: Invariant mass distributions using the 6 permutations of the b-jets in true $t\bar{t}H(H\rightarrow b\bar{b})$ events and the invariant mass of the Higgs candidate as constructed using the dM pairing method. The b-jets in the permuted distributions are labelled according to the true origin of the b partons matched to the b-jets: Hb = b-jet matched to a b quark originating from the Higgs boson, $H\bar{b}$ = b-jet matched to an anti-b quark originating from the Higgs boson, tb = b-jet matched to a b quark originating from the top quark, $t\bar{b}$ = b-jet matched to an anti b-quark originating from the anti-top quark.

The dR dM method is a sequential implementation of the dR and dM method. A small improvement is made on the efficiency of the dR method but it is again sub-percent seeing as the dM part of the method would only be used for the small fraction of events with 5 b-jets.

5.12 Boosted Decision Trees

In addition to the NN, input variable arrays are also used to train a boosted decision tree (BDT) for each pairing method, to test if an alternative classifier could improve upon the signal/background separation provided by the NN. This is done by making a like-for-like comparison between the BDT and NN trained using the same pairing method. BDT's were chosen as it is a competitive

Pairing Method Efficiencies		
Pairing Method	# Events Correctly Matched	Efficiency (%)
dRIso dM method	4115	48.52
dM method	3999	47.15
dRIso dR method	2706	31.91
dR dM method	2645	31.19
dR method	2625	30.95

Table 5.7: Table of the efficiencies of each of the pairing methods. The total number of true $t\bar{t}H(H\rightarrow b\bar{b})$ events with well-matched Higgs boson b-jets is 8481.

classifier [80] which is relatively transparent in its structure and quick to train. In Section 5.12.1, the basic premise behind the boosted decision tree is outlined. The TMVA package [81], which contains a BDT algorithm, is then used as an analysis tool for the rest of the study.

5.12.1 Decision Trees

A decision tree is a multivariate classifier that uses a series of cuts which can be optimised to distinguish between signal and background like events. A schematic of an example decision tree is shown in Figure 5.23. Events are passed to the primary node which splits events into background-like or signal-like events according to the splitting criterion on that node. The events are then passed to one of the two child nodes depending on the whether they satisfied the criterion or not. The fact that the event is not simply classified as background or signal based on one cut means that a multivariate analysis can recover signal sensitivity. The process is then repeated, with the sample once again being split according to the criterion on the two child nodes. This process is iterated until a minimum number of events, a maximum/minimum signal purity is reached or the tree depth is reached. Following this, the last nodes are labelled signal or background leaves according to the dominant classification type for events on that leaf.

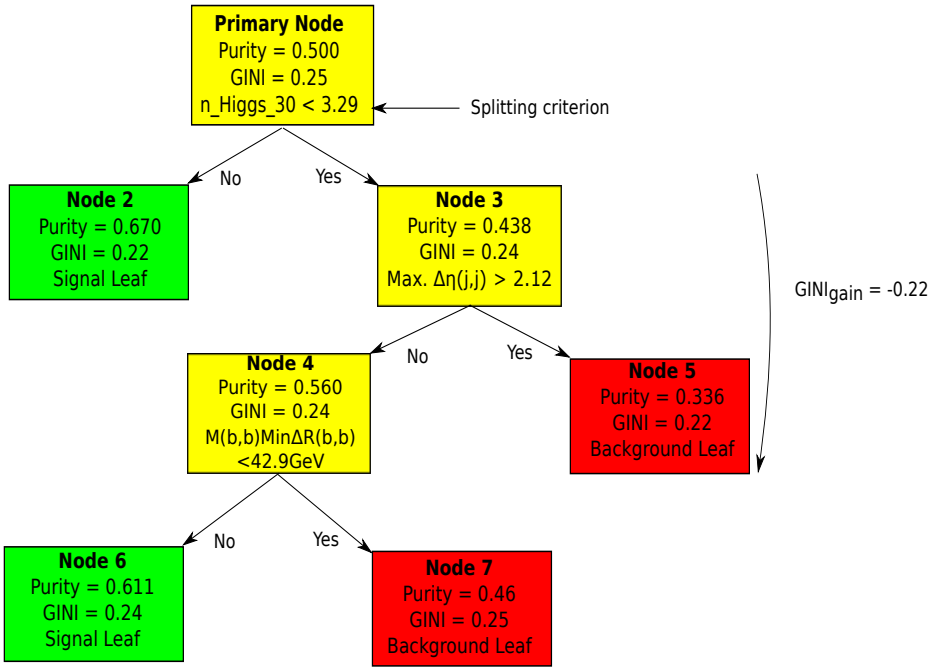


Figure 5.23: Schematic of an example decision tree. The primary node splits all events in the sample according to the primary criterion. Background events are scaled so that they represent the same initial proportion of the total sample as the signal events. Internal nodes are coloured yellow, background leaves are red and signal leaves are green. An example GINI_{gain} is shown on the left-hand side of the schematic.

5.12.2 Training

The order of the discriminating variables and the splitting criteria on the nodes are defined during training. The aim of the training is to maximise the purity on signal nodes and the impurity on the background nodes. This is done using a set of training events that are independent of those which will be used to test the NN after. The purity is defined as is shown in Equation 5.8, where S is the number of signal events and B is the number of background events on a node.

$$p = \frac{S}{S+B} \tag{5.8}$$

The primary node splits the events into two subsets according to some cut criteria. The variable and corresponding cut value that provide the best signal/background separation are chosen. At each node the separation index (GINI index) is calculated as shown in equation . Large GINI indices (maximum = 0.5) mean the sample is well mixed, whereas small indices mean the sample is well separated. All separation criteria such as the GINI index have a maximum where the samples are fully mixed ($p = 0.5$).

$$GINI_{node} = p(1-p) \tag{5.9}$$

Following the primary node, the training procedure selects the variables and cut criterion that increases the gain in separation index between the parent node and the sum of the indices of the child nodes. The gain in separation is calculated as shown in Equation 5.10. A negative gain means the separation of the two children combined is larger than the parent.

$$GINI_{gain} \cong GINI_{parent} - GINI_{child\ \#1} - GINI_{child\ \#2} \tag{5.10}$$

This process is repeated for subsequent nodes until one of three criterion are met: the number of events on a given node reaches a minimum percentage of the total number of events in the training sample which is set at the default value of 5%, the maximum depth of the decision tree is reached (max. depth = 3 internal node levels) or the maximum/minimum signal purity is reached. Leaves are then classified as signal or background depending on which holds the majority i.e. signal if $p > 0.5$.

The evaluation as to whether the event was correctly classified is done in training by comparing the true nature of the event which is known in simulation and the decision trees prediction.

Decision trees are considered useful because they can be trained very quickly and are relatively straightforward to understand. On the other hand, sometimes during training the algorithm can choose a cut or cut order based on a statistical fluctuation instead of actual signal or background phenomena. This is known as over-training and can result in an artificially enhanced performance of the classifier on training data that when tested on an independent data set performs worse. A good test to see if the classifier is over-trained is to compare the classifiers response in training and testing.

5.12.3 Boosting

Boosting combines a group of weak learning algorithms to make a stronger learning algorithm. A boosted decision tree is a collection (forest) of decision trees that are used to classify events as either background or signal depending on the majority classification of the individual decision trees.

The BDT algorithm used in this analysis is taken from the TMVA package. The classifier

is trained using adaptive boosting (AdaBoost) as it is the favoured boosting algorithm provided by TMVA. Adaptive boosting means misclassified events are given a higher event weight in the training of the following tree in the forest. The misclassification rate (*err*) of an individual tree is evaluated as the ratio of the sum of the weights of the misclassified events with respect to the sum of the weights of all events as shown in Equation 5.11.

$$err = \frac{\sum_i w_i \cdot M}{\sum_i w_i} \begin{cases} M = 0 & \text{correct classification} \\ M = 1 & \text{misclassified} \end{cases} \tag{5.11}$$

The original event weights used during the training of the first tree are modified according to the misclassification rate as shown in Equation 5.12.

$$\alpha = \frac{1 - err}{err} \tag{5.12}$$

The value α is known as the common boost weight and is multiplied into the weights of all events that were misclassified in the previous tree before training the next tree. The entire event sample is renormalised to retain the total number of events. The reweighting makes the new training concentrate on avoiding a misclassification of the same events.

The result of an individual decision tree is $h(x)$ where x is a vector of input variables for a single event. Values of $h(x)$ are designed to give +1 for signal events and 0 for background events. The resulting event classification of a boosted classifier is an averaged decision of all the individual trees which is calculated as shown in Equation 5.13. This is then transformed using Equation 5.5 to return a value between -1 and +1 where positive values are signal-like and negative values are background-like.

$$y_{BDT}(x) = \frac{1}{N_{collection}} \sum_{i \in forest} ln(\alpha_i) \cdot h_i(x) \tag{5.13}$$

AdaBoost produces the best results when used on small decision trees with a depth of 2 or 3 internal node layers. Individually these have very little discrimination power, however they are less prone to over training than larger decision trees and when combined in a BDT they outperform them [81]. A total of 850 trees with a depth of 3 internal node layers are used which provides sufficient statistical stability.

The performance is often further enhanced by forcing a slow learning. The aim of the mul-

tivariate classifier is to find the global minima in the multidimensional parameter space made up from the input variables, where the classifiers prediction is close to the target value. The learning rate defines the size of the steps in the parameter space a multivariate classifier can make as it tries to make its way to a global minima. Smaller learning rates mean the classifier is more likely to find the global minima as it will investigate a more detailed map of the phase space. However it can also mean the classifier takes longer to train and can get stuck in local minima. The suggested rate of for a BDT using the TMVA software is 0.03 and this is what is used here [81].

5.13 Performance of Multivariate Algorithms

Some of the NN input variables (Table 5.4) will distribute differently depending on which pairing method is used. Changing the pairing method can therefore effect the performance of the NN. To investigate this effect, independent multivariate classifiers are trained on input files that are individually generated using the different pairing methods. The classifiers are trained using $t\bar{t}H$ events as signal and a combined sample of $t\bar{t}$, $t\bar{t}+V$, $Z+$jets and single top (Wt process) (excluding fake events) as background. Cross-training is implemented so that the entire sample can be used for training and testing. This makes the algorithm more resilient to statistical fluctuations, without increasing the chances of over-training. Cross-training entails training two classifiers for any single sample; one on events that have an odd event number and another on events that have an even event number. These classifiers are then tested individually on the events they were not trained on (e.g. classifier trained on odd event numbers is tested on even event numbers and vice versa).

The NNs use the same packages and hyper-parameters that were used when training the NN in the Run I analysis. A more detailed description is found in Section 5.10. The NN's performance is evaluated on its ability to separate background and signal. This is done by calculating the separation from the classifier response distributions as:

$$S = \frac{1}{2}\sum_{i}^{N\,bins}\frac{(N_i^{signal}-N_i^{bckg})^2}{(N_i^{signal}+N_i^{bckg})}. \tag{5.14}$$

The classifier response distributions obtained from the training and testing of the NNs are shown in Figures 5.24, 5.25, 5.26, 5.27 and 5.28. Each distribution shows the response obtained from a classifier trained using a different pairing method.

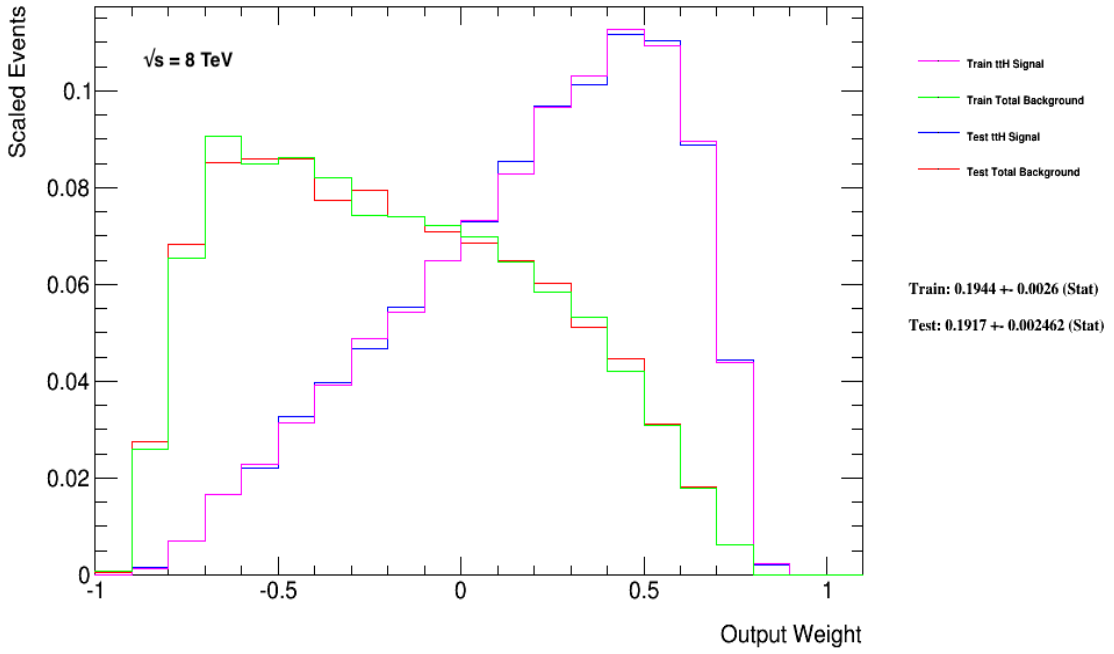


Figure 5.24: NN discriminant distributions for classifiers trained on simulated input ntuples of events with 10 variables constructed using the dR method.

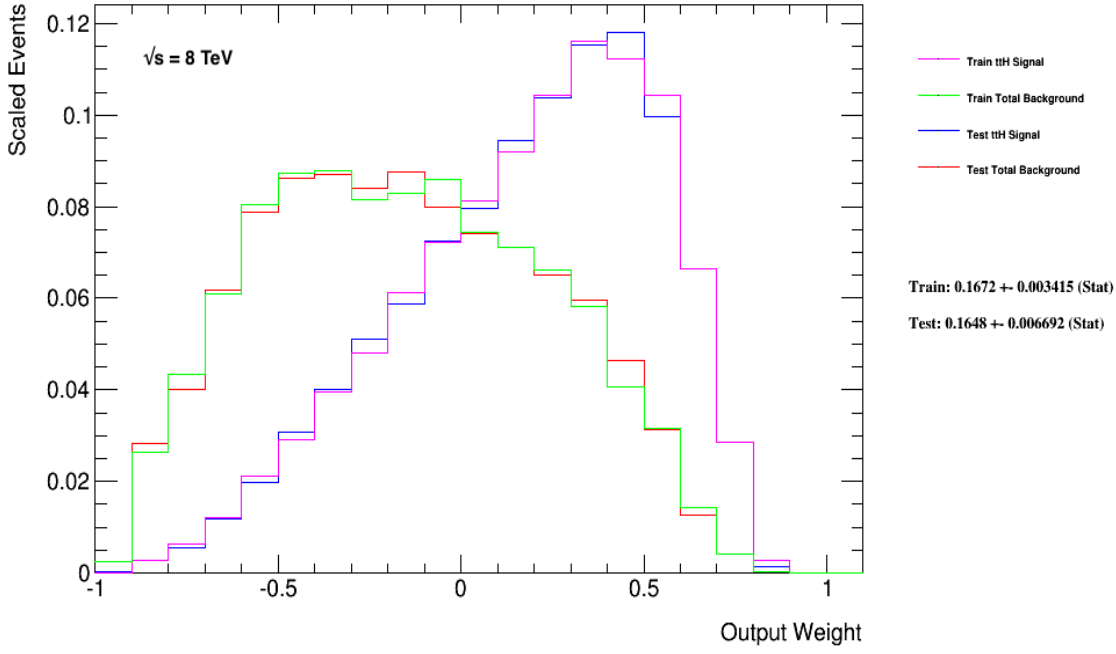


Figure 5.25: NN discriminant distributions for classifiers trained on simulated input ntuples of events with 10 variables constructed using the dM method.

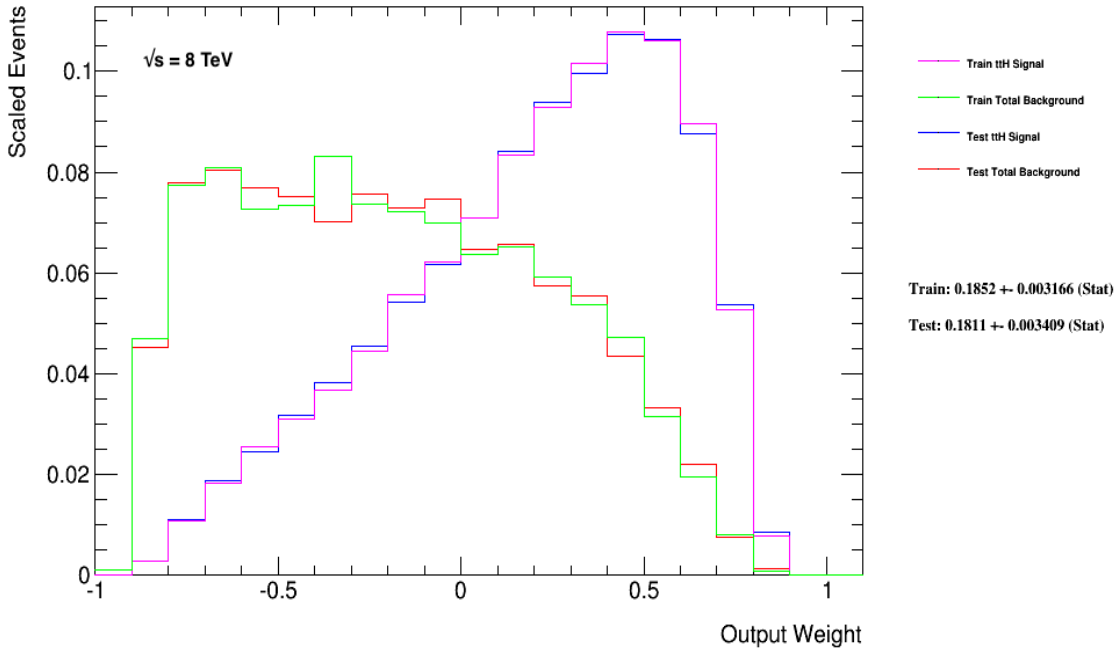


Figure 5.26: NN discriminant distributions for classifiers trained on simulated input ntuples of events with 10 variables constructed using the dRIso dR method.

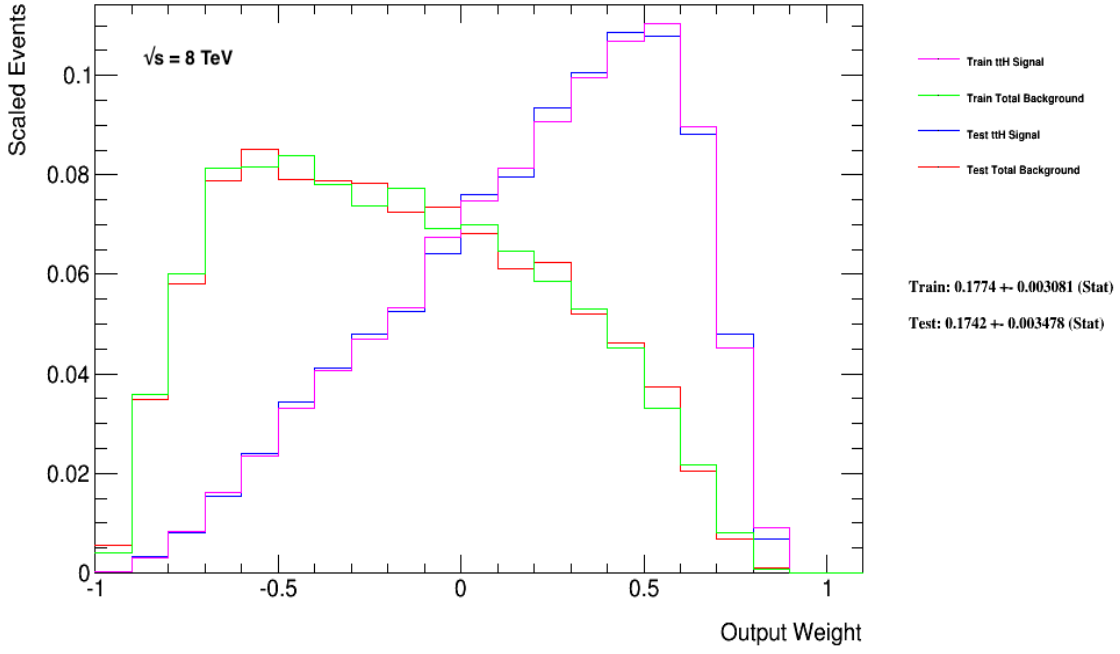


Figure 5.27: NN discriminant distributions for classifiers trained on simulated input ntuples of events with 10 variables constructed using the dRIso dM method.

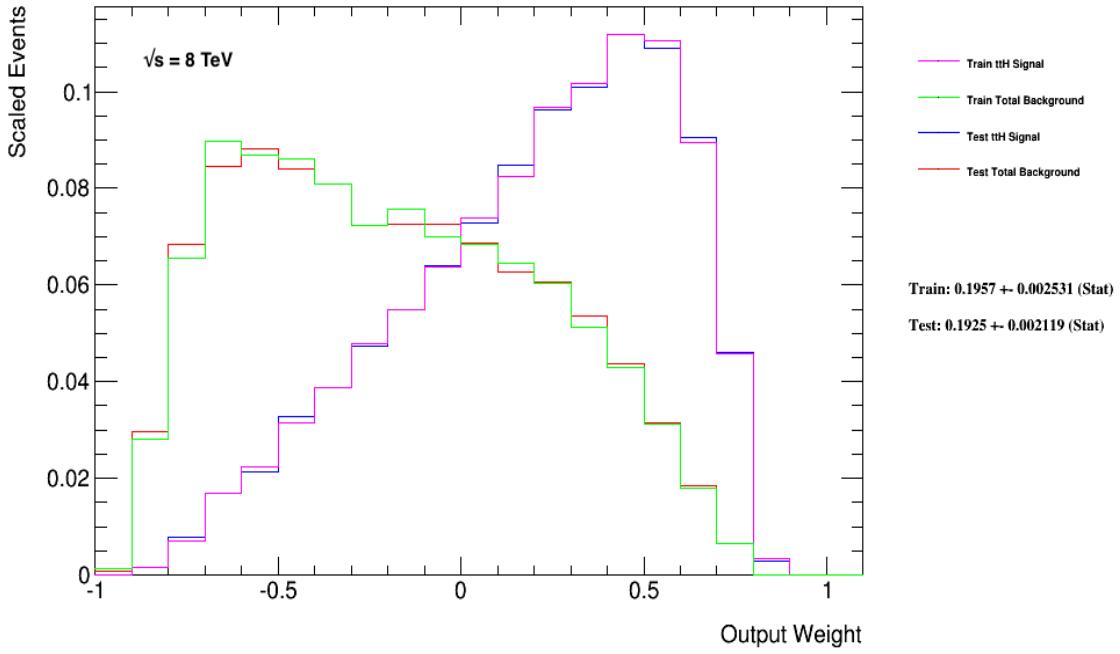


Figure 5.28: NN discriminant distributions for classifiers trained on simulated input ntuples of events with 10 variables constructed using the dR dM method.

BDTs were also trained using each of the pairing methods in order to allow like-for-like comparisons between NN and BDT classifiers for a single pairing method. The same 10 variables are used to train the BDT are also used to train the neural network. The training is performed on the same signal and background samples as the NN's and also implements cross-training. Distributions of the classifier responses obtained from training and testing the BDT on simulated samples of input variables is shown in Figure 5.29, 5.30, 5.31 , 5.32 and 5.33. Once again, each distribution shows the response obtained from a classifier trained using a different pairing method. Like the NN, the BDT's performance is evaluated on its ability to separate background and signal. The separation of the classifiers trained using each of the pairing methods is shown in Table 5.8.

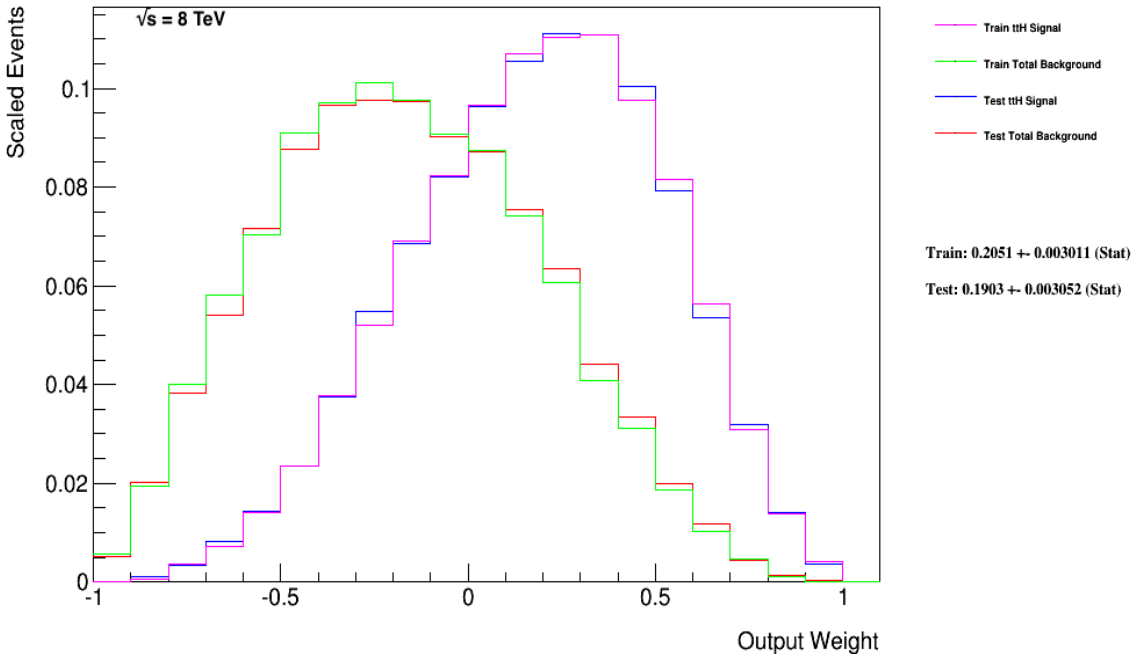


Figure 5.29: BDT discriminant distributions for classifiers trained on simulated input ntuples of events with 10 variables constructed using the dR method.

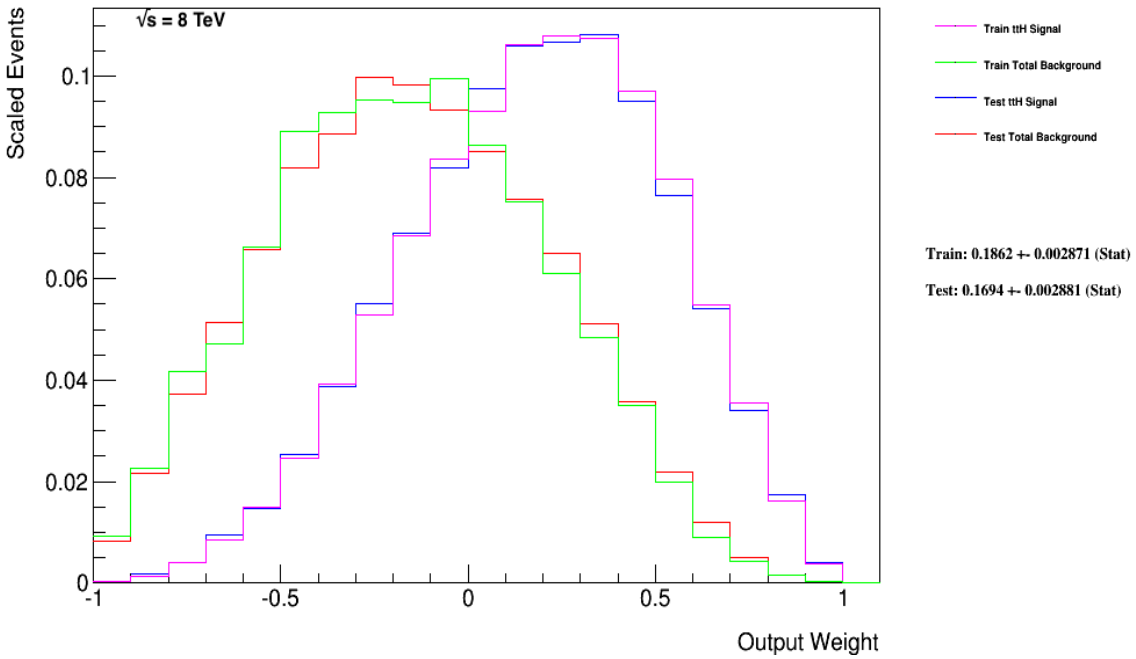


Figure 5.30: BDT discriminant distributions for classifiers trained on simulated input ntuples of events with 10 variables constructed using the dM method.

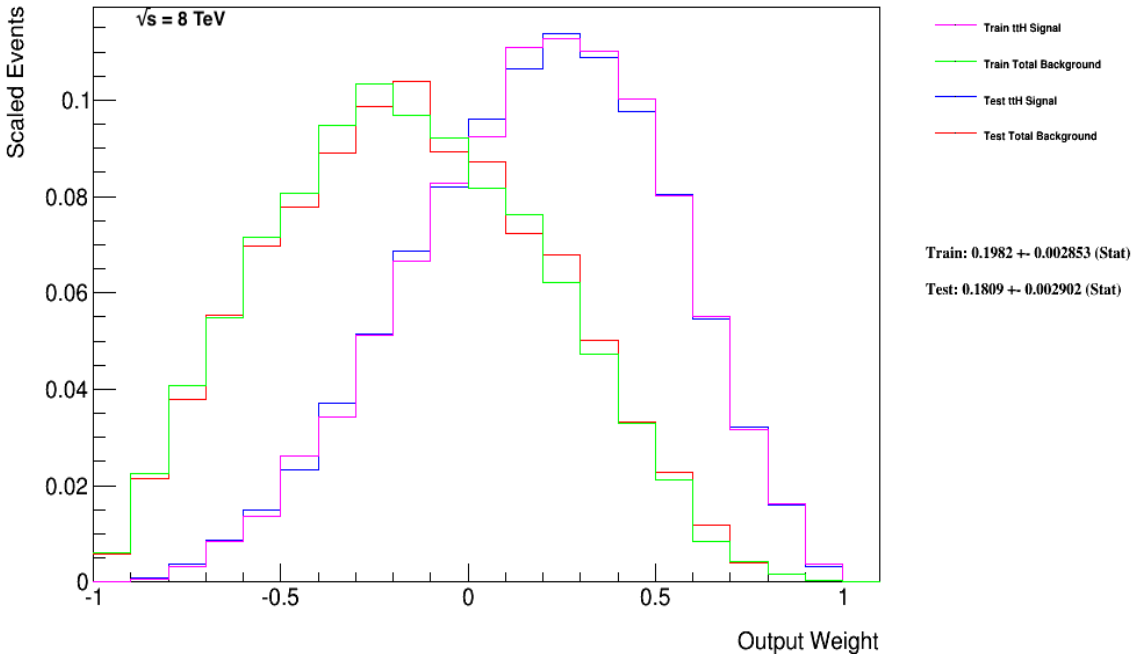


Figure 5.31: BDT discriminant distributions for classifiers trained on simulated input ntuples of events with 10 variables constructed using the dRIso dR method.

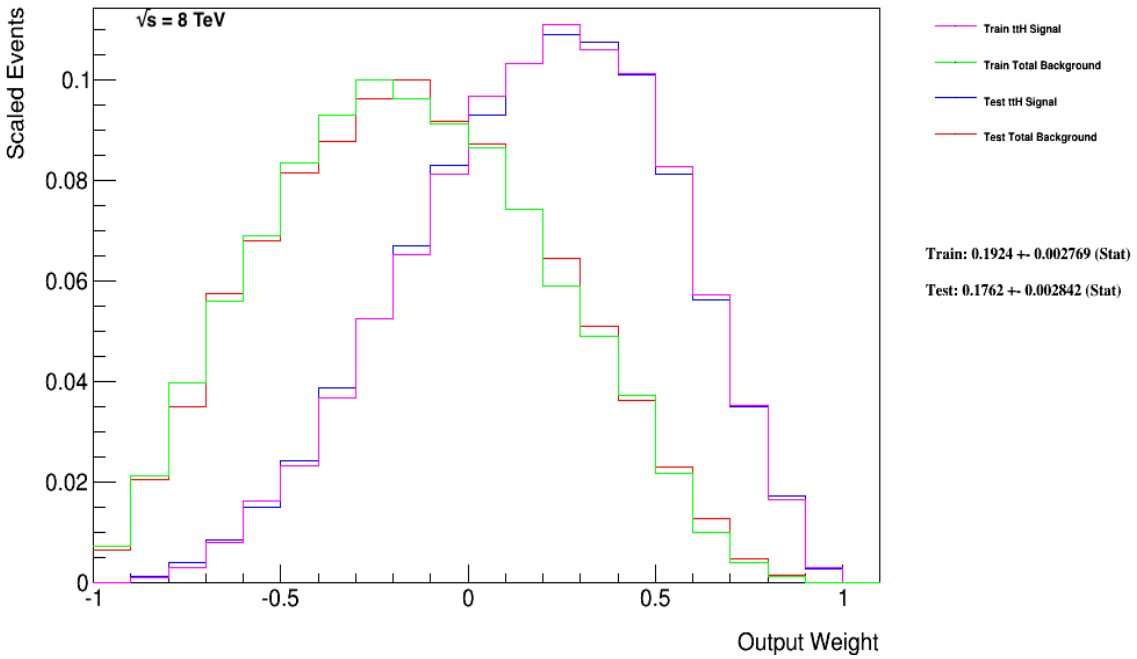


Figure 5.32: BDT discriminant distributions for classifiers trained on simulated input ntuples of events with 10 variables constructed using the dRIso dM method.

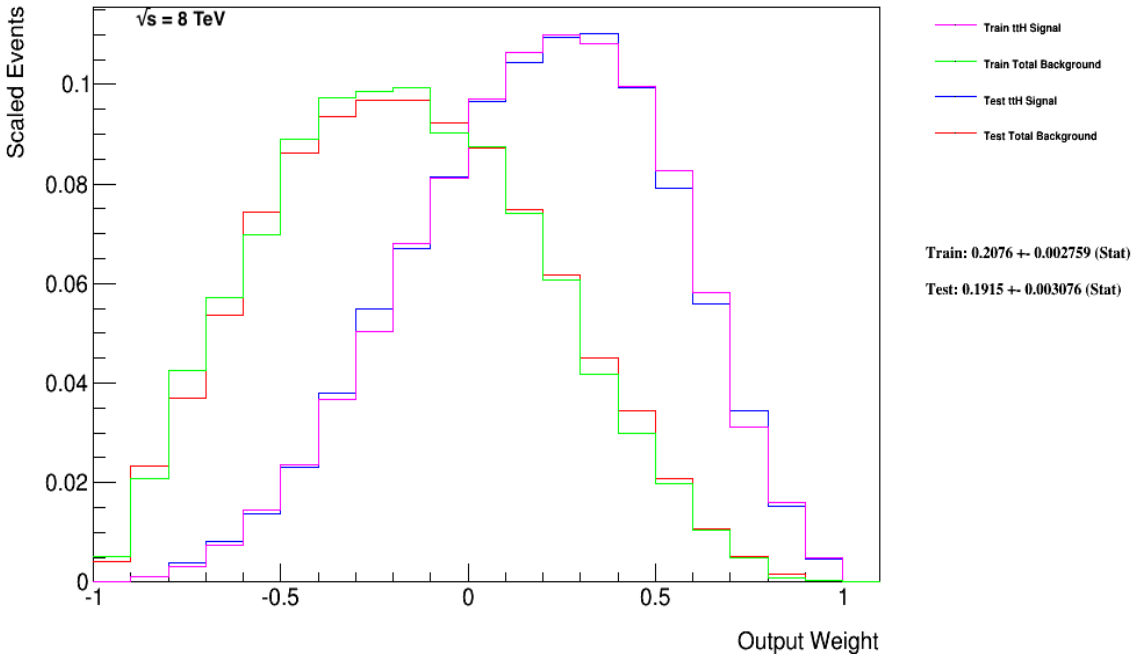


Figure 5.33: BDT discriminant distributions for classifiers trained on simulated input ntuples of events with 10 variables constructed using the dR dM method.

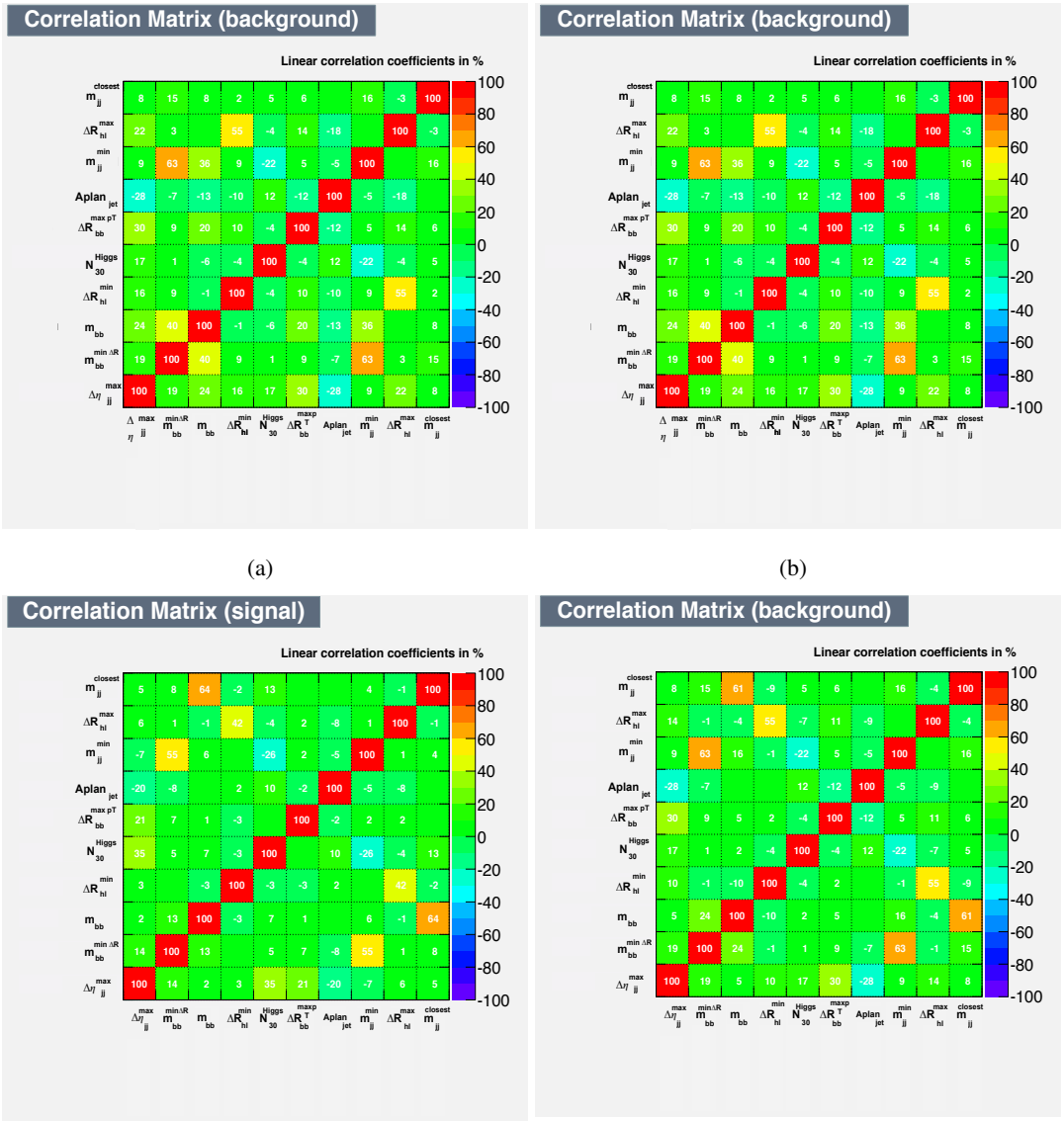
Classifier	**Pairing Method**	**Separation**
BDT	dRdM method	0.192 ± 0.003
	dR method	0.190 ± 0.003
	dRIso dR method	0.181 ± 0.003
	dRIso dM method	0.176 ± 0.003
	dM method	0.169 ± 0.003
NN	dRdM method	0.193 ± 0.002
	dR method	0.192 ± 0.002
	dRIso dR method	0.181 ± 0.003
	dRIso dM method	0.174 ± 0.003
	dM method	0.165 ± 0.007

Table 5.8: Separation of BDT and NN classifiers trained on ntuples of 10 input variables constructed using the different pairing methods, where the separation is defined in Equation 5.14. The statistical uncertainties are shown. For each classifier, pairing methods are ordered by their separation power. As the same event samples are used for all the separation values, any statistical fluctuations are very highly correlated between the quoted separation.

Overall, no significant improvement in the separation power of a classifier trained using the

dR pairing method is seen for either the NN or BDT when trained on 10 input variables. However, the wide range of results shows how much the pairing method can effect the performance of an individual classifier.

The dRIso dM method provides the best pairing efficiency but does not improve the separation of background from signal with respect to the Run I analysis method. In fact, any classifier trained using the dM method, performs poorly as shown in Table 5.8. This is explained by the correlation plots shown in Figure 5.34 which show the appearance of two highly correlated variables when using the dM method. The large correlation is between the variables 'mass of the dijet pair nearest to 125 GeV in the event' and 'Higgs boson mass reconstructed from the Higgs boson candidate b-jets in the event'. This correlation is a result of reconstructing the Higgs boson from the dijet pair with mass nearest 125 GeV (i.e. the dM Method).



(a) (b)

(c) (d)

Figure 5.34: Matrices showing the linear correlation coefficient between BDT input variables. The correlation between variables constructed using the dR method are calculated for (a) signal events and (b) background events. The correlations between variables constructed using the dM method are calculated for (c) signal events and (d) background events.

The highly correlated variables cause a redundancy of information and reduce the performance of the classifier. The variables being used were the 10 variables that provided the best discrimination as defined by NeuroBayes' internal ranking procedure. The ranking procedure in this case will typically remove highly correlated variables however, the correlation between variables can differ depending on how they are defined. In other words, different pairing methods can change

the correlation between variables. This set of variables, which was defined using the dR method, is therefore likely to change if the ranking procedure is rerun using variables made with another pairing method. A new set of variables could be chosen which would potentially increase the performance of the classifier by replacing correlated variables with a new discriminating variable.

When both classifiers are trained using 10 input variables and the same pairing method, the separation power of a BDT is the same as the NN taking into account statistical errors. The order in which the pairing methods appear in Table 5.8 for the BDT and NN is also the same. This is caused because the main driving force behind the ordering of the pairing methods is this correlation introduced by the dM method. The mass of the dijet pair nearest to 125 GeV in the event and the Higgs boson mass reconstructed from the selected Higgs boson candidate b-jets in the event, are more correlated if the dM method is the only method involved in the pairing procedure. Classifiers trained using this method therefore perform the worst. When the dM method is combined into the dRIso dM method, the dRIso part of the pairing procedure identifies one of the b-jets from the decay of the top quark which reduces the combinatorics of the problem. This subsequently reduces the impact of the dM method and improves the performance of the classifier. When combined into the dR dM method, the only events in which the dM method will actually have a decision to make is in events with more than 4 b-jets which make up less than $\approx 5\%$. This means the dM method is less of a detriment for the performance of the classifiers trained using the dR dM method.

A further study was made into the impact of including additional input variables. An additional 7 variables taken from the variables not used in the (4j,4b) region in Table 5.4 were added to the input variable list for both the BDT and the NN. New NNs and BDTs were trained for the two pairing methods which showed the best separation when trained on 10 input variables. The classifier distributions are shown in Figures 5.35 and 5.36, and 5.37 and 5.38. Comparing Table 5.9 with 5.8 shows that by adding these 7 variables, one can improve the performance of both classifiers. Larger improvements are seen in the separation power of the BDT than in the NN. A small improvement is seen between the separation power of the two best-performing BDT's when compared to the two best-performing NN's.

Classifier	Pairing Method	Separation
BDT	dRdM method	0.201 ± 0.003
	dR method	0.200 ± 0.003
NN	dRdM method	0.195 ± 0.003
	dR method	0.193 ± 0.002

Table 5.9: Separation of BDT and NN classifiers trained on 17 input variables constructed using the different pairing methods. The statistical uncertainties are shown. For each classifier, pairing methods are ordered by their separation power. As the same event samples are used for all the separation values, any statistical fluctuations are very highly correlated between the quoted separation.

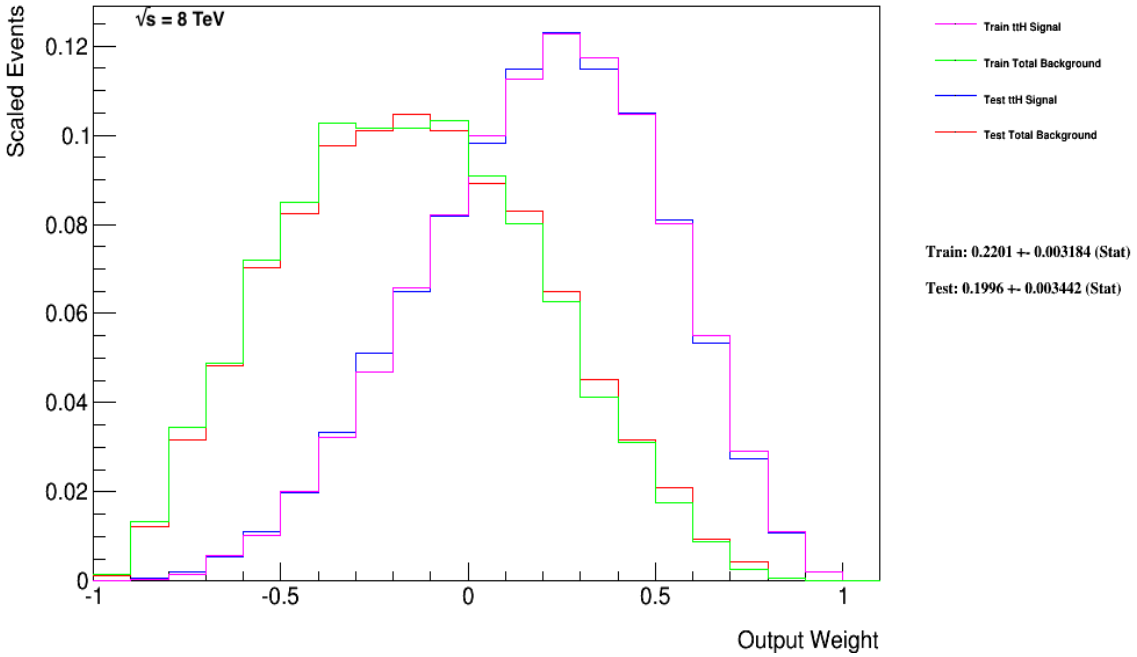


Figure 5.35: BDT discriminant distributions for classifiers trained on 17 variables constructed, using the dR method.

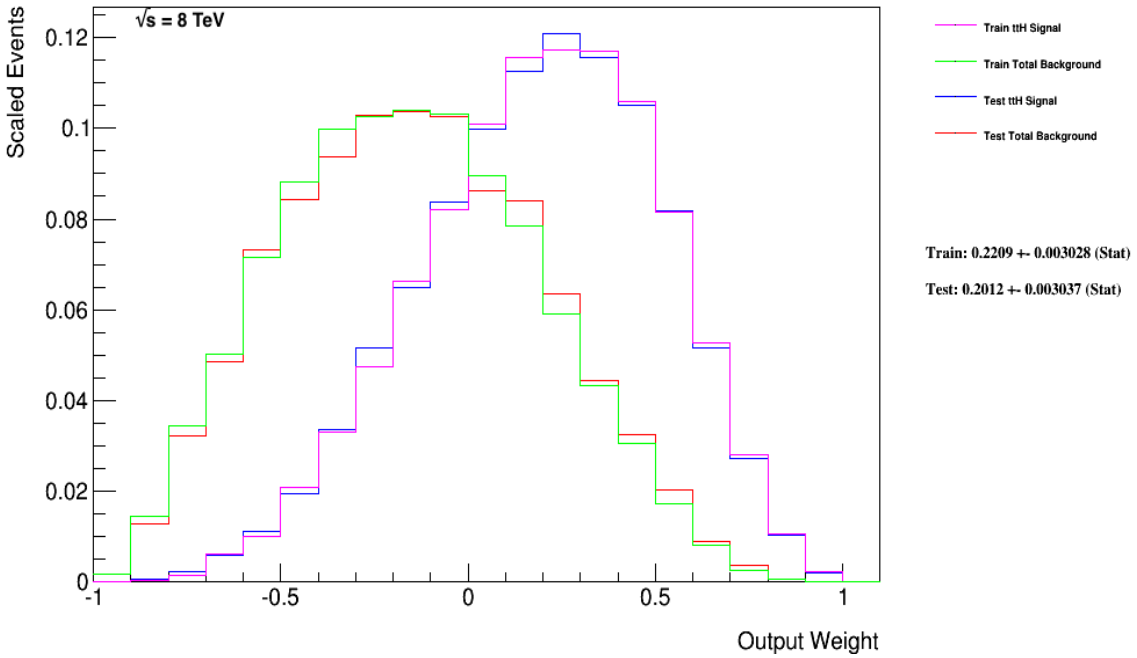


Figure 5.36: BDT discriminant distributions for classifiers trained on 17 variables constructed, using the dR dM method.

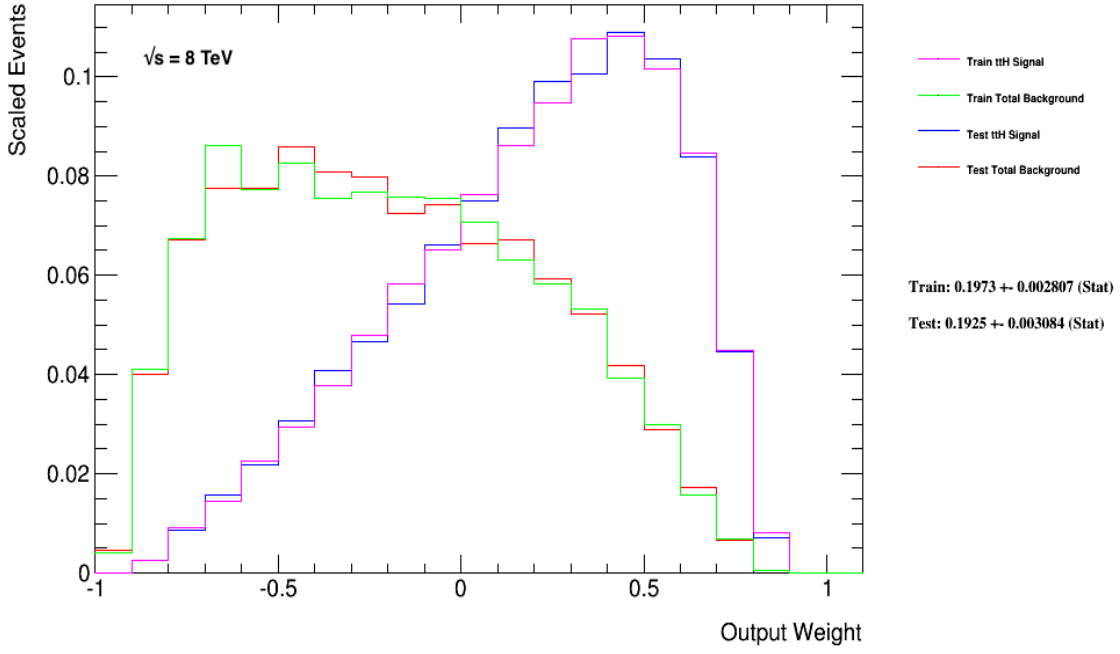


Figure 5.37: NN discriminant distributions for classifiers trained on simulated input ntuples of events with 17 variables constructed using the dR method.

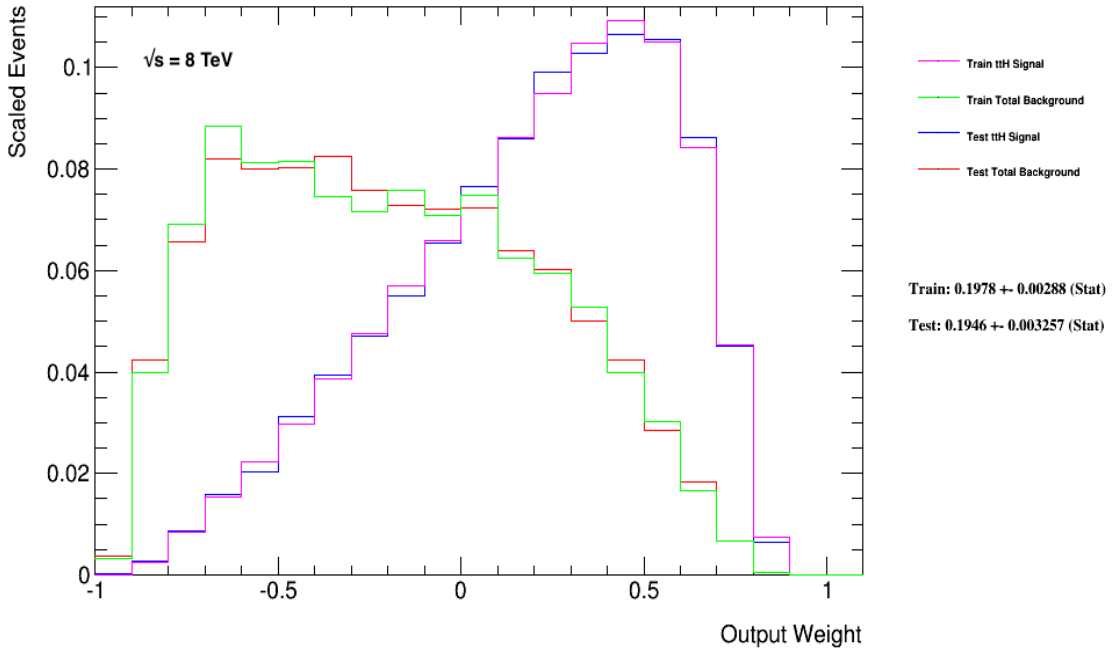
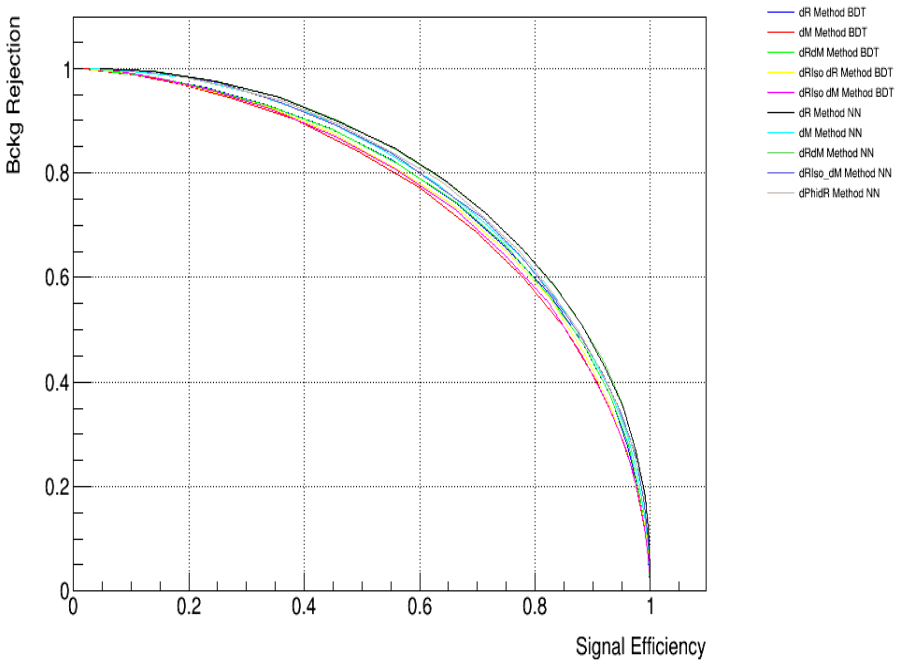


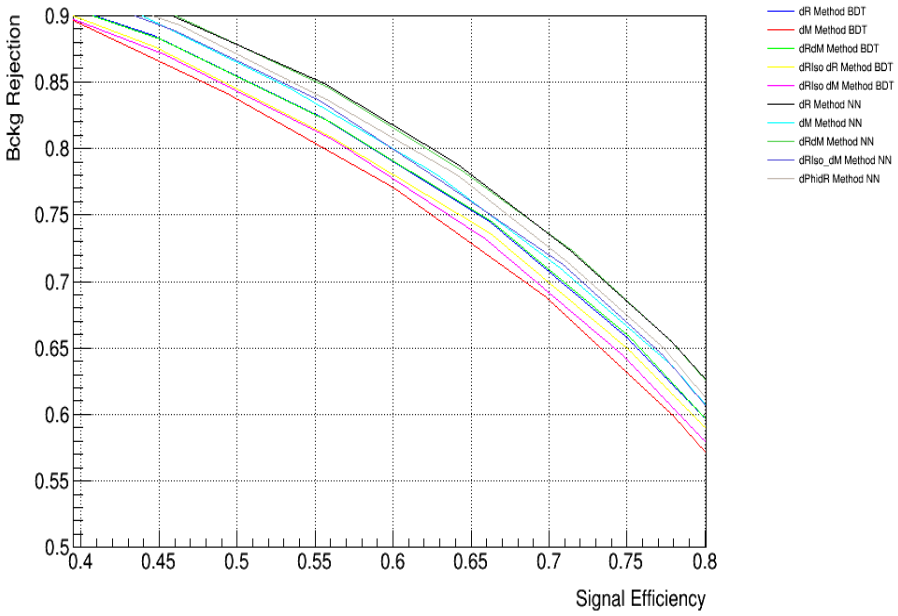
Figure 5.38: NN discriminant distributions for classifiers trained on simulated input ntuples of events with 17 variables constructed using the dR dM method.

The performance of the BDTs and NNs trained using each pairing method are shown in Figure 5.39. Each curve shows the receiver operating characteristics (ROC) for each method. The x-axis represents the signal efficiency (ε_{signal}). To calculate the signal efficiency, a cut is defined on the multivariate classifier output distribution which distinguishes between signal-like and background-like events. The cut is varied between the minimum and maximum of the range of the classifier output and at each point the signal efficiency is calculated as the fraction of signal events that pass the cut. The y-axis represents the background rejection. This is again calculated by varying the cut in the same range but for each cut the background efficiency (ε_{bckg}) is calculated as the fraction of background events passing the cut. The background rejection is then calculated as ($\text{rej.}_{bckg} = 1 - (\varepsilon_{bckg})$). The better the classifier's performance, the closer the apex of the curve is to the top right corner of the plot. These plots also allow one to check the effect of the statistical uncertainty on the performance of the classifier. Large statistical uncertainties would manifest themselves as statistical fluctuations in the ROC curve.

A comparison of the best performing 10 variables classifiers with the two best performing 17 variable classifiers can be made by comparing Tables 5.8 and 5.9.



(a)



(b)

Figure 5.39: ROC curves of the performance of BDT/NN classifiers trained using 10 input variables and showing all pairing methods. Figure (b) is simply a magnified region of Figure (a) to exemplify the differences between the curves.

5.14 Evaluation of the Choice of Signal Sample Parton Shower Systematic Uncertainty

The nominal $t\bar{t}H$ signal sample matrix element calculations are generated by POWHEL to NLO accuracy. The PYTHIA parton shower model then evolves the hard process generated by POWHEL into a multi-hadron final state. To evaluate the systematic uncertainty related with the choice of parton shower model for the signal sample, the full analysis is rerun with a signal sample that is generated using an alternative parton shower model (systematic sample). In the case of the $t\bar{t}H(H\rightarrow b\bar{b})$ dilepton analysis, the systematic signal sample used was generated using POWHEL to calculate the matrix element and HERWIG for the parton shower model. The most fundamental difference between the two parton shower models is that Pythia applies the string fragmentation model [82] and Herwig treats quark and gluon fragmentation according to the cluster fragmentation model [83],[84].

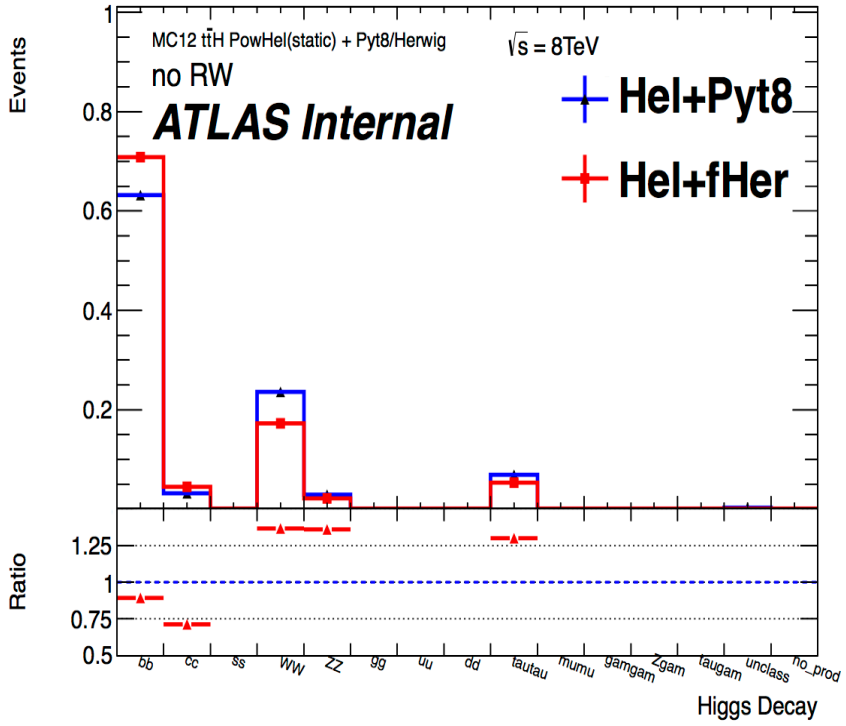
Higgs boson Decay channel	POWHEL+PYTHIA BR (%)	POWHEL+HERWIG BR (%)
$b\bar{b}$	57.62 ± 0.08	70.49 ± 0.08
$c\bar{c}$	2.92 ± 0.01	4.47 ± 0.02
$s\bar{s}$	0.024 ± 0.002	0.221 ± 0.005
WW	21.47 ± 0.05	17.19 ± 0.04
ZZ	2.65 ± 0.02	2.12 ± 0.01
gg	8.61 ± 0.03	0
$\tau\tau$	6.30 ± 0.03	5.24 ± 0.02
$\mu\mu$	0.021 ± 0.001	0.019 ± 0.001
$\gamma\gamma$	0.234 ± 0.005	0.189 ± 0.004

Table 5.10: Higgs boson branching ratios for the POWHEL+PYTHIA and POWHEL+HERWIG simulated samples.

The Higgs boson branching ratios are calculated by the Pythia and Herwig parton shower tools and are dependant on the mass of the Higgs. For the Pythia sample, the Higgs boson mass was set at 125 GeV, however in the Herwig sample this was mistakenly set to another value. While the Higgs boson decay branching ratios in the nominal samples are assumed to be accurate, the HERWIG sample was generated with significantly different branching ratios. The branching ratios for both samples are shown in Table 5.10. To use this sample for the evaluation of the choice of

hadronisation systematic uncertainty the HERWIG branching ratios are corrected to match those in PYTHIA. The original Higgs boson branching fractions (pre-selection) in the nominal and HERWIG samples are shown in the histogram on the left of Figure 5.40. In a given decay channel, each event in the HERWIG sample is multiplied by the ratio of the branching fraction in the PYTHIA sample with respect to the branching fraction in the HERWIG sample. The weights used are shown in Table 5.11. The branching fractions after this reweighting are shown on the right of Figure 5.40. All $t\bar{t}H(H\rightarrow gg)$ events were omitted in the generation of the HERWIG sample which is why the distributions in Figure 5.40 show branching ratios that don't quite match those in Table 5.10. Therefore, these are removed from the PYTHIA sample to avoid an overall global normalisation difference.

To ensure this has no major impact on the analysis, distributions of the H_{T} variable and other variables that are used as inputs to the NN are studied with and without the removal of $t\bar{t}H(gg)$ events. It is found that in the (4j,4b) region, 0.43% of signal events have a Higgs boson that decays to a gluon pair, which has a negligible effect. In the (4j,2b) region, 9.1% of signal events have a Higgs boson decaying to a gluon pair. Whilst this is a large fraction of the total number of signal events, their contribution to the H_{T} variable, which is used as the discriminating variable in this region, is negligible compared to the overwhelming background. It was therefore deemed reasonable to remove these events.



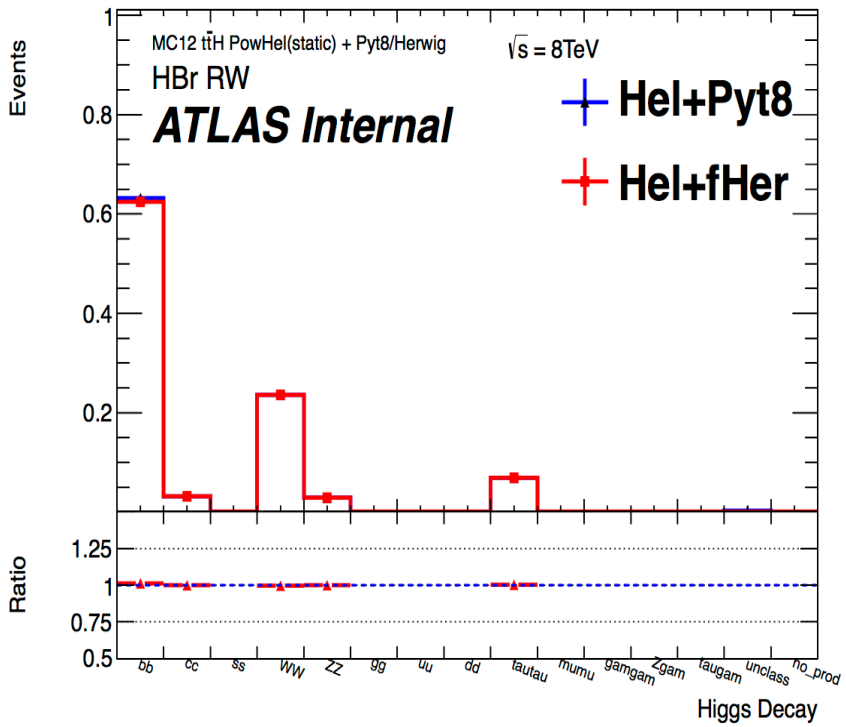


Figure 5.40: Branching ratios of the combined $t\bar{t}H$ dilepton, single lepton and all hadronic POWHEL+PYTHIA samples and POWHEL+HERWIG samples before (left) and after (right) the branching ratio correction. These distributions are pre-selection and are normalised to unity after all $t\bar{t}H(H\rightarrow gg)$ events removed from the POWHEL+PYTHIA sample.

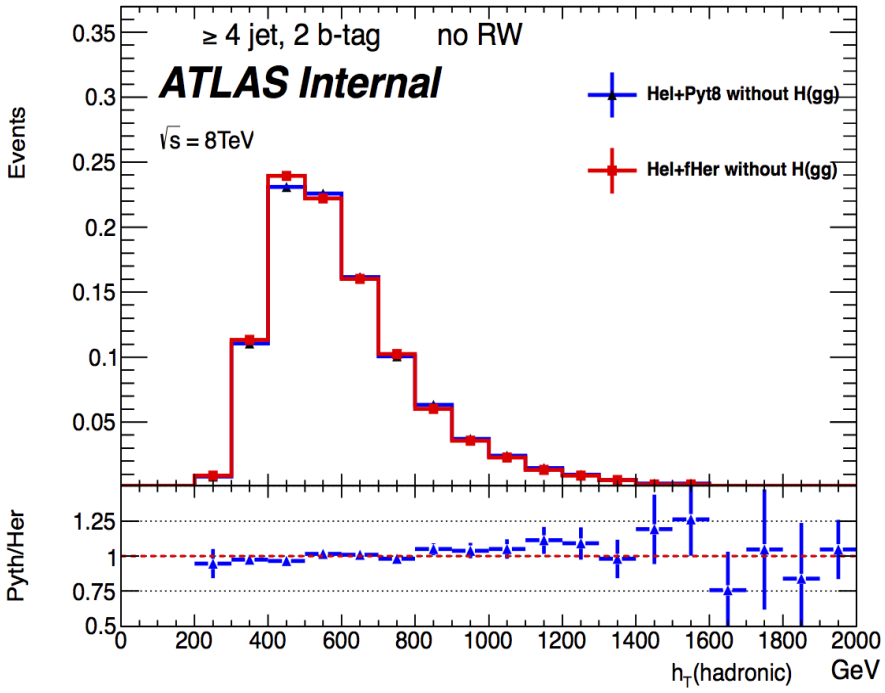
The HERWIG sample was generated with limited statistics which means one cannot use it to obtain a good estimate for the parton shower systematic uncertainty. Due to this, the nominal PYTHIA (with higher statistics and $t\bar{t}H(H\rightarrow gg)$ events removed) sample is reweighted so that the kinematic distributions match those of the corrected HERWIG sample. A variety of kinematic reweighting methods were derived pre-selection at truth level. The agreement between several post-selection kinematic distributions were studied to validate the procedure. $t\bar{t}H$ p_{T} reweighting provided the best agreement between the two samples. The reweighting values are shown appendix C.

In Figure 5.41 (a) one can see the disagreement between the H_{T} distributions of the two samples using different parton shower models before the reweighting is applied. This difference is essentially the parton shower systematic uncertainty. The H_{T} variable is used here to demonstrate the kinematic differences between the samples because it is the sum of the p_{T} of all of the jets and leptons in an event and is therefore sensitive to kinematic differences between the samples. After applying the $t\bar{t}H$ p_{T} reweighting to the PYTHIA sample, the distributions are seen to be in much better agreement. This is shown in Figure 5.41 (b) for events in the (4j,2b) analysis region, which has the largest number of signal events. The reweighted PYTHIA sample is then used in the final fit instead of the HERWIG sample to evaluate the systematic uncertainty.

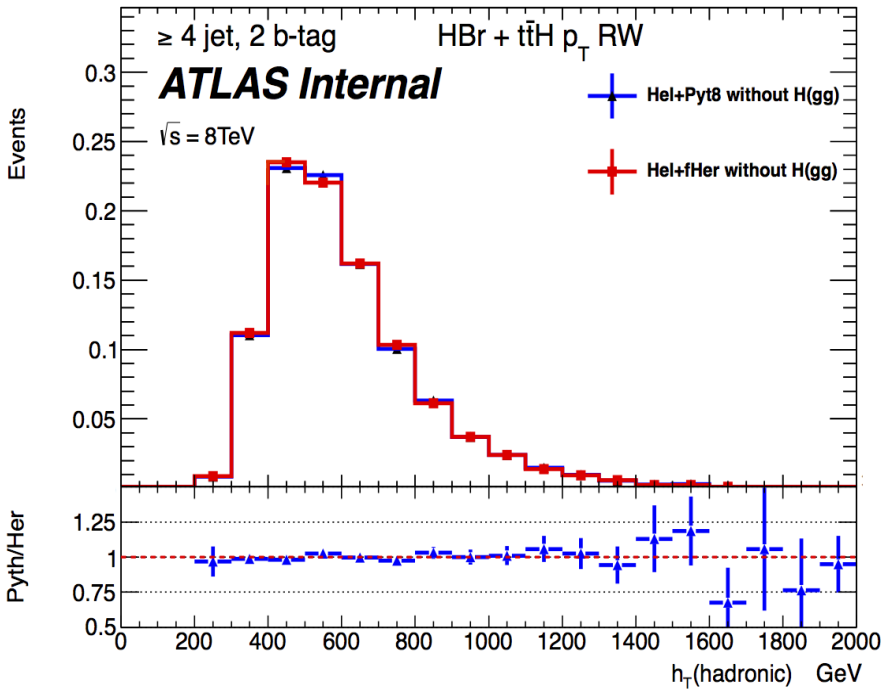
The percentage pre(post)-fit normalisation uncertainty associated with the choice of signal sample parton shower model, on the $t\bar{t}H$ sample alone for events in the ($\geq 4\mathrm{j}, \geq 4\mathrm{b}$) dilepton analysis region was found to be $\pm 1.8\%(\pm 0.01\%)$. Using the full set of simulated samples, this pre(post)-fit normalisation uncertainty was reduced to $\pm 0.03\%(< \pm 0.01\%)$ for the same region.

Higgs boson Decay channel	Channel Weight
$b\bar{b}$	0.892 ± 0.001
$c\bar{c}$	0.712 ± 0.005
WW	1.367 ± 0.004
ZZ	1.359 ± 0.001
$\tau\tau$	1.300 ± 0.007

Table 5.11: Higgs boson branching ratio reweighting values that are applied to the $t\bar{t}H$ HERWIG sample.



(a)



(b)

Figure 5.41: Distributions of $H_{\mathrm{T}}^{\mathrm{had}}$ in the $\geq 4\mathrm{j} = 2\mathrm{b}$ after applying the Higgs boson branching fraction correction, after removing $\mathrm{t\bar{t}H(H\rightarrow gg)}$ events and (a) before any reweighting and (b) after the $\mathrm{t\bar{t}H}$ p_{T} reweighting.

5.15 Statistical Analysis and Results

The discriminant distributions from all analysis regions are combined (dilepton analysis is also combined with the single lepton analysis) to test for the presences of a signal assuming a Higgs boson mass of 125 GeV. Each bin is considered to be like a Poisson counting experiment with N bins in the histogram, each providing a measurement. Poisson probability terms are calculated for each bin in the discriminant distributions using Equation 4.11. A likelihood function is constructed from the product of the Poisson probabilities over all bins combining distributions from all analysis regions.

The likelihood function is dependent on the ratio of the number of observed/expected events with respect to the number predicted by the SM known as the signal strength parameter defined as shown in Equation 5.15. It is also dependent on a set of nuisance parameters labelled θ, which represent the effects of systematic uncertainties on signal and background expectations. The number of events in any given bin is therefore dependent on μ and θ.

$$\mu = \frac{\sigma(t\bar{t}H)}{\sigma_{SM}(t\bar{t}H)} \tag{5.15}$$

The likelihood function is fit to data where the nuisance parameters will adjust the expectations for signal and background according to their associated systematic uncertainty. Nuisance parameters are implemented in the likelihood function as priors. The final fitted value of a nuisance parameter corresponds to the value that provides the best fit to data. The impact of systematic uncertainties upon the search sensitivity is constrained by using the highly populated signal-deficient regions in the fitting procedure. This requires a good understanding of the background processes and the effect each systematic uncertainty has on the discriminant distributions.

The different analysis regions have different contributions from various backgrounds and so the likelihood fit is used to constrain them. The highly populated $\geq 4\,\mathrm{j} = 2\,\mathrm{b}$ (4j,2b) channel provides a strong constraint on the overall $t\bar{t}$ background normalisation. The $(= 2\,\mathrm{j} = 2\,\mathrm{b})$(2j,2b), $(= 3\,\mathrm{j} = 2\,\mathrm{b})$ (3j,2b) and $(\geq 4\,\mathrm{j} = 2\,\mathrm{b})$ regions consist almost entirely of $t\bar{t}+$ light background events and provides a constraint on the $t\bar{t}$ modelling uncertainties. Regions with exactly 3 and ≥ 4 b-jets have different fractions of $t\bar{t}$+$b\bar{b}$ and $t\bar{t}+c\bar{c}$ backgrounds, so their interplay during the likelihood fit provides constraints on each of their normalisation uncertainties.

A comparison between the pre-fit and post-fit yields in data and MC for the various dilepton

analysis regions along with the total uncertainty are shown in Figures 5.42 and 5.43 respectively.

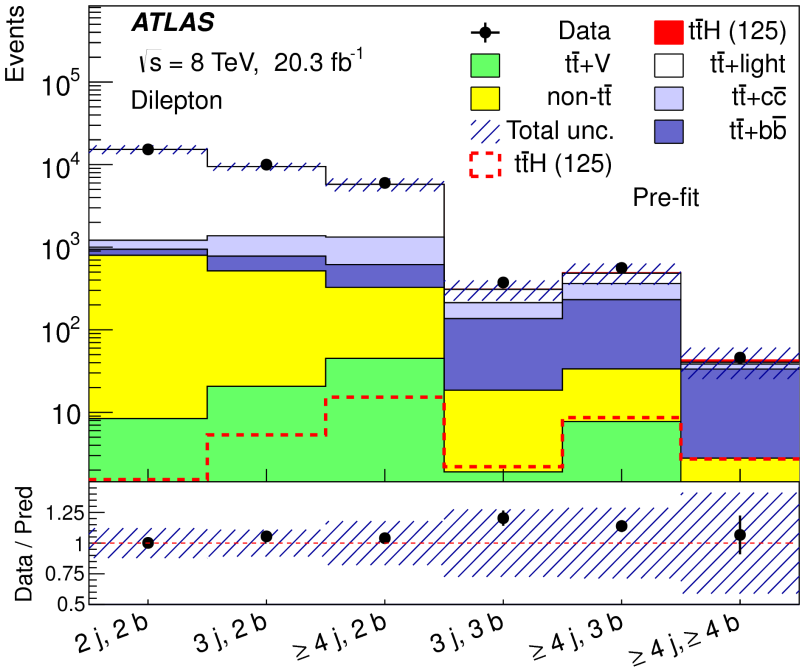


Figure 5.42: Data/MC comparison of pre-fit yields under the signal-plus-background hypothesis in each dilepton analysis region. The uncertainties are the statistical plus systematic uncertainties on the yields summed in quadrature. The signal is normalised to the fitted μ and is shown as both a filled area stacked on the backgrounds and separately as a dashed red line.

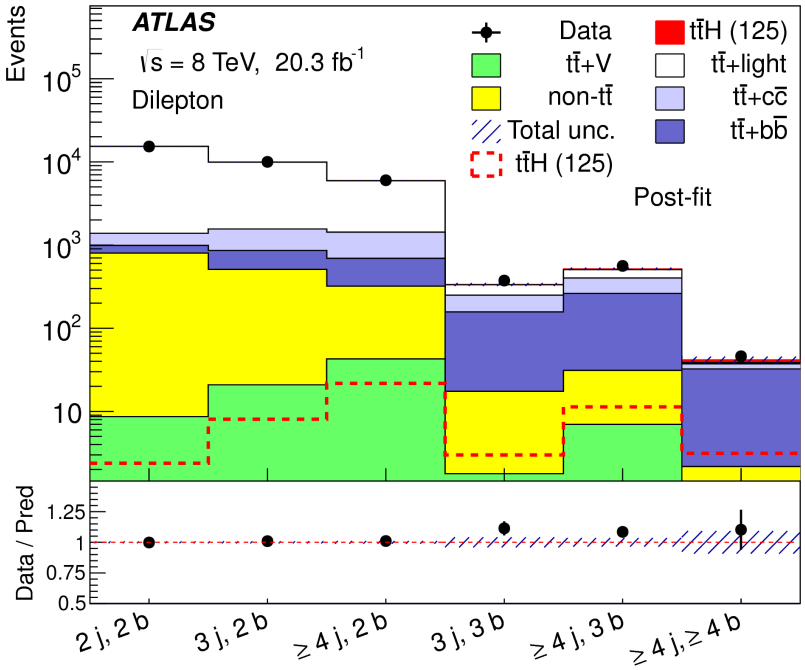


Figure 5.43: Data/MC comparison of post-fit yields under the signal-plus-background hypothesis in each dilepton analysis region. The uncertainties are the statistical plus systematic uncertainties on the yields summed in quadrature. The signal is normalised to the fitted μ and is shown as both a filled area stacked on the backgrounds and separately as a dashed red line.

The likelihood is used to calculate the test statistic q_μ as described in Section 5.1, which is used to measure the compatibility of the observed data with the background-only hypothesis ($\mu = 0$) and to make statistical inferences about μ such as upper limits using the CLs technique ([85],[86]). This provides a confidence interval which, upon repeating the experiment on new independent samples, frequently includes the value of an unobservable parameter of interest. The frequency with which the observed interval contains the parameter is determined by the % of the confidence level.

Both observed and expected confidence intervals are calculated. The observed confidence interval gives the upper limit on $\sigma(t\bar{t}H)/\sigma_{SM}(t\bar{t}H)$ using the experimental data observed. The expected confidence interval gives the expected upper limit on the value of $\sigma(t\bar{t}H)/\sigma_{SM}(t\bar{t}H)$ under either the background only hypothesis (signal strength parameter $\mu = 0$) or the signal-plus-background parameter (signal strength parameter $\mu = 1$) as predicted from simulations.

The discriminant distributions are fit to data in all regions simultaneously to produce the final result. This uses the signal-plus-background hypothesis and the signal strength parameter is allowed to float freely although it is constrained to be the same in all regions. The normalisation of each background is determined by fitting μ simultaneously. The fitted value of the signal strength

in each channel and their combined values along with their associated total uncertainties is shown in Figure 5.44.

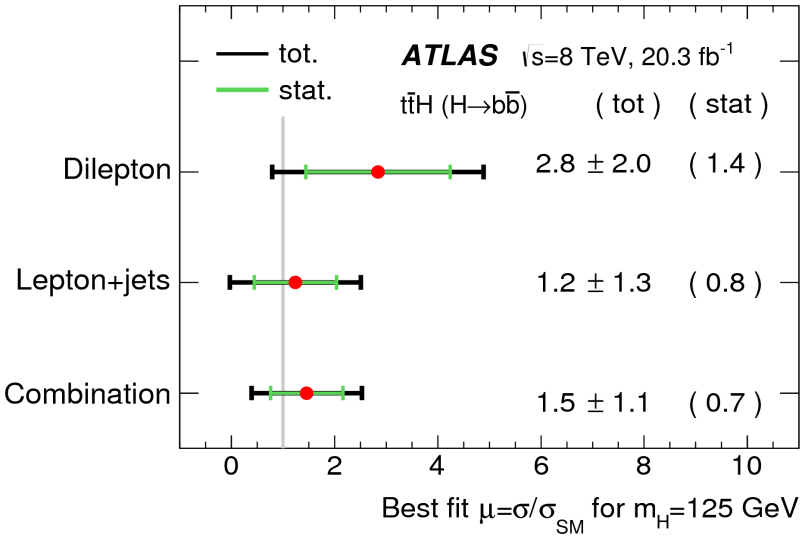


Figure 5.44: Values of the fitted signal strength (μ) in lepton+jets (i.e. single lepton channel) and dilepton channel along with the two channels combined and the associated uncertainties. The green line is the statistical uncertainty on μ.

95% CL upper limits on the $t\bar{t}H$ cross-section with respect to the SM prediction are shown in Figure 5.45 and Table 5.12. The black dashed line represents the expected upper limit under the background only hypothesis (signal strength parameter $\mu = 0$). The red dashed line represents the expected limit under the signal-plus-background hypothesis (signal strength parameter $\mu = 1$). The observed upper limit, represented by the solid black line is the upper limit obtained from data.

For the case where no Higgs boson exists, the possibility of observing a $t\bar{t}H$ signal in the dilepton channel 4.1 times stronger than that predicted by the SM Higgs boson in a repeated experiment is ruled out with 95% CL. For the case in which a SM Higgs boson exists, the possibility of observing a $t\bar{t}H$ signal in the dilepton channel 4.7 times stronger than that predicted by the SM Higgs boson is ruled out with 95% CL.

For the data observed in the dilepton analysis, the possibility of observing a $t\bar{t}H$ signal 6.7 times stronger than that predicted by the SM Higgs boson is ruled out with 95% CL. In all cases the SM Higgs boson is assumed to have $m_H = 125\,\text{GeV}$.

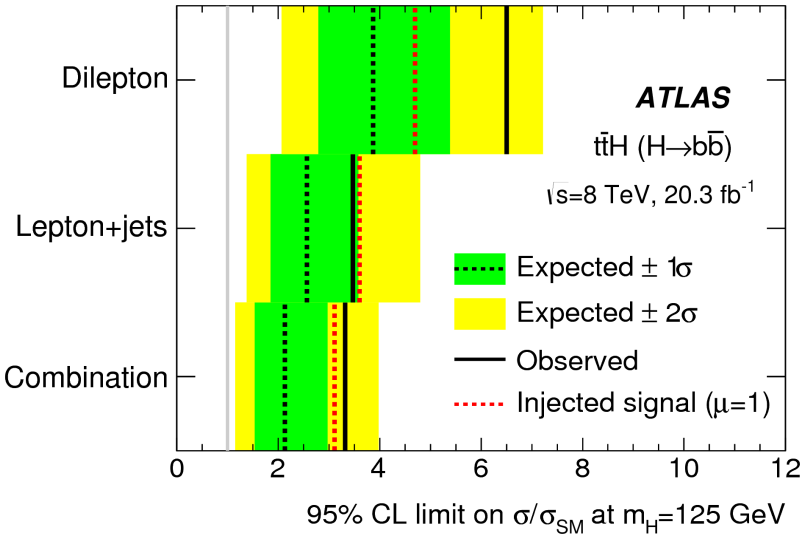


Figure 5.45: 95% CL upper limit on the cross-section $\sigma(t\bar{t}H)$ with respect to the SM prediction $\mu = \frac{\sigma(t\bar{t}H)}{\sigma_{SM}(t\bar{t}H)}$. Each channels individual contribution to the combined upper limit are shown along with the combined upper limit. The observed limits are represented by the solid lines. These can be compared with the expected limits under the background-only and signal-plus-background hypotheses. In all cases the backgrounds are predicted using their pre-fit distributions. 68% and 95% confidence intervals for the expected limit under the background only hypothesis are represented by the green and yellow bands.

In order to be able to claim an observation of the $t\bar{t}H$ process, one requires sensitivity to the Statndard Model $\sigma(t\bar{t}H)/\sigma_{SM}(t\bar{t}H) = 1$. This means the analyis needs to have an expected 95% CL upper limit under the signal-plus-background hypothesis (red dashed line in Figure 5.45) equal to 1. One also needs to be able to rule out the background only hypothesis, which means the observed limit has a 5σ significance from the limit expected under the background only hypothesis. The analysis does not have this sensitivity, and no significant excess of events above the background expectation can be claimed for a SM Higgs boson with mass 125 GeV. The observed and expected 95% CL upper limits are summarised in Table 5.12. The fitted value of the signal strength parameter in the dilepton channel is equal to $\mu = 2.8 \pm 2.0$.

Observed	-2σ	-1σ	Expected ($\mu = 0$)	1σ	2σ	Expected ($\mu = 1$)
6.7	2.2	3.0	4.1	5.8	7.7	4.7

Table 5.12: Observed and expected 95% CL upper limits on the $t\bar{t}H$ signal strength for the dilepton channel for a Higgs boson with $m_H = 125$ GeV. Confidence intervals about the expected upper limit for the background only hypothesis are shown as $\pm 1\sigma$ and $\pm 2\sigma$ along with the upper limit. The 95% CL upper limit for the signal strength under the signal-plus-background hypothesis is also shown.

Chapter 6

Conclusions

> *And that, my liege, is how we know the Earth to be banana shaped.*
>
> Sir Bedivere, *Monty Python and the Holy Grail*

In this thesis I have presented the first measurement of a potential bias in b-tagging calibrations performed on b-jets selected using the low momentum muon (soft muon) coming from the decay of the b hadron. Several calibration methods used by the ATLAS experiment are derived on this type of b-jet which highlights the importance of this study. To carry out this measurement, I selected an enriched sample of b-jets by targeting $t\bar{t}$ production with two charged leptons in the final state, due to the top quarks large branching fraction to b-quarks. The sample of b-jets was divided into two categories: those which decayed to a muon plus a neutrino, and those which decayed to another lepton (i.e. τ or electron) or hadronically. b-tagging efficiencies and scale factors were then derived separately for each category of jet. The ratio of the scale factors for each jet category is then used as a measure of the potential bias. This is done in two jet p_{T} bins of 0-90 GeV and 90-300 GeV due to the limited statistics available for jets that containing a low momentum muon. The ratio of scale factors is consistent with unity in both jet p_{T} bins. The uncertainty on the measurement of the bias is $\approx 8\%$ in both p_{T} bins.

I have also presented my work on the $\mathrm{t\bar{t}H(H\rightarrow b\bar{b})}$ analysis, part of which was a study of possible pairing algorithms that can be used to correctly match b-jets with the b-partons that originated from the Higgs boson. Several pairing methods were studied to try and improve upon the poor pairing efficiency of the pairing method used in Reference [58]. Pairing methods with better pairing

efficiencies allow one to reconstruct more complicated and better discriminating neural network input variables. The method used in Reference [58] was only able to correctly match the b-jets with the b hadrons originating from the Higgs boson in 31% of true $t\bar{t}H(H\rightarrow b\bar{b})$ events. However, the method I have called the dRIso dM method (see Section 5.11.3) obtained an efficiency of 49%, which is a significant improvement.

A study of alternative multivariate classifiers for the $t\bar{t}H(H\rightarrow b\bar{b})$ analysis is also presented. The aim of the study was to improve upon the neural networks ability to separate background from signal events. The performance of the neural network used in Reference [58] (10 input variables using dR pairing method) is compared with the performance of a boosted decision tree trained using the same events, variables and pairing method. The performance of each of the classifiers was also evaluated for all of the pairing methods individually. The ability of the multivariate classifiers to separate signal from background where compared using a metric called "speparation", which is defined in Equation 5.14. The best separation for the boosted decision tree was achieved when using the dR dM method which obtained a separation of 0.192 ± 0.003. The best separation for the neural network was achieved when using the so-called dR dM method (see Section 5.11.3) which obtained a separation of 0.193 ± 0.002. No significant improvement was seen when using a boosted decision tree trained using 10 input variables.

Further research into improving the performances of the multivariate classifiers was performed by looking at including additional input variables. Several additional variables, based on object pair properties, were added to the input variable list of both classifiers. The best separation for the 17 variable boosted decision tree was achieved when using the dR dM method which obtained a separation of 0.201 ± 0.003, whereas the best separation for the 17 variable neural network was achieved when using the dR dM method which obtained a separation of 0.195 ± 0.002. Therefore, a small improvement can be seen in the separation power of both classifiers when using additional variables. Furthermore, the 17 variable boosted decision tree has greater separation power than the 17 variable neural network, as is shown in Table 6.1.

Classifier	Separation
BDT (10 inputs)	0.192 ± 0.03
NN (10 inputs)	0.193 ± 0.02
BDT (17 inputs)	0.201 ± 0.03
NN (17 inputs)	0.195 ± 0.02

Table 6.1: Comparison of the separation power of the two classifiers investigated: boosted decision tree (BDT) and neural network (NN). The separation for the classifiers trained on 10 or 17 input variables using the dR dM method are compared. The uncertainties shown are the statistical only uncertainties on the separation.

This investigation into several possible methods that can be used to reconstruct $t\bar{t}H$ events provides an insight into a possible way in which the $t\bar{t}H$ analysis can be improved. Any future search for $t\bar{t}H$ production at the LHC could use these techniques (or similar methods) to reconstruct physics objects and improve the signal sensitivity of the analysis.

A full description of the evaluation of the signal sample parton shower systematic uncertainty in the $t\bar{t}H(H\rightarrow b\bar{b})$ analysis is also presented. This uncertainty is evaluated by rerunning the analysis using an alternative signal sample parton shower model. The systematic sample was originally generated using the HERWIG parton shower model, however this sample had particularly low statistics and did not provide a very good estimate of the systematic uncertainty. Instead the nominal signal sample is reweighted to provide a high statistics systematic sample. This reweighted sample is the used in the final fit to evaluate the impact of this uncertainty. A total post-fit normalisation systematic uncertainty of $\pm 0.01\%$ is found with respect to the nominal signal normalisation in dilepton events in the $(\geq 4\,\text{j}, \geq 4\,\text{b})$ region.

No evidence for $t\bar{t}H$ production was observed in data taken by the ATLAS detector during the 2009-2013 data taking period. The $t\bar{t}H$ analysis presented in Reference [58] did not have the sensitivity required to claim the observation of the $t\bar{t}H$ process. However, the additional techniques presented such as the alternative pairing methods and the optimisation of the multivariate methods could improve the sensitivity of the analysis for the next data taking period.

Appendices

Appendix A

Simulated Cross-Sections

Theoretical Cross-Sections and K-factors		
Process	**Cross-section** (pb)	**K-factor**
Top pair Powheg+Pythia (full sim)		
$t\bar{t}$	114.51	1.1992
Single Top AcerMC+Pythia (full sim)		
Single top: t-channel	25.750	1.1042
Single Top Powheg+Pythia (full sim)		
Single top: s-channel	1.6424	1.1067
Single top: Wt-channel Incl. (Diagram Removal)	20.461	1.0933
Diboson Alpgen+Herwig (full sim)		
$WW \to lvlv$ + Np0	2.4970	1.2307
$WW \to lvlv$ + Np1	1.2491	1.2307
$WW \to lvlv$ + Np2	0.59200	1.2307
$WW \to lvlv$ + Np3	0.32847	1.2307
$W(\to incl.)Z(\to ll)$ + Np0	0.80162	1.2785
$W(\to incl.)Z(\to ll)$ + Np1	0.52407	1.2785
$W(\to incl.)Z(\to ll)$ + Np2	0.29484	1.2785
$W(\to incl.)Z(\to ll)$ + Np3	0.18258	1.2785
$Z(\to incl.)Z(\to ll)$ + Np0	0.60660	1.3718
$Z(\to incl.)Z(\to ll)$ + Np1	0.28893	1.3718
$Z(\to incl.)Z(\to ll)$ + Np2	0.11853	1.3718
$Z(\to incl.)Z(\to ll)$ + Np3	0.056223	1.3718
$WW \to qqlv$ + Np0	9.9819	1.2832
$WW \to qqlv$ + Np1	5.0144	1.2832
$WW \to qqlv$ + Np2	2.3658	1.2832
$WW \to qqlv$ + Np3	1.3139	1.2832
Z+Jets Alpgen+Pythia unfiltered (full sim)		
$Z \to ee$ + Np0	718.89	1.18
$Z \to ee$ + Np1	175.60	1.18
$Z \to ee$ + Np2	58.849	1.18
$Z \to ee$ + Np3	15.560	1.18
$Z \to ee$ + Np4	3.9322	1.18
$Z \to ee$ + Np5	1.1994	1.18
$Z \to \mu\mu$ + Np0	718.91	1.18
$Z \to \mu\mu$ + Np1	175.81	1.18
$Z \to \mu\mu$ + Np2	58.805	1.18
$Z \to \mu\mu$ + Np3	15.589	1.18
$Z \to \mu\mu$ + Np4	3.9072	1.18
$Z \to \mu\mu$ + Np5	1.1933	1.18
$Z \to \tau\tau$ + Np0	718.80	1.18
$Z \to \tau\tau$ + Np1	175.83	1.18
$Z \to \tau\tau$ + Np2	58.630	1.18
$Z \to \tau\tau$ + Np3	15.508	1.18
$Z \to \tau\tau$ + Np4	3.9526	1.18
$Z \to \tau\tau$ + Np5	1.1805	1.18

Table A.1: The theoretical cross-sections and K-factors of the signal and background processes. “NP” stand for the number of additional hard partons generated by the Monte Carlo event generator.

Theotretical Cross-Sections and K-factors		
Process	**Cross-section (pb)**	**K-factor**
Z+jets+bb Alpgen+Pythia unfiltered samples		
$Z \to ee + bb$ + Np0	8.0397	1.18
$Z \to ee + bb$ + Np1	3.2353	1.18
$Z \to ee + bb$ + Np2	1.1388	1.18
$Z \to ee + bb$ + Np3	0.49066	1.18
$Z \to \mu\mu + bb$ + Np0	8.0422	1.18
$Z \to \mu\mu + bb$ + Np1	3.2155	1.18
$Z \to \mu\mu + bb$ + Np2	1.1400	1.18
$Z \to \mu\mu + bb$ + Np3	0.50943	1.18
$Z \to \tau\tau + bb$ + Np0	8.0358	1.18
$Z \to \tau\tau + bb$ + Np1	3.2299	1.18
$Z \to \tau\tau + bb$ + Np2	1.1445	1.18
$Z \to \tau\tau + bb$ + Np3	0.49266	1.18
Z+jets+cc Alpgen+pythia unfiltered samples		
$Z \to ee + cc$ + Np0	15.107	1.18
$Z \to ee + cc$ + Np1	7.2131	1.18
$Z \to ee + cc$ + Np2	3.0320	1.18
$Z \to ee + cc$ + Np3	1.1767	1.18
$Z \to \mu\mu + cc$ + Np0	15.115	1.18
$Z \to \mu\mu + cc$ + Np1	7.1980	1.18
$Z \to \mu\mu + cc$ + Np2	3.0303	1.18
$Z \to \mu\mu + cc$ + Np3	1.1738	1.18
$Z \to \tau\tau + cc$ + Np0	15.119	1.18
$Z \to \tau\tau + cc$ + Np1	7.2016	1.18
$Z \to \tau\tau + cc$ + Np2	3.0385	1.18
$Z \to \tau\tau + cc$ + Np3	1.1677	1.18
Z+jets Alpgen+herwig low mass		
$Z \to ee$ + Np0 excl. with 10 GeV $< M(ll) <$ 60 GeV	3477.9	1.195
$Z \to ee$ + Np1 excl. with 10 GeV $< M(ll) <$ 60 GeV	108.88	1.195
$Z \to ee$ + Np2 excl. with 10 GeV $< M(ll) <$ 60 GeV	52.651	1.195
$Z \to ee$ + Np3 excl. with 10 GeV $< M(ll) <$ 60 GeV	11.309	1.195
$Z \to ee$ + Np4 excl. with 10 GeV $< M(ll) <$ 60 GeV	2.5743	1.195
$Z \to ee$ + Np5 excl. with 10 GeV $< M(ll) <$ 60 GeV	0.69211	1.195
$Z \to \mu\mu$ + Np0 excl. with 10 GeV $< M(ll) <$ 60 GeV	3477.8	1.195
$Z \to \mu\mu$ + Np1 excl. with 10 GeV $< M(ll) <$ 60 GeV	108.63	1.195
$Z \to \mu\mu$ + Np2 excl. with 10 GeV $< M(ll) <$ 60 GeV	52.675	1.195
$Z \to \mu\mu$ + Np3 excl. with 10 GeV $< M(ll) <$ 60 GeV	11.283	1.195
$Z \to \mu\mu$ + Np4 excl. with 10 GeV $< M(ll) <$ 60 GeV	2.5690	1.195
$Z \to \mu\mu$ + Np5 excl. with 10 GeV $< M(ll) <$ 60 GeV	0.69425	1.195
$Z \to \tau\tau$ + Np0 excl. with 10 GeV $< M(ll) <$ 60 GeV	3478.1	1.195
$Z \to \tau\tau$ + Np1 excl. with 10 GeV $< M(ll) <$ 60 GeV	108.85	1.195
$Z \to \tau\tau$ + Np2 excl. with 10 GeV $< M(ll) <$ 60 GeV	52.777	1.195
$Z \to \tau\tau$ + Np3 excl. with 10 GeV $< M(ll) <$ 60 GeV	11.295	1.195
$Z \to \tau\tau$ + Np4 excl. with 10 GeV $< M(ll) <$ 60 GeV	2.5904	1.195
$Z \to \tau\tau$ + Np5 excl. with 10 GeV $< M(ll) <$ 60 GeV	0.69034	1.195

Table A.2: The theoretical cross-sections and K-factors of the signal and background processes. “NP” stand for the number of additional hard partons generated by the Monte Carlo event generator.

Appendix B

B-tagging Systematic Uncertainties

@70% channel_EM		
Systematic	0-90GeV	90-300GeV
$t\bar{t}$ Monte Carlo Generator 70%	-1.26653	-2.94177
$t\bar{t}$ parton Shower 70%	-1.99907	3.37943
$t\bar{t}$ Initial/final state radiation modelling 70% (+/-)	1.6105	3.53564
Baseline *in situ* (JES) down	-0.988626	0.592231
Baseline *in situ* (JES) up	1.08112	-0.746439
B-jet (JES) uncertainty down	0.145874	1.30732
B-jet (JES) uncertainty up	-0.042665	-1.1619
Close-by (JES) down	-0.158564	-0.0163813
Close-by (JES) up	0.0856373	0.17076
Electron energy resolution down	0.0673778	-0.00327586
Electron energy resolution up	-0.105339	-0.137636
Electron energy scale down	0.142956	-0.446219
Electron energy scale up	0.366614	-0.21031
Electron reconstruction/identification efficiency scale factor down	0.0255944	-0.130904
Electron reconstruction/identification efficiency scale factor up	-0.0506196	0.131527
Electron trigger efficiency scale factor down	0.00131601	0.0472447
Electron trigger efficiency scale factor up	-0.0325099	-0.0274437
η-intercalibration (JES) down	-0.274228	0.0455364
η-intercalibration (JES) up	0.408604	0.20139
Flavor composition (JES) uncertainty down	-1.04243	-1.02206
Flavour composition (JES) uncertainty up	0.977798	0.496057
Flavour response (JES) down	-0.774681	-0.615706
Flavour response (JES) up	0.522248	0.251007
Jet reconstruction efficiency	0.106594	0.0218507
Jet energy resolution	0.931561	0.361453
Jet vertex fraction scale factor down	-0.131803	0.0215466
Jet vertex fraction scale factor up	0.0374598	0.209874
Muon reconstruction/identification efficiency scale factor down	0.0129321	0.0754325
Muon reconstruction/identification efficiency scale factor up	-0.0428697	-0.0772957
Muon trigger scale factor down	0.137922	-0.00657513
Muon trigger scale factor up	-0.0373095	-0.0352224
Pileup Offset μ (JES) down	-0.380338	0.369905
Pileup Offset μ (JES) up	0.0672015	-0.138827
Pileup offset number of reconstructed primary vertices (JES) down	-0.0536857	0.0684607
Pileup offset number of reconstructed primary vertices (JES) up	0.0750484	-0.361198
Charm mistag scale factor down	0.0649064	0.0268843
Charm mistag scale factor up	-0.0536751	-0.021101
Light mistag scale factor down	0.180008	0.123278
Light mistag scale factor up	-0.177892	-0.121988
$t\bar{t}$ cross-section down	0.653905	0.461423
$t\bar{t}$ cross-section up	-0.504236	-0.354338
Z + jets cross-section down	-0.119135	-0.129545
Z + jets cross-section up	0.118806	0.12918
Diboson cross-section down	-0.0694956	-0.0206729
Diboson cross-section up	0.0693885	0.0206611
Single top (Wt) cross-section down	-0.0505563	-0.0187401
Single top (Wt) cross-section up	0.0502119	0.0186387
Fakes cross-section down	-0.37636	-0.0986341
Fakes cross-section up	0.372622	0.0977815
Semi-leptonic jet correction down	0.571734	-0.621832
Semi-leptonic jet correction up	-0.628812	0.514782
Non-SL Stat. Uncert.	0.880092	1.28891
SL Stat. Uncert.	3.06275	3.6219
Total Statistical	3.18669	3.84441
Total Systematic	3.56509	6.01269
Total Uncertainty	6.75179	9.85709

Table B.1: Systematic uncertainties in bins of jet p_{T}, on the final b-tagging bias $\beta^{SL,NSL}$. The numbers given are percentage variations from the nominal value. The naming convention dictates that any uncertainty with (JES) in the name contributes to the total jet energy scale uncertainty.

Appendix C

Sample Weights

$t\bar{t}H$ system pt	Weight (post branching ratio correction)	Error (stat only)
$0 < p_{\mathrm{T}} \leq 10$	0.969	$\pm$0.002
$10 < p_{\mathrm{T}} \leq 20$	0.973	$\pm$0.002
$20 < p_{\mathrm{T}} \leq 30$	0.981	$\pm$0.002
$30 < p_{\mathrm{T}} \leq 40$	0.989	$\pm$0.003
$40 < p_{\mathrm{T}} \leq 50$	0.999	$\pm$0.003
$50 < p_{\mathrm{T}} \leq 60$	1.003	$\pm$0.003
$60 < p_{\mathrm{T}} \leq 70$	1.010	$\pm$0.004
$70 < p_{\mathrm{T}} \leq 80$	1.027	$\pm$0.005
$80 < p_{\mathrm{T}} \leq 90$	1.029	$\pm$0.005
$90 < p_{\mathrm{T}} \leq 100$	1.017	$\pm$0.005
$100 < p_{\mathrm{T}} \leq 110$	1.035	$\pm$0.006
$110 < p_{\mathrm{T}} \leq 120$	1.025	$\pm$0.006
$120 < p_{\mathrm{T}} \leq 130$	1.031	$\pm$0.007
$130 < p_{\mathrm{T}} \leq 140$	1.036	$\pm$0.007
$140 < p_{\mathrm{T}} \leq 150$	1.036	$\pm$0.008
$150 < p_{\mathrm{T}} \leq 160$	1.052	$\pm$0.008
$160 < p_{\mathrm{T}} \leq 170$	1.048	$\pm$0.009
$170 < p_{\mathrm{T}} \leq 180$	1.040	$\pm$0.009
$180 < p_{\mathrm{T}} \leq 190$	1.06	$\pm$0.01
$190 < p_{\mathrm{T}} \leq 200$	1.06	$\pm$0.01
$200 < p_{\mathrm{T}} \leq 210$	1.06	$\pm$0.01
$210 < p_{\mathrm{T}} \leq 220$	1.058	$\pm$0.008
$220 < p_{\mathrm{T}} \leq 240$	1.058	$\pm$0.005
$240 < p_{\mathrm{T}} \leq 320$	1.098	$\pm$0.004
$320 < p_{\mathrm{T}}$	1.113	$\pm$0.008

Table C.1: $\mathrm{t\bar{t}H}$ p_{T} reweighting values (nominal / HERWIG).

Appendix D

$t\bar{t}H(H\to b\bar{b})$ Systematic Uncertainties

Systematic uncertainty	Type	Comp.
Luminosity	N	1
Physics Objects		
Electron	SN	5
Muon	SN	6
Jet energy scale	SN	22
Jet vertex fraction	SN	1
Jet energy resolution	SN	1
Jet reconstruction	SN	1
b-tagging efficiency	SN	6
c-tagging efficiency	SN	4
Light-jet tagging efficiency	SN	12
High-p_{T} tagging efficiency	SN	1
Background Model		
$t\bar{t}$ cross section	N	1
$t\bar{t}$ modelling: p_{T} reweighting	SN	9
$t\bar{t}$ modelling: parton shower	SN	3
$t\bar{t}$+heavy-flavour: normalisation	N	2
$t\bar{t}$+$c\bar{c}$: p_{T} reweighting	SN	2
$t\bar{t}$+$c\bar{c}$: generator	SN	4
$t\bar{t}$+$b\bar{b}$: NLO Shape	SN	8
W+jets normalisation	N	3
W p_{T} reweighting	SN	1
Z+jets normalisation	N	3
Z p_{T} reweighting	SN	1
Lepton misID normalisation	N	3
Lepton misID shape	S	3
Single top cross section	N	1
Single top model	SN	1
Diboson+jets normalisation	N	3
$t\bar{t}+V$ cross section	N	1
$t\bar{t}+V$ model	SN	1
Signal Model		
$t\bar{t}H$ scale	SN	2
$t\bar{t}H$ generator	SN	1
$t\bar{t}H$ hadronisation	SN	1
$t\bar{t}H$ PDF	SN	1

Figure D.1: Full list of systematic uncertainties considered in the dilepton $t\bar{t}H(H\to b\bar{b})$ analysis [58]. If an uncertainty is labelled with an "N", it means the uncertainty is taken on normalisation-only for all processes and in all channels. "S" means the uncertainty is taken on a shape-only uncertainty in all processes for all channels. "SN" means the uncertainty is taken on both shape and normalisation. The number in the column labelled "comp" is the number of components the systematic has been split into. Systematics are split into component parts if a more accurate treatment can be obtained by doing so.

Appendix E

$t\bar{t}H(H\rightarrow b\bar{b})$ Dilepton Cutflow

Cut	Data Events Passing
Initial	17030368
GRL (data only)	17030368
Cosmic	17030368
Good vertex	17030368
Two leptons of right type	17021716
Trigger	15553053
At least 1 lepton pt $>$ 25 GeV	15408782
Trigger matched	15349283
E-Mu overlap removal	15346482
Jet Cleaning	15329082
HT	15267668
Exactly 2 Jets	686783
OS leptons	681422
Dilepton mass $>$ 15 GeV	671863
Truth matched leptons	671863
Z mass veto	176322
1 btag	52169
2 btags	17582

Table E.1: Cutflow table showing the number of data events passing each cut in the $t\bar{t}H(H\rightarrow b\bar{b})$ dilepton analysis.

4 j, $\geq$ 4 b											
	$t\bar{t}H$ (125)	$t\bar{t}$ + light	$t\bar{t}+c\bar{c}$	$t\bar{t}+b\bar{b}$	$t\bar{t}+V$	W+jets	Z+jets	Single top	Diboson	Multijet	Total
BTAGBREAK0	±0.01	–	–	±0.01	–	±0.08	–	–	±0.03	–	±0.01
BTAGBREAK1	±0.10	±0.03	±0.03	±0.74	±0.11	±0.29	±0.20	±0.02	±0.21	–	±0.28
BTAGBREAK2	±2.1	±1.3	±1.2	±1.9	±1.8	±0.07	±0.27	±0.12	±0.20	–	±1.4
BTAGBREAK3	±2.7	±1.4	±1.3	±2.1	±2.5	±0.56	±1.1	±0.12	±1.0	–	±1.6
BTAGBREAK4	±2.2	±1.3	±1.3	±2.3	±2.0	±0.80	±1.7	±2.1	±1.7	–	±1.7
BTAGBREAK5	±3.9	±2.2	±2.0	±4.0	±3.6	±3.1	±2.6	±3.1	±1.5	–	±2.9
CTAGBREAK0	–	±0.04	±0.06	±0.01	±0.02	±0.16	±0.52	±0.08	±0.02	–	±0.02
CTAGBREAK1	±0.65	±2.9	±5.3	±1.6	±1.9	±2.1	±4.0	±2.2	±3.5	–	±2.7
CTAGBREAK2	±0.07	±0.75	±1.4	±0.54	±0.77	±0.14	±0.25	±0.19	±1.3	–	±0.71
CTAGBREAK3	±0.81	±3.5	±7.1	±1.9	±2.3	±3.1	±5.2	±3.1	±4.9	–	±3.4
Dibosons_XS	–	–	–	–	–	–	–	–	–	–	±0.00
Dibosons_XS_jet5	–	–	–	–	–	–	–	–	–	–	±0.00
Dibosons_XS_jet6	–	–	–	–	–	–	–	–	–	–	±0.00
Dibosons_XS_LJETS	–	–	–	–	–	–	–	–	±24.8	–	±0.04
ELE_ID	±0.91	±0.93	±0.91	±0.92	±0.91	±0.06	±0.05	±0.03	±0.06	–	±0.84
ELE_RECO	–	–	–	–	–	–	–	–	–	–	±0.00
ELE_RES	–	–	–	–	–	–	–	–	–	–	±0.00
ELE_SCALE	–	–	–	–	–	–	–	–	±19.0	–	±0.03
ELE_TRIG	–	–	–	–	–	–	±0.00	–	–	–	±0.00
JEFF	–	–	–	–	±2.1	–	–	–	–	–	±0.01
JER	±1.4	±0.61	±1.5	±1.4	±2.4	–	±12.9	±0.97	±10.6	–	±0.98
JetAF2Closure	–	–	–	–	–	–	–	–	–	–	±0.00
JetDet1	±0.35	±0.11	±0.12	±0.04	±0.69	–	–	±0.08	–	–	±0.05
JetDet2	–	±0.03	±0.04	–	–	–	–	–	–	–	±0.02
JetDet3	–	–	–	–	±0.06	–	–	–	±23.6	–	±0.04
JetEtaModel	±1.1	±0.63	±0.07	±0.03	±0.65	–	±3.5	–	–	–	±0.19
JetEtaStat	±0.08	±0.01	±0.03	–	±0.62	–	–	±0.12	–	–	±0.00
JetFlavB	±1.00	±1.1	±0.96	±2.3	±0.63	–	±3.6	±3.0	±2.7	–	±1.5
JetFlavComp	±3.4	±0.35	±1.2	±3.0	±2.8	±2.2	±2.4	–	–	–	±1.2
JetFlavResp	±2.4	±0.02	±1.1	±1.9	±2.4	–	–	–	–	–	±0.94
JetMixed1	–	–	–	–	–	–	–	–	–	–	±0.00
JetMixed2	–	–	±0.04	–	–	–	–	±0.08	–	–	±0.01
JetModel1	±1.3	±1.2	±0.02	±0.08	±1.1	±2.3	±4.5	±2.4	±2.8	–	±0.60
JetModel2	±0.13	±0.07	±0.09	–	±0.05	–	–	±0.12	–	–	±0.05
JetModel3	–	±0.04	±0.04	±0.03	±0.04	–	–	±0.09	–	–	±0.04
JetModel4	–	±0.01	–	–	–	–	–	–	–	–	±0.01
JetMu	–	–	–	–	±0.05	–	–	–	–	–	±0.00
JetNPV	±0.02	±0.00	–	±0.64	±0.16	–	–	±0.09	–	–	±0.23
JetPilePt	–	–	–	–	±0.19	–	–	–	–	–	±0.00
JetPileRho	±1.2	±1.0	±0.03	±0.19	±1.4	–	±4.5	±2.1	–	–	±0.52
JetSinglePart	–	–	–	–	–	–	–	–	–	–	±0.00
JetStat1	±0.04	±0.03	–	±0.19	±0.46	–	–	±0.13	–	–	±0.08
JetStat2	–	–	±0.03	±0.07	–	–	–	±0.08	–	–	±0.03
JetStat3	±0.08	±0.02	±0.01	–	±0.04	–	–	±0.06	–	–	±0.00
JVF	±1.2	±0.04	±0.83	±0.70	±0.73	±2.3	–	–	–	–	±0.35
LTAGBREAK0	–	–	–	–	–	–	–	–	–	–	±0.00
LTAGBREAK1	–	–	–	–	–	–	–	–	–	–	±0.00
LTAGBREAK10	–	±4.8	±0.82	±0.03	–	±0.19	±0.15	±0.19	±0.35	–	±1.9

Table E.2: Normalisation uncertainties (expressed in %) on signal and each of the background processes for the systematic uncertainties considered, after the fit to data in $= 4\,\text{jets} \geq 4\,\text{b}-\text{tags}$ region of the single lepton channel. The total uncertainty can be different from the sum in quadrature of individual sources due to the correlations between them [58].

4 j, $\geq$ 4 b											
	$t\bar{t}H$ (125)	$t\bar{t}$ + light	$t\bar{t}+c\bar{c}$	$t\bar{t}+b\bar{b}$	$t\bar{t}+V$	W+jets	Z+jets	Single top	Diboson	Multijet	Total
LTAGBREAK11	±0.30	±9.0	±1.7	±0.85	±1.00	±3.5	±4.5	±4.2	±4.2	–	±4.2
LTAGBREAK2	–	–	–	–	–	–	–	–	–	–	±0.00
LTAGBREAK3	–	–	–	–	–	–	±0.02	–	–	–	±0.00
LTAGBREAK4	–	±0.01	–	–	–	±0.03	±0.04	–	±0.05	–	±0.00
LTAGBREAK5	–	±1.1	–	–	–	±0.03	±0.02	–	–	–	±0.39
LTAGBREAK6	–	±1.4	–	–	–	±0.05	±0.08	–	–	–	±0.51
LTAGBREAK7	–	±0.05	–	–	–	±0.03	±0.03	±0.04	±0.03	–	±0.02
LTAGBREAK8	–	±1.9	–	–	–	±0.02	±0.09	±0.07	±0.02	–	±0.69
LTAGBREAK9	–	±2.3	±0.50	±0.03	–	±0.06	±0.09	±0.13	±0.16	–	±0.90
LUMI	±2.6	±2.6	±2.6	±2.6	±2.6	–	±2.6	±2.6	±2.6	–	±2.5
MUON_ID	–	–	–	–	–	–	–	–	–	–	±0.00
MUON_RECO	–	–	–	–	–	–	–	–	–	–	±0.00
MUON_RES_ID	–	–	–	–	–	–	–	–	–	–	±0.00
MUON_RES_MS	–	–	–	–	–	–	–	–	–	–	±0.00
MUON_SCALE	–	–	–	–	–	–	–	–	–	–	±0.00
MUON_TRIG	±0.96	±0.94	±0.96	±0.96	±0.96	±0.07	±0.04	±0.05	±0.01	–	±0.88
QCD_Norm_DILEP	–	–	–	–	–	–	–	–	–	–	±0.00
QCDmm_CRfake_LJETS	–	–	–	–	–	–	–	–	–	–	±0.00
QCDmm_EL_LJETS	–	–	–	–	–	–	–	–	–	±22.0	±0.24
QCDmm_MCsub_LJETS	–	–	–	–	–	–	–	–	–	±7.1	±0.08
QCDmm_MU_LJETS	–	–	–	–	–	–	–	–	–	±24.6	±0.27
SingleTop-DS	–	–	–	–	–	–	–	±5.9	–	–	±0.23
singleTop_XS	–	–	–	–	–	–	–	±4.2	–	–	±0.17
TAGEXTRAP	±0.01	±0.04	±0.15	±0.19	±0.01	±0.78	±0.05	±0.25	±0.59	–	±0.14
ttbar-DataRw-BTagEff	–	±0.09	±0.10	–	–	–	–	–	–	–	±0.05
ttbar-DataRw-Fragmentation	–	±0.06	±0.12	–	–	–	–	–	–	–	±0.04
ttbar-DataRw-IFSR	–	±0.94	±1.3	–	–	–	–	–	–	–	±0.12
ttbar-DataRw-JER	–	±0.77	±1.1	–	–	–	–	–	–	–	±0.10
ttbar-DataRw-JetCloseby	–	±0.07	±0.42	–	–	–	–	–	–	–	±0.05
ttbar-DataRw-JetDet1	–	±0.07	±0.11	–	–	–	–	–	–	–	±0.04
ttbar-DataRw-JetEtaCalibration	–	±0.02	±0.06	–	–	–	–	–	–	–	±0.02
ttbar-DataRw-JetFlavB	–	±0.01	±0.02	–	–	–	–	–	–	–	±0.01
ttbar-DataRw-MCgen	–	±0.68	±0.53	–	–	–	–	–	–	–	±0.16
ttbar-DataRw-Notoppt-HF	–	–	±0.14	–	–	–	–	–	–	–	±0.02
ttbar-DataRw-Nottbarpt-HF	–	–	±0.42	–	–	–	–	–	–	–	±0.07
ttbar_bb	–	–	–	±13.6	–	–	–	–	–	–	±4.9
ttbar_bb-CSSKINRW	–	–	–	–	–	–	–	–	–	–	±0.00
ttbar_bb-DEFAULTSCALE	–	–	–	–	–	–	–	–	–	–	±0.00
ttbar_bb-FSRSCALE	–	–	–	–	–	–	–	–	–	–	±0.00
ttbar_bb-MPISCALE	–	–	–	–	–	–	–	–	–	–	±0.00
ttbar_bb-MSTWRW	–	–	–	–	–	–	–	–	–	–	±0.00
ttbar_bb-NNPDFRW	–	–	–	–	–	–	–	–	–	–	±0.00
ttbar_bb-Q_CMMPSRW	–	–	–	–	–	–	–	–	–	–	±0.00
ttbar_bb-R_MBBRW	–	–	–	–	–	–	–	–	–	–	±0.00
ttbar_cc	–	–	±27.3	–	–	–	–	–	–	–	±4.6
ttbar_cc-DataRw-Notoppt	–	–	–	–	–	–	–	–	–	–	±0.00
ttbar_cc-DataRw-Nottbarpt	–	–	–	–	–	–	–	–	–	–	±0.00
ttbar_cc_MG	–	–	–	–	–	–	–	–	–	–	±0.00

Table E.2: Normalisation uncertainties (expressed in %) on signal and each of the background processes for the systematic uncertainties considered, after the fit to data in $= 4\,\text{jets} \geq 4\,\text{b}-\text{tags}$ region of the single lepton channel. The total uncertainty can be different from the sum in quadrature of individual sources due to the correlations between them [58].

4 j, ≥ 4 b											
	$t\bar{t}H$ (125)	$t\bar{t}$ + light	$t\bar{t}+c\bar{c}$	$t\bar{t}+b\bar{b}$	$t\bar{t}+V$	W+jets	Z+jets	Single top	Diboson	Multijet	Total
ttbar_cc_MG-MATCH	–	–	–	–	–	–	–	–	–	–	±0.00
ttbar_cc_MG-MC	–	–	–	–	–	–	–	–	–	–	±0.00
ttbar_cc_MG-Q2	–	–	–	–	–	–	–	–	–	–	±0.00
ttbar_MG-MATCH-HF	–	–	±0.06	–	–	–	–	–	–	–	±0.01
ttbar_MG-MC-HF	–	–	–	–	–	–	–	–	–	–	±0.00
ttbar_MG-PP-HF	–	–	–	–	–	–	–	–	–	–	±0.00
ttbar_MG-Q2-HF	–	–	±0.07	–	–	–	–	–	–	–	±0.01
ttbar_PartonShower-bb	–	–	–	±11.2	–	–	–	–	–	–	±4.1
ttbar_PartonShower-cc	–	–	±13.0	–	–	–	–	–	–	–	±2.2
ttbar_PartonShower-light	–	±2.0	–	–	–	–	–	–	–	–	±0.72
ttbar_XS	–	±4.2	±4.2	±4.2	–	–	–	–	–	–	±3.7
ttbarV_XS	–	–	–	–	±29.9	–	–	–	–	–	±0.16
ttbb_FSR	–	–	–	±5.1	–	–	–	–	–	–	±1.9
ttbb_MPI	–	–	–	–	–	–	–	–	–	–	±0.00
ttbbNLO_CSSKIN	–	–	–	±1.0	–	–	–	–	–	–	±0.37
ttbbNLO_MSTW	–	–	–	±0.23	–	–	–	–	–	–	±0.08
ttbbNLO_NNPDF	–	–	–	±0.46	–	–	–	–	–	–	±0.17
ttbbNLO_QCMMPS	–	–	–	±1.4	–	–	–	–	–	–	±0.52
ttbbNLO_RMBB	–	–	–	±2.5	–	–	–	–	–	–	±0.89
ttbbNLO_scale	–	–	–	±3.4	–	–	–	–	–	–	±1.3
ttH-Generator	±0.51	–	–	–	–	–	–	–	–	–	±0.01
ttH-PartonShower	±1.8	–	–	–	–	–	–	–	–	–	±0.03
ttH-PDF	–	–	–	–	–	–	–	–	–	–	±0.00
ttH-Scale_Dyn	±5.6	–	–	–	–	–	–	–	–	–	±0.10
ttH-Scale_Var	±0.42	–	–	–	–	–	–	–	–	–	±0.01
ttH-TTHPDFrel	±3.2	–	–	–	–	–	–	–	–	–	±0.06
ttV-Shape	–	–	–	–	–	–	–	–	–	–	±0.00
ttV_scale	–	–	–	–	±5.3	–	–	–	–	–	±0.03
Wjets_pt	–	–	–	–	–	±32.0	–	–	–	–	±0.73
Wjets_XS_jet5	–	–	–	–	–	–	–	–	–	–	±0.00
Wjets_XS_jet6	–	–	–	–	–	–	–	–	–	–	±0.00
Wjets_XS_LJETS	–	–	–	–	–	±36.0	–	–	–	–	±0.82
Zjets_pt	–	–	–	–	–	–	–	–	–	–	±0.00
Zjets_XS	–	–	–	–	–	–	–	–	–	–	±0.00
Zjets_XS_3	–	–	–	–	–	–	–	–	–	–	±0.00
Zjets_XS_4	–	–	–	–	–	–	–	–	–	–	±0.00
Zjets_XS_jet5	–	–	–	–	–	–	–	–	–	–	±0.00
Zjets_XS_jet6	–	–	–	–	–	–	–	–	–	–	±0.00
Zjets_XS_LJETS	–	–	–	–	–	–	±47.4	–	–	–	±0.27
ZjetsPythia_pt	–	–	–	–	–	–	±19.1	–	–	–	±0.11
SigXsecOverSM	±73.0	–	–	–	–	–	–	–	–	–	±1.3
Total	±73.9	±10.6	±24.6	±15.8	±32.6	±32.3	±55.2	±12.8	±43.1	±48.7	±8.2

Table E.2: Normalisation uncertainties (expressed in %) on signal and each of the background processes for the systematic uncertainties considered, after the fit to data in $= 4\,\text{jets} \geq 4\,\text{b}-\text{tags}$ region of the single lepton channel. The total uncertainty can be different from the sum in quadrature of individual sources due to the correlations between them [58].

Bibliography

[1] J. Thomas-Wilsker, I. Connelly, R. Goncalo, M. zur Nedden, Measurement of the Performance of b-jets with Muons using top-pair events at $\sqrt{s}$ = 8 TeV , Tech. Rep. ATL-COM-PHYS-2013-742, CERN, Geneva (Jun 2013).
URL https://cds.cern.ch/record/1554289

[2] G. Aad, J.-F. Arguin, H. Arnold, et al., Performance of b-Jet Identification in the ATLAS Experiment, Tech. Rep. ATL-COM-PHYS-2012-824, CERN, Geneva (Jun 2012).
URL https://cds.cern.ch/record/1455862

[3] S. Guindon, E. Shabalina, J. Adelman, et al., Search for the Standard Model Higgs boson produced in association with top quarks in pp collisions at 8 TeV with the ATLAS detector at the LHC , Tech. Rep. ATL-COM-PHYS-2014-1471, CERN, Geneva (Nov 2014).
URL https://cds.cern.ch/record/1969527

[4] G. Aad, et al., Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector , Eur. Phys. J. C75 (7) (2015) 349. arXiv:1503.05066, doi:10.1140/epjc/s10052-015-3543-1.

[5] K. Olive, et al., Review of Particle Physics , Chin. Phys. C38 (2014) 090001. doi:10.1088/1674-1137/38/9/090001.

[6] S. Willenbrock, Symmetries of the Standard Model (2004) 3–38TASI 2004. arXiv:hep-ph/0410370.

[7] J. Goldstone, A. Salam, S. Weinberg, Broken Symmetries , Phys. Rev. 127 (1962) 965–970. doi:10.1103/PhysRev.127.965.

[8] F. Bezrukov, M. Shaposhnikov, Why should we care about the top quark Yukawa coupling? , Z. Eksp. Teor. Fiz. 147 (arXiv:1411.1923. CERN-PH-TH-2014-218) (2014) 3. 8 p.
URL https://cds.cern.ch/record/1968356

[9] S. Dittmaier, S. Dittmaier, C. Mariotti, et al., Handbook of LHC Higgs Cross Sections: 2. Differential Distributions arXiv:1201.3084, doi:10.5170/CERN-2012-002.

[10] R. Godbole, A. Grau, G. Pancheri, Y. Srivastava, Soft gluon radiation and energy dependence of total hadronic cross-sections , Phys. Rev. D72 (2005) 076001. arXiv:hep-ph/0408355, doi:10.1103/PhysRevD.72.076001.

[11] R. Boughezal, Theoretical Status of Higgs Production at Hadron Colliders in the Standard Model, arXiv:0908.3641.

[12] J. Baglio, A. Djouadi, Higgs production at the lHC , JHEP 1103 (2011) 055. arXiv:1012.0530, doi:10.1007/JHEP03(2011)055.

[13] C. Anastasiou, R. Boughezal, F. Petriello, Mixed QCD-electroweak corrections to Higgs boson production in gluon fusion , JHEP 0904 (2009) 003. arXiv:0811.3458, doi:10.1088/1126-6708/2009/04/003.

[14] D. L. Rainwater, D. Zeppenfeld, Observing $H \to W^*W^* \to e^{\pm}\mu\mp \not{p}_T$ in weak boson fusion with dual forward jet tagging at the CERN LHC , Phys. Rev. D60 (1999) 113004. arXiv:hep-ph/9906218, doi:10.1103/PhysRevD.61.099901, 10.1103/PhysRevD.60.113004, 10.1103/PhysRevD.60.113004 10.1103/PhysRevD.61.099901.

[15] G. Aad, et al., Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector , JHEP 1501 (2015) 069. arXiv:1409.6212, doi:10.1007/JHEP01(2015)069.

[16] G. Aad, et al., Combined measurements of the mass and signal strength of the Higgs-like boson with the ATLAS detector using up to 25 fb^{-1} of proton-proton collision data , Tech. Rep. ATLAS-CONF-2013-014, CERN, Geneva (Mar 2013).
URL http://cds.cern.ch/record/1523727

[17] V. Khachatryan, et al., Combination of Standard Model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV , Tech. Rep.

CMS-PAS-HIG-12-045, CERN, Geneva (2012).
URL https://cds.cern.ch/record/1494149

[18] G. Aad, et al., Measurement of Higgs boson production in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector, Phys. Rev. D90 (2014) 112015. arXiv:1408.7084.

[19] http://atlas.ch/inner_detector.html.

[20] G. Aad, et al., The ATLAS Experiment at the CERN Large Hadron Collider , JINST 3 (2008) S08003. doi:10.1088/1748-0221/3/08/S08003.

[21] N. Nikiforou, Performance of the ATLAS Liquid Argon Calorimeter after three years of LHC operation and plans for a future upgrade , in: Proceedings, 3rd International Conference on Advancements in Nuclear Instrumentation Measurement Methods and their Applications (ANIMMA 2013), 2013. arXiv:1306.6756, doi:10.1109/ANIMMA.2013.6728060.
URL http://inspirehep.net/record/1240499/files/arXiv:1306.6756.pdf

[22] M. Shochet, L. Tompkins, V. Cavaliere, et al., Fast TracKer (FTK) Technical Design Report, Tech. Rep. CERN-LHCC-2013-007. ATLAS-TDR-021, CERN, Geneva, aTLAS Fast Tracker Technical Design Report (Jun 2013).
URL https://cds.cern.ch/record/1552953

[23] F. Fayette, Strategies for precision measurements of the charge asymmetry of the W boson mass at the LHC within the ATLAS experiment , Ph.D. thesis, Paris U., VI-VII (2009). arXiv:0906.4260.
URL http://inspirehep.net/record/823897/files/arXiv:0906.4260.pdf

[24] P. Strizenec, Performance of the ATLAS Liquid Argon Calorimeter after three years of LHC operation and plans for a future upgrade, JINST 9 (2014) C09007. doi:10.1088/1748-0221/9/09/C09007.

[25] G. Aad, et al., Commissioning of the ATLAS Muon Spectrometer with Cosmic Rays , Eur. Phys. J. C70 (2010) 875–916. arXiv:1006.4384, doi:10.1140/epjc/s10052-010-1415-2.

[26] A. Salvucci, Measurement of muon momentum resolution of the ATLAS detector, EPJ Web Conf. 28 (2012) 12039. arXiv:1201.4704, doi:10.1051/epjconf/20122812039.

[27] J. Allison, K. Amako, J. Apostolakis, et al., Geant4 developments and applications , IEEE Trans.Nucl.Sci. 53 (2006) 270. doi:10.1109/TNS.2006.869826.

[28] W. Lampl, S. Laplace, D. Lelas, et al., Calorimeter Clustering Algorithms: Description and Performance, Tech. Rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva (Apr 2008).
URL http://cds.cern.ch/record/1099735

[29] G. Aad, et al., Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV , Eur. Phys. J. C73 (3) (2013) 2304. arXiv:1112.6426, doi:10.1140/epjc/s10052-013-2304-2.

[30] L. Asquith, B. Brelier, J. M. Butterworth, et al., Performance of Jet Algorithms in the ATLAS Detector, Tech. Rep. ATL-PHYS-INT-2010-129, CERN, Geneva (Dec 2010).
URL https://cds.cern.ch/record/1311867

[31] M. Cacciari, G. P. Salam, G. Soyez, The Anti-k(t) jet clustering algorithm, JHEP 04 (2008) 063. arXiv:0802.1189, doi:10.1088/1126-6708/2008/04/063.

[32] S. Catani, Y. L. Dokshitzer, M. H. Seymour, B. R. Webber, Longitudinally invariant K_t clustering algorithms for hadron hadron collisions, Nucl. Phys. B406 (1993) 187–224. doi:10.1016/0550-3213(93)90166-M.

[33] Y. L. Dokshitzer, G. D. Leder, S. Moretti, B. R. Webber, Better jet clustering algorithms, JHEP 08 (1997) 001. arXiv:hep-ph/9707323, doi:10.1088/1126-6708/1997/08/001.

[34] G. Aad, et al., Calibration of *b*-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment , Tech. Rep. ATLAS-CONF-2014-004, CERN, Geneva (Feb 2014).
URL https://cds.cern.ch/record/1664335

[35] G. Aad, et al., Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data , Eur. Phys. J. C74 (11) (2014) 3130. arXiv:1407.3935, doi:10.1140/epjc/s10052-014-3130-x.

[36] G. Aad, et al., Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC protonproton collision data , Eur. Phys. J. C74 (11) (2014) 3130. arXiv:1407.3935, doi:10.1140/epjc/s10052-014-3130-x.

[37] G. Aad, et al., Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data , Tech. Rep. ATLAS-CONF-2014-032, CERN, Geneva (Jun 2014).
URL http://cds.cern.ch/record/1706245

[38] V. Lacuesta, Track and Vertex Reconstruction in the ATLAS Experiment, Tech. Rep. ATL-INDET-PROC-2012-015, CERN, Geneva (Sep 2012).
URL http://cds.cern.ch/record/1482119

[39] G. Aad, et al., Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics, Tech. rep. (2009). arXiv:0901.0512.

[40] 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).
URL https://cds.cern.ch/record/1369219

[41] Calibrating the b-Tag Efficiency and Mistag Rate in 35 pb-1of Data with the ATLAS Detector , Tech. Rep. ATLAS-CONF-2011-089, CERN, Geneva (June 2011).
URL https://cds.cern.ch/record/1356198

[42] C. Oleari, The POWHEG-BOX , Nucl. Phys. Proc. Suppl. 205-206 (2010) 36–41. arXiv:1007.3893, doi:10.1016/j.nuclphysbps.2010.08.016.

[43] T. Sjostrand, S. Mrenna, P. Z. Skands, PYTHIA 6.4 Physics and Manual , JHEP 0605 (2006) 026. arXiv:hep-ph/0603175, doi:10.1088/1126-6708/2006/05/026.

[44] G. Aad, et al., Simultaneous measurement of the ratio B(t→Wb)/B(t→Wq) and the top quark pair production cross section with the D0 detector at $\sqrt{s} = 1.96$ TeV , Phys. Rev. Lett. **129300**.

[45] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A. D. Polosa, ALPGEN, a generator for hard multiparton processes in hadronic collisions , JHEP 0307 (2003) 001. arXiv:hep-ph/0206293, doi:10.1088/1126-6708/2003/07/001.

[46] M. Bahr, S. Gieseke, M. Gigg, et al., Herwig++ Physics and Manual , Eur. Phys. J. C58 (2008) 639–707. arXiv:0803.0883, doi:10.1140/epjc/s10052-008-0798-9.

[47] B. P. Kersevan, E. Richter-Was, The Monte Carlo event generator AcerMC versions 2.0 to 3.8 with interfaces to PYTHIA 6.4, HERWIG 6.5 and ARIADNE 4.1, Comput. Phys. Commun. 184 (2013) 919–985. arXiv:hep-ph/0405247, doi:10.1016/j.cpc.2012.10.032.

[48] S. Frixione, E. Laenen, P. Motylinski, B. R. Webber, C. D. White, Single-top hadroproduction in association with a Wboson, JHEP 07 (2008) 029. arXiv:0805.3067, doi:10.1088/1126-6708/2008/07/029.

[49] B. Acharya, J. Adelman, S. Adomeit, et al., Object selection and calibration, background estimations and MC samples for the Winter 2013 Top Quark analyses with 2012 data , Tech. Rep. ATL-COM-PHYS-2013-088, CERN, Geneva (Jan 2013).
URL https://cds.cern.ch/record/1509562

[50] G. Aad, et al., Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with $5\,fb^{-1}$ of Data from the ATLAS Detector , Tech. Rep. ATLAS-CONF-2012-043, CERN, Geneva (Mar 2012).
URL https://cds.cern.ch/record/1435197

[51] G. Aad, J.-F. Arguin, H. Arnold, et al., Performance of b-Jet Identification in the ATLAS Experiment, Tech. Rep. ATL-COM-PHYS-2012-824, CERN, Geneva (Jun 2012).
URL https://cds.cern.ch/record/1455862

[52] W. Lukas, Fast Simulation for ATLAS: Atlfast-II and ISF , Tech. Rep. ATL-SOFT-PROC-2012-065, CERN, Geneva (Jun 2012).

[53] G. Aad, et al., Jet energy resolution in proton-proton collisions at $\sqrt{s} = 7$ TeV recorded in 2010 with the ATLAS detector , Eur. Phys. J. C73 (3) (2013) 2306. arXiv:1210.6210, doi:10.1140/epjc/s10052-013-2306-0.

[54] G. Aad, et al., Measurement of the top quark-pair production cross section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV , Eur. Phys. J. C71 (2011) 1577. arXiv:1012.1792, doi:10.1140/epjc/s10052-011-1577-6.

[55] J. Butterworth, E. Dobson, U. Klein, et al., Single Boson and Diboson Production Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV , Tech. Rep. ATL-COM-PHYS-2010-695, CERN, Geneva (Aug 2010).
URL https://cds.cern.ch/record/1287902

[56] 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. B716 (2012) 1–29. arXiv:1207.7214.

[57] S. Chatrchyan, et al., Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC , Phys. Lett. B716 (2012) 30–61. arXiv:1207.7235.

[58] G. Aad, et al., Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector , Tech. Rep. arXiv:1503.05066. CERN-PH-EP-2015-047, CERN, Geneva (Mar 2015).
URL https://cds.cern.ch/record/2001975

[59] , Tech. rep.

[60] V. M. Abazov, et al., Measurement of the t anti-t production cross section in p anti-p collisions at s**(1/2) = 1.96-TeV using secondary vertex b tagging, Phys. Rev. D74 (2006) 112004. arXiv:hep-ex/0611002, doi:10.1103/PhysRevD.74.112004.

[61] M. Czakon, P. Fiedler, A. Mitov, Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\frac{4}{S})$, Phys. Rev. Lett. 110 (2013) 252004. arXiv:1303.6254, doi:10.1103/PhysRevLett.110.252004.

[62] T. Gleisberg, S. Hoeche, F. Krauss, et al., Event generation with SHERPA 1.1 , JHEP 02 (2009) 007. arXiv:0811.4622, doi:10.1088/1126-6708/2009/02/007.

[63] F. Cascioli, P. Maierhofer, S. Pozzorini, Scattering Amplitudes with Open Loops , Phys. Rev. Lett. 108 (2012) 111601. arXiv:1111.5206, doi:10.1103/PhysRevLett.108.111601.

[64] M. Czakon, A. Mitov, Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders, Comput. Phys. Commun. 185 (2014) 2930. arXiv:1112.5675, doi:10.1016/j.cpc.2014.06.021.

[65] G. Aad, et al., Measurements of normalized differential cross sections for $t\bar{t}$ production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector , Phys. Rev. D90 (7) (2014) 072004. arXiv:1407.0371, doi:10.1103/PhysRevD.90.072004.

[66] K. Melnikov, F. Petriello, Electroweak gauge boson production at hadron colliders through O(alpha(s)**2), Phys. Rev. D74 (2006) 114017. arXiv:hep-ph/0609070, doi:10.1103/PhysRevD.74.114017.

[67] J. M. Campbell, R. K. Ellis, An Update on vector boson pair production at hadron colliders , Phys. Rev. D60 (1999) 113006. arXiv:hep-ph/9905386, doi:10.1103/PhysRevD.60.113006.

[68] J. M. Butterworth, J. R. Forshaw, M. H. Seymour, Multiparton interactions in photoproduction at HERA, Z. Phys. C72 (1996) 637–646. arXiv:hep-ph/9601371, doi:10.1007/s002880050286.

[69] N. Kidonakis, Two-loop soft anomalous dimensions for single top quark associated production with a W- or H- , Tech. Rep. arXiv:1005.4451 (May 2010). URL http://cds.cern.ch/record/1267752

[70] N. Kidonakis, NNLL resummation for s-channel single top quark production , Phys. Rev. D81 (2010) 054028. arXiv:1001.5034, doi:10.1103/PhysRevD.81.054028.

[71] N. Kidonakis, Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production , Phys. Rev. D83 (2011) 091503. arXiv:1103.2792, doi:10.1103/PhysRevD.83.091503.

[72] J. M. Campbell, R. K. Ellis, $t\bar{t}W^{+-}$ production and decay at NLO , JHEP 07 (2012) 052. arXiv:1204.5678, doi:10.1007/JHEP07(2012)052.

[73] M. V. Garzelli, A. Kardos, C. G. Papadopoulos, Z. Trocsanyi, t $\bar{t}$ W^{+-} and t $\bar{t}$ Z Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects , JHEP 11 (2012) 056. arXiv:1208.2665, doi:10.1007/JHEP11(2012)056.

[74] G. Bevilacqua, M. Czakon, M. V. Garzelli, et al., HELAC-NLO , Comput. Phys. Commun. 184 (2013) 986–997. arXiv:1110.1499, doi:10.1016/j.cpc.2012.10.033.

[75] G. Aad, et al., Summary of ATLAS Pythia 8 tunes , Tech. Rep. ATL-PHYS-PUB-2012-003, CERN, Geneva (Aug 2012).
URL http://cds.cern.ch/record/1474107

[76] S. Dittmaier, et al., Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables arXiv:1101.0593, doi:10.5170/CERN-2011-002.

[77] A. Denner, S. Heinemeyer, I. Puljak, D. Rebuzzi, M. Spira, Standard Model Higgs-Boson Branching Ratios with Uncertainties , Eur. Phys. J. C71 (2011) 1753. arXiv:1107.5909, doi:10.1140/epjc/s10052-011-1753-8.

[78] G. Aad, et al., Selection of jets produced in proton-proton collisions with the ATLAS detector using 2011 data , Tech. Rep. ATLAS-CONF-2012-020, CERN, Geneva (Mar 2012).
URL https://cds.cern.ch/record/1430034

[79] M. Feindt, U. Kerzel, The NeuroBayes neural network package, Nucl. Instrum. Meth. A559 (2006) 190–194. doi:10.1016/j.nima.2005.11.166.

[80] B. P. Roe, H.-J. Yang, J. Zhu, et al., Boosted decision trees, an alternative to artificial neural networks, Nucl. Instrum. Meth. A543 (2-3) (2005) 577–584. arXiv:physics/0408124, doi:10.1016/j.nima.2004.12.018.

[81] A. Hoecker, P. Speckmayer, J. Stelzer, et al., TMVA - Toolkit for Multivariate Data Analysis, ArXiv Physics e-printsarXiv:physics/0703039.

[82] B. Andersson, G. Gustafson, G. Ingelman, T. Sjostrand, Parton Fragmentation and String Dynamics, Phys. Rept. 97 (1983) 31–145. doi:10.1016/0370-1573(83)90080-7.

[83] R. D. Field, S. Wolfram, A QCD Model for e+ e- Annihilation, Nucl. Phys. B213 (1983) 65. doi:10.1016/0550-3213(83)90175-X.

[84] B. R. Webber, A QCD Model for Jet Fragmentation Including Soft Gluon Interference, Nucl. Phys. B238 (1984) 492. doi:10.1016/0550-3213(84)90333-X.

[85] T. Junk, Confidence level computation for combining searches with small statistics, Nucl. Instrum. Meth. A434 (1999) 435–443. arXiv:hep-ex/9902006, doi:10.1016/S0168-9002(99)00498-2.

[86] G. Cowan, K. Cranmer, E. Gross, O. Vitells, Asymptotic formulae for likelihood-based tests of new physics, Eur. Phys. J. C71 (2011) 1554, [Erratum: Eur. Phys. J.C73,2501(2013)]. arXiv:1007.1727, doi:10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z.