THE UNIVERSITY *of* EDINBURGH

Search for the Higgs boson decaying to bottom quarks and W boson tagging techniques at the ATLAS experiment at the LHC

Timothy Michael Bristow

Doctor of Philosophy
The University of Edinburgh
October 2016

Lay Summary

This thesis describes the author's contributions to the study of the Higgs boson particle at the ATLAS experiment at the Large Hadron Collider (LHC) at CERN. The author also presents a new analysis technique to identify energetic W bosons which could signify the presence of new, as yet undiscovered, fundamental particles being produced in LHC collisions.

The Large Hadron Collider (LHC) is a particle accelerator at CERN in Geneva. It is the most powerful accelerator on earth, with a circumference of 27 km, located between 50 and 175 metres underground. Protons are accelerated in opposite directions through the LHC beam pipe and collide at four points around the accelerator. When the protons collide they release large amounts of energy which is converted into heavy particles following the famous equation $E = mc^2$, which, in turn, decay in a cascade of lighter, more-stable particles. Detectors surround each collision point and make precise measurements of the final state particles produced in the collisions. The detectors - in a sense - work like large, 3-dimensional, 100 megapixel digital cameras, taking up to 40 million images per second. The ATLAS detector is the largest detector at the LHC; it is 46 metres long, 25 metres in diameter and weighs 7000 tonnes.

The Standard Model (SM) of particle physics currently represents the best understanding of the Universe on a subatomic scale. The SM provides a description of all fundamental particles and their interactions to a very high degree of accuracy. One of the biggest achievements at the LHC thus far has been the discovery of the Higgs boson particle, which confirms the SM mechanism by which fundamental particles acquire mass. The search for the Higgs boson started over four decades ago; it is the last of the fundamental particles predicted by the SM to be discovered. Its discovery was first announced in 2012 by the ATLAS and CMS experiments at the LHC.

The Higgs boson is not stable; it decays into other lighter particles almost immediately after it is produced in a collision. These lighter particles are generally well understood and can be identified efficiently by the ATLAS detector. The patterns, or signatures, left by multiple particles within the detector are used

to infer the presence of a Higgs boson in the collision. There are many sets of particles that the Higgs boson can decay to. The SM predicts that 58% of the time the Higgs boson will decay into a bottom quark - anti-bottom quark pair, written as $H \to b\bar{b}$. However, there are many other processes that present a very similar signature in the ATLAS detector, making it very difficult to observe the signature of the $H \to b\bar{b}$ decay directly. However, it is predicted that a Higgs boson can be produced together with another well particle, the W boson. The signature of the decay of the W boson particle is well understood, and therefore easy to identify in the ATLAS detector. This thesis details the search the decay of the Higgs boson through $H \to b\bar{b}$ where it is produced together with a W boson at the ATLAS experiment in LHC collisions in 2012. Unfortunately, no observation of the Higgs boson decaying as $H \to b\bar{b}$ has been made yet.

The LHC was shut down for upgrade to be carried out in 2013 and 2014, and restarted proton collisions in 2015, at a higher frequency and higher energies than in previous years. Whilst this means it is possible to produce heavier particles than before, and potentially discover new particles that are not predicted by the Standard Model, this does not come without challenges. When the higher energy and higher momentum particles decay their decay products can end up very close together in the ATLAS detector making it difficult to determine which decay process occurred. In this thesis, two possible machine learning techniques are investigated to improve the efficiency to identify energetic W bosons from their decay products.

Abstract

The Standard Model of particle physics is currently the most complete theory of subatomic particles. The discovery of the Higgs boson with a mass of 125 GeV in 2012 further validated the Standard Model, providing evidence for the theory that vector bosons obtain non-zero masses through the Higgs mechanism. Studies are ongoing to determine the exact nature and properties of the Higgs boson. A Higgs boson of this mass is predicted to decay to a pair of $b\bar{b}$ quarks with a branching ratio of 58%, however this decay mode has not yet been observed. This thesis presents a search for the associated production of a Higgs boson with a leptonically decaying W boson, $WH \to \ell\nu b\bar{b}$, using 20.3 fb^{-1} of Run 1 data collected by ATLAS at the LHC from pp collisions at a centre-of-mass energy of $\sqrt{s} = 8$ TeV. The observed (expected) significance of a Higgs boson with a mass of 125 GeV for the $WH \to \ell\nu b\bar{b}$ process is found to be 2.7σ (1.3σ). The measured cross section in units of the expected Standard Model cross section has a best-fit value of $\mu = \sigma/\sigma_{SM} = 2.2^{+0.67}_{-0.64}(\text{stat.})^{+0.7}_{-0.59}(\text{syst.}) = 2.2^{+0.97}_{-0.87}$. The results are combined with the search for $ZH \to \nu\bar{\nu}b\bar{b}$ and $ZH \to \ell^+\ell^-b\bar{b}$ to provide a best-fit value of $\mu = \sigma/\sigma_{SM} = 1.1^{+0.61}_{-0.56}$.

The start of Run 2 of the LHC in 2015 saw the collision energy being raised to $\sqrt{s} = 13$ TeV, increasing the probability of particles being produced with a large momentum boost. At these high energies there is also a possibility to discover new particles and interactions. An extension of the Standard Model, the Heavy Vector Triplet (HVT) model, describes new heavy vector bosons W' and Z', which can decay to pairs of heavy bosons (W, Z or Higgs bosons). If the W' and Z' bosons are sufficiently heavy, the hadronic decays of the diboson final states produce boosted jets. In this thesis, methods for identifying hadronically decaying boosted bosons are developed, based on techniques that examine the internal *substructure* of the jet.

Multiple substructure variables are combined into a single discriminant using two machine learning techniques: boosted decision trees and deep neural networks. Simulated events of $W' \to WZ \to q\bar{q}q\bar{q}$ are used to develop these *boosted W boson taggers*. An improvement in the background rejection power, whilst keeping 50%

of the signal, over previous boosted W boson taggers of up to 13%–when using deep neural networks–and 36%–when using boosted decision trees–is obtained. The performance of the new boosted W boson taggers are evaluated in a search for a narrow WW resonances from the decay of a Z' with boson-tagged jets in 3.2 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV collected with the ATLAS detector.

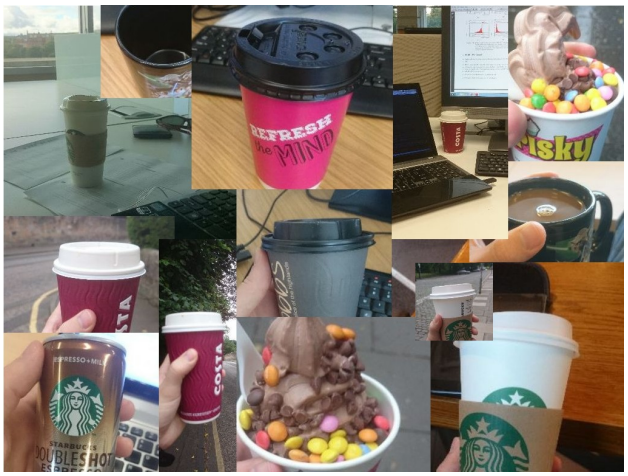
Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Parts of this work have been published in [1, 2].

(*Timothy Michael Bristow, October 2016*)

Acknowledgements



Oh a love ode to my cup of coffee... Your scent is as uplifting as a good night's sleep... Or something like that.

.

The four years I have spent doing my PhD have been some of the best (and worst... it's a PhD!!!) years of my life. I have gained so many friends and travelled all over the world. I even did some physics! There are far too many people to thank, but I will try to get most of them (in no particular order).

I have shared most of this journey with Dianne, Paul and Sebastian. Many great times were had along the way. Dianne had to put up with the majority of my PhD craziness, but more than that, she was my best friend throughout it all. Thank you, Dianne.

Thank you to Victoria Martin for her supervision, and putting up with my constant last minute shenanigans! Thank you for all the patience and support.

Throughout my PhD I have been lucky to have the support and advice from two of the most legit friends: Jacob and Simon (aka DBR Research). Thank you for all the late night chats, research ideas, and the casual existentialism. I too need to thank Daniel for these same things, although he is not part of DBR Research.

It has been a privilege to work with the Edinburgh PPE group, and in particular 5301 (the danger zone), Wahid, Christos, Ben Win, and those mentioned elsewhere in these acknowledgements.

I learnt a great deal whilst working with Gabriel Facini at CERN on the WH analysis. We had some stressful times, but we also had many laughs along the way. Thanks Gabriel!

A special thanks to Yanyan and Dave, my physics parents, for their help, support and spare bed.

If it were not for Trevor, I would not have been involved with ATLAS at all. So, thank you!

Thanks to Benji for encouraging me to apply for a PhD in Edinburgh. Excellent advice.

Thanks to Nathalie for keeping me sane over the last few months of thesis writing, keeping me from going down the rabbit hole, and the quote about the coffee above!

Ali J for all of the kind messages and laughs that we had. You guys are epic.

Sahal, David, and Katharine, you helped me get through my Masters degree sane enough to do a PhD.

I have thoroughly enjoyed playing cricket as part of the E=MCC, University Staff and CERN teams. Cricket in Scotland and Switzerland was an experience! Thanks guys!

Thanks, I think, to Chris, Rodney, Kieran and Jesse. I wish you guys laughed at my jokes. Thanks to Mansti, Shoe, and Nicky for all the talks (and for laughing at my jokes). The long distance phone calls from my grandparents checking in on me have been awesome, thank you!

This PhD would not have been possible without the financial support from SUPA, to whom I am extremely grateful. Thank you to the Oppenheimer Memorial Trust for the additional funding they provided that enabled me to get to Edinburgh in the first place.

I would not be here if it were not for the support I had from my parents. To you I owe the greatest thanks.

Contents

Lay Summary **i**

Abstract **iii**

Declaration **v**

Acknowledgements **vii**

Contents **ix**

1 Introduction **1**

2 Theoretical Motivation **7**
2.1 Introduction 8
2.2 Formalism of the Standard Model 8
2.2.1 Feynman Diagrams and Perturbation Theory 10
2.2.2 Particle Content of the Standard Model 11
2.2.3 Quantum Chromodynamics 13
2.2.4 Electroweak Theory 15
2.2.5 Spontaneous Symmetry Breaking and the Higgs Mechanism 19
2.2.6 Summary 23
2.3 Limitations of the Standard Model 23
2.4 Higgs Boson Phenomenology 24
2.4.1 Higgs Boson Production 25
2.4.2 Higgs Boson Decays 26
2.4.3 Higgs Discovery and Current Status 27
2.5 W Boson 28
2.6 Heavy Vector Triplet Model 30
2.7 Boosted W Boson Identification 32

3 The LHC and the ATLAS Detector **33**
3.1 Introduction 34
3.2 Large Hadron Collider 34
3.3 ATLAS Detector 37
3.3.1 Overview 37
3.3.2 Coordinate System 38
3.3.3 Inner Detector 40
3.3.4 Calorimeters 44

3.3.5 Muon Spectrometer 48
3.4 Trigger System 50
3.5 Computational Facilities 52

4 Particle Identification and Reconstruction 53
4.1 Introduction 54
4.2 Monte Carlo Simulation 54
4.3 Tracks 57
4.4 Primary Vertex 58
4.5 Electrons 59
4.6 Muons 59
4.7 Jets 61
4.7.1 Calorimeter Jets 62
4.7.2 Jet Calibration and Corrections 64
4.7.3 Track Jets 67
4.8 b jet Tagging 67
4.9 Missing Transverse Energy 69
4.10 Overlap Removal 69

5 $VH(\rightarrow b\bar{b})$ Analysis 71
5.1 Introduction 72
5.2 WH Analysis Strategy 74
5.3 Blinding Strategy 75
5.4 Event Selection Criteria 75
5.4.1 Trigger Selection 75
5.4.2 Object Selection 76
5.4.3 Geometrical and Kinematic Selections 78
5.4.4 Multivariate Analysis 80
5.4.5 Control Regions 81
5.5 Data and Simulated Samples 82
5.6 Backgrounds 84
5.6.1 W+jets 84
5.6.2 $t\bar{t}$ and single top 86
5.6.3 Diboson Processes 87
5.7 Multi-Jet Background 87
5.7.1 Multi-Jet Template 88
5.7.2 Truth Tagging 90
5.7.3 Multi-Jet Fitting Procedure 92
5.8 Systematic Uncertainties 99
5.8.1 Experimental Uncertainties 99
5.8.2 Multi-Jet Systematic Uncertainties 100
5.8.3 Monte Carlo Simulation Uncertainties 101
5.9 Statistical Analysis 106
5.9.1 Input Distributions 111
5.10 Results 111
5.10.1 Introduction 111
5.10.2 Event Yields 114

5.10.3 WH Dijet Mass Analysis Results 116
5.10.4 MVA Results 121
5.10.5 Diboson Crosscheck 123
5.11 Conclusions 126
5.12 Future Work 127

6 Machine Learning Techniques 129
6.1 Introduction 130
6.2 Data Preparation 132
6.2.1 Cleaning 132
6.2.2 Training and Testing Datasets 133
6.3 Deep Learning 133
6.3.1 Artificial Neural Networks 135
6.3.2 Autoencoders 139
6.3.3 Stacked Autoencoders 141
6.4 Decision Trees 143
6.4.1 Boosted Decision Trees 145
6.4.2 Random Forests 146
6.5 Tuning Classifier Hyperparameters 147
6.5.1 Classifier Metrics 147
6.5.2 Grid Search 148
6.5.3 Validation Curves 148
6.5.4 Feature Selection 149
6.5.5 Scikit-Learn 149
6.5.6 AGILEPack 150

7 Boosted W Boson Tagging 151
7.1 Introduction 152
7.2 Jet Grooming Algorithms 154
7.3 Jet Substructure Variables 157
7.3.1 Jet Shapes 158
7.3.2 Splitting Scales 160
7.3.3 Subjettiness 161
7.3.4 Centre-of-Mass Jet Shapes 163
7.4 Monte Carlo Samples 164
7.5 Run 1 W Tagger 166
7.5.1 Introduction 166
7.5.2 Event Selection 168
7.5.3 Tagger Optimisation 169
7.5.4 Results 170
7.6 Run 2 Tagger Crosscheck 172
7.6.1 Introduction 172
7.6.2 Event Selection 172
7.6.3 Crosscheck Results 173
7.7 BDT and DNN W Boson Taggers 180
7.7.1 Training and Testing Samples 181
7.7.2 BDT and DNN W Tagger Tuning 183

7.7.3 Results 190
7.7.4 Discussion 190
7.8 HVT Z' Prospects 194
7.8.1 Object and Event Selection 196
7.8.2 Data-to-Monte Carlo Comparisons 197
7.8.3 Discussion of Results 200
7.8.4 Discussion and Conclusion 201

8 Conclusions **205**

A GPU-based HLT **209**
A.0.1 Graphics Processing Unit 209
A.0.2 ATLAS Trigger System 211
A.0.3 GPU HLT Studies 212
A.0.4 Conclusion and Summary 216

B Truth Tagging **217**

C Mismodelling of H_T and $\Delta\phi(l, E_T^{\mathbf{miss}})$ **219**

D Input Distributions for $WH \to \ell\nu b\bar{b}$ **221**

E Event Yield Ratios **233**

F Event Yields in $WH \to \ell\nu b\bar{b}$ **247**

G Substructure Variable Definitions **257**
G.1 Thrust 257
G.2 Fox-Wolfram Moments 258
G.3 Dipolarity 258

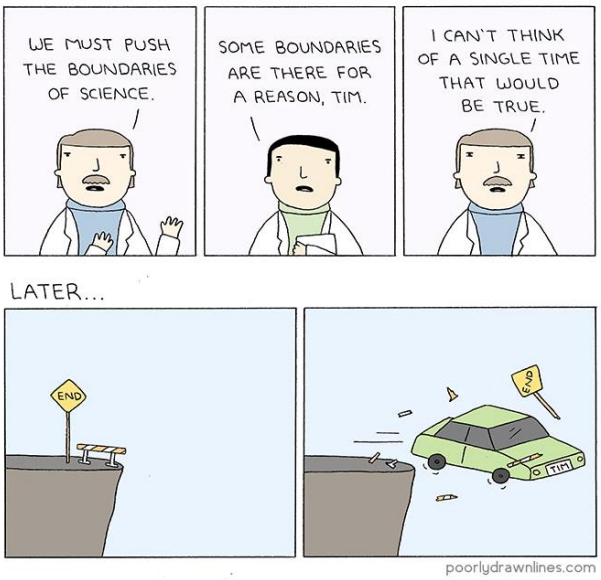
H Boosted Boson Tagging Decision Functions and Statistics **261**
H.1 Training Folds Statistics 261

I BDT Results Check **273**

Bibliography **277**

Chapter 1

Introduction



The story so far: Tim decided to do a PhD. This has made a lot of people very angry and been widely regarded as a bad move. Image used with permission from [3], caption adapted from [4].

The *how* and *why* of the Universe has been a matter of philosophical debate for many centuries. Many people have devoted their lives to discovering the underlying physical laws that govern the Universe. This has proceeded through a combination of theoretical models and experimental observations. One key breakthrough has been atomic theory—the idea that all matter can be broken up into constituent atoms[1]. As it is now widely understood, atoms can be broken up into fundamental subatomic particles. Understanding the properties and interactions of these fundamental particles is an important step towards an explanation of how the Universe works.

The Standard Model of particle physics is currently the most complete theory describing the interactions and behaviour of the fundamental elementary particles [6–8]. All the known fundamental particles and three of the four known fundamental forces are included in the Standard Model, however, gravitational interactions between fundamental particles are currently not included. Over many decades, the predictions of the Standard Model have shown exceptional agreement with experimental observations. In studying high energy proton-proton collisions at the Large Hadron Collider [9] (LHC), with experiments such as the ATLAS experiment [10], the Standard Model can be tested to new extremes. Further validation of the Standard Model with experimental evidence is of crucial importance in order to ensure that the theory is correct. There is also the possibility for New Physics to be seen which is not included in the Standard Model. Simulations of proton-proton collisions are performed according to the Standard Model expectation, which are compared with data collected by the detector. Many measurements must be taken to ensure that any observations are not statistical fluctuations.

One of the challenges in the formulation of the Standard Model was the explanation of non-zero masses of the fundamental particles, which are forbidden by certain underlying symmetries of the theory. By incorporating interactions with the *Higgs field*, these symmetries are spontaneously broken and the fundamental particles gain mass [11–14]. Excitations of the Higgs field correspond to *Higgs bosons*. The discovery of the Higgs boson was one of the primary goals

[1] As Richard Feynman once said [5], "If, in some cataclysm, all of scientific knowledge were to be destroyed, and only one sentence passed on to the next generation of creatures, what statement would contain the most information in the fewest words? I believe it is the atomic hypothesis (or the atomic fact, or whatever you wish to call it) that all things are made of atoms — little particles that move around in perpetual motion, attracting each other when they are a little distance apart, but repelling upon being squeezed into one another. In that one sentence, you will see, there is an enormous amount of information about the world, if just a little imagination and thinking are applied."

of the LHC physics program when it began in 2010. In 2012, the ATLAS and CMS [15] experiments announced the discovery of a new particle in proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 7$ and 8 TeV that was consistent with a Standard Model Higgs boson, with a mass of approximately 125 GeV [16, 17]. Following the discovery of the Higgs boson, studies are required to establish that it has the properties predicted by the Standard Model. The current measurements of its spin, couplings, and mass all indicate that it is consistent with these predictions [18]. However, it has not yet been observed to couple to bottom quarks. Since a Higgs boson with the observed mass is expected to decay to a bottom quark-antiquark pair ($H \to b\bar{b}$) almost 58% of the time, this is a vital measurement to be performed.

At the LHC, there are large backgrounds to the $H \to b\bar{b}$ process. These backgrounds can be reduced by considering the case where the Higgs boson is produced in association with a W or Z boson that decays leptonically. This thesis presents a search for $WH \to \ell\nu b\bar{b}$ using 20 fb^{-1} of data at a centre-of-mass energy of $\sqrt{s} = 8$ TeV collected by the ATLAS detector during the first experimental run of the LHC between 2010 and 2013, also known as Run 1 of the LHC. A combination of this process with $ZH \to \nu\bar{\nu}b\bar{b}$ and $ZH \to \ell^+\ell^-b\bar{b}$ is also presented.

In attempting to observe processes such as $H \to b\bar{b}$, the identification of hadronic decays is critical. Hadronic decays are observed in the detector as narrow cones of energy deposits, which are clustered together to form *jets*, roughly corresponding to the decay products of a single particle. In the $WH \to \ell\nu b\bar{b}$ search, it was observed that there was greater sensitivity in the regions where the W boson had a higher transverse momentum (p_{T}). During Run 2 of the LHC, which began in 2015, the higher collision energies of up to $\sqrt{s} = 14$ TeV increases the likelihood of such high p_{T} particles being produced. New challenges are encountered in identifying these hadronic final states in this *boosted* regime. If a heavy particle decays into multiple jets, these jets can become collimated to the point that they are clustered together and classified as a single jet. Examining the *substructure* of the jet can help to identify so-called *subjets*, where each of the subjets corresponds to one hadronic decay.

The need for improved jet substructure techniques is motivated by the potential for new heavy particles to be produced during the second run, Run 2, of the LHC. For example, Heavy Vector Triplet (HVT) models [19, 20] extend the Standard Model and predict the existence of heavy vector bosons W' and Z', which are

degenerate in mass. If such particles exist, it is possible that they would decay into a diboson final state of W, Z or Higgs bosons. In this thesis, the author presents a method of improving identification of boosted bosons by implementing machine learning techniques in the form of boosted decision trees and deep neural networks. These studies are performed in the context of W bosons, considering simulations of the HVT process $V' \to VV \to JJ$, where V' is a W' or Z', V is a W, Z or Higgs boson, and J is a boosted jet corresponding to each of the W/Z bosons. The performance of these W boson *taggers* is demonstrated using Run 2 data at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 3.2 fb^{-1}. Whilst these studies are performed in the context of boosted W bosons, these techniques could be applied to other particles, such as the Higgs boson. This would have direct relevance for $H \to b\bar{b}$ searches.

The work contained in this thesis is a combination of the author's personal work and his contributions within the ATLAS Collaboration. In particular, the $WH \to \ell\nu b\bar{b}$ search presented here was done in collaboration with other members of the ATLAS 'Higgs Sub-Group 5' (HSG5) research group. The work on the boosted boson tagging was performed in part with the Boosted Boson Tagging research group. In the following outline of the thesis, the author's personal contributions are highlighted. In general, plots presented which contain an 'ATLAS' label are public results, and plots without such a label are plots produced by the author himself. The work presented in this thesis has been included in two publications from the ATLAS experiment: References [1] and [2], as detailed in Chapters 5 and 7, respectively.

Natural units are used throughout this thesis, such that $c = \hbar = 1$, and charges are given in units of the magnitude of the electric charge. The thesis is structured as follows:

Chapter 2 presents the Standard Model of particle physics and the theoretical motivation for the Higgs boson. The predicted properties and phenomenology of the Higgs boson are presented, followed by a review of the discovery of the Higgs boson and its current status. Properties of the W boson are discussed and the Heavy Vector Triplet model is introduced.

Chapter 3 introduces the LHC and the ATLAS detector. The separate components of the ATLAS detector and their role in particle identification are described in detail. A version of the High Level Trigger (HLT) that runs on graphical processing units (GPUs), to which the author contributed, is included

in Appendix A.

Chapter 4 describes Monte Carlo event simulation, and the reconstruction and identification of physics objects that are used for the analyses in Chapters 5 and 7. Particular attention is paid to the reconstruction of jets.

Chapter 5 details the search for $WH \to \ell\nu b\bar{b}$ production at the ATLAS experiment using data from proton proton collisions at a centre-of-mass energy $\sqrt{s} = 8$ TeV during Run 1. The results for the $ZH \to \ell^+\ell^- b\bar{b}$ and $ZH \to \nu\nu b\bar{b}$ analyses are also presented, and these are combined with the WH result. This work was performed in collaboration with the HSG5 research group. In particular, the author worked on multi-jet background estimation and *truth tagging* of b-jets for the WH search, and on the development and maintenance of the analysis framework used to produce a common set of data containers (ntuples) used by multiple researchers within the HSG5 group.

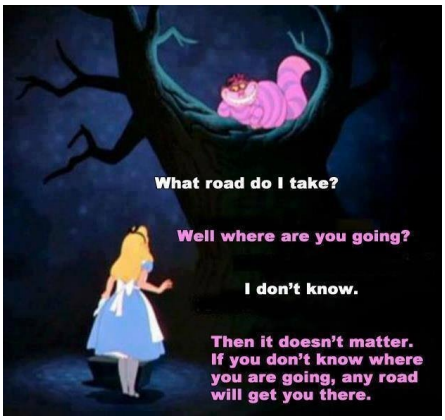
Chapter 6 covers in detail machine learning and data analysis techniques implemented in Chapter 7. Two *classifiers* are introduced: deep neural networks based on *stacked autoencoders* [21–23], and *boosted decision trees* [24, 25]. Data preparation methods, and methods to tune the *hyperparameters* of the classifiers to prevent overfitting are also discussed.

Chapter 7 describes techniques used by ATLAS for identifying boosted W bosons. The author worked within the Boosted Boson Tagging research group to find an optimal combination of *jet grooming algorithm* and *jet substructure variable* to use for boosted W boson identification in Run 2. Studies are performed using $\sqrt{s} = 8$ TeV simulations, with cross checks performed by the author on $\sqrt{s} = 13$ TeV simulations. The author was the sole developer of two machine learning classifiers presented here that are trained to identify boosted W bosons at $\sqrt{s} = 13$ TeV. Training, testing and evaluation of these classifiers is described in detail, using simulations of HVT W' and Z' diboson decays. Data-to-simulation comparisons of the W boson tagger in a QCD background enriched region are performed using selection criteria that are used in the search for Z' bosons decaying to dibosons in ATLAS Run 2 data.

A summary and conclusion of all the results presented in this thesis is given in Chapter 8.

Chapter 2

Theoretical Motivation



Cheshire Cat teaches Alice about quantum field theory [26, 27].

2.1 Introduction

The Standard Model (SM) of particle physics [6–8] is a gauge quantum field theory providing a unified description of all fundamental particles and their interactions under the known forces, with the exception of gravity. Electromagnetism and the weak interaction are described collectively by electroweak theory, and strong interactions are described by quantum chromodynamics (QCD). Electroweak theory and QCD are combined to form the Standard Model, which models all particle interactions. Over many decades, the SM has been tested experimentally and has shown exceptional agreement with data. With the observation of the Higgs boson at the LHC in 2012, all particles predicted by the Standard Model have been discovered[1].

Summaries of the Standard Model and electroweak symmetry breaking are given here to motivate the study of the Higgs boson[2]. The expected properties of the Higgs boson are discussed and the current status of the experimental results for the Higgs boson is highlighted in Section 2.4. Some properties of the W boson are discussed in Section 2.5. These descriptions provide important background to the search for the decay of the Higgs boson to $b\bar{b}$ with the associated production of a W boson in Chapter 5. The Heavy Vector Triplet model describing the W' and Z' bosons is presented in Section 2.6, and Section 2.7 considers a search for the Z' using boosted W boson identification.

2.2 Formalism of the Standard Model

A key aspect in the formulation of the SM is the interpretation of fundamental particles and interactions, or forces, as excitations of quantum fields. The forces are mediated by the exchange of force-carrying gauge bosons, as opposed to the classical view, where the field mediates the forces. The dynamics and interactions of the fundamental particles and fields of the SM are described in terms of a Lagrangian[3], $\mathcal{L}$, describing the free and interaction terms separately. Symmetries obeyed by a Lagrangian imply conservation laws with conserved currents and quantum numbers, according to Noether's first theorem [28, 29].

[1] Studies are ongoing to determine the exact nature of the Higgs boson and ensure it has the properties as predicted by the Standard Model.

[2] Unless explicitly specified otherwise, the text refers to a Standard Model Higgs boson.

[3] The Standard Model Lagrangian is a sum of Lagrangians for each type of interaction.

These symmetries are global or local transformations applied to the system, or Lagrangian, that leave it invariant. *Symmetry groups* contain groups of such transformations constructed from a set of group generators. In the SM, these symmetries provide the fundamental interactions. In a gauge symmetry, which describes local interactions between particles, local symmetry is a *requirement*. In this case, the Lagrangian is said to be *locally gauge invariant*.

A consequence of requiring local gauge invariance is that for every generator of the symmetry group there is a gauge field included in the Lagrangian that interacts with the matter fields[4]. Depending on the symmetry group under consideration, these represent the photon field, gluon fields or the weak isospin and hypercharge fields, each of which have corresponding gauge bosons: the photon, the gluons, and the $W^{\pm}$ and Z bosons, respectively. The photon and gluons are both massless, but adding in mass terms to the Lagrangian for the $W^{\pm}$ and Z bosons does not preserve gauge invariance; the masses can be introduced through *electroweak symmetry breaking*, which is achieved via the *Higgs mechanism* in the SM [11–14].

The Standard Model is a locally gauge invariant quantum field theory (QFT) under the $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetry, consistent with both special relativity and quantum mechanics. $SU(3)_C$ is the colour symmetry group of QCD, and $SU(2)_L \times U(1)_Y$ is the weak-isospin-hypercharge gauge group of the electroweak interactions, such that electroweak interactions are invariant under weak isospin $SU(2)_L$ and weak hypercharge $U(1)_Y$ transformations[5]. The electroweak symmetry group is reduced to the $U(1)_{EM}$ symmetry group of quantum electrodynamics (QED) under spontaneous symmetry breaking, induced by the Higgs mechanism. In this formalism, the QCD and electroweak interactions are described by two symmetry groups, each with independent couplings, rather than a single unified symmetry group. The corresponding forces are described in terms of fields, with excitations of the fields corresponding to virtual particles.

[4]In fact, a Lagrangian containing only the free fermion fields, requires the introduction of the gauge fields and gauge bosons when enforcing local gauge invariance. It also gives an interaction term between the fermion fields and gauge fields.

[5]Weak-isospin is defined according to $\boldsymbol{T}^i = \frac{1}{2}\boldsymbol{\sigma}^i$, where $\boldsymbol{\sigma}^i$ are the three Pauli matrices. Hypercharge is defined as $Y = 2(Q - T^3)$, where T^3 is the magnitude of the third weak isospin component and Q is the electric charge.

2.2.1 Feynman Diagrams and Perturbation Theory

The interactions between particles and fields in QFT can be represented by *Feynman diagrams*. Each diagram represents a perturbative contribution to the amplitude of the transition between two states. The diagrams themselves consist of a number of lines representing particles, which are joined at vertices representing interactions. Internal lines in the diagrams correspond to virtual particles that are exchanged to mediate interactions. A set of Feynman rules are associated with every particle and vertex within any Feynman diagram[6], which can be used to calculate the corresponding transition amplitude as a perturbation series expansion. An example of a diagram for a *leading order* contribution is shown in Figure 2.1(a). Higher order contributions, or perturbations, can be included by considering diagrams that include additional interaction vertices from virtual particles, such as in Figure 2.1(c)-(d). These enter the transition amplitude as an integral over the four-momenta of all particles within the closed loop. There are also higher order corrections from *real emissions*, where for example, an electron radiates a photon, as shown in Figure 2.1(b). In general, an infinite number of such additional diagrams exist, and in the path integral formulation, all possible diagrams must be summed in the calculation of the overall transition amplitude.

One of the difficulties in QFT is that the higher order calculations, represented by the loop diagrams, can lead to divergences. All possible energy and momentum combinations for the virtual particles must be considered, and these integrals can become infinite. The procedure of *renormalisation* works by redefining the charge (effectively the coupling constant) and mass terms in the Lagrangian to account for loop effects. This results in energy-dependent renormalised values of the charge and mass, which contain compensating divergences. The new renormalised values should be used in place of the 'bare' charge and mass. This allows perturbation theory to be used up until an unphysical *renormalisation scale*, μ_R, which sets the cut-off above which loop contributions are included in the renormalised quantities. Calculations of masses and coupling constants will have a μ_R dependence, due to the subsequent perturbative corrections being truncated. Renormalisation was first introduced in the late 1940s to remove divergences in QED based on work by Feynman, Schwinger, Tomonaga and Dyson [33]. It was then shown by 't Hooft in 1971 that both the electroweak theory and QCD are

[6]The factors for the different vertices and particles that enter into the transition amplitude can be found in any of the following References [30–32].

also renormalisable [34, 35]. QCD exhibits asymptotic freedom of quarks and gluons within hadrons, where the coupling constant decreases at high energies. At low energies the coupling constant is large and perturbation theory breaks down [32].

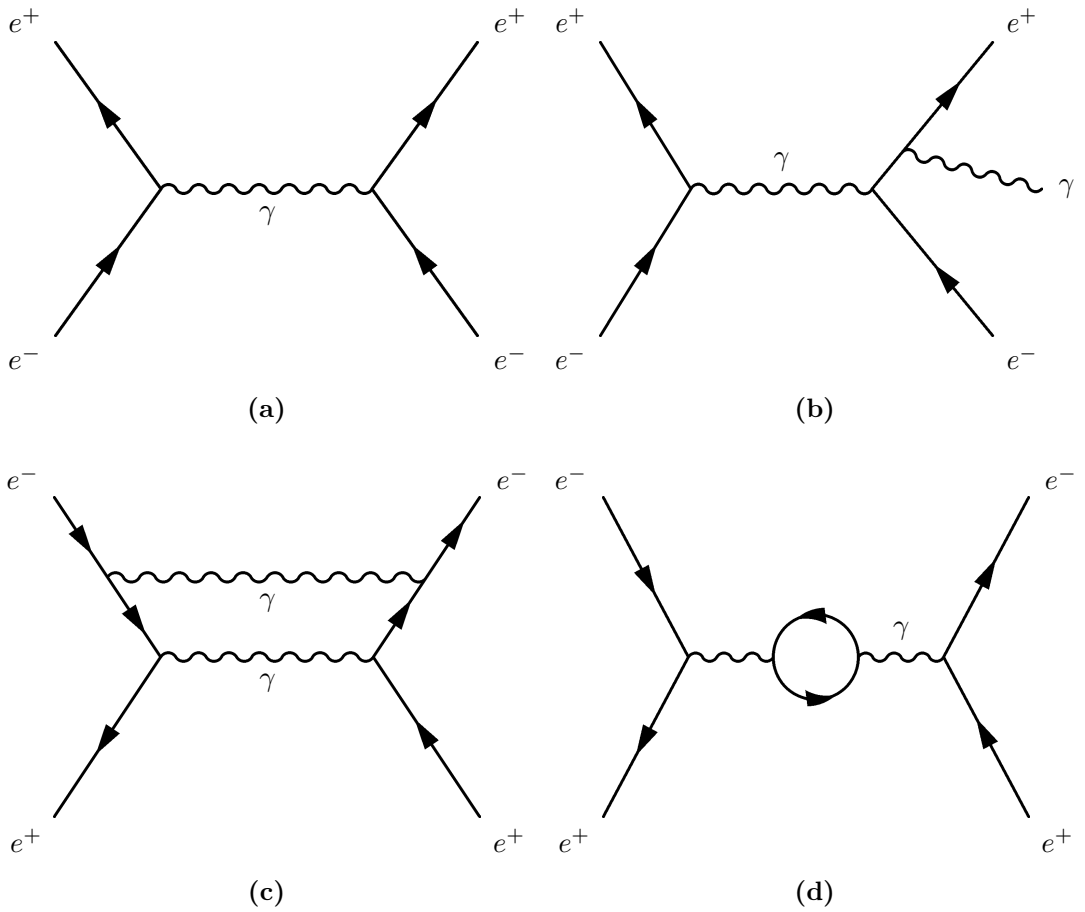


Figure 2.1 *The leading order Feynman diagram for the process $e^+e^- \to e^+e^-$ is shown in (a), with a real emission correction in (b), and two possible virtual corrections in (c) and (d).*

In the following sections, the particle content of the Standard Model will be introduced, followed by a description of quantum electrodynamics, quantum chromodynamics, weak interactions, electroweak unification, and the Higgs mechanism.

2.2.2 Particle Content of the Standard Model

Fundamental particles within the SM are divided into spin-$\frac{1}{2}$ *fermions* and integer-spin *bosons*. The fermions provide the matter content of the model,

and the bosons mediate the forces. Each particle possesses a number of charges corresponding to the fundamental forces they experience: the electromagnetic force acts on all particles with non-zero electric charge, the weak force acts on all particles with non-zero weak isospin, and QCD acts on all particles with colour charges.

The fermions, listed in Table 2.1, are divided into six *flavours* of leptons and quarks, which are grouped into three generations of increasing mass. Within each lepton generation, there is a charged lepton and a corresponding neutrino, and within each quark generation there is an up-type and a down-type quark. In addition, each fermion has a corresponding antiparticle that has opposite signs for all additive quantum numbers. Quarks and charged leptons both interact through the weak and the electromagnetic forces, whilst neutrinos only experience the weak force. The fermions are described in terms of fields with left- and right-handed chiral components (see Equations 2.5 and 2.6). Left-handed leptons from each generation are grouped into doublets with a total weak isospin of $T = \frac{1}{2}$ and a third component of weak isospin of $T^3 = \pm\frac{1}{2}$, consisting of a charged lepton and associated neutrino, and right-handed singlets with zero total weak isospin and $T^3 = 0$. Left-handed components of the up- and down-type quarks from each generation are grouped into doublets with $T = \frac{1}{2}$ and $T^3 = \pm\frac{1}{2}$, and right-handed singlets with zero total weak isospin and $T^3 = 0$. Antiparticles have opposite chirality of their corresponding particles and a change in sign of T^3. The weak force acts on those particles and antiparticles with a non-zero T^3: left-handed particles and right-handed antiparticles only, excluding right-handed particles and left-handed antiparticles from weak interactions. Quarks possess an additional colour charge of red, blue, green (and corresponding anti-colours) and experience strong interactions. Quarks combine to form colourless hadrons, which can be classified as either mesons or baryons; mesons contain a quark and antiquark, and baryons contain three quarks. Quarks have not been observed to exist as free particles.

The electromagnetic and weak interactions of the fermions are mediated by the massless photon and the massive $W^{\pm}$ and Z bosons, respectively, and are unified within electroweak theory (see Section 2.2.4). The scalar Higgs boson is predicted by the Higgs mechanism, which gives masses to the fermions, and the $W^{\pm}$ and Z bosons, and mediates interactions of particles with the Higgs field. The $W^{\pm}$ and Z bosons interact with the fields of the leptons and quarks, and additionally self-interact. The strong interactions of quarks are described

by quantum chromodynamics (QCD), mediated via eight massless gluons, which also self-interact. Of the gauge bosons, only the $W^{\pm}$ has non-zero electric charge, and all but the Higgs boson (spin-0) have spin-1.

Fermions	Generation	Name	Charge
Leptons	1st	Electron (e^-)	-1
		Electron neutrino (ν_e)	0
	2nd	Muon (μ^-)	-1
		Muon neutrino (ν_μ)	0
	3rd	Tau (τ^-)	-1
		Tau neutrino (ν_τ)	0
Quarks	1st	Up (u)	$\frac{2}{3}$
		Down (d)	$-\frac{1}{3}$
	2nd	Charm (c)	$\frac{2}{3}$
		Strange (s)	$-\frac{1}{3}$
	3rd	Top (t)	$\frac{2}{3}$
		Bottom (b)	$-\frac{1}{3}$

Gauge Bosons
Photon (γ)
Gluon (g)
Z Boson (Z^0)
W Bosons ($W^{\pm}$)
Higgs Boson (H)

Table 2.1 *Fundamental particles and gauge bosons in the Standard Model. Charge here refers to the electric charge.*

2.2.3 Quantum Chromodynamics

QCD is a non-Abelian gauge theory based on a $SU(3)$ symmetry group that describes strong interactions, based on the assumption that quarks obey an exact $SU(3)$ colour symmetry. It is invariant under transformations of the form

$$q(x) \rightarrow q'(x) = e^{ig_s\alpha_a(x)\lambda^a}q(x), \tag{2.1}$$

where $x = (\vec{r}, t)$, $q(x)$ represents a quark field, g_s is the coupling strength term, λ_a (where $a = 1, \ldots, 8$) represents the eight Gell-Mann matrices corresponding to the non-Abelian generators of the $SU(3)$ colour symmetry group, and $\alpha_a(x)$ are the group parameters that specify the eight generators of the group that represent the gauge transformations. The λ_a generators are given by the relation $[\lambda^a, \lambda^b] = 2\sum_c f^{abc}\lambda^c$, where f^{abc} are the structure constants.

Enforcing the local gauge invariance of the QCD Lagrangian requires the introduction of a massless vector field for each of the eight generators, which

corresponds to eight massless gluons. These gluons possess colour charge and can self-interact. The locally gauge invariant QCD Lagrangian is given by

$$\mathcal{L}_{QCD} = \bar{q}_j(i\gamma^\mu\partial_\mu - m_{q_j})q_j - g_s\bar{q}_j\gamma^\mu\lambda_a G^a_\mu q_j - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a, \tag{2.2}$$

where μ and ν are Lorentz indices (with values of 0, 1, 2, 3)[7], q_j represents the quark field (with the flavour of the quark $j = 1, \ldots, 6$), m_{q_j} is the mass of quark q_j, and γ^μ are the Dirac matrices. The strength of the coupling constant is a function of the four-momentum transfer in the interaction, $q^2 \equiv -Q^2$, and is typically written in terms of the strong coupling constant, $\alpha_S(Q^2) = g_s^2(Q^2)/4\pi$. $G^a_{\mu\nu}$ represents the massless gluon fields, defined as

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g f_{abc} G^b_\nu G^c_\nu, \tag{2.3}$$

where f_{abc} are the $SU(3)$ structure constants, and the third term represents self-interactions between gluons.

There are two properties of QCD which require further mention: *asymptotic freedom* and *confinement*. As the energy scale of the interaction increases (i.e. an increase of Q^2), the running coupling constant $\alpha_S(Q^2)$ decreases. At high energies, or equivalently at short distances, $\alpha_S(Q^2) \ll 1$, and quarks and gluons behave as free particles. Importantly, perturbative calculations can be used in this regime where the quarks interact weakly. Confinement is the hypothesis that quarks and gluons do not exist as free particles outside of bound colourless states, which is in agreement with experimental evidence.

There is currently no analytical proof of confinement, however, the observed behaviour can be explained by considering the situation where a quark and antiquark are separated. The force between them is mediated via the exchange of virtual gluons, which are themselves colour charged. The attractive interactions between the virtual gluons have the effect of forcing the colour field between the quarks into a 'tube'. The energy density within this tube is constant at large distances, thus the total amount of energy in the colour field increases linearly with separation. This translates to an infinite amount of energy needed to pull a quark and antiquark apart. It will become more favourable, at some point, to pull a quark-antiquark pair from the vacuum, creating two new hadrons. This

[7]Equations presented throughout the thesis use the Einstein summation convention. Indices labelled with Greek letters imply a sum over values of 0, 1, 2, 3. Roman letters take the values 1, 2, 3, unless otherwise specified.

process can further continue with these new hadrons, up until the quark and antiquark have low enough energy to form a colourless hadron. This process is known as *hadronisation*. In high energy collisions, such as at the LHC, if a hadron is given sufficient energy the quarks produced in the hard scatter will move apart, radiating gluons and $q\bar{q}$ pairs, producing collimated showers of hadrons, which are observable in the detector as *jets*. The jets are identified by clustering energy deposits and particle tracks from showers in the detector, intending to capture the decay products from the quarks produced in the hard scatter. This is discussed in more detail in Section 4.7.

2.2.4 Electroweak Theory

Electroweak theory unifies electromagnetism and the weak force as described by a $SU(2)_L \times U(1)_Y$ gauge symmetry: a unification of the $SU(2)$ symmetry of the weak interactions with the $U(1)$ symmetry of QED. One of the challenges encountered is creating a theory that can provide masses to the $W^{\pm}$ and Z gauge bosons, whilst keeping the Lagrangian locally gauge invariant. There are a number of other phenomena that have to be included in electroweak theory: fermion flavour change, parity and CP violation. Some of the history of electroweak unification is given below, leading to a solution that gives mass to the vector gauge bosons – the Higgs mechanism.

Quantum Electrodynamics

Electromagnetism was the first of the forces to be described by a quantum field theory: the theory of quantum electrodynamics (QED). QED is a gauge theory that is invariant under transformations of a $U(1)$ symmetry group, where the electric charge is the group generator. The Lagrangian for QED describes the dynamics of the fermions, photons and their interactions. The Lagrangian for a free particle is given by $\mathcal{L} = \bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi$, where the fermion fields are described as complex Dirac spinors $\psi(x)$, separated into left- and right-handed chiral components. Under local $U(1)$ transformations of the form $\psi(x) \rightarrow e^{-i\chi(x)}\psi(x)$, where $\chi(x)$ is any rotation in $U(1)$ space, this is not invariant, due to the derivative acting on $\chi(x)$, which is non-zero. The unwanted terms from the gauge transformations of $\psi(x)$ can be absorbed into local transformations of a new field, A^{μ}, the electromagnetic field. The Lagrangian is made locally gauge invariant by replacing the derivative, ∂_{μ}, with a covariant derivative:

$D_\mu = \partial_\mu + ieA_\mu$, where the photon field transforms as $A_\mu \to A'_\mu = A_\mu - \frac{1}{e}\partial_\mu\chi$, absorbing the additional $\chi(x)$ terms from the gauge transformation of $\psi(x)$. In enforcing local gauge invariance and introducing an additional gauge field, the form of the interaction between the fermions and electromagnetic field is found, giving the interacting Lagrangian. The photon field couples to the fermion fields ψ with a coupling constant of $\alpha(Q^2) = \frac{e^2(Q^2)}{4\pi}$, where $e(Q^2)$ is the electric charge of ψ. $\alpha(Q^2)$ *increases* slowly with an increase in Q^2, leading to a running coupling constant. The Lagrangian of QED is then given by

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{2.4}$$

where γ^μ represents the Dirac matrices, $\bar{\psi} = \psi\gamma^0$ is the adjoint spinor, m is the mass of the electron, and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the EM field tensor. Notably, there is no mass term for the photon (i.e. no term of the form $\frac{1}{2}m_A^2 A_\mu A^\mu$) as this would not be gauge invariant: $\frac{1}{2}m_A^2 A_\mu A^\mu \to \frac{1}{2}m_A^2(A_\mu - \partial_\mu\chi)(A^\mu - \partial^\mu\chi) \neq \frac{1}{2}m_A^2 A_\mu A^\mu$. This is precisely what the Higgs mechanism addresses for the electroweak interactions.

Weak Theory

Fermi originally introduced the weak force in 1934 to explain β decay [32, 36, 37]. The weak force was described as a contact force with an interaction strength of about $G_F = 1.2 \times 10^{-5}$ GeV^{-2}. The theory was extended to explain observations in 1957 that parity was not conserved in weak interactions [38, 39]. This was incorporated by giving the weak force a vector−axial (V−A) structure: the interaction vertices contain both vector and axial vector components[8]. This is equivalent to different couplings for left- and right-handed chiral components.

Fermi's theory is not valid at energies where $Q^2 \approx m_W^2$, however. The short range of the interaction indicates that the force should be mediated by a massive particle. This can be formalised by assigning weak isospin values; left-handed fermions are assigned a weak-isospin of magnitude $T = \frac{1}{2}$ grouped into doublets with a third component of $T_3 = \pm\frac{1}{2}$, and the right-handed fermions are assigned $T = 0$ and $T_3 = 0$, excluding them from the weak interactions. The right-handed anti-fermions are assigned $T = \frac{1}{2}$ and $T_3 = \pm\frac{1}{2}$. Left-handed anti-fermions are

[8]There are only five possible Lorentz invariant bilinear covariants that can be constructed with Dirac γ matrices and spinors: scalar, pseudoscalar, vector, vector-axial and tensor. Vector quantities change sign under parity, but axial vectors do not.

assigned $T = 0$ and $T_3 = 0$. In creating a QFT describing these interactions, local gauge invariance requires three new gauge fields to be introduced: W_ν^k ($k = 1, 2, 3$), corresponding to three new gauge bosons W^1, W^2 and W^3. The physical charged $W^\pm$ bosons are given by a linear combination of W_ν^1 and W_ν^2. A third neutral gauge boson, W^3, is predicted as well, which, from experimental evidence from LEP and SLAC [40], is clearly not the Z boson, since it does not couple to right-handed particles as the Z boson does. The Z boson is therefore not accommodated in this field theory.

Since only left-handed particles and right-handed antiparticles participate in the charged weak interactions, the chiral components are included separately in any Lagrangian formalism. Within a generation, the lepton fields are described in left-handed doublets and right-handed singlets, with the doublets containing particles of different flavour:

$$\psi_{Lk}^l = \begin{pmatrix} \nu_\ell \\ \ell \end{pmatrix}_L, \quad \psi_{Rk}^l = \ell_{Rk}, \tag{2.5}$$

where ℓ and ν_ℓ are the lepton flavours, k is the generation (k=1,2,3), and L and R refer to the handedness. There is no observed or predicted right-handed neutrino in the Standard Model.

For quarks, there are two singlets and a doublet for each generation – one for the up-type and one for the down-type quarks:

$$q_L = \begin{pmatrix} u_k \\ d'_k \end{pmatrix}_L, \quad q_R^{(u)} = u_{Rk}, \quad q_R^{(d)} = d_{Rk}, \tag{2.6}$$

where k is the generation and u refers to u, c, t and d' refers to $\{d', s', b'\}$, which are the weak eigenstates of the down-type quark.

Electroweak Unification

In 1961, Glashow first proposed that the weak and electromagnetic forces could be unified in a single theory by considering a QFT invariant under the $SU(2)_L \times U(1)_Y$ symmetry group [6]. In addition to unifying the forces, this allowed for a description of neutral current interactions in the weak sector by predicting the Z boson. Glashow's work still did not provide a mechanism to give masses to the W and Z bosons, however.

Under this proposal, weak hypercharge, Y, with an associated $U(1)$ symmetry, replaces the $U(1)$ symmetry of QED. For invariance under transformations of $SU(2)_L \times U(1)_Y$, hypercharge must be the same for both chiralities of a particle, and an associated gauge field, B_μ, is introduced. Thus, the left-handed doublets, and the right-handed singlets, of the matter fields remain invariant under the transformations:

$$\psi_L \rightarrow \psi'_L = e^{i\alpha_i(x)T_i + i\beta(x)Y}\psi_L, \tag{2.7}$$

$$\psi_R \rightarrow \psi'_R = e^{i\beta(x)Y}\psi_R, \tag{2.8}$$

where $T_i = \sigma_i/2$ are the generators of the $SU(2)$ group and σ_i are the three Pauli matrices, $\alpha_i(x)$ are the $SU(2)$ group parameters, and $\beta(x)$ the $U(1)_Y$ group parameter. These transformations preserve the nature of both the weak and QED interactions. The locally gauge invariant electroweak Lagrangian is given below in Equation 2.9.

$$\mathcal{L}_{EW} = i\sum_f \bar{\psi} i\gamma^\mu D_\mu \psi - \frac{1}{4}W^a_{\mu\nu}W^{a\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}, \tag{2.9}$$

where f runs over all flavours, $D_\mu = \partial_\mu + igT_iW^i_\mu + ig'\frac{Y}{2}B_\mu$ is the covariant derivative, ϵ_{abc} is the Levi-Civita symbol, the weak gauge fields are given by $W^a_{\mu\nu} = \partial_\nu W^a_\nu - \partial_\nu W^b_\mu - g\epsilon_{abc}W^b_\mu W^c_\nu$, and $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$ is the hypercharge gauge field. The couplings to the W^a_μ and B_μ fields are given by g and g', respectively. The third term in $W^a_{\mu\nu}$ is a self-interaction term that predicts a coupling between the W and Z bosons.

The physical $W^\pm$, Z and photon are found by mixing the W^a_μ and B_μ gauge fields:

- $W^\pm_\mu = (W^1_\mu \mp iW^2_\mu)/\sqrt{2}$ ($W^\pm$ bosons)
- $Z_\mu = \cos\theta_W W^3_\mu - \sin\theta_W B_\mu$ (Z boson)
- $A_\mu = \sin\theta_W W^3_\mu + \cos\theta_W B_\mu$ (photon)

Here θ_W is the weak mixing angle, defined as $\tan\theta_W = g'/g$. Moreover, $g\sin\theta_W = g'\cos\theta_W = e$, the electric charge. The value of θ_W must be determined experimentally, after which it can be used, in conjunction with e, to find the coupling constants. The ratio of weak-to-EM coupling constants in terms of θ_W is found, experimentally, to be $\sin^2\theta_W = \frac{\alpha}{\alpha_W} = \frac{e^2}{g^2} \approx 0.23$. From the definitions of A_μ and Z_μ, it can be seen that these fields couple to both the B_μ field and

the third weak gauge field W_μ^3. The Z boson couples to both chiralities as it has contributions from both the B_μ field and W_μ^3, but with different strengths. It has both vector and axial-vector couplings which differ in strength, and therefore does not preserve parity.

The electroweak Lagrangian (Equation 2.9) is still missing mass terms for the $W^\pm$ and Z bosons; adding mass terms to $\mathcal{L}_{EW}$ is not gauge invariant. Similarly, for the fermion mass term:

$$m\bar{\psi}\psi = m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R) \tag{2.10}$$

is not invariant because of the different transformations of the left- and right-handed terms under $SU(2)_L$ (see Equation 2.7). When Glashow first introduced the theory, he made the comment in the introduction to his paper that this was "a stumbling block that we must overlook." This was resolved by spontaneous symmetry breaking and the introduction of the Higgs mechanism by Weinberg and Salam in 1967 [7, 41].

2.2.5 Spontaneous Symmetry Breaking and the Higgs Mechanism

The electroweak Lagrangian in Equation 2.9 is sufficient for massless $W^\pm$ and Z bosons. Masses of the $W^\pm$ and Z bosons can be generated through the Higgs mechanism, a method of spontaneous electroweak symmetry breaking, proposed by Higgs, Brout and Englert [11–14]. In this model, an additional scalar field is introduced into the Lagrangian, with a potential that is invariant under $SU(2)_L \times U(1)_Y$ transformations, and with a non-zero ground state energy, (or *vacuum expectation value (vev)*). Weinberg [7] and Salam [41] expanded on the work of Glashow by introducing this idea of spontaneously breaking *local* symmetries into the electroweak theory to form the Glashow-Weinberg-Salam (GWS) model. Essentially, in the Standard Model, the Higgs mechanism breaks the $SU(2)_L \times U(1)_Y$ electroweak symmetry, giving mass to the vector bosons in the process, whilst keeping the $U(1)$ symmetry of QED intact and remaining locally gauge invariant. A brief overview of this process is discussed below.

The Higgs scalar field is introduced as an isospin doublet of four complex scalar fields, $\phi(x)$, with weak hypercharge $Y = 1$, total weak isospin $T = \frac{1}{2}$, and $T_3 = \pm\frac{1}{2}$

for the electrically charged upper and neutral lower components, respectively:

$$\phi(x) = \begin{pmatrix} \phi^+(x) \\ \phi^0(x) \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1(x) + i\phi_2(x) \\ \phi_3(x) + i\phi_4(x) \end{pmatrix}. \tag{2.11}$$

The potential of this field is given by

$$V(\phi) = \frac{1}{2}\mu^2\phi^\dagger\phi + \frac{1}{4}\lambda(\phi^\dagger\phi)^2, \tag{2.12}$$

and is illustrated in Figure 2.2. The μ^2 term in $V(\phi)$ can be either negative or positive, but λ is required to be positive so that the potential is bounded from below. For $\mu^2 > 0$ the potential is parabolic, and for $\mu^2 < 0$ it is shaped like the bottom of a wine bottle. There is a local maximum at $\phi(x) = 0$, with minima along the circumference where $\phi^\dagger\phi = v^2 = -\frac{\mu^2}{\lambda} \neq 0$. At any of the minima, the symmetry of the potential is broken. In expressing the ground state, choosing any three of the four ϕ_i independently to be 0 and the other to be $v = \sqrt{-\frac{\mu^2}{\lambda}}$ results in no loss of generality.

In the GWS model, the vacuum expectation value, ϕ_0, of $\phi(x)$ is chosen as

$$\phi_0 \equiv \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{2.13}$$

The motivation for such a choice of ϕ_0 is that any symmetry that it breaks will generate a mass for the corresponding gauge boson. The minimum of the potential must correspond to a non-zero vacuum expectation value for *only* the neutral scalar field ϕ^0 component of $\phi(x)$. The assigned hypercharge and weak isospin of ϕ_0 breaks both the $SU(2)$ and $U(1)_Y$ symmetries, but the $U(1)_{EM}$ symmetry remains unbroken. This can be seen in the context of the electroweak transformations given in Equation 2.7. For each of the $SU(2)$ generators T_i, and the weak hypercharge generator, acting on ϕ_0 is non-zero. The $U(1)_{EM}$ generator, $Q = T_3 + \frac{Y}{2}$, is not broken:

$$Q\phi_0 = \frac{1}{2}(\sigma^3 + I) \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} = 0. \tag{2.14}$$

Thus, the choice of potential $V(\phi)$ has broken the electroweak symmetry, whilst the electromagnetic symmetry is maintained. Importantly, the neutral component

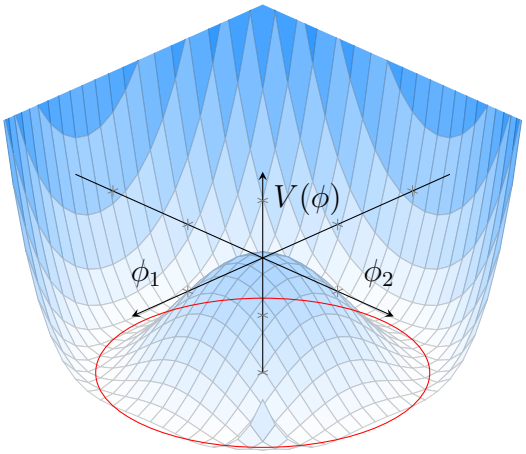


Figure 2.2 *The $V(\phi)$ potential of the complex scalar field $\phi(x)$, with $\mu^2 < 0$. There are degenerate minima along the circumference of the red circle. This choice of a non-symmetrical potential spontaneously breaks the symmetry of the Lagrangian [42].*

of ϕ_0 has a non-zero vev, such that the remaining unbroken symmetry leaves the photon massless. As is shown in the expansion below in Equation 2.16, masses are obtained for the gauge bosons corresponding to the broken symmetries.

Since the vacuum expectation value is non-zero, ϕ is redefined such that the ground state is effectively close to zero, allowing perturbative calculations to be performed. Excitations about the ground state describe the particles. Where the vev of the field is zero, such as ψ in electromagnetism, the field already corresponds to the particle, such as fermions or photons. Since in this case the vev is non-zero, the field ϕ is redefined, without any loss of generality, such that the new vev is *close* to zero. Excitations about this ground state can then be calculated. Since $\phi(x)$ is a complex field, expanding about the ground state could be performed in terms of real and complex components, $h(x)$ and $\xi(x)$, such that an expansion would be about $v + h(x) + i\xi(x)$. In the unitary gauge, $\alpha(x) = -\xi(x)/v$, the ξ dependence in the Lagrangian disappears. In the unitary gauge the perturbative calculations can be performed by setting

$$\phi(x) = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \tag{2.15}$$

where $h(x)$ represents excitations about the ground state, which, as shown in Equation 2.18, correspond to massive Higgs bosons.

Particle Masses

The masses of the gauge bosons can be seen explicitly by expanding the locally gauge invariant Lagrangian about the ground state in Equation 2.15 and identifying the quadratic terms:

$$\begin{aligned}\mathcal{L}_H &= (D^\mu\phi)^\dagger(D_\mu\phi) - \frac{1}{2}\mu^2\phi^\dagger\phi - \frac{1}{4}\lambda(\phi^\dagger\phi)^2 \\ &= \frac{1}{2}(\partial^\mu h)(\partial_\mu h) - \underbrace{\lambda v^2 h^2 + \frac{v^2}{4}g^2 W^{+\mu}W^-_\mu + \frac{v^2}{8}(g^2+g'^2)Z^\mu Z_\mu}_{\text{mass terms}} \\ &\quad + \underbrace{\frac{1}{2}g^2 W^{+\mu}W^-_\mu(vh+h^2) + \frac{1}{8}(g^2+g'^2)Z^\mu Z_\mu(vh+h^2) - \lambda v h^3 - \frac{\lambda}{4}h^4}_{\text{interaction terms with the Higgs field}} \\ &\quad + \text{derivatives} + \text{constants},\end{aligned} \tag{2.16}$$

where $D_\mu = \partial_\mu + igT_aW^a_\mu + ig'\frac{Y}{2}B_\mu$. Reading the mass terms from this gives masses for the W and Z:

$$m_W = \frac{1}{2}gv \tag{2.17a}$$

$$m_Z = \frac{1}{2}v\sqrt{g^2+g'^2}, \tag{2.17b}$$

and an additional massive particle from the h^2 term, the *Higgs boson*, with mass

$$m_H = \sqrt{2\lambda v^2}. \tag{2.18}$$

The theory also predicts couplings between $h(x)$ and the other gauge fields, such that the Higgs boson will couple to itself, and the $W^\pm$ and Z bosons, which allows for indirect couplings to photons too, through intermediate loops of the massive gauge bosons.

The GWS model has only four free parameters, which need to be determined from experimental measurements: the weak coupling constant g, the weak hypercharge coupling g', and the λ and μ terms in the Higgs potential. From the relation in Equation 2.18, and $m_W = \frac{1}{2}gv$, and the measured values of g and m_W, a value of v is predicted to be $v = 246$ GeV. From this, a measurement of m_H provides a value for λ.

The masses of the fermions can be given by including Yukawa couplings between

the Higgs field and the fermions, and included as a separate Lagrangian term [30]:

$$\mathcal{L}_{Yuk} = -G_f \left(\bar{\psi}_L \phi \psi_R + \bar{\psi}_R \phi^\dagger \psi_L \right), \tag{2.19}$$

where G_f is the coupling for a given fermion, and ψ_L, ψ_R are the fermion doublet and singlet from Equations 2.5 and 2.6.

Considering the Higgs potential, and expanding around the minimum this becomes

$$\mathcal{L}_{Yuk} = -\frac{G_f}{\sqrt{2}}(\bar{\psi}_L \psi_R + \bar{\psi}_R \psi_L)(v + h(x)). \tag{2.20}$$

Giving a mass for the fermion of $m_f = \frac{G_f v}{\sqrt{2}}$ in both the charged lepton and quark sectors.

2.2.6 Summary

The final Standard Model Lagrangian is a sum of all of the Lagrangians from QCD, electroweak theory, the Higgs sector and the Yukawa terms:

$$\mathcal{L}_{SM} = \mathcal{L}_{QCD} + \mathcal{L}_{EW} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yuk}. \tag{2.21}$$

In summary, the Higgs mechanism provides masses to the $W^\pm$ and Z bosons in the context of electroweak interactions through spontaneous local symmetry breaking. A new complex scalar field, the Higgs field, is introduced, which mixes with the gauge fields and generates masses for the $W^\pm$ and Z bosons when spontaneous symmetry breaking occurs. Three of the four degrees of freedom in the Higgs field couple to the previously massless bosons and give masses to the $W^\pm$ and Z bosons, whilst the final degree of freedom predicts a new particle: the Higgs boson. Yukawa couplings are included between the Higgs field and the fermions, providing mass terms for the fermions.

2.3 Limitations of the Standard Model

Whilst the Standard Model has been outstanding in many of its predictions, there are strong indications, both from experimental observations and theoretical considerations, that the SM is not complete. Some of these are listed here:

- Neutrino masses – The masses for the neutrinos are not included a priori. The observation of massive neutrinos by Super-Kamiokande [43] and SNO [44, 45] can be accommodated by the introduction of the PMNS matrix [46], which describes the neutrino mixing in weak interactions, analogous to the CKM matrix in the quark sector [47, 48].

- Matter-antimatter asymmetry – The *observable* Universe is constituted almost entirely of matter, with relatively little antimatter. This overall abundance of matter over antimatter in the Universe is not explained by the SM. CP violation in the Standard Model can arise through the CKM and PMNS matrices, although it has only been observed in the quark sector. The required asymmetry between the number of baryons observed cannot be accounted for by the CP violation seen so far [20].

- Dark Matter – Based on recent observations by the WMAP [49] and Planck [50] experiments, ordinary matter accounts for 4.9% of the observed universe, the remaining Dark Energy (68.3%) and Dark Matter (26.8%) are not explained by the SM and cosmological models.

- The Hierarchy Problem – This is also known as naturalness. When trying to formulate a Grand Unified Theory, or introduce New Physics, loop corrections to the Higgs boson mass diverge at high energies. Without fine-tuning the parameters this implies that the Standard Model cannot be used up to high energy scales. Related to the hierarchy problem is the lack of an explanation for the large differences in masses for each generation of quarks and leptons. For example, the electron is 200 times lighter than the muon, and 3500 times lighter than the tau. Gravity is not accounted for at all, and no explanation is provided for gravity being far weaker than the three other forces.

2.4 Higgs Boson Phenomenology

The Higgs boson is predicted by the SM to be a scalar, parity conserving particle. However, its mass is a free parameter in the theory and must be determined experimentally. Production cross sections and branching ratios of the Higgs boson are predicted as a function of its mass. Discovering the Higgs boson was one of the major physics goals at the LHC. Prior constraints on the mass were available from LEP [51] ($m_H < 114.4$ GeV was excluded at 95% confidence level) and the

Tevatron [52] ($149 < m_H < 182$ GeV and $90 < m_H < 109$ GeV were excluded at 95% confidence level). The ATLAS and CMS collaborations both announced on the 4th of July 2012 that a Higgs-like boson of mass 125 GeV (see Section 2.4.3) had been discovered [16, 17]. This discovery was based on the decays $H \to \gamma\gamma$, $H \to ZZ^* \to 4\ell$ and $H \to WW^* \to \ell\nu\ell\nu$. The Higgs boson can decay into a number of additional final states, which are discussed below. Finding evidence in each of these decays is important to ascertain the nature of the observed boson, especially in the decays to fermions, such as $H \to b\bar{b}$.

2.4.1 Higgs Boson Production

The production of a SM Higgs boson at hadron colliders can occur in multiple ways, as shown in Figure 2.3. At the LHC, the dominant production channels are through gluon fusion (ggF), vector boson fusion (VBF) and Higgs-Strahlung with an associated vector boson [53].

The associated vector boson production at the LHC is studied in Chapter 5 in the context of a search for a Higgs boson decaying to $b\bar{b}$, with a short motivation for this production mechanism provided here.

The Higgs-Strahlung production mechanism has a significantly lower cross section than the other production mechanisms, but it offers the benefit of having an experimental signature of a massive vector boson. In attempting to identify the decay $H \to b\bar{b}$, there are challenges associated with each of the production mechanisms, in particular due to large QCD backgrounds. The associated production with a W/Z boson, where the vector boson decays leptonically, provides a handle on the large QCD backgrounds.

Of the other production mechanisms, gluon fusion has the largest cross section, dominated by a top quark loop, due to the large coupling of the top quark to the Higgs boson. Vector boson fusion can be identified by two *forward* jets in the detector, although its cross section is an order of magnitude lower than gluon fusion. Associated production with a top-quark pair, $t\bar{t}H$, has a relatively small cross section, but it can probe the Higgs boson fermion couplings directly, which is one of the motivations for the search for a Higgs boson in this channel. The production cross sections at $\sqrt{s} = 8$ TeV are shown in Figure 2.4.

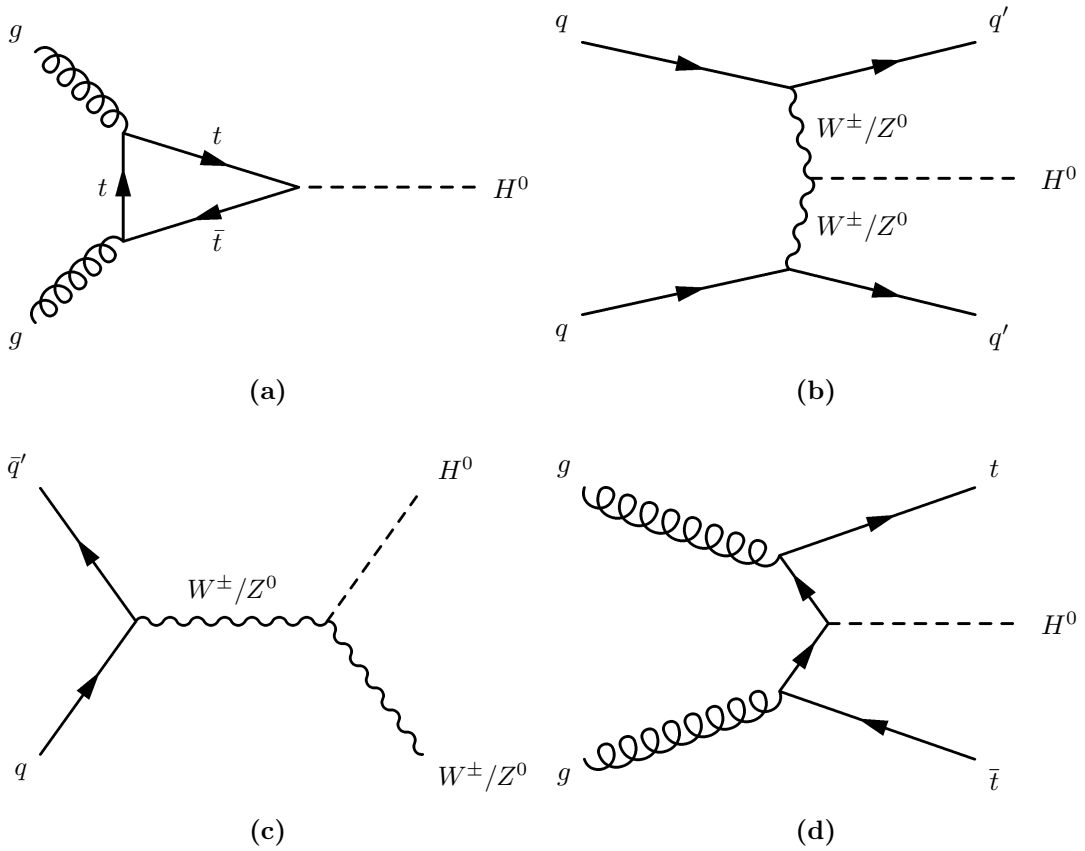


Figure 2.3 *The Higgs production mechanisms at hadron colliders are shown in (a)-(d). Gluon-gluon fusion (ggF) is shown in (a), where two gluons interact via top quarks to produce a Higgs boson. Vector boson fusion (VBF) is shown in (b), where two vector bosons produce a Higgs boson and two final state quarks. Higgs-Strahlung is shown in (c), where a W or Z boson is produced in association with a Higgs boson. $t\bar{t}H$ production is shown in (d), where a top quark pair is produced in association with a Higgs boson.*

2.4.2 Higgs Boson Decays

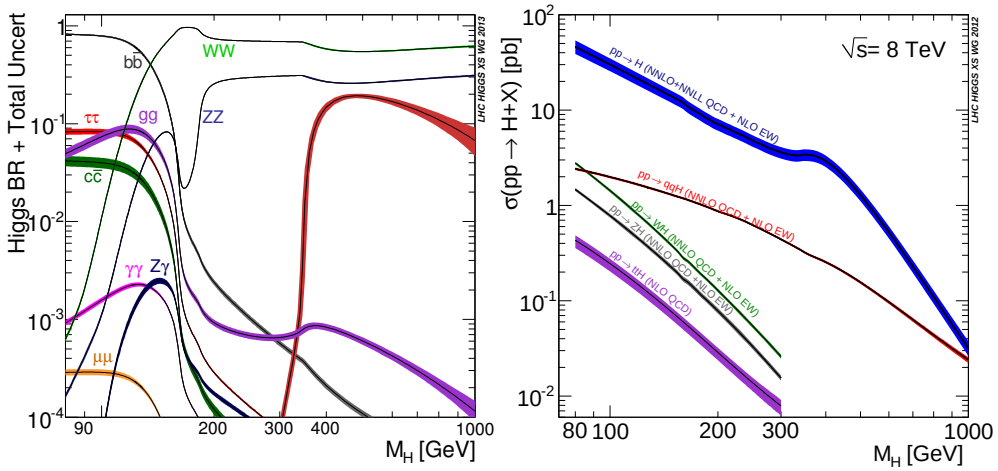
The Higgs boson is short-lived, with a lifetime on the order of 10^{-22} seconds[9] and thus decays close to the interaction point in the ATLAS detector. It must therefore be identified from its decay products.

The Higgs boson couples to all massive fermions and massive vector bosons, including self-couplings. Couplings to gluons and photons are possible indirectly, through intermediate loops of other particles. An exception to this is the direct decay of the Higgs boson to the top quark, since the top quark is too heavy. The

[9]The lifetime is given by $\tau = \frac{\hbar}{\Gamma}$. For a Higgs boson of mass 125.09 GeV the predicted decay width is 4.1×10^{-3} GeV, giving a lifetime of 1.61×10^{-22} seconds [54].

Higgs boson can decay into W and Z bosons, via off-shell contributions for W/Z bosons. The branching ratios for the decay modes as a function of the Higgs boson mass are shown in Figure 2.4.

The dominant decay channel for a Standard Model Higgs boson with a mass of 125.09 GeV (the combined mass measurement from ATLAS and CMS [55]) is to $b\bar{b}$, with a branching ratio of 58.1%. Observing the $H \to b\bar{b}$ process would allow for a direct measurement of the coupling strength of the Higgs field to fermions.



(a) **(b)**

Figure 2.4 *The branching ratios for different masses of Higgs boson are shown in (a). The favoured branching ratio at a mass of $m_H = 125$ GeV is to $b\bar{b}$. The bosonic channels WW, ZZ and $\gamma\gamma$, have lower branching ratios, but they also have much lower backgrounds. On the right hand side in (b) the cross sections for the different production mechanisms at $\sqrt{s} = 8$ TeV are shown [54].*

2.4.3 Higgs Discovery and Current Status

Plots of the latest Higgs boson mass distributions for the three discovery channels, $H \to ZZ^* \to 4\ell$, $H \to \gamma\gamma$ and $H \to WW^* \to \ell\nu\ell\nu$, are shown in Figure 2.5 with the full Run 1 dataset at $\sqrt{s} = 7$ and 8 TeV, corresponding to integrated luminosities of 4.5-4.6 fb^{-1} and 20.1 fb^{-1}, respectively.

Subsequent studies have been carried out using more data and improved techniques to probe the properties of this newly discovered particle. All the current results provide further evidence for this being the Standard Model Higgs boson with a spin-0 nature and positive parity [18].

The most recent results, combined with CMS, can be found in Reference [55]. This uses the combination of $H \to ZZ$, WW, $\gamma\gamma$, $\tau^+\tau^-$, $b\bar{b}$ and $\mu^+\mu^-$ using the full datasets from Run 1. The combined mass measurement from ATLAS and CMS is $125.09 \pm 0.21(stat.) \pm 0.11(syst.)$ GeV with a measured signal strength[10] of $\mu = 1.09 \pm 0.11$, in agreement with the SM. ATLAS sees $H \to \tau^+\tau^-$ decays with a significance of 4.5σ [56], which, when combined with CMS, increases to 5.4σ and shows evidence of VBF production with a significance of 5.5σ. There is still no discovery of $H \to b\bar{b}$, although ATLAS sees a significance of 1.4σ at 125 GeV with the full Run 1 dataset [1], with a signal strength of $\mu^{bb} = 0.65^{+0.43}_{-0.40}$. Combined with CMS there is a significance of 2.6σ with a signal strength of $\mu^{bb} = 0.69^{+0.29}_{-0.27}$ [55]. A signal significance of 2.8σ is seen for $H \to b\bar{b}$ by the Tevatron [57].

2.5 W Boson

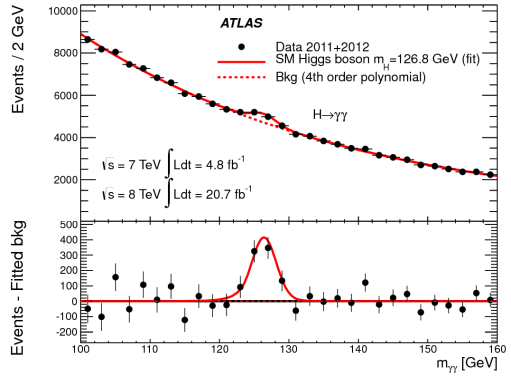
As this thesis focuses on WH production and the implementation of improved W boson identification at high transverse momentum, some properties of the W boson are discussed here. The two charged $W^\pm$ bosons are referred to as the W boson for most parts of the text, as both are considered at the same time.

W bosons were first discovered in 1983 by the UA1 and UA2 experiments at CERN [60, 61]. They had been predicted by the GWS model in the 1960s, along the with Z boson. The mass and spin have been measured as $m_W = 80.385 \pm 0.015$ GeV and spin-1 [20], respectively, from the Tevatron and LEP.

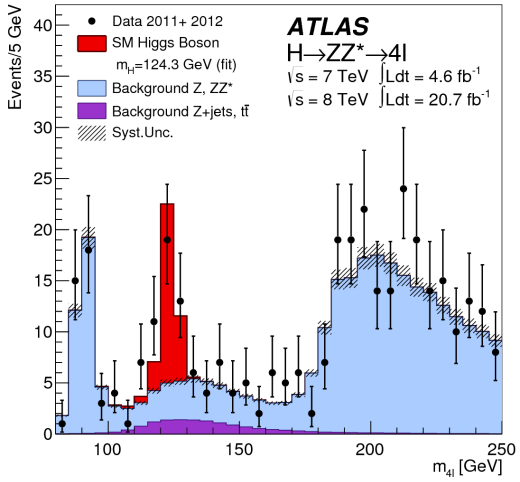
The W boson decays leptonically to a left-handed fermion and a right-handed antifermion in the same isospin doublet. The W boson also decays hadronically to pairs of quarks, except for the top quark, whichare not necessarily in the same isospin doublet due to CKM mixing. The W boson has a short lifetime on the order of 10^{-25} seconds, and as such is identified by its decay products. Leptonic decays of the W boson account for about 32.6% of all decays, and hadronic decays for about 67.4%. The W boson decay modes and their branching ratios are listed in Table 2.2.

There are some experimental benefits from considering leptonic final states; triggering on electrons and muons can be done with high efficiency with the

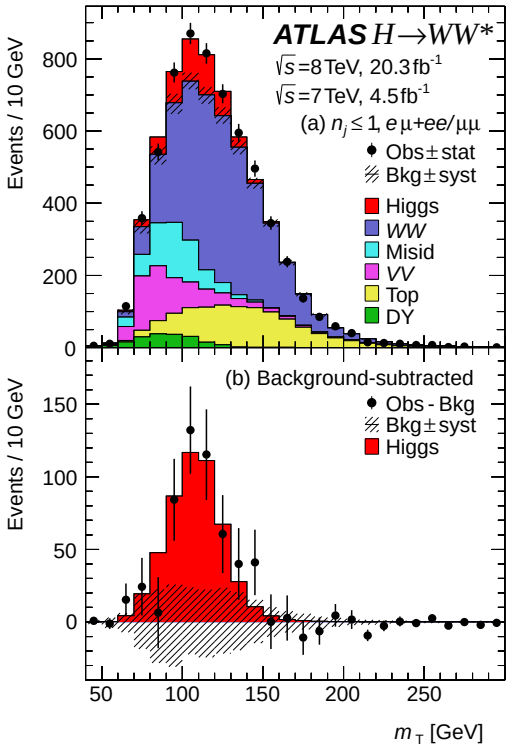
[10] Signal strength is defined in Section 5.1.



(a) $H \to \gamma\gamma$



(b) $H \to ZZ^* \to 4\ell$



(c) $H \to WW^* \to \ell\nu\ell\nu$

Figure 2.5 *Mass plots for the three channels $H \to ZZ^*$, $H \to WW^*$ and $H \to \gamma\gamma$, which contributed to the initial discovery of the Higgs boson [58, 59]. The plots shown here use the full Run 1 ATLAS dataset. A clear excess is observed in each of these channels. The mass of the Higgs boson is estimated from the ZZ and $\gamma\gamma$ channels, which have excellent mass resolution.*

ATLAS detector. The neutrino cannot be detected in the detector, and thus missing momentum in an event can be used to identify the W boson decay. Since the τ is heavy enough to decay hadronically, the $W \to \tau\nu$ process, where the τ does decay hadronically, difficult to study experimentally due to the large backgrounds at the LHC. The leptonic decays of the τ can be identified by the presence of electrons or muons and missing energy in the event due to neutrinos. Since there is already missing energy from the neutrinos is the W boson decay, this process can be misidentified as $W \to e/\mu\nu$.

Decay modes	Fraction Γ_i/Γ
$\ell^-\bar{\nu}_\ell$	$10.86 \pm 0.09\%$
$e^-\bar{\nu}_e$	$10.71 \pm 0.16\%$
$\mu^-\bar{\nu}_\mu$	$10.63 \pm 0.15\%$
$\tau^-\bar{\nu}_\tau$	$11.38 \pm 0.21\%$
$q\bar{q}$	$67.41 \pm 0.27\%$

Table 2.2 *The different decay modes and branching ratios for the W^- boson [20]. The W^+ decay modes are given by the charge conjugates of these. In the table ℓ refers to the average lepton branching ratio. q refers to any quark except the top quark.*

Hadronic decays of the W boson are identified by two jets. As discussed in Section 2.2.3, quarks hadronise and are identified by clustering the energy deposits from the decays; so-called jets. In general, jets are identified by considering a cone of a specific radius and looking for energy deposits within the detector.

2.6 Heavy Vector Triplet Model

Given the shortcomings of the Standard Model discussed in Section 2.3, many models *beyond the Standard Model (BSM)* are proposed. Many of these theories contain multiple free parameters which must be found from a direct comparison with data. By considering direct experimental manifestations of these models, such as new heavy particles or resonances, many of the free parameters can be neglected. Resonance searches are typically not sensitive to all the free parameters in the model; they are only sensitive to the parameters which affect the mass of the resonance, and the interactions that provide the production and decay of the resonance. The resonance can be given a simplified description in terms of a phenomenological Lagrangian that only considers the relevant couplings and mass parameters. The Lagrangian must be constructed to describe the phenomenology of a broad range of models, but it is not needed to meet any additional theoretical requirements. If a resonance is found, the observation can be expressed in terms of this simplified description, and compared with numerous models that allow for the phenomenological parameters of the Lagrangian to be calculated. One such model is the Heavy Vector Triplet (HVT) Model A [19, 20].

In the HVT model, a triplet of heavy particles is introduced: the $W'^{\pm}$ and Z', which are degenerate in mass and have comparable production rates. In

Model A, the new triplet field arises from the spontaneous symmetry breaking of two additional $SU(2)$ gauge groups that reduce to the electroweak gauge group: $SU(2)_1 \times SU(2)_2 \times U(1)_Y \to SU(2)_L \times U(1)_Y$.

The parameterisation of the couplings in the HVT Model A Lagrangian allows for the phenomenological description of a large number of BSM models, and the generality of the model allows it to be used as a framework for interpreting the results in terms of these BSM models. The triplet field in Model A couples to the SM vector bosons, fermions and the Higgs boson, with couplings given by the parameters $g^2 C_F/g_V$, where g is the SM $SU(2)_L$ gauge coupling, $C_F \sim 1$ is a multiplicative factor of the fermion couplings, and g_V is the coupling strength to the new triplet. The coupling to the Higgs boson is given by $g_V C_H$, where $C_H \sim -g^2/g_V^2$, is a multiplicative factor of the Higgs boson coupling. The branching ratios of the processes $W' \to WZ$, $W' \to WH$, $Z' \to WW$ and $Z' \to ZH$ are all approximately 2% for masses in the range 1 to 3 TeV. Cross sections, branching fractions and particle decay widths are given in Table 2.3 for a number of HVT signal masses with $g_V = 1$.

Any W, Z and H bosons produced in the decay of the W' and Z' will have a large transverse momentum. When particles are produced with momentum greatly exceeding their mass, their decay products are *boosted* in the lab frame. Some of the methods for identifying hadronic decays of such boosted particles are discussed in the context of the W boson in Section 2.7.

Table 2.3 *The resonance width (Γ) and the product of cross section times branching ratios (BR) with $g_V = 1$, where two vector bosons decay hadronically in model A for the HVT W' and Z' for several values of resonance pole masses (m).*

		$W' \to WZ$	$Z' \to WW$
m	Γ_{HVT}	$\sigma \times$BR	$\sigma \times$BR
[TeV]	[GeV]	[fb]	[fb]
1.3	33.3	62.7	28.7
1.6	40.9	23.3	10.6
2.0	51.0	7.6	3.35

2.7 Boosted W Boson Identification

As mentioned in Section 2.5, the hadronic decay of W bosons often results in two jets in the detector. If the two jets are sufficiently close in the detector, they can be *clustered* together, resulting in what appears to be a single jet [62]. This is likely to happen in the case of the W boson if it has a high transverse momentum. Specialised identification methods are thus required for boosted W bosons.

In identifying boosted W bosons, it is important to discriminate between a jet that comes from the hadronisation of a single quark (QCD backgrounds at the LHC, for example), and a jet that is made up of multiple quarks. There are a number of ways of approaching this challenge, as discussed in Chapter 7. These rely on characterising jet *substructure* to identify when a jet contains multiple *subjets*, where each subjet corresponds to a single quark or gluon. These methods take advantage of the topology of the W boson decays, where the mass of the subjets is a lot smaller than the mass of the W boson.

Diboson decays of the W' and Z' from the HVT model introduced in Section 2.6 are used in Chapter 7 to study boosted W bosons.

The methods developed are not specific to boosted W bosons and are applicable to Z and Higgs bosons. In the $WH(\rightarrow b\bar{b})$ search shown in Chapter 5, the highest sensitivity is found in regions where the W boson has a large transverse momentum. Whilst this search considers leptonic decays of the W boson, the $H \rightarrow b\bar{b}$ process could stand to benefit from the techniques which are presented here for boosted W boson tagging.

Chapter 3

The LHC and the ATLAS Detector

"The Hadron Collider is a place of science. Various research takes place in the facility and it offers education to citizens. A city with a Hadron Collider doesn't have to worry about education, the facility provides it for all citizens." - Cities: Skylines PC Game [63, 64]. Image © 2015-2016 Paradox Interactive AB.

3.1 Introduction

The Large Hadron Collider (LHC) at CERN is the largest and most powerful particle accelerator ever built. The high energy proton collisions offer a glimpse into the conditions present in the primordial Universe, immediately after the Big Bang, and an opportunity to probe elementary particles and their interactions at an unprecedented energy scale. The primary goals for the LHC are to test and verify the Standard Model of particle physics, in particular to study the Higgs boson, and to find any hints of New Physics. ATLAS is one of two general purpose particle detectors at the LHC that measure the outcomes of the proton collisions. The amount of data produced in these collisions is vast, and immense computational power and storage is required to process the data. This is handled by a trigger system consisting of multiple levels in hardware and software, and multi-tiered distributed computing systems.

This chapter gives a brief description of the LHC in Section 3.2, followed by a description of ATLAS in Section 3.3, and finally the trigger systems and computational systems are described in Sections 3.4 and 3.5, respectively.

3.2 Large Hadron Collider

The LHC [9] is a subterranean hadron collider at CERN near Geneva on the border of Switzerland and France. Beams of protons or heavy ions (Pb^+) are accelerated in separate beam-pipes in opposite directions in a 27 km circular tunnel between 50 and 175 m underground, which previously housed the Large Electron Positron collider (LEP). These beams are made to collide at four interaction points within particle detectors, which record the output of these collisions. The design allows for proton-proton collisions with a centre-of-mass energy of $\sqrt{s} = 14$ TeV with an *instantaneous luminosity* of $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1}, or heavy ions with an energy of up to 2.76 TeV per beam and $\mathcal{L} = 10^{27}$ cm^{-2}s^{-1}. The instantaneous luminosity, $\mathcal{L}$, is the collision rate per unit area, and defined as $\mathcal{L} = \frac{1}{\sigma}\frac{dN}{dt}$, the number of events N detected in a time dt in a given cross section σ. The LHC is predominantly used for proton-proton collisions, with shorter, dedicated heavy-ion runs. During 2011 the collision energy was $\sqrt{s} = 7$ TeV and during 2012 it was $\sqrt{s} = 8$ TeV. In 2015 for Run 2, this was increased to $\sqrt{s} = 13$ TeV.

The protons used at the LHC are obtained from a hydrogen gas bottle. Electrons are stripped from the hydrogen atoms by applying an electric field to the gas, and the remaining protons go through a multi-staged process of acceleration to reach their maximum velocity. This begins with a linear accelerator, LINAC2, accelerating the protons up to 50 MeV. They are subsequently injected into the Proton Synchrotron Booster (PSB) where they reach energies of up to 1.4 GeV. They are then fed to the Proton Synchrotron (PS) which accelerates them up to 25 GeV, and then into the Super Proton Synchrotron (SPS) which accelerates them up to 450 GeV. At this point, they are injected into the main LHC ring and accelerated by 16 radio-frequency (RF) cavities up to their maximum velocity. The protons are accelerated in bunches and collided at intervals of 25-50 ns. The layout of the collider is shown in Figure 3.1.

Superconducting magnets are used to bend the path of the protons around the collider (the beams do not follow a perfect circle; there are straight sections). There are 1232 dipole magnets, with a field strength of up to 8.3 T, which guide the proton bunches around the main ring. There are a further 392 quadrupole magnets of 6.5 T each, which focus the beam, and more specialised magnets at places such as the beam injection points. The massive field strength of these magnets necessitates them being superconducting, cooled by liquid helium to an operating temperature of 1.9 K.

Many of the physics processes that are being searched for at the LHC have a tiny cross section. In order to study such rare processes, the amount of data collected must be maximised. One of the ways in which to accomplish this is by increasing instantaneous luminosity. The instantaneous luminosity depends on a number of factors including the number of particles per bunch, bunches per beam, and the horizontal and vertical beam size at the point of interaction, characterised by β^* functions that are properties of the accelerator and the focussing magnets. The LHC luminosity design goals are achieved with a maximum of 2808 bunches in circulation at any one time, each with up to $\approx 10^{11}$ protons, and a bunch spacing of 25 ns (Run 1 from 2010-2012 ran with a bunch spacing of 50 ns [65]). For any given process, the number of events is dependent on its cross section and the *integrated luminosity*. The integrated luminosity, $L = \int \mathcal{L} dt$, is a measure of the total amount data collected expressed in terms of an inverse cross section (typically fb^{-1}). During Run 1 of the LHC the ATLAS experiment recorded 4.7 fb^{-1} at $\sqrt{s} = 7$ TeV and 20.3 fb^{-1} at $\sqrt{s} = 8$ TeV, and during Run 2 in 2015 3.2 fb^{-1} of data was recorded at $\sqrt{s} = 13$ TeV. The relation between the rate of

a type of event and L is given by:

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

where N_{event} is the number of a type of event occurring per second, L is the total integrated luminosity and σ_{event} is the cross section of the event, a measure of the probability for the interaction. Increasing the luminosity makes additional proton-proton interactions more likely, known as *in-time pile-up*, which makes it more difficult to separate single interactions. Additionally, the high frequency of collisions (up to 40 MHz in the ATLAS detector), and the inherent latency of the hardware used in the detectors causes further *out-of-time pile-up*. This is discussed in more detail in Section 4.4.

The collisions are recorded by seven detector experiments placed around the beam at four interaction points (see Figure 3.1). Here, magnets near the detectors bring the particles from the opposing beams closer to each other in order to cause a collision. There are two multi-purpose detectors: A Toroidal LHC Apparatus (ATLAS) Experiment [10] and the Compact Muon Solenoid (CMS) Experiment [15]. These are used to search for New Physics and provide Standard Model measurements. The LHC-beauty (LHCb) Experiment [66] is primarily used for investigating flavour physics and CP violation through b-hadron interactions. A Large Ion Collision Experiment (ALICE) [67] is designed to collect data from heavy-ion collisions, which are used to investigate quark gluon plasma and QCD processes. The LHC-forward (LHCf) detector records collisions in the forward regions, almost parallel with the beam-pipe, investigating the origin of ultra-relativistic cosmic rays through neutral pion production. The TOTal cross-section, Elastic scattering and diffraction dissociation Measurment (TOTEM) experiment measures the total cross section and the luminosity of the LHC [68]. The Monopole and Exotics Detector (MoEDAL) [69] experiment searches for direct evidence of magnetic monopoles and highly ionising massive (pseudo-)stable charged particles.

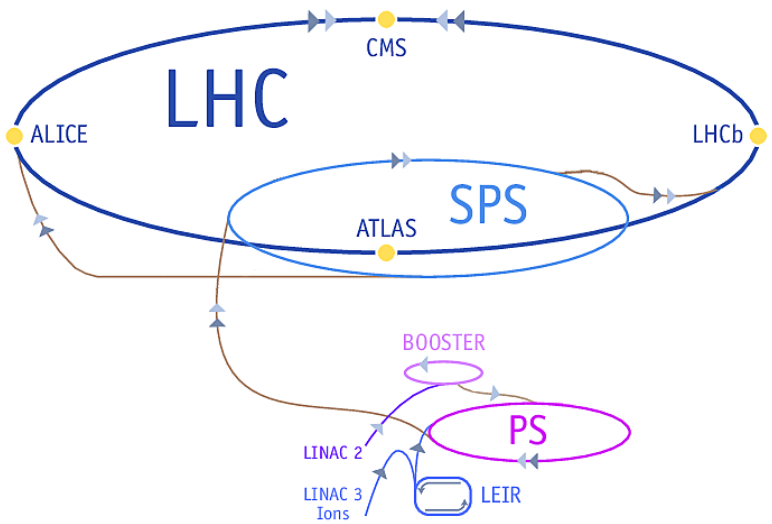


Figure 3.1 *Overview of the LHC showing the four main experiments and the accelerator chain. LINAC2 initially accelerates the protons. They are then injected into the Proton Synchrotron Booster (PSB), then from the PSB into the Proton Synchrotron (PS) and on into the Super Proton Synchrotron (SPS). After this, they are injected into the main LHC ring and accelerated up to their maximum velocity. Adapted from [70].*

3.3 ATLAS Detector

3.3.1 Overview

The ATLAS detector [10] (Figure 3.2) is the largest particle detector at the LHC, measuring 44 m in length, 25 m in diameter and weighing 7000 metric tonnes. It is designed to study many diverse physics processes from both the Standard Model and Beyond the Standard Model (BSM). One of the most important goals in the ATLAS physics programme has been the search for the Higgs boson. Some Standard Model physics goals include precision QCD measurements, flavour physics and electroweak physics studies. Particles produced in these interactions leave characteristic patterns within the detector; charged tracks or energy deposits, for example. The detector thus requires efficient, high resolution particle reconstruction and identification hardware capable of measuring as many of the particle properties as possible. Additional requirements include good vertex resolution for flavour tagging and pile-up rejection, and a fast trigger system for filtering out uninteresting events.

ATLAS has a forward-backward cylindrical symmetry and near hermetic cov-

erage. It is divided into three main subdetectors, each of which is designed to identify specific types of interactions and properties. The detector has an 'onion' type structure centred around the point of the collisions with the subdetectors in numerous layers parallel to the beam (the *barrel region*) and *end-cap regions* perpendicular to the beam. This is illustrated in Figure 3.2.

The *Inner Detector* (ID) is designed to track the paths of charged particles as they travel through the detector, and to reconstruct the interaction vertices. It is subdivided into layers of different technologies, with the highest precision provided closest to the interaction point, and it is surrounded by a 2 T solenoidal magnet.

The calorimetry system surrounds the ID and inner solenoid, absorbing and measuring the energy of interacting charged and neutral particles. The inner *electromagnetic calorimeter* measures the energy deposits from showers of electromagnetically interacting particles, and the outer *hadronic calorimeter* measures the energy deposits from hadrons.

High momentum muons have a low energy loss when traversing the detector, and coupled with their relatively long lifetime of 2.2 μs, are the only charged particles that are frequently able to pass through both the ID and calorimeters. The muon system surrounds the calorimeters and ID, extending out to the maximum radius of the detector. A large toroidal magnet, which gives ATLAS its characteristic shape (and from which the name ATLAS is derived), bends the trajectory of the muons so that their momentum can be inferred. The muon system consists of a number of subdetectors for precise muon tracking.

An illustration of particle interactions moving through the detector is shown in Figure 3.3.

The following sections describe the ATLAS subdetectors in more detail, the technologies used and their dimensions.

3.3.2 Coordinate System

A right-handed coordinate system (illustrated in Figure 3.4) is used with the x-axis pointing towards the centre of the LHC ring, the y-axis pointing upwards and the z-axis along the beam direction. The polar (θ) and azimuthal (ϕ) angles are defined according to these axes. Rapidity and pseudorapidity are used as

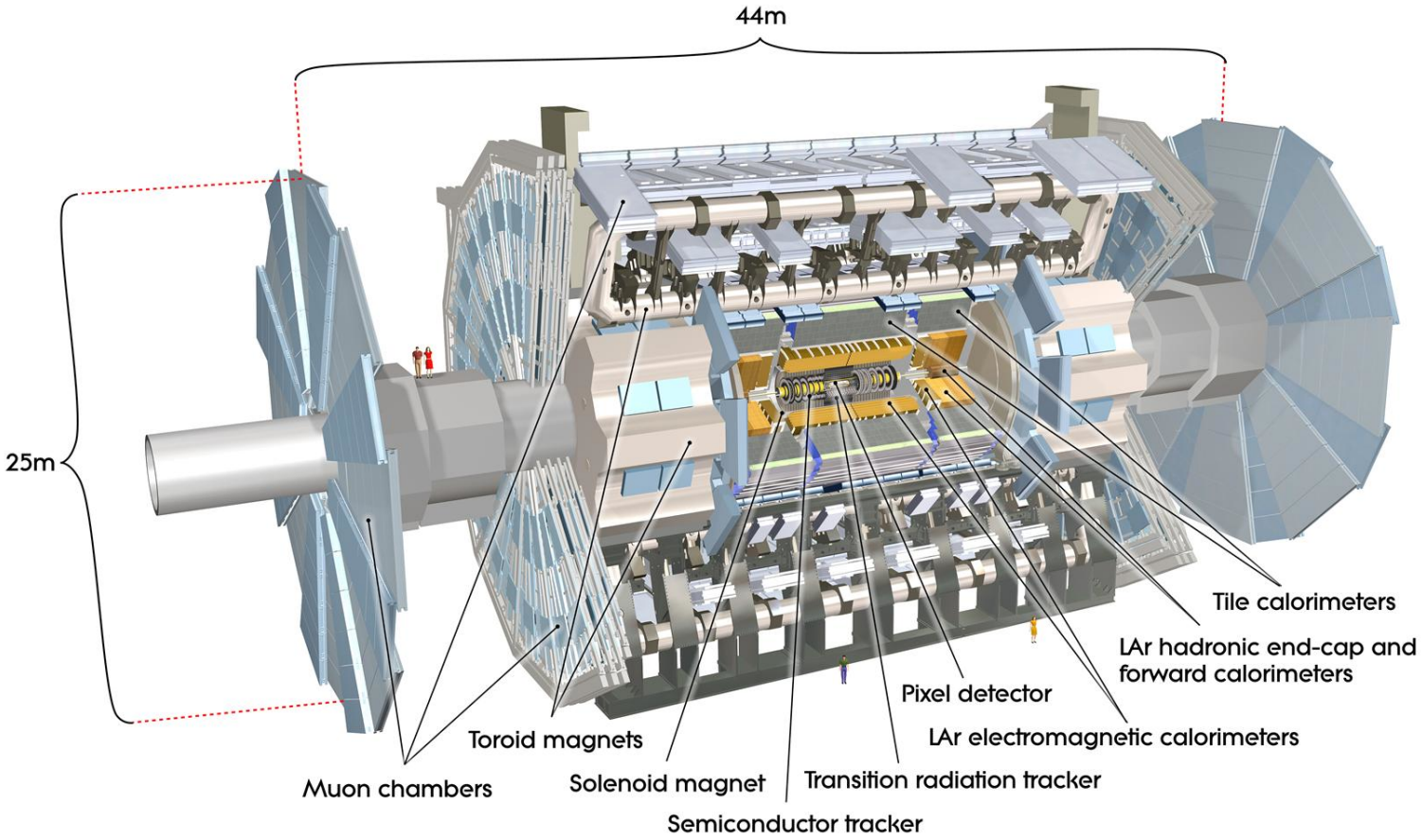


Figure 3.2 *The ATLAS detector [71] and the main subdetector systems. The Inner Detector is enclosed by the solenoid magnet and consists of the Pixel Detector, and the SCT and TRT trackers. The calorimeter system consists of the LAr electromagnetic calorimeter, and the hadronic tile, LAr end-cap and forward calorimeters. ATLAS measures 44 m in length and 25 m in diameter.*

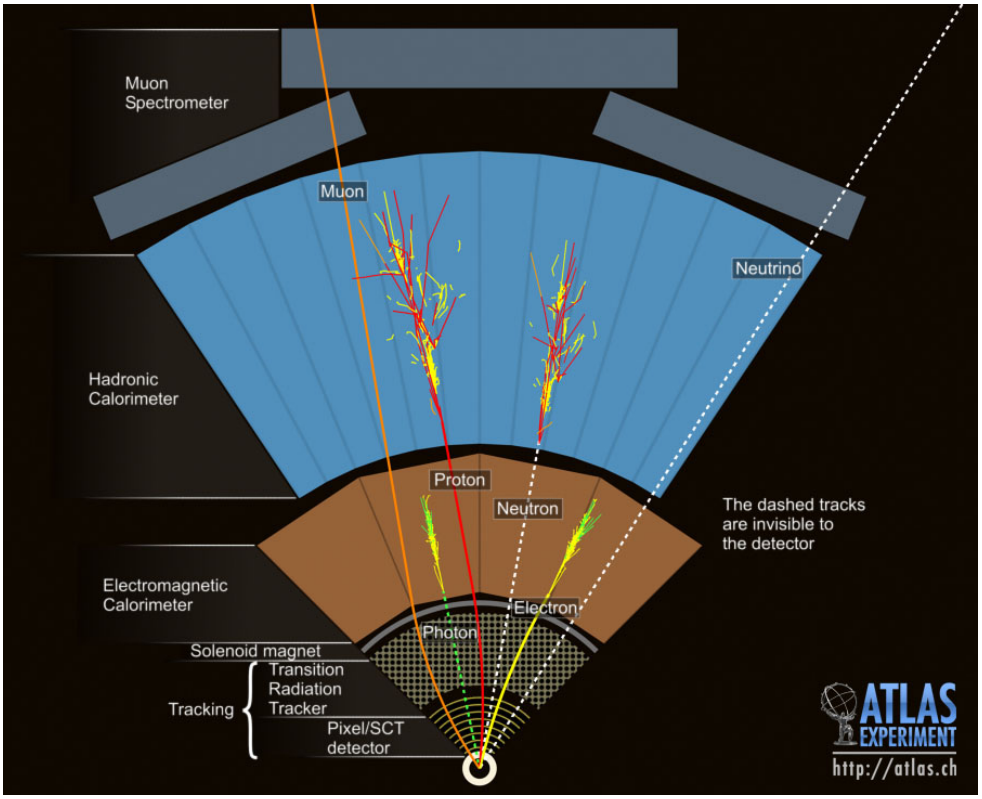


Figure 3.3 *Schematic of particle interactions in the detector. Electrons leave a track in the Inner Detector and shower in the electromagnetic calorimeter. Photons shower in the electromagnetic calorimeter, but leave no tracks in the Inner Detector. Protons leave tracks and showers in the hadronic calorimeter, whilst neutrons are only detected in the calorimeter. Muons move through all the subdetectors before reaching the Muon Spectrometer. Neutrinos are not detected directly. ATLAS Experiment © 2013 CERN.*

a measure of the polar angle. Rapidity is defined as $y = \frac{1}{2}\ln\frac{E+p_z}{E-p_z}$, where E is energy and p_z is the z component of the momentum. Pseudorapidity is defined as $\eta = -\ln(\tan\frac{\theta}{2})$, where vectors perpendicular to the beam axis have $\eta = 0$, increasing as they become more parallel with the beam.

The distance between two particles in the $\eta - \phi$ plane is often used. This is given by:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \tag{3.2}$$

3.3.3 Inner Detector

The Inner Detector [10] is a tracking detector that measures the paths taken by charged particles; the curvature of these trajectories allows for the calculation

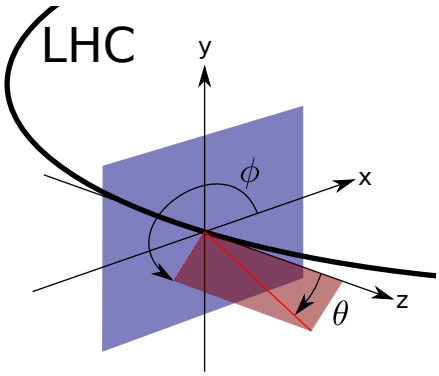


Figure 3.4 *The ATLAS coordinate system relative to the LHC beam [72]. The $x-y$ plane is shown in blue. A projection in the $x-y$ plane along the z-axis is shown in red. The polar angle (θ) and azimuthal angle (ϕ) are also indicated relative to the axes.*

of the charge and momentum of the particle. The ID is the closest subdetector to the interaction point, which requires a high granularity detector capable of precision tracking in order to identify short-lived particles and to reconstruct the interaction vertices. It consists of, in increasing distance from the interaction point, a silicon Pixel Detector, the Semi-Conductor Tracker (SCT), which uses silicon strips, and the Transition Radiation Tracker (TRT), as shown in Figure 3.5. The Pixel Detector is the smallest and most precise of the three. The SCT and TRT cover a much larger volume and are further away from the interaction point, and consequently have a lower granularity. These subdetectors are described in further detail below.

Pixel Detector The Pixel Detector [10] is found closest to the interaction point and is the most sensitive and precise part of the ID with about 80 million electronic readout channels (one per pixel). This provides critical tracking information required for identifying short lived particles such as b-hadrons, which decay within a few centimetres of the interaction point. Each pixel is a silicon wafer with an area of 50×400 μm^2 and 256 ± 3 μm thick[1]. Each pixel has an intrinsic resolution of 10 μm in the $R-\phi$ plane and 115 μm along z (barrel) and R (end-caps). Pixels are reverse-biased $p-n$ diodes; ionising particles passing through the silicon creates electron-hole pairs, which causes a short pulse of current, recorded as a *hit*.

The pixels are arranged into 1744 modules of 46080 pixels each, each with an area of $\approx$ 10 cm^2, with a combined coverage of 1.7 m^2. The Pixel Detector

[1]About 10% of the wafers are slightly larger at 50×600 μm^2.

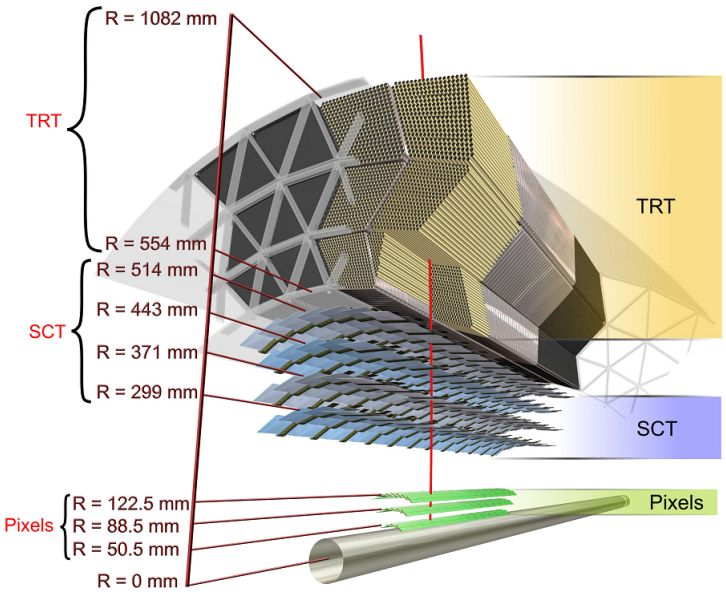


Figure 3.5 *A diagram of the Inner Detector barrel region [73]. The pixels offer the most precise measurements, found right next to the beam-pipe. The SCT and TRT extend to a much larger radius and cover a larger volume, but at a lower resolution.*

consists of three barrel layers ($|\eta| < 2$) and three pixel disks in each end-cap ($|\eta| < 2.5$). The barrel layers contain 1456 modules (for a total of almost 67 million pixels) in three cylindrical tubes 1.4 m long at radii of 50.5, 88.5 and 122.5 mm. The three disk modules in the end-cap region found at $|z| = 495$, 580 and 650 mm ensure near hermeticity. Each end-cap region contains 288 modules accommodating approximately 13 million pixels. These detectors are exposed to huge amounts of radiation which affect the hit efficiency and signal-to-background noise [74, 75]. The radiation modifies the doping concentration which can lead to type inversion and a higher required depletion voltage. The leakage currents are increased, which affects the power consumption and background noise.

An additional barrel layer has been inserted into the Pixel Detector for Run 2, called the Insertable B-Layer (IBL) [76]. This is placed at a radius of 31 mm, offering improved vertex resolution and b-jet reconstruction.

Semi-Conductor Tracker The Semi-Conductor Tracker (SCT) [10] is a silicon strip detector surrounding the Pixel Detector. The SCT operates on a similar principle to the Pixel Detector; however, long silicon strips are used in place of pixels. The longer strips allow the SCT to cover a larger area than the Pixel Detector, 63 m^2 compared with 1.7 m^2, but at a lower resolution. There are 15912

silicon micro-strip sensors each consisting of 768 active strips with a strip pitch width of 80 μm and a length of 6 cm (these are daisy chained to form 12 cm long sensors). Modules are created from two silicon strip sensors glued back-to-back at a small relative angle, allowing for a measurement of the coordinate parallel to the strip. The SCT modules have an intrinsic accuracy of 17 μm perpendicular to the strips in the $R-\phi$ plane, and 580 μm parallel to the beam in z (barrel) and R (end-cap). As with the Pixel Detector, there are similar effects on the efficiency and leakage currents due to radiation [77].

The SCT consists of four barrel layers and nine end-cap disks on each side. In the barrel region, double-sided modules are placed at radii of 299, 371, 443 and 514 mm, aligned parallel to the beam-pipe and perpendicular in ϕ. The end-cap disks are centred at $|z| = 853.8$, 934, 1091.5, 1299.9, 1399.7, 1771.4, 2115.2, 2505 and 2720.2 mm, and extend the coverage of the SCT to $|\eta| < 2.5$.

In total, there are 4088 modules (some of which are single-sided) giving about 6.2 million readout channels.

Transition Radiation Tracker The Transition Radiation Tracker (TRT) [10] is a collection of 370 000 drift chambers forming the outermost part of the ID. The drift chambers, or *straws*, are coated[2], carbon-fibre reinforced, Kapton tubes 4 mm in diameter and 144 cm in length (39 cm in the end-cap region) with a gold-plated tungsten wire of diameter 31 μm running down the centre, and filled with a gas mixture of Xenon (70%), Carbon Dioxide (27%) and Oxygen (3%). Straws are grouped into modules of 52544 straws in the barrel region, and modules of 122880 straws in the end-cap regions.

The straws are aligned parallel to the beam axis in the barrel region and radially in the end-caps. There are 73 layers in the barrel region, each consisting of 96 modules, at radii of 554-1082 mm, and $|z| = 0 - 780$ mm. There are 160 layers in the end-caps, each consisting of 20 modules, at radii of $615 - 1106$ mm and $|z| = 827 - 2744$ mm. Together these modules provide 4.2×10^5 readout channels. The straws do not provide tracking information parallel to themselves.

When a charged particle traverses a straw it ionises the gas, freeing electrons, which drift to the tungsten wire, causing an electric current which is recorded. The time taken for the electrons to drift to the nearest wire core is used to

[2]The tubes are coated with a $5 - 6$ μm graphite-Kapton surface layer, which protects an aluminium cathode of thickness 0.2 μm.

calculate the impact parameter of the charged particle with respect to the anode. Each straw has an intrinsic accuracy of 130 μm (compared with the straw diameter of 4 mm), but on average a charged particle with $p_{\mathrm{T}} > 0.5$ GeV and $|\eta| < 2$ will traverse 35 straws (22 in the region $0.8 < |\eta| < 2.0$) [78]. The combined accuracy allows the TRT to provide measurements of the transverse momentum of charged tracks to a high degree of precision of approximately 50 μm [10].

The layers of straws are interleaved with polypropylene fibres (barrel region) and foils (end-caps) so that X-ray transition radiation (TR) is emitted as charged particles move between the media of differing dielectric constants, which is subsequently absorbed by the gas in the straw tube. The more relativistic a particle is the more TR will be emitted; lighter particles, like electrons, will produce stronger signals than hadrons, for example. The amount of TR emitted coupled with momentum measurements is used for discrimination between electrons and pions.

Inner Solenoid The ID is surrounded by a 2 T solenoid magnet used to bend the trajectories of charged particles moving through the detector, so that momentum and charge can be deduced using the Lorentz force. However, particles that are not travelling with high momentum (< 400 MeV) escape detection because their paths are so tightly curved by the magnetic field that they do not move far enough away from the interaction point in the radial direction to be detected. The magnet is 5.3 m long, 2.4 m in diameter, 4.5 cm thick and weighs 5 metric tonnes.

3.3.4 Calorimeters

The calorimeter system is located outside the solenoidal magnet that surrounds the ID (Figure 3.6) [10]. The inner layer, the electromagnetic calorimeter (ECal), is used for the measurement of the energy of photons and electrons. The outer layer, the hadronic calorimeter (HCal), is used to find the direction and energy of hadrons, which are reconstructed as jets.

Both the ECal and HCal are sampling calorimeters, which use alternating layers of an absorbing material and an active sampling material. Particle showers are induced by high energy particles coming into contact with the dense absorbing material, which results in multiple lower energy particles interacting with the

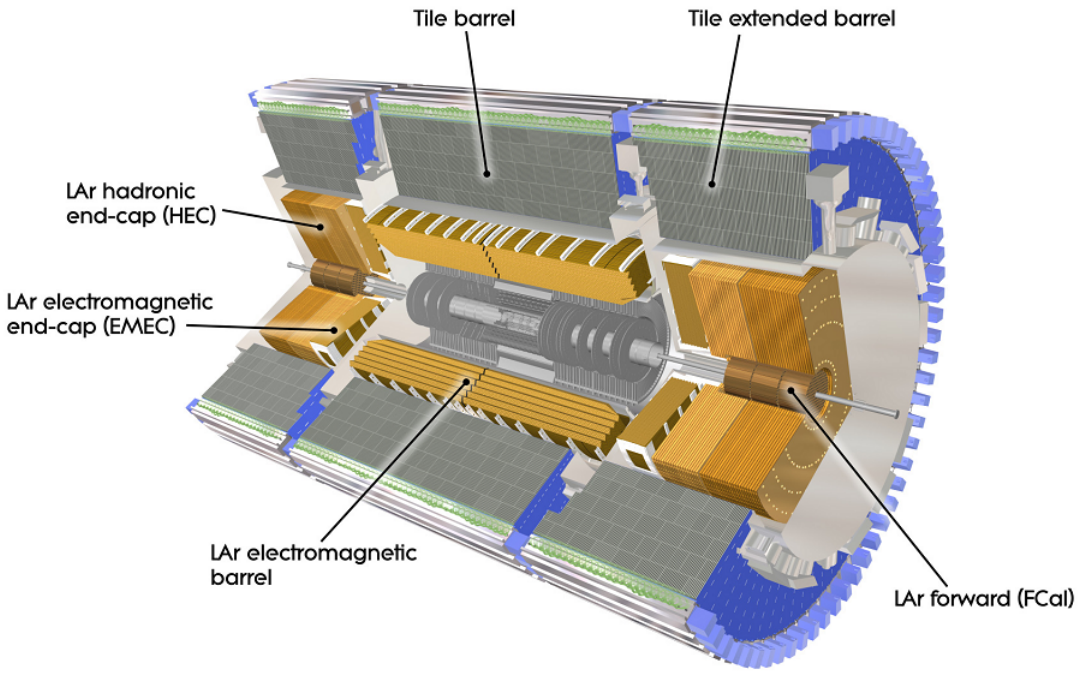


Figure 3.6 *The ATLAS calorimeter system [79]. The electromagnetic calorimeter is composed of the EMEC and the LAr electromagnetic barrel. The hadronic calorimeter uses scintillating tiles in the barrel region, and LAr in the end-cap (HEC) and forward regions (FCal).*

sampling material. This produces a signal proportional to the initial energy in the form of photons or an electric current. The energy of the interacting particle can be calculated from these measurements in consideration with the length scale of energy losses for the particles in the material. Electrons (or positrons) lose $1/e$ of their total energy in a single radiation length, X_0, through bremsstrahlung, whilst photons have 7/9 chance of producing an $e^{\pm}$ pair [80]. Apart from neutral pions, which create electromagnetic showers, hadrons will lose energy through inelastic interactions, which is parameterised as the mean free path, λ, and gives the characteristic scale of the hadronic showers, where $\lambda \sim 35A^{1/3}$ gm^2 and A is the atomic weight of the absorption material. Hadronic showers can also be caused by other nuclear decays.

Calorimeters are also used to infer the presence of particles such as muons and neutrinos. Muons can leave an ionisation signal that, if the muon is solitary, can be identified as having come from the muon. This track can then be followed into the Muon Spectrometer to check that it was indeed a muon. Neutrinos can also be inferred with the calorimeter by considering momentum conservation of particles within the calorimeter. This requires high precision, so there must be little leakage out of the calorimeters.

The calorimeters are designed to stop as many particles as possible, except for muons and neutrinos. The two parameters X_0 and λ are chosen to achieve this goal, with the ECal having a total thickness of more than $22X_0$, where the liquid argon sampling material has $X_0 = 14$ cm, and the lead absorber $X_0 = 0.5$ cm. The entire calorimeter has a total thickness of approximately 10λ. Coverage of up to $|\eta| < 4.9$ ensures that there is near hermetic coverage within the detector. The electromagnetic and hadronic calorimeters are described in more detail below.

Electromagnetic Calorimeter The electromagnetic calorimeter [10] is a liquid argon (LAr) sampling calorimeter with a lead absorber. It covers the range $|\eta| < 3.2$, with the barrel region covering the range $|\eta| < 1.475$ and two end-caps covering $1.375 < |\eta| < 2.5$ and $2.5 < |\eta| < 3.2$. The calorimeter has an accordion geometry, which provides full azimuthal coverage.

The barrel region is divided into three longitudinal layers of radius between 2.8 and 4 m, and a length of 6.4 m, which are split in half at $z = 0$ with a gap of 6 mm. The first two layers have the highest resolution, finely segmented into cells of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ ($\Delta\eta \times \Delta\phi = 0.025/8 \times 0.1$ for $|\eta| < 1.4$ in the first layer), with a resolution of $\Delta\eta = 0.05$ in the third layer. The first two

layers are designed to separate charged and neutral pions and the third for high energy electrons and photons producing large showers. The end-caps are divided into two co-axial wheels 0.63 m thick, covering a radius of 0.330 m to 2.098 m, with a 3 mm gap between the wheels at $|\eta| = 2.5$. Granularity reaches up to $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ for $2.5 < |\eta| < 3.2$ in the end-cap region. There are 101760 readout channels in the barrel region and in the end-caps a further 62208.

To correct for energy lost through the calorimeter a presampler is included before the absorption plates in the $|\eta| < 1.8$ region, with a granularity of $\Delta\eta \times \Delta\phi = 0.025 \times 0.1$. The presampler adds another 7808 readout channels in the barrel region and 1536 readout channels in the end-caps.

The ECal should fully contain EM showers to avoid leakage into the hadronic calorimeters, and as such it a thickness of at least $22X_0$ in the barrel region, and $24X_0$ in the end-caps. However, at the transition between the barrel and end-caps, $1.37 < |\eta| < 1.52$, the ECal measurements are not used as there is a large amount of material in front of the calorimeter (about $7X_0$).

Hadronic Calorimeter The hadronic calorimeter [10] is divided into three regions covering up to $|\eta| < 4.9$. A tile calorimeter in the barrel region covers $|\eta| < 1.7$, hadronic end-caps (HEC) cover $1.5 < |\eta| < 3.2$, and forward calorimeters (FCal) cover $3.1 < |\eta| < 4.9$.

The barrel region consists of a central barrel 5.8 m long covering $|\eta| < 1.0$, and two extended barrels 2.6 m long covering $0.8 < |\eta| < 1.7$. These barrels are divided into three layers, extending from a radius of 2.28 m to 4.23 m corresponding to a mean free path of $\approx 7.4\lambda$ (with a maximum of 9.2λ at $\eta = 0$ at the outer edge of the tile region). The tile calorimeter used in the barrel sections uses a steel absorber and plastic scintillating tiles, which produce photons that are sent along fibre optic cables and measured by photomultiplier tubes. Tiles are staggered around the barrel between steel absorption sections in 64 wedge-shaped modules of size $\Delta\phi \approx 0.1$. Fibre optic cables are arranged into cells of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ in the first two layers, with a granularity of $\Delta\eta = 0.2$ in the third layer. Over 500000 scintillating tiles and fibre optic cables are used.

In the HEC, liquid argon is used as the sampling material, and copper as the absorption material. The HEC consists of two wheels of thickness 0.8 m and 1.0 m with a maximum radius 2.03 m on either side of the detector. The front wheels have 24 copper plates of 25 mm thickness and a thinner front plate, and

the first nine have an inner radius of 0.372 m, and the rest 0.475 m. The back wheels have 16 copper plates of thickness 50 mm. In both wheels, the copper plates are separated by 8.5 mm. Each of the two wheels on either end consists of 32 wedge-shaped modules of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ (0.2×0.2 in the $|\eta| \geq 2.6$). In total, the HEC has 5632 readout channels.

The FCal uses liquid argon as the sampling material and copper and tungsten as the absorption materials. It is divided into three layers, the first of which is used for electromagnetic measurements with a copper absorption material and radiation length of $27.6X_0$. The second and third layers are used for hadronic measurements and have tungsten absorbers, with $\lambda = 3.6$. Layers are each 0.45 m thick with a radius of 0.455 m and are found at $|z| = 4.7$ m. The FCal has a total of 1762 readout channels.

3.3.5 Muon Spectrometer

The Muon Spectrometer (MS) [10] is the outermost subsystem, consisting of three air-core superconducting toroidal magnet systems (one in the barrel region and two end-caps) and four types of tracking chambers covering a region of up to $|\eta| < 2.7$. Muons have a relatively long lifetime, interact weakly, and are, in general, produced with relativistic momentum, thus allowing them to pass through all inner layers of the detector.

The MS extends from a radius of 4.25 m to 11 m. Trajectories of the muons are curved in the $R - z$ plane by the large barrel toroid magnet (0.5 T) within the range $|\eta| < 1.4$, and by two smaller end-cap magnets (1 T) in the range $1.6 < |\eta| < 2.7$. The MS is designed to measure the muon p_{T} to within 3 GeV for a 100 GeV muon, and to within 100 GeV for a 1 TeV muon, which corresponds to a tracking resolution of ≤ 50 μm in the z direction.

The four types of detectors used in the MS cover different radii and pseudorapidities. These can be divided into two distinct types: tracking chambers that offer high precision tracking, and trigger chambers that have a fast response for triggering and association of muons with a certain bunch crossing. Monitored Drift Tubes Chambers (MDT) and Cathode Strip Chambers (CSC) are used to provide high precision tracking and momentum measurements. Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC) are less precise, but have a

much faster readout[3]. The components are aligned with a precision better than 40 μm using around 12000 optical sensors to measure the component positions. The magnetic field is monitored with 1800 Hall probes.

There are three concentric barrel layers consisting of MDTs and RPCs with radii of approximately 5 m, 7.5 m and 10 m. The precision measurements in the barrel region are performed by the MDTs that cover $|\eta| < 2.0$ in the first layer, and up to $|\eta| < 2.7$ in the outer two layers. The RPCs are attached to the MDTs, covering $|\eta| < 1.05$. In the end-caps, three-layered large wheels of CSCs and TGCs are found at distances of $|z| = 7.4$, 10.8, 14 and 21.5 m. CSCs are used for precision measurements in the innermost layer in a region of $2 < |\eta| < 2.7$ and TGCs cover the region $1.05 < |\eta| < 2.4$.

The MDTs are chambers containing drift tubes 3 cm in diameter, between 0.85 and 6.5 m long, and filled with a combination of Ar-CH_4-N_2 gas. These work on a similar principle to the straw detectors in the ID. They have a single gold-plated W-Re wire, held at a voltage of 3200 V, running through the centre of the tube. The MDTs are positioned perpendicular to the beam-axis, measuring the curvature within the $R-z$ plane with a precision of 35 μm per chamber, or 80 μm per tube. MDTs take up to 20 measurements per track, providing an extremely precise momentum measurement. There are about 350000 tubes contained within 1152 MDT chambers.

The CSCs are multi-wire proportional drift chambers, filled with a gas mixture of Ar-CO_2-CF_4. Each chamber contains interleaving cathode strips placed parallel and perpendicular to the anode multi-wire layers. The parallel cathode strips provide precision measurements in the bending plane with a resolution of 40 μm, whilst the perpendicular cathode strips give measurements in ϕ with a resolution of 5 mm. There are 32 chambers in total, providing approximately 60000 readout channels.

The TGCs operate similarly to the CSCs, with slightly different dimensions, smaller gaps between electrodes for faster readout, but they only provide measurements of the ϕ coordinate. The TGCs have a resolution of 2-6 mm in R and 3-7 mm in ϕ. There are 3588 chambers with a readout of 318000 channels.

The RPCs are wire-free, gaseous detectors containing $C_2H_2F_4$ and C_4H_{10}. Two parallel, highly resistive plates are held at a constant high voltage, with metallic

[3]The MDT has a response time of 700 ns, the CSC of 50 ns, the RPC of 1.5 ns and the TGC of 4 ns.

strips attached to their outer faces. The metallic strips are placed perpendicular to each other to provide measurements in the bending plane and in the ϕ coordinate. They are attached to the MDTs, providing ϕ measurements that complement the MDT measurements in the bending plane. The RPCs have a resolution of 10 mm in both the z and ϕ coordinates. The 606 RPC chambers have a combined readout of about 370000 channels.

Outer Magnets Each of the toroid magnets consists of eight coils, arranged symmetrically in the azimuthal direction, as illustrated in Figure 3.7. The outer air-core toroidal magnet provides a 0.5 T magnetic field within the barrel region of the MS [10]. Two smaller magnets provide a 1 T magnetic field in the end-cap regions. These provide a curvature of the muon trajectory for tracking in the $R - z$ plane. The toroidal magnet is 25 m long with an inner (outer) radius of 9.4 m (20.1 m). The end-cap magnets are 5 m in length with an inner (outer) radius of 1.65 m (10.7 m). These magnets are superconducting, operating at 4.7 K.

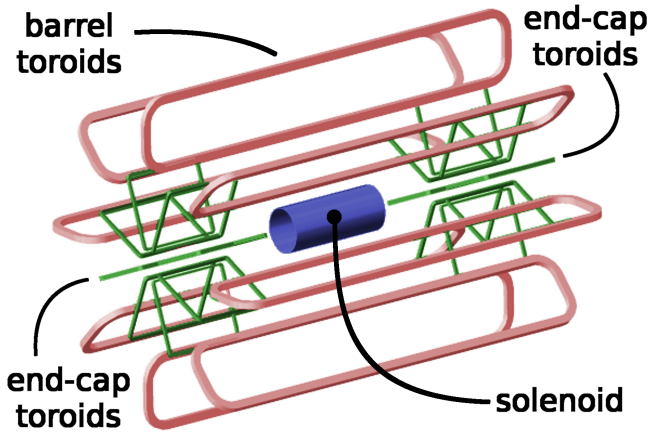


Figure 3.7 *Illustration of the ATLAS magnet system [81]. The barrel region toroid magnet is shown in red and the two end-cap toroid magnets are shown in green. The inner solenoid is shown in blue, which is parallel to the beam-pipe.*

3.4 Trigger System

The LHC has a design collision rate of 40 MHz (during Run 1 the increased bunch-spacing of 50 ns reduced the rate to 20 MHz), and with each event on the order of 1.5 MB this corresponds to an immense amount of data [10]. It

is impractical, and not necessary, to store all of this data, which would require a bandwidth of 64 TB/s. The trigger system is used to reduce the event rate in real-time to a more manageable level on the order of a few hundred MB/s to GB/s. A combination of hardware and software-based triggers are used to perform decisions at high speed.

During Run 1 ATLAS used a three-tier trigger system: the hardware-based Level 1 trigger (L1), software-based Level 2 trigger (L2) and the software-based Event Filter (EF). Together L2 and the EF are known as the High Level Trigger (HLT). For Run 2 the previously distinct elements of the HLT are combined into a single trigger. Each level in the trigger significantly reduces the data rate, with only the events accepted by the HLT being stored for offline processing. The L1 trigger is described below, followed by the L2 and EF used in Run 1, and the HLT for Run 2.

Level 1 The L1 trigger is built from custom electronics and is physically within the ATLAS detector. It provides fast, coarse identification of high momentum physics objects like leptons, photons, jets and large missing transverse energy. The p_{T} thresholds for the triggers change as the running conditions of the LHC change. A trigger 'menu' contains the selection criteria for which events are to be kept for further processing. The calorimeters and muon trigger chambers (there is no tracking information used) provide fast readouts with a reduced granularity, which are compared with the selections in the trigger menu. In addition to this, the trigger also identifies Regions-of-Interest (RoI) in η and ϕ, which are used as seeds for the L2 trigger. The L1 trigger was initially designed to reduce the initial rate from 40 MHz down to 70 kHz within 2.5 μs per event. For Run 2 this has been increased to 100 kHz [82].

Level 2 Trigger (Run 1) The software-based L2 trigger further analyses the RoIs identified by the L1 trigger at a higher resolution. Full detector granularity, including tracking information, is used within the RoI to identify events of interest. The RoIs only represent approximately 2% of the total event data. The L2 trigger was designed to reduce the output from the L1 trigger to 3.5 kHz within 40 ms per event.

Event Filter (Run 1) The Event Filter has access to the full event data with reconstruction and analysis of events done in much more detail, almost to the

level of offline analyses. The Event Filter was originally designed to reduce the output from the L2 trigger to 200 Hz within 4 s per event, although the actual achieved rate was higher than 200 Hz during data taking. In 2011, it ranged from 200 to over 300 Hz, whilst in 2012 it was between 300 and 600 Hz [83].

High Level Trigger (Run 2) The HLT in Run 1 was composed of the L2 and EF running in separate computing farms. These have now been merged together and run a single unified process in the same software. With the new design the output rate has been increased to 1 kHz [82]. Currently the HLT uses CPU-based software; a study was done into the feasibility of instead using Graphical Processing Units (GPUs) with many hundreds or thousands of processing cores. This is shown in Appendix A.

3.5 Computational Facilities

The data recorded by ATLAS after trigger selection is on the order of a few hundred MB/s [10], which still requires a large amount of storage space and computing power in order to be processed and analysed. ATLAS uses a multi-tiered distributed computing network that spans multiple countries and continents [84]. This forms part of the much larger LHC Computing Grid, which is used by all the LHC experiments at CERN. Not only is this used for analysis of the data, it is also used for creating Monte Carlo event simulations, which requires considerable computing resources.

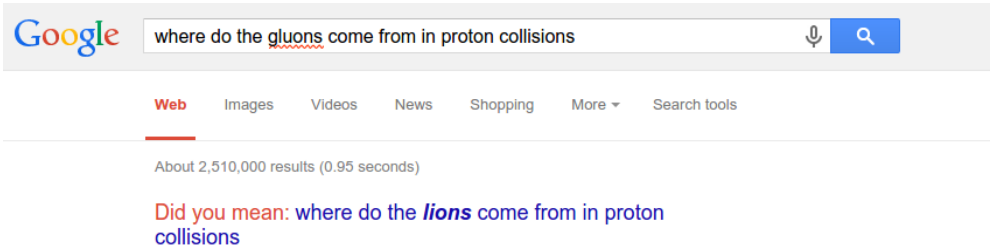
A copy of all data from the LHC is kept at Tier-0, the CERN Data Centre. The raw data from the ATLAS HLT is transferred directly to this site, where initial reconstruction is performed, and the data moved to other file formats and copied to other Grid sites.

Tier-1 consists of high performance computing centres situated around the world. These are used for storing a large proportion of raw data, reprocessing the raw data, Monte Carlo production and storage, and offering large scale analysis capabilities. This provides redundancy and allows for faster access to data.

Tier-2 and Tier-3 are smaller computing centres found at many universities. Tier-2 sites are used for Monte Carlo production and analyses. Tier-3 sites are used by small scientific groups, offering access to the resources of the upstream Tiers.

Chapter 4

Particle Identification and Reconstruction



** No lions have been observed at the LHC as of 2016.*

4.1 Introduction

Many types of particles are produced at the LHC in high energy proton-proton (pp) collisions. These particles deposit energy in the detector as they move outwards from the collision point. The measurements of these energy deposits are used to define *physics objects* using reconstruction and identification algorithms. These can refer to individual particles (like photons), collections of particles (like jets), or the missing transverse energy (any energy or momentum imbalanced in an event). The reconstruction and identification of physics objects is a complex task that has many complicated aspects. High energy particles produce radiation, and decay in the detector, leaving a collection of energy deposits. These issues can be addressed by employing Monte Carlo simulations to assess the performance of different reconstruction and identification algorithms, and to model how these objects change between processes.

This chapter describes how the Monte Carlo simulation is performed, and the reconstruction and identification algorithms for the physics objects used in the subsequent chapters.

4.2 Monte Carlo Simulation

During a proton-proton collision, the constituent quarks and gluons, or *partons*, within each proton are involved in parton-parton interactions. These interactions can be separated into the hard scattering process and soft emissions, which can be treated independently according to the factorisation theorem [85]. Perturbation theory is used to calculate properties of the hard scatter, but produces divergent contributions when considering soft emissions. These effects must be considered separately when simulating the interactions.

Monte Carlo simulations are generated in a series of successive steps, starting with an initial state, incoming partons that participate in the hard scatter, followed by the creation of stable final state particles that interact with the detector.

The event generation process is described below.

Parton Distribution Function The parton distribution function (PDF), given by $F(x, Q^2)$, provides the probability of finding a given flavour of parton with

a fraction x of the momentum of the proton at the energy scale of the hard scatter, Q, where $Q^2 \equiv -q^2$ is the momentum transfer in the interaction. Due to the non-perturbative nature of QCD, these are not predicted and must be parameterised in terms of fits to experimental data, including data from the HERA electron-positron collider at DESY and the Tevatron and LHC hadron colliders. They are dependent on the energy scale and the quark model being used. The soft non-perturbative emissions in the proton-proton collisions can be included in the parton distribution function to contain their divergences; the unphysical factorisation scale, μ_F, sets a cut-off below which emissions are included in the parton distribution function, and above which they are considered part of the hard scatter. The scaling of a parton distribution function with μ_F is described by the DGLAP equations [86–88]. Parton distribution functions are available from several fitting collaborations such as CTEQ [89], HERA [90, 91], MSTW [92], and NNPDF [93].

Hard scatter process The hard scatter process of coloured quarks and gluons from the proton-proton collision are described using Matrix Element calculations at fixed order in perturbative QCD. The specific process is selected in the hard scatter, for example, the associated production of the W boson and Higgs boson.

Parton showering High energy coloured partons from the initial hard scattering process can be treated using perturbative QCD. These partons fragment into lower energy objects by emitting QCD radiation. At leading order, gluons can radiate one or two gluons, or a quark-antiquark pair, and quarks can radiate gluons, which themselves can then radiate further. This process continues until the partons reach the factorisation scale, $Q = \mu_F$, where QCD perturbation theory breaks down. The ordering of the showering is not strictly defined and is done differently in some MC generators.

Hadronisation When partons from the showering have reached a low enough Q^2 they hadronise due to colour confinement. The hadrons that are produced in this step may be unstable. The decay of these unstable particles must be simulated according to known branching ratios. As an example, in the Lund string model of hadronisation [94] strings are stretched between coloured partners until the potential energy is large enough to produce a $q\bar{q}$ pair. This

continues until the point that there is no longer sufficient energy to create these pairs. In the cluster model [95] all gluons are split into $q\bar{q}$ pairs. After splitting the gluons, the colourless clusters are identified, which are then decayed to hadrons.

Initial and final state radiation The incoming and outgoing partons of the scattering process both emit radiation. *Final state radiation* (*FSR*) from the outgoing partons, in the form of soft and collinear emissions, provides additional showers in the event, which subsequently results in hadronisation. *Initial state radiation* (*ISR*) from the incoming partons is provided similarly by a parton shower.

Multiple parton interactions (MPI) Additional soft partons from the protons, not involved in the hard scatter, can interact, resulting in additional particles in the final state. This is known as the *underlying event*, and is not well described by perturbative methods. *Multiple partonic interaction* models [96], tuned to LHC data, are used to simulate this underlying event activity.

In the following chapters, the event generation is done with a number of general purpose generators. They use different hadronisation models, parton showering models and treat MPI differently. Pythia 6 [97] and 8 [98], Herwig [99] and Herwig++ [100] all use LO matrix elements to describe parton scattering with two initial and two final state partons. Sherpa [101] is used to model events with more final state partons, describing the perturbative part of the interaction, at up to next-to-leading-order (NLO) accuracy. In this thesis, a number of specialist generators are used, for example, Powheg [102–104], AcerMC [105], Madgraph [106] and MCFM [107]. Powheg provides parton showers with NLO accuracy. Madgraph and MCFM are used in conjunction with general purpose generators to provide better estimates (up to NLO) of the cross section, the matrix elements calculations and the showering process. AcerMC provides a library of matrix elements for common Standard Model background processes.

The interactions of the outgoing particles with the detector are simulated by passing each generated event through a full simulation of the ATLAS detector based on Geant4 [108], which allows for the detector response to be calibrated and efficiencies to be estimated. The simulation estimates how much energy is deposited in different parts of the detector along a particle's trajectory. In particular, modelling the calorimeter response is a complex and time-consuming

task. In some cases, ATLFASTII [109] is used instead of GEANT4, where the calorimeter response is parameterised.

4.3 Tracks

Charged particles traverse the Inner Detector leaving a chain of hits in the pixel, SCT and TRT detectors. These hits are combined to form *tracks* that show the trajectory of the particle as it moves through the ID [110, 111].

Track finding is performed using two complementary strategies; the *inside-out*, and *outside-in* algorithms [112, 113]. In the inside-out algorithm, track seeds are identified in the first three layers of the Pixel Detector and the first layer of the SCT. These seeds are extended through the SCT to form track candidates and an initial track fit is performed with a Kalman filter. At this step, outliers are removed and fake track candidates are rejected by track quality cuts. The track candidates are extended into the TRT, after which the track candidate is again fitted with input from the Pixel Detector, SCT and TRT. The outside-in algorithm is used to improve the tracking of particles that decay at a displaced vertex. Unused track segments in the TRT are extrapolated into the SCT and Pixel Detector.

Tracks are defined in terms of five parameters: the closest approach of the track to the reconstructed primary vertex given by the transverse and longitudinal impact parameters, d_0 and z_0, respectively, and the track momentum expressed in terms of the azimuthal angle ϕ, the polar angle θ, and the charge multiplied by the inverse transverse momentum q/p_{T}. Quality criteria require reconstructed tracks to have:

- $p_{\mathrm{T}} > 1$ GeV,
- impact parameters of $|d_0| < 2$ mm and $|(z_0 - z_v)\sin\theta| < 10$ mm (where z_v is the z coordinate of the primary vertex). For b jet identification these are tightened to $|d_0| < 1$ mm, $|(z_0 - z_v)\sin\theta| < 1.5$ mm for tracks not associated with the secondary vertex,
- at least seven pixel or SCT hits,
- a track χ^2/DOF of < 5.

Tracks are essential in the identification of charged particles, finding the location of primary vertices (see Section 4.4) from collisions, and secondary or tertiary vertices from particle decays.

4.4 Primary Vertex

Vertices are points where the trajectories of at least two reconstructed tracks point to the same origin. The *primary vertex* refers to the location of the hard scatter process. The primary vertex with the largest scalar sum $\sum p_{\mathrm{T}}^2$, and at least three associated tracks, is taken as the vertex of the hard pp collision. *Secondary vertices* occur some distance away from the primary vertex in the transverse and beam direction, where particles that were produced in the initial collision decay.

Pile-up

Aside from the hard process of interest, there can be multiple further soft interactions (low q^2) during each bunch crossing, known as *pile-up*. The number of these interactions is Poisson distributed with a mean value of $\langle\mu\rangle$. *In-time pile-up* refers to the case where these soft interactions are from the same bunch crossing. Due to the inherent properties of the detector subsystems, there can be overlap between separate bunch crossings, or *out-of-time pile-up*. Some of the detector subsystems have a read-out integration time which is longer than the bunch spacing, for example, recording activity in LAr calorimeter components takes on the order of 500 ns, compared with the bunch spacing of 25-50 ns.

As a result of pile-up, a number of primary vertices, N_{PV}, can be identified in an event, which gives a good indication of the level of pile-up.

The average pile-up during 2012 was $\langle\mu\rangle = 20.7$ [114]. In 2015, for Run 2, there was an average pile-up of $\langle\mu\rangle = 13.5$ (19.6) for a 50 ns bunch spacing (25 ns) [115].

In the MC samples used for the studies in this thesis, in-time pile-up is simulated by overlaying energy deposits from low q^2 pp collisions. The out-of-time pile-up is simulated by overlaying events with a time shift, which is meant to simulate the inherent delay of the detector subsystems.

4.5 Electrons

Electron candidates [116, 117] are identified by matching a track in the Inner Detector with a cluster of cells in the EM calorimeter. The reconstruction methods for the electrons and muons described below are given in terms of the $WH \to \ell\nu b\bar{b}$ analysis in Chapter 5. Three electron selection categories are defined, referred to as *loose*, *medium* or *tight* selection. The looser selections form a subset of the tighter selections; tighter selections have higher selection purity. The selection efficiency is dependent on the E_{T} and η of the electron, increasing for all η values at higher E_{T}. The selection depends on a number of kinematic properties. A likelihood based identification method [117] combines the shower shape, hadronic leakage (ratio of energy in the transverse plane in the hadronic calorimeter to the EM calorimeter), the number of hits in the ID, the impact parameter, track matching (using cuts on $\Delta\phi$, E/p and η) and rejection of electrons matching photon conversions.

In addition to the likelihood identification, there are cuts applied on the transverse energy (E_{T}) and $|\eta|$, and track and calorimeter based isolation criteria must be satisfied. The track isolation criteria for the lepton selection requires that the scalar sum of the track momenta within a cone of radius $\Delta R = 0.2$ that are not associated with the lepton are restricted to a chosen fraction. This is done similarly in the calorimeter, considering instead energy deposits within a cone of radius $\Delta R = 0.3$ that are not associated with the lepton that are below a chosen threshold. The isolation cuts are used to reduce the number of jets that are misidentified as leptons, since jets have a wider footprint. The requirements for each of the categories are given in Table 4.1. Corrections are applied to the energy resolution, reconstruction and identification efficiencies and calorimeter isolation to account for mismodelling in MC simulations [118].

4.6 Muons

Muons pass through the ID and calorimeters, leaving minimal energy deposits, and into the Muon Spectrometer (MS). The muons are identified by the interactions measured in the MS, and, for $|\eta| < 2.5$, matched ID tracks.

Initially, the trigger chambers (see Section 3.3.5) identify activity in the MS [119]. Hits from the surrounding regions in the MDT and CSC are used to reconstruct

Property	Loose Category	Medium Category	Tight Category
E_{T} (GeV)	> 7	> 25	
$\lvert\eta\rvert$	< 2.47		
Track isolation %	< 10		< 4
Calo isolation %	-		< 4
Likelihood	Very Loose	Loose	Very Tight

Table 4.1 *The requirements for electron selection in the loose, medium and tight categories. The transverse energy, likelihood criteria and the track isolation requirements become more stringent as the selection is tightened. Tight selection introduces an isolation cut on the calorimeter energy deposits. Where there is a '-', this indicates that these criteria are not applied.*

track segments, which are combined between multiple layers of the MS to identify tracks. If ID tracks are used for muon identification, the tracks in the MS are extrapolated to the primary vertex, matched to tracks in the ID, and then combined into a single track. Isolation criteria are imposed to suppress muons from hadronic decays and hadrons that continue through the calorimeters (known as *punch through*).

Three different strategies are used:

1. *Combined*: Muons are reconstructed using both the MS and ID,
2. *Calo*: Muons with $p_{\mathrm{T}} > 20$ GeV are identified in the calorimeter and that are matched to tracks found in the ID with $|\eta| < 0.1$,
3. *Standalone*: Muons with $|\eta| > 2.5$ are identified by the MS but are not matched to the ID.

As with the electron identification, *loose*, *medium* and *tight* categories are used to identify muons [120].

The category of muon is based on kinematics of the muon candidates and the strategy used to identify the candidate. For the *loose* category, any of the three strategies above can be used to identify the muon, but for the other two categories only the *combined* strategy is used, as *standalone* and *calo* muons can be more easily satisfied by other objects, such as jets. The requirements on the p_{T}, η, track and calorimeter isolation change between categories and are shown for the combined identification strategy in Table 4.2. As for the electron, there

Property	Loose Category	Medium Category	Tight Category
p_{T} (GeV)	> 7	> 25	
$\lvert\eta\rvert$	< 2.7	< 2.5	
Track isolation %	< 10		< 4
Calo isolation %	-		< 4
$\lvert d_0\rvert$ (mm)	< 0.1		
$\lvert z_0\rvert$ (mm)	< 10		

Table 4.2 *The requirements for muon selection in the loose, medium and tight categories for the combined muon identification strategy.*

are corrections applied to the p_{T}, identification efficiency and the calorimetric isolation.

4.7 Jets

Partons can only exist in a colourless state, which is due to QCD confinement; any quarks or gluons from the fragmentation of a particle during high energy collision must hadronise. The hadronisation process produces a collimated shower in the particle detector. The energy deposits and tracks from these showers can be clustered together into narrow cones, called *jets*, to measure the momentum and energy of the original quark or gluon of origin. The jet mass is calculated by taking the difference between the sum of the square of the energy E_i and square of the momenta p_i of each of the i jet constituents (the calorimeter clusters or tracks, as described in the following text), given by the equation

$$M^2 = \left(\sum_i E_i\right)^2 - \left(\sum_i p_i\right)^2 . \tag{4.1}$$

An example of jets identified in an event in the ATLAS detector is shown in Figure 4.1.

A *jet algorithm* defines a set of rules for combining the large number of final state particles from the shower into a single object. There is no strict definition of a jet, however, but the algorithm used should satisfy a number of criteria. Two of the most important criteria being infrared and collinear safety (IRC-safety) [122]. The collinear splitting of quarks and gluons, or soft emissions, can alter the detailed measurement of the jet mass or p_{T} in IRC-unsafe algorithms,

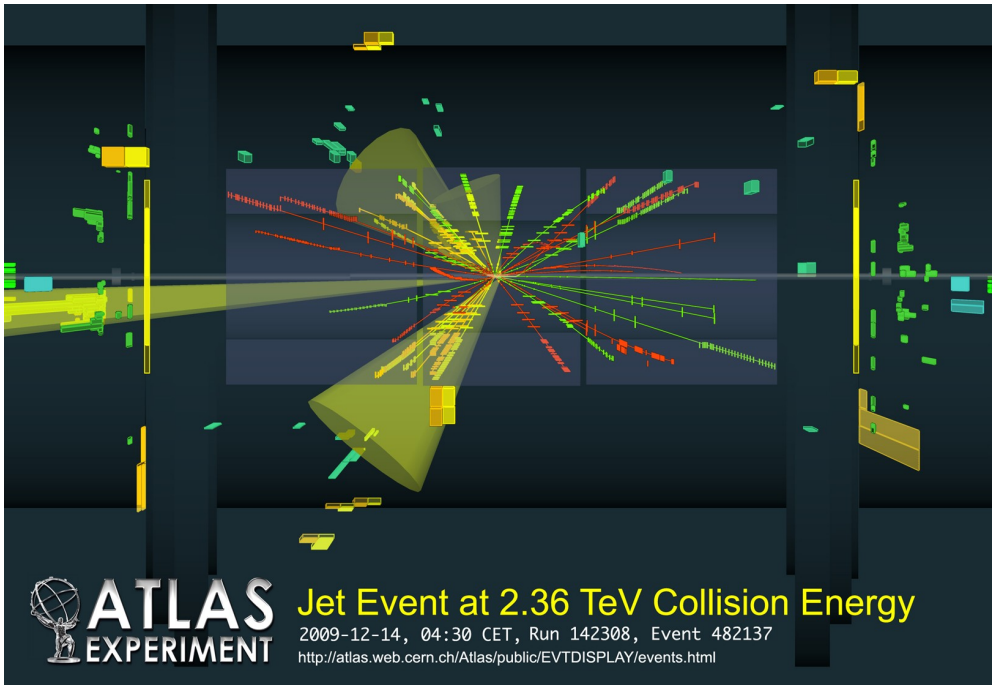


Figure 4.1 *An event display from a collision recorded by the ATLAS detector [121]. The yellow cones indicate energy deposits that have been clustered together and identified as jets.*

although the set of reconstructed jets does not change.

Jets can be reconstructed using measurements from either the calorimeter (*calorimeter jets*) or from tracks (*track jets*). In MC simulation, *truth jets* are reconstructed from stable particles, with a lifetime τ such that $\tau c > 10$ mm [20]. There are two different types of calorimeter-based jet algorithms used in general: cone based and sequential recombination algorithms. Both of the algorithms consider jets within a defining radius R, which sets the size of the jets. Typically, when identifying a one quark decay, a radius of $R = 0.4$ is used, as in Chapter 5. In Chapter 7, large-R jets are considered where $R = 1.0$ or $R = 1.2$.

4.7.1 Calorimeter Jets

The inputs for the jet algorithms come from three-dimensional clusters formed from cells in the calorimeters [123]. This proceeds through clustering of neighbouring cells in the calorimeter in a way designed to capture the particle showers, whilst suppressing noise from the electronics in the detector and from pile-up. The algorithm for clustering the cells was developed for use at DØ [124], and has been adapted for use at ATLAS [125]. Clusters are grown iteratively

around calorimeter cells with significant seed signals. The ratio of the cell signal to the average expected noise defines a cell signal significance, $\varsigma_{\text{cell}}^{\text{EM}}$. The thresholds, in terms of this ratio, for the seeding of, the growth of, and the boundary features of the clusters are defined in terms of three parameters: $\{S,\ N,\ P\}$. S is the signal-to-noise ratio, used for the primary seed threshold, N is the growth control threshold, and P is the cell filter. These take the default values at ATLAS of $S = 4$, $N = 2$, $P = 0$, derived from optimisations of the response and the relative energy resolution for charged pions in test-beam experiments [125]. A set of seed cells for the clusters with $S \geq 4$ are identified initially. Neighbouring cells of the seeds with a signal-to-noise ratio of $N \leq \varsigma_{\text{cell}}^{\text{EM}} < S$ are considered as secondary seeds and added to the cluster. Any cells neighbouring any of the seeds with a signal-to-noise ratio of $P \leq \varsigma_{\text{cell}}^{\text{EM}} < N$ are added to the cluster and define the boundary of the cluster. Once the clusters are identified, these clusters are evaluated to identify if they contain multiple effective clusters, in which case they are split up. Once all cells have been assigned to a cluster, the cluster is split if there are multiple cells that have energy greater than 500 MeV.

Sequential recombination algorithms are IRC-safe by construction. These define an algorithm-dependent distance parameter between each pair of clusters, successively combining together pairs until all clusters are deemed too far apart. The distance parameter is calculated between each pair of clusters i and j, cluster i and the beam B, according to

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2},$$

$$d_{iB} = k_{ti}^{2p},$$

where $\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$, and k_{ti} and k_{tj} are the transverse momenta of the clusters being compared. The k_t [126], anti-k_t [127] and Cambridge/Aachen (C/A) [128] algorithms each use a variation of this distance parameter: $p = 1$, $p = -1$ and $p = 0$, respectively.

Sequential recombination algorithms follow an iterative process, finding the minimum distance parameter at each step. If the minimum distance is between the two clusters, they are combined by adding their four-momenta and creating a single *proto-jet* that replaces them. If the minimum distance is between a cluster and the beam line, then the cluster is labelled as a jet and removed from the list of remaining clusters. This continues until all the clusters or proto-jets have been incorporated into jets. In the case of boosted systems, where the decay

products become sufficiently close together, multiple particles may be included in the definition of the jet. By examining the substructure of the jet, some of the proto-jets can be identified as *subjets*, which correspond (approximately) to individual decay products.

The three sequential recombination algorithms have slightly different characteristics. The k_t algorithm tends to cluster together softer particles first. This follows the QCD evolution of the jet, but can result in irregular shaped jets. The Cambridge/Aachen (C/A) algorithm clusters on the angular separation only and does not depend on energy or momentum. This can be beneficial for examining the substructure of the jet, as de-clustering the jet would reveal subjets at each step. As the jets become more collimated the de-clustering becomes less efficient at revealing the subjets. As with k_t jets, C/A jets can be irregularly shaped. The anti-k_t algorithm is used for most analyses at ATLAS as they are IRC-safe. The clustering is performed on hard particles first, giving the jets a conical shape. Additionally, the jet boundaries are not affected by soft emissions.

An illustration of jets clustered with the k_t, anti-k_t, and Cambridge/Aachen algorithms is shown in Figure 4.2. The irregular, complex shapes of the k_t and Cambridge/Aachen algorithms can be seen clearly. The anti-k_t algorithm gives conical shaped jets. In the $WH \to \ell\nu b\bar{b}$ analysis presented in Chapter 5 anti-k_t jets with a radius parameter of $R = 0.4$ are used. Both the Cambridge/Aachen and anti-k_t algorithms are used for the substructure studies of boosted W bosons in Chapter 7.

4.7.2 Jet Calibration and Corrections

Reconstructed jets based on the calorimeter clusters are calibrated such that the reconstructed energy matches the true energy on average. The correction and calibration process can be separated into a number of separate steps [130, 131].

Origin correction Jets are corrected such that they originate directly from the primary vertex, rather than the centre of the ATLAS coordinate system. This does not affect the energy of the jet, but it improves the angular resolution.

Pile-up subtraction Two corrections are applied to account for pile-up contributions. A correction is applied using a technique based on jet areas [132],

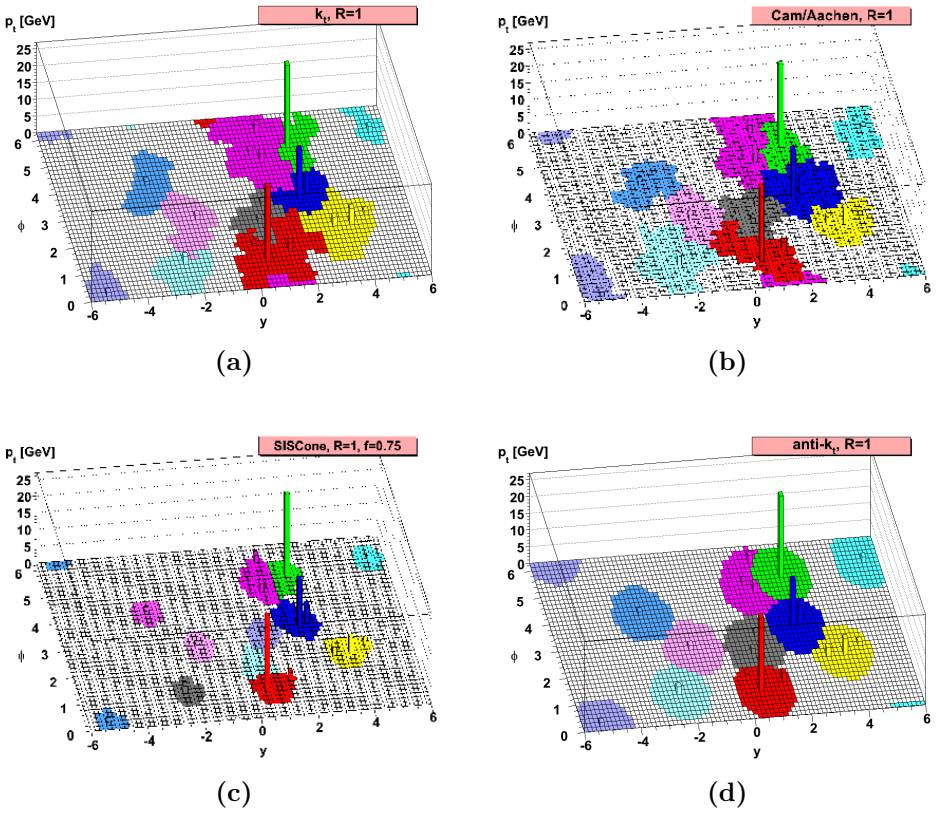


Figure 4.2 *The clustering done with the anti-k_t method, shown in (d), produces conical jets with regular boundaries except for the boundaries between two jets of approximately equal momentum [127]. Also shown here are the clustering as performed with the (a) k_t, (b) Cambridge/Aachen, and (c) the cone-based SISCone [129] algorithms.*

which adjusts the p_{T} spectrum by subtracting the average energy deposits from additional low q^2 pp interactions in the event [133]. The jet area, A, provides a measure of the jet susceptibility to soft emissions. The median of the momentum density of the jet, $\rho = \mathrm{median}(p_{\mathrm{T}}/A)$ (for k_t jets with $R = 0.4$), measures the contributions to the p_{T} due to pile-up. The corrected p_{T} is given by a $p_{\mathrm{T}}^{\mathrm{corr}} = p_{\mathrm{T}} - \rho A$, which allows for jets with large pile-up contributions to be removed by introducing a p_{T} threshold. Additional corrections are applied which depend on both in-time and out-of-time pile-up.

Energy Calibration There are two energy calibrations given here: *Electromagnetic and jet energy scale* and *local cluster weighting* calibration. These are both used for jets in later chapters.

EM+JES Electromagnetic and jet energy scale (EM+JES) calibration [130, 131] is applied to jets to correct for the differences in response for the hadronic and electromagnetic showers in the calorimeter, energy lost in dead materials and *out-of-cone effects*. Out-of-cone effects refer to particles of a jet that were not identified by the jet reconstruction. The corrections are p_{T} and η dependent, taken from simulation, with residual corrections obtained using in-situ measurements providing corrections to data. An additional calibration, Global Sequential Calibration (GSC) [130], can be used to correct for the response based on variables other than p_{T} and η, for example, the fraction of the jet energy in the third ECal layer and the first HCal layer will be correlated to energy losses in the poorly instrumented region between them.

LCW In the local cluster weighting (LCW) [130] scheme the clusters are calibrated first before being used to identify jets, with a finalisation of the calibration taking place on the jet afterwards. A number of weights are applied which account for the difference between hadronic and electromagnetic showers, energy lost outside the cluster and dead material. All weights are taken from simulation.

The jet energy resolution (JER) is measured in-situ from dijet events in data. The bisector method projects the imbalance in the dijet p_{T} in the direction bisecting the two jets [134] and in the orthogonal direction. Differences between the variance of the two components in simulations and at detector-level in data are then used to evaluate the energy resolution. The JER uncertainties are found as functions of p_{T} and η.

The effect of pile-up is further reduced by requiring a cut on the jet vertex fraction (JVF) [135]. JVF is the p_{T}-weighted fraction of associated tracks that originate from the hard scatter, where an associated track refers to a track which, when extrapolated outside the ID, matches a jet identified in the calorimeter within a cone of size ΔR.

Jets are selected for further consideration after they have passed a number of quality criteria and kinematic requirements, such as cuts on p_{T} or η.

4.7.3 Track Jets

Charged particle tracks originating from the primary vertex of the event are used to define track jets, which are robust against the effects of pile-up [130]. Tracks with $p_{\mathrm{T}} > 500$ MeV and $|\eta| < 2.5$ are considered for track jets. The transverse impact parameter must be $|d_0| < 1.5$ mm, and the longitudinal impact parameter $|z_0 \sin\theta| < 1.5$ mm. The tracks are used as input to one of the jet clustering algorithms, such as the anti-k_t algorithm, instead of calorimeter clusters as explained previously. A track jet must have a $p_{\mathrm{T}} \geq 3$ GeV, and be within $|\eta| < 2.5$.

4.8 b jet Tagging

Identifying jets originating from b quarks, referred to as b *jets*, is a challenging task as there are huge backgrounds from light jets. b quarks hadronise into b hadrons (B^+, B^0, B_s, B_d and Λ_b), which have masses of $\gtrsim 5$ GeV and relatively long lifetimes (an average of 1.568 ± 0.009 ps over all channels), decaying at a secondary vertex on the order of a few millimetres away from the beam-line [20]. The dominant decay mode of the b hadrons is to c-hadrons. The c-hadrons have a lifetime an order of magnitude smaller than b hadrons (apart from the D^+ which is on the same order) and can be identified by a tertiary vertex. A significant fraction of b hadrons decay directly into an electron or muon with a branching ratio of $b \to c\ell\nu_\ell$ (where ℓ is an electron or muon) of 10.67% per lepton flavour [20]. The decays involving muons can be identified by a high p_{T} muon close to the jet axis. An illustration of a b jet decay is shown in Figure 4.3.

A number of algorithms exist which take advantage of these properties to identify b jets [136]. The IP3D algorithm uses jets with tracks that have a large transverse impact parameter d_0. The SV1 algorithm is used to identify secondary vertices from b hadron decays, by considering all tracks associated to this point. It uses the invariant mass and energy of the tracks and the number of two-track vertices (two-track pairs forming a good vertex) as discriminating variables. In both cases, numerous tracks are required to reconstruct the primary vertex accurately, as pile-up can cause significant errors. The JetFitter algorithm reconstructs the decay chain topology, finding a common line on which all vertices (primary and secondary vertices from b hadrons and their daughter c-hadron vertices) lie. The positions of these vertices are used to reconstruct the b hadron flight path.

The MV1c [136–138] algorithm combines the IP3D, SV1 and JetFitter algorithms and the jet p_{T} and η in a neural network. The MV1c algorithm is designed to offer a high c-jet rejection. The b-tagging algorithm is calibrated using simulated $t\bar{t}$ events to find operating points which give average b-tagging algorithm efficiencies of 50% (*tight*), 70% (*medium*) and 80% (*loose*) for b jets with $p_{\mathrm{T}} > 20$ GeV.

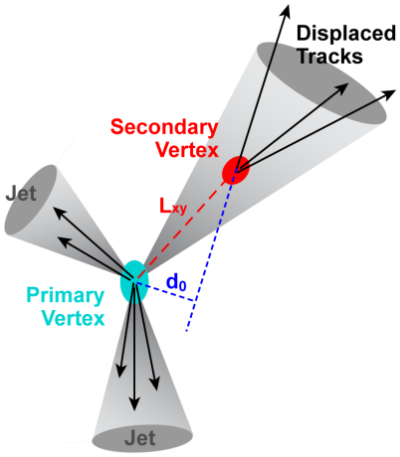


Figure 4.3 *An illustration of an event with a b jet [139]. The b hadron formed from the b quark decays at a secondary vertex a distance L_{xy} from the primary vertex. Tracks that originate from the secondary vertex have a large transverse impact parameter d_0 with respect to the primary vertex.*

Tagging efficiencies

The efficiencies for the b tagging of b jets, c jets and light jets are determined from both data and simulation. Control regions are defined using regions that are dominated by $t\bar{t}$ (for b jets) and D^* mesons (for c jets and multi-jet events for light jets). Differences in these control regions between data and MC simulations are used to derive scale factors as a function of p_{T} and η. To account for the dependence of this on the MC generator used for the simulation, additional MC-to-MC scale factors are applied.

The tight (loose) operating points of the MV1c algorithm deliver rejection factors against c jets of 26 (3) and of 1400 (30) against light jets.

4.9 Missing Transverse Energy

In many instances, not all of the decay products are detected in an event. This can be due to neutrinos in the final state of the decay (or perhaps even a new, unknown particle), energy lost in dead materials within the detector or from uncovered regions of the detector. The colliding protons have a nearly zero transverse component, and so the decay products should have almost zero net total transverse momenta. The negative vector sum of the transverse momenta from all energy clusters in the calorimeter (within $|\eta| < 4.9$) and muon detector measures the missing transverse momentum $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ [140, 141] (with magnitude $E_{\mathrm{T}}^{\mathrm{miss}}$). The magnitude and direction of $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ can be used to infer the presence of invisible particles in an event, such as neutrinos.

Corrections to energy clusters associated to reconstructed objects in the event (jets, electrons, photons and muons) are taken from the object calibrations. Muons can deposit energy in the muon calorimeters as well as the muon detector, so these calorimeter deposits are removed to avoid double counting. Pile-up can have a significant effect on $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$ due to the difficulty in associating calorimeter clusters with the primary vertex.

Missing transverse momentum can also be measured using tracks associated with the primary vertex, denoted $\mathbf{p}_{\mathrm{T}}^{\mathrm{miss}}$, with magnitude $p_{\mathrm{T}}^{\mathrm{miss}}$. This is only available within a region of $|\eta| < 2.4$, the coverage of the ID. The track-based $\mathbf{p}_{\mathrm{T}}^{\mathrm{miss}}$ is robust against pile-up, but does not include neutral particles.

4.10 Overlap Removal

A reconstructed particle can sometimes pass the requirements for more than one type of object. A precedence for identification is defined in order to avoid double counting. In general, the particle type with a higher identification efficiency will be chosen. For example, if an electron and a jet overlap within an angle of $\Delta R < X$ (where the value of X depends on the analysis), then the jet will be discarded.

Specific requirements for overlap removal for the $WH \to \ell\nu b\bar{b}$ analysis presented in this thesis is discussed in Chapter 5.

Chapter 5

$VH(\rightarrow b\bar{b})$ Analysis

B.B. has been observed through vibrating strings, although not at the LHC. Image from [142].

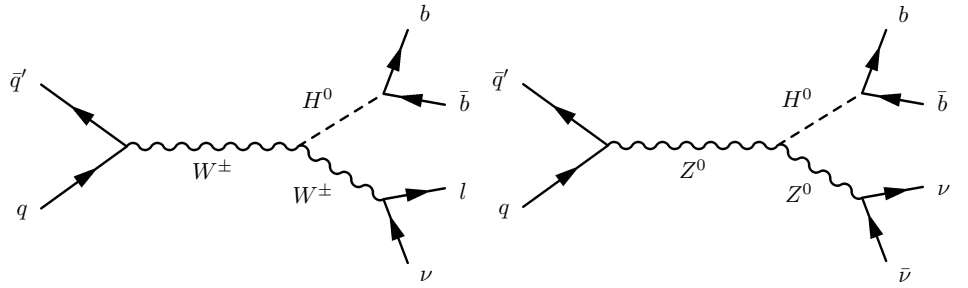
5.1 Introduction

The largest expected branching ratio of the Higgs boson is to a $b\bar{b}$ pair, however, at the LHC there is a large irreducible background in this channel. A better handle on this process is obtained by considering a Higgs boson produced in association with a W or Z boson (collectively referred to as V), where the Higgs boson decays into $b\bar{b}$ and the vector boson decays leptonically. Feynman diagrams for these three processes are shown in Figure 5.1. The final state signature of the leptonically decaying vector boson provides background suppression over QCD jet activity due to the added requirement of the muon or electron and missing transverse momentum from the neutrino, however, this comes at the price of a lower cross section compared with the gluon-gluon Higgs production mechanism. Hadronic decays of the vector bosons do not offer a clean signature in the detector and are not considered in this case.

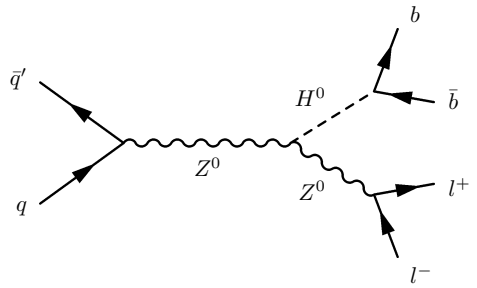
The ATLAS $VH(\to b\bar{b})$ search using Run 1 data is described in Reference [1] with two distinct methods: the nominal results are found using a multivariate analysis (MVA) which combines a number of variables in a boosted decision tree (BDT), and as a crosscheck, a cut-based analysis is performed using the dijet mass, m_{bb}. A mass range of between 110-140 GeV is considered for the signal region when considering a Higgs boson of mass 125 GeV, since the dijet mass resolution is around 11% (14 GeV). As the Higgs boson has not yet been observed in this decay channel, masses other than 125 GeV are considered for setting exclusion limits. The analysed data is taken from the ATLAS experiment accumulated during Run 1 of the LHC from proton-proton collisions at centre-of-mass energies of $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV, corresponding to integrated luminosities of 4.7 fb^{-1} and 20.3 fb^{-1}, respectively. The author contributed to the WH cut-based analysis of the $\sqrt{s} = 8$ TeV data, and as such, this is described in detail here. Brief overviews of the MVA and the cut-based analysis of the ZH channels are also given, as these affect the WH analysis.

The analysis proceeds by applying a number of selection criteria to enhance the signal-to-background ratio. The final number of signal and background events in the data sample are determined by performing a binned maximum likelihood fit.

The focus of this chapter is on the $WH \to \ell\nu b\bar{b}$ analysis, however, the results from the $ZH \to \nu\nu/\ell\ell b\bar{b}$ analyses are also given, and are combined in a binned likelihood fit to improve the sensitivity of the search.



(a) $WH \to \ell\nu b\bar{b}$ **(b)** $ZH \to \nu\bar{\nu}b\bar{b}$



(c) $ZH \to \ell^+\ell^- b\bar{b}$

Figure 5.1 *Feynman diagrams for the three $VH(b\bar{b})$ channels considered.*

At $\sqrt{s} = 8$ TeV the expected VH cross section with next-to-next-to-leading order (NNLO) QCD and next-to-leading order (NLO) EW corrections for a Higgs boson with $m_H = 125$ GeV is $\sigma(WH) = 696.6^{+3.7\%}_{-4.1\%}$ fb, and for ZH is $\sigma(ZH) = 394.3^{+5.1\%}_{-5.0\%}$ fb [54]. The W and Z bosons have branching ratios of $\mathrm{Br}(W \to \ell\nu_\ell) = (10.86 \pm 0.09)\%$, and $\mathrm{Br}(Z \to \ell^+\ell^-) = (3.3658 \pm 0.0023)\%$ where ℓ refers to each lepton flavour individually (not a sum over them), and $\mathrm{Br}(Z \to \nu\bar{\nu}) = (20.00 \pm 0.06)\%$ [20] for the combination of all three ν flavours. Since the τ is heavy enough to decay hadronically, which is difficult to identify, hadronic τ decays are not considered. There is no dedicated selection designed to identify leptonically decaying τ leptons, however, these decays will have some acceptance in the other leptonic decay channels considered. The three channels are distinguished from each other based on the number of final state leptons:

- **0-lepton channel:** targeting $ZH \to \nu\bar{\nu}b\bar{b}$,
- **1-lepton channel:** targeting $WH \to \ell\nu b\bar{b}$,
- **2-lepton channel:** targeting $ZH \to \ell^+\ell^- b\bar{b}$.

Here, ℓ refers to either an electron or muon. These channels, whilst aiming to target specific processes, have overlapping contributions, as detailed in Table 5.2.

5.2 WH Analysis Strategy

The WH signal region is defined as an event with an electron or muon, at least two b jets and missing transverse momentum. The largest backgrounds come from W/Z+jets and $t\bar{t}$ production. Other significant backgrounds come from single top quark and diboson (WZ and ZZ) production and multi-jet events. The backgrounds to the analysis are estimated using both Monte Carlo (MC) and data-driven techniques, which are discussed in Sections 5.5 and 5.6. Particular attention is paid to the multi-jet background estimation in Section 5.7.

The analysis procedure is validated by using a similar diboson final state signature, $VZ(\to b\bar{b})$, which has a softer p_{T} spectrum and a lower peak in the invariant mass distribution of the two b quarks. The diboson production has a cross section that is five times larger than the expected 125 GeV Higgs boson cross section.

The binned maximum likelihood fit (also referred to as the *global likelihood fit*) is used to perform a fit to m_{bb} (in some control regions which are described in the following text, the fit is performed on the MV1c distribution of the leading jet in the event) in the dijet mass analysis, or the BDT output for the MVA (see Section 5.9 for the fitting procedure). The inputs to the likelihood fit are taken from 81 signal regions and 11 control regions determined by the number of jets in the event, and divided into bins of the p_{T} of the W or Z boson, as described in Section 5.4. The primary metric used to identify the presence of a signal is the signal strength parameter $\mu = \sigma/\sigma_{\mathrm{SM}}$, where σ is the measured cross section, and σ_{SM} is the expected cross section for a Standard Model Higgs boson.

The observed m_{bb} distributions from data are fitted to those from simulated signal and background processes. The impact of systematic uncertainties on the shape and normalisation of the expected signal and background processes (as discussed in Section 5.8) is described in terms of a set of nuisance parameters that are constrained by Gaussian or log-normal prior probability distributions. The likelihood is profiled, with the nuisance parameters treated as floating parameters in the fit, constraining them. An initial fit is performed where estimates of some of the background normalisations are found by leaving them to float in the fit. The backgrounds with the corrected normalisation are used for the subsequent fitting procedures where the nuisance parameters are constrained and the final event yields are determined. In the text that follows, the results obtained before

the global likelihood fit are referred to as *prefit*, and after the fit as *postfit*. The analysis is performed *blinded*, as described in Section 5.3.

5.3 Blinding Strategy

In order to prevent data-driven or personal bias in an analysis it can be performed 'blinded'. In this type of analysis, the final result in a predefined signal region is kept hidden until the details of analysis are fully understood and described, such as the selection criteria and systematic uncertainties.

The analysis presented here was performed blind. The data in the signal dijet mass window (i.e. 110-140 GeV) was not revealed whilst the analysis procedure was being validated. The likelihood fits that were performed before unblinding still *included* the data, but the postfit signal strength, μ, was not looked at, nor was the 110-140 GeV region in the dijet mass distributions. In this way, it was ensured that the fit was performing as it should, without any knowledge of what the signal looked like. After the fit model was deemed sufficiently well understood, the data was unblinded.

After unblinding the data, disagreement was seen between data and MC that was not understood in the regions where the W boson candidate had low transverse momentum (p_{T}^{W}) in the 1-lepton electron channel. No cuts or selection criteria were changed, however, this region is not included in the final global likelihood fit presented in this thesis.

5.4 Event Selection Criteria

Selections are applied to the data and Monte Carlo samples to enhance the signal-to-background ratio. Initially, the events are characterised based on trigger selections, followed by a set of geometrical and kinematic cuts on reconstructed physics objects, which are detailed in Table 5.1 and motivated in the text below.

5.4.1 Trigger Selection

0-lepton events (targeting $ZH \to \nu\bar{\nu}b\bar{b}$) are selected from events that pass an $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger with a threshold of 80 GeV. The turn-on of the trigger reaches

maximum efficiency (of almost 100%) at the plateau region where the offline $E_{\mathrm{T}}^{\mathrm{miss}} > 120$ GeV[1]. Below this point the coarse granularity, and resolution, of the Level 1 trigger results in some events with lower $E_{\mathrm{T}}^{\mathrm{miss}}$ being underestimated and discarded. 1-lepton selection (targeting $WH \to \ell\nu b\bar{b}$) uses two sets of single lepton triggers. The first set has an electron E_{T} threshold of 24 GeV, muon p_{T} threshold of 24 GeV and includes track isolation cuts in order to reduce the effect of high pileup, although this requirement lowers the efficiency of the trigger. The second set is intended to recover some of this efficiency by removing the track isolation cuts and imposing much higher thresholds of 60 GeV (electrons) and 36 GeV (muons). The 2-lepton selection (targeting $ZH \to \ell^+\ell^- b\bar{b}$) uses the same triggers as the 1-lepton selection with additional di-electron and di-muon triggers with thresholds of 12 GeV and 13 GeV, respectively.

5.4.2 Object Selection

After the initial trigger selection, further selections are performed on the reconstructed objects in the event. The objects considered are the number of loose, medium and tight leptons (as discussed in Sections 4.5 and 4.6), the number of jets (and b jets, specifically) and the $E_{\mathrm{T}}^{\mathrm{miss}}$ in the event. The lepton selections are slightly different for each of the VH channels as they have different kinematics and final states; however, they have the same jet selections.

Selections on the number of leptons are as follows:

- the 0-lepton channel must contain no loose leptons,
- the 1-lepton channel must contain one tight lepton and no further leptons,
- the 2-lepton channel must contain two loose leptons, of which at least one is also a medium lepton.

Jets are reconstructed using the anti-k_t algorithm with a distance parameter of $R = 0.4$ and calibrated using the EM+JES calibration scheme with a further correction applied from GSC (discussed in Section 4.7.2). Events with any *forward jets* ($|\eta > 2.5|$) of $p_{\mathrm{T}} > 30$ GeV are vetoed in order to reduce backgrounds from top quark processes. Jets that are considered in the analysis, so called *selected*

[1]The online $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger does not contain Inner Detector or Muon Spectrometer information.

jets, must be within $|\eta| < 2.5$ and have $p_{\mathrm{T}} > 20$ GeV. To reduce the effects of pile-up there is a cut of $|\mathrm{JVF}| > 0.5$ for jets with $p_{\mathrm{T}} < 50$ GeV and $|\eta| < 2.4$. Exactly two or three selected jets are required, and the events are then further categorised by the number of b jets. Events with more than two b jets are vetoed and at least one of the b jets must have $p_{\mathrm{T}} > 45$ GeV. In the the 3-jet region, events where the lowest p_{T} jet is b tagged are also vetoed. The b jets are selected according to either loose, medium or tight operating points, as discussed in Section 4.8. Events with 0, 1 or 2 b jets are divided into five regions:

- **0-tag:** Events with exactly zero b jets,
- **1-tag:** Events with exactly one loose b jet,
- **2-tag:** Events with exactly two b jets. In order to improve sensitivity and to get a better handle on the flavour composition of the tagged jets, this is subdivided based on the b-jet classification (see Figure 5.2):
  - **2L category:** 2 loose b jets, one of which can be medium or tight,
  - **2M category:** 2 medium b jets, one of which can be tight,
  - **2T category:** 2 tight b jets.

Overlap removal (as discussed in Section 4.10) is performed to prevent objects being identified as multiple types of objects. The procedure is as follows:

- If an electron and a jet overlap within $\Delta R < 0.4$, the jet is discarded.
- If a muon and a jet overlap within $\Delta R < 0.4$, the jet is discarded if it has ≤ 3 associated tracks, as this is likely to have come from a muon; otherwise, the muon is discarded.
- If a muon and electron overlap within $\Delta R < 0.2$ the muon is discarded if it is only identified in the calorimeter, otherwise, the electron is discarded.

Any muons, overlapping with a jet, that are discarded are used for jet energy corrections (see b-jet p_{T} correction in Section 5.4.3) and in the calculation of $E_{\mathrm{T}}^{\mathrm{miss}}$. The additional criteria in the muon-jet overlap removal are chosen to identify the case where a muon from the vector boson decay falls within a jet.

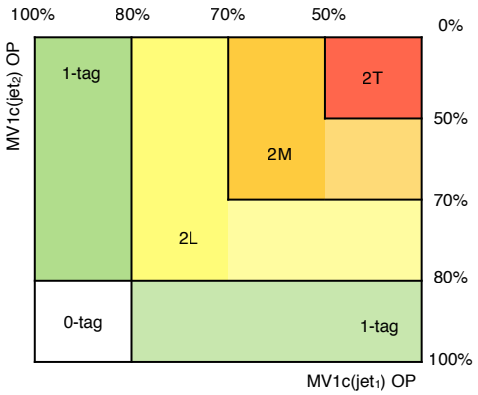


Figure 5.2 *Summary of the event classification criteria as a function of the MV1c scores for the two highest p_T jets. The percentages correspond to the b-tagging efficiency of the MV1c operating points (OP) which define the loose, medium and tight jets [1].*

5.4.3 Geometrical and Kinematic Selections

Further geometrical and kinematic cuts are applied to reduce specific backgrounds, which are summarised in Table 5.1. In general, the cuts are looser for the MVA selection, allowing the BDT to find the optimal cuts using more information, and at the same time increasing the number of events available for training.

Events are further categorised based on the magnitude of the transverse momentum of the vector boson candidate (p_{T}^{V}), taking advantage of the increased signal-to-background ratio and sensitivity at high p_{T}^{V}. In the 1-lepton channel, the transverse momentum of the W boson, $\mathbf{p}_{\mathrm{T}}^{W}$ (with magnitude p_{T}^{W}), is the vector sum of the lepton transverse momentum and the $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$. In the 0-lepton channel the transverse momentum of the Z boson, $\mathbf{p}_{\mathrm{T}}^{Z}$ (with magnitude p_{T}^{Z}), is given by $\mathbf{E}_{\mathrm{T}}^{\mathrm{miss}}$, and in the 2-lepton channel it is the vector sum of the two leptons' transverse momentum.

In the 0-lepton channel the lowest p_{T}^{V} bin begins at 100 GeV, which is possible due to a parameterisation of the $E_{\mathrm{T}}^{\mathrm{miss}}$ trigger, allowing for the trigger threshold to be lowered from 120 GeV without a significant loss in efficiency. In the 1-lepton channel, only the muon channel is used for the $p_{\mathrm{T}}^{W} < 120$ GeV bins. As discussed in Section 5.3, the electron channel has large contributions in this region from the difficult-to-model multi-jet background and is not considered. Only two p_{T}^{V} bins are defined in the MVA in order to maximise the use of available training events, and since the MVA is given p_{T}^{V} as a training variable. The p_{T}^{V} bins are

Variable	Dijet-mass analysis					Multivariate analysis	
Common selection							
p_{T}^{V} [GeV]	0–90	$90^{(*)}$–120	120–160	160–200	> 200	0–120	> 120
$\Delta R(\mathrm{jet}_1, \mathrm{jet}_2)$	0.7–3.4	0.7–3.0	0.7–2.3	0.7–1.8	< 1.4	> 0.7 (p_{T}^{V}<200 GeV)	
0-lepton selection							
$p_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]	NU	> 30	> 30			NU	> 30
$\Delta\phi(\vec{E}_{\mathrm{T}}^{miss}, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$		$< \pi/2$	$< \pi/2$				$< \pi/2$
$\min[\Delta\phi(\vec{E}_{\mathrm{T}}^{miss}, \mathrm{jet})]$		–	> 1.5				> 1.5
$\Delta\phi(\vec{E}_{\mathrm{T}}^{miss}, \mathrm{dijet})$		> 2.2	> 2.8				> 2.8
$\sum_{i=1}^{N_{\mathrm{jet}}=2(3)} p_{\mathrm{T}}^{\mathrm{jet}_i}$ [GeV]		> 120 (NU)	> 120 (150)				> 120 (150)
		See text	–				–
1-lepton selection							
m_{T}^{W} [GeV]	< 120					–	
H_{T} [GeV]	> 180		–			> 180	–
$E_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]	–		> 20		> 50	–	> 20
2-lepton selection							
$m_{\ell\ell}$ [GeV]	83-99					71-121	
$E_{\mathrm{T}}^{\mathrm{miss}}$ [GeV]	< 60					–	

Table 5.1 *Event selections for the different channels.* $(*)$ *In the 0-lepton channel, the lower edge of the second* p_T^V *bin is set at 100* GeV *instead of 90* GeV. *For the 1-lepton channel, only the 1-muon sub-channel is used in the* $p_T^V < 120$ GeV *bins. Where there is a value in brackets for a specific cut (*$\Delta R(\mathrm{jet}_1, \mathrm{jet}_2)$ *and* $\sum_{i=1}^{N_{\mathrm{jet}}=2(3)} p_T^{\mathrm{jet}_i}$*), the value in brackets refers to a different cut used in the electron channel. 'NU' indicates that the cut is not used.*

given in Table 5.1.

The angular separation $\Delta\phi(j_1, j_2)$ between the leading two jets depends on p_{T}^{V}, and as such this cut is changed for each bin. These cuts are removed at high p_{T}^{V} due to the relatively small background contributions.

In the 1-lepton channel, cuts are made on the transverse mass of the W boson, m_{T}^{W}, the scalar sum of the full set of reconstructed objects, H_{T}, and $E_{\mathrm{T}}^{\mathrm{miss}}$. Due to the presence of neutrinos in the W boson decay, its mass cannot be reconstructed precisely, therefore the transverse mass is used, defined as $m_{\mathrm{T}}^{W} = \sqrt{2p_{\mathrm{T}}^{\ell}E_{\mathrm{T}}^{\mathrm{miss}}(1 - \cos(\phi^{\ell} - \phi^{\mathrm{miss}}))}$, where p_{T}^{ℓ} and ϕ^{ℓ} refer to the transverse momentum and azimuthal angle of the lepton, $E_{\mathrm{T}}^{\mathrm{miss}}$ and its azimuthal angle ϕ^{miss}. The m_{T}^{W} selection is used to identify W bosons and reduce the $t\bar{t}$ background and QCD at low m_{T}^{W}. H_{T} is the scalar sum of the $E_{\mathrm{T}}^{\mathrm{miss}}$, the p_{T} of the leading two

jets and lepton. The H_{T} cut is used in the $p_{\mathrm{T}}^{V} < 120$ GeV bin, and the $E_{\mathrm{T}}^{\mathrm{miss}}$ cut in $p_{\mathrm{T}}^{V} > 120$ GeV, in order to reduce the multi-jet background.

The final discriminant in the cut-based analysis is the dijet invariant mass. Excellent mass resolution of jets is therefore paramount; energy and p_{T} calibrations are performed to improve this [1]. Energy from muons, which were removed during overlap removal with jets, is added to the energy of the b jet, minus associated measurements of the muon energy in the calorimeter. The p_{T} resolution of the b jets is improved by applying a p_{T}-dependent correction derived from simulated VH events. This improves the dijet mass resolution by around 14%, resulting in a dijet mass resolution of around 11% (or ≈14 GeV).

The acceptance and cross section times branching ratio for all three channels after the full event selection is shown in Table 5.2.

$m_H = 125$ GeV at $\sqrt{s} = 8$TeV				
Process	$\sigma\times$ BR [fb]	Acceptance [%]		
		0-lepton	1-lepton	2-lepton
$q\bar{q} \to (Z \to \ell^+\ell^-)(H \to b\bar{b})$	14.9 ± 0.70	–	1.1	10.9
$gg \to (Z \to \ell^+\ell^-)(H \to b\bar{b})$	1.3 ± 0.85	–	0.7	8.1
$q\bar{q} \to (W \to \ell\nu)(H \to b\bar{b})$	131.7 ± 6.23	0.3	3.7	–
$q\bar{q} \to (Z \to \nu\bar{\nu})(H \to b\bar{b})$	44.2 ± 2.08	3.8	–	–
$gg \to (Z \to \nu\bar{\nu})(H \to b\bar{b})$	3.8 ± 2.51	5.0	–	–

Table 5.2 *The cross section, σ, times branching ratio (BR) and acceptance for the three channels at 8* TeV*. The $q\bar{q}$- and gg-production modes are shown separately. The branching ratio for $Z \to \ell^+\ell^-$ refers to just electrons and muons, $W \to \ell\nu$ to all three lepton flavours and decays to neutrinos for $Z \to \nu\bar{\nu}$. Acceptance is defined here as the fraction of events remaining after the full 2-tag signal event selection. Other production and decay modes of the Higgs boson are negligible after these selections. The uncertainties on $\sigma\times BR$ include scale and PDF uncertainties on the production cross section and uncertainties on the branching ratios [1, 54].*

5.4.4 Multivariate Analysis

In tandem with the dijet mass analysis, a multivariate analysis (MVA) was performed using a boosted decision tree (BDT). The BDT takes as inputs additional kinematic, geometrical and b-tagging properties and accounts for correlations between these variables.

Variable	0-Lepton	1-Lepton	2-Lepton
p_{T}^{V}		×	×
$E_{\mathrm{T}}^{\mathrm{miss}}$	×	×	×
$p_{\mathrm{T}}^{b_1}$	×	×	×
$p_{\mathrm{T}}^{b_2}$	×	×	×
m_{bb}	×	×	×
$\Delta R(b_1, b_2)$	×	×	×
$\lvert\Delta\eta(b_1, b_2)\rvert$	×		×
$\Delta\phi(V, bb)$	×	×	×
$\lvert\Delta\eta(V, bb)\rvert$			×
H_{T}	×		
$\min[\Delta\phi(\ell, b)]$		×	
m_{T}^{W}		×	
$m_{\ell\ell}$			×
MV1c(b_1)	×	×	×
MV1c(b_2)	×	×	×
	Only in 3-jet events		
$p_{\mathrm{T}}^{\mathrm{jet}_3}$	×	×	×
m_{bbj}	×	×	×

Table 5.3 *Variables used in the multivariate analysis for the 0-, 1- and 2-lepton channels. The variables b_1 and b_2 refer to the leading and sub-leading b jets according to p_T.*

In addition to looser selection criteria, there are fewer p_{T}^{V} bins and the electron and muon sub-channels are combined. Only two p_{T}^{V} bins are used for the 1- and 2-lepton channels: $p_{\mathrm{T}}^{V} < 120$ GeV and $p_{\mathrm{T}}^{V} > 120$ GeV, and only one bin for the 0-lepton channel: $p_{\mathrm{T}}^{V} > 120$ GeV. This provides more events for training the BDT.

The full list of variables used for the MVA is listed in Table 5.3.

5.4.5 Control Regions

Events in the 0-tag and 1-tag regions are used as *control regions* to constrain the main backgrounds. Events in the 1-lepton channel with three selected jets and 2-tags provide a control region since this largely consists of $t\bar{t}$ events. The 1-tag

Process	Generator
Signal	
$q\bar{q} \to ZH \to \nu\bar{\nu}b\bar{b}/\ell^+\ell^-b\bar{b}$	PYTHIA8
$gg \to ZH \to \nu\bar{\nu}b\bar{b}/\ell^+\ell^-b\bar{b}$	POWHEG+PYTHIA8
$q\bar{q} \to WH \to \ell\nu b\bar{b}$	PYTHIA8
Vector boson + jets	
$W \to \ell\nu$	SHERPA 1.4.1
$Z/\gamma* \to \ell^+\ell^-$	SHERPA 1.4.1
$Z \to \nu\bar{\nu}$	SHERPA 1.4.1
Top-quark	
$t\bar{t}$	POWHEG+PYTHIA6
t-channel	ACERMC+PYTHIA6
s-channel	POWHEG+PYTHIA6
Wt	POWHEG+PYTHIA6
Diboson	POWHEG+PYTHIA8
WW	POWHEG+PYTHIA8
WZ	POWHEG+PYTHIA8
ZZ	POWHEG+PYTHIA8

Table 5.4 *The generators used for the simulation of the signal and background processes. [1]*

control regions and 2-tag signal regions are both included in the global likelihood fit, however, the 0-tag control regions are not; they are only used to improve the background modelling.

5.5 Data and Simulated Samples

The data used for the analysis presented here is from pp collision data recorded by ATLAS with stable beams and all subsystems providing high quality data during 2012, corresponding to an integrated luminosity of 20.3 fb^{-1} at a centre-of-mass energy of $\sqrt{s} = 8$ TeV.

Monte Carlo simulated samples are produced to model the signal and the majority of the backgrounds. The multi-jet background is estimated from data, however, it does depend on the simulation of other backgrounds. All detector simulations are produced using the ATLFAST-II simulation [109].

The $q\bar{q}$-initiated WH and ZH production is simulated with the PYTHIA8 [98] generator with the CTEQ6L1 [143] PDFs and AU2 tune [144, 145] for the underlying event. The tune specifies the parton shower, hadronisation, and multiple parton interactions. QED final-state radiation is simulated with PHOTOS[146]. gg initiated ZH production is simulated to leading order (LO) in QCD using the POWHEG generator [102–104] within the MiNLO approach [147] using the CT10 PDFs [148], interfaced to PYTHIA8 with the AU2 tune. MiNLO improves the parton showering predictions of POWHEG by choosing the renormalisation scale dynamically. The same setup is used for the $q\bar{q}$ processes as a crosscheck [149].

The p_{T} distribution of the Higgs boson is different for the two ZH production processes and requires different calculations. For the $q\bar{q}$ processes the total production cross sections and uncertainties are calculated at NNLO in QCD [150–152], and with vector-boson-p_{T}-dependent NLO electroweak corrections [153, 154]. For gg initiated ZH production, NLO corrections in QCD [155] are also taken into account. The $q\bar{q}$ and gg initiated ZH samples are then combined and weighted according to cross section. In this analysis only the $H \to b\bar{b}$ decay mode is considered, and the branching ratios are calculated with HDECAY [156]. All processes are simulated with a range of Higgs boson masses from 100 to 140 GeV in steps of 5 GeV.

The V+jets backgrounds are simulated at LO in QCD with the SHERPA generator [101] and the CT10 PDFs. The cross sections are calculated at NNLO for (W/Z)+jets [157].

For $t\bar{t}$ production the POWHEG generator is used with the CT10 PDFs, and then interfaced with PYTHIA6 [97] using the CTEQ6L1 PDFs and the Perugia2011C tune [144, 145], with the cross section normalised to NNLO [158]. The Perugia2011C tunes are designed to provide a best-guess prediction of the charged track multiplicity.

The single top and diboson backgrounds are simulated with POWHEG in the same way as for $t\bar{t}$, however, the t-channel exchange is simulated with the AcerMC generator [105] interfaced with PYTHIA6, using the CTEQ6L1 PDFs and the Perugia2011C tune. For the diboson channel, PYTHIA8 is used instead of PYTHIA6 [159]. The cross sections are calculated in the following references [160–162].

5.6 Backgrounds

The backgrounds to the analysis are taken from simulation, apart from the multi-jet background, which is estimated from data. The background normalisations for W+jets and $t\bar{t}$ are left to float freely in a preliminary global likelihood fit, with prior constraints applied from theoretical cross sections, which are used to optimise the initial selection. In subsequent fits these background normalisation factors are fixed. The diboson normalisation is taken from theory, and the multi-jet normalisation is estimated from data. Relative background contribution estimates in the inclusive 2L+2M+2T b tag region, after the global likelihood fit, are shown in Table 5.5.

p_{T}^{V} bin (GeV)	$p_{\mathrm{T}}^{V} < 90$	$90 \leq p_{\mathrm{T}}^{V} < 120$	$120 \leq p_{\mathrm{T}}^{V} < 160$	$160 \leq p_{\mathrm{T}}^{V} < 200$	$p_{\mathrm{T}}^{V} \geq 200$	Incl.
Z+jets	2.66%	3.17%	1.73%	1.36%	0.85%	2.52%
W+jets	36.4 %	34.4 %	31.2 %	37.2 %	46.6 %	35.5 %
$t\bar{t}$	30.3 %	39.4 %	45.8 %	38.5 %	26.9 %	34.5 %
s-top	17.4 %	13.7 %	10.7 %	9.86%	8.55%	15.3 %
VV	1.32%	1.48%	2.07%	3.67%	5.98%	1.64%
MJ (el)	0 %	0 %	2.93%	2.62%	1.09%	2.68%
MJ (μ)	6.65%	2.32%	0.23%	0.15%	0 %	4.53%

Table 5.5 *The relative contribution of each major background is shown here in the inclusive 2-tag region (2L+2M+2T). These values are after the global likelihood fit. The inclusive MJ contribution in the electron channel is calculated for* $p_T^V > 120$ GeV.

5.6.1 W+jets

The W+jets background includes contributions from b jets, c jets and light jets (l). The $W + b\bar{b}$ background is irreducible for the analysis, whilst charm, strange, up and down quarks can be misidentified as heavy flavoured b jets, providing an additional background contribution. This region is referred to as the W + heavy-flavour(hf) background, where hf = $bb + bc + bl + cc$. The entire W+jets contribution accounts for about 45% of the total background in the inclusive 2L+2M+2T b-tag region, and is particularly dominant where $p_{\mathrm{T}}^{W} > 200$ GeV, consisting of over 60% of the background.

The relative flavour composition of the jets is not well known. There are significant contributions to the 1-lepton channel from $W + l/cl$ in the 0 and 1-tag regions. The W + hf background becomes dominant for the 2 b-jet regions,

increasing as the b-tagging requirement is tightened and p_{T}^{W} increases. Separating the 2-tag region into 2L, 2M and 2T offers increased sensitivity to the flavour composition. The c jets and light jets are not directly tagged; a parameterisation of their probability to be b tagged as a function of p_{T} and η is used. This is applied as a reweighting factor for truth-c-jets and light jets in simulated $W + l$, $W + cl$ and WW backgrounds.

In the 0-tag and 1-tag regions, which have a large contribution from W+light jets, a mismodelling of the $\Delta\phi(j_1, j_2)$ (Figure 5.3) and p_{T}^{W} (Figure 5.4) is seen in both the electron and muon channels. The source of this is the MC generators; it is seen in both Sherpa and Alpgen (although an improvement is seen with NLO generators [163]). A slope in the data-to-MC comparison of these variables shows a deficit in the low p_{T}^{W} bins and an excess in the higher bins. This effect is seen in the $\Delta\phi(j_1, j_2)$ distributions as well, where high p_{T}^{W} corresponds to low $\Delta\phi(j_1, j_2)$. Subsequently, corrections are applied, parameterised by p_{T}^{W}. The effect of the corrections in the muon channel for $\Delta\phi(j_1, j_2)$ is shown in Figure 5.3 and for p_{T}^{W} in Figure 5.4.

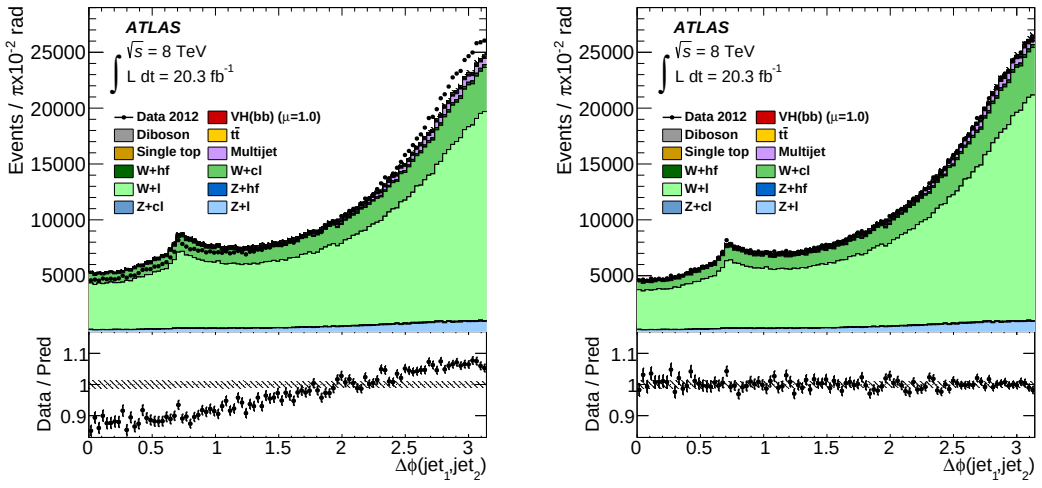
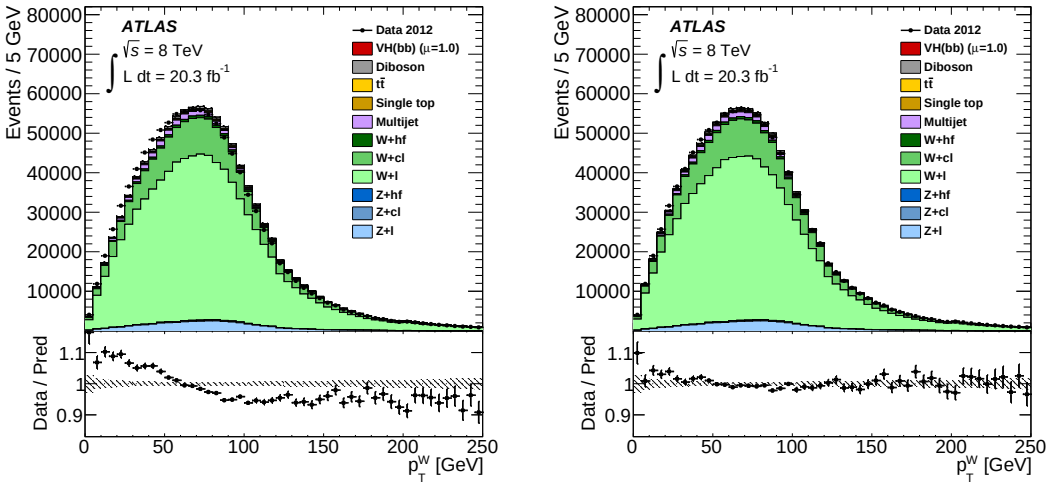


Figure 5.3 *The $\Delta\phi(j_1, j_2)$ distribution observed in the 2-jet 0-tag control region of the 1-muon sub-channel (a) before and (b) after reweighting to correct the mismodelling of the background Monte Carlo samples. All the p_T^W bins are combined.*

A strong dependence of the b-tagging efficiency is seen on ΔR between the jet being considered and the closest jet to it, where large differences are seen between parameterised tagging and direct tagging for $W + cc$ events with $\Delta R < 1$. As such, a dedicated correction depending on ΔR is applied to the $W + cc$ events



(a) **(b)**

Figure 5.4 *The p_T^W distribution in the 2-jet 0-tag control region of the 1-muon sub-channel (a) before and (b) after $\Delta\phi(j_1, j_2)$ reweighting to correct the mismodelling of the background Monte Carlo samples.*

when truth tagging is used (see Section 5.7.2).

5.6.2 $t\bar{t}$ and single top

The $t\bar{t}$ background is a major source of background in the 2-jet and 3-jet categories. The top quark is expected to decay exclusively to Wq ($q = d, s, d$) and dominated by Wb, with a measured ratio of $\mathrm{Br}(t \to Wb)/\sum_q \mathrm{Br}(t \to Wq) = 1.014 \pm 0.003(stat.) \pm 0.032(syst.)$ [20]. The three major final states of the $t\bar{t}$ background are determined by the decay of the W bosons: di-leptonic, semi-leptonic (where one W boson decays leptonically and the other hadronically), or fully hadronic.

The semi-leptonic decay provides the largest background to the 1-lepton analysis, although there are also contributions from the di-leptonic decays at low p_{T}^W. This final state produces four jets, two of which are b jets, a single lepton and $E_{\mathrm{T}}^{\mathrm{miss}}$. The contribution in the 2-jet region is not significant, but it is the dominant background for the 3-jet region. The contribution of this background ranges from between 21% of the total background at high p_{T}^V and up to 40% in the $120 < p_{\mathrm{T}}^V < 160$ GeV bin.

The simulations of the top p_{T} from the POWHEG generator interfaced to PYTHIA produce a spectrum that is too hard. It is corrected by applying a

correction at the level of generated top quarks in the $t\bar{t}$ production process.

In addition to $t\bar{t}$ production, single top production contributes between 14% of the total background at low p_{T}^{V}, and 7% at high p_{T}^{V}. This occurs in three channels:

- s-channel, where a top and bottom quark are produced by the decay of a virtual W boson,
- t-channel, where a top quark is produced along with a bottom quark and a light quark,
- Wt-channel, where the top quark is produced in association with a W boson.

5.6.3 Diboson Processes

Diboson processes, WW, WZ and ZZ, are backgrounds to the search. The main contribution to the 1-lepton channel comes from $W \to \ell\nu$ and $Z \to b\bar{b}$. Smaller contributions arise from $Z \to \ell^{+}\ell^{-}$, through the misclassification or reconstruction of one of the leptons, and $W \to q\bar{q}$, where the W is mistagged. The background contribution from the diboson processes is between 1.4% at low p_{T}^{V} and 5% at high p_{T}^{V}. The $V(Z \to b\bar{b})$ process is predicted to have a cross section almost 5 times that of $V(H \to b\bar{b})$, which offers an opportunity to validate the analysis.

5.7 Multi-Jet Background

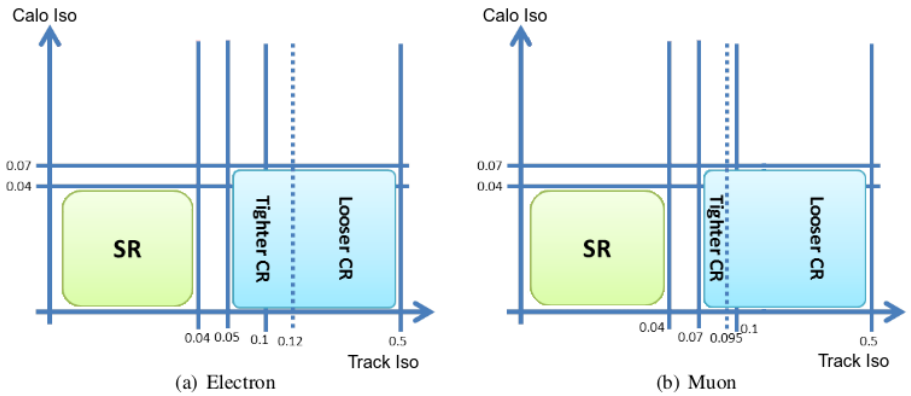
There is a huge production cross-section from QCD multi-jet (MJ) events, which is over five orders of magnitude greater than the VH processes [164]. The VH selections are designed to remove the majority of the MJ events, and although MJ events have a low acceptance, the huge cross-section means that the MJ contribution is non-zero. The multi-jet background arises predominantly from QCD processes where jets and photon conversions are incorrectly identified as leptons or $E_{\mathrm{T}}^{\mathrm{miss}}$. It is particularly apparent in the electron channel as jets are misidentified as electrons more often than as muons. Lepton isolation criteria provide crucial discrimination between vector boson decay products and multi-jet events in these situations. The MJ background contribution is measured using data driven methods as generating sufficient amounts of simulated events would

be unfeasible. The following sections describe a number of aspects of the MJ background estimation in the 1-lepton channel. Firstly, the template used to estimate the multi-jet background is introduced in Section 5.7.1, in Section 5.7.2 a method of increasing the number of events with tight-tagged b jets is discussed. The normalisation of the multi-jet background is discussed in Section 5.7.3 and the systematic error calculations are given in Section 5.8.2.

5.7.1 Multi-Jet Template

A multi-jet-enriched template region is defined from a region orthogonal to the signal region (defined in Table 5.1) by applying all the event selection cuts and a modified lepton isolation selection, as illustrated in Figure 5.5. Events with multiple leptons (including the leptons identified with the nominal isolation requirements and the modified isolation requirements) are vetoed. After these selections have been applied to both data and all other backgrounds, any differences seen in the yields between data and the other backgrounds are assumed to come from multi-jet events, since the region is signal-free. This process is described below, which provides the shape of the background. For reference, the shape of the dijet mass from the MJ background is shown in Figure 5.13. The estimate of the MJ contribution then needs to be extrapolated from the template region into the signal region and the normalisation is determined through a fit to data. This is described in detail in Section 5.7.3.

Electrons and muons in the MJ-enriched template region are selected with looser cuts, inverted track isolation cuts and a looser calorimeter isolation cut. In the MJ enriched region, in order to identify jets that are being misidentified as leptons, the isolation criteria are modified to $0.05(0.07) < iso_{track} < 0.5$ and $0.07 < iso_{calo}$ for the electron (muon) channel. After overlap removal has been performed on the leptons identified with the modified isolation, further requirements are imposed on the isolation cuts to construct nominal and systematic variations of the MJ template. In the electron channel, the iso_{track} region is divided into two regions which have almost the same number of events being selected: $0.05 < iso_{track} < 0.12$ is used for the nominal template, and $0.12 < iso_{track} < 0.5$ is used for a systematic variation which is symmetrised to give the up and down systematic variations. In the muon channel, the full iso_{track} region is used for the nominal template, and two regions with an almost equal number of events, $0.05 < iso_{track} < 0.12$ and $0.12 < iso_{track} < 0.5$, are used for the up and down



(a) Electron (b) Muon

Figure 5.5 *Lepton selection for the WH analysis in the (iso_{track}, iso_{calo}) plane [1]. The green box corresponds to tight lepton selection as used for the nominal event selection (signal region or SR) and the blue box is the MJ lepton selection. Nominal MJ electrons are those in the tighter control region (CR), and those in the looser control region are used for systematic uncertainty determination. All MJ muons in the area defined by the blue box are considered as nominal MJ muons, whilst the tigher and looser control regions are used for upper and lower systematic variations.*

systematic variations, respectively. The nominal template values are shown in Table 5.6 and the systematics are further discussed in Section 5.8.2.

Property	Nominal	MJ electron	MJ muon
iso_{track}	< 0.04	$0.05 < iso_{track} < 0.12$	$0.07 < iso_{track} < 0.5$
iso_{calo}	< 0.04	< 0.07	
Electron selection	Tight	Medium	-

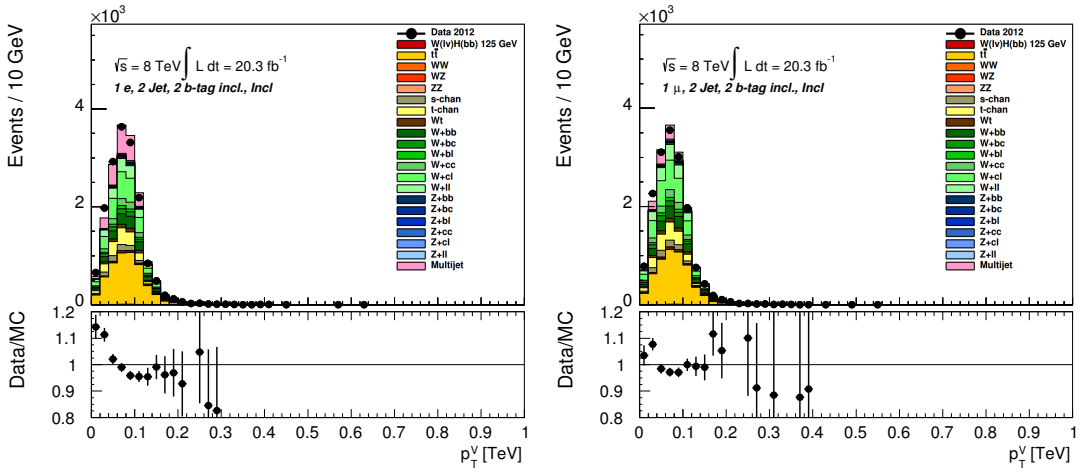
Table 5.6 *The modified isolation requirements for lepton selection in the MJ-enriched template region.*

After requiring the two b jets in the event selection, the number of events is severely reduced, so one b-tag events are used in this two b-tag region (discussed in Section 5.7.2).

By applying the event selection defined by the MJ template, the initial estimate of the MJ background is found to be concentrated at low p_{T}^{W}, as can be seen in Figure 5.6. The H_{T} cut in the $p_{\mathrm{T}}^{W} < 120$ GeV bin reduces the MJ contribution, but it is still substantial in the electron channel, contributing up to 11% for the 2L category and 6% in the 2T category. As such, since the background cannot be well modelled or constrained for this bin in the electron channel, it is not used in the signal selection.

Case	Order	Keep μ	Keep e	Keep jet
$\Delta R(jet, e) < 0.4$	1	-	Yes	No
$\Delta R(jet, \mu) < 0.4$	2	$N_{trk} \leq 3$	-	$N_{trk} \geq 4$
$\Delta R(\mu, e) < 0.2$	3	if not calo μ	if calo μ	-

Table 5.7 *A summary of the overlap removal in the multi-jet selection and the order in which this is done. Here, N_{trk} refers to the number of ghost matched tracks with $p_T > 0.5$ GeV. In ghost matching, the tracks are treated as infinitesimally soft, low-p_T particles and added to the list of inputs to the calorimeter jets. They do not affect the jet clustering, but afterwards it is possible to identify which tracks were clustered into each jet.*



(a) **(b)**

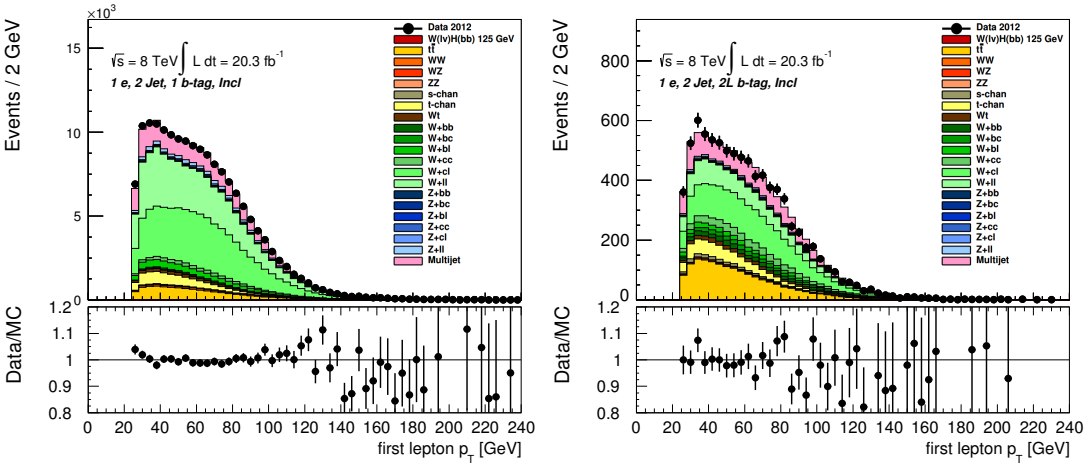
Figure 5.6 *Distributions of p_T^W in the inclusive 2 b-tag region (2L+2M+2T) in the electron (a) and muon channel (b). The multi-jet contribution (shown in pink) is concentrated in the $p_T^W < 120$ GeV region.*

5.7.2 Truth Tagging

When applying the two b-jet selection many events are removed by the b-tagging selection. This results in relatively few events from the other simulated backgrounds in the template region, which makes shape comparisons with data difficult to perform. Since the MJ contribution is calculated by taking the difference in yields from all other backgrounds (from MC) and data, this leads to a poor estimate of the number of MJ events. In order to increase the number of simulated events available, a method called *truth tagging* is used. The basic concept of truth tagging is the promotion of 1-tag events to 2-tag events, motivated by the similar kinematics seen in these regions. For example,

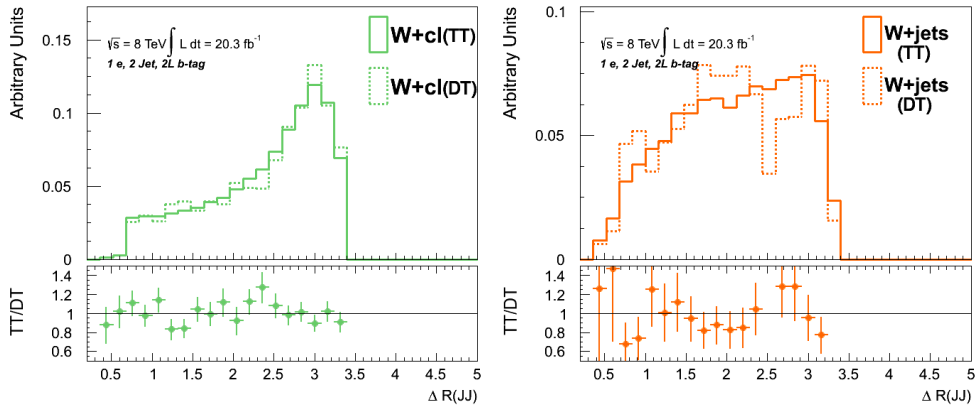
a comparison of the shape of the lepton p_{T} in 1-tag and 2-tag events shows good agreement in the MJ region for the 2L electron channel, as seen in Figure 5.7. Simulated background events that contain exactly one b jet (passing at least the loose criteria) and one non-b-jet are promoted to 2-tag events by assigning an MV1c value to the other non-b-jet such that it passes the b-jet requirement: the non-b-jet is truth tagged. The assigned MV1c value is estimated from the 2-tag MJ region based on the b-tagged jet's MV1c value and the rank of the other jet (leading or subleading) with an accompanying scale factor. The event is reweighted such that the overall yield and the event kinematics are consistent with the 2-tag region. The effect of truth tagging can be seen in Figure 5.8, where a smoother shape is obtained after applying truth tagging.

There are large differences in the kinematics in the electron channel between the 1-tag and 2T-tag regions due to the heavy flavour of the contribution in the 2T channel compared with the dominant light flavour in the 1-tag region. To correct for this a reweighting is applied based on the p_{T}^{W} and $\Delta\phi(j_1, j_2)$ (as discussed in Section 5.6.1). An additional correction is applied to account for biasing the results with truth tagged jets. The details of the truth tagging are given in Appendix B. The contribution of the truth tagged events accounts for almost 90% of the MJ estimate, with the remaining 10% coming from events that have two b-tagged jets.



(a) **(b)**

Figure 5.7 *A comparison of electron p_T between the 1 b-tag and 2L b-tag regions shows good agreement. The shape and relative normalisation motivates the method of truth tagging: promoting 1 b-tag events in to the 2 b-tag region to increase the number of available events.*



(a) **(b)**

Figure 5.8 *Distributions of $\Delta R(jj)$ for (a) $W+cl$ and (b) W+jets MC events with (TT) and without truth tagging (DT), both normalised to the event yield. Distributions with the solid lines are with truth tagging. Truth tagging helps to mitigate the effects of having fewer b-tagged jets.*

5.7.3 Multi-Jet Fitting Procedure

The shape of the MJ contribution is modelled in the region defined by the MJ-enriched template and then extrapolated in the region where the nominal event selection is applied (referred to here as the *signal region*), so as to get the normalisation correct. The following details how this extrapolation is performed.

In the signal region, the goal is to minimise the sum $\text{DATA}_{sig} - \text{MJ}_{sig} - \text{EW}_{sig}$ in bins of a well modelled kinematic variable, where DATA_{sig}, MJ_{sig} and EW_{sig} refer to the contributions in the signal region from data, the multi-jet background, and all other backgrounds, including top quark processes, respectively. The estimate of the MJ contribution in the MJ template region is first calculated using

$$\text{MJ}_{MJ} = \text{DATA}_{MJ} - \text{EW}_{MJ}, \tag{5.1}$$

where DATA_{MJ} and EW_{MJ} are the contributions in the MJ template region (i.e. the modified lepton isolation requirements were used). However, since the normalisation of MJ is not known exactly and it is coupled to the EW background, scale factors SF_{EW} and SF_{MJ} need to be applied to their distributions in the signal region. In order to find the correct normalisation, these two scale factors are left free to float in a fit performed to data in the signal region where the data and EW are selected using the signal region isolation requirements, and the MJ taken

directly from the template region:

$$\mathrm{MJ}_{sig} = \mathrm{DATA}_{sig} - \mathrm{SF}_{EW} * \mathrm{EW}_{sig} - \mathrm{SF}_{MJ} * \mathrm{MJ}_{MJ}. \tag{5.2}$$

The scale factors are estimated by minimising a χ^2 function with Minuit [165]. The fit is performed inclusively in p_{T}^{W}, where, in the signal region, the overall expected Higgs boson signal-to-background ratio is at the per-mill level, with an expected MJ yield that is an order of magnitude larger than the Higgs boson signal yield (all postfit yields are shown in Appendix F). This procedure should be safe to use as the MJ background is not the dominant background in the signal region and additionally has relatively large systematic uncertainties (as calculated in Section 5.8.2) that are larger than the expected Higgs boson signal yields. The full contribution due to the EW and MJ backgrounds per bin is $\mathrm{SIM}_{sig} = \mathrm{EW}_{sig} * \mathrm{SF}_{EW} + \mathrm{MJ}$, where EW_{sig} is the contribution in the signal region.

The fitting must be done using the distribution in data of a variable that provides stable results. This is performed separately in the 0-, 1- and 2-tag regions for both 2- and 3-jet selections.

The baseline fitting procedure in the signal region (2-tags) was originally performed after the H_{T} cut in the two b-tag inclusive region [166], combining all three exclusive 2L+2M+2T b-tag regions: the *inclusive fit* region. Mismodelling was observed after the application of the H_{T} cut (discussed in Appendix C), motivating a study to investigate whether performing the fit before or after this cut produced better results[2]. In addition, a check was performed to see if performing the fit separately in the 2L/2M/2T b-tag regions (three *exclusive fits*) produced more stable fits. The following variables were investigated as the variable to perform the fit on: $E_{\mathrm{T}}^{\mathrm{miss}}$, m_{T}^{W} and p_{T}^{ℓ}. In total this gives 24 fit regions: {inclusive, exclusive}×{pre-H_{T} cut, post-H_{T} cut}×{$E_{\mathrm{T}}^{\mathrm{miss}}$, m_{T}^{W} , p_{T}^{ℓ}}×{electron channel, muon channel}.

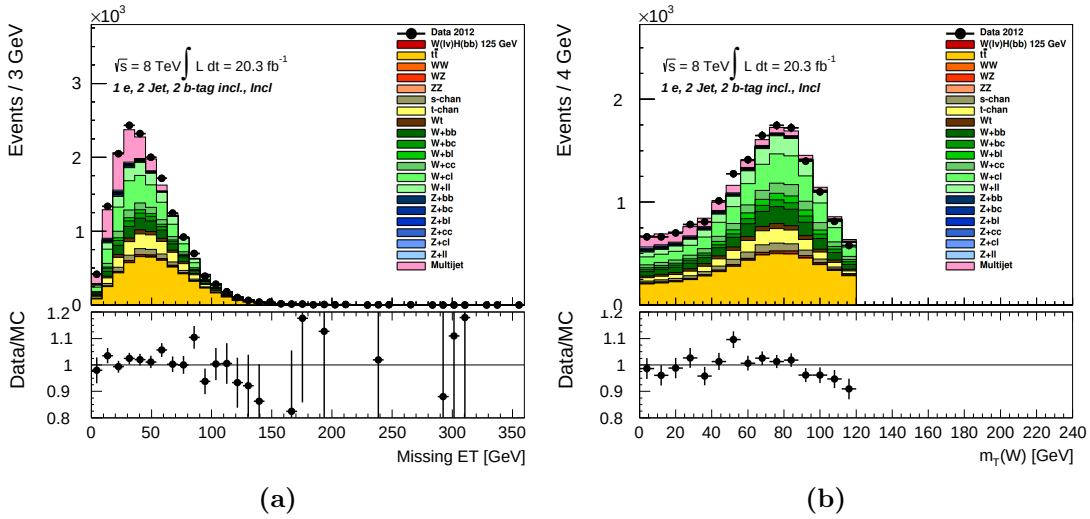
Choice of Fitting Variables

Due to mismodelling of $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$, there were disagreements seen between $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T}^{W} ; fitting on m_{T}^{W} improves the m_{T}^{W} distribution, but gives worse agreement for $E_{\mathrm{T}}^{\mathrm{miss}}$, whilst fitting on $E_{\mathrm{T}}^{\mathrm{miss}}$ gives worse agreement for m_{T}^{W}. The MJ events

[2]One of the findings was that $\Delta\phi(\ell, E_{\mathrm{T}}^{\mathrm{miss}})$ is poorly modelled, subsequently affecting the modelling of m_{T}^{W} . This is shown in Appendix C, Figure C.2.

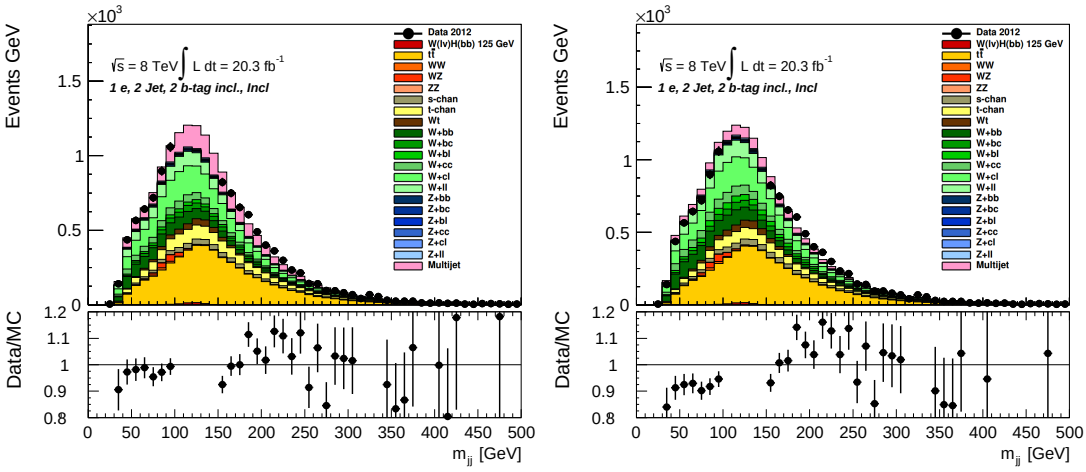
are concentrated at lower values of $E_{\mathrm{T}}^{\mathrm{miss}}$, whilst it is more broadly distributed in m_{T}^{W}. This allows for a greater separation of the MJ contribution from the other backgrounds, as can be seen in Figure 5.9. Performing the fit on m_{T}^{W} tends to underestimate the MJ contribution, especially in the lower m_{T}^{W} region. A comparison of the effect on the dijet mass distribution when fitting on $E_{\mathrm{T}}^{\mathrm{miss}}$ and m_{T}^{W} is shown in Figure 5.10. The better shape of $E_{\mathrm{T}}^{\mathrm{miss}}$ variable, and the correspondingly lower uncertainties, motivated the choice of this being the variable on which to perform the fit.

The inclusive 2-tag fitting procedure assumes that the 2L, 2M and 2T regions have the correct ratio of events in the inclusive region. This has been shown to be a good assumption for the MJ background, but not for the EW background. The muon channel has a small MJ contribution, and illustrates the incorrect ratios in EW background in Figure 5.11 (a) and (c). However, this adds a lot of complexity to the analysis, so the exclusive fits are only used for systematics. In addition, the low event count in the 2T region affects the exclusive fit and the normalisation, particularly for MJ. The effect of fitting to the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution both inclusively and exclusively is shown for p_{T}^{ℓ} in Figure 5.11 (b) and (d). Fitting before and after the H_{T} cut gives similar results and as such, the original choice of fitting after the H_{T} cut was kept.



(a) **(b)**

Figure 5.9 *Distributions of E_T^{miss} and m_T^W are shown in (a) and (b), respectively. These are both fitted to data using the inclusive 2 b-tag region, after the H_T cut. The MJ contribution is concentrated towards lower E_T^{miss}, with a much broader distribution in m_T^W .*



(a) **(b)**

Figure 5.10 *The invariant mass of the two jets in the event, m_{bb}, when fitting on E_T^{miss} (a) and fitting on m_T^W (b). The signal region is not shown in the data to keep the analysis blind.*

Fitted Distributions

The final fit which provides the normalisation of the MJ contribution in the signal region is performed on the $E_{\mathrm{T}}^{\mathrm{miss}}$ distribution using the inclusive 2 b-tagging region for all 2L/2M/2T b-tag regions. The $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions for some tagging regions are shown in Figure 5.12. The $E_{\mathrm{T}}^{\mathrm{miss}}$ variable is also used for the fitting procedure in the 1 b-tag region. Scale factors (see Equation 5.2) which are applied to the MJ yields, obtained in the MJ enriched template region, are shown in Table 5.9. These scale factors are used to extrapolate the MJ yield into the signal region.

The MJ background in the $p_{\mathrm{T}}^W > 120$ GeV bins in the electron channel is found to be approximately 3% and 2% for the 2L and 2T categories, respectively. For the muon channel, the contributions from the MJ background in the signal region with $p_{\mathrm{T}}^W < 90$ and $90 \leq p_{\mathrm{T}}^W < 120$ are between 2% and 7%, respectively, dropping to negligible levels at higher p_{T}^W. The relative MJ background contributions in the inclusive 2 b-tag region, before the global likelihood fit has been performed, are detailed in Table 5.8 and after the global likelihood fit in Table 5.14.

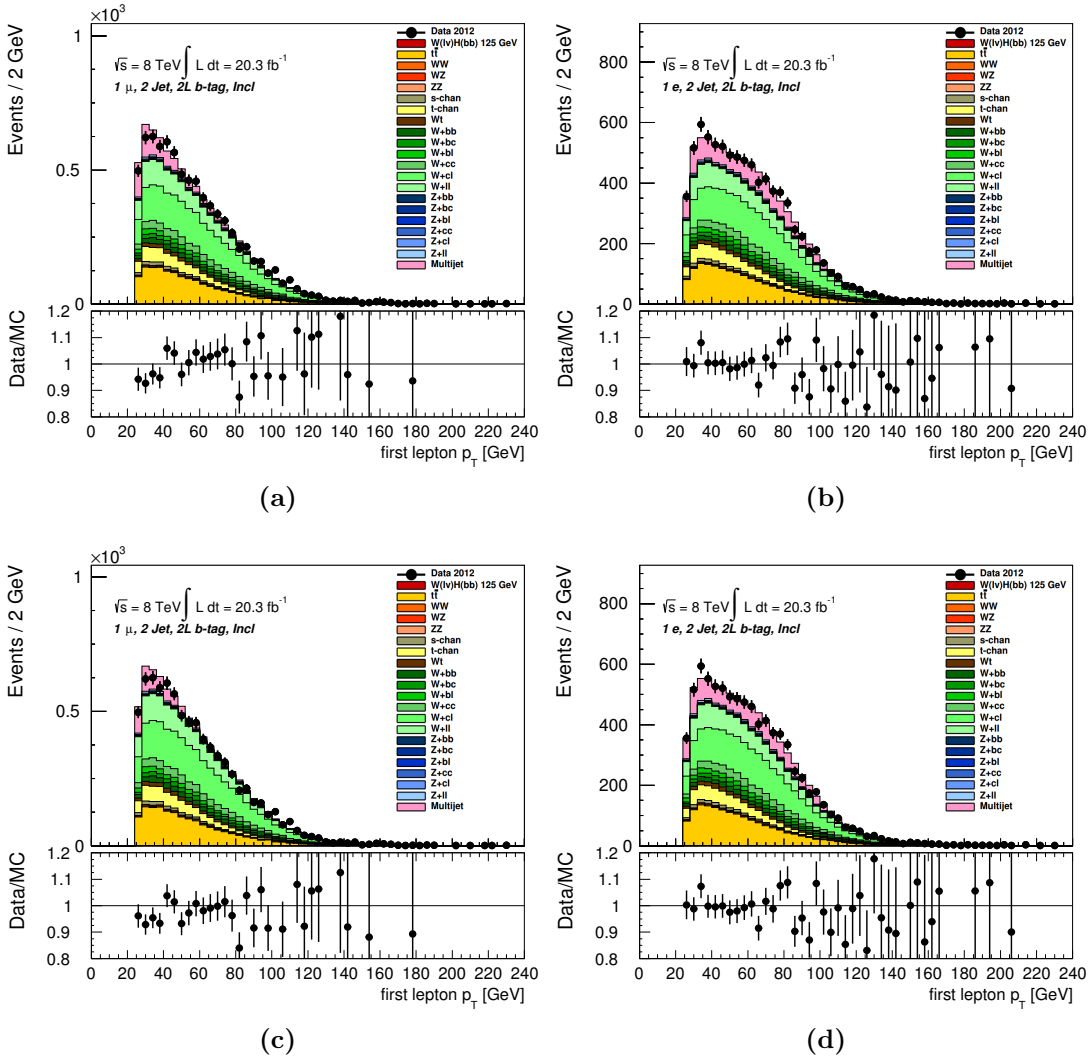


Figure 5.11 *Distributions of the muon (left) and electron (right) p_T fitted to E_T^{miss} and using exclusive fitting (top) and inclusive fitting (bottom) in the 2L b-tag region.*

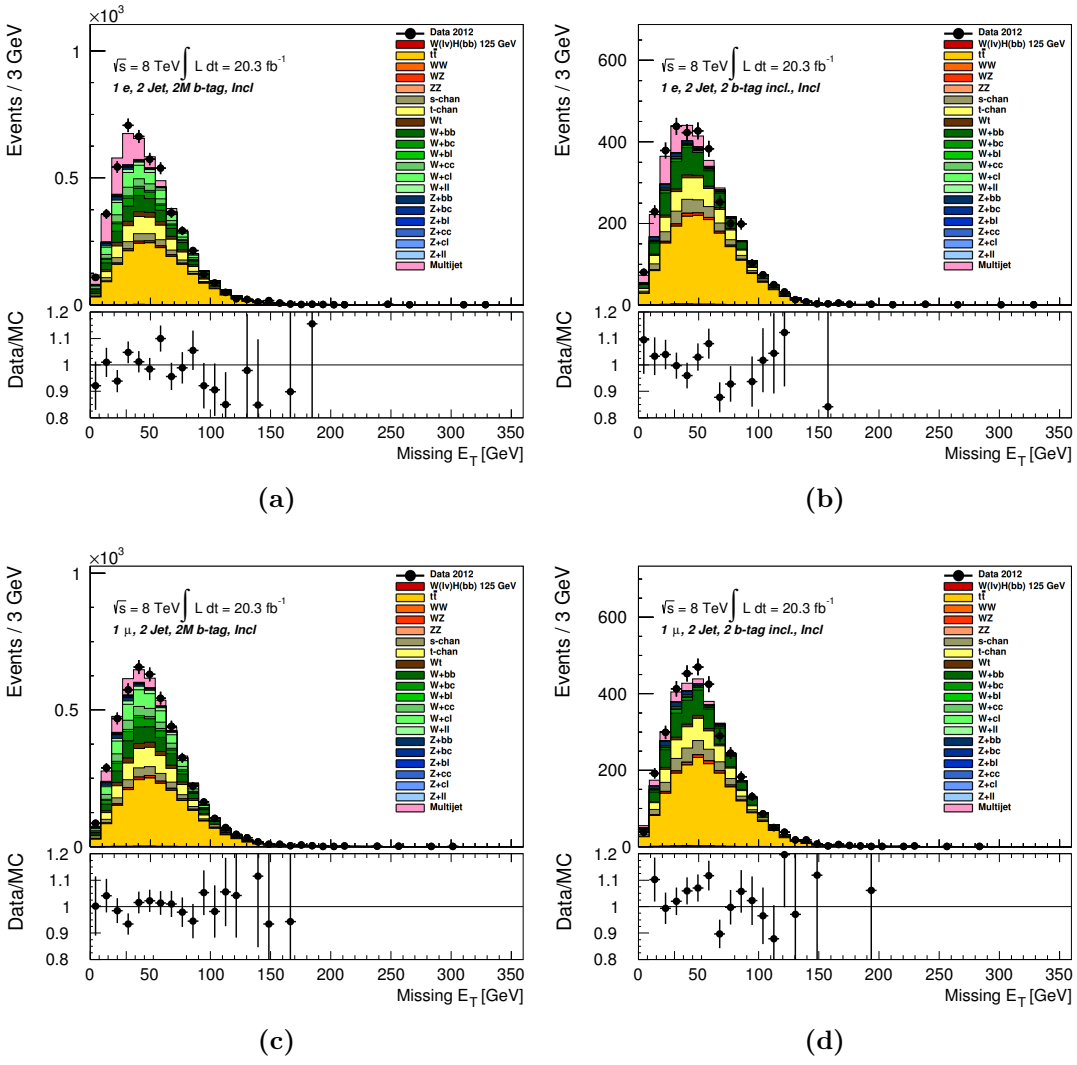


Figure 5.12 *Distributions of E_T^{miss} used for the MJ fitting in the electron ((a) and (b)) muon channel ((c) and (d)). The E_T^{miss} is shown in the 2M (left) and 2T (right) b-tag regions, scaled by the factor found when fitting in the inclusive 2 b-tag region.*

2 b-tag inclusive (2L+2M+2T) MJ prefit contributions					
Region	$p_{\mathrm{T}}^{V} < 90$	$90 \leq p_{\mathrm{T}}^{V} < 120$	$120 \leq p_{\mathrm{T}}^{V} < 160$	$160 \leq p_{\mathrm{T}}^{V} < 200$	$p_{\mathrm{T}}^{V} \geq 200$
	2-jets				
MJ (el)	-	-	3.22%	2.82%	1.08%
MJ (μ)	7.60%	2.71%	0.25%	0.15%	0.00%
	3-jets				
MJ (el)	-	-	0.90%	0.92%	0.37%
MJ (μ)	2.36%	0.74%	0.05%	0.16%	0.00%

Table 5.8 *The estimated prefit percentage of background events in the 2 b-tag inclusive region (2L+2M+2T) from multi-jet events. These are based on the event yields that are provided in Appendix F. These results are shown postfit in Table 5.14.*

Region	Electron Channel (Stat. Error)		Muon Channel (Stat. Error)	
	MJ	EW	MJ	EW
0 b tag	0.942 ± 0.009	1.029 ± 0.002	1.2 ± 0.1	1.059 ± 0.002
1 b tag	0.88 ± 0.02	1.021 ± 0.004	1.1 ± 0.1	1.034 ± 0.005
2L b tag	0.079 ± 0.003	1.02 ± 0.01	0.19 ± 0.02	1.01 ± 0.01
2M b tag	0.08 ± 0.01	1.03 ± 0.02	0.13 ± 0.05	1.07 ± 0.03
2T b tag	0.096 ± 0.007	1.0160 ± 0.0003	0.15 ± 0.09	1.08 ± 0.03
2 b tag incl.	0.079 ± 0.005	1.03 ± 0.10	0.16 ± 0.01	1.0514 ± 0.0002

Table 5.9 *Multi-jet scale factors obtained in different tagging regions for the electron and muon channel when fitting to the E_T^{miss} distribution. The final MJ fit is taken from the inclusive (loose, medium and tight) 2 b-tag region. The error quoted is the statistical error from the fit.*

5.8 Systematic Uncertainties

The main uncertainties to the analysis are discussed in the following sections. These are divided into experimental uncertainties, multi-jet background uncertainties and Monte Carlo simulation uncertainties. The Monte Carlo simulation uncertainties include those on all electroweak backgrounds, the $t\bar{t}$ background, and the signal. For all the backgrounds, including the multi-jet background, uncertainties are considered that affect the shape and the normalisation of the background m_{bb} distribution. Shape uncertainties are assessed in each bin of a distribution (m_{bb} or MV1c, as will be discussed in Section 5.9) as the difference between the systematic variation and the nominal value. The normalisation uncertainties provide an error on the event yield. The uncertainties are included as nuisance parameters in the global likelihood fit, constrained by Gaussian or log-normal prior probabibility distributions (referred to as *priors*). The global likelihood fit provides an overall systematic and statistical error that is determined as a combination of all input uncertainties, having been adjusted and constrained by the fit.

5.8.1 Experimental Uncertainties

The largest uncertainties affecting the dijet mass come from the jet energy scale (JES) and heavy flavour tagging. There are additional uncertainties that affect trigger selection, object reconstruction and identification, and the subsequent energy and momentum calibrations and resolutions. These are discussed below.

The uncertainties on the corrections for the trigger, reconstruction, identification and isolation efficiencies, and energy resolution when applied to the electron and muon candidates are less than 1% and therefore negligible.

The uncertainty on the jet energy scale (JES) depends on a number of independent factors, including the flavour composition of the jet, pile-up corrections and calibration analyses [130]. Overall, there are 24 components to account for these sources. For central jets, the total error ranges from $\approx 3\%$ for jets with $p_{\mathrm{T}} = 20$ GeV, to $\approx 1\%$ for jets with $p_{\mathrm{T}} = 1$ TeV. A further correction to the b jet energy calibration of $1-2\%$ is also applied. The jet energy resolution is considered independently for b jets and non-b-jets [134]. These uncertainties are η and p_{T} dependent, and can have a large effect, especially for lower p_{T} jets. It

ranges from $10-20\%$ for $p_{\mathrm{T}} = 20$ GeV jets to 5% for $p_{\mathrm{T}} > 200$ GeV.

The scale factors for the b-tagging efficiency for the b jets are ≈ 1 with uncertainties of between 2% ($p_{\mathrm{T}} \leq 200$ GeV) and 8% ($p_{\mathrm{T}} > 200$ GeV). Half of the correction applied at low ΔR in the Wcc samples is also used as a systematic uncertainty.

The integrated luminosity has an uncertainty of 2.8%, and there is a 4% uncertainty on the number of interactions per bunch crossing [167].

5.8.2 Multi-Jet Systematic Uncertainties

Systematic uncertainties on the shape and normalisation of the MJ contribution are calculated separately. The general philosophy taken in calculating these uncertainties is as follows: for the shape uncertainty, the MJ template is changed by varying the isolation requirements in the MJ template. The statistical error is represented by the uncertainty from the MJ scale factor fit, provided by the Minuit χ^2 fit. For the 2 b-tag region, the uncertainties are calculated in each of the 2L/2M/2T regions separately by using the statistical error from the exclusive fit (fitting in 2L/2M/2T as opposed to 2L+2M+2T). The other simulated backgrounds used in the MJ calculation are scaled with scale factors from the global likelihood fit; a further systematic is obtained by removing these scale factors and observing the change in the MJ yield.

The change in m_{bb} when changing the track and calorimeter isolation can be seen in Figure 5.13. The difference in each bin between the nominal shape and the variation is used to define the shape uncertainty. The track isolation selection in the electron channel is changed to $0.12 < iso_{track} < 0.5$ and the effect is symmetrised about the nominal shape to give the *up* and *down* systematics. In the muon channel, the isolation cut is chosen to separate the events in the MJ template into two regions of roughly the same number of events; the upward variation is found using $0.07 < iso_{track} < 0.095$, and the downward variation by using $0.095 < iso_{track} < 0.5$. In the electron channel, the calorimeter isolation requirement is changed to $iso_{calo} < 0.04$ for the downward variation and $0.04 < iso_{calo} < 0.07$ for the upward variation. In the muon channel, changes in the calorimeter and track isolation requirements show the same variations; no additional uncertainty is added for the calorimeter isolation change.

Since the MJ scale factor is fitted for each shape variation, there is an

accompanying normalisation change and a migration of events between the 2L/2M/2T regions. Each of the fits in the 1 b-tag and inclusive 2 b-tag regions has a statistical error provided by the Minuit χ^2 fit. Since there is a migration of events (and thus jet flavour) between b-tag regions, the statistical errors on the heavy flavour contribution in each of the 2 b-tag regions are found by performing the fit in the 2L/2M/2T regions exclusively[3] and taking the statistical error from the Minuit χ^2 fit. A second uncertainty is found by comparing the yields when applying and not applying the EW scale factors that come from the profile likelihood fit. The normalisation uncertainty is calculated by adding the statistical errors in quadrature with the change in yields after changing the EW scale factor. The statistical and EW uncertainties are detailed in Table 5.10. In general, the errors in the muon channel are quite high due to the relatively small number of events.

There is an uncertainty introduced when reweighting events that have been promoted from 1-tag to 2-tag events due to the $\Delta\phi(j_1, j_2)$ and p_{T}^{W} mismodelling described in Section 5.6.1. These are considered as shape uncertainties.

The calculated normalisation and shape uncertainties are listed in Table 5.11. The MJ contributions in the 2 b-tag 2 jet region are larger than those in the 3-jet region, and generally have larger uncertainties. In the 2 b-tag 2 jet region for the electron channel, the MJ contribution is no more than 3% in any p_{T}^{V} bin, with an uncertainty on the normalisation of between 5-22%, depending on the p_{T}^{V} bin. In the muon channel, there is a sizeable contribution to the total background in the lowest p_{T}^{V} bin of up to 7%, however, the contribution to the background at $p_{\mathrm{T}}^{V} > 120$ GeV is negligible. There is an uncertainty of between 12 and 60% in the 2 b-tag 2 jet region. Given the fairly low contribution to the total background, these uncertainties do not have a large effect on the sensitivity of the analysis.

5.8.3 Monte Carlo Simulation Uncertainties

The uncertainties for the MC simulation of the following processes are assessed separately: the $t\bar{t}$, single top, V+ jets, and diboson backgrounds, and the Higgs boson signal. The calculation of the systematic effects is focused on the observables that are used in the global likelihood fit, in particular p_{T}^{V}, the jet flavour and the dijet mass. For the background, where possible, the uncertainty

[3] 3-jet region is done inclusively (2L+2M+2T) as there are a relatively small number of events in each of the exclusive 2L/2M/2T regions.

Channel	Region	Stat error (%)	EW error (%)	Total (%)
e 2-jet	1-tag	2	1	3
	2-tag 2L/2M/2T	10/10/20	5/10/10	11/14/22
e 3-jet	1-tag	3.5	1	4
	2-tag incl.	5	3	6
μ 2-jet	1-tag	11	5	12
	2-tag 2L/2M/2T	25/25/45	12/34/40	28/42/60
μ 3-jet	1-tag	7	8	11
	2-tag incl.	5	13	14

Table 5.10 *A summary of the normalisation uncertainties on the multi-jet background estimation. The statistical error is obtained from the fit and the EW error is given by the relative changes in yields when removing the EW scale factors in the fit. These are added in quadrature to obtain the normalisation uncertainty.*

Systematic	Description	Region	Value (%)
SysMJMuNorm	Normalisation μ channel	2 jet 1/2L/2M/2T-tag	12/28/42/60
		3 jet 1/2	11/14
SysMJElNorm	Normalisation el channel	2 jet 1/2L/2M/2T-tag	3/11/14/22
		3 jet 1/2	4/6
SysMJMuTrkIso	iso_{track} changed	2/3-jet 1/2-tag	Shape
SysMJElTrkIso	iso_{track} changed	2/3-jet 1/2-tag	Shape
SysMJElCaloIso	iso_{calo} changed	2-jet 1/2-tag	Shape
SysMJDR	Spoof reweighting (el only)	2-jet 2-tag	Shape
SysMJPtV	Spoof reweighting (el only)	2-jet 2-tag	Shape

Table 5.11 *Summary of the MJ systematics. The normalisation uncertainties are given separately for the electron and muon channels and split into 1-tag, 2L, 2M, and 2T for the 2-jet region, and 1-tag and the inclusive 2-tag selection in the 3-jet region. The shape uncertainties are found in the 1-tag and inclusive 2-tag selections for all regions. iso_{track} is varied in both channels, and iso_{calo} is varied for the electron channel. Shape uncertainties are included from the truth tagging reweighting applied in the electron channel. Half of the reweighting value of $\Delta\phi(j_1, j_2)$ and p_T used in the reweighting is used as a systematic.*

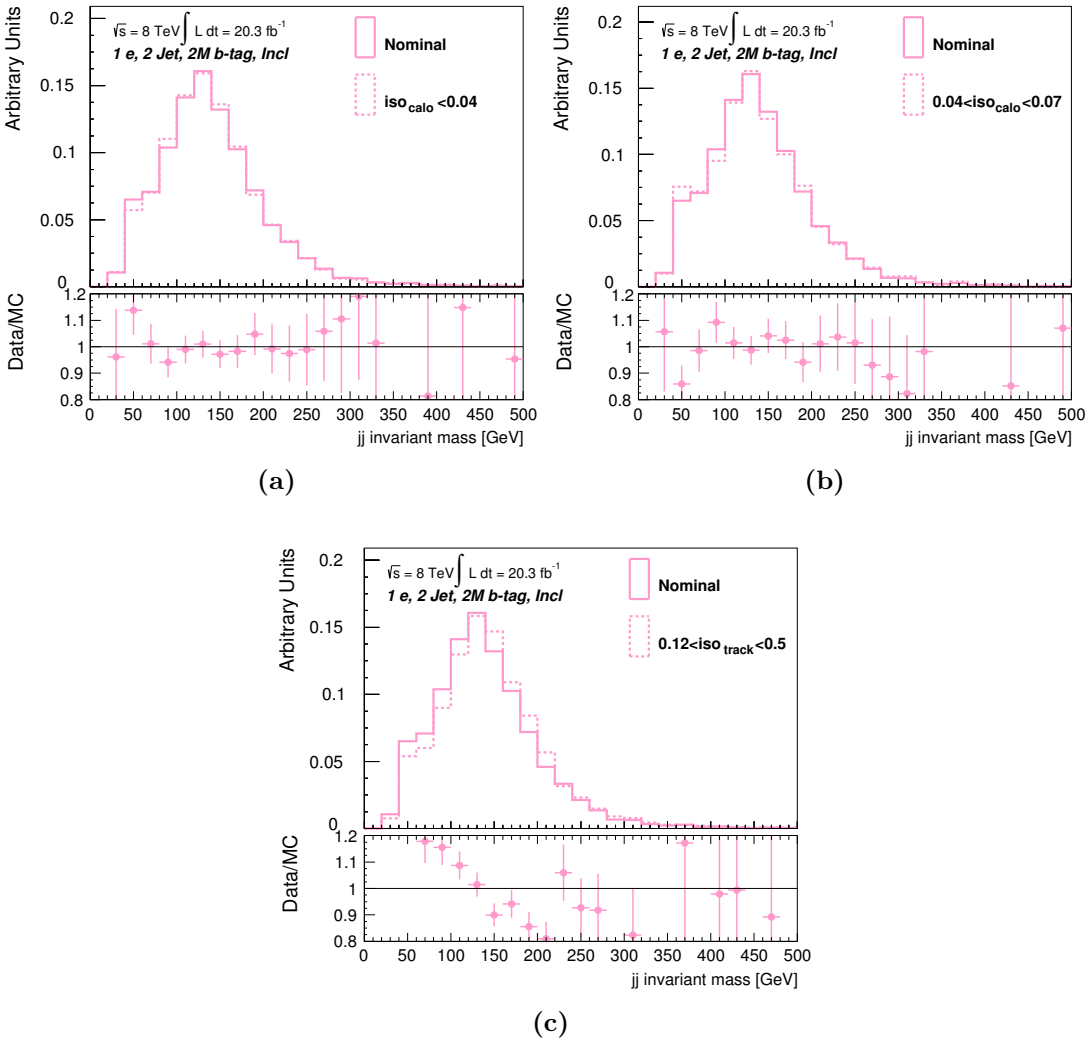


Figure 5.13 *Changes in the dijet mass distribution, m_{bb}, from changes to the track isolation and the calorimeter isolation in the MJ template used for systematics. In these plots, the solid pink lines show the nominal selection, and the dotted lines show the systematic variations. Plot (a) is the result of changing the calorimeter isolation requirement to* < 0.04 *(downward systematic), (b) requires* $0.04 < iso_{calo} < 0.07$ *(upward systematic), and (c) has the track isolation requirement* $0.12 < iso_{track} < 0.5$.

is assessed using dedicated control regions in data, which are extrapolated into the signal region. If this is not possible, then different MC generators are compared with the nominal ones. A brief overview of each background is given here with a full description available in Reference [1].

***t*$\bar{t}$**: Following the discussion in Section 5.6.2, an uncertainty of half of the p_{T} correction is applied. The normalisation is found from the global likelihood fit, however, it is determined almost entirely by the large top contribution in the 3-jet

region of the 1-lepton analysis. An uncertainty is applied for the extrapolation of the normalisation from the 3-jet to 2-jet signal region, calculated by comparing the results from multiple MC generators. A shape systematic is added to the m_{bb} and p_{T}^{V} distributions by considering the 3-jet to 2-jet ratio.

Single top: Theoretical cross section uncertainties across the different production channels have uncertainties between 4% and 7% [168]. Different generators are used to quantify the uncertainties on the m_{bb} distribution, p_{T}^{V}, and the 3-jet to 2-jet ratio.

V+jets: Similar methods for calculating the uncertainties are used for both Z+jets and W+jets. Uncertainties are included to account for the $\Delta\phi(j_1, j_2)$ and p_{T}^{V} corrections discussed in Section 5.6.1. The normalisation and 2/3-jet ratio for the light jet contribution is taken from simulation, and for the heavier flavours it is left to float in the final fit. It is not possible to define a control region in data that is dominated by $W + bb$; instead generator level comparisons are done to find uncertainty estimates. The uncertainty on the shape of the dijet mass distribution is taken from the side band region in data for Z+jets and generator level comparisons for W+jets.

Diboson: The cross section uncertainty is estimated by varying the parameters used for the simulation at the parton level. Differences between generators are also assessed. Uncertainties in p_{T}^{V} come from the comparison of LO and NLO prediction, and similarly for the 3-jet to 2-jet ratio.

Signal: The uncertainties in the signal processes are all obtained from MC. They are all calculated for the $m_H = 125$ GeV mass point, and these are used for all mass points considered in the limit setting. The cross section normalisations are taken from the theoretical predictions at NLO (the signal simulated is discussed in Section 5.5). An uncertainty on the branching ratio and the cross section are considered [156, 169].

The systematic uncertainties for both signal and background are summarised in Table 5.12. The systematic uncertainties for each process are included as nuisance parameters in the final likelihood fit. This is discussed in Section 5.9.

WH Signal ($m_H = 125$ GeV)	
Cross section (scale)	1% ($q\bar{q}$), 50% (gg)
Cross section (PDF)	2.4% ($q\bar{q}$), 17% (gg)
Branching ratio	3.3 %
Acceptance (scale)	1.5%–3.3%
3-jet acceptance (scale)	3.3%–4.2%
p_{T}^{V} shape (scale)	S
Acceptance (PDF)	2%–5%
p_{T}^{V} shape (NLO EW correction)	S
Acceptance (parton shower)	8%–13%
Z+jets	
Zl normalisation, 3/2-jet ratio	5%
Zcl 3/2-jet ratio	26%
Z+hf 3/2-jet ratio	20%
Z+hf/Zbb ratio	12%
$\Delta\phi(j_1, j_2)$, p_{T}^{V}, m_{bb}	S
W+jets	
Wl normalisation, 3/2-jet ratio	10%
Wcl, W+hf 3/2-jet ratio	10%
Wbl/Wbb ratio	35%
Wbc/Wbb, Wcc/Wbb ratio	12%
$\Delta\phi(j_1, j_2)$, p_{T}^{V}, m_{bb}	S
$t\bar{t}$	
3/2-jet ratio	20%
High/low-p_{T}^{V} ratio	7.5%
Top-quark p_{T}, m_{bb}, $E_{\mathrm{T}}^{\mathrm{miss}}$	S
Single top	
Cross section	4% (s-,t-channel), 7% (Wt)
Acceptance (generator)	3%–52%
m_{bb}, $p_{\mathrm{T}}^{b_2}$	S
Diboson	
Cross section and acceptance (scale)	3%–29%
Cross section and acceptance (PDF)	2%–4%
m_{bb}	S
Multi-jet	
0-, 2-lepton channels normalisation	100%
1-lepton channel normalisation	2%–60%
Template variations, reweighting	S

Table 5.12 *Summary of the systematic uncertainties on the signal and background modelling. S indicates that only a shape uncertainty is assessed.*

5.9 Statistical Analysis

A binned maximum likelihood estimate is used for the statistical model in this analysis. One of the primary results obtained from the fitting procedure is the signal strength parameter, μ, a multiplicative factor of the expected signal yield. The signal yield is normalised to the expected Standard Model Higgs boson cross section, so any change from $\mu = 1$ indicates a deviation from the SM, and $\mu = 1$ indicates perfect agreement with the presence of a SM Higgs boson. The fit also constrains the systematic uncertainties and finds the background normalisations. Three forms of hypothesis testing are employed: exclusion limit setting, discovery testing and the signal measurement.

In this analysis, the parameters in the likelihood function $\mathcal{L}(\boldsymbol{\alpha}; \boldsymbol{x})$ are separated into $\boldsymbol{\alpha} = (\mu, \boldsymbol{\theta})$, where μ is the signal strength parameter that multiplies the expected signal yield, and $\boldsymbol{\theta}$ represents nuisance parameters (NP) from uncertainties in the analysis.

As this is a counting experiment, the observed rate of signal and background events is assumed to follow a Poisson distribution, which is incorporated into the likelihood function as an overall term of $\text{Pois}(n|\mu S(\boldsymbol{\theta}; \boldsymbol{x}) + B(\boldsymbol{\theta}; \boldsymbol{x}))$. Here $S(\boldsymbol{\theta}; \boldsymbol{x})$ and $B(\boldsymbol{\theta}; \boldsymbol{x})$ represent the total signal and background event yields as functions of the NPs.

The NPs are constrained by Gaussian prior probability functions, or *priors*, with the mean given by the expected value of the NP, and the width given by its uncertainty. For the normalisation uncertainties, a log-normal prior is used to maintain a positive probability density function. Some NPs have been determined from other datasets and are not completely unknown. This knowledge is added to the likelihood as an auxiliary measurement that increases as the NP is shifted from the nominal value; a penalty term. Formally, this is given as $\text{Pois}(m|B_{CR})$, where m and B_{CR} indicate the measured and expected event yield in the corresponding dataset from which it was estimated.

The combination of the signal and background probability density function for each bin, the overall Poisson distribution, and the penalty term leads to the

binned likelihood function of

$$\mathcal{L}(\mu, \boldsymbol{\theta}; \boldsymbol{x}) = \text{Pois}(n|\mu S(\boldsymbol{\theta}; \boldsymbol{x}) + B(\boldsymbol{\theta}; \boldsymbol{x})) \times \left[\prod_{b \in \text{bins}}^{N} \frac{\mu f(\boldsymbol{\theta}; \boldsymbol{x})_b^{sig} + f(\boldsymbol{\theta}; \boldsymbol{x})_b^{bkg}}{\mu S(\boldsymbol{\theta}; \boldsymbol{x}) + B(\boldsymbol{\theta}; \boldsymbol{x})} \right] \text{Pois}(m|B_{CR}). \tag{5.3}$$

The nominal value for μ is found by maximising the log-likelihood estimate (MLE), $\ln\mathcal{L}(\mu, \boldsymbol{\theta}); \boldsymbol{x}$, for all $\boldsymbol{\theta}$ using the Minuit minimisation program [165].

The inputs to the likelihood function for the analysis are taken from the expected background-only distributions of the invariant mass of the two signal region b jets in both the 2-tag (2- and 3-jet) and 1-tag regions. As the flavour composition is not well modelled in the W+jets background, the MV1c distribution of the single tagged jet in the 1-tag region is also included in the fit, however, only two p_{T}^V bins are used: $p_{\text{T}}^V < 120$ GeV and $p_{\text{T}}^V > 120$ GeV. In total, this comes to 81 2-tag regions in the fit, and 11 1-tag control regions.

The binning of the dijet mass distribution is optimised for every Higgs boson mass point to increase the sensitivity, as can be seen in Figure 5.16. In the tail regions, wider bins are used to reduce statistical uncertainties, and in the signal region sensitivity is increased by reducing the bin widths. The transformation applied to the distributions can be found in Reference [1].

The normalisation parameters for the $t\bar{t}$ and W/Z+jets backgrounds are left free to float in a preliminary global likelihood fit. The difference in yields before and after the preliminary fit is used to provide scale factors, as shown in Table 5.13, which are then applied to correct the normalisation, and subsequent fits use these corrected normalisations. The postfit multi-jet background contribution in the signal region, as a percentage of the total background, is given in Table 5.14.

The effects of the systematic uncertainties on the signal strength parameter are shown in Figures 5.14 and 5.15. In these figures, the systematic uncertainties are listed in descending order of their impact on $\hat{\mu}$. The pull of each systematic uncertainty is given by the deviation of the fitted nuisance parameters $\hat{\theta}$ from their nominal values θ_0, given as the number of standard deviations with respect to their nominal uncertainties $\Delta\theta$: $(\hat{\theta} - \theta_0)/\Delta\theta$. The pull of the systematic uncertainties on $\hat{\mu}$ are shown by the filled circles, and the associated error bars show the postfit nuisance parameter uncertainties relative to their nominal uncertainties. The red open circles, and error bars, show the fitted values and uncertainties of the floating normalisation parameters in the fit, which have a

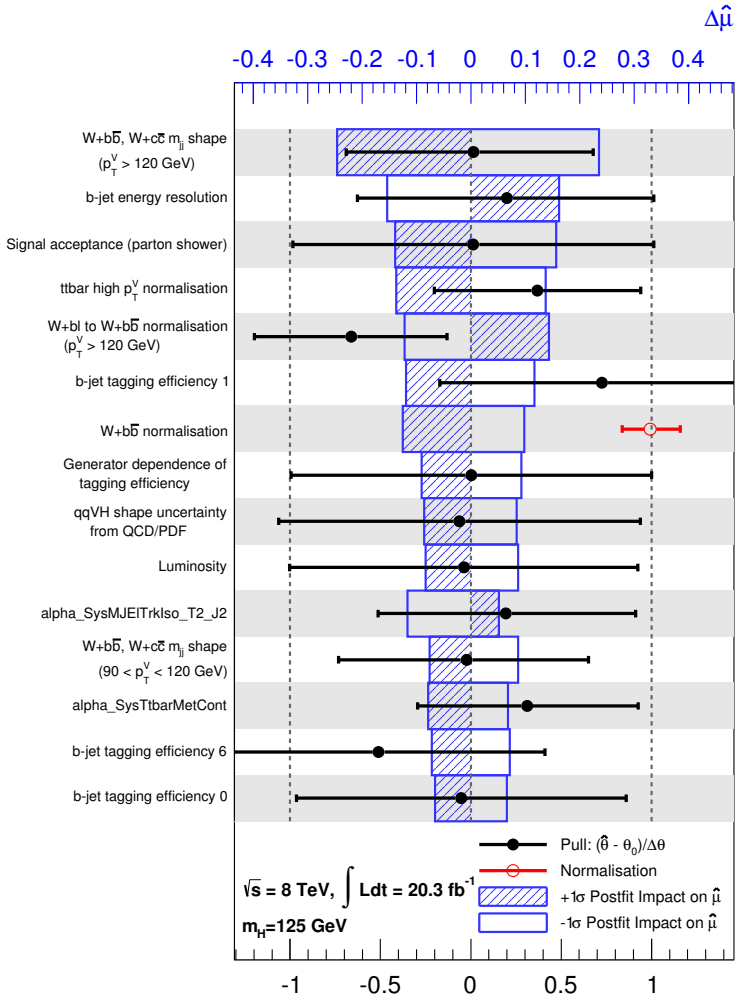


Figure 5.14 *Impact of the systematic uncertainties on the fitted signal-strength parameter $\hat{\mu}$ for the 1-lepton channel in the dijet mass analysis. The systematic uncertainties are listed in descending order of their impact on $\hat{\mu}$. The largest impact comes from the shape uncertainties on m_{jj} for the largest backgrounds, $W + b\bar{b}$ and $W + c\bar{c}$. The b-jet energy resolution and the parton showering also have a similar impact on $\hat{\mu}$. Normalisation uncertainties from the backgrounds that are left to float in the preliminary global likelihood fit have a similar impact to each other. The largest pull comes from the $W + bl$ to $W + b\bar{b}$ normalisation for $p_T^W > 120$ GeV, but the pulls are generally well behaved and a decrease is seen in the relative uncertainty on the nuisance parameters . The normalisation of the largest background, although it has one of the larger impacts on $\hat{\mu}$, has a normalisation factor of 1 with relatively small errors.*

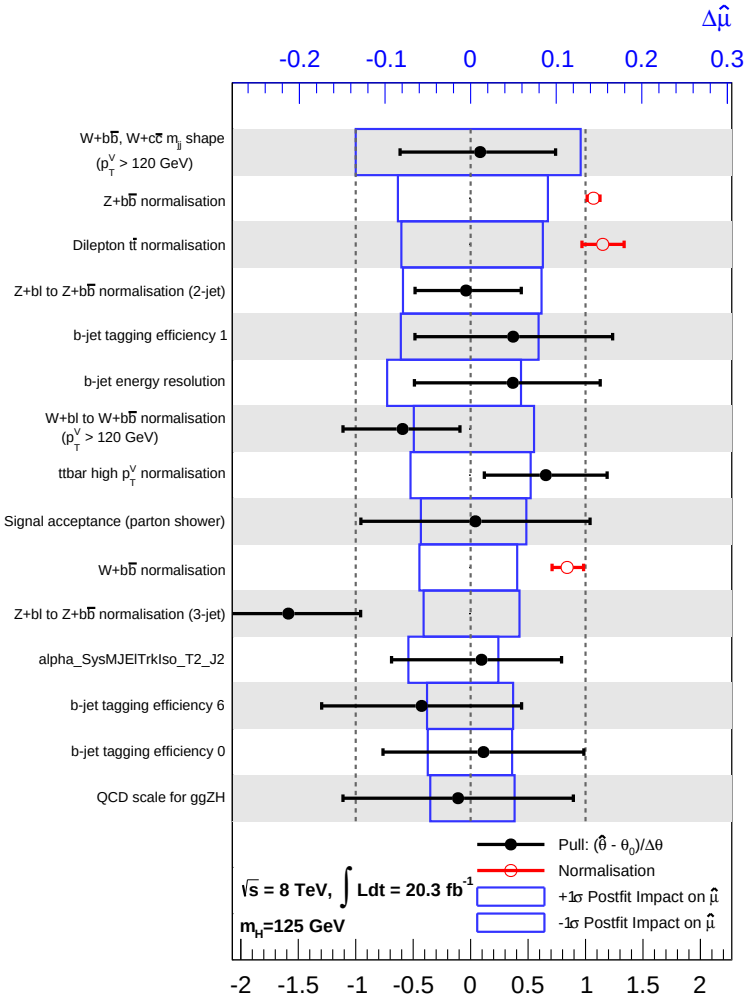


Figure 5.15 *Impact of the systematic uncertainties on the fitted signal-strength parameter $\hat{\mu}$ for the combined 0-, 1- and 2-lepton channel in the dijet mass analysis. The systematic uncertainties are listed in descending order of their impact on $\hat{\mu}$. As seen in the 1-lepton channel, the largest impact comes from the shape uncertainties on m_{jj} for the largest backgrounds, $W + b\bar{b}$ and $W + c\bar{c}$. However, the normalisations of the other backgrounds have a relatively larger impact on $\hat{\mu}$ than in the 1-lepton channel. This is expected since these backgrounds contribute more in the combined channel. The normalisation factors themselves are well behaved and are all found to be close to 1. The b-jet tagging efficiency and b-jet energy resolution have a relatively large impact on $\hat{\mu}$. The pulls are well behaved, however, the pull on the $Z + bl$ to $Z + b\bar{b}$ normalisation is large, although the impact on $\hat{\mu}$ is relatively small.*

Process	Scale factor
$t\bar{t}$ 0-lepton	1.36 ± 0.14
$t\bar{t}$ 1-lepton	1.12 ± 0.09
$t\bar{t}$ 2-lepton	0.99 ± 0.04
Wbb	0.83 ± 0.15
Wcl	1.14 ± 0.10
Zbb	1.09 ± 0.05
Zcl	0.88 ± 0.12

Table 5.13 *Normalisation scale factors obtained from minimising the likelihood that are applied to the $t\bar{t}$, Wbb, Wcl, Zbb, and Zcl backgrounds. These are obtained from the global likelihood fit to the 8* TeV *data using the MVA. Errors include statistical and systematic uncertainties.*

2 b-tag inclusive (2L+2M+2T) MJ postfit contributions					
p_{T}^{V} region	$p_{\mathrm{T}}^{V} < 90$	$90 \leq p_{\mathrm{T}}^{V} < 120$	$120 \leq p_{\mathrm{T}}^{V}160$	$160 \leq p_{\mathrm{T}}^{V}200$	$p_{\mathrm{T}}^{V} \geq 200$
			2-jets		
MJ (el)	-	-	2.83%	2.43%	1.04%
MJ (μ)	7.01%	2.53%	0.23%	0.15%	0.00%
			3-jets		
MJ (el)	-	-	0.87%	0.90%	0.37%
MJ (μ)	2.31%	0.94%	0.05%	0.16%	0.00%

Table 5.14 *The estimated postfit percentage of background events in the 2 b-tag inclusive region (2L+2M+2T) from multi-jet events. These are based on the event yields that are provided in Table 5.15 and Appendix F. The corresponding prefit estimates are shown in Table 5.8.*

prefit value of one. Both the pulls and normalisation uncertainties refer to the bottom x-axis.

The blue boxes in Figures 5.14 and 5.15, referring to the top x-axis, show the effect of the uncertainties in the nuisance parameters on $\hat{\mu}$. Individual nuisance parameters θ are fixed to their postfit value $\hat{\theta}$ and modified upwards (hatched boxes) or downwards (open boxes) by their postfit uncertainty, and subsequently repeating the fit with all other nuisance parameters floating freely.

As discussed above in Section 5.1, a crosscheck is performed to measure the

VZ diboson signal strength, μ_{VZ}. This factor changes the normalisation of the diboson contributions with respect to the SM expectation. Higgs boson production is included as a background here. The results of this are given in Section 5.10.5.

5.9.1 Input Distributions

The inputs for the dijet mass analysis are the MV1c distributions of the b-tagged jet in the 1-tag regions and the transformed m_{bb} distribution for the 2-tag regions. The transformed m_{bb} distribution in the 2-jet region for 2T b tags is shown in Figure 5.16 (a) and (b). The untransformed m_{bb} distribution in this region, which is not used as an input, but is a useful illustration of the effect of the fitting, is shown in Figure 5.16 (c) and (d). Further distributions in the other b-tagging regions can be found in Appendix D.

5.10 Results

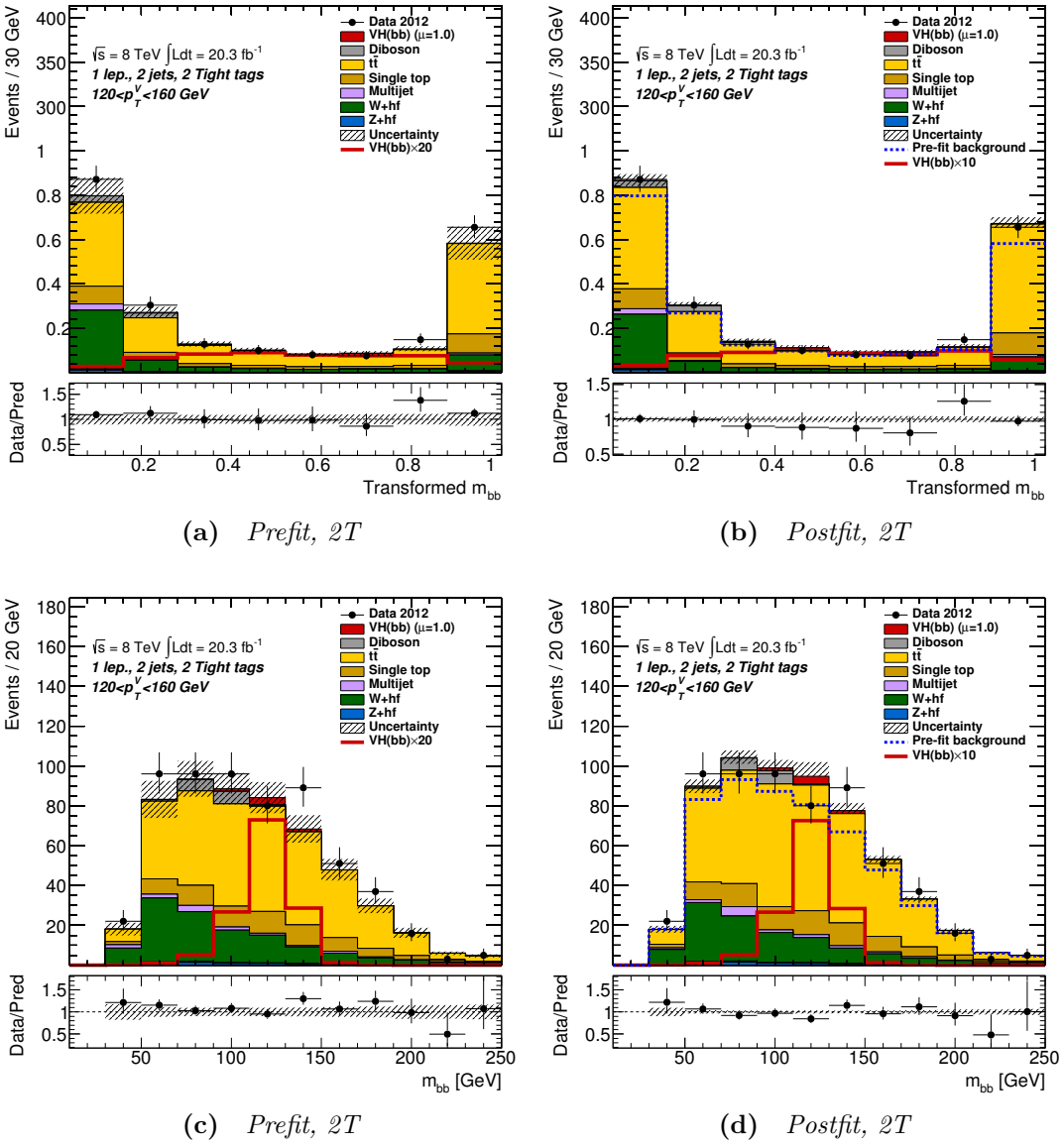
5.10.1 Introduction

Results are obtained using the maximum likelihood estimate fits to data discussed in Section 5.9. The statistical significance of any observed signal needs to be quantified in some way. The sensitivity of the result is examined by looking at the expected significance for a number of hypotheses.

The hypothesis testing is done using a test statistic which characterises the full dataset as a function of a number of parameters. This allows for compatibility measurements between the background-only hypothesis, H_0, and the observed data. Test statistics used here are based on the profile likelihood ratio $\Lambda(\mu)$, where μ is the signal strength. The ratio is defined as

$$\Lambda(\mu) = (\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}}_\mu)/\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})), \tag{5.4}$$

where $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ maximise the likelihood (such that $0 \leq \hat{\mu} \leq \mu$), and $\hat{\hat{\boldsymbol{\theta}}}_\mu$ are the NP values that maximize the likelihood for a fixed μ (conditional maximum likelihood estimate). A value of $\Lambda(\hat{\mu}) = 1$ gives perfect agreement with the data, and the NPs determine the shape. $\hat{\mu}$ is often restricted to be positive as in the case that



(a) *Prefit, 2T* **(b)** *Postfit, 2T*

(c) *Prefit, 2T* **(d)** *Postfit, 2T*

Figure 5.16 *The transformed (top) and untransformed (bottom) m_{bb} distribution in the 2-jet, 2-tag, $120 < p_T^V < 160$ GeV bin with two tight-tagged b jets in the 1-lepton channel before (left) and after (right) the fit is performed.*

it becomes negative the MLE is effectively $\mu = 0$. Variations of the test statistic are used for discovery, measurement and exclusion, which are discussed below.

The null-hypothesis, or background-only hypothesis, H_0, describes the known processes. A signal hypothesis, H_1, describes H_0 in addition to a signal process, for example, the Higgs boson. The compatibility of the observed data with either hypothesis is given by the p-value: the probability of observing the data with equal or greater incompatibility than expected under the null-hypothesis. The p-

value can be converted into an equivalent significance, Z, which is the probability of rejecting H_0 if it is true. A p-value is computed for each test statistic.

The test statistic used for discovery is

$$q_0 = \begin{cases} -2\ln\Lambda(0) & \text{if } \hat{\mu} \geq 0 \\ 0 & \text{if } \hat{\mu} < 0. \end{cases} \tag{5.5}$$

Under this scenario, q_0 will show any disagreement with the background-only hypothesis if there are upward fluctuations in the data. The corresponding p-value, showing the disagreement with the null-hypothesis, is given by

$$p_0 = \int_{q_{0,\text{obs}}}^{\text{inf}} f(q_0|0, \hat{\hat{\theta}}_0) dq_0, \tag{5.6}$$

where $f(q_0|\mu, \hat{\hat{\theta}}_0)$ is the probability density function of q_0 under the assumption of no signal ($\mu = 0$) and $q_{0,\text{obs}}$ is the observed value of q_0. In particle physics, the convention for discovery (the rejection of H_0) requires a significance of $Z = 5$, or 5σ, where σ is one standard deviation, corresponding to $p = 2.87 \times 10^{-7}$.

An upper limit on the signal strength is found using a single-sided test statistic which assumes a null-hypothesis of signal plus background

$$q_\mu = \begin{cases} -2\ln\Lambda(\mu) & \text{if } \hat{\mu} \leq \mu \\ 0 & \text{if } \hat{\mu} > \mu. \end{cases} \tag{5.7}$$

The corresponding p-value of this test statistic is given by

$$p_\mu = \int_{q_{\mu,\text{obs}}}^{\text{inf}} f(q_{\mu,obs}|\mu, \hat{\hat{\theta}}_\mu) dq_\mu, \tag{5.8}$$

where $f(q_\mu|\mu, \hat{\hat{\theta}}_\mu)$ is the probability density function of q_μ under the assumption of signal strength μ and $q_{\mu,obs}$ is the observed value of q_μ. A 95% confidence level upper limit is found for μ where it satisfies $p_\mu = 0.05$.

However, if there is a downward fluctuation in the data, models with little expected sensitivity can be excluded incorrectly. At ATLAS a modified version of this is used, the CL_s method [170]. Here, rather than p_μ, the p-value used is

$$p'_\mu = \frac{p_\mu}{1 - p_b}, \tag{5.9}$$

where p_b refers to the p-value for the background-only hypothesis ($\mu = 0$). Incompatibilies with the null-hypothesis (signal+background) are thus down-weighted if there is also an incompatibility with the background-only hypothesis. A Standard Model Higgs boson would be excluded at the 95% confidence level if $\mu < 1$ with $p'_\mu - 0.05$.

For measurement of the observed best-fit signal strength $\hat{\mu}$ a double-sided test statistic is used, under the assumption of a null-hypothesis that includes signal plus background:

$$t = -2 \ln \Lambda(\mu). \tag{5.10}$$

. Here, no restraints are placed on the signal strength. The corresponding p-value gives the compatibility of the data with the signal-plus-background null-hypothesis. The measurement sensitivity is evaluated by finding the difference in μ between the nominal value $\hat{\mu}$ and the signal-plus-background hypothesis with $\mu = 1$.

Expected results are obtained in a similar way with the data replaced by the expectations from simulation, with all NPs set to the best fit values, also known as the Asimov dataset [171].

The dijet mass analysis has an expected measurement sensitivity of 1.9σ for the signal-plus-background hypothesis which includes a 125 GeV Higgs boson, compared with the MVA-based analysis which has a higher expected sensitivity of 2.5σ, which was the motivation for using the MVA as the main ATLAS analysis in Reference [1], with the dijet mass analysis used as a crosscheck. All the following results refer to a Higgs boson mass of 125 GeV, unless otherwise specified.

5.10.2 Event Yields

The event yields in data and from Monte Carlo for the 2- and 3-jet regions with 2T b tags are shown below in Table 5.15. As can be seen for the 2-jet region there is a large contribution from W+jets and top backgrounds at low p_{T}^V. For the 3-jet region the top background is again the dominant contribution.

The background yields increase in other regions, but the signal yields decrease. In all regions, the yields in data are within the error on the estimated background yields. The full set of event yields for all regions is shown in Appendix F.

Sample	2T-tag				
p_{T}^V bin (GeV)	$p_{\mathrm{T}}^V < 90$	$90 < p_{\mathrm{T}}^V < 120$	$120 < p_{\mathrm{T}}^V < 160$	$160 < p_{\mathrm{T}}^V < 200$	$p_{\mathrm{T}}^V > 200$
	2-jet				
$Z + l$	0.1	0.1	0.0	0.0	0.0
$Z + cl$	0.4	0.1	0.1	0.0	0.0
$Z + hf$	43.4	15.3	6.8	1.7	0.5
$W + l$	2.0	0.4	0.3	0.1	0.1
$W + cl$	13.8	3.4	2.3	0.5	0.3
$W + hf$	407.6	116.0	106.3	30.0	27.6
s-top	486.2	106.5	72.9	13.3	4.7
MJ_μ	136.6	12.8	0.0	0.0	0.0
MJ_e	0.0	0.0	11.6	2.0	1.1
$t\bar{t}$	1172.9	431.8	377.9	44.5	10.3
VV	36.9	12.4	15.4	6.8	6.1
Total Bkg.	2300.0 ± 49.8	698.7 ± 15.5	593.6 ± 15.2	98.6 ± 3.6	50.7 ± 4.2
Total Signal	30.0 ± 11.8	10.0 ± 3.9	15.6 ± 6.1	8.3 ± 3.2	8.7 ± 3.4
Data	2364.0	700.0	591.0	112.0	59.0
	3-jet				
$Z + l$	0.0	0.0	0.0	0.0	0.0
$Z + cl$	0.2	0.0	0.0	0.0	0.0
$Z + hf$	22.2	6.1	3.7	0.9	0.4
$W + l$	0.7	0.2	0.1	0.0	0.0
$W + cl$	4.5	1.1	1.0	0.3	0.1
$W + hf$	142.7	41.6	48.5	19.7	21.9
s-top	336.0	71.7	50.5	10.5	7.3
MJ_μ	35.5	1.9	0.0	0.3	0.0
MJ_e	0.0	0.0	2.4	0.7	0.3
$t\bar{t}$	2638.0	642.7	485.4	61.3	15.6
VV	9.2	3.4	5.8	2.8	2.7
Total Bkg.	3189.0 ± 42.6	768.7 ± 15.7	597.5 ± 14.8	96.5 ± 3.8	48.3 ± 4.1
Total Signal	8.3 ± 3.4	3.3 ± 1.3	5.6 ± 2.2	3.4 ± 1.4	4.1 ± 1.6
Data	3161.0	779.0	590.0	101.0	53.0

Table 5.15 *Number of events obtained after performing the unconditional fit in the 1-lepton channel 2- and 3-jet 2T b-tag regions, with a Signal of $m_H = 125$. The uncertainties are the full postfit errors including all NPs with priors, floating normalisations, and correlations.*

m_H (GeV)	Obs.	Exp.	$+2\sigma$	$+1\sigma$	-1σ	-2σ
110	1.68	0.99	1.84	1.37	0.71	0.53
115	1.95	1.09	2.03	1.51	0.78	0.58
120	2.74	1.27	2.36	1.76	0.91	0.68
125	3.88	1.63	3.05	2.27	1.18	0.88
130	4.12	2.02	3.76	2.81	1.45	1.08
135	4.47	2.59	4.84	3.61	1.87	1.39
140	5.95	3.70	6.90	5.15	2.67	1.99

Table 5.16 *The observed and expected 95% CL upper limits on the cross section ratio σ/σ_{SM} for the 1-lepton channel in the dijet mass analysis for different Higgs masses, m_H.*

5.10.3 WH Dijet Mass Analysis Results

In the following, upper limits are placed on the cross section times branching ratio for a Standard Model Higgs boson, the best-fit signal strength measurements are calculated, and the significance of the observations is found.

Expected results are obtained by replacing the data by the expectations from simulation, with all NPs set to the best-fit values, also known as the Asimov dataset [171].

Confidence Limits

Figure 5.17 shows the 95% confidence level (CL) upper limits on the cross section times branching ratio, $\sigma/\sigma_{\mathrm{SM}}$, for the 1-lepton channel and the combination of the 0-, 1- and 2-lepton channels in the Higgs boson mass range 110–140 GeV. The observed limit for $m_H = 125$ GeV in the 1-lepton channel is 3.9 times the SM value, to be compared to an expected background-only limit of 1.6. For the combined 0-, 1-, 2-lepton channels, the observed and expected limits are 2.1 and 1.1 times the expected SM value, respectively. Numerical values of the limits are shown in Tables 5.16 and 5.17.

The values obtained in this analysis are compared with those from a number of analyses in Table 5.19.

m_H (GeV)	Obs.	Exp.	$+2\sigma$	$+1\sigma$	-1σ	-2σ
110	0.76	0.66	1.23	0.91	0.47	0.35
115	0.77	0.73	1.36	1.02	0.53	0.39
120	1.42	0.84	1.57	1.17	0.61	0.45
125	2.14	1.06	1.98	1.48	0.76	0.57
130	2.48	1.29	2.40	1.79	0.93	0.69
135	2.96	1.71	3.20	2.38	1.23	0.92
140	3.85	2.38	4.45	3.32	1.72	1.28

Table 5.17 *The observed and expected 95% CL upper limits on the cross section ratio σ/σ_{SM} for the combined 0-, 1- and 2-lepton channels in the dijet mass analysis for different Higgs masses, m_H.*

Signal Strength Results

The individual μ values for the lepton channels are obtained from a simultaneous fit with the signal strength for each lepton channel floating independently. Similarly, for the combined $(W/Z)H$ processes, a simultaneous fit with the signal strength is performed for each of the WH and ZH processes floating independently. The fitted signal strength values for $m_H = 125$ GeV are shown in Figure 5.18. The upper plot shows the signal strengths for the three lepton channels and their combination, and the bottom plot shows the signal strengths for the WH and ZH processes. As shown in Table 5.2, although the 1-lepton channel targets the WH process, there is some overlap with other processes, which accounts for the difference seen in the signal strength for the 1-lepton channel and the WH process in Figure 5.18. The signal strength in the 1-lepton channel is higher than the SM expectation at $\mu_{1-lep} = 2.2^{+0.97}_{-0.87}$, however, when combined with the 0- and 2-lepton channels this is reduced to $\mu_{012-lep} = 1.1^{+0.61}_{-0.56}$, which is consistent with a Standard Model Higgs boson. A similar result is seen with the WH signal strength and the combined WH and ZH signal strength.

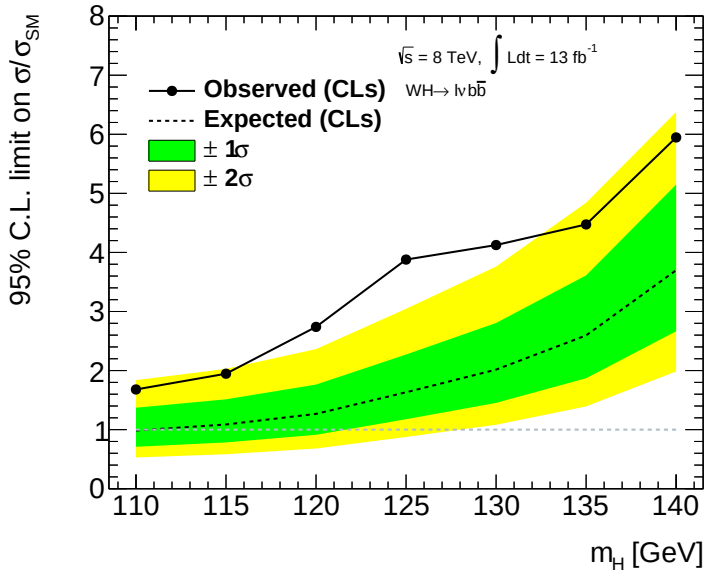
The signal strength obtained in this analysis is compared with those from a number of analyses in Table 5.18.

Signal Significance

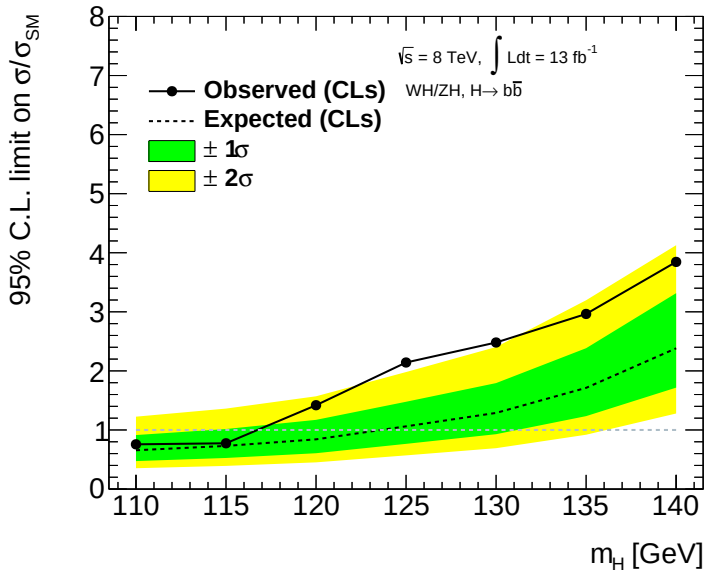
To claim discovery, a signal significance of at least 5σ is the convention used for discovery, which is equivalent to $p = 2.87 \times 10^{-7}$.

For a Higgs boson with mass 125 GeV in the 1-lepton channel, there is an excess observed (expected) with a significance of 2.7σ (1.3σ). For the combination of the 0-, 1-, and 2-lepton channels there is an observed (expected) significance of 2.01σ (1.94σ).

The signal significance obtained in this analysis is compared with those from a number of analyses in Table 5.19.

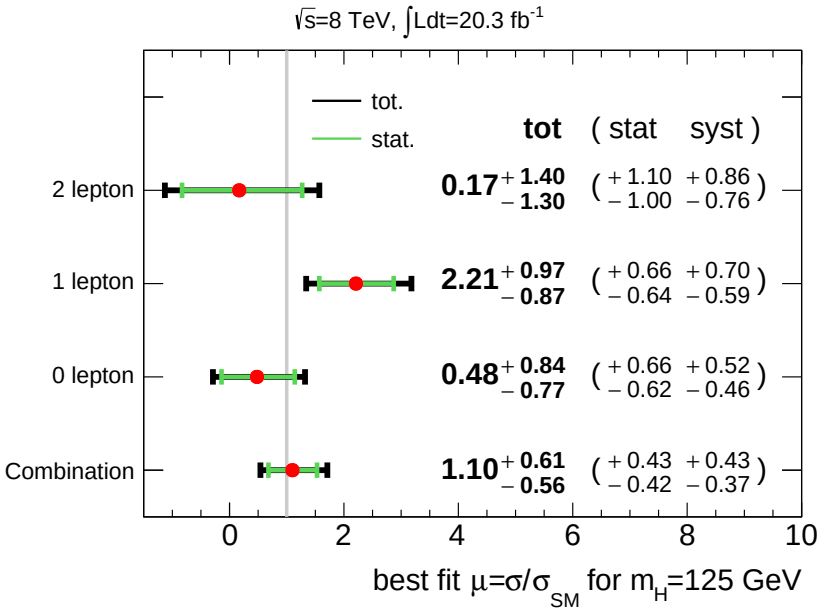


(a) *1-lepton channel*

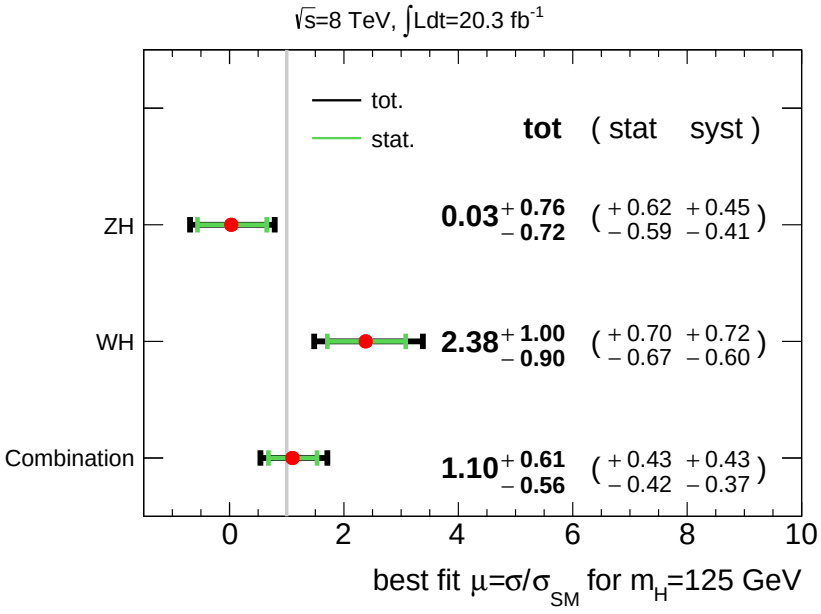


(b) *Combined 0-, 1-, and 2-lepton channel*

Figure 5.17 *Observed (solid) and expected 95% CL cross-section upper limits, normalised to the SM Higgs boson production cross section, as a function of m_H for the 1-lepton channel (top) and the combined 0-, 1-, and 2-lepton channels, as obtained using the dijet-mass analysis. The expected upper limit is given for the background-only hypothesis (dashed). The dark and light shaded bands represent the 1σ and 2σ ranges of the expectation in the absence of a signal. The numerical values of the data points are shown in Tables 5.16 and 5.17.*



(a) *μ for the 0-, 1-, and 2-lepton channels*



(b) *μ for the WH and ZH channels*

Figure 5.18 *The fitted values of the Higgs boson signal strength parameter μ for $m_H = 125$ GeV for the separate lepton channels and their combination (top), WH and ZH processes and their combination (bottom), with the 8 TeV dataset. The individual μ values for the $(W/Z)H$ processes are obtained from a simultaneous fit with the signal strength for each of the WH and ZH processes floating independently.*

5.10.4 MVA Results

The nominal analysis in Reference [1] was performed using an MVA selection for the 8 TeV data, and combined with a dijet mass analysis for the 7 TeV data. The results for the combination of the 0-, 1- and 2-lepton channels, which has the highest sensitivity, are quoted here for comparison with the dijet mass analysis results. The signal strength values are listed in Table 5.18, and the cross section limits and significance are listed in Table 5.19.

The 95% CL upper limits on the cross section times branching ratio for a number of Higgs boson mass points are shown in Figure 5.19. The observed (expected) limit on $\sigma/\sigma_{\mathrm{SM}}$ for a Higgs boson of mass 125 GeV with the MVA provides a stronger limit compared to the dijet mass analysis for the 8 TeV data (see Table 5.19).

The fitted signal strength for the MVA using the 8 TeV dataset is lower than that of the dijet mass analysis, both of which are consistent with a Standard Model Higgs boson. The combined signal strength from the 8 TeV dataset and the dijet mass analysis of the 7 TeV dataset is lower, due to a deficit seen in the 7 TeV data, but is still consistent with the Standard Model.

The data corresponds to an observed (expected) significance of 1.7σ (2.5σ) for $m_H = 125$ GeV. When combined with the dijet mass analysis of the 7 TeV data, this lowers to 1.4σ (2.6σ).

The consistency of the signal strength results obtained from the dijet mass analysis and the MVA was assessed by using a *bootstrap* method [172]. A large number of events are selected randomly from the Monte Carlo samples with the signal strength is set to $\mu = 1$. These are chosen such that they are representative of the integrated luminosity of the data used in the main analysis. The fitted $\hat{\mu}$ values for both the dijet mass analysis and the MVA are compared and their statistical correlation is calculated. Similarly, this is done for the expected results. The expected results for the dijet mass analysis and the MVA have a 67% correlation, whilst the observed results are statistically consistent at the 8% level.

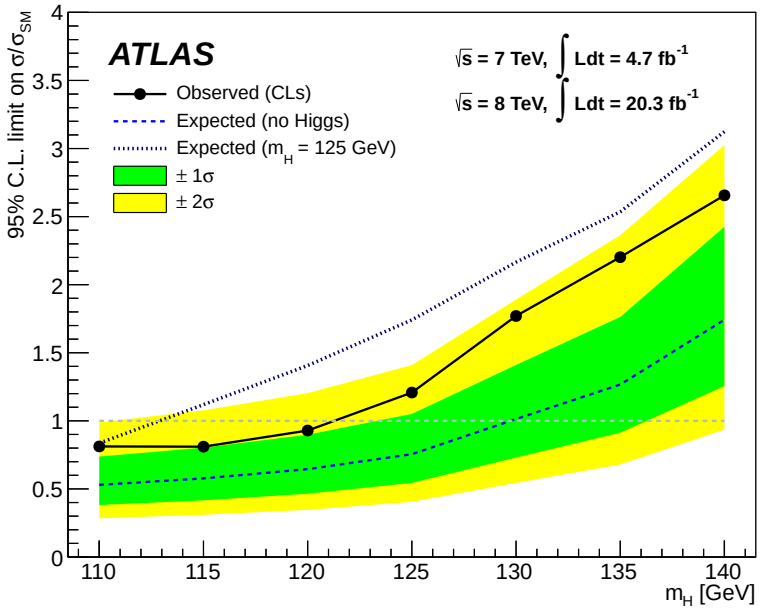


Figure 5.19 *Observed (solid) and expected 95% CL cross-section upper limits on the normalised SM Higgs boson production cross section, as a function of m_H for all channels and data-taking periods combined. These are obtained using the dijet-mass analysis for the 7* TeV *dataset and the MVA for the 8* TeV *dataset. The dashed line shows the expected upper limit for the background-only hypothesis, the dotted line shows the injection of a SM Higgs boson with $m_H = 125$* GeV*, and the shaded bands represent the 1σ and 2σ ranges of the background-only expectation.*

Analysis Method	Signal strength μ
Dijet mass analysis (8 TeV data only)	
1-lepton channel	$\mu = 2.2^{+0.67}_{-0.64}(\text{stat.})^{+0.7}_{-0.59}(\text{syst.}) = 2.2^{+0.97}_{-0.87}$
Combined lepton channels	$\mu = 1.10^{+0.43}_{-0.42}(\text{stat.})^{+0.43}_{-0.37}(\text{syst.}) = 1.10^{+0.61}_{-0.56}$
MVA (7 and 8 TeV data)	
Combined lepton channels (8 TeV only)	$\mu = 0.65^{+0.33}_{-0.32}(\text{stat.})^{+0.28}_{-0.24}(\text{syst.}) = 0.65^{+0.43}_{-0.40}$
Combined lepton channels (7+8 TeV)(*)	$\mu = 0.51^{+0.31}_{-0.30}(\text{stat.})^{+0.25}_{-0.22}(\text{syst.}) = 0.51^{+0.40}_{-0.37}$
CMS [173]	$\mu = 1.0 \pm 0.5$

Table 5.18 *The best-fit signal strength obtained from a number of different analyses. The dijet mass analysis of the 1-lepton channel targets the $WH \to \ell\nu b\bar{b}$ process using only 8* TeV *data. The remaining results show the signal strength for the $VH(H \to b\bar{b})$ process, using the combination of the 0-, 1-, and 2-lepton channels. The superscript of mbb indicates that a value is found using the dijet mass analysis. (*) For the analysis of the 7* TeV *data, a dijet mass analysis is used. The CMS result uses the full $\sqrt{s}$ =7 and 8 TeV datasets.*

Analysis Method	$\sigma/\sigma_{\text{SM}}$	Significance
Dijet mass analysis (8 TeV data only)		
1-lepton channel	3.9 (1.6)	2.7σ (1.3σ)
Combined lepton channels	2.1 (1.1)	2.0σ (1.9σ)
MVA (7 and 8 TeV data)		
Combined lepton channels (8 TeV)	1.4 (0.8)	1.7σ (2.5σ)
Combined lepton channels (7+8 TeV)(*)	1.2 (0.8)	1.4σ (2.6σ)
CMS [173]	0.95 (1.89)	2.1σ (2.1σ)

Table 5.19 *The observed (expected) values of σ/σ_{SM} and the observed (expected) significance obtained from a number of different analyses. The dijet mass analysis of the 1-lepton channel targets the $WH \to \ell\nu b\bar{b}$ process using only 8* TeV *data. The remaining results target the $VH(H \to b\bar{b})$ process, using the combination of the 0-, 1-, and 2-lepton channels. (*) For the analysis of the 7* TeV *data, a dijet mass analysis is used. The CMS result uses the full $\sqrt{s}$ =7 and 8 TeV datasets.*

5.10.5 Diboson Crosscheck

The analysis procedures are validated by performing a fit to extract the VZ signal strength, μ_{VZ}. The difference in data in the m_{bb} distribution between all backgrounds and the diboson backgrounds is shown in Figure 5.20 (a).

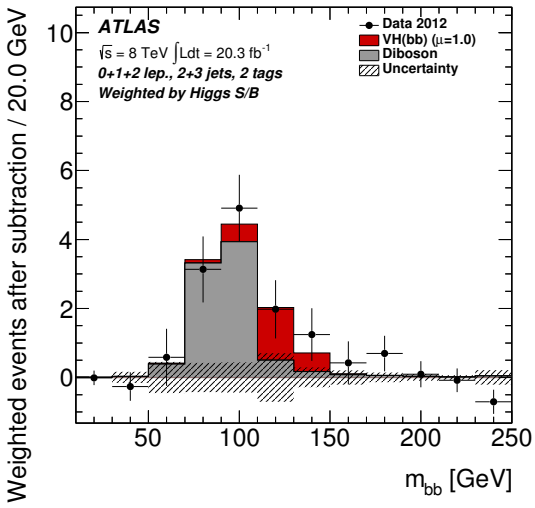
The signal strengths obtained using the dijet-mass analysis at 8 TeV, for both the 1-lepton channel and the combination of the 0-, 1- and 2-lepton channels is shown in Figure 5.20 (b).

The observed and expected significances obtained for these channels are shown in Table 5.20.

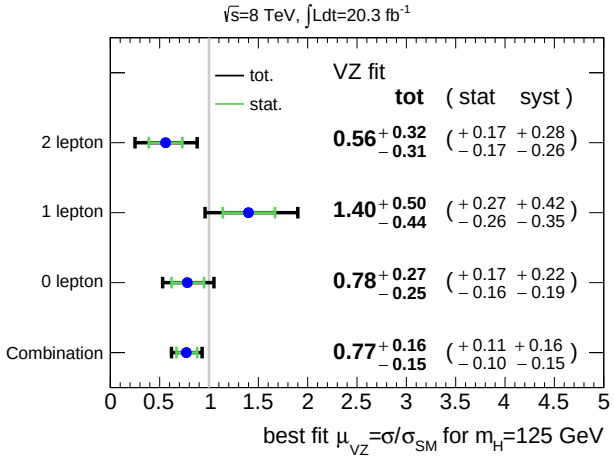
The signal strengths and significances obtained with the dijet mass analysis of the 8 TeV data are compared in Table 5.20 with those obtained with a dijet mass analysis of the 7 TeV and combined with the MVA 8 TeV analysis. The observed signal strength is compatible with the Standard Model expectation, validating the analysis procedure.

Analysis Method	Signal strength μ_{VZ}	Significance
Dijet mass analysis (8 TeV data only)		
1-lepton channel	$\mu_{VZ} = 1.40^{+0.27}_{-0.26}(\text{stat.})^{+0.42}_{-0.35}(\text{syst.}) = 1.40^{+0.50}_{-0.44}$	3.5σ (2.7σ
Combined lepton channels	$\mu_{VZ} = 0.77^{+0.11}_{-0.10}(\text{stat.})^{+0.16}_{-0.15}(\text{syst.}) = 0.77^{+0.16}_{-0.15}$	4.7σ (5.7σ)
MVA (8 TeV) combined with dijet mass analysis (7 TeV data)		
MVA+dijet mass	$\mu_{VZ} = 0.74 \pm 0.09(\text{stat.}) \pm 0.14(\text{syst.})$	4.9σ (6.3σ)

Table 5.20 *The signal strength μ_{VZ} and the observed (expected) significance obtained from three analyses: the dijet mass analysis of the 8* TeV *data in the 1-lepton channel, the combination of the 0-, 1-, and 2-lepton channels, and the combined MVA (for the 8* TeV *data) and dijet mass (for the 7* TeV *data) analyses. The dijet mass analysis of the 1-lepton channel targets the $WZ \to \ell\nu b\bar{b}$ process. The remaining results target the $VZ(Z \to b\bar{b})$ process.*



(a)



(b)

Figure 5.20 *Plot (a) shows the distribution of m_{bb} after subtraction of all backgrounds except for the diboson processes, using the combination of the three lepton channels in the dijet-mass analysis for the 8* TeV *data. All p_T^V bins, 2- and 3-jet regions and 2-tag b-tagging categories are summed and weighted by the respective ratios of expected Higgs boson signal to fitted background in each region. The contribution of the WH and ZH signal processes with a SM Higgs boson with $m_H = 125$* GeV *is shown with a signal strength of $\mu = 1.0$. The total uncertainty on the fitted background is indicated by the hatched band. Plot (b) shows fitted values of the diboson signal strength μ_{VZ} using the dijet mass analysis for the 8* TeV *dataset for the separate lepton channels and their combination.*

5.11 Conclusions

A search has been presented in this thesis for a Standard Model Higgs boson decaying into $b\bar{b}$ and produced in association with a leptonically decaying vector boson. The focus of this work is the $WH \to \ell\nu b\bar{b}$ process where the reconstruction of the dijet invariant mass of the two b quarks from the Higgs boson candidate is used as the final discriminant. The combination of the associated W and Z processes are also presented. The results presented are obtained from pp collisions recorded by the ATLAS detector at $\sqrt{s} = 8$ TeV during Run 1 of the LHC in 2012, corresponding to an integrated luminosity of 20.3 fb^{-1}. This is performed in tandem with, and as a cross-check of, a multivariate analysis. The two methods are found to be consistent, with an expected correlation of the signal strengths obtained with either method of 67%.

The dijet mass analysis uses categories based on the number of leptons, jets, b jets, and the transverse momentum of the vector boson. Separate analysis categories, the 0-lepton, 1-lepton and 2-lepton channels, are designed to target the $ZH \to \nu\bar{\nu}b\bar{b}$, $WH \to \ell\nu b\bar{b}$ and $ZH \to \ell^+\ell^- b\bar{b}$ processes, respectively. The analysis method is validated using a measurement of the $VZ(\to b\bar{b})$ yield.

The results obtained are consistent with a Standard Model Higgs boson of mass 125 GeV decaying to $b\bar{b}$, however, there is not enough data to be able to make a statistical claim for discovery. The observed (expected) significance of a Higgs boson with a mass of 125 GeV in the 1-lepton channel (targeted at the $WH \to \ell\nu b\bar{b}$ process) is 2.7σ (1.3σ). Combining all three lepton channels, an observed (expected) significance of 2.01σ (1.94σ) is found. The observed significance using dijet mass analysis is higher than that of the MVA method (1.7σ), however, it has a lower expected significance (1.94σ as opposed to 2.5σ).

An observed (expected) 95% confidence upper limit on $\sigma/\sigma_{\mathrm{SM}}$ for $m_H = 125$ GeV in the 1-lepton channel is found at 3.9 (1.6). For the combination of all three lepton channels, the observed (expected) limit is 2.1 (1.1). Further, a Higgs boson with a mass of between 110 and 115 GeV has been excluded at the 95% confidence level. The limits from the MVA are tighter, with observed and expected limits of 1.4 and 0.8, respectively.

The best-fit signal strength for a 125 GeV SM Higgs boson in the 1-lepton channel using data from $\sqrt{s} = 8$ TeV collisions, obtained with a dijet mass analysis, is $\mu^{\mathrm{mbb}}_{\text{1-lep}} = 2.2^{+0.97}_{-0.87}$. For the combination of the three lepton channels it is

$\mu^{\text{mbb}}_{\text{012-lep}} = 1.10^{+0.61}_{-0.56}$. The MVA signal strength is lower, at $\mu = 0.65^{+0.43}_{-0.40}$, however, they are found to be within errors and statistically consistent at the level of 8%.

The reported results from other experiments and combinations are consistent with these results, as discussed here. An analysis of $VH(H \to b\bar{b})$ was performed by the CMS experiment [173], reporting an observed (expected) significance of 2.1σ (2.1σ) and a signal strength of $\mu_{bb}^{CMS} = 1.0 \pm 0.5$. All results from the $t\bar{t}H(H \to b\bar{b})$ and $VH(H \to b\bar{b})$ Higgs production from ATLAS and CMS have been combined to obtain an observed (expected) significance for the process $H \to b\bar{b}$ of 2.6σ (3.7σ) with a signal strength of $\mu_{bb}^{ATLAS+CMS} = 0.70^{+0.29}_{-0.27}$ [174]. A combined analysis from the CDF and DØ experiments at the Tevatron reported an excess of 2.8σ [57], and a signal strength of $\mu = 1.59^{+0.69}_{-0.72}$ [52], using 10 fb^{-1} of data from proton-antiproton collisions with a centre-of-mass energy of $\sqrt{s} = 1.96$ TeV.

The combined ATLAS signal strength measured from $H \to WW$, ZZ, $\gamma\gamma$, $\tau^+\tau^-$, $b\bar{b}$, and $\mu^+\mu^-$ for a Higgs boson with a mass of 125.36 GeV using both 7 and 8 TeV data is $\mu = 1.18^{+0.15}_{-0.14}$ [175]. When combined with the CMS results, a value of $\mu = 1.09^{+0.11}_{-0.10}$ is obtained [174]. These results are consistent with the Standard Model.

5.12 Future Work

The latest $VH(\to b\bar{b})$ results from ATLAS were presented at ICHEP in 2016 [176]. Using 13.2 fb^{-1}of data collected at $\sqrt{s} = 13$ TeV, an observed (expected) signal significance of 0.42σ (1.94σ) for a Higgs boson of mass 125 GeV was reported, with a signal strength of $\mu = 0.21^{+0.36}_{-0.35}(stat.) \pm 0.36(syst.)$. These results are consistent with previous results in the $VH(\to b\bar{b})$ channels.

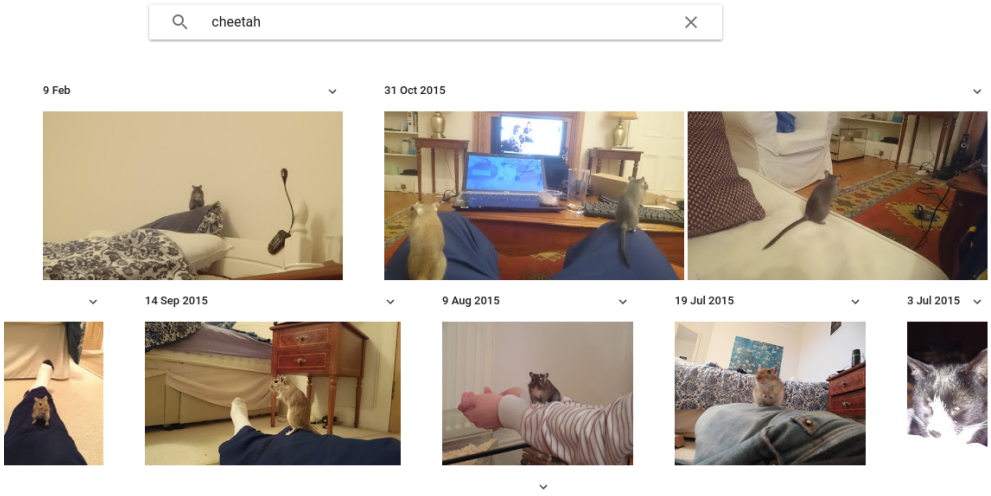
The remainder of the Run 2 schedule for 2016, 2017 and 2018 will see proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV, producing an expected 100 fb^{-1}. 2018 will also have a dedicatd Pb-Pb run.

Running at the design centre-of-mass energy of $\sqrt{s} = 14$ TeV, the LHC is expected to deliver up to 300 fb^{-1} by 2022 (with an average pile-up of 60). After an upgrade in 2022 to the High Luminosity LHC (HL-LHC), a total of 3000 fb^{-1} will be delivered by 2030 (with an average pile-up of 140) [177]. This additional data is needed to be able to claim a 5σ discovery of the Higgs boson decaying

into $b\bar{b}$. A cut-based study was carried out to determine the sensitivity to the $WH \to \ell\nu b\bar{b}$ and $ZH \to \ell^+\ell^- b\bar{b}$ processes, at the HL-LHC [178], which estimates an expected sensitivity of 2.6σ and $\mu = 1 \pm 39\%$ for 300 fb^{-1}, and 5.9σ and $\mu = 1 \pm 19\%$ for 3000 fb^{-1}. These results use a conservative estimate of the systematic uncertainties. An extrapolation of the expected improvements in the understanding of the systematic uncertainties, improvements in b tagging and the use of a multivariate analysis, the significance improves to 3.9σ with $\mu = 1 \pm 27\%$ for 300 fb^{-1}, and 8.8σ with $\mu = 1 \pm 14\%$ for 3000 fb^{-1}.

Chapter 6

Machine Learning Techniques



Machine learning, and deep learning in particular, has enjoyed success in recent times in fields such as image recognition.

6.1 Introduction

In Chapter 7 of this thesis, an analysis is presented which optimises the selection of boosted, hadronically decaying W bosons at ATLAS. This is accomplished by making use of *machine learning* techniques, which describe algorithms that offer semi-automated, multivariate techniques to build predictive models. Two machine learning algorithms are employed here for classifying W bosons: *boosted decision trees* [24, 25] and *deep neural networks*, in the form of *stacked autoencoders* [21–23].

In general, a machine learning algorithm takes an input dataset, or set of *input vectors*, $\{\boldsymbol{x}\}$ of dimension n, and applies a set of *weighted* transformations to the input vector and outputs a vector $\{\hat{\boldsymbol{x}}\}$ of dimension m, where typically $m \leq n$. Each of the n elements of an input vector are referred to as a *feature*. The elements of the output vector can be interpreted individually, or by considering a linear combination of each element. The goal of the algorithm is to be able to 'learn' a function of a given input dataset, by tuning the weights of the set of transformations, such that the algorithm can classify or provide predictions on previously unseen input datasets. In high energy particle physics, the input dataset is generally a set of observables, either measured or simulated, that describe a collision event. Regularities in the dataset are used to classify events as coming from different processes or for the identification of particles. In the context of the work presented here, these algorithms are used as *classifiers*, where the algorithms learn to classify if the input dataset corresponds to a W boson or a QCD jet.

Training is an essential part of the process of implementing machine learning algorithms. The weights of the transformations that are applied to an input vector are determined and tuned by training the algorithms on a set of input vectors referred to as *training datasets*, in an attempt to minimise a predefined *loss function*. In particle physics, and particularly in the work presented in Chapter 7, Monte Carlo simulations are used as training datasets. Machine learning can be separated into broad categories based on the learning paradigm that is used for training. In *supervised learning*, the elements of the training dataset are *labelled* as belonging to a specific class, such as signal or background. This allows for a comparison of the classification provided by the algorithm with the true label. By iteratively updating the set of weighted transformations applied to an input dataset, the algorithms are trained to classify datasets with greater accuracy.

Artificial neural networks and decision trees are examples of algorithms which employ supervised learning. *Unsupervised learning* overlaps with data mining, where the goal is to find unknown or hidden structure in the data. The training datasets used in unsupervised learning are not necessarily labelled. Unsupervised algorithms include *clustering* (grouping together data points that are similar and appear to belong to a particular class) and *dimensionality reduction* (reducing the representation of the data into a subset of features that are uncorrelated). The autoencoders discussed in Section 6.3.2 are used for learning representations of data and dimensionality reduction in an unsupervised manner.

After an algorithm has been trained, it should be evaluated on a *test dataset* that is independent from the training dataset. The performance of the algorithm when it is applied to these test datasets provides a method of validating that the training process has been successful. The algorithm should ideally perform equally well on both the test and training datasets.

Trained algorithms are typically referred to as *models*, each of which has a number of tuneable *hyperparameters* that can additionally affect their performance.

Before training a machine learning algorithm, the input datasets must be prepared. The preparation includes *cleaning* the data by identifying missing or inaccurate values. In the case of algorithms such as stacked autoencoders, the data must also be standardised. A method known as *stratified k-fold cross validation* can be employed to ensure that when training an algorithm, it does not *overfit*, whereby it learns to identify patterns in the training dataset and cannot generalise to other datasets, or *underfit*, whereby the algorithm is not learning a suitable representation of the input data. The tuning of the hyperparameters is another important tool that can be used to prevent overfitting and underfitting.

In previous studies at the LHC, artificial neural networks have proved useful for identifying boosted top quarks [179]. Deep networks based on stacked autoencoders have also been used in exotics searches [180]. In this chapter, artificial neural networks, (stacked) autoencoders and boosted decision trees are introduced. These techniques are implemented in Chapter 7 for classifying jets which come from a boosted W boson.

Data preparation, cross validation, algorithm evaluation and hyperparameter optimisation are also discussed in Sections 6.2 and 6.5. These are common to both the deep neural networks and boosted decision trees.

6.2 Data Preparation

In the studies performed in Chapter 7, the datasets consist of a number of collision events, each with a number of attributes describing the event. The events are labelled as coming from either *signal* or *background* processes. Expanding on the previous notation of a set of input vectors, these labelled datasets can be represented as vectors $\{\boldsymbol{x}, \boldsymbol{t}\}$. $\boldsymbol{x}$ describes all the events, each of which has n features, and $\boldsymbol{t}$ contains the corresponding label of the event.

The following sections describe how the datasets are cleaned and prepared. This involves removing extreme *outliers* which can negatively impact training and may sometimes indicate an erroneous value, standardising the values, and splitting the datasets into orthogonal training and testing datasets, ideally with no missing values or data.

6.2.1 Cleaning

It is possible that some datasets have extreme outliers. These outliers might be statistical fluctuations or an indication of an error. Such outliers can have a large effect on the performance of a classifier if they are used for training. Similarly, there can be missing values, which again will affect the classifier performance. In the process of cleaning the data, either these events can be removed or a value inferred for the missing entries given by an average or some other function.

Part of the cleaning process involves searching for correlations between variables and choosing which ones are useful for classification purposes. This is covered in Section 6.5.4.

For some algorithms, such as stacked autoencoders, it is necessary to standardise or normalise the samples [181]. If there are two input variables that have different scales, they can affect the classification disproportionately. By scaling all the input variables to be within a certain range this is mitigated. This is not necessary for decision tree classifiers [181]. For the training of the stacked autoencoders in Chapter 7, each variable is standardised to unit variance and zero mean:

$$x_{std}^{(i)} = \frac{x^{(i)} - \mu_x}{\sigma_x}, \tag{6.1}$$

where $x^{(i)}$ refers to the i-th entry of the sample, μ_x is the mean of the variable x,

and σ_x is the standard deviation.

6.2.2 Training and Testing Datasets

When implementing a machine learning classifier, it is essential that there are independent datasets for training and testing. This is to ensure that the classifier has learnt a model of the training dataset that has captured underlying relationships that are applicable to new datasets. If the classifier overfits to the training dataset, random fluctuations in the training dataset are effectively incorporated into the model. Often the amount of simulated data available for training and testing a classifier is limited. Such Monte Carlo simulations must be separated into two datasets for this, and it may be that the dataset chosen for training is not representative. The relative proportion of signal and background events should be the same in the training and test datasets as it is in the full dataset from which these training and test datasets are derived. This is known as *stratification* [182]. Another procedure that can be used is *cross-validation*. The dataset is split into a number of equal partitions, or *folds*, all of which are random. Each partition is used in turn for testing, with the remaining folds being used for training. In the studies presented here, these two methods are used in conjunction in what is called *stratified k-fold cross-validation* [182]. This can heavily mitigate any bias caused by the partitioning process. The classifier is trained and tested on each of the folds individually, and the overall performance of the classifier is obtained by averaging the performance on each of the folds.

6.3 Deep Learning

Deep learning refers to a set of machine learning algorithms that use multiple processing layers to create a high-level, abstract representation of a given dataset [183]. By finding complex representations and correlations within the dataset, models can be created for predictions or classification of unseen datasets. In the case of high energy physics data, classification models could use low-level kinematic variables to describe higher-level processes, for example.

Artificial neural networks (ANNs), as described in Section 6.3.1, consist of a number interconnected *neurons*, or *units*, which are arranged into a number of *layers*. Each unit applies a transformation to any input it receives and sends

this output to one or more units in the following layer. Although there is no strict definition, these types of *multi-layer architectures* are often considered *deep neural networks* when there are two or more layers in between the initial and final layers [183]. In a multi-layered architecture, each layer essentially computes a non-linear function of the previous layer, learning an increasingly more complex function that describes relationships between the features of a given dataset.

Training deep neural networks using supervised learning requires all the samples to be labelled. When training a neural network, the weight of the transformation for each unit is adjusted by attempting to minimise a loss function, or *error function*. The loss function provides a way in which to quantify how closely the classifications provided by the algorithm match the target labels. By adjusting the weights associated with each unit, the accuracy of the classification can be improved. Typically, these adjustments to the weights are propagated backwards one layer at a time. *Gradient descent* methods (described in Section 6.3.1) are often used to minimise the loss function and adjust the weights, however, the relative value of the adjustments become smaller at each layer; when there are sufficiently many layers the earlier layers do not get updated as quickly as other layers. In general, the training requires a significant number of data points on which to train.

A solution to the training difficulties encountered for deep neural networks is to implement a *greedy layer-wise approach* [184, 185]. Using this approach, a layer is added to the network and trained, and this trained layer is used as the input for the subsequent layer. It is said to be "greedy" in the sense that at any point in the training, only a single layer is considered, and any following layers in the network are ignored; a locally optimum result is found each time a layer is added. This can be done in a supervised manner, or, as illustrated in Section 6.3.2, it can be done unsupervised. Unsupervised learning has the benefit that the network is able to learn initial features of the training sample. The two learning techniques can be used in tandem, using the network trained with unsupervised learning as a set of initial weights for the units in a neural network, which is subsequently trained with supervised learning. This is also useful given a limited dataset. This is implemented as a stacked autoencoder, which is described in Sections 6.3.2 and 6.3.3.

In Chapter 7, a deep neural network, in the form of a stacked autoencoder, is used for classifying boosted W bosons. The following introduces artificial neural networks, autoencoders, and stacked autoencoders. A software package that

implements this, AGILEPack, is used in subsequent studies, and is introduced here.

6.3.1 Artificial Neural Networks

Artificial neural networks were originally inspired by models that sought to explain information processing in biological systems [186]. Neural networks are modelled as a number of interconnected neurons, where the connections can be weighted and adjusted to allow the network to 'learn'. Mathematically, neural networks are linear combinations of variable non-linear basis functions, where each basis function is a non-linear function of a linear combination of inputs with variable coefficients, which are fitted to the training dataset [187]. Training the neural network on a set of features of the training dataset allows for such a function of the inputs to be learnt, identifying complex, non-linear relationships between a number of attributes of the dataset. Neural networks are often used for classification tasks, where the classifier predicts if the data belongs to one of K discrete classes (signal or background, for example). In these classification tasks, there are K or fewer neurons in the output layer, which correspond to each class.

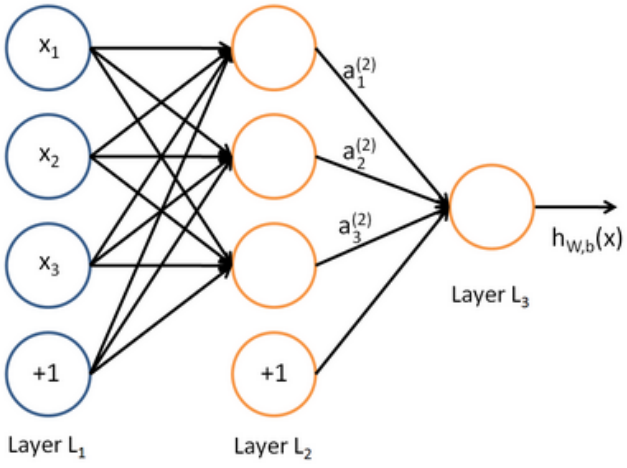


Figure 6.1 *A basic feedforward neural network [183]. In this diagram, the layers run from left to right, and the neurons per layer run vertically. The leftmost layer is the input layer, where the neurons labelled x_n accept as input the n features of the data. The '+1' neurons are the bias units. The right-most layer is the output layer, which may have one or more unit. In binary classification, the output layer can make use of a single unit which produces a continuous output, with lower and higher values indicating the probability of the class. Layer 2 is a hidden layer.*

The basic components of a neural network are single neurons, or *units*. Each

neuron takes an input vector $\boldsymbol{x}$ with n features[1] and is parameterised by a weight W, which must be fit to a training dataset. A neural network is comprised of a number of layers of neurons, with each layer l taking as input the output from the *activation function* (defined in Equations 6.2 and 6.3) of neurons in the previous layer. The left-most layer, or *input layer*, has $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ inputs. The *output layer* consists of n or fewer outputs. Any intermediate layers are *hidden layers*. This is illustrated in Figure 6.1.

In each layer, a neuron may be connected to one or more neurons in the next layer, or skip the next layer completely. Each neuron accepts a linear combination of output from the neurons it is connected to in the preceding layer, such that any layer will consist of M such linear combinations, where $M \leq$ the number of neurons in that layer.

Transformations are applied to each input a neuron receives, and each of these is linearly combined to give a value $h_{W,b}(\boldsymbol{x})$:

$$h_{W,b}(\boldsymbol{x}) = f(\sum_{i=1}^{n} W_i x_i + b), \quad f : \Re \mapsto \Re, \text{ for } x_1, \ldots, x_n \tag{6.2}$$

where $f(\cdot)$ is the activation function. The bias term b can be included as a term $x_0 = 1$, by summing from $i = 0$ and incorporating it into a weight term W_0, i.e. $h_W(\boldsymbol{x}) = h(\sum_{i=0}^{D} W_i x_i)$. In the context of the work presented in this thesis, the function $f(\cdot)$ is given by the softmax function. The softmax function is a generalised logistic function, defined within $[0, 1]$, operating over a K-dimensional vector $\boldsymbol{z}$, given by:

$$f(\boldsymbol{z}_j) \equiv \sigma(\boldsymbol{z}_j) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \quad \text{for } j = 1, \ldots, K. \tag{6.3}$$

The softmax function has a well-defined derivative, which is used when considering error propagation, discussed in the following sections. The derivative of the function is given as

$$\frac{\partial \sigma(\boldsymbol{z}_i)}{\partial z_j} = \begin{cases} \sigma(\boldsymbol{z}_i)_i(1 - \sigma(\boldsymbol{z}_i)), & \text{if } i = j \\ -\sigma(\boldsymbol{z}_i)\sigma(\boldsymbol{z}_j) & \text{otherwise.} \end{cases} \tag{6.4}$$

[1]There is an additional input in the form of a bias term b, which itself takes no inputs. This is referred to as a *bias unit*. The bias term is analogous to the intercept term in a linear equation of the form $y = mx + c$.

Feedforward networks map a set of inputs x_n to a set of outputs y_m using a set of linear combinations of multiple neurons, determined by an adjustable weight vector

$$\boldsymbol{W} = W_{ji}, \tag{6.5}$$

where W_{ji} is the weight between neuron i in layer l and neuron j in layer $l+1$. This is illustrated in Figure 6.1.

The value obtained for a single neuron by applying the activation function is referred to as the *activation*, a, of the neuron. The activation will depend on the input the neuron receives from preceding layers. For layer k, the activation of neuron j is given by the recursive relation:

$$a_j^{(k)} = \begin{cases} \sum_{i=0}^{D} x_i, & \text{for } k = 1 \\ \sum_{i=0}^{D} W_{ji} z_i, & \text{for } k = 2, \ldots, n_l \end{cases} \tag{6.6}$$

where $z_j = h_{W,b}(x_j)$ refers to the overall input from the previous layer, i is one of the D units in the previous layer, $j = 1, \ldots, M$ is the number of neurons in the current layer, and W_{ji} is the weight between neuron j in layer $k+1$ and neuron i in the current layer, and n_l is the total number of layers. Thus, for a given set of $\boldsymbol{W}$ and $\boldsymbol{x}$ for a neural network, a single real number is output for each neuron. This method of calculating the activations for each layer sequentially is known as *forward propagation*, providing the definition of a *feedforward* neural network [187].

Training

The goal of using a neural network for classification is to provide a mapping such that, given input vectors $\{\boldsymbol{x}\}$, where $\boldsymbol{x} = \{x_1, \ldots, x_n\}$, the output $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{W}) = \{\hat{\boldsymbol{x}}\}$ from the neural network matches some set of target output values $\{\boldsymbol{t}\}$ as closely as possible. This can be achieved by minimising a loss, or error function, defined as:

$$E(\boldsymbol{W}) = \frac{1}{2} \sum_{n=1}^{N} \left\| \boldsymbol{g}(\boldsymbol{x}_n, \boldsymbol{W}) - \boldsymbol{t}_n \right\|^2 . \tag{6.7}$$

In the case of a weighted input dataset, such as those used in Chapter 7, the weights are included in the error function.

Minimising $E(\boldsymbol{W})$ proceeds by adjusting $\boldsymbol{W}$. However, $E(\boldsymbol{W})$ can have multiple

local optima. Applying a small update to the weight, $\boldsymbol{W} \rightarrow \boldsymbol{W} + \delta\boldsymbol{W}$ gives a corresponding change in E of

$$\delta E \approx \delta \boldsymbol{W}^T \nabla E(\boldsymbol{W}), \tag{6.8}$$

where $\nabla E(\boldsymbol{W})$ is a derivative, indicating the greatest rate of increase in $E(\boldsymbol{W})$.

The updates of $\boldsymbol{W}$ are done over τ iterations, following $\boldsymbol{W}^{(\tau+1)} = \boldsymbol{W}^{(\tau)} + \delta\boldsymbol{W}^{(\tau)}$. In gradient descent optimisation, this update is given by:

$$\boldsymbol{W}^{(\tau+1)} = \boldsymbol{W}^{(\tau)} - \eta \nabla E(\boldsymbol{W}), \tag{6.9}$$

where $\eta > 0$ is the *learning rate*. The learning rate controls the rate at which weights are updated. In *stochastic gradient descent* methods [188, 189], the weights are given random initial values, and updated according to $\boldsymbol{W}^{(\tau+1)} = \boldsymbol{W}^{(\tau)} + \boldsymbol{V}^{(\tau+1)}$, where $\boldsymbol{V}^{(\tau+1)} = \mu\boldsymbol{V}^{(\tau)} - \eta\nabla E(\boldsymbol{W})$, initialised to random values, and μ is the *momentum*.

This can be generalised further to include a *regularisation* term. In order to reduce the likelihood of overfitting due to overly large weights, the error function is redefined as

$$\tilde{E}(\boldsymbol{W}) = E(\boldsymbol{W}) - \frac{\lambda}{2}(\boldsymbol{W})^2, \tag{6.10}$$

where λ decreases the magnitude of the weights, controlling the relative importance of the weights and the original error function.

The learning rate, momentum, and regularisation are important hyperparameters that can have a large effect on the training of a neural network. These hyperparameters are considered in Chapter 7 when optimising the stacked autoencoders.

Before training begins, the elements of $\boldsymbol{W}$ must be set to initial random values[2]. As shown in Equation 6.9, updating $\boldsymbol{W}$ requires calculating the derivative of the error function with respect to the weights. An efficient way of updating $\boldsymbol{W}$ is through the *back-propagation* algorithm [189]. In this algorithm, the updates to the weights of each layer are dependent on the following layer. The weights are calculated for the final layer, and then the updates are calculated for each preceding layer.

[2]If the weights all start with the same initial values, then all hidden layers learn the same function.

For a given layer l, the derivative of $E(\boldsymbol{W})$ is given by

$$\frac{\partial \tilde{E}}{\partial W_{ji}} = \delta_j z_i. \tag{6.11}$$

where W_{ji} is defined as in Equation 6.5, δ_j is an *error term* for each unit j in layer $l+1$ (how much a unit is contributing to the error function), and z_i is the output of neuron i in layer l. The error term is defined as

$$\delta_j = \frac{\partial \tilde{E}}{\partial a_j} = \sum_k \frac{\partial \tilde{E}}{\partial a_k} \frac{\partial a_k}{\partial a_j}. \tag{6.12}$$

The back-propagation algorithm can be summarised as follows:

1. Calculation of all activations for every neuron for a given input vector $\boldsymbol{x}$, propagated forward through the network.
2. Calculate the error function (Equation 6.10).
3. Evaluation of δ_j for each unit.
4. *Back-propagate* the δ terms from the output layer through to the input layer.
5. Evaluate all the derivatives in Equation 6.11 using δ terms and z_i.
6. Update the weight vector.

The above method is iterated multiple times in training to reduce the error function, Equation 6.10, to within a pre-defined threshold.

6.3.2 Autoencoders

An autoencoder is a multi-layered, unsupervised, neural network using back-propagation trained on a dataset which is not necessarily labelled [23, 183, 190]. The target values of the network are set to the same values as the inputs, allowing the autoencoder to learn an approximation to the identity function, i.e. $h_W(\boldsymbol{x}) = \hat{\boldsymbol{x}} \approx \boldsymbol{x}$. Limiting the number of neurons in the hidden layers (*hidden neurons*), or limiting the number of connections, places constraints on the network, forcing it to learn a compressed representation of the input. If the events in the dataset

are independent and identically distributed, this is not likely to work, but if the dataset is structured and contains correlations between the inputs, then this method can be used to attempt to find correlations. The following description is given in terms of the implementation found in the AGILEPack software [191] used in Chapter 7, and described in Section 6.5.6. Here there is a single hidden layer in the autoencoder, but multiple hidden neurons.

The autoencoder consists of an encoding step, which encodes values from the input layer to a compressed representation, and decoding step, which decodes the compressed representation and reconstructs the output. Encoding maps the input $\boldsymbol{x}$ into a new representation $\boldsymbol{\varphi}(\boldsymbol{x})$:

$$\boldsymbol{\varphi}(\boldsymbol{x}) = f(\boldsymbol{W}_1\boldsymbol{x} + b_1), \tag{6.13}$$

where f is a softmax activation function, and $\boldsymbol{W}_1$ and b_1 are the weight and bias terms, respectively, for the connection between the hidden layer and input layer. Decoding takes the compressed representation $\boldsymbol{\varphi}(x)$ and maps it onto the original vector space according to:

$$\boldsymbol{\rho}(\boldsymbol{x}) = g(\boldsymbol{W}_2\boldsymbol{\varphi}(x) + b_2), \tag{6.14}$$

where g is the activation function, and $\boldsymbol{W}_2$ and b_2 are the weight vector and bias term, respectively, for the connection between the hidden layer and the output layer.

6.3.2.1 Training

As before with the neural network training, an error function (or loss function) is defined:

$$L(\boldsymbol{x}) = \frac{1}{2n}\sum_{i=1}^{n} \|x_i - \rho_i(x_i)\|^2 + \lambda_1\|\boldsymbol{W}_1\|^2 + \lambda_2\|\boldsymbol{W}_2\|^2, \tag{6.15}$$

where λ_1 and λ_2 are regularisation parameters (usually on the order of 10^{-4} [192]) which can be optimised to prevent overfitting. Here, stochastic gradient descent in combination with back-propagation are used to find the weights and bias terms that minimise the function [188].

The same method of back-propagation is used as described in Section 6.3.1, where the gradient of the error function in Equation 6.15 is minimised with respect to

each element of $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$. Using these definitions, the weights and bias terms can be adjusted using back-propagation for a number of iterations or *epochs*. In Chapter 7, the number of epochs is varied to find an optimal value.

6.3.3 Stacked Autoencoders

Autoencoders can be stacked into multiple layers, creating a deep network that can learn a higher order representation of the data than a single autoencoder. They are stacked in such a way that the hidden layer of each autoencoder provides the input to the next layer. Stacked autoencoders can be trained in an unsupervised manner to learn an initial representation of the data. The final layer can then feed into a standard feedforward neural network used for classification, where the stacked autoencoder provides the initial weights for the classifier.

The motivation for using a deep learning approach to identifying boosted W bosons is that patterns are found in the data *automatically*, rather than relying on high-level features that must be constructed individually. Adding additional layers to a neural network introduces challenges to the training procedure, for example, when using gradient descent methods the propagation of the updates to the earlier layers become less pronounced as more layers are added. Using this method of stacking autoencoders and training them separately mitigates some of these difficulties by pre-training the network. When the full network is constructed and trained on labelled data, the initial layers have already learned a representation of the data and require less tuning than they would otherwise, allowing deeper networks to be trained more easily, which is especially useful when there are not many events to be used for training, as is the case in Chapter 7. This is the motivation for the use of a stacked autoencoder in Chapter 7 to create a deep learning solution to identifying boosted W bosons.

Training Stacked Autoencoders

In a stacked autoencoder, a greedy layer-wise approach is used for training [184]. At each step, the target vector is equal to the initial the input vector. The first layer (that is, the first autoencoder) is trained on a set of input vectors $\{\boldsymbol{x}\}$ to learn a first order representation of the input. Formally, this is the training of an autoencoder with weights $\boldsymbol{W}^{(1,1)}$, $\boldsymbol{W}^{(1,2)}$ and bias terms $b^{(1,1)}$, $b^{(1,2)}$, where the first term in the superscript (i, j) indicates the layer in the stacked

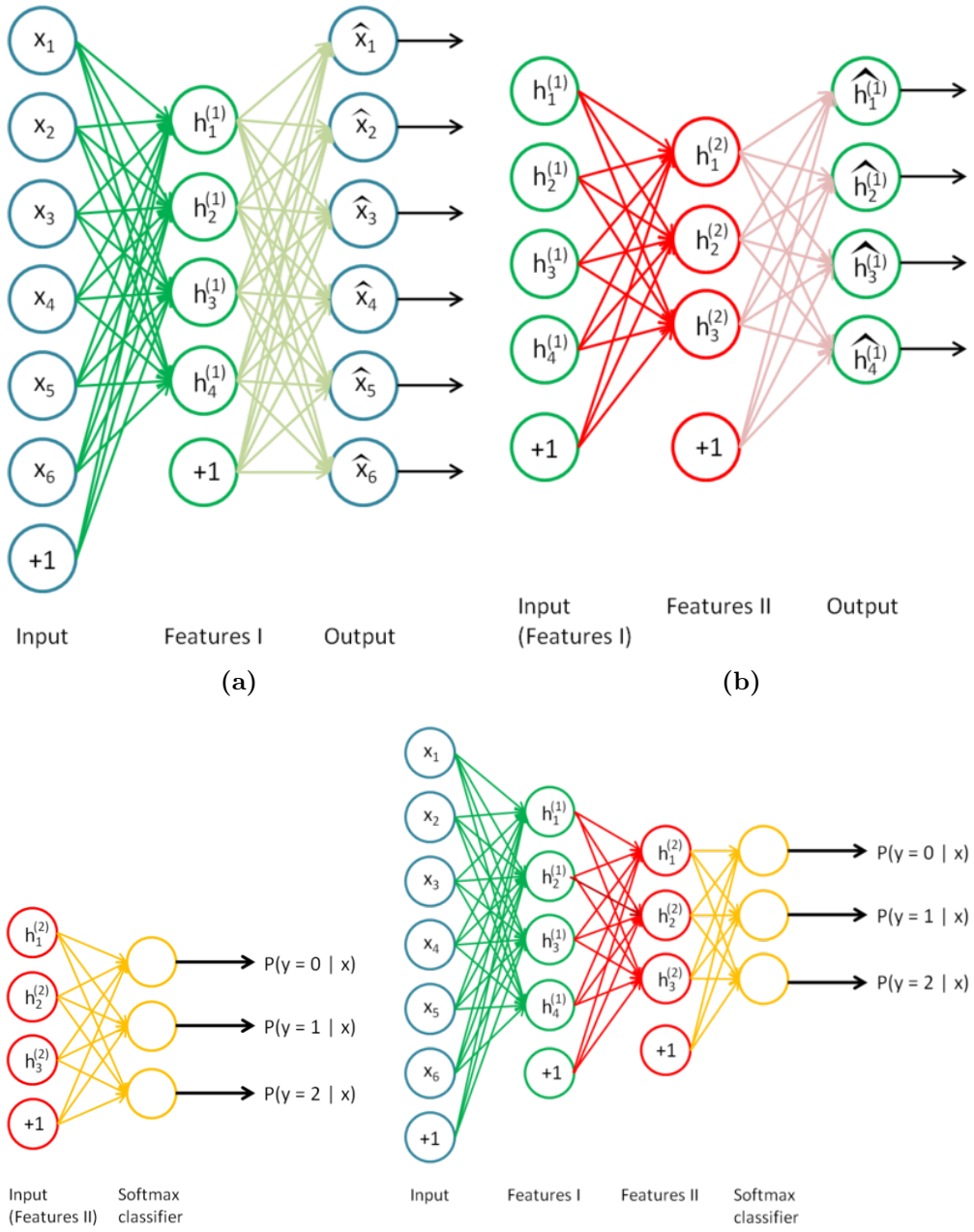


Figure 6.2 *The training of a stacked autoencoder follows a greedy layer-wise approach [23]. In (a) the initial autoencoder has a single hidden layer $h^{(1)}$ and is trained to reconstruct the input x_i. The hidden layer $h^{(1)}$ is used as an input to the second autoencoder in (b), training a second layer $h^{(2)}$ to reconstruct the input $h^{(1)}$. In (c), $h^{(2)}$ is used as the input to a softmax classifier. In (d) the entire classifier is shown, which takes an input and transforms this through two autoencoders and a softmax layer to produce a prediction.*

autoencoder, and the second term is the i^{th} layer in the current autoencoder. The training process is more easily understood by referring to Figure 6.2. Initially, in Figure 6.2(a), the first autoencoder is trained such that it is able to produce an output vector $\boldsymbol{h}_{\boldsymbol{W}}^{(1)}(\boldsymbol{x}) = \hat{\boldsymbol{x}} \approx \boldsymbol{x}$, where the superscript indicates the autoencoder number. At this point, all weights and bias terms are frozen before adding a second autoencoder. The second autoencoder takes as input the units in the hidden layer of the first autoencoder, and is trained to output a vector $\hat{\boldsymbol{h}} \approx \boldsymbol{h}$, as shown in Figure 6.2(b). This can continue further to create a network with more layers. In Figure 6.2(c), the hidden layer from the final autoencoder is fed into a neural network with a softmax activation function for classification, referred to as a *softmax classifier*, with the number of output units proportional to the number of classes in the dataset. At this point, the complete classifier, consisting of the stacked autoencoder and a softmax classifier, can be constructed, as shown in Figure 6.2(d). The classifier formed from this combination is trained further using standard back-propagation in a supervised manner for a number of epochs, adjusting the weights of both the (previously frozen) stacked autoencoder and the final layer of the softmax classifier, in what is known as *fine-tuning*. The process of training the network with an unsupervised algorithm can significantly improve deep learning algorithms [193].

6.4 Decision Trees

Classification and regression trees [24], or *decision trees*, use a 'divide-and-conquer' approach to classification using supervised learning. For the work presented here, decision trees are constructed as *binary trees*, which provide one of two decisions: *signal* or *background*. A decision tree consists of a single *root node*, which is connected via a number of intermediate *decision nodes*, to a set of *leaf nodes*, as illustrated in Figure 6.3. The root and internal nodes represent a single feature or attribute of an input dataset, whilst the leaf nodes represent classification outcomes. The path from the root node to a leaf node is known as a *branch*. At each decision node, a simple criterion is defined, based on the feature the decision node represents, which splits the current branch into two regions: signal-like or background-like. This continues until a leaf node is reached. Thus, each branch specifies a sequence of selection criteria on features of a dataset, which will classify a single data point as signal or background. In general, applying a decision tree to a dataset will not fully separate signal and background events at

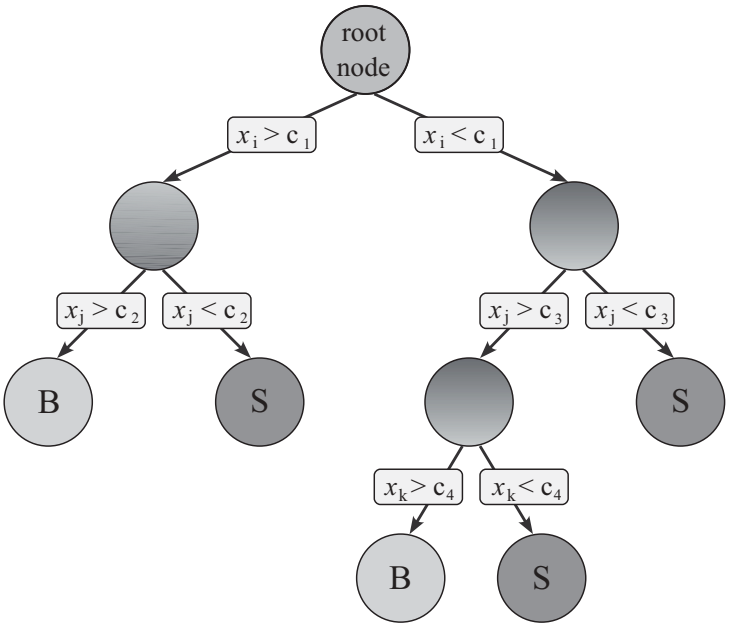


Figure 6.3 *A binary decision tree is shown here with a single root node, followed by a set of nodes which provide yes/no decisions (binary splits), ending in a set of leaf nodes which are labelled according to the class which they are most likely to be able to classify - either signal (S) or background (B) [182].*

the leaf nodes. The goal is to be able to find a suitable choice of selection criteria such that a large proportion of events are classified as one or the other.

A decision tree is built from the top down, adding a single decision node at a time, which then spawns two leaf nodes: one leaf node for each class. At each decision node, the feature from the training dataset is chosen which gives the greatest separation between background and signal. A common metric for this, which is used here, is the *Gini index* [187]. For a given tree T, the Gini index is calculated for a decision node according to

$$Q_\tau(T) = \sum_{k=0}^{1} p_{\tau k}(1 - p_{\tau k}), \quad \tau = \{0, 1\} \tag{6.16}$$

where τ refers to the two new leaf nodes which will be created by splitting on the decision node, $k = 0, 1$ refers to signal or background, respectively, and $p_{\tau k}$ is the weighted proportion[3] of events in the training dataset of class k which will be classified by leaf node τ. When $Q_\tau(T) = 0.5$, its maximum, there is

[3]In the studies performed in this thesis, the training dataset is weighted according to cross section, Monte Carlo event weights, the number of events in the Monte Carlo sample, and a p_{T} reweighting factor.

no discrimination between classes. The value of $Q_\tau(T)$ is evaluated for a given decision node for all the features of the input dataset which do not exist in the current branch. This process is repeated for each leaf node until a stopping criterion is met, such as the number of data points or events belonging to each leaf node dropping below a threshold, or the training dataset is fully divided into signal and background. A maximum depth of the tree can be implemented to prevent the tree from growing too large, which can lead to overfitting. The maximum depth is a hyperparameter of the decision tree that requires tuning. After training, the tree will have learnt a function $h(\boldsymbol{x}) \in \{-1, 1\}$, where -1 indicates background and 1 indicates signal, taken as the weighted majority in the leaf node. This can be modified to give a probability given by the ratio of signal and background events in the leaf node.

Decision trees are easily interpretable and are deterministic. However, since they are sensitive to the details of the dataset, small changes can give rise to vastly different tree structures [194]. One possible solution to this is to use an *ensemble* of trees to improve the generisability and robustness. Ensemble methods are separated into *boosting* or *averaging* ensemble methods, such as boosted decision trees and random forests, respectively. These are discussed in the following sections.

6.4.1 Boosted Decision Trees

Boosting is a technique used to combine ensembles of relatively weak *base classifiers* to create a significantly stronger classifier [195]. The basic idea is to train M classifiers, or *estimators*, on a training dataset in sequence where each following training iteration uses a modified version of the dataset. In the modified dataset, any events that were misclassified previously are given a larger weight, forcing the classifier to focus more on these instances. The final classifier is then given by a linear combination of all the base classifiers. In the work presented here, a method of boosting called *adaptive boosting*, or *AdaBoost* is used [25]. When using a decision tree as the base classifier, referred to as a *boosted decision tree* or *BDT*, the classification performance can be improved significantly [194]. This is the method employed for the work presented here.

The training dataset is given an initial set of weights (in the datasets used here, each event is weighted based on a number of parameters of the Monte Carlo simulation). For the first decision tree these weights are used, thereafter any

misclassified events are multiplied by $\exp(\alpha^{\beta})$, where β is the learning rate and α is the boost weight. At each iteration, the dataset with the adjusted weights is used to build the next decision tree. By iteratively adjusting the weights, the classifier learns a better separation of signal and background. The number of base classifiers and the learning rate are hyperparameters that can be tuned to improve the performance. In general, a slower learning rate with an increase in the number of boosts, or number of classifiers, can help to improve classification.

The AdaBoost classifier calculates a linear combination of m base classifiers, $Y_M(\boldsymbol{x})$ and produces a real value between -1 and 1, with a value nearer to -1 indicating a more background-like event and closer to 1 a more signal-like event. Different cuts on the output of the classifier will yield different signal and background efficiencies, and this must be tuned to each problem individually. The classifier can also be used to predict the probability of an event being in a specific class. The *predicted class probability* for a single base classifier in the ensemble is given by the fraction of events of the same class in a leaf node. The predicted class probability from the ensemble is a weighted mean of all class probabilities of the base classifiers in the ensemble.

6.4.2 Random Forests

In averaging ensemble methods, a number of estimators (in this instance a decision tree) are created independently and the predictions from each estimator are averaged. This reduces the variance of the prediction compared with the individual decision trees. Random forests [196] are one type of averaging method, which consist of a large collection of de-correlated decision trees.

Decision trees in a random forest are built using data points drawn from the training dataset using *replacement*, such that any data point may be used multiple times when building a single decision tree. Splitting a branch at a given decision node is done in a randomised way by only choosing the feature on which to perform the split from a random subset of all features. This increases the bias of the forest, but the variance decreases from the averaging procedure.

Random forests, and decision tree models in general, can be used to identify the relative importance of all features in the model. In AdaBoost trees, the depth of the decision node corresponding to a feature within the tree is used to assess its importance. Features at a low depth contribute to more predictions

and have a higher relative importance. In a randomised forest, another method is used since the depth of a variable is chosen at random. The mean decrease in impurity (Gini importance) is calculated by taking a sum of the Gini impurity over the number of splits and averaged over all trees in the ensemble for a given feature [197]. The averaging of these values over all trees reduces the variance of the importance, improving feature selection. Random forests are less able to detect relevant features if their correlation is high, generally only identifying one of the correlated variables as important and the rest given lower importance [198]. In Chapter 7, random forests are used to select which features are used for training the BDTs and stacked autoencoders.

6.5 Tuning Classifier Hyperparameters

The classifiers discussed so far learn a set of weights from a fit to the training dataset, but there are some parameters that must be chosen by hand, such as the input features to train on and the model hyperparameters. The model hyperparameters would be the maximum depth of the tree or the learning rate, for example. Classifiers have a bias, which is given by its average error over different training datasets, and a variance, which is an indication of how the classifier changes performance over the training datasets. The parameters have to be tuned to get the bias and variance as low as possible. The performance of a classifier must be evaluated on some previously unseen sample to ensure its generality and to check for over- and underfitting. A number of techniques are available for assessing and improving the performance of a classifier, some of which are discussed below.

6.5.1 Classifier Metrics

Model validation requires a *scoring function* or set of metrics on which to evaluate its performance. These are often given in terms of error rates, or the number of correctly classified events. In the work presented here, a number of scoring functions are defined in terms of the true and false positive rates [199]. True positives (TP) and true negatives (TN) are those events that are labelled signal and background which are actually signal and background, respectively. False positives (FP) and false negatives (FN) are those events labelled as signal and background that are actually background and signal, respectively. The metrics

considered in the work presented here are the following:

$$\text{accuracy} \equiv \frac{\text{TP+TN}}{\text{total events}} \tag{6.17a}$$

$$\text{weighted accuracy} \equiv \frac{\frac{\text{TP}}{\text{all signal}} + \frac{\text{TN}}{\text{all background}}}{\text{total events}} \tag{6.17b}$$

$$\text{signal efficiency} \equiv \frac{\text{TP}}{\text{total events}} \tag{6.17c}$$

$$\text{background efficiency} \equiv \frac{\text{TN}}{\text{total background}} \tag{6.17d}$$

$$\text{background rejection efficiency} \equiv 1 - \text{background efficiency} \tag{6.17e}$$

$$\text{background rejection power} \equiv \frac{1}{\text{background efficiency}} \tag{6.17f}$$

Receiver Operating Characteristic (ROC) curves offer a graphical tool for selecting and evaluating multiple models [182, 200]. These are generally given in terms of signal efficiency on the x-axis and the background rejection efficiency on the y-axis. This offers an intuitive way to find a working point, or cut, with a maximum signal efficiency and background rejection. These are implemented in the following chapter to find optimal models.

6.5.2 Grid Search

In some cases, it may be that certain parameters or input features can be eliminated without losing classification performance, or even improving it. A grid search is a brute-force approach to finding an optimal set of hyperparameters and input features that maximises a chosen metric. Given a set of possible values for the hyperparameters and input features, it trains and tests a classifier with every combination of these. This can be used in conjunction with cross validation to ensure the validity of the classifier.

6.5.3 Validation Curves

The influence of a single hyperparameter on the classifier can be evaluated using a validation curve [199]. The value from the chosen metric obtained on the training and test datasets for a number of values for the hyperparameter are plotted together to look for indications of overfitting or underfitting. If the scores are both showing poor performance (low accuracy or signal efficiency, for example),

this indicates that the classifier is underfitting and not learning any suitable function of the data. If the training score is high and the validation score is low this indicates overfitting. In the validation curves plotted in Chapter 7, the accuracy is chosen as the scoring function.

A similar technique is used to show the performance of a classifier as a function of training samples. A learning curve shows how the validation score and training score change. If these are both converging to a low value as the number of samples increases, then adding more events for training will not be useful.

6.5.4 Feature Selection

Aside from the methods above, where a grid search could be used to find a set of features, or observed variables in high energy physics datasets, to be used as inputs for training, there are other methods available. When training a decision tree, the Gini index (Equation 6.16) gives an indication of how much separation a given feature will provide. Each feature is then given an *importance*, or rank, based on this. This allows for the identification of the most important features in the training of a BDT or random forest [181, 199]. Unimportant features can generally be removed without negatively affecting the performance of the classifier. If multiple features are correlated then there is not necessarily any informational gain if all of them are used. A visual inspection of the distribution of the features between signal and background events can also be helpful in deciding if the variables should be used.

6.5.5 Scikit-Learn

In Chapter 7, Scikit-learn [199] is used to implement a set of BDTs. The AdaBoost algorithm is used with a decision tree classifier as the base classifier. The tuning is performed on three hyperparameters: number of estimators (the number of base classifiers), maximum decision tree depth, and the learning rate. A grid search is used in conjunction with stratified 10-fold cross validation.

Other software used includes: ROOT [201], root_numpy [202], IPython [203], Pandas [204], Matplotlib [205] and Numpy [206].

6.5.6 AGILEPack

The studies performed in Chapter 7 employ the AGILEPack software package [191] to implement a stacked autoencoder with a softmax classifier (as described in Section 6.3.3) for identifying boosted W bosons. Stacked autoencoders with a depth of four layers are used in conjunction with a softmax classifier. A grid search is implemented to determine the best values for the hyperparameters: regularisation, momentum, unsupervised training epochs of the stacked autoencoder, supervised training epochs of the stacked autoencoder and softmax layer in the fine-tuning step, and the learning rate. The same set of features in the input dataset are used as for the BDT for consistency. All training and testing datasets are standardised according to Equation 6.1, and stratified 5-fold cross validation is used. This was chosen over the conventional 10-fold as each network can take a significant amount of time to train (up to an hour each), and much longer to train than the BDT. Increasing the number of folds can decrease the variance, however, this also decreases the fold sizes and the sample size available for training, which also increases the variance, although at a lower rate. 5-fold cross validation was chosen as a compromise between the time needed for training and the variance of the trained networks.

Chapter 7

Boosted W Boson Tagging

Commercial W-jet grooming is available at only £2 for 6 minutes.

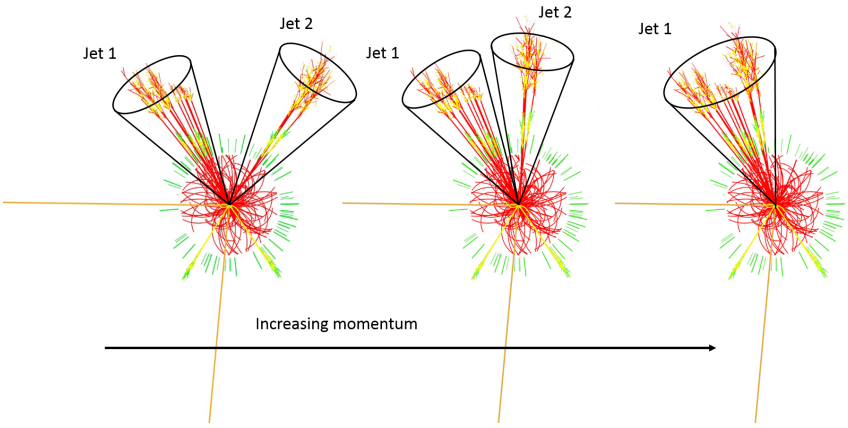


Figure 7.1 *Jets are identified as clusters of hadronic activity from the decay of a parent particle. If a particle decays into two jets, these jets get closer together at higher momentum. This can result in two jets being clustered together and identified as a single jet. Adapted from [207].*

7.1 Introduction

At high collision energies at the LHC, particularly during Run 2, W and Z bosons can be produced with large transverse momenta, or *boosted*. Hadronically decaying W and Z bosons are generally identified as two jets within the detector. However, these jets can become sufficiently close together such that they are identified as a single jet when the parent particle is highly boosted, as is illustrated in Figure 7.1. Being able to identify these boosted W and Z bosons correctly requires specialised techniques that examine the internal structure, the *substructure*, of the jets.

The W boson is important for many analyses and decays hadronically 67.5% of the time. Furthermore, there are many Beyond the Standard Model theories that predict new heavy particles at the LHC, which decay into boosted W bosons. Having an efficient way of *tagging* boosted W bosons is thus well motivated. These substructure techniques will also be a useful tool in the boosted regime when considering the $WH(H \to b\bar{b})$ search, by extending these techniques to boosted $H \to b\bar{b}$ decays [208], and possibly even hadronically decaying W bosons in this search channel.

A key parameter in jet reconstruction is the radius of the jet R (see Section 4.7). W bosons decay hadronically into two quarks which generally produce two jets, which can be captured in a single large-R jet [209]. In this context, large-R typically refers to a radius of $R \geq 1.0$. For reference, many analyses which do not

consider boosted jets use a radius of $R = 0.4$, as in Chapter 5. Smaller radius jets are not able to contain all the activity of the jets, unless the W boson is highly boosted. The two prong decays of the W boson provide hard substructure within the jet, as well as soft contributions from pile-up and radiative effects. Jets from QCD production, which consists of mainly light quarks and gluons (referred to as *multi-jet* events), are predominantly soft and do not exhibit the same hard substructure. Selectively removing the soft radiation within the jet can reveal the hard substructure more distinctly. This process is known as *grooming*, described in detail in Section 7.2. Numerous grooming algorithms are available which attempt to remove the soft emissions within jets, thereby providing a more efficient way of discriminating between jets from light quarks or gluons, and hadronic decays of heavy vector bosons.

Examining the internal constituents of the jet is another useful method that can reveal the hard substructures within the jet. In substructure studies, the jets are identified using the k_t, C/A, anti-k_t algorithms with a large radius [123] that captures the entire decay. Substructure variables (Section 7.3) are used to characterise and quantify the hard substructure within a jet, often used in conjunction with jet grooming algorithms.

The methods that are currently used to identify boosted W bosons at ATLAS can be broadly separated into jet grooming and substructure techniques, typically in the context of large-R jets [2, 209–211]. In Reference [2], cut-based methods using groomed jets and jet substructure variables for tagging boosted W bosons are presented based on Run 1 MC and data at a centre-of-mass energy of $\sqrt{s} = 8$ TeV, which provide the baseline technique for ATLAS Run 2 analyses. The results from this paper advocate the use of a combination of applying a mass cut on a jet that has been *groomed*, and a cut on a single substructure variable. This is summarised in Section 7.5.

The results and recommendations from Reference [2] are re-evaluated at higher energies of up to $\sqrt{s} = 13$ TeV in Section 7.6. Subsequently, a study into the use of machine learning techniques is introduced. This is used as a feasibility study to identify the maximum performance gain over the baseline that can be achieved using two machine learning classifiers which were introduced in Chapter 6: boosted decision trees (BDTs) and deep neural networks (DNNs).

This chapter is structured as follows: Grooming algorithms are introduced in Section 7.2 and jet substructure variables are described in Section 7.3. The

simulated Monte Carlo samples are listed in Section 7.4. The boosted W boson tagger based on Run 1 $\sqrt{s} = 8$ TeV collisions is discussed in Section 7.5. In Section 7.6 the W boson tagging is studied in Run 2 conditions at $\sqrt{s} = 13$ TeV using machine learning (ML) techniques. The setup for the ML taggers and their results are described in Section 7.7.

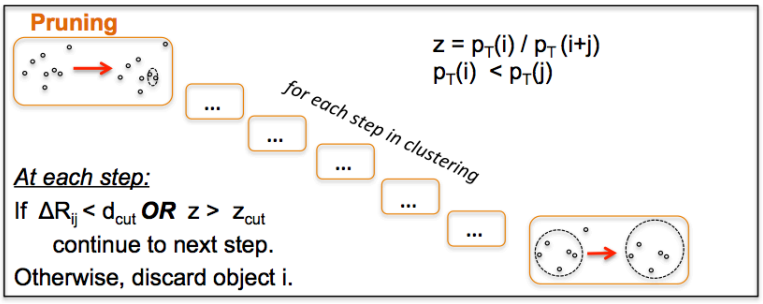
7.2 Jet Grooming Algorithms

Jet grooming is a technique to identify hard substructure within large radius jets. Jet grooming 'undoes' the last stages of the jet clustering to enable soft emissions to be identified and removed. The k_t and C/A jet clustering algorithms (see Section 4.7) cluster on low-p_{T} constituents and angular separation first, respectively, whereas the anti-k_t algorithm clusters on high-p_{T} constituents first. The k_t and C/A clustering algorithms therefore provide valuable information about the large-scale substructure of the jets. By examining the clustering in reverse order, the hardest *subjets* are revealed by the k_t algorithm, and those with a large angular separation are revealed by the C/A algorithm.

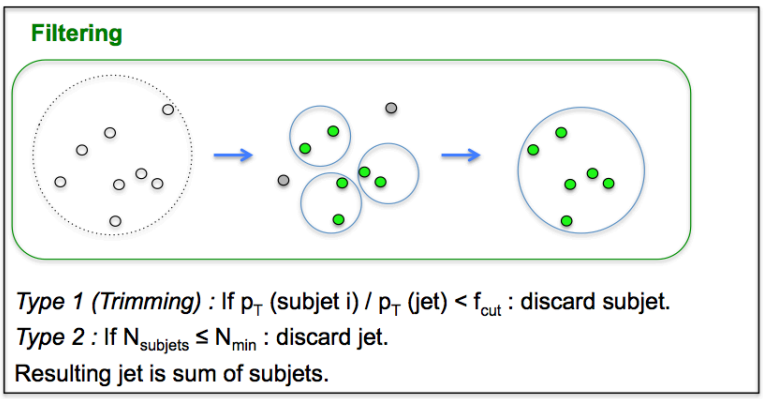
Groomed jets generally have better mass resolution than ungroomed jets (up to $\approx 10\%$ [209] for top quarks). This improved resolution is due to the removal of soft and wide-angled emissions, essentially reducing the effective jet area and defining the jet mass (defined later in the text in Equation 4.1) by the hard substructure. Additionally, since these soft contributions are removed, there is reduced sensitivity to pile-up. The jet mass for multi-jet events is shifted lower after grooming, as constituents are removed. In jets from decaying W or Z bosons this is less pronounced since the higher p_{T} subjets survive, with the smaller contributions from soft emissions being removed. This is found to be particularly apparent in the *trimming algorithms*, discussed below.

There are three main categories of jet grooming algorithms (referred to as *groomers*) which are considered for W boson tagging at ATLAS: trimming [212], pruning [213, 214] and split filtering [208]. The performance of the grooming algorithms depends on the jet clustering algorithm and the radius parameter, R. The different groomers are discussed below and illustrated in Figure 7.2.

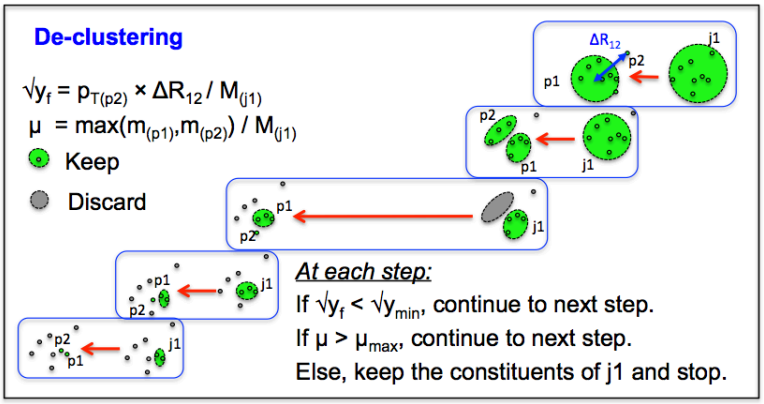
- Trimming - Trimming algorithms are designed to exploit the difference in the jet constituents from multiple sources. Pile-up, multiple parton



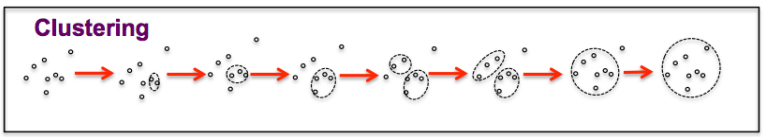
(a)



(b)



(c)



(d)

Figure 7.2 *Illustrations of the three jet grooming algorithms discussed in Section 7.2 [211]. (a) is the pruning jet grooming algorithm, (b) shows trimming, (c) shows the declustering (or splitting) part of the split filtering grooming algorithm, and (d) is the corresponding filtering stage.*

interactions and initial state radiation (see Chapter 4) are much softer in jets than the products of the hard scatter. The original constituents of large-R jets (C/A or anti-k_t) are reconstructed into subjets using the k_{T} algorithm [126], using $R = R_{\mathrm{sub}}$ as a distance parameter. Subjets which have $p_{\mathrm{T}} < f_{\mathrm{cut}} \times p_{\mathrm{T}}^{\mathrm{jet}}$, where $p_{\mathrm{T}}^{\mathrm{jet}}$ is the p_{T} of the original jet, are removed. The trimming algorithm with anti-k_t, $R = 1.0$, $f_{\mathrm{cut}} \geq 5\%$ and $R_{\mathrm{sub}} = 0.3$ is the groomer that was recommended by ATLAS in Reference [210] for top quark tagging at $\sqrt{s} = 7$ TeV, whilst using $R_{\mathrm{sub}} = 0.2$ has also been found to perform well for W boson tagging and pile-up reduction in Reference [2] at $\sqrt{s} = 8$ TeV. The trimming algorithm with anti-k_t, $R = 1.0$ was also shown to provide good jet mass resolution and pile-up rejection for top quark tagging at $\sqrt{s} = 7$ TeV [209].

- Pruning - Pruning algorithms remove relatively small p_{T} constituents, similar to trimming, with the additional removal of wide-angled constituents or radiation. The original constituents of large-R jets (C/A or anti-k_t) are reconstructed with the C/A algorithm, cutting on the angular distance between two subjets (denoted by R_{cut}) and the fraction of the p_{T} carried by the lighter subjet (Z_{cut}). During the clustering, pairs of constituents are considered. If $\Delta R_{12} > R_{\mathrm{cut}} \times 2M/p_{\mathrm{T}}$, where ΔR_{12} is the angle between the two constituents[1] (see Figure 7.3) and M is the jet mass, then the second constituent is discarded. Additionally, the jet constituent is discarded if it is soft, having $f_2 < Z_{\mathrm{cut}}$, where f_2 is the p_{T} fraction of the softer constituent. The pruning step is performed at each recombination in the jet clustering, such that the wide-angle and soft terms that are removed are from the proto-jets (defined in Section 4.7), rather than the original jet. Pruning with C/A jets, $R = 0.8$, Z_{cut} =10% and $R_{\mathrm{cut}} = 0.5$ has been shown to work well for W boson tagging, as recommended by CMS [215, 216].

- Mass-drop filtering (split filtering) - The algorithm is designed to identify symmetrical subjets within a jet, where each subjet has a much smaller mass than their sum. There are two distinct stages to the algorithm: the *splitting* stage and the *filtering* stage. In the splitting stage, the constituents of large-R C/A jets are *declustered* using an exact reverse of the initial clustering steps, splitting the clusters into two pieces at each step. The mass drop, μ_{12}, (mass of the hardest piece as a fraction of the current jet's mass, see Equation 7.1) and momentum balance, $\sqrt{y_{12}}$ (see Equation 7.2),

[1] In general, the opening angle between decay products is given by $2M/p_{\mathrm{T}} \sim R$ [209].

are calculated at each step in the declustering process. The mass drop is defined as

$$\mu_{12} = \frac{\max(m_1, m_2)}{m_{12}}, \tag{7.1}$$

where the index 1 (2) refers to the leading (subleading) piece in the declustering step and m_{12} refers to the invariant mass of the two pieces.

The momentum balance is defined as

$$\sqrt{y_{12}} = \frac{\min(p_{\mathrm{T}1}, p_{\mathrm{T}2})}{m_{12}} \Delta R_{12}, \tag{7.2}$$

where ΔR_{12} is the angle between the two pieces. If there is a large mass drop such that $\mu_{12} < \mu_{\max}$ or the momentum balance is $\sqrt{y_{12}} > \sqrt{y_{\min}}$, this piece is presumed to be a hard structure and is returned as a subjet[2]. Otherwise, the declustering procedure is repeated on the highest momentum piece. This is continued iteratively.

In the filtering stage, all subjets identified in the declustering process are *reclustered* using a new radius parameter $R_{\mathrm{sub}} = \min(0.3, \Delta R_{12})$, where ΔR_{12} is the value calculated during the splitting stage in Equation 7.2. The three or fewer hardest subjets that remain after the declustering and filtering process are taken as the final constituent subjets of the original jet. All other subjets are discarded which removes any soft radiation from the subjets, whilst retaining any hard perturbative radiation from the decay products. For two prong decays, such as hadronically decaying vector bosons, the mass drop filtering algorithm with C/A jets of radius $R = 1.2$ has been shown to provide good discrimination against QCD backgrounds [209].

7.3 Jet Substructure Variables

Substructure variables can be categorised into jet shapes, splitting scales, subjettiness, and centre-of-mass jet shapes (where the jet is at rest, i.e. $p_{\mathrm{T}}^{\mathrm{jet}} = 0$), as defined in the following sections. The substructure variables are often defined with respect to the axis of the jet and depend on a distance measure in relation to this, as shown in Figure 7.3. The jet axis is taken as the axis in the direction of the jet momentum, unless otherwise specified. N-subjettiness, for example, uses

[2] $\mu_{\max}$ and $\sqrt{y_{\min}}$ are parameters that need to be configured for the algorithm.

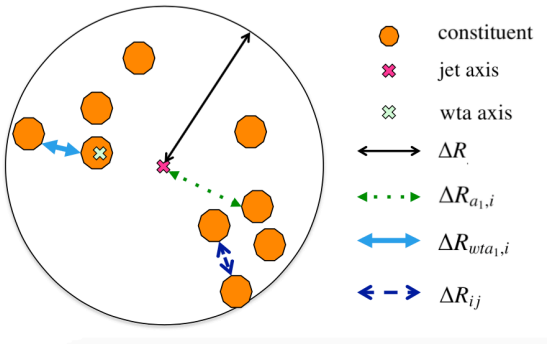


Figure 7.3 *The distance measures used in the calculation of the substructure variables [2]. This represents a jet cluster in (η, ϕ) space. The filled (orange) circles represent the constituents which are clustered together to form the jet. Here 'wta' stands for 'winner-takes-all', which is the axis along the hardest constituent.*

the direction of the subjet momentum to define axes of the individual subjets.

7.3.1 Jet Shapes

Jet shapes describe the relative positions and momenta of the jet constituents independently of subjets, essentially characterising the distribution of the energy within the jet. The jet shape variables under study in this work are the jet mass, energy correlation ratios, mass normalised angularity [217], and the planar flow. Their definitions are described in detail below.

7.3.1.1 Jet Mass

The jet mass was first introduced in Equation 4.1. In two prong decays, such as hadronic W boson decays, the jet mass can be approximated as

$$M^2 \approx p_{\mathrm{T}1} p_{\mathrm{T}2} \Delta R_{12}^2.. \tag{7.3}$$

7.3.1.2 Planar Flow

The planar flow [217] P quantifies the geometric distribution in η and ϕ of the jet energy perpendicular to the jet axis and how uniform the distribution is. P

is defined as

$$P = 4 \times \frac{\det(I_{ab})}{\mathrm{Tr}(I_{ab})^2}, \tag{7.4}$$

where I_{ab} is a matrix of momentum correlations,

$$I_{ab} = \frac{1}{M} \sum_{i \in J} \frac{p_{i,a} p_{i,b}}{E_i}, \tag{7.5}$$

and i is the index taken over all jet (J) constituents and a, b are the components of the p_{T} of the i^{th} jet constituent perpendicular to the jet axis. QCD jets are expected to be more isotropic than jets from W boson decays, having a value close to $P = 1$. W boson decays will have $P < 1$.

7.3.1.3 Energy Correlation Variables

Energy correlation variables are used to identify N-prong jet substructure (or N hard subjets) by considering the energy and pair-wise angles of the jet constituents. $(N+1)$-point correlators are defined which are sensitive to N-prong substructure. The energy correlation variables considered here are defined in terms of ratios of the 1-point, 2-point and 3-point energy correlation functions of the jet. The energy correlation functions are given by:

$$E_{\mathrm{CF1}}(\beta) = \sum_{i \in J} p_{\mathrm{T}_i}, \tag{7.6a}$$

$$E_{\mathrm{CF2}}(\beta) = \sum_{i<j \in J} p_{\mathrm{T}_i} p_{\mathrm{T}_j} (\Delta R_{ij})^{\beta}, \tag{7.6b}$$

$$E_{\mathrm{CF3}}(\beta) = \sum_{i<j<k \in J} p_{\mathrm{T}_i} p_{\mathrm{T}_j} p_{\mathrm{T}_k} (\Delta R_{ij} \Delta R_{ik} \Delta R_{jk})^{\beta}, \tag{7.6c}$$

where i is the index of the jet constituent in the jet J, and β is the weight of the angular separation between jet constituents. The 1-point energy correlation function, $E_{\mathrm{CF1}}(\beta)$, is approximately the jet p_{T}. If $\beta = 2$, then $E_{\mathrm{CF2}}(\beta = 2)$ is equivalent to the jet mass in a two prong decay. In the results presented here a value of $\beta = 1$ is used.

The energy correlation variables are converted into the dimensionless ratios:

$$e_2^{(\beta)} = \frac{E_{CF2}(\beta)}{E_{CF1}(\beta)^2}, \tag{7.7a}$$

$$e_3^{(\beta)} = \frac{E_{CF3}(\beta)}{E_{CF1}(\beta)^3}. \tag{7.7b}$$

The $e_2^{(\beta)}$ term is sensitive to radiation from jets with a single hard core, and $e_3^{(\beta)}$ to radiation from jets with two hard cores. For a jet with a single hard core, $e_2^{(\beta)} \neq 0$, whilst $e_3^{(\beta)} \to 0$. For two hard cores, both of these values will be non-zero, but they will grow at different rates, such the ratios between the two quantities can be useful to discriminate between single and two prong decays.

The quantities $e_2^{(\beta)}$ and $e_3^{(\beta)}$ are combined into the ratios $C_2^{(\beta)}$ [218] and $D_2^{(\beta)}$ [219, 220]:

$$C_2^{(\beta)} = \frac{e_3^{(\beta)}}{(e_2^{(\beta)})^2}, \tag{7.8a}$$

$$D_2^{(\beta)} = \frac{e_3^{(\beta)}}{(e_2^{(\beta)})^3}. \tag{7.8b}$$

These two variables have been shown to provide good separation between decays with one or two hard cores. For example, in W boson tagging, these both produce distributions which tend to peak at lower values than seen in multi-jet backgrounds, allowing for good separation between these processes [2] (see Figure 7.13 later in this chapter).

7.3.2 Splitting Scales

Splitting scale variables are designed to quantify the relative momenta and mass of subjets within a jet. They are defined from subjets identified through examining the clustering history of the jet, and the subsequent reclustering of the constituents using either the k_t or C/A algorithm. The variables that are considered here, and defined below, are $\sqrt{d_{12}}$, $\sqrt{z_{12}}$, YFilt, mass drop $\sqrt{\mu_{12}}$ and momentum balance $\sqrt{y_{12}}$. The variables $\sqrt{\mu_{12}}$ and $\sqrt{y_{12}}$ are defined in Equations 7.1 and 7.2.

The splitting scale $\sqrt{d_{12}}$ is calculated for a jet (re)clustered with the k_t-clustering

algorithm [221]. It is the k_t distance between the two proto-jets of the final clustering step:

$$\sqrt{d_{12}} = \min(p_{\mathrm{T}_1}, p_{\mathrm{T}_2}) \times \Delta R_{12}, \tag{7.9}$$

where 1 and 2 are the two proto-jets. $\sqrt{d_{12}}$ can be used to distinguish the products of the W boson, which tend to have a symmetric energy distribution and larger values of $\sqrt{d_{12}}$, from QCD jets, which are more asymmetric on average. The dimensionless YFilt variable is given by:

$$\mathrm{YFilt} = \sqrt{d_{12}}/M. \tag{7.10}$$

The dimensionless variable $\sqrt{z_{12}}$ is a mass-related variation of $\sqrt{d_{12}}$ [222], given by:

$$\sqrt{z_{12}} = \left(\frac{d_{12}}{d_{12} + M^2}\right)^{\frac{1}{2}}. \tag{7.11}$$

7.3.3 Subjettiness

Subjettiness discriminates between jets that exhibit well-formed substructure and those that do not. A boosted jet containing a W boson decaying to two quarks should have two distinct hard subjets with an invariant mass of approximately 80 GeV. Boosted QCD jets of this mass can originate from a single hard parton, acquiring mass through large angle soft splittings. Subjettiness attempts to exploit this difference by identifying the number of hard subjets within the boosted jet.

N-subjettiness (τ_N) variables consider the subjet multiplicity, describing the degree to which the jet substructure resembles N or fewer subjets [223, 224]. The constituents of the jet J are clustered with the exclusive k_t algorithm into exactly N subjets, defining N subjet axes a_N. The τ_N variables are defined as a sum over all of the original jet constituents, multiplied by the p_{T}-weighted distance between each constituent and its nearest subjet. If the constituents are localised near the subjet axes, a lower τ_N value is obtained, describing how well the jet can be described as having N or fewer subjets.

The subjettiness variables relevant for two prong decays are defined as:

$$\tau_0(\beta) = \sum_{i \in J} p_{\mathrm{T}_i} \Delta R_0^{\beta}, \tag{7.12a}$$

$$\tau_1(\beta) = \frac{1}{\tau_0} \sum_{i \in J} p_{\mathrm{T}_i} \Delta R_{a_1,i}^{\beta}, \tag{7.12b}$$

$$\tau_2(\beta) = \frac{1}{\tau_0} \sum_{i \in J} p_{\mathrm{T}_i} \min(\Delta R_{a_1,i}^{\beta}, \Delta R_{a_2,i}^{\beta}), \tag{7.12c}$$

where i is the sum over the constituents of the original jet J, ΔR_0 is the original jet radius, $\Delta R_{a_N,i}$ is the angular separation between jet constituent i and the axis of one of the N exclusive subjets, τ_0 is a normalisation factor, and β provides a weight to the angular separation.

When calculating τ_1 and τ_2, the jet constituents are reclustered using the k_t algorithm into exactly one and two subjets, respectively. The labels a_1 and a_2 in Equations 7.12b and 7.12c refer to the axes of the subjets in these two cases, and $\Delta R_{a_{1/2},i}$ is the angular distance between jet constituent i and the subjet axes. For τ_2, $\min(\Delta R_{a_1,i}, \Delta R_{a_2,i})$ is the distance between the jet constituent i and the closest subjet.

For two prong decays, the k_t algorithm usually identifies two subjets with all jet constituents clustered around their axes, having relatively small angular separation. This translates to much smaller values of τ_2 than τ_1. Whereas for one prong decays there should be a similar value for both τ_2 and τ_1.

An alternative implementation of this method can be defined with a different jet axis definition in the determination of $\Delta R_{a_N,i}$. In the standard definition, the subjet axis is used. In the 'winner-takes-all' definition, the hardest constituents of the subjets are chosen, instead of the subjet axes, as the a_N axes in Equations 7.12b and 7.12c (see Figure 7.3). This choice has shown improved discrimination power in recent studies [225].

The dimensionless ratio of τ_2 and τ_1 is useful in identifying two prong decays, defined as

$$\tau_{21} = \frac{\tau_2}{\tau_1}, \ \tau_{21}^{\mathrm{wta}} = \frac{\tau_2^{\mathrm{wta}}}{\tau_1^{\mathrm{wta}}}. \tag{7.13}$$

The superscript 'wta' denotes that the 'winner-takes-all' axes are used. As for the τ_2 variable, the W boson decays should exhibit smaller values of these ratios.

7.3.4 Centre-of-Mass Jet Shapes

The centre-of-mass jet shapes use jet constituents that are transformed from collider coordinates into the rest frame of the jet. This includes properties such as thrust major and minor axes ($T_{\mathrm{maj/min}}$) [226–229], Fox-Wolfram moments [230], sphericity and aplanarity.

The jet sphericity and aplanarity are a measure of the jet topology. This is described in terms of an eigenvector problem, where the jet sphericity tensor $S^{\alpha,\beta}$ must be diagonalised.

$$S^{\alpha,\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i |\vec{p}_i|^2}, \tag{7.14}$$

where $\alpha, \beta = \{1, 2, 3\}$ correspond to the x, y and z components of the momenta of the jet energy clusters in the jet rest frame.

Diagonalisation of the tensor yields three eigenvalues and eigenvectors, constrained by $\lambda_1 \geq \lambda_2 \geq \lambda_3$ and $\lambda_1 + \lambda_2 + \lambda_3 = 1$.

The sphericity S and aplanarity A are given by linear combinations of the eigenvalues.

$$S = \frac{3}{2}(\lambda_2 + \lambda_3),\ A = \frac{3}{2}\lambda_3. \tag{7.15}$$

The sphericity is a measure of the sum of the squares of the transverse momenta of the jet constituents with respect to the event axis (the first eigenvector). By construction, $0 \leq S \leq 1$. Two back-to-back subjets in the jet rest frame have $S \to 0$, whereas an isotropic distribution of energy clusters would have $S \to 1$. The more jets there are in an isotropic distribution, the more spherical it will be. The aplanarity is a measure of the transverse momentum component out of the jet plane constrained to the range $0 \leq A \leq \frac{1}{2}$. A planar event, where there is a highly directional distribution of the clusters has $A \to 0$, whereas an isotropic distribution has $A \to 0.5$. The QCD background is expected to more spherical and isotropic than W boson decays.

The substructure variables defined here are summarised in Table 7.1. Examples of the substructure variables are shown in Figure 7.13 for the datasets defined in the following section.

Variable	W boson	QCD
Jet mass M	$M \to 80$ GeV	$M << 80$ GeV
Planar flow P	$P \ll 1$	$P \to 1$
Energy correlation $C_2^{(\beta=1)}$	Peak closer to 0 for W bosons	
Energy correlation $D_2^{(\beta=1)}$	$D_2^{(\beta=1)} \gg C_2^{(\beta=1)}$, peak closer to 0 for W bosons	
$\sqrt{d_{12}}$	Symmetric distribution	Asymmetric
YFilt	Asymmetric, skewed to lower values	Asymmetric, skewed to higher values
$\sqrt{z_{12}}$	Asymmetric, skewed to higher values	Asymmetric, skewed to lower values
τ_{21}	Peak closer to 0 for W bosons	
τ_{21}^{wta}	Lower values for W boson	
Sphericity S	$S \to 1.0$	$S \to 0$
Aplanarity A	$A \to 0.5$	$A \to 0$

Table 7.1 *Summary of the jet substructure variables described in Section 7.3.*

7.4 Monte Carlo Samples

The Monte Carlo samples used for these studies consist of one set of samples generated for $\sqrt{s} = 8$ TeV and another set generated for 13 TeV, corresponding to conditions during Run 1 and Run 2, respectively. The samples are discussed in further detail below and the 13 TeV samples are listed in Table 7.2. The samples are produced with the same generators in both cases, but with different tunes and PDFs. All generated MC events are passed through a GEANT4 [108] simulation of the ATLAS detector [109]. In the following, the $\sqrt{s} = 8$ TeV samples are those used in the cut-based W boson tagging studies in Reference [2] which are discussed in Section 7.5. The $\sqrt{s} = 13$ TeV samples are those used for the machine learning studies presented in Section 7.6.

For the $\sqrt{s} = 8$ TeV samples, the pile-up in events is emulated by overlaying the generated hits with hits taken from minimum-bias events in data. In the $\sqrt{s} = 13$ TeV samples, pile-up is not overlaid. It has been shown that there is only a weak dependence of the jet substructure and jet grooming algorithms (in particular the trimming algorithms) on pile-up, and as such this is not considered [2, 231]. A further consideration is due to the reweighting procedures that are used to account for pile-up. One of the concerns in the training of the ML taggers that are used is the number of available events for training and testing, and the reweighting for the pile-up contributions will effectively reduce the number of events available.

In the presented analysis, a number of different jet collections, using different

clustering and grooming algorithms, are reconstructed for each event. In order to ensure that the same jets are being identified in different collections, the jets in each of these collections are matched within a radius of ΔR.

The collection of ungroomed C/A jets with $R = 1.2$, intended to capture the entire decay of a hadronically decaying boosted W boson in a single jet, are used for the initial selection. Further selections are applied on the collections of groomed jets (including trimmed, pruned and split filtered jets) with radii of between $R = 1.0$ and $R = 1.2$.

Signal Samples

Samples of high-p_{T} W bosons are obtained at $\sqrt{s} = 8$ TeV from generated $W' \to WZ \to q\bar{q}\ell^+\ell^-$ events with different W' masses between 400 and 2000 GeV. The MC samples are generated using Pythia8 (8.165) with the AU2 [232] tune and the MSTW20080LO [233] PDF set.

The signal samples for the process $W' \to WZ \to q\bar{q}q\bar{q}$ are simulated at $\sqrt{s} = 13$ TeV for multiple W' masses between 1.5 TeV and 3 TeV. These are generated with Pythia8 (8.186), using the A14 [234] tune and the NNPDF2.3LO PDF set [93]. The properties of the bottom and charm hadron decays are simulated with EvtGen (1.2.0) [235].

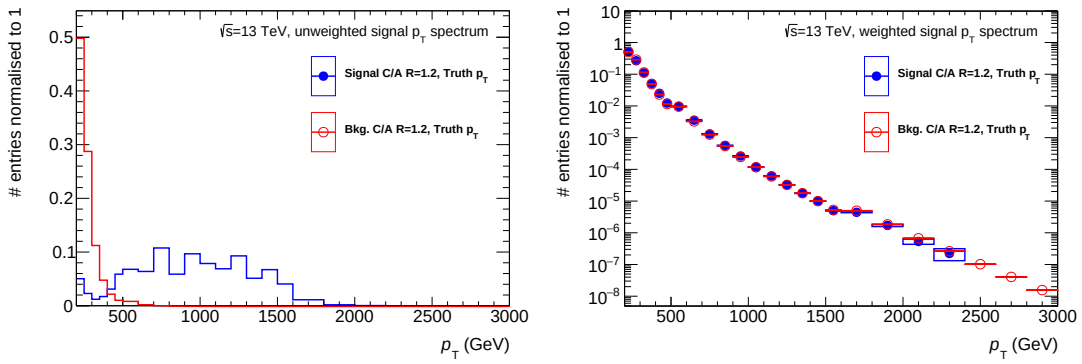
The W' boson signal sample has a much harder p_{T} spectrum than the multi-jet background. In order to have a more uniform comparison, the W' samples are reweighted such that the p_{T} of the highest momentum ungroomed C/A ($R = 1.2$) jet matches that of the multi-jet background sample introduced below. This is illustrated in Figure 7.4.

Background Samples

Background samples of high-p_{T} multi-jet events are generated. These are generated in sets according to the p_{T} of the leading ungroomed C/A ($R = 1.2$) jet in the event: JZ4: [400, 800] GeV, JZ5: [800, 1300] GeV, JZ6: [1300, 1800], and JZ7: [1800, 2500]. All of these samples are weighted by their relative cross-sections to produce a smoothly falling p_{T} distribution. For $\sqrt{s} = 13$ TeV, the samples are produced using Pythia with the same tune and PDF as the $\sqrt{s} = 13$ TeV signal samples. The $\sqrt{s} = 8$ TeV samples are produced using

Pythia with the same tune as the $\sqrt{s} = 8$ TeV signal samples, and the CT10 [89] PDF set.

The total cross sections and number of events created for the $\sqrt{s} = 13$ TeV signal and background samples are listed in Table 7.2.



(a) **(b)**

Figure 7.4 *The $\sqrt{s} = 13$ TeV W' signal and multi-jet background leading ungroomed C/A ($R = 1.2$) p_T distributions before and after p_T reweighting has been applied to the signal.*

7.5 Run 1 W Tagger

7.5.1 Introduction

To achieve an optimal separation of boosted W bosons and the QCD background, the jet grooming algorithms and jet substructure variables introduced earlier in this chapter were studied, based on Run 1 conditions at $\sqrt{s} = 8$ TeV [2]. In these studies, the boosted boson tagging was performed using a mass window requirement on the groomed jet mass, combined with a subsequent cut on a jet substructure variable. This is referred to as the *cut-based* tagger in the following. The results from these studies were presented as the baseline recommendation for identifying boosted W bosons in Run-2 analyses at ATLAS [2]. These recommendations are evaluated on $\sqrt{s} = 13$ TeV MC samples in Section 7.6, from which the groomer is chosen for the machine learning implementation in Section 7.7.

Sample, $m_{W'}$ (GeV)	Cross section (fb)	Filter Eff.	# Events
$W' \to WZ \to q\bar{q}q\bar{q}$ mass points (GeV)			
1500	32.33	1	30000
1700	17.78	1	29000
1800	13.42	1	30000
1900	10.27	1	29000
2000	7.87	1	50000
2100	6.12	1	30000
2400	2.97	1	30000
2500	2.37	1	30000
2600	1.89	1	30000
2800	1.23	1	29000
2900	0.99	1	29000
3000	0.81	1	29800
Multi-jet JZx samples			
JZ4	2.55×10^8	5.3×10^{-4}	1997000
JZ5	4.55×10^6	9.2×10^{-4}	1995000
JZ6	2.58×10^5	9.4×10^{-4}	1997000
JZ7	1.62×10^4	3.9×10^{-4}	1990000

Table 7.2 *The $\sqrt{s} = 13$ TeV MC samples used for the machine-learning-based W tagging studies. Listed here are the cross sections, filter efficiencies and number of events of each of the samples used. The JZx samples refer to different p_T slices of the multi-jet background which are given in the text, where x runs from 4 to 7. In the JZx samples, generator level cuts are applied on the jet p_T. The filter efficiency is the fraction of the total phase space that this specific sample corresponds to, with the generator level cuts.*

7.5.2 Event Selection

Event selections are defined based both on ungroomed and groomed jets. The ungroomed jet, regardless of the grooming algorithm under consideration, is reconstructed using the C/A algorithm with a radius parameter of $R = 1.2$. Unless otherwise indicated, the jet being referred to in the following text is the reconstructed jet with no jet calibration applied. Correlations between the W boson p_{T} and its jet substructure are accounted for by categorising the events by the p_{T} of the leading ungroomed C/A jet with $R = 1.2$, using stable, truth level particles as inputs. This large-R C/A jet is used as a rough identification of the hadronically decaying W boson. Since jets reconstructed with each grooming algorithm could potentially have different p_{T}, this ensures that each grooming algorithm is compared in the same region of phase space.

Preselection

Selection is first performed based on the ungroomed C/A jet, thereafter the selection is based on groomed jet properties. The leading ungroomed $R = 1.2$ C/A truth jet must be within $|\eta| < 1.2$ and have $p_{\mathrm{T}} > 50$ GeV, such that it is within tracking acceptance. Events where the ungroomed reconstructed C/A jet is within $\Delta R = 0.9$ of the ungroomed C/A truth jet are used for the p_{T}-based reweighting. The truth matching also allows for the jet mass response of uncalibrated jets to be evaluated.

The leading groomed reconstructed jet must be within $|\eta| < 1.2$ and have less than 99% of the total energy of the jet in the EM cluster. This removes contamination from $Z \to ee$ events which are far more collimated. The leading reconstructed groomed jet must be within $\Delta R < 0.75 \times R$ of the leading groomed truth jet, where R is the radius of the grooming algorithm. Approximately 2% of W bosons are unmatched using this procedure.

The substructure of the W boson jets is expected to be correlated with the p_{T}. As such, bins are defined using the generated p_{T} of the highest p_{T} jet, as reconstructed by the C/A algorithm, before any grooming has been applied. The ranges are defined in terms of the truth p_{T} ($p_{\mathrm{T}}^{\mathrm{truth}}$) of the leading jet: 200-350 GeV, 350-500 GeV and 500-1000 GeV.

Selection

After the preselection, a jet mass window is defined as the smallest mass window, calculated over a binned mass histogram, that contains 68% of all signal events. This mass window is based on the groomed jet mass and is different for each groomer. All events where the leading groomed jet has a mass outside the mass window are removed; this removes a large portion of the background. The mass windows for three p_{T} ranges in Table 7.5. An example of the jet mass distribution for one of the groomers is shown in Figure 7.10. The mass window gives a baseline signal efficiency of $\epsilon_W^{\mathrm{G}} = 68\%$, where 'G' indicates that grooming has been applied, but no substructure variables have been considered yet. The background efficiency of the mass cut, $\epsilon_{\mathrm{QCD}}^{\mathrm{G}}$, is defined as the fraction of background events passing preselection that fall within the 68% mass window. Minimising $\epsilon_{\mathrm{QCD}}^{\mathrm{G}}$ is used as the primary criterion for assessing the performance of the grooming algorithm here.

Finally, a cut is applied on a substructure variable. The value of the cut is chosen such that it gives a signal efficiency of $\epsilon_W^{\mathrm{G\&T}} = 50\%$, where 'T' stands for 'Tagger' (the substructure variable) and indicates that a cut has been applied and the jet has been groomed. After applying this cut, the inverse of the background efficiency, or the *background rejection power*, is calculated and used as the metric of performance for this combination of groomer and substructure variable. The background rejection factor is given by

$$\text{Background rejection power} = 1/\epsilon_{\mathrm{QCD}}^{\mathrm{G\&T}}. \tag{7.16}$$

7.5.3 Tagger Optimisation

Initially, jet algorithms were selected with more than 500 configurations of grooming algorithms with different choices for the parameters such as R, R_{sub}, as shown in Table 7.3. The 27 best performing grooming algorithms were chosen based on the minimisation of $\epsilon_{\mathrm{QCD}}^{\mathrm{G}}$. This was performed separately for every p_{T} bin.

Each of the selected 27 jet collections were tested in combination with 26 different substructure variables, as presented in Section 7.3. For each of these combinations of groomer and substructure variable the background rejection power at 50% signal efficiency was used to evaluate their performance in every p_{T} bin.

Trimming configurations

Jet algorithms	R	R_{sub}	f_{cut} (%)
C/A, anti-k_t	0.6, 0.8, 1.0, 1.2	0.1, 0.2, 0.3	1, 2, 3, 4, 5, 7, 9, 11, 13, 15

Pruning configurations

Jet algorithms	R	Reclust. alg.	Z_{cut} (%)	R_{cut}
C/A, anti-k_t	0.8, 1.0, 1.2	C/A	10, 15, 20, 25, 30	$\frac{1}{100}$, $\frac{1}{10}$, $\frac{1}{8}$, $\frac{1}{4}$, $\frac{1}{2}$, 1.0

Split filtering configurations

Jet algorithms	R	R_{sub}	μ_{max}	y_{cut}
C/A	0.8, 1.0, 1.2	0.3, min(0.3, $\frac{\Delta R}{2}$)	67, 78, 89, 100	0.06, 0.07, ..., 0.20

Table 7.3 *Details of the different trimming, pruning and split filtering configurations used for the jet grooming algorithms. All combinations of these parameters were explored in Reference [2].*

7.5.4 Results

The results of the study in Reference [2] show that a pairwise combination of the groomed jet mass and a single substructure variable can achieve 50% identification efficiency for W bosons with $p_{\text{T}} > 200$ GeV, whilst keeping the multi-jet background selection efficiency at only 2-4%. The best performing grooming algorithm configuration is the anti-k_t, $R = 1.0$ trimmed jet groomer with f_{cut}=0.05 and R_{sub} =0.2, and is thus the recommended strategy proposed in Reference [2]. Energy correlation variables $C_2^{(\beta)}$ and $D_2^{(\beta)}$, and the N-subjettiness ratio τ_{21}^{wta} all provide equally good performance over the entire p_{T} range which was evaluated.

Four of the 27 jet collections (given in Table 7.4) were selected for further study. These four collections showed consistent performance over the full p_{T} range, with roughly equal efficiencies for each algorithm in each p_{T} bin. The effect of pile-up is evaluated by the correlation between the average groomed jet mass, $\langle M \rangle$, and the number of primary vertices, NPV. For ungroomed jets, an increase of $\approx$ 2 GeV per additional vertex is seen in $\langle M \rangle$, whilst for the groomed jets this is almost negligible, as shown in Table 7.5. The anti-k_t, $R = 1.0$ trimmed jet groomer with f_{cut}=0.05 and R_{sub} =0.2 displays less dependence on pile-up than

the other groomers do, with an increase of $0.1 - 0.2$ GeV per additional vertex. This algorithm is referred to as *R2-trimming*.

The best three performing substructure variables over all p_{T} ranges for the four jet collections were consistently found to be τ_{21}^{wta}, $C_2^{(\beta=1)}$ and $D_2^{(\beta=1)}$ (as defined in Equations 7.13, 7.8a and 7.8b, respectively). The cut values for these are shown for R2-trimming in Table 7.5. These cuts give background rejection powers of ≈ 50, corresponding to $\epsilon_{QCD}^{\mathrm{G\&T}} \approx 2\%$, with relative systematic uncertainties of between 16% (low p_{T} bin) and 25% (high p_{T} bin).

Grooming configuration	$\epsilon_{\mathrm{QCD}}^{\mathrm{G}}$	$\delta\langle M\rangle/\delta\mathrm{NPV}$
anti-k_t, $R = 1.0$, trimmed, f_{cut}= 0.05, $R_{\mathrm{sub}} = 0.2$	$(11 \pm 1)\%$	$0.1 - 0.2$ GeV
anti-k_t, $R = 1.0$, trimmed, f_{cut}= 0.05, $R_{\mathrm{sub}} = 0.3$	$(16 \pm 1)\%$	$0.5 - 0.6$ GeV
C/A, $R = 1.0$, pruned, $Z_{\mathrm{cut}} = 0.15$, $R_{\mathrm{cut}} = 0.5$	$(16 \pm 2)\%$	$0.9 - 1.1$ GeV
C/A, $R = 1.2$, split-filt, $\sqrt{y_{12}} = 0.15$, $R_{\mathrm{sub}} = 0.3$	$(13 \pm 1)\%$	$0.1 - 0.3$ GeV

Table 7.4 *The four best performing grooming configurations and their background efficiencies, and pile-up dependence for $\epsilon_W^{\mathrm{G}} = 68\%$ in the range $200 < p_T < 350$ GeV. The uncertainty on $\epsilon_{\mathrm{QCD}}^{\mathrm{G}}$ indicates combined statistical and systematic uncertainties. $\delta\langle M\rangle/\delta NPV$ indicates the average increase in $\langle M\rangle$ per additional primary vertex (δNPV). Ungroomed jets showed an increase of $\approx$ 2 GeV per additional vertex.*

Variable		Tagging criteria in p_{T} range ($\epsilon_{\mathrm{QCD}}^{\mathrm{G\&T}}$%)		
		200–350 GeV	350–500 GeV	500–1000 GeV
$\epsilon_W^{\mathrm{G}} = 68\%$ mass range		61–93 GeV	71–91 GeV	73–91 GeV
$\epsilon_W^{\mathrm{G\&T}} = 50\%$	$C_2^{(\beta=1)}$	< 0.18 (3.5%)	< 0.13 (2.1%)	< 0.10 (2.1%)
	$D_2^{(\beta=1)}$	< 1.14 (3.1%)	< 1.23 (2.6%)	< 1.35 (2.3%)
	τ_{21}^{wta}	< 0.32 (3.1%)	< 0.36 (3.0%)	< 0.40 (2.6%)

Table 7.5 *The mass windows for calibrated R2-trimmed jets that provide $\epsilon_W^{\mathrm{G}} = 68\%$, and the requirements on the three substructure variables that result in the lowest background efficiencies $\epsilon_{\mathrm{QCD}}^{\mathrm{G\&T}}$ (indicated as a percentage in bracketed terms), when combined with the mass window requirement. These cuts are chosen such that $\epsilon_W^{\mathrm{G\&T}} = 50\%$.*

7.6 Run 2 Tagger Crosscheck

7.6.1 Introduction

The work introduced in Section 7.5 was based on MC simulations using Run 1 conditions at $\sqrt{s} = 8$ TeV. The conditions during Run 2 at $\sqrt{s} = 13$ TeV are somewhat different and particles can obtain a greater boost in p_{T}. The recommendations from the Run 1 studies were derived as functions of p_{T} up to 1 TeV. However, higher p_{T} ranges need to be explored. A crosscheck of the Run 1 recommendations as applied to Run 2 MC is shown in Section 7.6.2.

The cut-based tagger introduced for the Run 1 studies presented in the previous section is limited to using simple two-variable taggers: the mass of the groomed jet, and a single substructure variable. A method of combining several substructure variables, using machine learning classifiers from Chapter 6, is investigated in Section 7.7 to create *machine learning (ML) taggers*. Two types of ML tagger are presented: deep neural networks that are based on a set of stacked auto-encoders, and boosted decision trees. These taggers are used in conjunction with the same groomed jet mass requirement as the cut-based tagger.

The baseline grooming algorithms and substructure variables are tested using the $W' \to WZ \to q\bar{q}q\bar{q}$ signal process and a multi-jet background (listed in Table 7.2), using jets with transverse momentum $p_{\mathrm{T}} > 200$ GeV. The grooming algorithms are first evaluated using the same criteria as defined in Section 7.5.2: the minimisation of the background efficiency $\epsilon^{\mathrm{G}}_{\mathrm{QCD}}$. The grooming algorithm that performs the best is chosen for further studies in Section 7.7.

7.6.2 Event Selection

The majority of the preselection criteria are the same as those given in Section 7.5.2, with additional truth matching of the selected jets and the parent W boson to reduce background from Z boson jets when considering the signal samples:

- at least 2 tracks from the primary vertex are required,
- the leading, ungroomed, truth C/A jet must have $p_{\mathrm{T}} > 50$ GeV and $|\eta| < 1.2$ (to be within tracking acceptance),

- background events where the leading ungroomed C/A jet is within a cone of radius $\Delta R = 0.75 \times R$ (where R is the radius parameter of the grooming algorithm) of the above truth, ungroomed jet are used for the p_{T} reweighting of the signal sample,
- the leading groomed jet must be within $|\eta| < 1.2$ (such that large-R jets are within tracking acceptance),
- the leading groomed jet must be matched within a cone of radius $\Delta R = 0.75 \times R$ of truth W boson parent, where the truth W boson parent has $p_{\mathrm{T}} > 5$ GeV,
- the leading groomed jet must be matched within a cone of radius $\Delta R = 0.75 \times R$ of the truth groomed jet,
- a 68% mass window selection is applied to the mass of the leading reconstructed groomed jet.

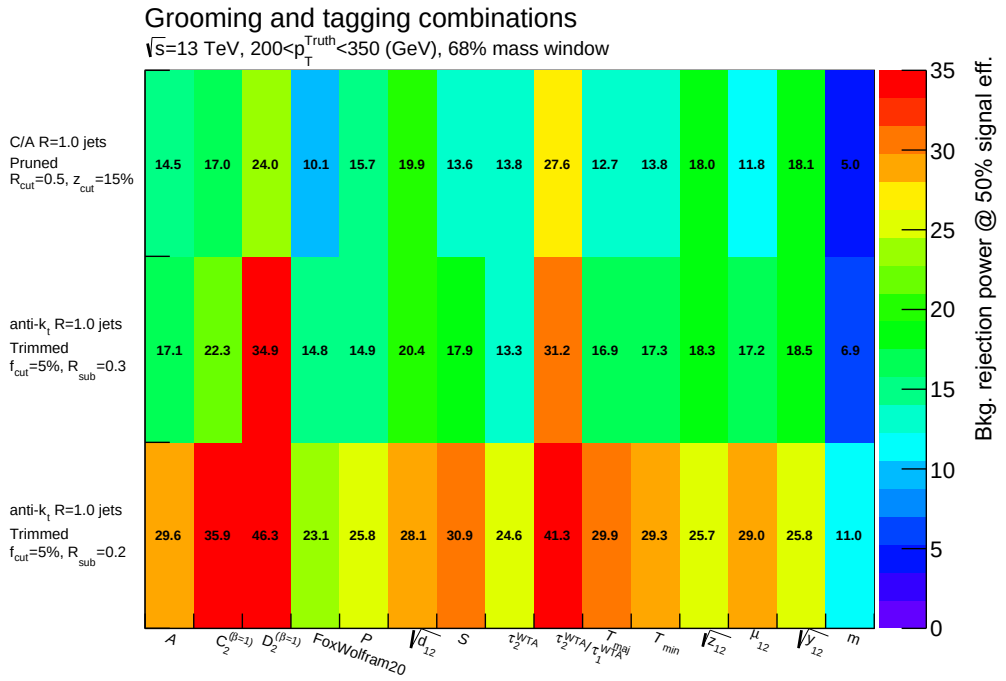
The four grooming algorithms that were recommended in Section 7.5, given in Table 7.4, were re-evaluated on $\sqrt{s} = 13$ TeV MC samples.

7.6.3 Crosscheck Results

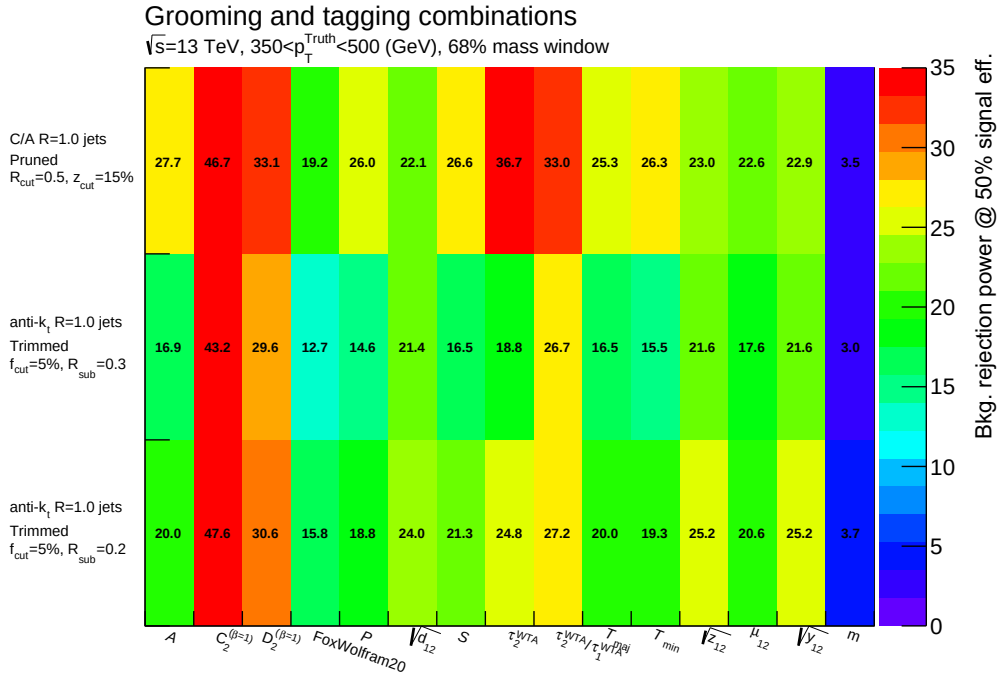
In the Run 1 cut-based studies (Section 7.5) the performance of the tagger was evaluated in three p_{T} bins. In the crosscheck of the cut-based results using $\sqrt{s} = 13$ TeV MC, an additional high p_{T} bin of 1000-1500 GeV is used. The performance of multiple cut-based taggers using three of the jet groomers, with a 68% mass window cut on the groomed jet mass and the pairwise combination with multiple substructure variables, is shown in Figures 7.5 and 7.6. The performance of the C/A, $R = 1.2$ split-filtering groomer with $\sqrt{y_{12}} = 0.15$ is compared with the anti-k_t, $R = 1.0$ trimming groomer with f_{cut}= 0.05 and $R_{\mathrm{sub}} = 0.2$ in Figure 7.7. The background rejection factors in these figures show similar results to the Run 1 cut-based taggers in Section 7.5.4.

From these background rejection power plots (Figures 7.5 and 7.6), it is observed that the variables $C_2^{(\beta=1)}$ and $D_2^{(\beta=1)}$ offer the highest rejection factors, typically within a range of 40-50, over all p_{T} bins. The variable τ_{21}^{wta} offers a high rejection power of between 35 and 40 for all groomers in the bins where $p_{\mathrm{T}} > 500$ GeV (and the lowest p_{T} bin for the R2-trimming configuration). Apart from the p_{T} bin

$350 < p_{\mathrm{T}} < 500$ GeV, shown in Figure 7.5(b), the R2-trimming groomer shows very consistent performance in all three of these variables, and in general has higher background rejection power than the other groomers. The full ROC curves of the background rejection power for three of the groomers in combination with the variables $C_2^{(\beta=1)}$, $D_2^{(\beta=1)}$ and τ_{21}^{wta} are shown in Figures 7.8 and 7.9. Based on these results, anti-k_t, $R = 1.0$ trimmed jets with $f_{\mathrm{cut}}= 0.05$ and $R_{\mathrm{sub}} = 0.2$ was chosen as the default groomer to be used for the BDT and DNN W boson taggers in Section 7.7.

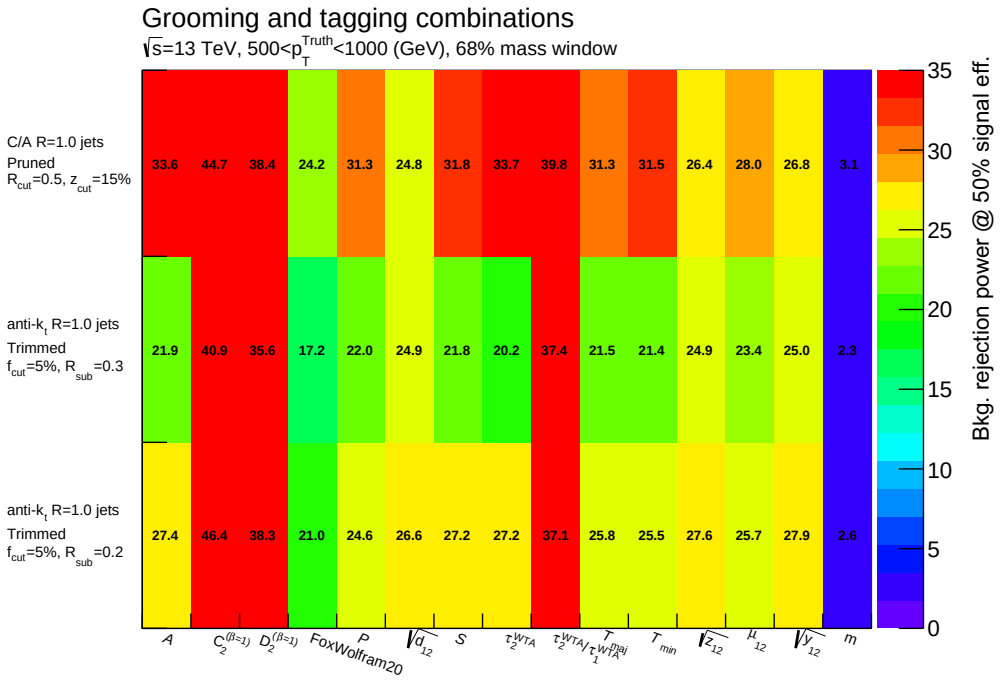


(a) $200 < p_T^{Truth} < 350$ GeV

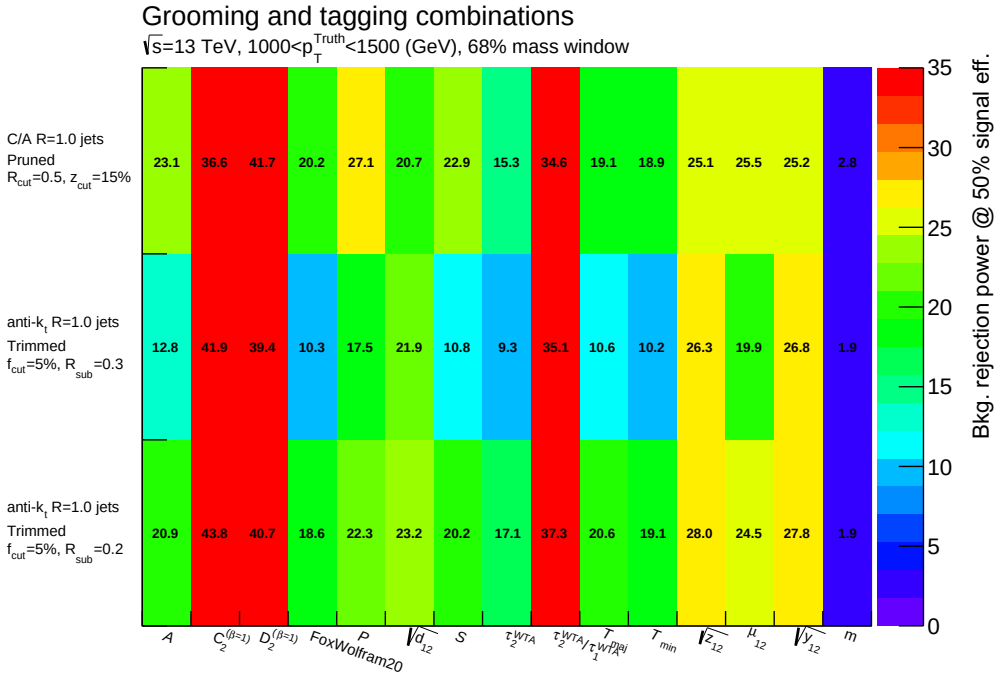


(b) $350 < p_T^{Truth} < 500$ GeV

Figure 7.5 *Background rejection power* $(1/\epsilon_{\text{QCD}}^{\text{G\&T}})$ *for* W *boson identification for different combinations of jet grooming algorithms and substructure variables. There is a 68% mass window cut on the groomed jet mass and a cut on the substructure variable, which gives a signal efficiency of 50%. Plots (a) and (b) show the rejection power for events with a leading ungroomed C/A jet with* $200 < p_T^{Truth} < 350$ GeV*, and* $350 < p_T^{Truth} < 500$ GeV*, respectively. Additional definitions for the substructure variables on the x-axis can be found in Appendix G.*



(a) $500 < p_T^{Truth} < 1000$ GeV



(b) $1000 < p_T^{Truth} < 1500$ GeV

Figure 7.6 *Background rejection power* $(1/\epsilon_{\mathrm{QCD}}^{\mathrm{G\&T}})$ *for* W *boson identification for different combinations of jet grooming algorithms and substructure variables. There is a 68% mass window cut on the groomed jet mass and a cut on the substructure variable, which gives a signal efficiency of 50%. Plots (a) and (b) show the rejection power for events with a leading ungroomed C/A jet with* $500 < p_T^{Truth} < 1000$ GeV*, and* $1000 < p_T^{Truth} < 1500$ GeV*, respectively. Additional definitions for the substructure variables on the* x*-axis can be found in Appendix G.*

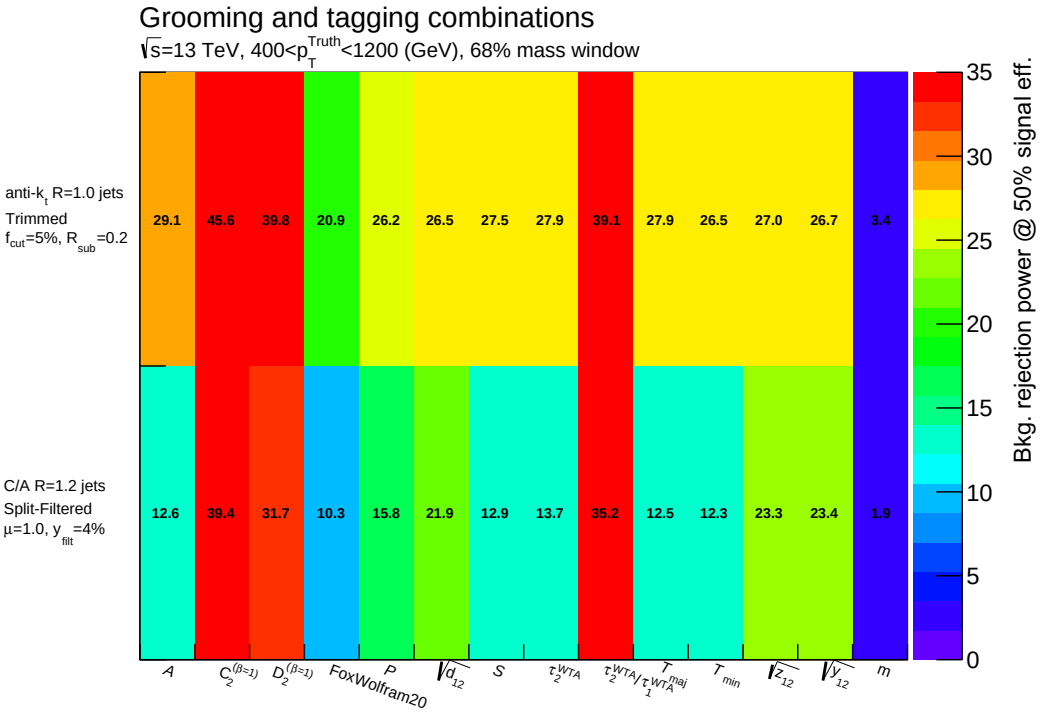
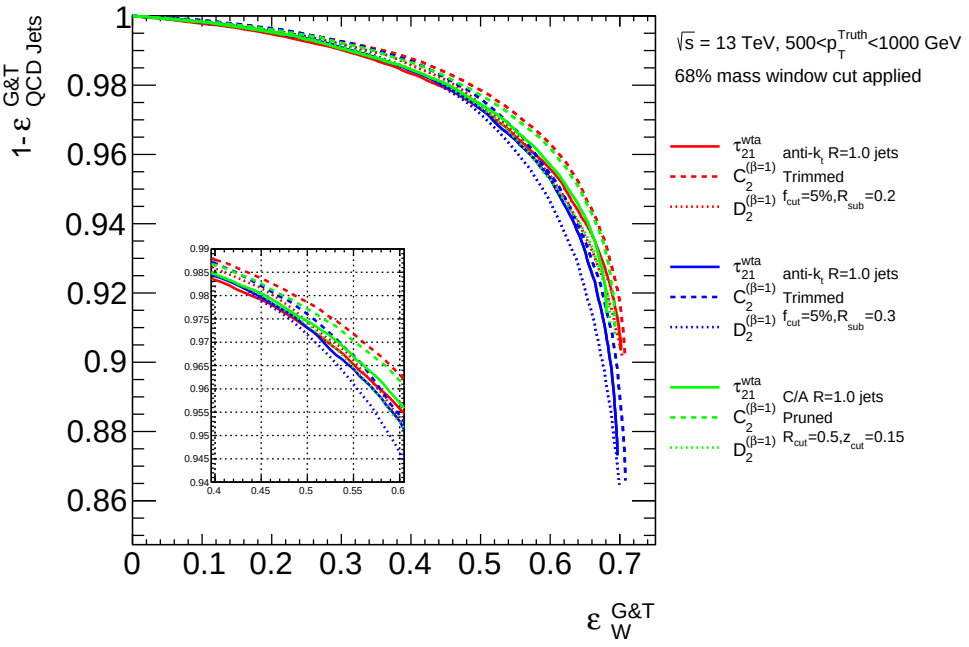
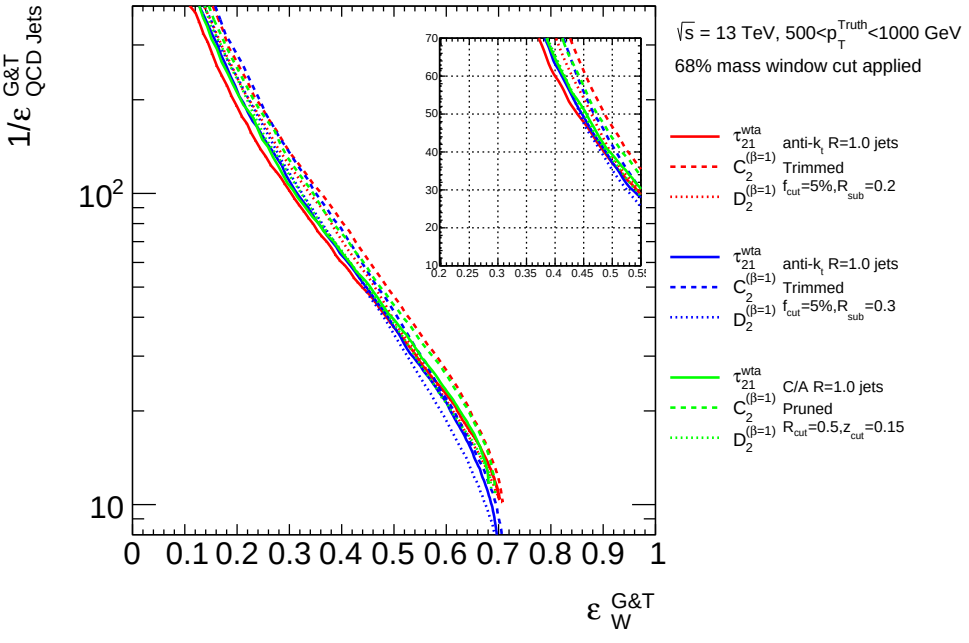


Figure 7.7 *Background rejection power (1/background efficiency, or $1/\epsilon_{\mathrm{QCD}}^{\mathrm{G\&T}}$) values for W boson identification when using different combinations of jet grooming algorithms and substructure variables. There is a 68% mass window cut on the groomed jet mass, followed by a cut on the substructure variable, which gives a signal efficiency of 50%. These show the rejection power for events with a leading ungroomed C/A jet with $400 < p_T^{Truth} < 1200$ GeV. Additional definitions for the substructure variables on the x-axis can be found in Appendix G.*

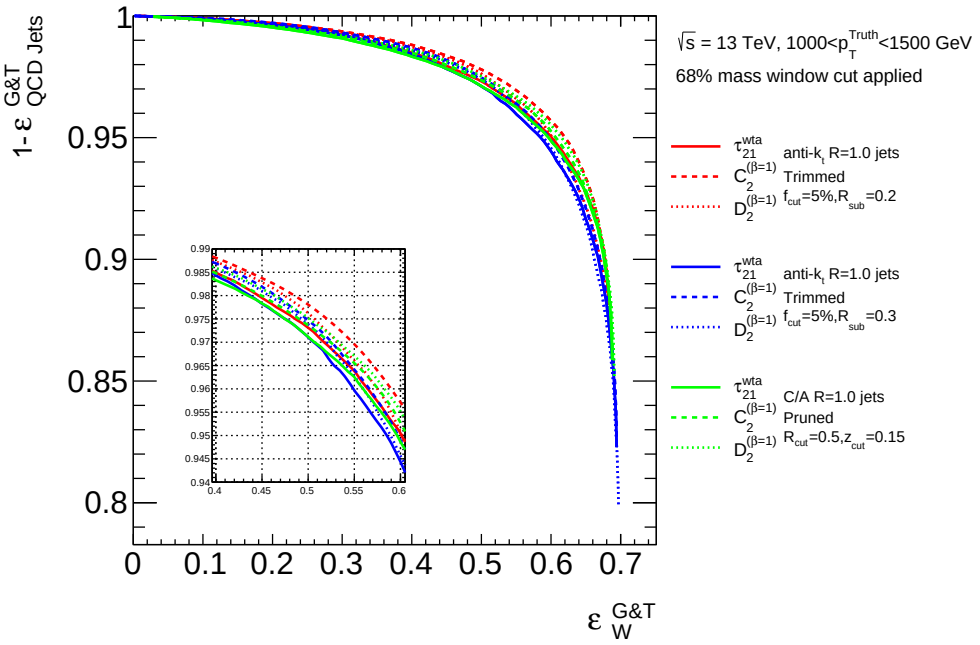


(a)

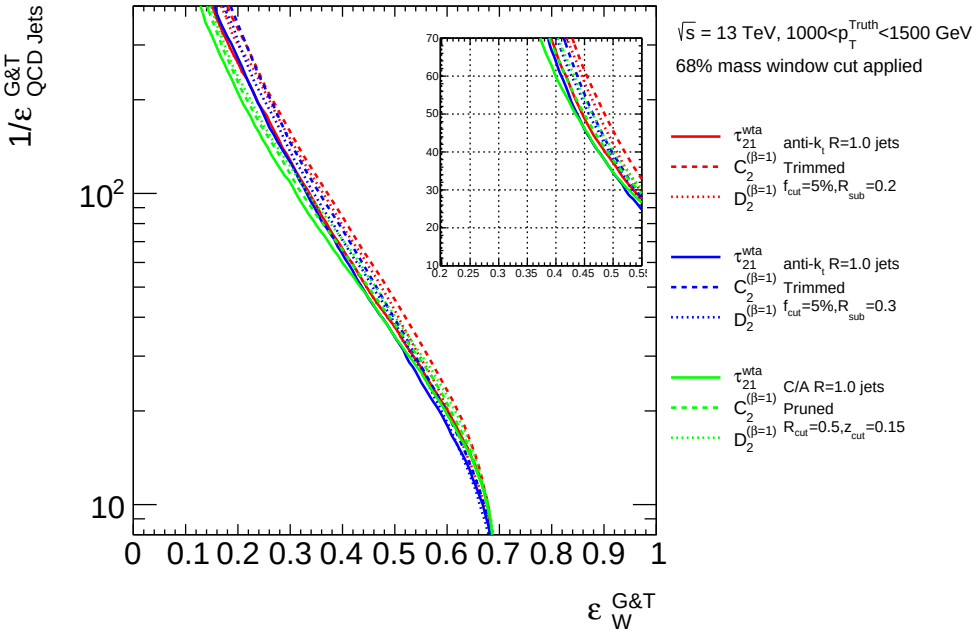


(b)

Figure 7.8 *(a) Background efficiency versus signal efficiency and (b) background rejection power versus signal efficiency for multiple grooming configurations and the best performing substructure variables where* $500 < p_T^{Truth} < 1000$ GeV. *The maximum signal efficiency is restricted to 68% due to the mass window cut on the groomed jet mass. The inset enlarges the area around 50% signal efficiency, which is the working point for these studies.*



(a)



(b)

Figure 7.9 *(a) Background efficiency versus signal efficiency and (b) background rejection power versus signal efficiency for multiple grooming configurations and the best performing substructure variables where* $1000 < p_T^{Truth} < 1500$ GeV. *The maximum signal efficiency is restricted to 68% due to the mass window cut on the groomed jet mass. The inset enlarges the area around 50% signal efficiency, which is the working point for these studies.*

7.7 BDT and DNN W Boson Taggers

The goal of the machine learning algorithms presented here is to create a model that can separate a boosted W boson in a W' decay from a multi-jet background by learning a general representation of the W' and multi-jet MC samples, which can then be applied to data. The idea is to combine a number of substructure variables into a single discriminant by using boosted decision trees and deep neural networks to create W boson taggers. In the following, these are referred to as BDT W taggers and DNN W taggers, respectively, and collectively as Machine Learning (ML) taggers. The methods employed to develop these taggers are discussed below.

Initially, the same selection criteria are applied as in Section 7.6.2. Using the results from the crosscheck in Section 7.6, the jet grooming algorithm used here is that with anti-k_t, $R = 1.0$ trimmed jets with f_{cut}= 0.05 and $R_{\text{sub}} = 0.2$. After applying the 68% mass window selection criterion on the leading groomed jet (as shown in Figure 7.10), several substructure variables are selected for use in the ML taggers. The choice of variables used in the ML taggers is based on a number of criteria: *a)* the performance of these variables when used pairwise with the groomed jet mass cut, *b)* the correlations between the variables and the sample type (signal or background), and *c)* their relative importance as reported by a random forest classifier. The selection of the variables is performed in Section 7.7.1.2.

Once a subset of variables has been selected, the hyperparameters of the classifiers are tuned (as discussed in Section 6.5). The tuning step effectively requires several models to be trained, which are compared with each other to find an optimal model. The ML taggers are then evaluated against the cut-based taggers in Section 7.7.3.

The ML taggers are trained within a range of $400 < p_{\text{T}} < 1200$ GeV and contain background events from the JZ4 and JZ5 MC datasets (defined in Section 7.4). The motivation for choosing this p_{T} range was to ensure there were enough events available for training the BDT and DNN taggers, and additionally, that the event weights associated with the higher p_{T} MC background samples are close to 1 for the training of the DNN W taggers. The selection of training and testing samples is discussed in Section 7.7.1.

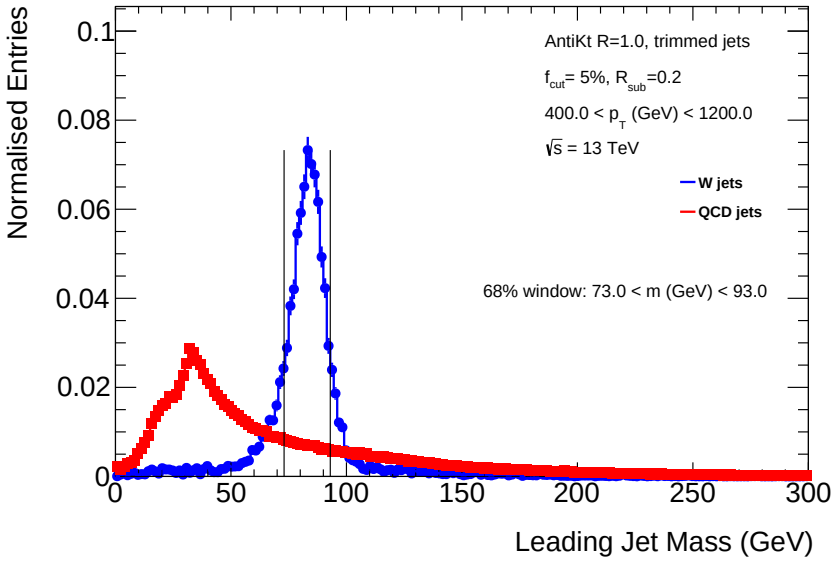


Figure 7.10 *Distributions of the leading groomed jet's mass with no selection on the mass applied. The solid vertical lines represent the 68% mass window for the signal sample. Shown here is the mass of the leading ungroomed C/A jet with* $400 < p_T^{Truth} < 1200$ GeV.

7.7.1 Training and Testing Samples

In this analysis, stratified k-fold cross validation (see Section 6.2.2) is used. The MC samples are cleaned and split into 75% for training and 25% for testing; this is done k times ($k = 10$ for the BDTs and $k = 5$ for the DNNs). Further, the performance is heavily affected by the hyperparameters chosen for the model, the input features that are used and how the samples are weighted. For the DNNs, the variables are scaled to a mean of zero and a standard deviation of one. Each of the data preparation techniques is addressed in the following subsections.

7.7.1.1 Stratified Cross Validation

After preselection, shown in Section 7.6.2, and the mass window requirement, there are a significant number of events that are kept; however, these events need to be processed further and there is a subsequent reduction. In some cases, a variable is not defined and is given a default value that has no physical meaning; these events are removed from the training sample during cleaning. The number of events passing the selections is shown in Table 7.6.

The 10-fold (for the BDTs) and 5-fold (for the DNNs) stratified cross validation

Selection	Background	Signal	Total
Raw number of events	3293505	354483	3647988
Preselection	2015119	93291	2108410
Mass window and p_{T} requirement	185902	57252	243154
Cleaning	158054	53971	212025

Table 7.6 *The number of events available for the BDT and DNN training. The cutflow shows how many events are removed after preselection has been applied, the 68% mass window requirement and a* $400 < p_T <$ 1200 GeV *cut on the anti-k_t, $R = 1.0$ trimmed jets with $f_{\mathrm{cut}} = 0.05$ and $R_{\mathrm{sub}} = 0.2$, and finally the number of events removed in cleaning.*

folds are created after the preselection has been done, the mass window cut has been applied, and the datasets are cleaned. For each fold this translates to approximately 158000 training events (40000 signal and 118000 background) and 53000 testing events (14000 signal and 39000 background). In order to check that the folds were representative, the mean value and standard deviation of each fold was compared with the full dataset. These values are shown in Appendix H.

7.7.1.2 ML W Tagger Input Variables

14 variables are initially considered as inputs for the ML taggers. In machine learning classifiers, adding more input features to the model does not necessarily increase the performance; a subset of the features is often sufficient. In order to find a set of variables to use as inputs, a number of sources were considered: a visual inspection of the distributions of the variables (as shown in Figure 7.13), the performance of the variables in the cut-based method (as shown in Figure 7.7), the correlations between these variables (as shown in Figure 7.11), the feature importances from a random forest classifier (as shown in Figure 7.12), and the performance of the ML taggers in testing.

Considering the results in Figure 7.12, it is observed that the top variables from the cut-based method, $C_2^{(\beta=1)}$, $D_2^{(\beta=1)}$ and τ_{21}^{wta}, are the top performing variables here. The variable $\sqrt{d_{12}}$ is the next best variable, and has a similar importance to τ_2^{wta}. The thrust variables and sphericity do not perform well, and are thus not considered further. The remaining variables do not have vastly different importance and it is not clear from this plot alone which of them should be chosen.

The linear correlations between the top performing substructure variables (after the mass window cut and jet grooming) and the class label (i.e. signal or background) are computed and shown in Figure 7.11. The three variables $\sqrt{d_{12}}$, $\sqrt{z_{12}}$ and YFilt are all correlated, as expected, since they are all variants of $\sqrt{d_{12}}$. τ_{21}^{wta} and τ_{2}^{wta} are also highly correlated, however τ_{2}^{wta} has a lower feature importance and potentially less information. Aplanarity has a correlation with the thrust variables and sphericity; however, it is kept as an input variable as it has a much higher feature importance. Additional considerations are that $\sqrt{z_{12}}$ and YFilt both have a jet mass dependence, and μ_{12} is dependent on the grooming algorithm.

Considering the above results, the following variables are dropped: $\sqrt{z_{12}}$, YFilt, μ_{12}, τ_{2}^{wta}, thrust variables and sphericity. Therefore, the final set of variables chosen for the inputs to the BDTs and DNNs is: $C_2^{(\beta=1)}$, $D_2^{(\beta=1)}$, $\sqrt{d_{12}}$, τ_{21}^{wta}, P and A.

The variables used for training and testing the DNNs, standardised to a mean of zero and a standard deviation of one, are shown in Figure 7.14.

7.7.2 BDT and DNN W Tagger Tuning

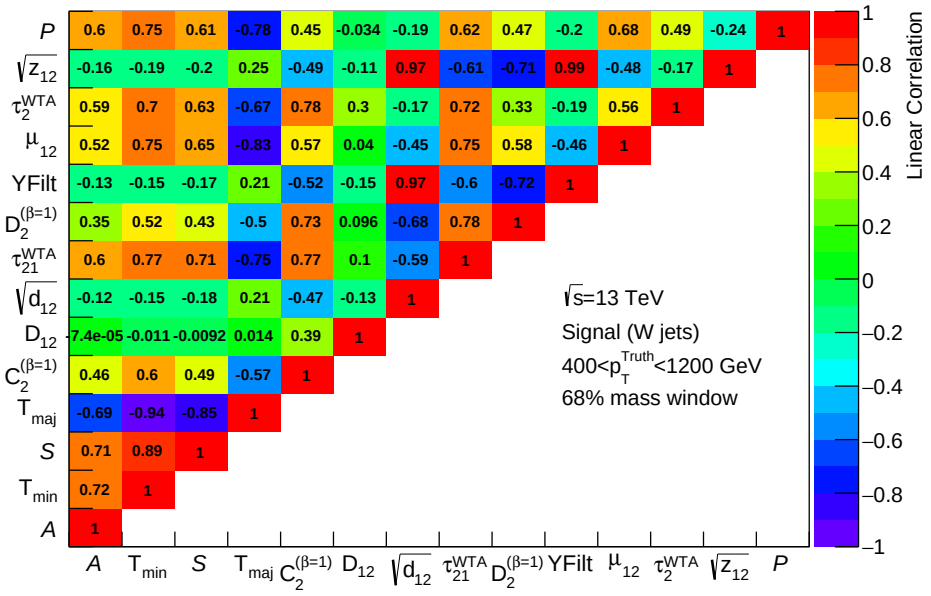
To tune the hyperparameters (see Section 6.5), a grid search is performed over multiple parameter combinations and all cross validation folds to find the optimal choice. In the results tables that follow, the accuracy or *score* of the model is quoted, which is defined as the weighted average of all the correct predictions (see Equation 6.17c).

BDT Parameters

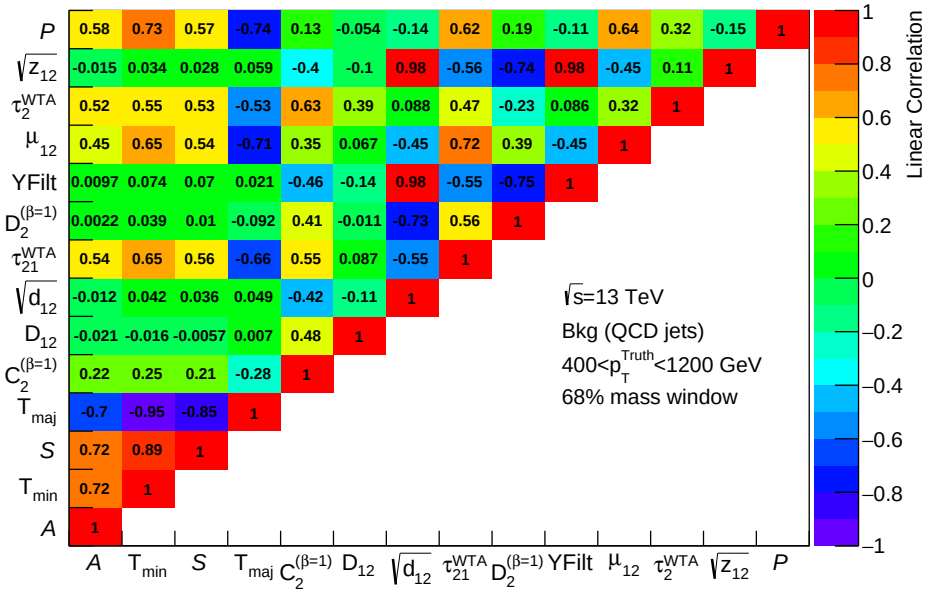
Three hyperparameters are tuned for the BDTs, where training the BDT takes on the order of a minute for each configuration of hyperparameters[3]. The values used in the grid scan are given below:

- base classifiers: Decision Tree Classifier with depth $\in \{3, 4, 5, 6, 8, 10, 15\}$,
- number of estimators $\in \{20, 35, 50, 65, 80\}$,
- learning rate $\in \{0.1, 0.2, 0.3\}$.

[3]Training was performed using an Intel Core i7 2600S CPU with 16 GB of RAM.



(a)



(b)

Figure 7.11 *The matrix plots show the linear correlation coefficients between the possible feature inputs for the BDT and DNNs. There is a 68% mass window cut on the groomed jet mass, where the groomed jets are anti-k_t, $R = 1.0$ trimmed jets with $f_{\text{cut}} = 0.05$ and $R_{\text{sub}} = 0.2$. Here the correlations are shown for separate signal and background samples where the leading ungroomed C/A jet has $400 < p_T^{Truth} < 1200$ GeV. The general features in the correlations are the same between signal and background.*

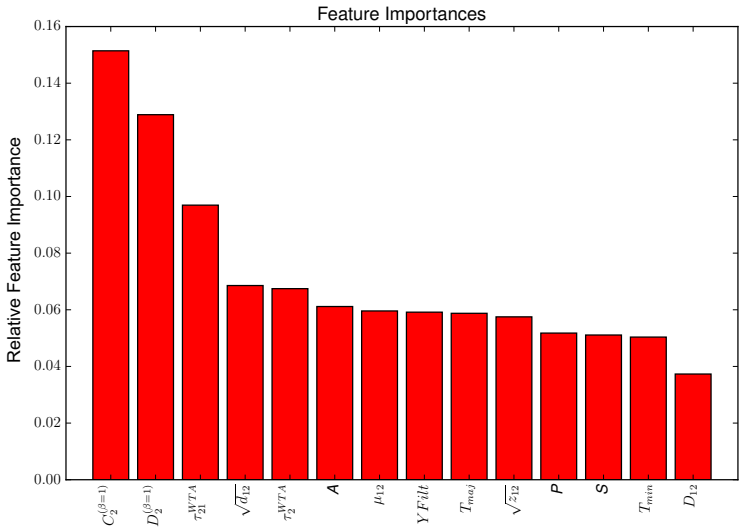


Figure 7.12 *The feature importances taken from a Random Forest Classifier with 500 trees.*

The effect of the maximum tree depth on the overfitting of the BDTs is evaluated by considering the validation curves that are shown in Figure 7.15. These results indicate that trees with a maximum depth of greater than five are overfitting, and as such, all trees with a depth greater than five are excluded. The top five hyperparameter sets from the grid scans, excluding those with maximum depth greater than five, are listed in Table 7.7.

ID	Max. depth	Learning rate	No. estimators	Ave. Accuracy	Ave. Bkg. rej. power
31	5	0.1	35	0.81	61.5
35	5	0.2	20	0.81	60.7
32	5	0.1	50	0.81	58.8
30	5	0.1	20	0.81	58.2
40	5	0.3	20	0.81	56.9

Table 7.7 *Training parameters and background rejection power for the top 5 BDTs with* $400 < p_T^{Truth} < 1200$ GeV. *The final two columns give the average accuracy and background rejection power at 50% signal efficiency obtained when running over all of the training folds.*

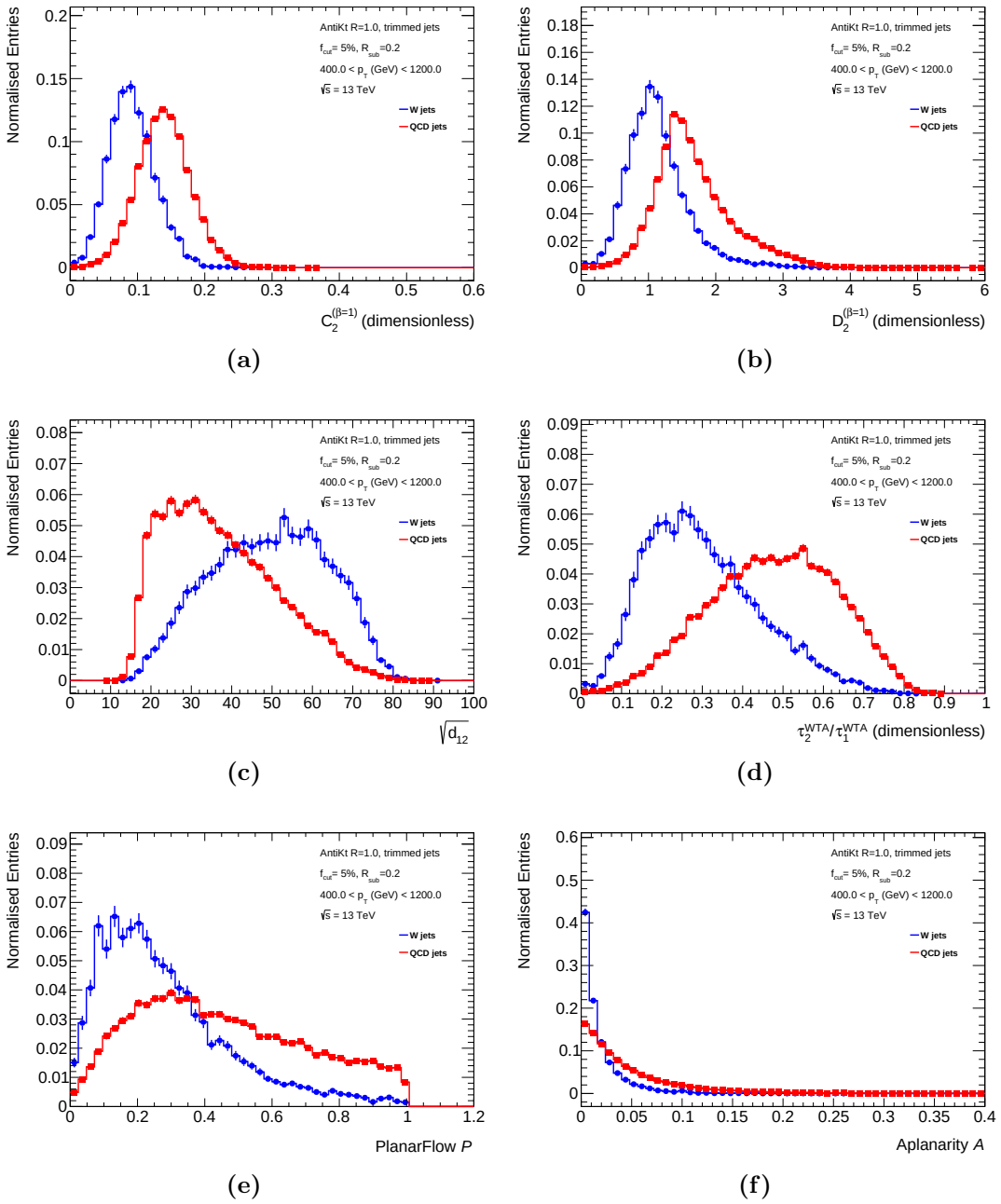


Figure 7.13 *The distributions of the input variables used in the BDTs. These have the 68% mass window cut applied to the leading groomed jet, where the leading C/A $R = 1.2$ jet has $400 < p_T^{Truth} < 1200$ GeV.*

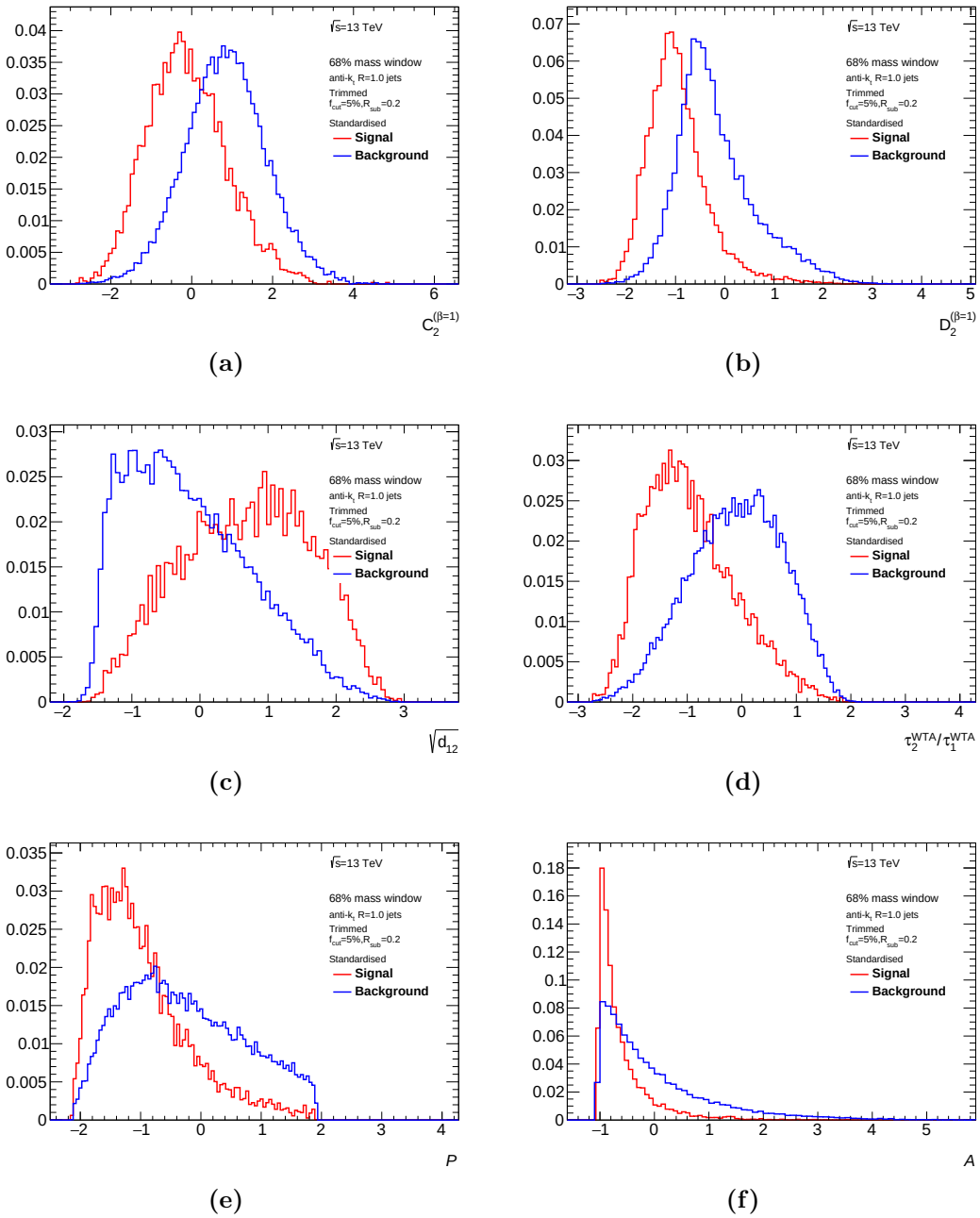


Figure 7.14 *The distributions of the input variables used in the DNNs. These have the 68% mass window cut applied to the leading groomed jet, where the leading C/A $R = 1.2$ jet has $400 < p_T^{Truth} < 1200$ GeV. These distributions have been standardised such that the combined signal and background sample has a standard deviation of 1 and a mean of 0.*

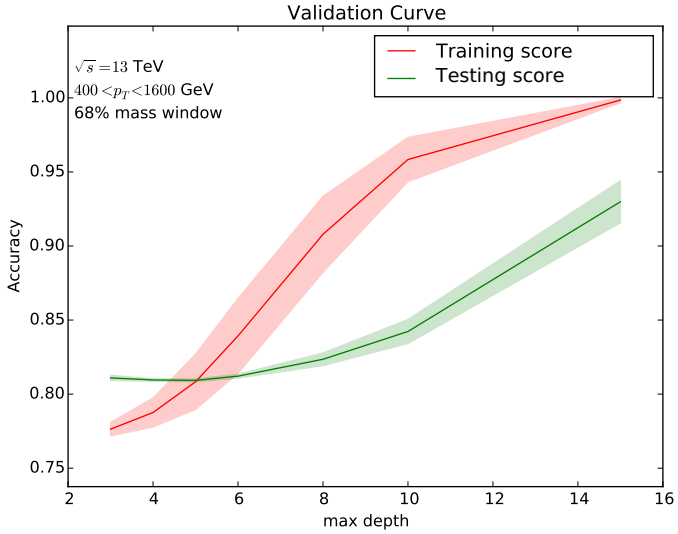


Figure 7.15 *Validation curves (see Section 6.5.3) for the BDTs, showing accuracy as a function of the maximum depth of the tree. The shaded bands show the the standard deviation of the mean accuracy of all 10 cross validation folds, and the vertical line indicates the optimal maximum depth of five. The diverging curves indicate that there is a high variance problem. BDTs with a large maximum depth are learning the training sample well (the training accuracy, or training score), but are not generalising well. From this plot, it can be seen that a high value for maximum depth gives inconsistent results. There is a 68% mass window cut on the groomed jet mass. Shown here are the plots for combined signal and background datasets where the leading ungroomed C/A jet has* $400 < p_T^{Truth} <$ 1200 GeV.

DNN Parameters

For the DNNs, fewer hyperparameters were tested, due to the relative time taken to train each model, which is on the order of 30 minutes to an hour[4]. However, some guidelines from the author of the software that was used, AGILEPack, were used to narrow down the search space [192]. The list of hyperparameters tested (as discussed in Sections 6.3.1 and 6.3.2.1) are:

- logarithm of the learning rate $\in \{-10, -8, -6, -4, -2\}$,
- momentum $\in \{0.7, 0.85\}$,
- logarithm of the regularisation (or shrinkage) $\in \{-10, -7\}$,
- unsupervised training epochs $\in \{20, 60\}$,
- supervised training epochs $\in \{40, 80\}$.

The top five sets of hyperparameters from the grid scans are shown in Table 7.8. The accuracy and background rejection power given is the average over all folds of the test datasets.

ID	Momentum	Log regularisation	Log learning rate	Unsup. epochs	Sup. epochs	Ave. accuracy	Ave. bkg. rej. power
35	0.7	-10	-6	60	80	0.69	52.3
32	0.7	-10	-6	20	40	0.69	50.7
17	0.7	-10	-8	60	40	0.690	50.6
33	0.7	-10	-6	60	40	0.69	50.5
12	0.85	-7	-10	20	40	0.70	50.2

Table 7.8 *Training parameters and background rejection power for the top 5 DNNs with* $400 < p_T^{Truth} < 1200$ GeV. *'Unsup.' and 'Sup.' refer to the number of unsupervised and supervised training epochs. The final two columns give the accuracy and background rejection power at 50% signal efficiency obtained when running over all of the training folds.*

[4]The same hardware was used for training the DNN classifiers as for the BDT: Intel Core i7 2600S CPU with 16 GB of RAM.

7.7.3 Results

The training of the BDTs produces a decision function that provides discrimination between signal and background events. Each event is assigned a value by the decision function of between -1 and 1 with a corresponding signal probability, which is then used to determine if it is signal or background; the value of the decision function at which an event is deemed signal is chosen such that it meets a given threshold on the signal-to-background ratio. The decision functions for the top three variations of the BDTs from Table 7.7 are shown in Figure 7.16 (a,c,e). A good separation between signal and background is observed on the uncleaned samples (i.e. only event selection has been applied, no events with outliers are removed). The predicted signal probability (see Section 6.4.1) is used as a class *discriminant* for the BDTs, as shown in Figure 7.16 (b,d,f). Again, a good separation is seen between signal and background.

The classification probabilities from the top three DNNs listed in Table 7.8 are shown in Figure 7.17. There is excellent separation between signal and background, and in general the model has prediction probabilities that are different from the BDTs. However, there is a larger overlap, with both distributions having a long tail, especially the signal. This means that there is more chance of rejecting a signal event.

The ROC curves in Figures 7.18 and 7.19 show the signal efficiency versus 1−background efficiency and signal efficiency versus background rejection power for the best BDT and DNN taggers, respectively. In each of these, the performance on the uncleaned MC dataset is shown. The top performing cut-based taggers are shown for comparison; in all cases the BDTs and DNNs outperform them. At 50% signal efficiency, the highest background rejection power is just over 50 for the DNN taggers, just over 60 for the BDT taggers, and around 45 for the cut-based taggers.

7.7.4 Discussion

All top five BDT taggers have similar background rejection power, although there is a large variance in the performance when considering all configurations tested. BDTs with a maximum depth of greater than five have higher rejection power (of up to 1200 in some cases), but they heavily overfit the training sample. This could potentially be addressed by training with more events, since the number

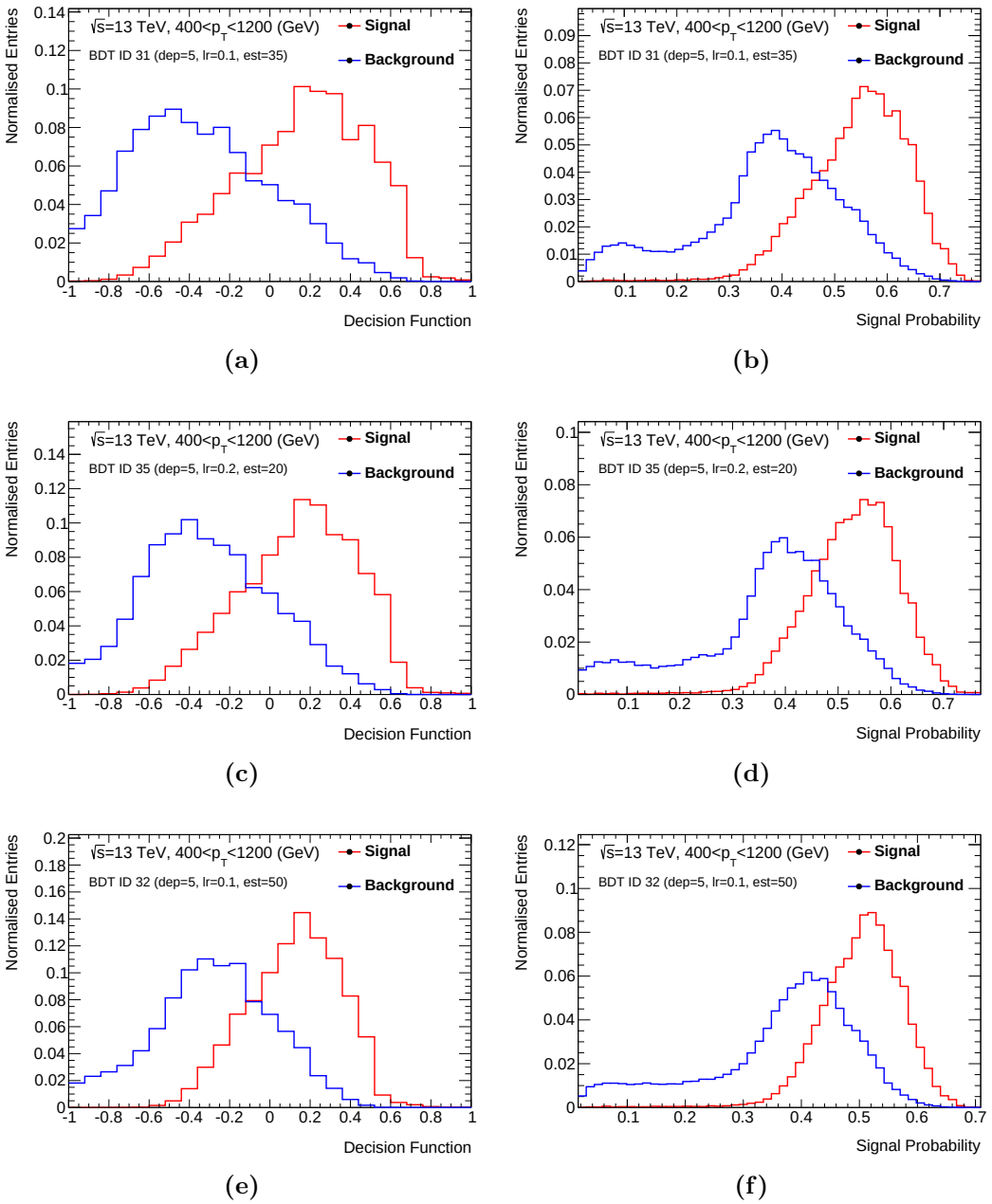


Figure 7.16 *Plots on the left show the decision functions for the top three performing BDTs where* $400 < p_T^{Truth} < 1200$ GeV. *This shows the scores assigned during to training to signal and background events. Plots on the right show the signal probability corresponding to the configurations in the left-hand side plots when testing on the uncleaned dataset, after preselection.*

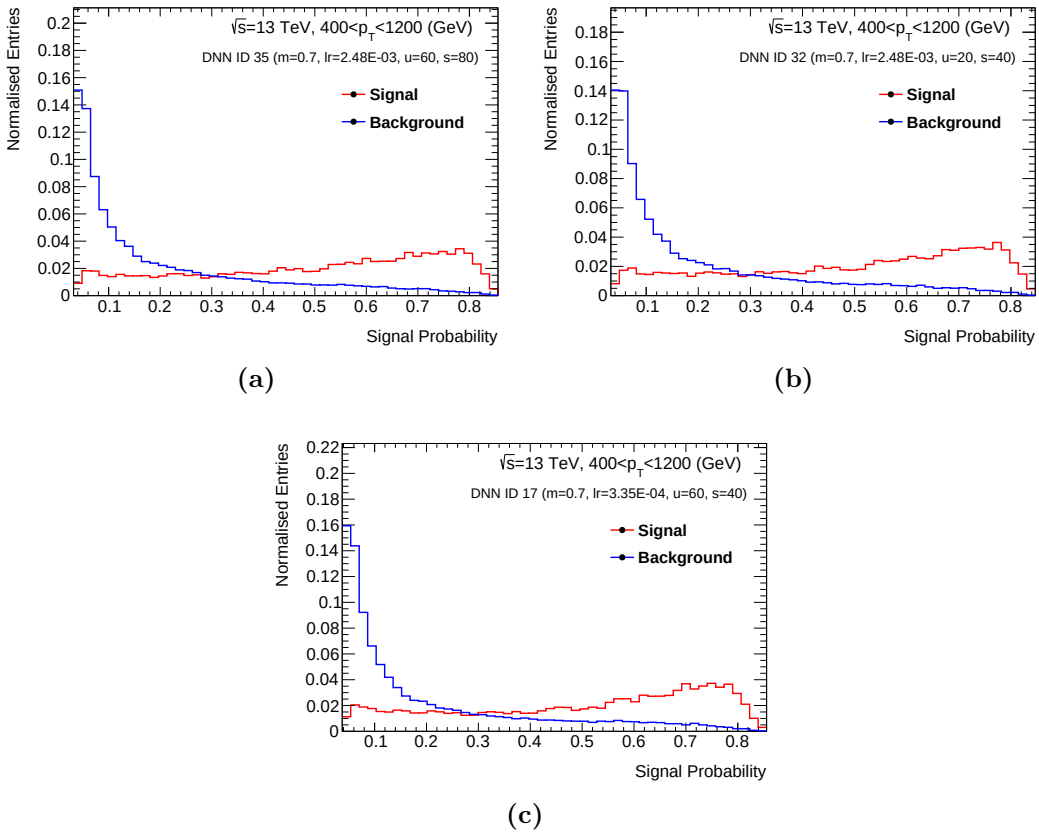


Figure 7.17 *Plots (a)-(c) show the signal probability for each event as classified by the top three performing DNNs where* $400 < p_T^{Truth} < 1200$ GeV.

of events needed to populate each base classifier in the BDT doubles for every additional layer.

In the signal distributions in Figure 7.16 (the red line) it is observed that there is a small peak in the background at low values. These low values are found to come from lower p_{T} events that have large Monte Carlo weights, as shown in Figure I.1 in Appendix I. These events are found in the JZ4 background sample. The lowest regions where the BDT signal discriminant is < 0.16 contain only $\approx 0.02\%$ of the total background events, however, these events have much larger weights than the other regions, as shown in Figure I.2 in Appendix I. After weighting the distributions, this low region contribution accounts for almost 12% of the total background. BDT signal discriminant is plotted against the jet p_{T} in Figure I.3 in Appendix I, which suggests that high p_{T} jets are more likely to be tagged as boosted W bosons. A possible explanation for this is the Monte Carlo weights being used in the training of the BDTs. Since the low p_{T} background events have larger weights, the BDT will focus more on discriminating against these events.

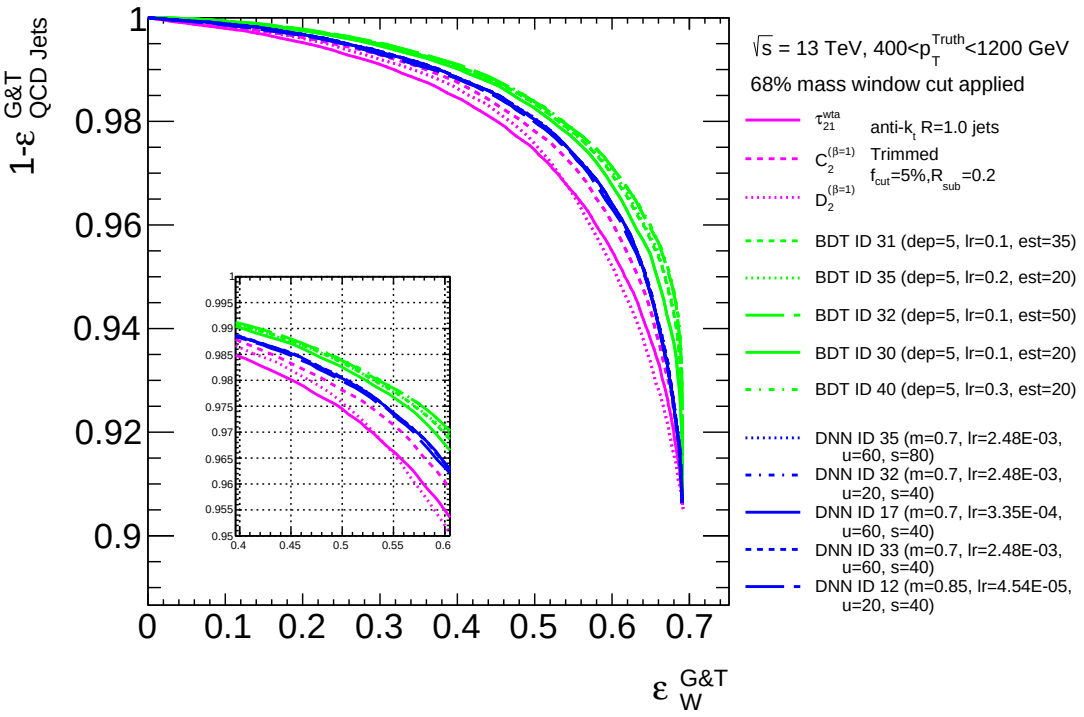


Figure 7.18 *The ROC curves show the signal efficiency on the x-axis and the background efficiency on the y-axis for the top five performing BDTs, the top five performing DNNs and the top three substructure variables. A 68% mass window cut is applied on the groomed jet mass. The corresponding background rejection plot is shown in Figure 7.19.*

For the DNN taggers, although all top five configurations provide similar background rejection power, there is less variance over the full set of hyperparameters that were tested, compared with the variance seen in the BDT taggers. This can be explained by the smaller set of combinations in the grid search. For a number of combinations of hyperparameters, the DNN taggers do not converge, particularly when the learning rate is set to larger values.

Aside from showing a large gain in background rejection power, the ML taggers showed an increase in performance over a relatively large p_{T} range of 800 GeV. The ML taggers are not, however, p_{T} independent, and training them in smaller p_{T} bins is expected to improve the performance overall; although this would require a lot more simulated events. Additionally, the boundaries between the p_{T} bins need to be considered such that the classification efficiency is continuous over the boundaries. The results here consider events where the leading jet has p_{T} of up to 1200 GeV, but this should be extended for searches of resonances beyond 3 TeV. At higher p_{T} some variables will quite likely change behaviour

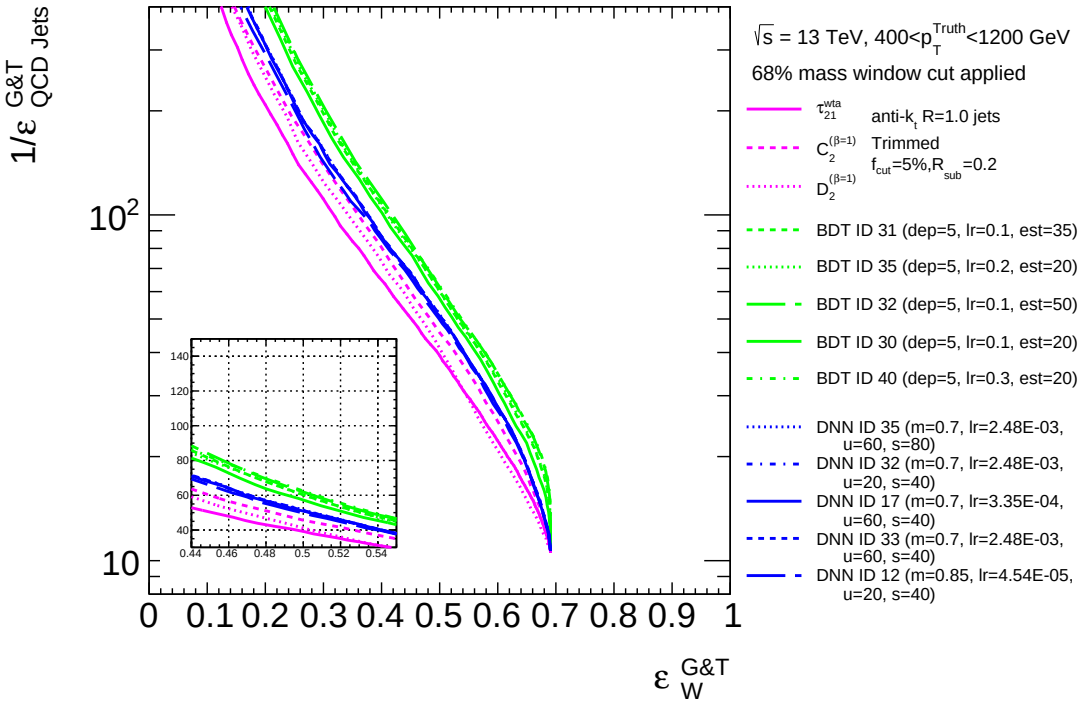


Figure 7.19 *The ROC curves in (a) show the signal efficiency on the x-axis and the background rejection power on the y-axis for the top five performing BDTs, the top five performing DNNs and the top three substructure variables. A 68% mass window cut is applied on the groomed jet mass. At 50% signal efficiency the BDT and DNN obtain a background rejection power of ≈ 60 and ≈ 50, respectively, compared with ≈ 45 for the single substructure variables.*

and the setup used here would need to be re-evaluated.

In the next section, the BDT and DNN taggers offering the highest background rejection power are applied to the HVT $Z' \to WW \to q\bar{q}q\bar{q}$ search. This is used as an opportunity to validate the results obtained here using data.

7.8 HVT Z' Prospects

Many extensions to the SM predict diboson resonances that would be accessible at the energies achieved during Run 2. One such model is the Heavy Vector Triplet model (HVT) [19], which introduces additional heavy $W'^{\pm}$ and Z' bosons. This model is used as the signal process when developing the ML W boson taggers in this chapter. The W' and Z' can decay into diboson resonances. Hints of such a heavy diboson resonance were found during Run 1 when an excess of events

with a global significance of 2.5σ was seen at 2 TeV from hadronic decays in WZ events [236], however, this has not been confirmed in Run 2 [237].

Searches for diboson resonances can benefit from improved identification of boosted vector bosons. One such search is the diboson resonance search at ATLAS at a centre-of-mass energy of $\sqrt{s} = 13$ TeV presented in Reference [237], using data collected in 2015 by ATLAS from pp collisions at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 3.2 fb^{-1}. This analysis focuses on the decay of TeV scale resonances with a mass of between 1.2 and 3 TeV, which decay into highly boosted W and Z bosons. A Z' of masses between $1.38-1.6$ TeV is excluded at 95% CL from the analysis of the $Z' \to WW$ decays. This analysis has since been updated to include a further 12.3 fb^{-1} (for a total of 15.5 fb^{-1}) of data collected by ATLAS in 2016 [238]. With this additional data, a Z' of masses between $1.2 - 1.8$ TeV are excluded at 95% CL. These analyses both use the same selection criteria and assess systematic uncertainties in the same way. The application of the ML taggers to the process $Z' \to WW \to q\bar{q}q\bar{q}$ is investigated in the following sections and validated using the 2015 dataset, corresponding to an integrated luminosity of 3.2 fb^{-1}.

The signal process $Z' \to WW \to q\bar{q}q\bar{q}$ is simulated at $\sqrt{s} = 13$ TeV in the same way as the $W' \to WZ \to q\bar{q}q\bar{q}$ process in Section 7.4. The same set of mass points is also created. The dominant background contribution is from multi-jet events; the samples used here are listed in Table 7.2.

A subset of the event selection criteria used in Reference [237] is applied to simulated events and data. The identification of the vector bosons in Reference [237] uses large-R groomed jets and the substructure variable $D_2^{(\beta=1)}$ (described in Section 7.3.1.3) in order to identify the boosted W bosons. In the application of the ML taggers, $D_2^{(\beta=1)}$ is replaced by a class prediction probability from the BDT and DNN W boson taggers developed above. In the following sections, the input variables used in training and the output of the ML taggers are shown for data and Monte Carlo.

A summary is given below of the object reconstruction and event selection from [237], and the modifications that are made to the selection criteria for the evaluation of the ML W boson tagger. This is followed by comparisons of data and Monte Carlo simulations. Finally, prospects and future steps are discussed.

7.8.1 Object and Event Selection

The primary goal is to identify two high p_{T} large-R jets that are consistent with a W boson. Large-R jets with $R = 1.0$ are reconstructed from clusters using the anti-k_t algorithm. These jets are trimmed using the same configuration used in Section 7.6: anti-k_t, $R = 1.0$, $f_{\mathrm{cut}} \geq 5\%$ and $R_{\mathrm{sub}} = 0.2$. Two groomed large-R jets are required, one of which has $p_{\mathrm{T}} > 450$ GeV and the other at least $p_{\mathrm{T}} > 200$ GeV, and have a separation in rapidity of $|\Delta y_{12}| < 1.2$. The two jets must both be within $|\eta| < 2.0$, and have a jet mass of greater than 50 GeV. The mass of the dijet pair, m_{JJ}, is required to be greater than 1 TeV. An approximate transverse momentum balance between the two jets, as expected from a heavy object decay, is enforced by the cut $\frac{p_{\mathrm{T}_1} - p_{\mathrm{T}_2}}{p_{\mathrm{T}_1} + p_{\mathrm{T}_2}} < 0.15$, where the subscripts 1 and 2 refer to the leading and subleading jets, respectively.

Any events with electrons or muons (using the medium working point and loose isolation criteria from Sections 4.5 and 4.6) are rejected. Additionally, any events with a large missing transverse momentum of $E_{\mathrm{T}}^{\mathrm{miss}} < 250$ GeV are rejected.

A 30 GeV symmetrical mass window is defined using the trimmed jet mass, centred on the mass peaks of the W boson observed in the simulated MC, after local cluster weighting calibration of the jet mass (see Section 4.7.2), of 84 GeV. Both the leading and subleading groomed jets must have a mass within this window. Note that the definition of the mass window used here is different from that used in the previous sections.

After the mass window requirement, the identification of the jet as a W boson is performed: a cut-based method as described in Reference [237], similar to that presented in Chapter 5, or using an ML W tagger. In the cut-based method, a cut on $D_2^{(\beta=1)}$, in conjunction with the W mass requirement, is applied such that there is a 50% signal efficiency.

For both jets, there is a requirement of $n_{trk} < 30$ (the number of tracks associated with the ungroomed jet which corresponds to the large-R groomed jet), which has been shown to provide discrimination between W boson and multi-jet events, improving the expected signal sensitivity in Reference [237] by 30%. This is applied for each jet after the substructure cut and the mass window requirement.

In order to compare the search results when using the cut-based W tagger and the ML taggers, the cut on $D_2^{(\beta=1)}$ is removed for both jets and replaced with a cut on the output from either the BDT or the DNN. As shown earlier in Figures 7.16

and 7.17, the output from the BDT/DNN is a continuous variable. A cut must be chosen on these output variables such that a given selection efficiency is obtained. For the ML W boson tagger, the BDT and DNN configurations chosen are the best performing configurations trained in the range $400 < p_{\mathrm{T}} < 1200$ GeV, as shown in Section 7.7.2 (the first entries in Tables 7.7 and 7.8).

7.8.2 Data-to-Monte Carlo Comparisons

The output of the ML taggers is evaluated after the mass window requirement and all other event selection criteria listed in Section 7.8.1. Before the boosted W boson selection has been performed, the region is heavily dominated by QCD backgrounds. This offers an opportunity to compare the output of the taggers on data and background simulations. In the following, a Z' of mass 1.2 TeV is considered as the signal. A cut on the invariant dijet mass of $1.1 < m_{JJ} < 1.3$ TeV is implemented. This mass cut corresponds to a majority of the corresponding large-R jets having a range of $500 < p_{\mathrm{T}} < 700$ GeV.

The variables used as inputs for the two ML W boson taggers for the leading and subleading jets are shown in Figures 7.20 and 7.21, respectively, where the signal process is a Z' of mass 1.2 TeV. These distributions are shown after the mass window cuts have been applied, but without a cut on the ML tagger output or the n_{trk} variable. The distributions show excellent agreement between the background expectation and data, and there are no substantial differences seen between the two jets. A comparison of these variables with those used for training the BDT and DNN taggers (shown earlier in Figures 7.13 and 7.14) reveals some differences. The differences seen between the signal distributions in Figures 7.20 and 7.21, compared with those used for training, is because of the different Z' masses used. There is a slight shift in the peaks in the distributions of τ_{21}^{wta}, compared with the training samples. The peak of $D_2^{(\beta=1)}$ in the signal sample used for training is centred around 1, as opposed to about 1.3 here, and has a greater separation from the background.

The output of the BDT and DNN taggers shows excellent agreement between the multi-jet background expectation and data in the QCD enriched region, as seen in Figure 7.22.

The 50% signal point is re-derived here for $D_2^{(\beta=1)}$, the BDT W tagger and the

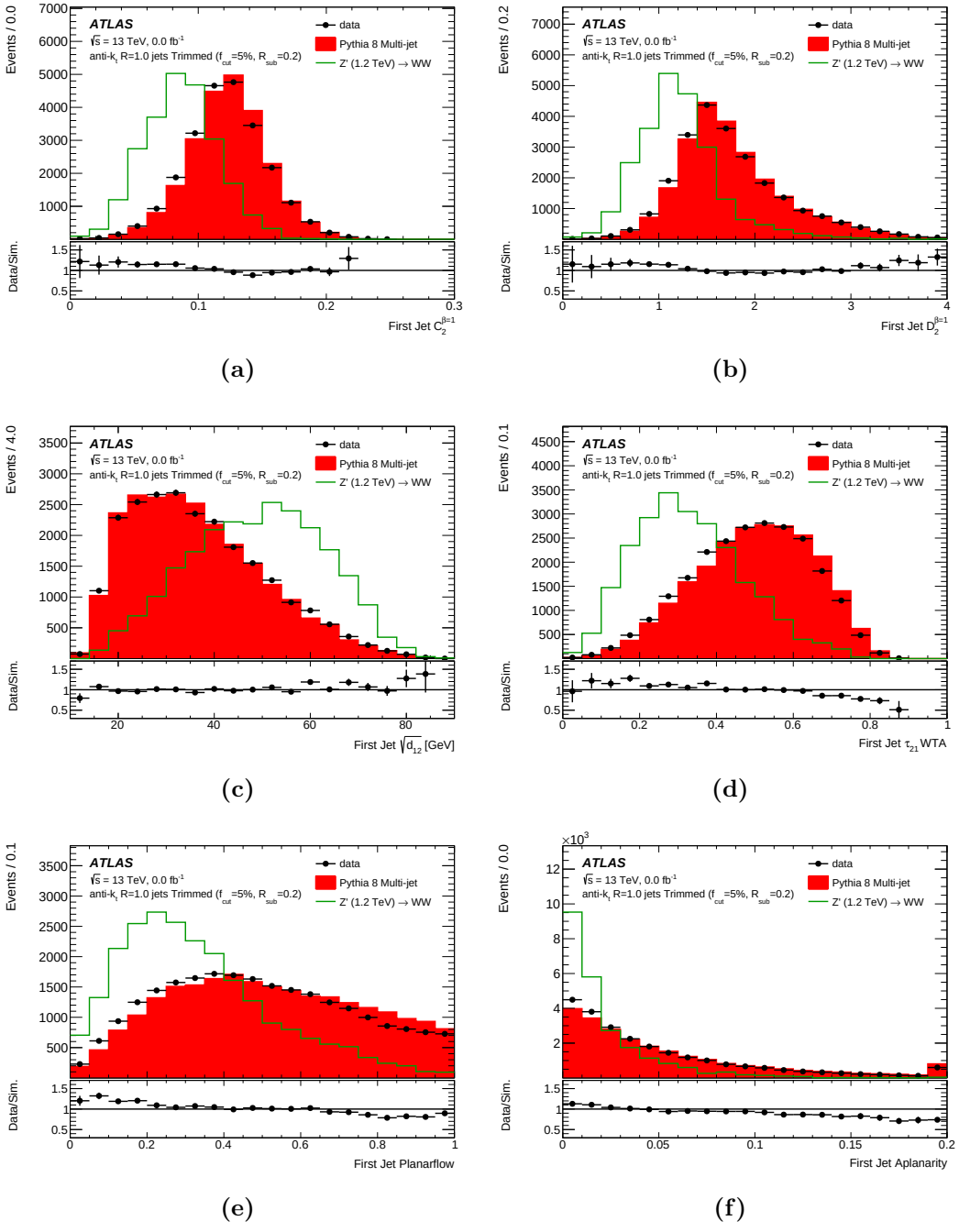


Figure 7.20 *The distributions of the input variables which are used as inputs for the ML W tagger for the leading jet after preselection in the HVT $Z' \to WW$ search. A Z' of mass 1.2* TeV *is considered here, and an invariant dijet mass requirement of* $1.1 < m_{JJ} < 1.3$ TeV *is imposed.*

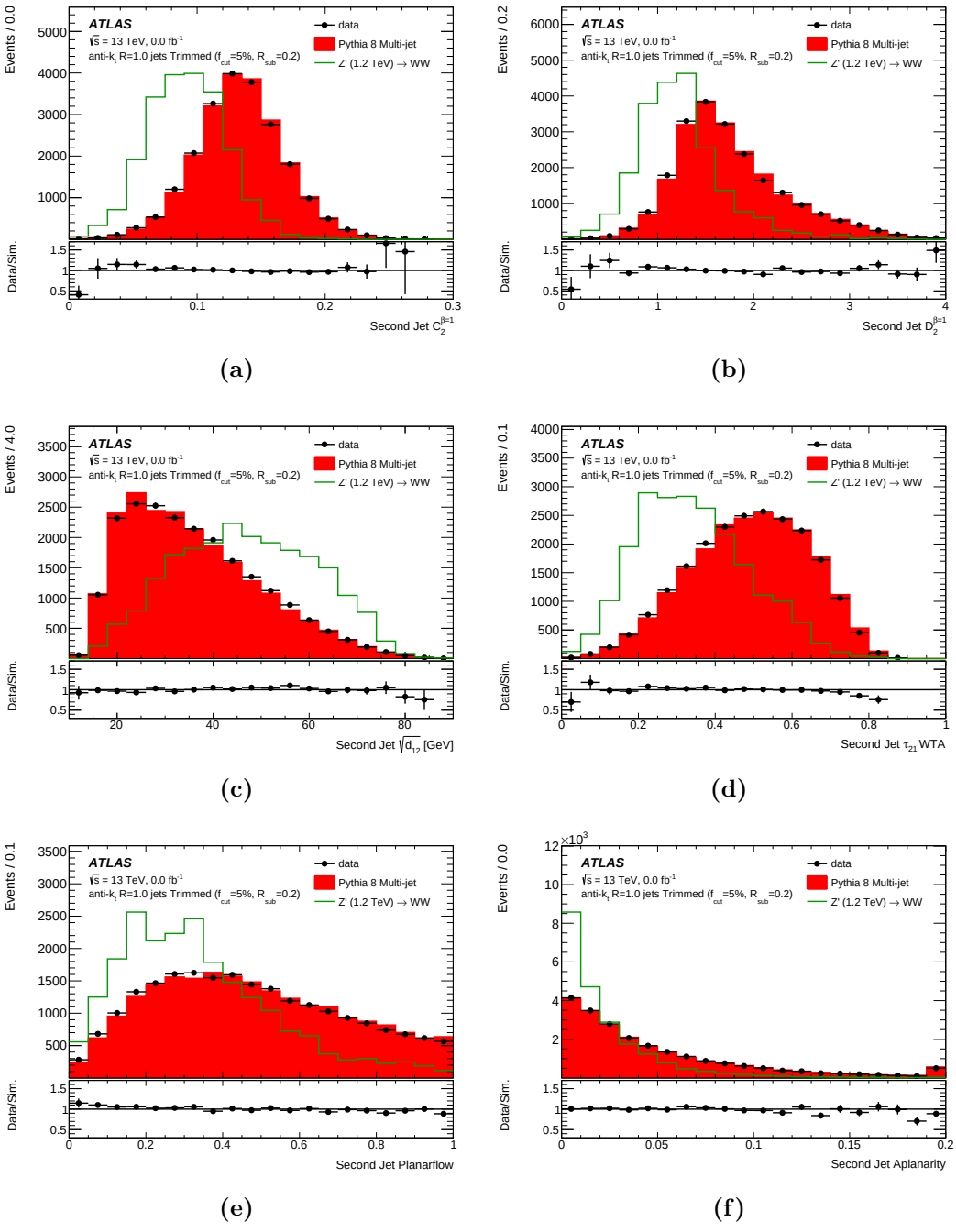


Figure 7.21 *The distributions of the input variables which are used as inputs for the ML W tagger for the second-to-leading jet after preselection in the HVT $Z' \to WW$ search. A Z' of mass 1.2* TeV *is considered here, and an invariant dijet mass requirement of* $1.1 < m_{JJ} < 1.3$ TeV *is imposed.*

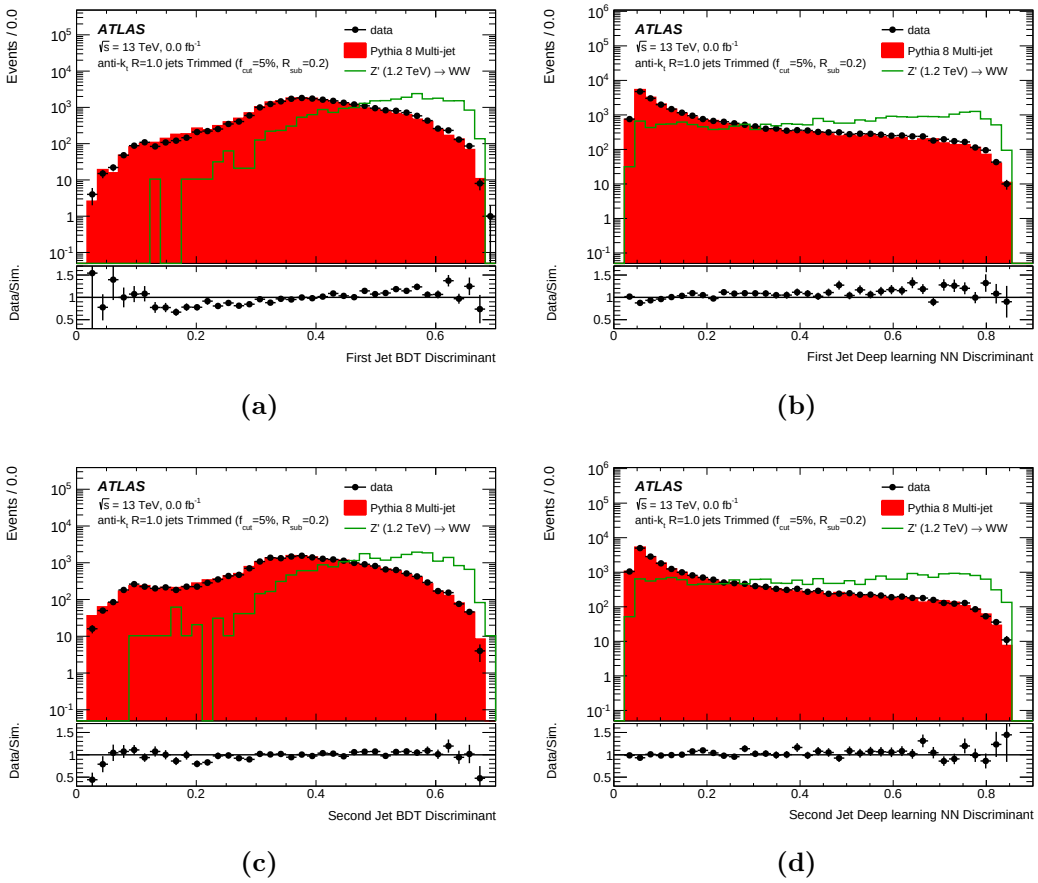


Figure 7.22 *Plots (a) and (b) show the output of the BDT and DNN taggers for the leading jet in background, data and for the HVT model with a Z' of mass 1.2* TeV*. An invariant dijet mass requirement of* $1.1 < m_{JJ} < 1.3$ TeV *is imposed. Below this in plots (c) and (d) the output is shown for the subleading jet. Good agreement is seen between data and the background estimation.*

DNN W tagger on the leading jet[5]. Similar conclusions to Section 7.7.3 are found. There is a lower gain in background rejection power for the BDT tagger, however. There is approximately the same gain of about 10% for the DNN tagger. The gain in background rejection power after selecting the leading jet is shown in Table 7.9.

7.8.3 Discussion of Results

The selection criteria reduce the number of background events drastically when tagging both jets (a factor in the range 1000-2000), and as such, it is difficult

[5]In the analysis in Reference [237], the working point for the $D_2^{(\beta=1)}$ cut is p_{T} dependent.

	No-tag	1-tag		
	Pre-tag selection	$D_2^{(\beta=1)} < 1.22$	BDT> 0.53	DNN> 0.49
Signal events	67 ± 1.3	33 ± 0.9	33 ± 0.9	33 ± 0.9
Background events	136540 ± 476	3791 ± 66	3300 ± 58	3450 ± 61
Signal efficiency (%)	-	50 ± 3	50 ± 3	50 ± 3
Bkg. efficiency (%)	-	2.8 ± 1.79	2.4 ± 1.80	2.5 ± 1.80
Bkg. rejection gain (%)	-	-	13.0 ± 1.40	9.0 ± 3.90

Table 7.9 *Event yields after selecting the leading jet in the event at 50% signal efficiency with a cut on the $D_2^{(\beta=1)}$ variable, and the BDT and DNN W boson taggers. The signal here is a Z' of mass 1.2* TeV*, and there is a dijet invariant mass cut of* $1.1 < m_{JJ} < 1.3$ TeV*. No cut on n_{trk} is applied here. As can be seen in the final row, there is a gain in background rejection for the BDT and DNN W boson taggers over the $D_2^{(\beta=1)}$ selection on the order of 10%. The statistical errors quoted are due to limited Monte Carlo events.*

to provide a statistically conclusive comparison of the impact of the ML taggers on the event yields after requiring two W bosons. With the available QCD simulated events, after the first jet is tagged, there are between 3300 and 3500 background events passing the selection, however, after the second jet is tagged, this is reduced to just over 100 events. At this point the statistical uncertainties on the background efficiency and rejection are larger than the absolute yield. In the event selection in the cut-based analysis there is a further cut on n_{trk} at this point, reducing the background yield further. With more simulated events it will be feasible to estimate the gain in sensitivity using the ML taggers in this analysis. The results from Table 7.9 after tagging a single jet, indicate that, given more simulated background events, there is likely to be a significant gain in the sensitivity over using a single substructure variable.

7.8.4 Discussion and Conclusion

The systematic uncertainties in the diboson search in Reference [237] are calculated differently for the signal and background processes. The smoothly falling dijet invariant mass in the multijet background is modelled with a parametric function with three parameters that are determined by a binned maximum likelihood fit. The uncertainties on the background expectation are determined by a maximum likelihood fit to the dijet mass spectrum. However, there are known differences in the fit parameters between data and simulation for

the shape. These differences will need to be addressed when implementing the analysis with the ML taggers.

For each signal process, the systematic uncertainties with the largest effect on the signal expectation are the jet mass and p_{T}, and $D_2^{(\beta=1)}$, and on the integrated luminosity. The scale uncertainty on $D_2^{(\beta=1)}$ is assessed using data-to-MC comparisons of track and calorimeter measurements of $D_2^{(\beta=1)}$. Similar methods are used for the scale uncertainties on the jet mass and p_{T}. The scale uncertainties for $D_2^{(\beta=1)}$, the jet mass and the jet p_{T} are 5%, 6% and 5%, respectively. The systematic uncertainty on their resolutions are 20% each.

The signal uncertainties are included as nuisance parameters in a binned maximum likelihood fit, using the parameterised background expection as the background, and a template from simulation for the signal shape.

To estimate the effect of systematic uncertainties on the background when using the ML taggers a similar method to the current estimate can be used. A control region must be constructed to assess how well the output from ML tagger describes the data, and the uncertanties estimated by a fit to data in this region.

For the signal systematic uncertainties there are additional considerations. Six substructure variables are used, each of which will introduce its own uncertainty to the output of the ML tagger. Since the final discriminant is the dijet invariant mass, and only a cut is applied to the ML tagger, the changes to the shape of the ML tagger output should be assessed. This can be done by varying the input distributions and observing the change to the distribution.

The increased signal yield that can be achieved using the ML taggers, coupled with more data being collected, will provide not only an increased signal sensitivity, but also allow for the signal systematics to be constrained. The aim of this study is to assess any potential gains from ML tagging, and as such, the study of associated systematic effects is beyond the scope of this thesis. A possible approach to evaluating the systematics for the ML taggers is to test the ML taggers with the *up* and *down* systematic variations of the individual input substructure variables. The systematic studies of all the substructure variables has not yet been performed, however, $D_2^{(\beta=1)}$ systematics have been calculated in Reference [237]. Systematics are calculated using data-to-simulation comparisons of the ratio of $D_2^{(\beta=1)}$ measured using tracks and energy clusters. Similar methods could be used here.

In Reference [237] (and introduced in Section 7.8.1) an additional variable, n_{trk}, is used in the boosted boson identification and offers good discrimination against QCD backgrounds. However, it has been shown to be poorly modelled [239, 240] and it introduces large systematic uncertainties. The variable is used in the cut-based analyses in References [237, 238] as it offers an increase in the sensitivity of 30%, which is large enough to offset the systematic uncertainties. Including this variable in the ML taggers could further improve their performance, although the effects of the systematic uncertainties will need to be assessed carefully. An initial assessment of the improvement offered by including this variable in the ML taggers showed a difference in background rejection power of around 20%.

The improvements shown in this proof of concept show that this search can stand to benefit by using the ML taggers. The gain in signal sensitivity will need to be quantified when more data and more simulated QCD samples become available, however, these initial results are promising. The ML taggers show potential for increased identification efficiency at higher p_{T}, as discussed in Section 7.7.4, which is especially relevant for searches for heavy mass particles that decay into W bosons. These gains in signal sensitivity and high p_{T} performance will have a direct impact on the setting of exclusion limits, and, in particular, setting upper limits on the cross section of a given signal process. Currently, the Z' from HVT model A (with $g_V = 1$), is excluded at 95% CL for masses between $1.2 - 1.8$ TeV, but the ML taggers could be used to extend this to higher mass points.

The techniques shown here are general, and can be used for identifying boosted Z bosons or Higgs bosons. The improved identification efficiency will benefit many physics searches where these boosted bosons are present.

Chapter 8

Conclusions

The second Timeless Test ended with the submission of this thesis. Image © ESPNCricInfo Ltd. [241]

The final missing particle in the Standard Model of particle physics, the elusive Higgs boson, was discovered in 2012, almost 50 years after it was first predicted by Brout, Englert and Higgs. Finding evidence of the Higgs boson decaying into b quarks is an important undertaking towards further validating the Standard Model of particle physics and determining the exact nature of the Higgs boson.

In this thesis, a search has been presented that considers a Higgs boson produced in association with a W boson, where the Higgs boson decays to $b\bar{b}$ and the W boson decays leptonically. The search is performed using data from proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 8$ TeV collected by the ATLAS experiment at the LHC, corresponding to an integrated luminosity of 20.1 fb^{-1}. The reconstructed invariant mass of two b-quark jets was used as the main criteria for the analysis, the so-called dijet mass analysis. The combination of the $WH \to \ell\nu b\bar{b}$, $ZH \to \nu\bar{\nu}b\bar{b}$ and $ZH \to \ell^+\ell^- b\bar{b}$ production modes was also presented. In the $WH \to \ell\nu b\bar{b}$ search, particular attention was drawn to the multi-jet background estimation and the corresponding systematic uncertainties, where the author made a large contribution to the analysis.

A Standard Model Higgs boson decaying to $b\bar{b}$ is not observed in the search presented. An observed (expected) 95% confidence upper limit on the measured cross section, in units of the expected Standard Model cross section, σ/σ_{SM}, for $m_H = 125$ GeV in the $WH \to \ell\nu b\bar{b}$ channel is found at 3.9 (1.6). For the combination of all three channels, the observed (expected) limit is 2.1 (1.1).

The observed (expected) significance from $WH \to \ell\nu b\bar{b}$ for a Higgs boson with a mass of 125 GeV is 2.7σ (1.3σ). When combined with the associated Z boson channels, an observed (expected) significance of 2.01σ (1.94σ) is seen.

These results are consistent with a Standard Model Higgs boson of mass $m_H = 125$ GeV; however, not enough data was available to make a conclusive statistical claim as to whether or not the Higgs boson does decay to $b\bar{b}$. Data taken during Run 2 of the LHC will be essential in the measurement of this decay.

Run 2 of the LHC began in 2015 with proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. Heavily boosted particles are likely to be produced in these high energy collisions and bring additional challenges to analyses. For example, a boosted Higgs boson decaying to $b\bar{b}$ becomes more difficult to observe due to the collimation of the two b-quark jets. This challenge is not unique to Higgs boson decays, and generic techniques are required for identifying hadronic decays of boosted bosons. Methods of combining jet grooming algorithms with a

jet substructure variable have been presented, with a focus on boosted W bosons in $\sqrt{s} = 13$ TeV collisions. These boosted W boson taggers form a baseline identification technique against which further improvements can be tested.

This thesis presented a study of the potential gain in identification of boosted W bosons by using machine learning techniques. Two machine learning algorithms have been implemented: boosted decision trees (BDT) and deep neural networks (DNN). These algorithms are used to combine six substructure variables into a single discriminant that can be used to separate boosted W bosons from a QCD multi-jet background. Improvements of the BDT and DNN W boson taggers over the baseline W boson taggers in the background rejection power at 50% signal efficiency of between 36% and 13%, respectively, have been achieved in testing. No systematics have yet been considered for these W boson taggers. Possible improvements to these methods were also discussed.

The Standard Model of particle physics, although exceptionally successful, has a number of shortcomings. Many new physics models, containing additional particles or interactions, have been proposed to address some of these limitations. The high energy collisions during Run 2 have the possibility to produce heavy particles predicted by these new physics models. Narrow diboson resonances have been considered in this thesis in the context of a Heavy Vector Triplet model, which predicts heavy W' and Z' bosons that can decay into boosted W bosons. The selection from the $Z' \to WW \to q\bar{q}q\bar{q}$ search (where Z' has a mass of 1.2 TeV) is used as a way to validate the BDT and DNN W boson taggers. Using Run 2 data corresponding to an integrated luminosity 3.2 fb^{-1}, these W boson taggers are compared with a baseline selection that uses a single jet substructure variable. Data-to-Monte Carlo comparisons show good agreement for both of the W boson taggers in a QCD background enriched region. The performance of the W boson taggers as applied to this analysis is in agreement with the results obtained in testing.

Run 2 of the LHC promises to deliver an immense amount of data, up to 100 fb^{-1}, at higher energies than ever before. Many exciting and interesting prospects to further measure the properties of the Higgs boson and to search for New Physics lie ahead.

Appendix A

GPU-based HLT

A.0.1 Graphics Processing Unit

Graphics Processing Units (GPUs) are primarily used for graphical applications and computer physics engines [242]. However, due to their highly parallel architecture they have found use in high performance computing. Current GPUs from NVIDIA have thousands of processing cores, far more than are found in multi-core server [1] or desktop CPUs, which typically have up to eight cores. The GPU cores are effective at processing data in parallel with many concurrent threads, whilst the CPUs are designed to run fewer threads, faster. GPUs are able to produce performance in the TeraFLOPs range.

The parallel computing platform used on NVIDIA GPUs is called CUDA (Compute Unified Device Architecture) [243]. This provides a framework for accessing the GPU through C/C++ code. CUDA is only available for NVIDIA hardware, however, and it changes between hardware versions. Another more general framework that can be used for this as well called OpenCL [244]. This is not supported or maintained by NVIDIA, so whilst it can be used to run on NVIDIA devices it is not able to use the specialised NVIDIA features. It does offer the ability to port code between architectures and can also be run on multi-core processors.

The features and implementations in NVIDIA GPUs differ between architectures, but the overall schema is the same. Threads are run on the GPUs in groups called warps which are collections of 32 threads, which are then grouped into thread blocks, which in turn are grouped into grids (see Figure A.1) [245, 246]. All threads in a warp must execute the same instructions, or kernel. The maximum sizes of the thread blocks and grids are limited by the maximum number of

[1] The Intel Xeon-Phi multi-processor has 48 cores.

threads and warps that the GPU can run at once and by the number of thread blocks a single Streaming Multiprocessor can run. The GPU is split into a number of Streaming Multiprocessors (SM) which each contains a number of the cores. The memory on the GPU is divided into local or global memory, shared memory and registers. Registers are limited to a single thread, shared memory is available to a thread block and global memory is available to all threads. The speeds of the different memories decrease with increasing scope, but the largest bottleneck comes from copying data from the host device to the actual GPU memory. Generally, code that is written to run on a normal CPU has to be rewritten and redesigned to run optimally on a GPU [245, 247].

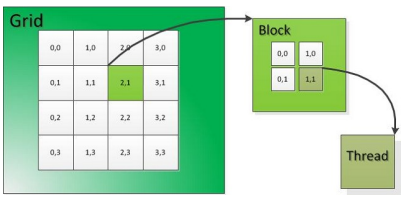


Figure A.1 *The abstraction of threads being grouped into blocks that are grouped into grids in the CUDA programming framework.*

Two recent architectures are Fermi [248, 249] and Kepler [250, 251][2]. Fermi offers 32 (version 2.0) or 48 (version 2.1) cores per SM and Kepler has up to 192 cores per SM. There are general increases to the amount of memory, the maximum number of threads and the scheduling. Kepler offers dynamic parallelism so threads can spawn new threads, no longer relying on the CPU to do this.

GPUs offer access to hundreds of cores with thousands of threads, which can be used to do parallel processing on a large scale where SIMT (Single Instruction Multiple Threads) can be utilised. Algorithms where a single instruction can be carried out on multiple data or where all data is independent are good candidates to be run in parallel. GPUs offer a lot of on-device memory that can be accessed from the host device at high speed. However, these can be difficult to program optimised code for and the memory management can be restrictive. The time taken to read and write data between the host memory and GPU memory is a lot slower than processing time in many cases, so this adds a further constraint.

[2]At the time this study was done Kepler was the most recent. Maxwell and Pascal have since superseded it.

A.0.2 ATLAS Trigger System

The basic outline of the ATLAS trigger was given in Section 3.4. Some of this is expanded on here in the context of the GPU-based HLT.

Triggers make fast decisions about whether or not to keep events. Events that do not contain interesting physics or do not pass quality requirements are discarded. The LHC has a design collision rate of 40 MHz, which must be reduced to ≈ 1000 Hz by the trigger system (for Run 2). Triggers also help to reject QCD multi-jet events and keep low cross section events.

There are three hardware and software triggers implemented in the ATLAS detector:

- Level 1 (L1) is a hardware based trigger,
- Level 2 (L2) is a software based trigger,
- and the Event Filter (EF) is also based in software.

L2 and EF together are known as the High Level Trigger (HLT).

When a region-of-interest (RoI) is identified by the L1 trigger, the data are transferred to Readout Buffers (ROB) that store the data pending the L2 trigger decision. The ROBs are detector specific each containing partial event information.

The L2 trigger runs fast software algorithms to analyse the RoIs from L1. L2 has access to the full granularity of the detector within the RoI including tracking data. The Readout Buffers are grouped into Readout Systems (ROS) that contain data from all parts of the detector and the L2 trigger can access any of this data. The L2 trigger is the first level of the trigger system to get access to data from the pixel and SCT detectors. The data from the pixel and SCT modules is stored in ROBs, which is handled by Readout Drivers (ROD). The ROBs are encoded into a bytestream, which must be decoded when the L2 trigger wants access to the module information. The decoded information must be analysed to find all the modules with hits, the hits clustered and converted to actual spacepoints.

The decoding and clustering lend themselves to parallel processing, however, the decoding is currently done sequentially. A previous study was done where the bytestream decoding is parallelised at the level of ROB fragments, but it was

highlighted that this could be parallelised further by moving to the level of module readouts. The same study did the clustering in parallel as well. This study used NVIDIA Tesla C2050 GPU cards to run the data preparation in parallel, showing massive speed gains over CPU based methods.

The EF uses seeds from the L2 trigger, but has access to the complete set of data for tracking, typically running the same sort of analysis as offline reconstruction software. The Event Builder combines all the fragments from the ROBs to give the EF access to the full event. Events that pass the full trigger system are written into data streams and stored offline.

The L2 and EF were originally designed such that they would be standalone using different software and provide slightly different functionality. The current upgrade of the ATLAS trigger system will see L2 and the EF merged and done on a single node, instead of having high-speed network connections between L2 and EF nodes [252]. This reduces time spent decoding data and running the same software on different nodes. The EF does not need to decode the data in the RoI if L2 has already requested it. This allows the entire HLT to benefit from parallel processing done at the L2 level. In addition, the hardware needs to be upgraded to increase processing power and speed to deal with increased luminosity at the LHC. The use of GPUs for the HLT upgrade is studied.

A.0.3 GPU HLT Studies

Previous studies were performed on NVIDIA Tesla C2050 (Fermi 2.0 architecture) High Performance Computing (HPC) cards at RAL [253, 254]. The benchmark software used was written to perform the data preparation chain for the Pixel and SCT detectors, going from the incoming bytestream to calculating regions of interest and finding the associated spacepoints. These Tesla cards are 2nd generation CUDA cards from NVIDIA with the Fermi 2.0 architecture. The current top of the line HPC cards from NVIDIA use the Kepler architecture. The Kepler K20 card should give much better performance in theory, offering more cores, memory, and threads. However, the benchmark code was written and optimised for the Fermi range.

The testing was done on an NVIDIA Kepler K20, NVIDIA GeForce GT 650M (Kepler architecture), an NVIDIA GeForce GT 630M (Fermi 2.1 architecture) and a first generation NVIDIA Tesla C1050. The initial results showed that without

Method Name	Card Type	Register	Shared Memory	Block Size
Pixel Clustering	Fermi 630M	83%	100%	100%
	Kepler K20	50%	73%	100%
SCT Clustering	Fermi 630M	77%	36%	33%
	Kepler K20	64%	27%	25%
SCT Space Points	Fermi 630M	65%	100%	17%
	Kepler K20	59%	73%	12.5%

Table A.1 *Theoretical occupancy limits for the NVIDIA Kepler K20 and NVIDIA Fermi 630M GPUs. The occupancy limits are shown for the three most expensive methods in the HLT benchmark code and are calculated separately for the registers, shared memory and block size limitations.*

significant optimisation of the software (i.e. using it in the same configuration), Kepler K20 does not perform well, even being outperformed by laptop GPUs (630M). Whilst this is not a good indication of the actual performance of the hardware, it is a good indication of the amount of effort that has to go into maintaining the code for this. If the Kepler cards were to be used for the HLT upgrade, the code for this would have to be redone whenever the architecture or card was changed in the future.

NVIDIA profiling tools were used for optimising the code and finding bottlenecks [255–257]. Finding the pieces of code that take the longest to run and then analysing those methods in detail is a useful way of approaching this. One of the indicators of performance is the occupancy. This gives an indication of how close the card is to being fully utilised. Having a low occupancy increases idle time for the GPU and lowers throughput. It is dependent on how many threads are actually being run and how well the shared memory and registers are being filled and distributed. Occupancy can be defined as the active number of threads / the maximum number of threads possible. In general, newer GPUs have more cores than a CPU, which run at lower clock speeds. If a GPU has a low occupancy it is likely that it will run more slowly due to the slower clock speeds. The number of threads per block and the memory usage have been analysed on the different cards using NVIDIA's profiling software, Visual Profiler, to find the occupancy, as shown in Table A.1.

The different methods in the benchmark software are analysed separately. The methods are ranked according to their run-time and are similarly ranked between architectures, but the occupancy changes. This is due to a number of factors:

the maximum number of threads changing between architectures, the number of registers (fast memory allocated to a single thread) assigned to each thread, not making proper use of shared memory and the thread block sizes being too small. The thread block sizes range from 32 up to 256. The block sizes of 32 do not come close to the maximum allowance and the Streaming Multiprocessors are heavily underutilised. Reducing the number of registers per thread will increase the number threads being run at once and increase the occupancy, at the cost of slower memory access, but this needs to be tested and quantified [258].

The theoretical occupancy for the most expensive three methods in the data preparation was calculated for both the 630M and the Kepler and these are listed below. The three methods, pixel clustering, SCT clustering and SCT space point making, were found to have the theoretical occupancies listed in Table A.1.

Changing the number of threads per block and the number of blocks per grid to optimise theoretical occupancy did not show big improvements. A big improvement has been seen by changing the number of blocks per grid to a multiple or factor of the number of Streaming Multiprocessors on the GPU (two for the 630M and 13 for the K20) and by changing the number of threads per block to a multiple or factor of the number of cores per Streaming Multiprocessor. The results for the K20 are quantified in Figure A.2. Overall, this sees a decrease in actual computation time of a factor of about 10% for the 630M and 33% for the K20. There are relatively large memory copy overheads on K20 that are not yet understood that add significantly to the overall run-time. These add up to about 3.5 s for the K20, whereas these are only about 100 ms for the 630M. The time per method and the calls to each method are both decreased by changing the thread and block sizes.

The benchmark software has some built-in timing methods that are less accurate than the profiling software; however, it is the only comparison that was available with the Tesla C2050 GPU. The mean processing time for the C2050 is 9.868 ms, divided amongst all calls to the methods which is still much lower than the optimised results for the 630M. The 630M sees a mean processing time of approximately 20 ms and for the K20 it is higher at 23 ms. The K20 showed dramatic improvement on this benchmark after the HLT code was tuned, dropping to 23 ms from 37 ms. This benchmark is not accurate, however it does give an indication of performance increase.

Parts of the bytestream decoding and clustering code were rewritten to increase

a)

Unoptimised 630M				Optimised 630M			
Time(%)	Time	Calls	Name	Time(%)	Time	Calls	Name
33.18	1.41s	4179	PixelClusterizationKernel	34.08	1.30s	1800	PixelClusterizationKernel
27.37	1.16s	9640	SctClusterizationKernel	29.7	1.14s	8445	SctClusterizationKernel
26.4	1.12s	4865	SctSpacePointMakingKernel	23.48	898.21ms	4266	SctSpacePointMakingKernel
3.15	133.64ms	200	SctPostProcessingKernel	3.36	128.44ms	200	SctPostProcessingKernel
2.44	103.45ms	200	PixelPostProcessingKernel	2.57	98.51ms	200	SctDecodingKernel
2.33	98.59ms	200	SctDecodingKernel	2.49	95.41ms	2179	PixelSpacePointMakingKernel
2.28	96.54ms	2179	PixelSpacePointMakingKernel	1.84	70.28ms	200	PixelPostProcessingKernel
1.68	71.25ms	200	PixelDecodingKernel	1.2	45.76ms	200	PixelDecodingKernel
0.54	22.75ms	1002	[CUDA memcpy HtoD]	0.59	22.73ms	1002	[CUDA memcpy HtoD]
0.39	16.55ms	1000	[CUDA memcpy DtoH]	0.42	16.15ms	1000	[CUDA memcpy DtoH]
0.24	10.25ms	400	[CUDA memset]	0.27	10.29ms	400	[CUDA memset]
0	18.43us	4	[CUDA memcpy HtoA]	0	18.37us	4	[CUDA memcpy HtoA]

b)

Unoptimised K20				Optimised K20			
Time(%)	Time	Calls	Name	Time(%)	Time	Calls	Name
39.76	2.77s	4179	PixelClusterizationKernel	37.58	1.49s	1076	PixelClusterizationKernel
29.34	2.04s	4865	SctSpacePointMakingKernel	24.57	976.49ms	2417	SctSpacePointMakingKernel
18.22	1.27s	9640	SctClusterizationKernel	22.89	909.56ms	4815	SctClusterizationKernel
3.03	210.89ms	200	SctPostProcessingKernel	3.31	131.68ms	200	PixelPostProcessingKernel
2.92	203.08ms	200	SctDecodingKernel	3.14	124.96ms	200	SctPostProcessingKernel
2.76	192.04ms	200	PixelPostProcessingKernel	2.56	101.88ms	200	SctDecodingKernel
1.36	94.42ms	2179	PixelSpacePointMakingKernel	2.28	90.73ms	2076	PixelSpacePointMakingKernel
1.32	92.04ms	200	PixelDecodingKernel	1.4	55.69ms	200	PixelDecodingKernel
0.64	44.40ms	1002	[CUDA memcpy HtoD]	1.12	44.48ms	1002	[CUDA memcpy HtoD]
0.61	42.75ms	1000	[CUDA memcpy DtoH]	1.08	42.81ms	1000	[CUDA memcpy DtoH]
0.03	2.39ms	400	[CUDA memset]	0.06	2.42ms	400	[CUDA memset]
0	20.42us	4	[CUDA memcpy HtoA]	0	20.99us	4	[CUDA memcpy HtoA]

Figure A.2 *Comparison of the run-times for the different methods of the HLT benchmark software on the a) NVIDIA GeForce GT 630M and b) NVIDIA Kepler K20 GPUs. The largest areas of improvement are highlighted.*

the occupancy and be more efficient, with substantial gains seen in some methods (see Figure A.3). Additional changes were made to the algorithms in order to take advantage of NVIDIA Kepler's Dynamic Parallelism [259], whereby threads can launch other threads.

Unoptimised Tesla K20				Optimised Tesla K20			
Time(%)	Time	Calls	Name	Time(%)	Time	Calls	Name
39.76	2.77s	4179	PixelClusterizationKernel	29.05	978.44ms	2417	SctSpacePointMakingKernel
29.34	2.04s	4865	SctSpacePointMakingKernel	26.98	908.76ms	4815	SctClusterizationKernel
18.22	1.27s	9640	SctClusterizationKernel	26.82	903.33ms	1076	PixelClusterizationKernel
3.03	210.89ms	200	SctPostProcessingKernel	3.95	133.02ms	200	PixelPostProcessingKernel
2.92	203.08ms	200	SctDecodingKernel	3.69	124.23ms	200	SctPostProcessingKernel
2.76	192.04ms	200	PixelPostProcessingKernel	3.03	101.90ms	200	SctDecodingKernel
1.36	94.42ms	2179	PixelSpacePointMakingKernel	2.66	89.65ms	2076	PixelSpacePointMakingKernel
1.32	92.04ms	200	PixelDecodingKernel	1.65	55.68ms	200	PixelDecodingKernel
0.64	44.40ms	1002	[CUDA memcpy HtoD]	1.24	41.79ms	1000	[CUDA memcpy DtoH]
0.61	42.75ms	1000	[CUDA memcpy DtoH]	0.86	28.88ms	1002	[CUDA memcpy HtoD]
0.03	2.39ms	400	[CUDA memset]	0.07	2.35ms	400	[CUDA memset]
0	20.42us	4	[CUDA memcpy HtoA]	0	20.80us	4	[CUDA memcpy HtoA]
Total	6.93s	24069			3.37s	13590	

Figure A.3 *A substantial decrease in run times after optimising the bytestream decoding and clustering to more efficiently occupy all cores on the GPU and reduce the number of calls to different methods.*

A.0.4 Conclusion and Summary

There is a strong case for using GPUs for the HLT at ATLAS. Significant gains can be achieved. There can be substantial changes needed to maintain this performance when moving to newer generations of the hardware. However, this is unlikely to be required often, if at all, and since these GPUs would be in the HLT farm at CERN, there is no added constraint of having to write the software to run on multiple setups.

Currently, one of the major bottlenecks in this solution is the relatively slow transfer of data to the GPU. This will improve as the GPUs move to faster memory technologies, but this will likely remain the bottleneck. Some additional constraints are due to the difficulty of merging the GPU-based code with the current software framework used by ATLAS, ATHENA. Which processes to offload to the GPU needs to be investigated, as not all processes are suited to running in parallel.

Appendix B

Truth Tagging

The truth tagging method generates a random MV1c value for a jet such that it is forced to pass the loose cut (a number greater than 0.4050). This is described below:

1. Each jet is tagged with an efficiency dependent on its MV1c value and kinematics. Loop through all the MV1c bins and find the efficiency of the jet as if its MV1c value was the median of these two bins. Add all the efficiencies together where the jet passes the loose cut to get the efficiency that it passes the loose cut. The actual MV1c value for the jet isn't known (hence the spoofing) so a number of values in different OPs are used to evaluate the efficiency.

2. Get a random value for the jet corresponding to the cumulative efficiency of this jet passing in each MV1c OP bin.

3. Then calculate a weight value by looping through each MV1c bin until the cumulative efficiency passes the random value above. This is then taken as the final MV1c value of the jet. If the random value does not match any of the above conditions, the MV1c value corresponding to the loose cut is used.

Once an MV1c value is found for the jet a scale factor is applied to account for the efficiency of forcing the jet to pass the loose cut.

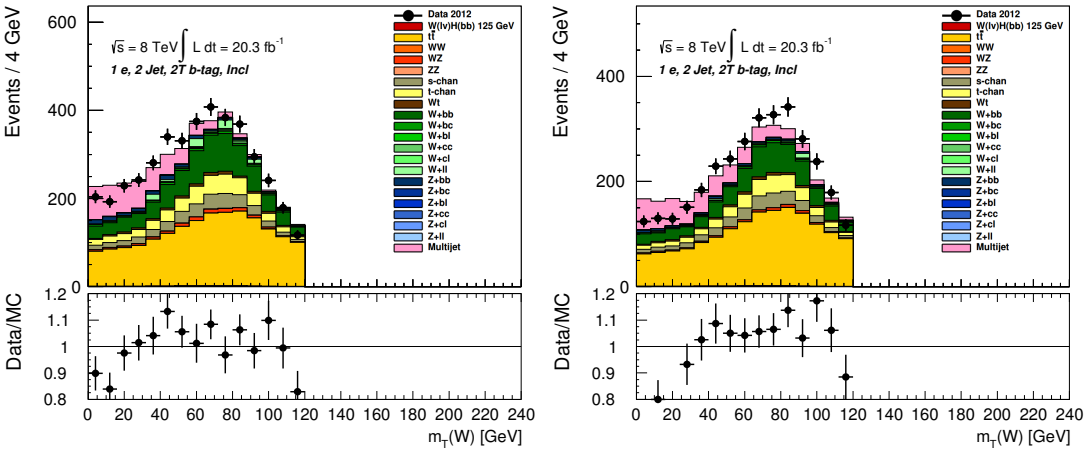
A weight is calculated to account for the truth tagging, which takes the probability of the second jet failing, provided the first passed, and vice versa and adds these two values. The jet is tested in each of the MV1c bins and the tagging efficiency is calculated, and this is multiplied by an additional scale factor (SF) taken from the b-tagging calibration interface [137, 138]. The failure weight, w_f, is given

by the scale factor multiplied by the efficiency for the jet when it does not pass the loose cut. The pass weight, w_p, is the sum of all products scale factor and efficiency for the jet in each of the bins of the MV1c operating points from loose to tight. The overall weight is then given by $w = w_p(j_1) * w_f(j_2) + w_p(j_2) * w_f(j_1)$, where $j_{1/2}$ refers to the leading or sub-leading jet in the event.

Appendix C

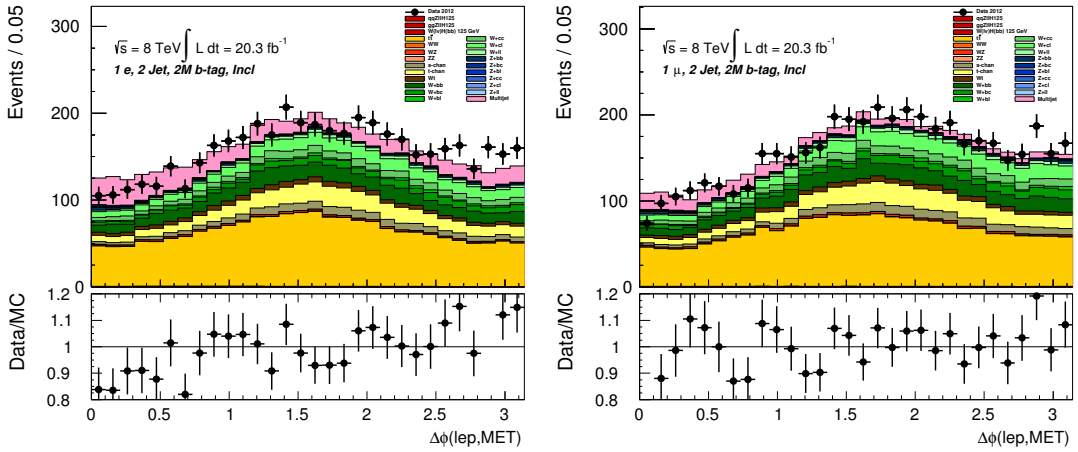
Mismodelling of H_T and $\Delta\phi(l, E_T^{\text{miss}})$

Mismodelling can be seen after the H_T selection cut is applied. In particular, the H_T cut broadens most kinematic distributions, shifting the peaks and removing a large part of the multi-jet background. The transverse mass of the W boson, m_T^W, has much worse modelling after the H_T cut (see Figure C.1), which originates from the angular variable $\Delta\phi(l, E_T^{\text{miss}})$ (the angle between the neutrino and the lepton in the W decay), which itself is poorly modelled (see Figure C.2). When cutting on H_{T} this mismodelling becomes more pronounced due to the low number of events. The source of the broadening distributions is the result of the multi-jet contribution being removed, which in most distributions has a pronounced peak. The shifting of the peaks of the distributions is due to the H_T cut removing softer contributions from the leptons, jets and E_T^{miss}.



(a) *Pre-H_T el channel* **(b)** *Post-H_T el channel*

Figure C.1 *Distributions of m_T^W before (a) and after (b) the H_T cut in the electron channel with 2 tight b-tags.*



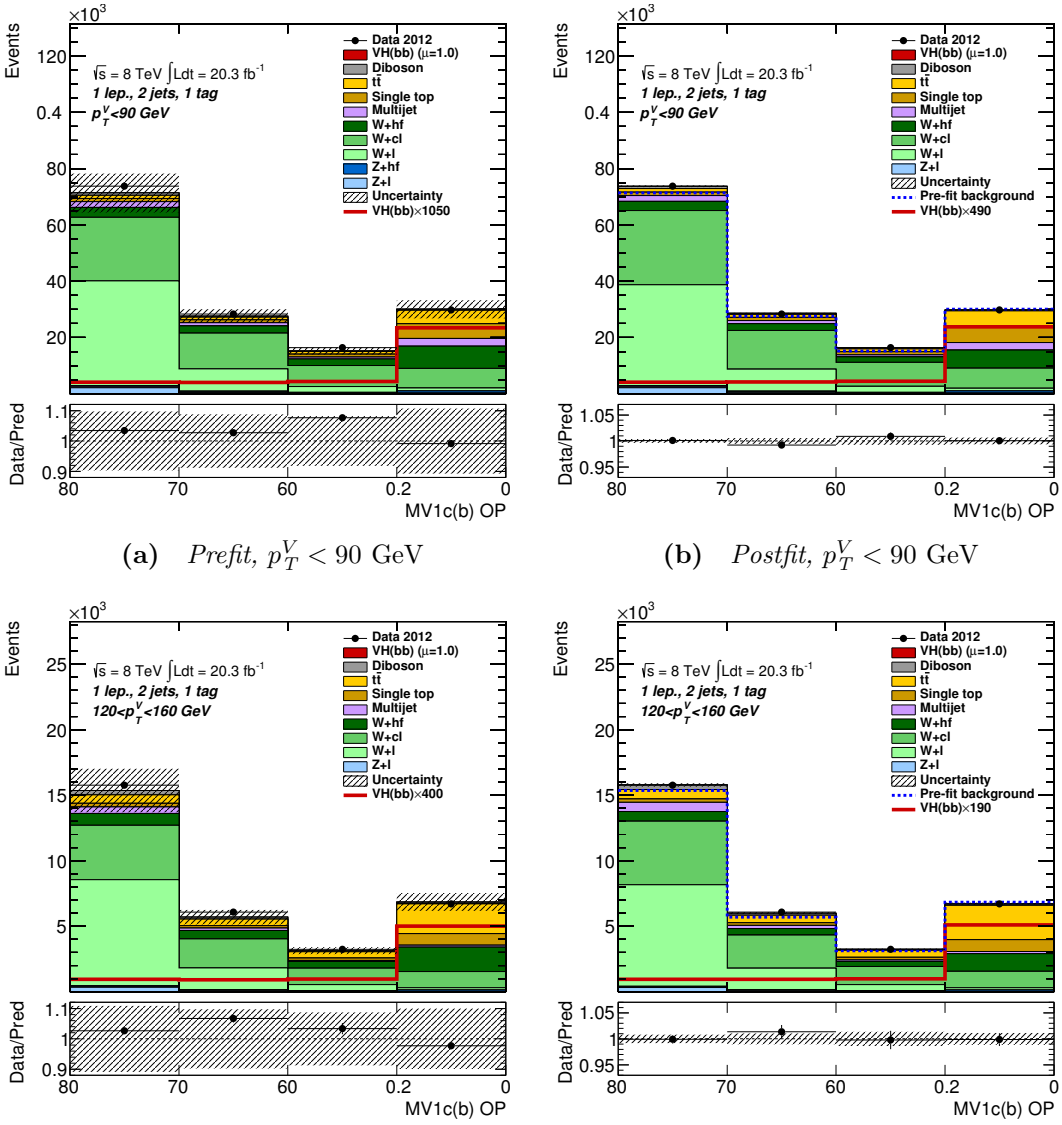
(a) **(b)**

Figure C.2 *Poor modelling of $\Delta\phi(\ell, E_T^{miss})$ is seen in both the electron (shown here in (a)) and muon channels. This leads to poor modelling in the m_T^W distribution as well (b). Shown here is the distribution in the 2 medium b-tag region.*

Appendix D

Input Distributions for $WH \to \ell\nu b\bar{b}$

The input distributions in the 1-lepton channel for the global fit are shown here. The MV1c distribution is used in the 1-tag region and the transformed m_{bb} distribution is used for the 2-tag regions.



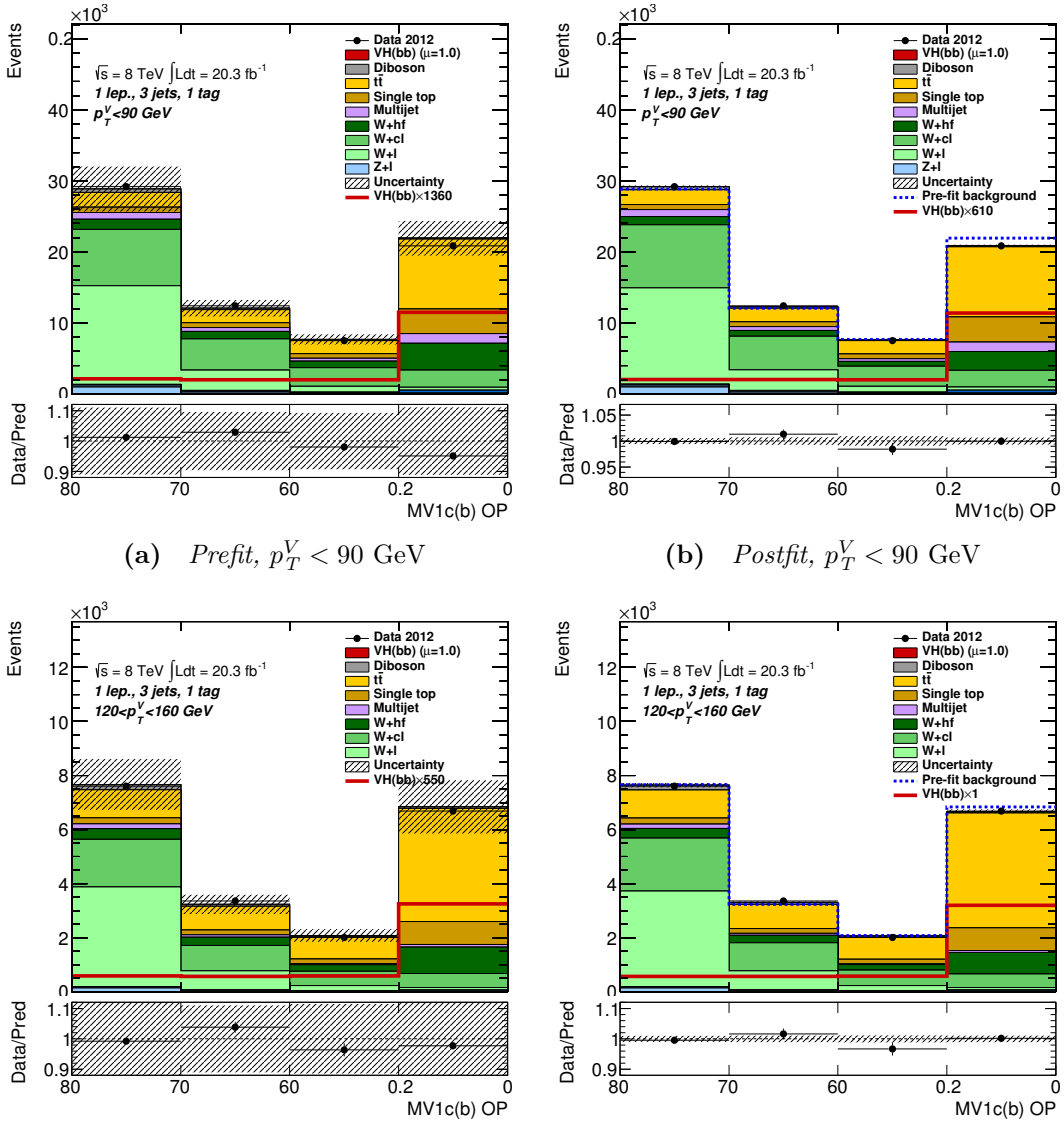
(a) *Prefit*, $p_T^V < 90$ GeV

(b) *Postfit*, $p_T^V < 90$ GeV

(c) *Prefit*, $p_T^V > 120$ GeV

(d) *Postfit*, $p_T^V > 120$ GeV

Figure D.1 *The MV1c distribution in the 2-jet, 1-tag region and the* $p_T^V < 90$ GeV *(top row) and* $p_T^V > 120$ GeV *(bottom row) intervals used as an input for the global fit in the 1-lepton channel before (left) and after (right) the fit is performed.*



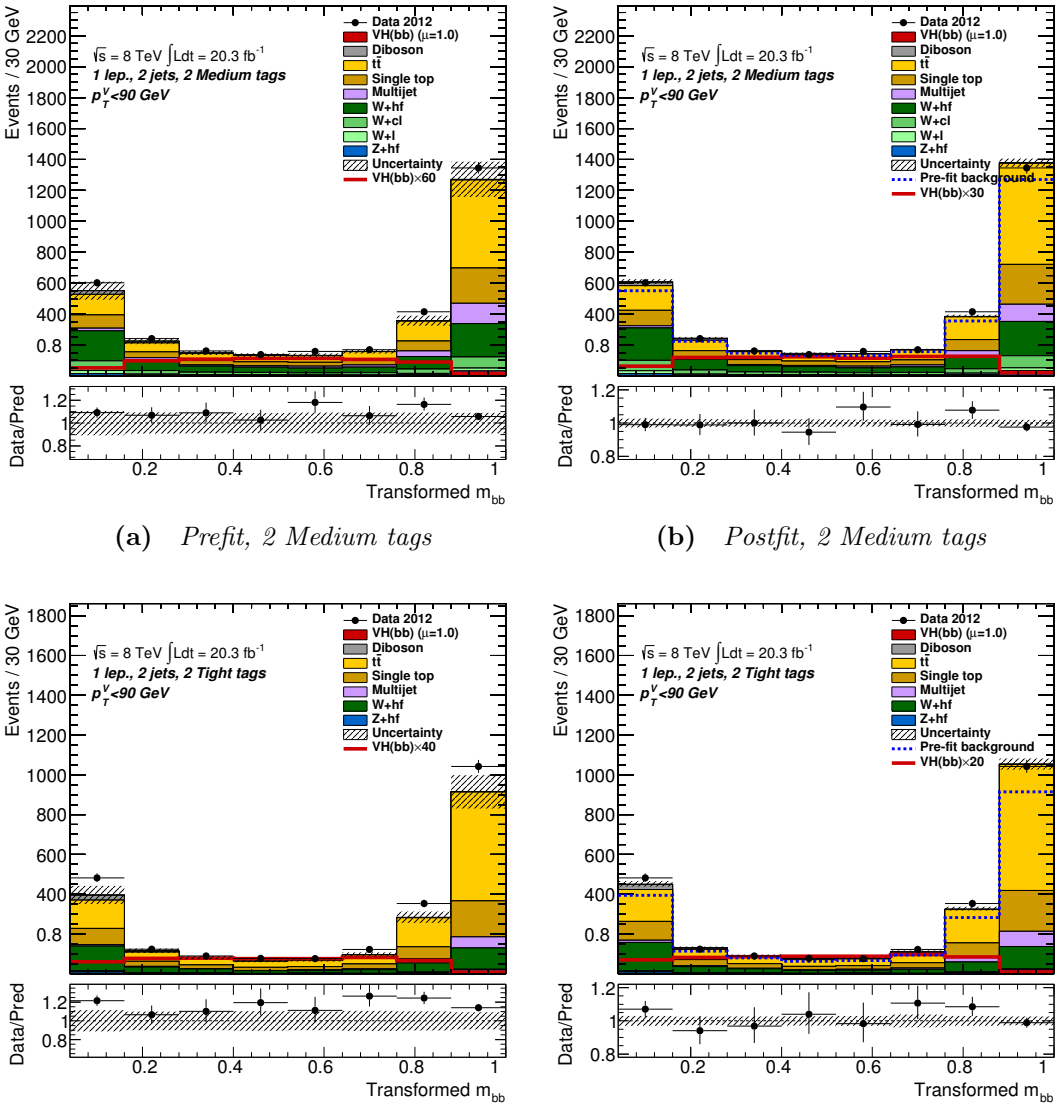
(a) *Prefit*, $p_T^V < 90$ GeV

(b) *Postfit*, $p_T^V < 90$ GeV

(c) *Prefit*, $p_T^V > 120$ GeV

(d) *Postfit*, $p_T^V > 120$ GeV

Figure D.2 *The MV1c distribution in the 3-jet, 1-tag region and the* $p_T^V <$ 90 GeV *(top row) and* $p_T^V > 120$ GeV *(bottom row) intervals used as an input for the global fit in the 1-lepton channel before (left) and after (right) the fit is performed.*



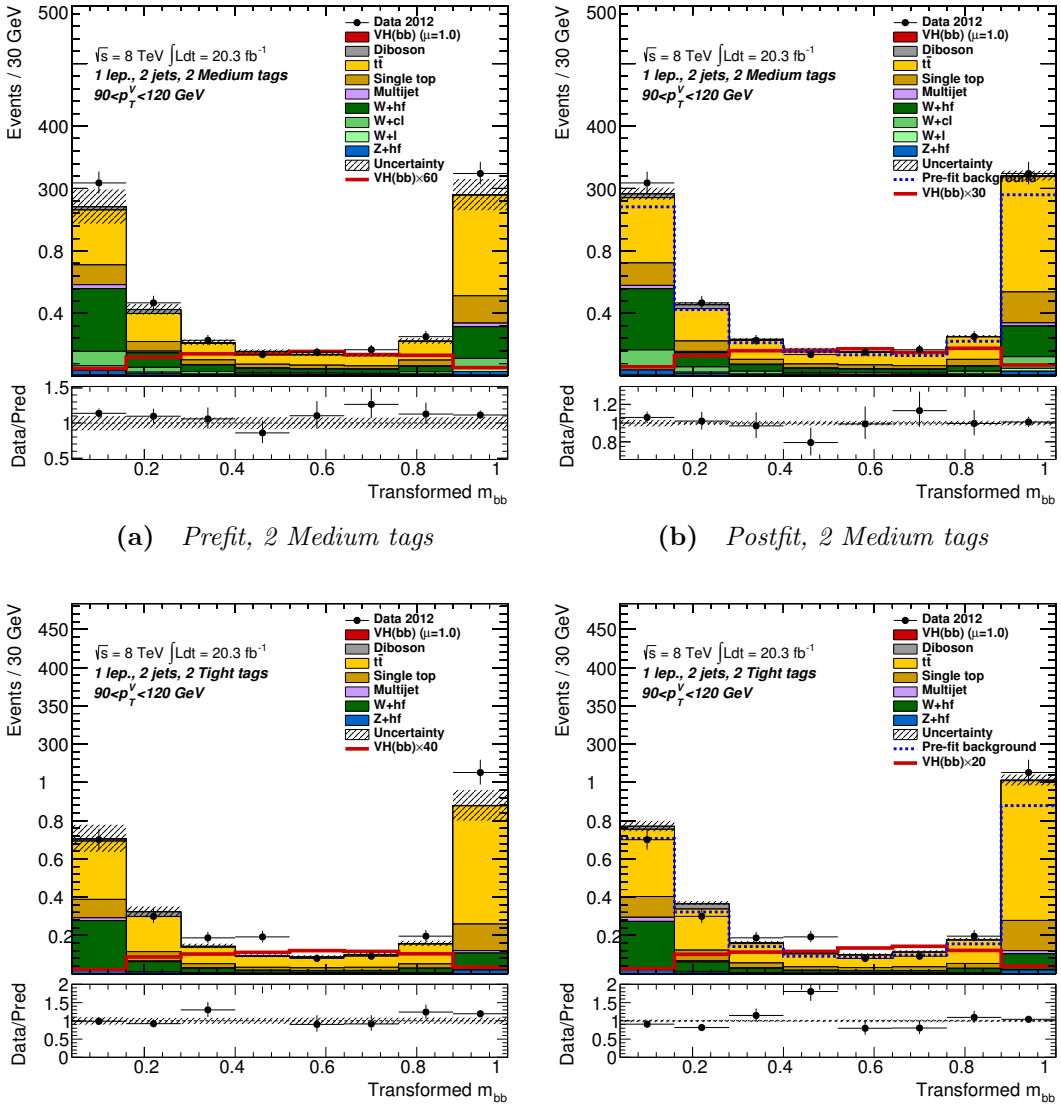
(a) *Prefit, 2 Medium tags*

(b) *Postfit, 2 Medium tags*

(c) *Prefit, 2 Tight tags*

(d) *Postfit, 2 Tight tags*

Figure D.3 *The m_{bb} distribution in the 2-jet, 2-tag, $p_T^V < 90$ GeV region with two medium tagged b-jets (top row) and two tight tagged b-jets (bottom row) used as an input for the global fit in the 1-lepton channel before (left) and after (right) the fit is performed.*



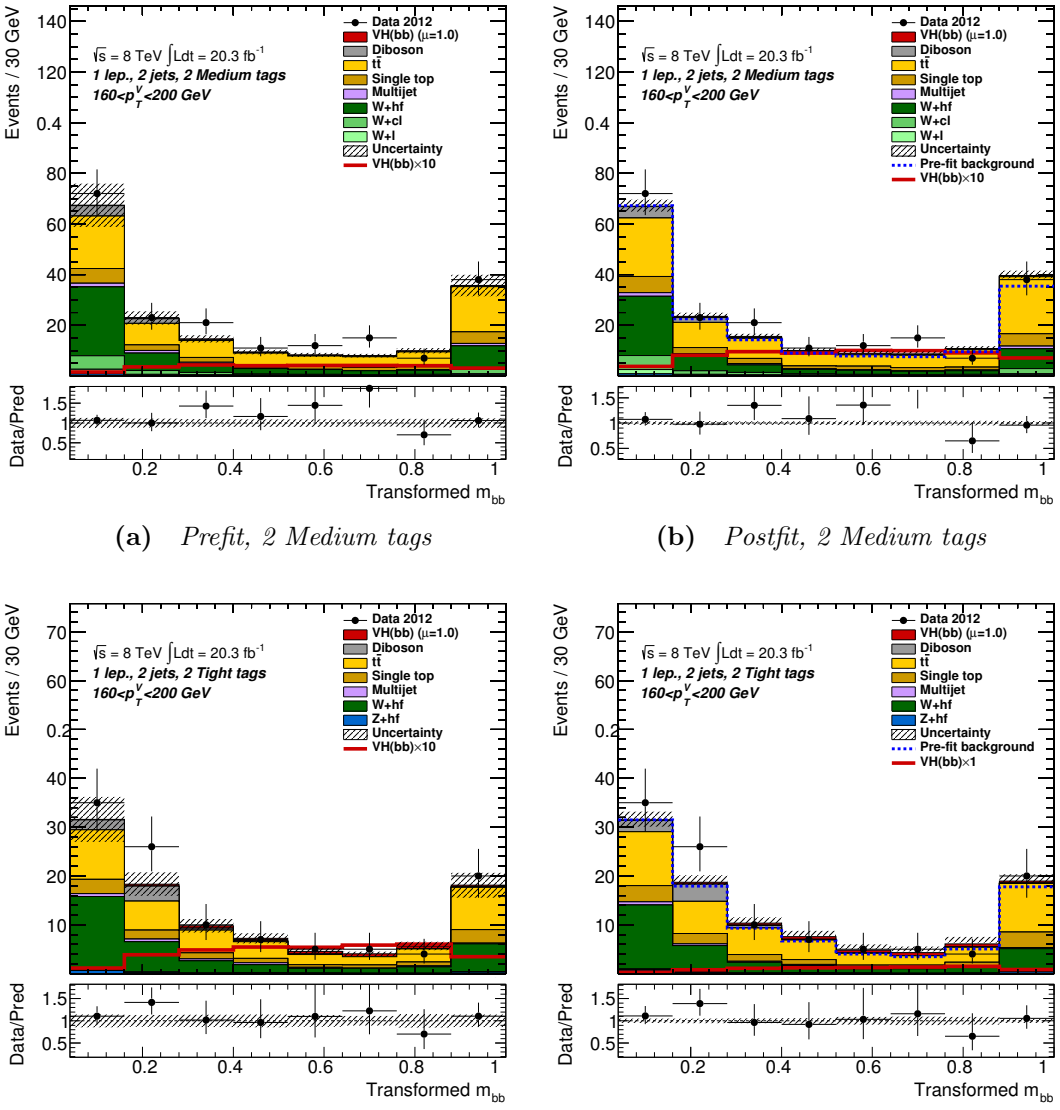
(a) *Prefit, 2 Medium tags*

(b) *Postfit, 2 Medium tags*

(c) *Prefit, 2 Tight tags*

(d) *Postfit, 2 Tight tags*

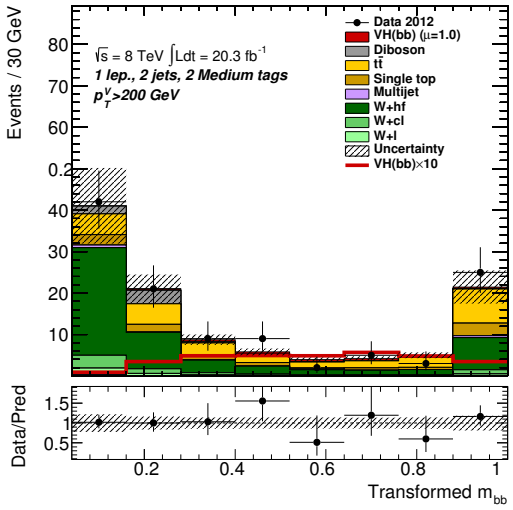
Figure D.4 *The m_{bb} distribution in the 2-jet, 2-tag, $100 < p_T^V < 120$ GeV region with two medium tagged b-jets (top row) and two tight tagged b-jets (bottom row) used as an input for the global fit in the 1-lepton channel before (left) and after (right) the fit is performed.*



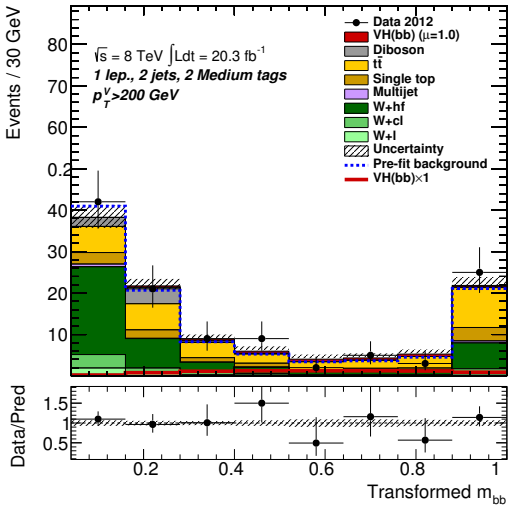
(a) *Prefit, 2 Medium tags* **(b)** *Postfit, 2 Medium tags*

(c) *Prefit, 2 Tight tags* **(d)** *Postfit, 2 Tight tags*

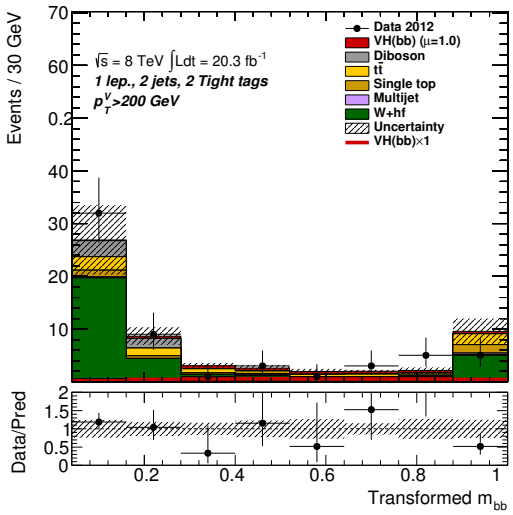
Figure D.5 *The m_{bb} distribution in the 2-jet, 2-tag, $160 < p_T^V < 200$ GeV region with two medium tagged b-jets (top row) and two tight tagged b-jets (bottom row) used as an input for the global fit in the 1-lepton channel before (left) and after (right) the fit is performed.*



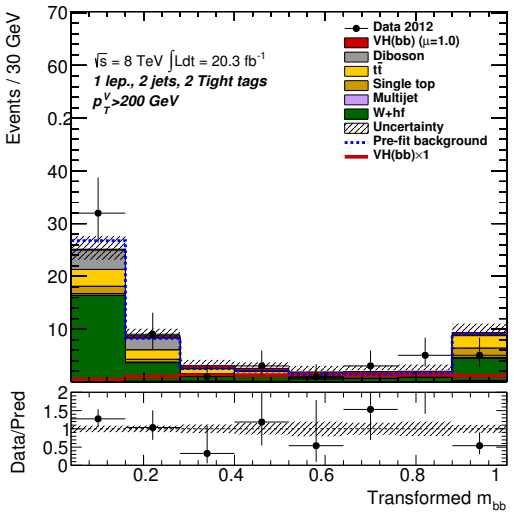
(a) *Prefit, 2 Medium tags*



(b) *Postfit, 2 Medium tags*

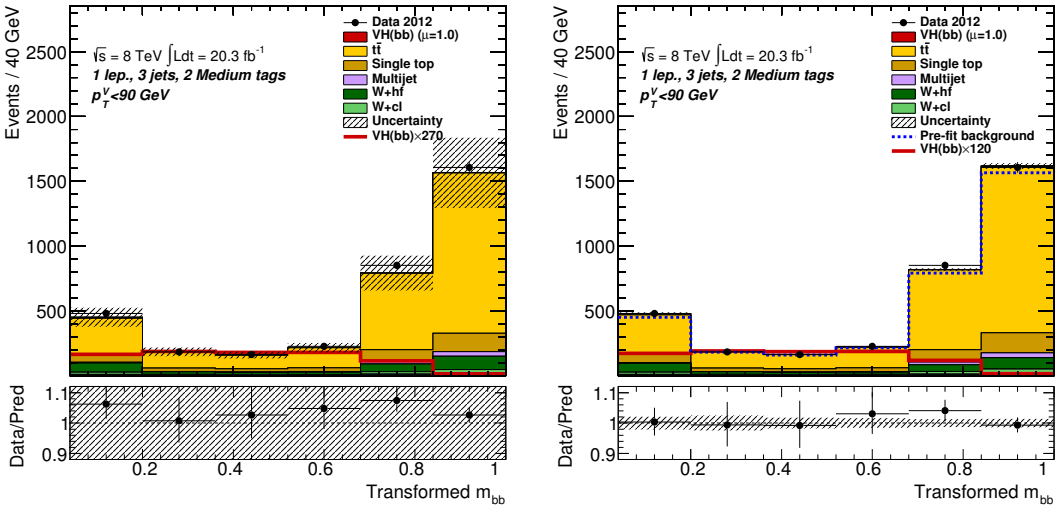


(c) *Prefit, 2 Tight tags*

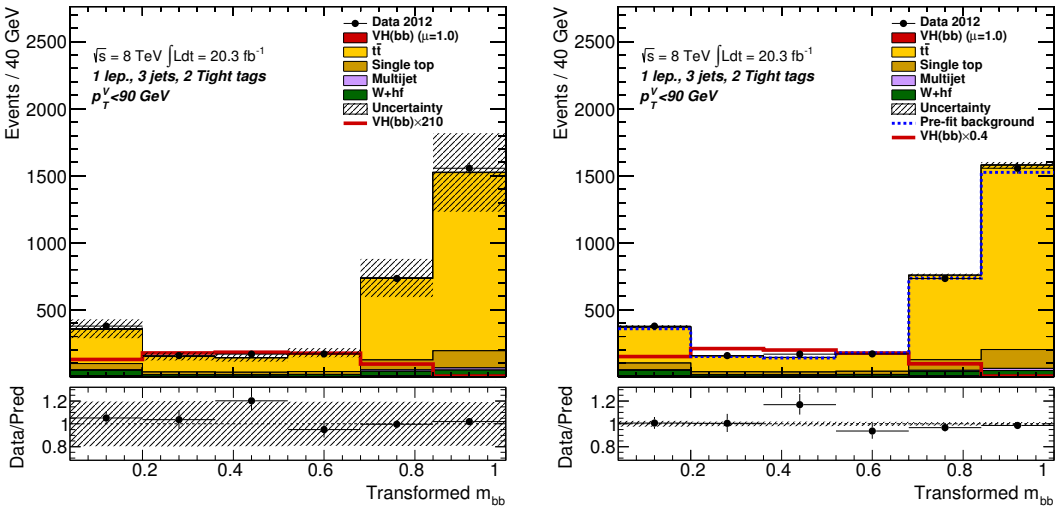


(d) *Postfit, 2 Tight tags*

Figure D.6 *The m_{bb} distribution in the 2-jet, 2-tag, $p_T^V > 200$ GeV region with two medium tagged b-jets (top row) and two tight tagged b-jets (bottom row) used as an input for the global fit in the 1-lepton channel before (left) and after (right) the fit is performed.*

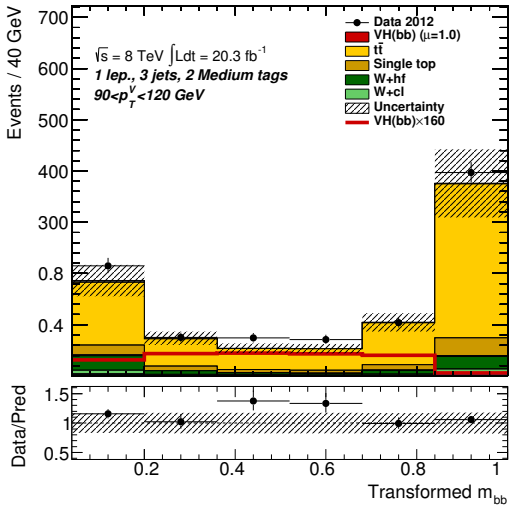


(a) *Prefit, 2 Medium tags* **(b)** *Postfit, 2 Medium tags*

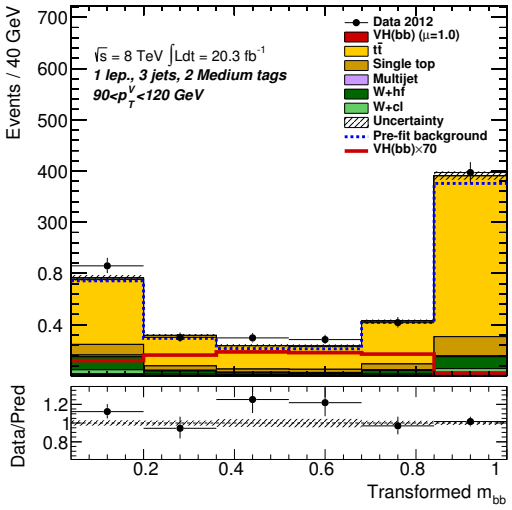


(c) *Prefit, 2 Tight tags* **(d)** *Postfit, 2 Tight tags*

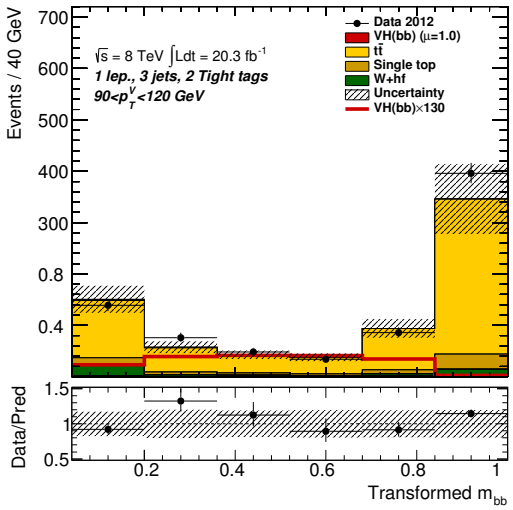
Figure D.7 *The m_{bb} distribution in the 3-jet, 2-tag, $p_T^V < 90$ GeV region with two medium tagged b-jets (top row) and two tight tagged b-jets (bottom row) used as an input for the global fit in the 1-lepton channel before (left) and after (right) the fit is performed.*



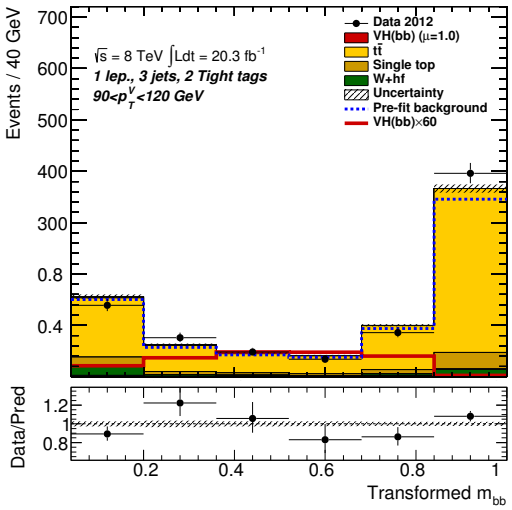
(a) *Prefit, 2 Medium tags*



(b) *Postfit, 2 Medium tags*

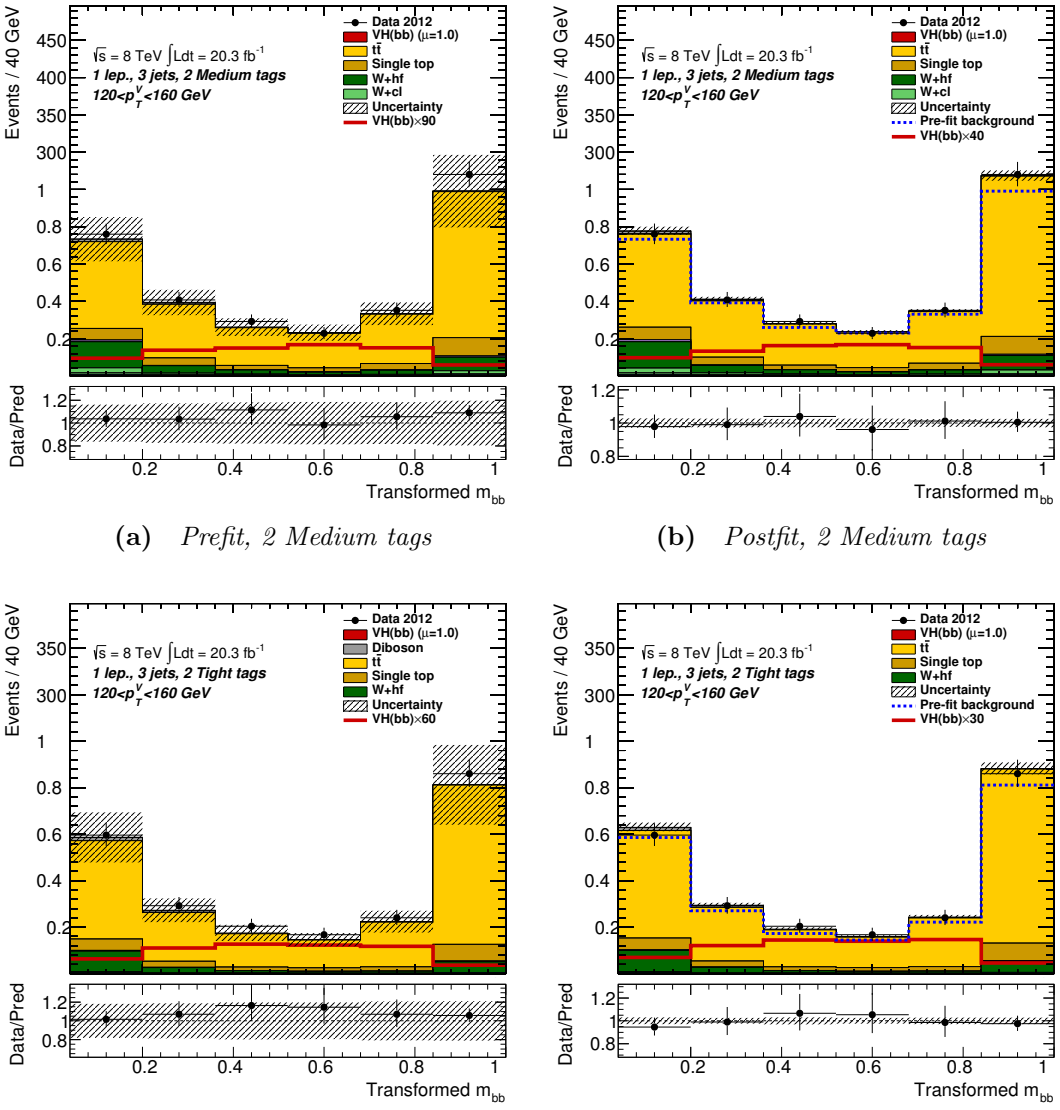


(c) *Prefit, 2 Tight tags*



(d) *Postfit, 2 Tight tags*

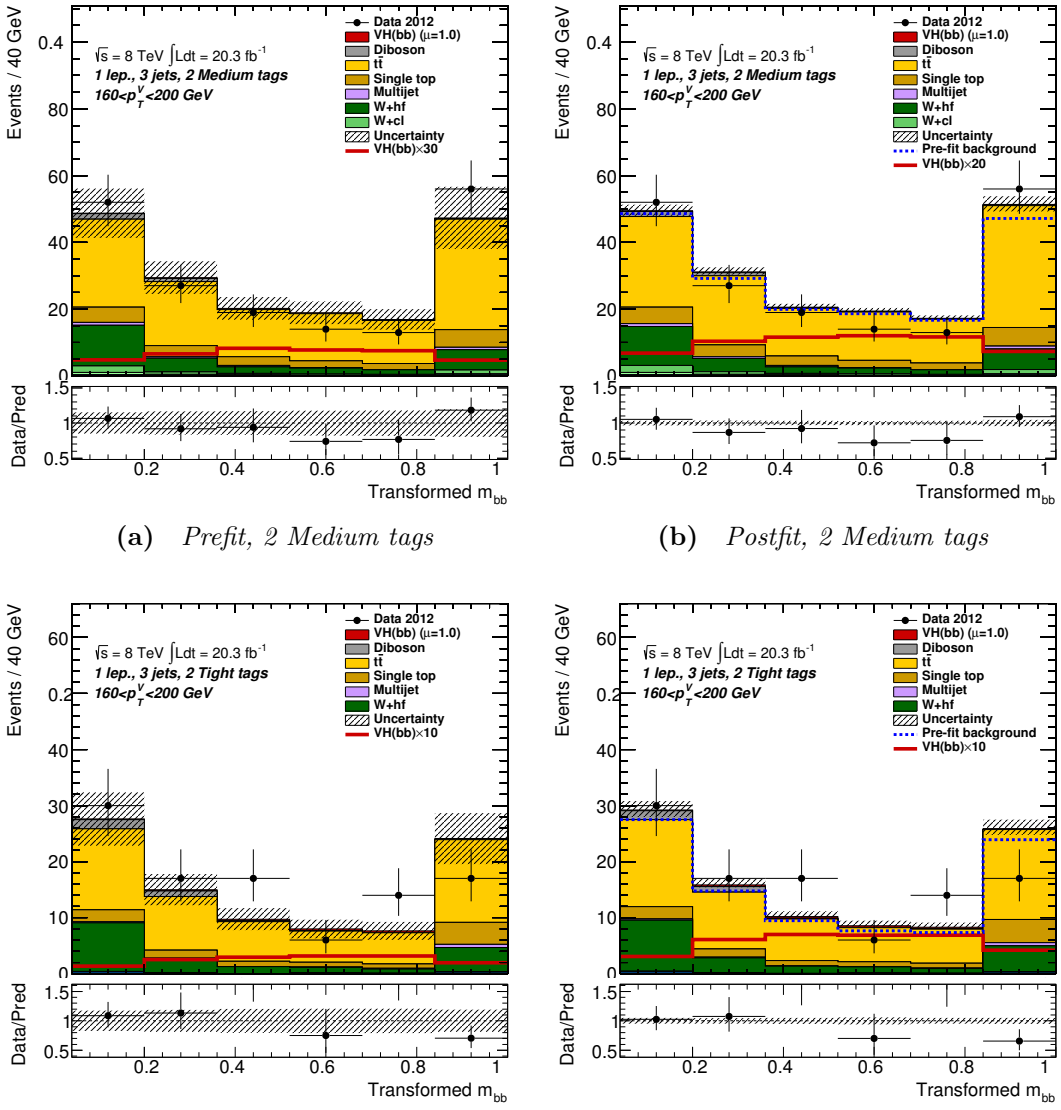
Figure D.8 *The m_{bb} distribution in the 3-jet, 2-tag, $100 < p_T^V < 120$ GeV region with two medium tagged b-jets (top row) and two tight tagged b-jets (bottom row) used as an input for the global fit in the 1-lepton channel before (left) and after (right) the fit is performed.*



(a) *Prefit, 2 Medium tags* **(b)** *Postfit, 2 Medium tags*

(c) *Prefit, 2 Tight tags* **(d)** *Postfit, 2 Tight tags*

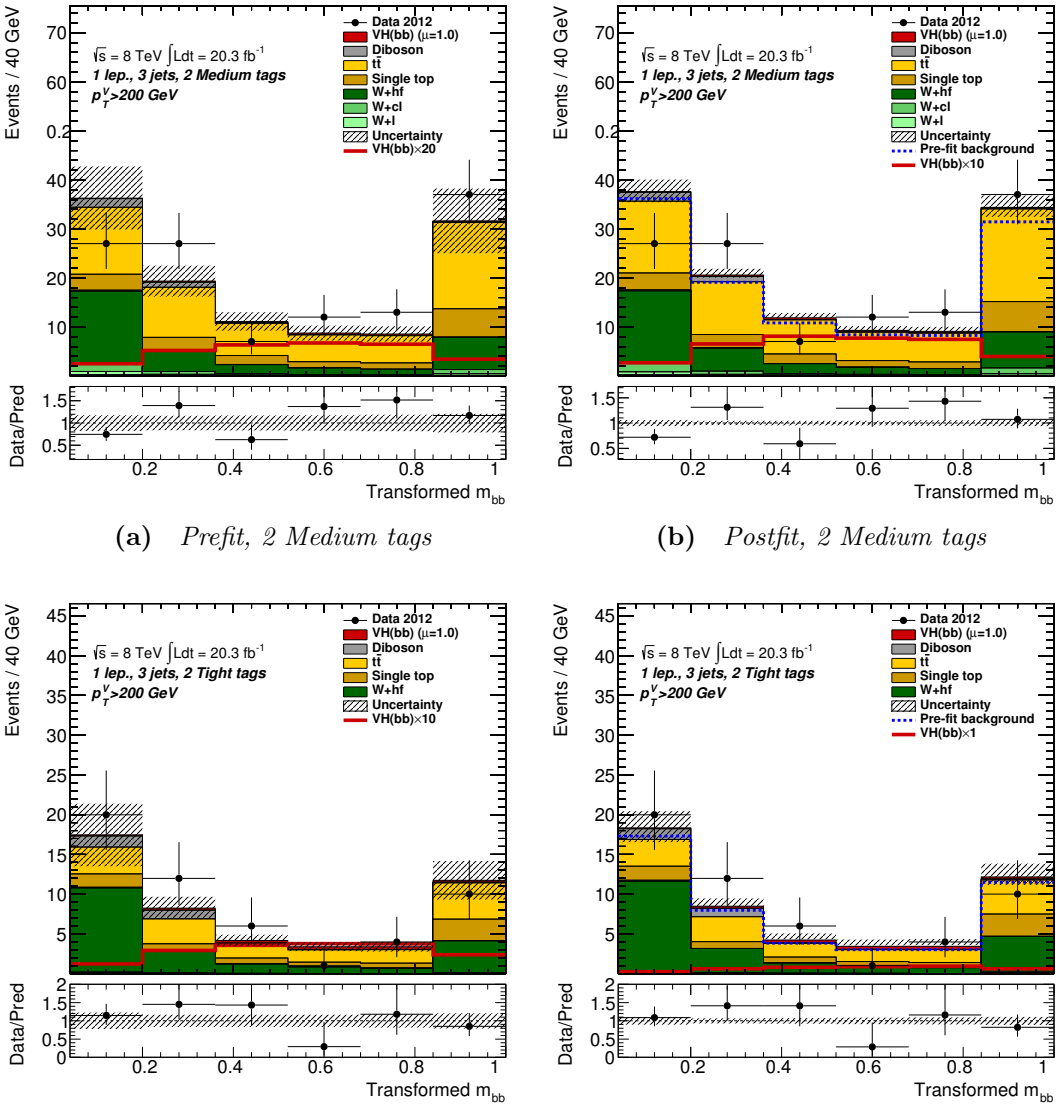
Figure D.9 *The m_{bb} distribution in the 3-jet, 2-tag, $120 < p_T^V < 160$ GeV region with two medium tagged b-jets (top row) and two tight tagged b-jets (bottom row) used as an input for the global fit in the 1-lepton channel before (left) and after (right) the fit is performed.*



(a) *Prefit, 2 Medium tags* **(b)** *Postfit, 2 Medium tags*

(c) *Prefit, 2 Tight tags* **(d)** *Postfit, 2 Tight tags*

Figure D.10 *The m_{bb} distribution in the 3-jet, 2-tag, $160 < p_T^V < 200$ GeV region with two medium tagged b-jets (top row) and two tight tagged b-jets (bottom row) used as an input for the global fit in the 1-lepton channel before (left) and after (right) the fit is performed.*



(a) *Prefit, 2 Medium tags*

(b) *Postfit, 2 Medium tags*

(c) *Prefit, 2 Tight tags*

(d) *Postfit, 2 Tight tags*

Figure D.11 *The m_{bb} distribution in the 3-jet, 2-tag, $p_T^V > 200$ GeV region with two medium tagged b-jets (top row) and two tight tagged b-jets (bottom row) used as an input for the global fit in the 1-lepton channel before (left) and after (right) the fit is performed.*

Appendix E

Event Yield Ratios

The following tables show the ratio of the yields after unconditional fits over the prefit yields for the separate lepton channels. These are further separated for the different tagging regions.

Sample	1-tag	
	$100 < p_{\mathrm{T}}^{V} < 120$ GeV	$p_{\mathrm{T}}^{V} > 120$ GeV
	2-jet	
VH	1.08	1.06
VV	1.0	1.0
$t\bar{t}$	1.39	1.43
s-top	1.3	1.11
W+l	1.04	0.99
W+cl	1.13	1.11
W+hf	0.83	0.68
Z+l	0.99	0.94
Z+cl	0.87	0.87
Z+hf	1.1	1.08
MJ	1.35	0.56
Total	1.04	0.98
	3-jet	
VH	–	1.07
VV	–	1.0
$t\bar{t}$	–	1.28
s-top	–	0.99
W+l	–	0.97
W+cl	–	1.05
W+hf	–	0.74
Z+l	–	0.88
Z+cl	–	1.03
Z+hf	–	0.9
MJ	–	1.04
Total	–	0.99

Table E.1 *Table of post unconditional fit over prefit yields for 0-lepton 2 and 3-jet 1-tag events in the dijet mass selection.*

Sample	2L-tag			
	$100 < p_{\mathrm{T}}^{V} < 120$ GeV	$120 < p_{\mathrm{T}}^{V} < 160$ GeV	$160 < p_{\mathrm{T}}^{V} < 200$ GeV	$p_{\mathrm{T}}^{V} > 200$ GeV
	2-jet			
VH	1.11	1.12	1.12	1.11
VV	1.04	1.06	1.1	1.08
$t\bar{t}$	1.42	1.48	1.41	1.41
s-top	1.27	1.1	1.05	1.0
W+l	1.01	1.0	1.01	1.02
W+cl	1.15	1.15	1.13	1.1
W+hf	0.86	0.8	0.77	0.71
Z+l	1.01	0.99	0.97	0.95
Z+cl	0.91	0.91	0.88	0.86
Z+hf	1.18	1.14	1.16	1.07
MJ	0.99	1.22	1.22	1.22
Total	1.1	1.08	1.06	1.0
	3-jet			
VH	–	1.1	1.1	1.12
VV	–	0.98	1.0	0.94
$t\bar{t}$	–	1.35	1.28	1.26
s-top	–	1.01	1.01	0.97
W+l	–	1.0	1.01	1.0
W+cl	–	1.09	1.07	1.06
W+hf	–	0.79	0.82	0.81
Z+l	–	0.91	0.92	0.89
Z+cl	–	1.06	1.07	1.02
Z+hf	–	1.05	1.01	0.96
MJ	–	0.53	0.53	0.53
Total	–	1.1	1.06	0.99

Table E.2 *Table of post unconditional fit over prefit yields for 0-lepton 2 and 3-jet 2L-tag events in the dijet mass selection.*

Sample	2M-tag			
	$100 < p_{\mathrm{T}}^{V} < 120$ GeV	$120 < p_{\mathrm{T}}^{V} < 160$ GeV	$160 < p_{\mathrm{T}}^{V} < 200$ GeV	$p_{\mathrm{T}}^{V} > 200$ GeV
	2-jet			
VH	1.11	1.12	1.12	1.13
VV	1.06	1.08	1.09	1.15
$t\bar{t}$	1.38	1.51	1.5	1.45
s-top	1.22	1.15	1.2	1.05
W+l	1.1	1.06	1.1	1.03
W+cl	1.16	1.18	1.19	1.21
W+hf	0.89	0.82	0.83	0.78
Z+l	1.08	1.07	1.05	1.01
Z+cl	0.93	0.93	0.89	0.93
Z+hf	1.15	1.15	1.12	1.12
MJ	0.99	1.22	1.22	1.22
Total	1.17	1.17	1.13	1.08
	3-jet			
VH	–	1.12	1.12	1.11
VV	–	0.96	0.99	0.97
$t\bar{t}$	–	1.34	1.32	1.37
s-top	–	1.02	1.07	1.03
W+l	–	1.13	1.09	1.09
W+cl	–	1.02	1.07	1.12
W+hf	–	0.8	0.86	0.85
Z+l	–	1.11	0.97	0.96
Z+cl	–	1.04	1.11	1.08
Z+hf	–	1.04	1.1	1.05
MJ	–	0.53	0.53	0.53
Total	–	1.16	1.13	1.07

Table E.3 *Table of post unconditional fit over prefit yields for 0-lepton 2 and 3-jet 2M-tag events in the dijet mass selection.*

Sample	2T-tag			
	$100 < p_{\mathrm{T}}^{V} < 120$ GeV	$120 < p_{\mathrm{T}}^{V} < 160$ GeV	$160 < p_{\mathrm{T}}^{V} < 200$ GeV	$p_{\mathrm{T}}^{V} > 200$ GeV
	2-jet			
VH	1.1	1.1	1.1	1.11
VV	1.04	1.08	1.09	1.1
$t\bar{t}$	1.38	1.46	1.5	1.33
s-top	1.24	1.15	1.07	1.15
W+l	1.0	1.0	1.0	1.0
W+cl	1.2	1.11	1.13	1.11
W+hf	0.83	0.89	0.82	0.81
Z+l	1.0	1.0	1.0	1.0
Z+cl	0.88	0.88	0.88	0.88
Z+hf	1.13	1.12	1.12	1.08
MJ	0.99	1.22	1.22	1.22
Total	1.18	1.17	1.12	1.04
	3-jet			
VH	–	1.1	1.11	1.1
VV	–	0.98	0.9	0.95
$t\bar{t}$	–	1.39	1.33	1.27
s-top	–	1.06	1.11	0.96
W+l	–	1.0	1.0	1.0
W+cl	–	1.11	1.11	1.11
W+hf	–	0.9	0.91	0.92
Z+l	–	1.0	1.0	1.0
Z+cl	–	0.88	0.71	0.88
Z+hf	–	1.07	1.07	1.06
MJ	–	0.53	0.53	0.53
Total	–	1.22	1.11	1.04

Table E.4 *Table of post unconditional fit over prefit yields for 0-lepton 2 and 3-jet 2T-tag events in the dijet mass selection.*

Sample	1-tag	
	$p_{\mathrm{T}}^{V} < 120$ GeV	$p_{\mathrm{T}}^{V} > 120$ GeV
	2-jet	
VH	2.18	2.14
VV	1.0	1.04
$t\bar{t}$	1.12	1.13
s-top	1.11	1.03
W+l	0.97	0.96
W+cl	1.12	1.13
W+hf	0.86	0.76
Z+l	1.01	0.96
Z+cl	1.0	1.0
Z+hf	1.0	1.0
MJ_e	–	1.34
MJ_μ	0.99	1.0
Total	1.03	1.02
	3-jet	
VH	2.21	2.16
VV	1.0	0.94
$t\bar{t}$	1.0	1.01
s-top	1.02	0.99
W+l	0.98	0.97
W+cl	1.09	1.08
W+hf	0.74	0.86
Z+l	1.02	1.0
Z+cl	1.0	1.0
Z+hf	1.0	1.0
MJ_e	–	0.8
MJ_μ	0.99	1.0
Total	0.99	0.99

Table E.5 *Table of post unconditional fit over prefit yields for 1-lepton 2 and 3-jet 1-tag events in the dijet mass selection.*

Sample	2L-tag				
	$p_{\mathrm{T}}^{V} < 90$ GeV	$90 < p_{\mathrm{T}}^{V} < 120$ GeV	$120 < p_{\mathrm{T}}^{V} < 160$ GeV	$160 < p_{\mathrm{T}}^{V} < 200$ GeV	$p_{\mathrm{T}}^{V} > 200$ GeV
	2-jet				
VH	2.26	2.26	2.26	2.21	2.32
VV	1.0	1.0	1.0	1.04	1.12
$t\bar{t}$	1.14	1.15	1.15	1.16	1.15
s-top	1.12	1.1	1.07	1.03	1.05
W+l	0.95	0.96	0.95	0.96	0.97
W+cl	1.11	1.12	1.14	1.12	1.12
W+hf	1.01	0.91	0.86	0.84	0.71
Z+l	0.99	0.97	1.0	1.0	1.0
Z+cl	1.0	1.0	1.0	1.0	1.0
Z+hf	1.0	1.0	1.0	1.0	1.0
MJ_e	–	–	0.93	0.89	0.91
MJ_μ	1.0	1.0	1.0	1.0	–
Total	1.06	1.05	1.04	1.03	0.97
	3-jet				
VH	2.23	2.35	2.24	2.42	2.23
VV	1.0	1.0	1.0	1.0	0.95
$t\bar{t}$	1.02	1.03	1.06	1.06	1.04
s-top	1.04	1.04	1.02	1.04	1.03
W+l	0.97	0.96	0.97	0.97	0.96
W+cl	1.09	1.07	1.09	1.09	1.07
W+hf	0.85	0.92	0.97	0.99	0.97
Z+l	1.0	1.0	1.0	1.0	1.0
Z+cl	1.0	1.0	1.0	1.0	1.0
Z+hf	1.0	1.0	1.0	1.0	1.0
MJ_e	–	–	1.03	1.0	1.0
MJ_μ	1.0	1.0	1.0	–	–
Total	1.01	1.02	1.04	1.05	1.02

Table E.6 *Table of post unconditional fit over prefit yields for 1-lepton 2 and 3-jet 2L-tag events in the dijet mass selection.*

Sample	2M-tag				
	$p_{\mathrm{T}}^{V} < 90$ GeV	$90 < p_{\mathrm{T}}^{V} < 120$ GeV	$120 < p_{\mathrm{T}}^{V} < 160$ GeV	$160 < p_{\mathrm{T}}^{V} < 200$ GeV	$p_{\mathrm{T}}^{V} > 200$ GeV
	2-jet				
VH	2.34	2.34	2.31	2.3	2.4
VV	1.05	1.08	1.12	1.11	1.2
$t\bar{t}$	1.15	1.17	1.18	1.18	1.19
s-top	1.14	1.12	1.1	1.09	1.1
W+l	0.99	0.95	0.96	0.96	1.0
W+cl	1.08	1.09	1.11	1.05	1.13
W+hf	1.05	0.98	0.9	0.86	0.81
Z+l	1.0	1.0	1.0	1.0	1.0
Z+cl	1.0	1.0	1.0	1.0	1.0
Z+hf	1.05	1.09	1.04	1.0	1.0
MJ_e	–	–	0.93	0.94	0.99
MJ_μ	0.84	0.84	0.84	–	–
Total	1.09	1.1	1.09	1.07	1.03
	3-jet				
VH	2.32	2.32	2.29	2.26	2.38
VV	1.0	1.0	1.0	0.93	1.05
$t\bar{t}$	1.04	1.05	1.07	1.05	1.06
s-top	1.07	1.07	1.02	1.08	1.07
W+l	1.0	1.0	1.0	1.0	0.94
W+cl	1.1	1.14	1.06	1.11	1.16
W+hf	0.9	0.93	1.04	0.97	1.05
Z+l	1.0	1.0	1.0	1.0	1.0
Z+cl	1.0	1.0	1.0	1.0	1.0
Z+hf	1.0	1.0	1.0	1.0	1.0
MJ_e	–	–	1.04	1.05	1.05
MJ_μ	1.0	1.86	–	1.0	–
Total	1.04	1.05	1.07	1.05	1.08

Table E.7 *Table of post unconditional fit over prefit yields for 1-lepton 2 and 3-jet 2M-tag events in the dijet mass selection.*

Sample	2T-tag				
	$p_{\mathrm{T}}^{V} < 90$ GeV	$90 < p_{\mathrm{T}}^{V} < 120$ GeV	$120 < p_{\mathrm{T}}^{V} < 160$ GeV	$160 < p_{\mathrm{T}}^{V} < 200$ GeV	$p_{\mathrm{T}}^{V} > 200$ GeV
	2-jet				
VH	2.3	2.28	2.29	2.29	2.35
VV	1.07	1.07	1.12	1.15	1.23
$t\bar{t}$	1.15	1.18	1.19	1.13	1.21
s-top	1.15	1.13	1.1	1.13	1.09
W+l	1.0	1.0	1.0	1.0	1.0
W+cl	1.09	1.08	1.07	0.9	0.91
W+hf	1.1	0.97	0.91	0.87	0.83
Z+l	1.0	1.0	1.0	1.0	1.0
Z+cl	1.0	1.0	1.0	1.0	–
Z+hf	1.02	1.01	1.1	1.1	0.92
MJ_e	–	–	1.0	0.81	0.92
MJ_μ	1.38	1.37	–	–	–
Total	1.15	1.13	1.13	1.08	1.05
	3-jet				
VH	2.29	2.34	2.3	2.26	2.35
VV	1.0	1.0	1.0	0.99	0.97
$t\bar{t}$	1.03	1.06	1.1	1.08	0.97
s-top	1.09	1.11	1.03	1.05	1.02
W+l	1.0	1.0	1.0	1.0	1.0
W+cl	1.13	1.13	1.13	1.13	0.97
W+hf	0.93	1.0	1.03	1.06	1.09
Z+l	1.0	1.0	1.0	1.0	1.0
Z+cl	1.0	1.0	1.0	1.0	1.0
Z+hf	1.0	1.0	1.0	0.99	1.46
MJ_e	–	–	1.03	1.07	1.05
MJ_μ	1.0	1.01	–	1.0	–
Total	1.03	1.06	1.09	1.08	1.08

Table E.8 *Table of post unconditional fit over prefit yields for 1-lepton 2 and 3-jet 2T-tag events in the dijet mass selection.*

Sample	1-tag	
	$p_{\mathrm{T}}^{V} < 120$ GeV	$p_{\mathrm{T}}^{V} > 120$ GeV
	2-jet	
VH	1.08	1.06
VV	1.0	1.08
$t\bar{t}$	1.18	1.15
s-top	1.0	1.0
W+l	1.0	1.0
W+cl	1.11	1.11
W+hf	0.84	0.84
Z+l	1.01	1.0
Z+cl	0.92	0.91
Z+hf	1.11	1.11
MJ	0.18	0.18
Total	1.01	1.02
	3-jet	
VH	1.08	1.08
VV	0.98	0.95
$t\bar{t}$	1.23	1.22
s-top	1.0	1.0
W+l	1.0	1.0
W+cl	1.11	1.11
W+hf	0.84	0.84
Z+l	0.97	0.92
Z+cl	1.13	1.07
Z+hf	0.97	0.93
MJ	0.18	0.18
Total	1.0	0.96

Table E.9 *Table of post unconditional fit over prefit yields for 2-lepton 2 and 3-jet 1-tag events in the dijet mass selection.*

Sample	2L-tag				
	$p_{\mathrm{T}}^{V} < 90$ GeV	$90 < p_{\mathrm{T}}^{V} < 120$ GeV	$120 < p_{\mathrm{T}}^{V} < 160$ GeV	$160 < p_{\mathrm{T}}^{V} < 200$ GeV	$p_{\mathrm{T}}^{V} > 200$ GeV
	2-jet				
VH	1.13	1.13	1.12	1.12	1.13
VV	1.06	1.07	1.07	1.11	1.1
$t\bar{t}$	1.21	1.18	1.28	1.19	1.15
s-top	1.0	1.0	1.0	1.0	1.04
W+l	1.0	1.0	1.0	1.0	1.0
W+cl	1.11	1.11	1.11	1.11	1.11
W+hf	0.84	0.84	0.84	0.79	0.84
Z+l	1.04	1.03	1.02	1.01	1.03
Z+cl	0.93	0.93	0.93	0.92	0.93
Z+hf	1.19	1.18	1.17	1.15	1.12
MJ	0.18	0.18	1.0	1.0	–
Total	1.09	1.1	1.1	1.08	1.06
	3-jet				
VH	1.14	1.13	1.13	1.11	1.11
VV	0.99	1.0	1.01	1.04	0.97
$t\bar{t}$	1.25	1.31	1.33	1.17	–
s-top	1.0	1.0	1.0	0.95	–
W+l	1.0	1.0	1.0	1.0	1.0
W+cl	1.11	1.11	1.11	1.11	1.11
W+hf	0.84	0.84	0.84	0.84	0.84
Z+l	0.99	0.96	0.95	0.95	0.95
Z+cl	1.18	1.11	1.1	1.09	1.07
Z+hf	1.13	1.08	1.05	1.01	1.02
MJ	0.18	1.0	0.18	–	–
Total	1.11	1.08	1.05	1.02	1.01

Table E.10 *Table of post unconditional fit over prefit yields for 2-lepton 2 and 3-jet 2L-tag events in the dijet mass selection.*

Sample	2M-tag				
	$p_{\mathrm{T}}^{V} < 90$ GeV	$90 < p_{\mathrm{T}}^{V} < 120$ GeV	$120 < p_{\mathrm{T}}^{V} < 160$ GeV	$160 < p_{\mathrm{T}}^{V} < 200$ GeV	$p_{\mathrm{T}}^{V} > 200$ GeV
	2-jet				
VH	1.13	1.13	1.12	1.11	1.13
VV	1.05	1.06	1.08	1.08	1.14
$t\bar{t}$	1.2	1.18	1.28	1.2	1.15
s-top	1.0	0.98	0.95	1.0	–
W+l	1.0	1.0	1.0	1.0	1.0
W+cl	1.11	1.11	1.11	1.11	1.11
W+hf	0.84	0.84	0.78	–	0.84
Z+l	1.12	1.11	1.1	1.09	1.06
Z+cl	0.92	0.95	0.96	0.95	0.99
Z+hf	1.17	1.15	1.14	1.13	1.14
MJ	0.18	0.18	1.0	–	–
Total	1.14	1.14	1.14	1.12	1.13
	3-jet				
VH	1.14	1.15	1.12	1.13	1.15
VV	0.97	1.04	0.96	0.94	1.0
$t\bar{t}$	1.28	1.3	1.24	1.22	1.5
s-top	1.0	1.03	1.0	–	1.0
W+l	1.0	1.0	1.0	1.0	1.0
W+cl	1.11	1.11	1.11	1.11	1.11
W+hf	0.84	0.84	0.84	0.83	0.84
Z+l	1.07	1.04	1.03	1.03	1.02
Z+cl	1.2	1.1	1.13	1.12	1.18
Z+hf	1.16	1.12	1.08	1.1	1.06
MJ	0.18	1.0	–	–	–
Total	1.17	1.15	1.09	1.09	1.06

Table E.11 *Table of post unconditional fit over prefit yields for 2-lepton 2 and 3-jet 2M-tag events in the dijet mass selection.*

Sample	2T-tag				
	$p_{\mathrm{T}}^{V} < 90$ GeV	$90 < p_{\mathrm{T}}^{V} < 120$ GeV	$120 < p_{\mathrm{T}}^{V} < 160$ GeV	$160 < p_{\mathrm{T}}^{V} < 200$ GeV	$p_{\mathrm{T}}^{V} > 200$ GeV
	2-jet				
VH	1.11	1.1	1.1	1.09	1.1
VV	1.05	1.04	1.06	1.07	1.11
$t\bar{t}$	1.19	1.19	1.25	1.28	–
s-top	1.0	1.0	1.13	1.28	1.07
W+l	1.0	–	–	–	–
W+hf	0.84	0.85	0.84	0.84	–
Z+l	1.0	1.0	1.0	1.0	1.0
Z+cl	0.88	0.88	0.94	0.88	0.88
Z+hf	1.14	1.13	1.11	1.1	1.11
MJ	0.18	0.18	0.18	–	–
Total	1.14	1.13	1.12	1.1	1.11
	3-jet				
VH	1.12	1.13	1.11	1.1	1.11
VV	0.99	0.97	0.98	0.94	0.98
$t\bar{t}$	1.27	1.26	1.28	1.15	1.29
s-top	1.0	1.0	0.92	1.0	–
W+l	1.0	–	–	–	–
W+hf	0.84	–	0.84	0.84	–
Z+l	1.0	1.0	1.0	1.0	1.0
Z+cl	0.88	0.88	0.88	0.88	0.88
Z+hf	1.15	1.11	1.09	1.07	1.06
MJ	0.18	1.0	–	–	–
Total	1.18	1.14	1.1	1.06	1.05

Table E.12 *Table of post unconditional fit over prefit yields for 2-lepton 2 and 3-jet 2T-tag events in the dijet mass selection.*

Appendix F

Event Yields in $WH \to \ell\nu b\bar{b}$

The following tables contain the event yields for the 1-lepton channel both pre and post-fit. This is separated into the different b-tagging regions: 1-tag, 2 loose/medium/tight.

Sample	1-tag	
p_{T}^{V} region (GeV)	$p_{\mathrm{T}}^{V} < 120$	$p_{\mathrm{T}}^{V} > 120$
	2-jet	
$Z + l$	2926.8	455.4
$Z + cl$	1193.3	151.2
$Z + hf$	1448.9	181.2
$W + l$	48 093.7	10 421.4
$W + cl$	49 714.6	8865.1
$W + hf$	16 338.4	3920.3
s-top	8418.6	1543.2
MJ_{μ}	6607.5	55.5
MJ_{e}	0.0	853.8
$t\bar{t}$	7827.7	3919.5
VV	1672.9	649.9
Total Background	144242.3 ± 11487.82	31016.3 ± 2745.90
Total Signal	34.4 ± 3.80	19.6 ± 2.30
Data	148390.0	31777.0
	3-jet	
$Z + l$	1324.9	188.8
$Z + cl$	588.2	79.0
$Z + hf$	690.4	88.5
$W + l$	18 013.9	4672.3
$W + cl$	17 295.8	3749.5
$W + hf$	7240.0	1915.7
s-top	5534.8	1433.0
MJ_{μ}	3241.0	25.3
MJ_{e}	0.0	381.9
$t\bar{t}$	15 652.4	6917.2
VV	894.2	356.4
Total Background	70475.6 ± 6147.76	19807.6 ± 2153.23
Total Signal	12.9 ± 2.02	9.1 ± 1.14
Data	69956.0	19649.0

Table F.1 *Number of events obtained before the global likelihood fit in the 1-lepton 2- and 3-jet 1 b-tag regions, with a Signal of* $m_H = 125$ GeV. *The uncertainties are from the prefit NPs except the floating normalisations.*

Sample	2L-tag				
p_{T}^{V} region (GeV)	$p_{\mathrm{T}}^{V} < 90$	$90 < p_{\mathrm{T}}^{V} < 120$	$120 < p_{\mathrm{T}}^{V} < 160$	$160 < p_{\mathrm{T}}^{V} < 200$	$p_{\mathrm{T}}^{V} > 200$
	2-jet				
$Z + l$	55.6	20.4	10.3	1.7	0.7
$Z + cl$	35.0	11.9	5.8	1.1	0.4
$Z + hf$	81.9	28.7	12.8	2.4	1.0
$W + l$	908.4	253.7	196.1	56.0	47.2
$W + cl$	1404.2	364.7	263.1	68.5	43.7
$W + hf$	1077.6	316.6	269.6	80.1	77.7
s-top	573.7	138.6	106.6	26.9	16.6
MJ_{μ}	457.0	54.1	5.1	1.0	0.0
MJ_{e}	0.0	0.0	45.5	10.2	2.1
$t\bar{t}$	871.8	334.4	361.4	94.6	52.3
VV	68.9	22.1	25.0	9.6	10.3
Total Background	5534.0 ± 513.13	1545.2 ± 139.32	1301.3 ± 121.95	351.9 ± 37.80	252.0 ± 37.06
Total Signal	7.4 ± 0.92	2.6 ± 0.32	3.7 ± 0.44	1.8 ± 0.22	2.1 ± 0.31
Data	5891.0	1570.0	1355.0	375.0	239.0
	3-jet				
$Z + l$	24.5	6.8	3.7	0.8	0.5
$Z + cl$	17.9	4.6	2.5	0.6	0.3
$Z + hf$	41.9	9.7	6.8	1.2	0.7
$W + l$	336.5	83.4	75.6	24.9	24.8
$W + cl$	502.8	118.9	101.3	31.4	25.7
$W + hf$	480.9	113.3	108.8	40.2	45.9
s-top	389.8	94.3	85.4	28.7	23.9
MJ_{μ}	138.3	10.8	1.2	0.0	0.0
MJ_{e}	0.0	0.0	11.9	2.5	0.8
$t\bar{t}$	1938.2	537.2	533.4	162.9	105.7
VV	31.3	9.7	11.5	5.2	5.7
Total Background	3902.1 ± 477.92	988.5 ± 130.27	942.3 ± 135.03	298.3 ± 43.80	233.9 ± 36.84
Total Signal	2.1 ± 0.36	0.8 ± 0.14	1.3 ± 0.15	0.7 ± 0.11	1.0 ± 0.17
Data	3882.0	1004.0	1051.0	313.0	238.0

Table F.2 *Number of events obtained before the global likelihood fit in the 1-lepton 2- and 3-jet 2L b-tag regions, with a Signal of* $m_H = 125$ GeV*. The uncertainties are from the prefit NPs except the floating normalisations.*

Sample	2M-tag				
p_{T}^V region (GeV)	$p_{\mathrm{T}}^V < 90$	$90 < p_{\mathrm{T}}^V < 120$	$120 < p_{\mathrm{T}}^V < 160$	$160 < p_{\mathrm{T}}^V < 200$	$p_{\mathrm{T}}^V > 200$
	2-jet				
$Z + l$	4.3	1.5	0.7	0.1	0.0
$Z + cl$	7.3	2.5	1.2	0.2	0.0
$Z + hf$	63.5	21.4	9.8	1.5	0.7
$W + l$	66.4	18.7	13.4	3.6	3.0
$W + cl$	243.1	63.6	44.8	11.0	6.5
$W + hf$	702.0	213.3	183.0	54.1	51.0
s-top	521.8	123.3	88.8	19.1	10.5
MJ_μ	242.4	19.1	1.4	0.0	0.0
MJ_e	0.0	0.0	26.8	4.9	1.2
$t\bar{t}$	1073.5	403.0	387.1	72.4	29.1
VV	44.2	14.0	16.6	6.6	5.7
Total Background	2968.6 ± 251.23	880.4 ± 66.21	773.5 ± 64.67	173.5 ± 18.21	107.8 ± 18.62
Total Signal	11.6 ± 1.36	3.9 ± 0.46	5.9 ± 0.67	2.9 ± 0.34	3.3 ± 0.48
Data	3230.0	984.0	868.0	199.0	116.0
	3-jet				
$Z + l$	1.6	0.5	0.3	0.1	0.0
$Z + cl$	3.0	0.8	0.4	0.1	0.0
$Z + hf$	30.0	8.0	5.0	1.2	0.5
$W + l$	22.6	5.8	5.1	1.5	1.4
$W + cl$	83.5	19.7	17.0	4.8	3.6
$W + hf$	290.4	73.4	77.1	27.6	30.0
s-top	371.0	83.5	70.0	19.5	16.0
MJ_μ	70.5	6.6	0.0	0.6	0.0
MJ_e	0.0	0.0	5.8	2.1	0.4
$t\bar{t}$	2484.8	642.3	547.4	119.9	59.1
VV	14.5	4.9	6.3	3.0	3.2
Total Background	3371.9 ± 563.78	845.4 ± 143.26	734.4 ± 129.13	180.3 ± 29.62	114.2 ± 19.64
Total Signal	3.2 ± 0.52	1.3 ± 0.22	2.1 ± 0.25	1.3 ± 0.17	1.5 ± 0.24
Data	3519.0	936.0	780.0	181.0	123.0

Table F.3 *Number of events obtained before the global likelihood fit in the 1-lepton 2- and 3-jet 2M b-tag regions, with a Signal of* $m_H = 125$ GeV. *The uncertainties are from the prefit NPs except the floating normalisations.*

Sample	2T-tag				
p_{T}^V region (GeV)	$p_{\mathrm{T}}^V < 90$	$90 < p_{\mathrm{T}}^V < 120$	$120 < p_{\mathrm{T}}^V < 160$	$160 < p_{\mathrm{T}}^V < 200$	$p_{\mathrm{T}}^V > 200$
	2-jet				
$Z + l$	0.1	0.0	0.0	0.0	0.0
$Z + cl$	0.4	0.1	0.1	0.0	0.0
$Z + hf$	42.6	15.1	6.2	1.5	0.5
$W + l$	2.0	0.4	0.3	0.1	0.1
$W + cl$	12.6	3.2	2.1	0.5	0.4
$W + hf$	372.1	119.0	117.2	34.4	33.2
s-top	424.2	94.2	66.1	11.7	4.3
MJ_μ	99.1	9.4	0.0	0.0	0.0
MJ_e	0.0	0.0	11.6	2.4	1.2
$t\bar{t}$	1019.9	367.2	316.5	39.2	8.5
VV	34.6	11.6	13.8	5.9	5.0
Total Background	2007.4 ± 184.13	620.1 ± 50.31	533.8 ± 50.25	95.8 ± 11.89	53.1 ± 11.49
Total Signal	13.1 ± 1.62	4.4 ± 0.55	6.8 ± 0.80	3.6 ± 0.45	3.7 ± 0.55
Data	2364.0	700.0	591.0	112.0	59.0
	3-jet				
$Z + l$	0.0	0.0	0.0	0.0	0.0
$Z + cl$	0.2	0.0	0.0	0.0	0.0
$Z + hf$	22.2	6.1	3.7	0.9	0.3
$W + l$	0.7	0.2	0.1	0.0	0.0
$W + cl$	3.9	1.0	0.9	0.2	0.2
$W + hf$	153.1	41.7	46.9	18.6	20.0
s-top	309.2	64.8	48.8	10.0	7.1
MJ_μ	35.5	1.9	0.0	0.3	0.0
MJ_e	0.0	0.0	2.3	0.7	0.3
$t\bar{t}$	2551.9	607.8	443.2	57.0	16.1
VV	9.2	3.4	5.8	2.8	2.7
Total Background	3085.8 ± 586.93	726.8 ± 138.12	551.9 ± 108.52	90.6 ± 16.54	46.7 ± 8.72
Total Signal	3.6 ± 0.61	1.4 ± 0.23	2.4 ± 0.31	1.5 ± 0.19	1.8 ± 0.29
Data	3161.0	779.0	590.0	101.0	53.0

Table F.4 *Number of events obtained before the global likelihood fit in the 1-lepton 2- and 3-jet 2T b-tag regions, with a Signal of* $m_H = 125$ GeV. *The uncertainties are from the prefit NPs except the floating normalisations.*

Sample	1-tag	
p_{T}^{V} region (GeV)	$p_{\mathrm{T}}^{V} < 120$	$p_{\mathrm{T}}^{V} > 120$
	2-jet	
$Z + l$	2954.2	435.4
$Z + cl$	1192.0	151.1
$Z + hf$	1454.5	181.0
$W + l$	46 616.2	10 045.2
$W + cl$	55 794.8	10 012.5
$W + hf$	14 001.6	2983.9
s-top	9310.9	1585.4
MJ_{μ}	6525.3	55.5
MJ_{e}	0.0	1148.2
$t\bar{t}$	8806.0	4434.4
VV	1671.1	675.6
Total Background	148326.6 ± 408.00	31708.2 ± 199.06
Total Signal	75.0 ± 30.21	41.9 ± 17.05
Data	148 390.0	31 777.0
	3-jet	
$Z + l$	1356.5	188.6
$Z + cl$	587.6	78.9
$Z + hf$	689.7	88.4
$W + l$	17 680.2	4519.5
$W + cl$	18 774.6	4065.9
$W + hf$	5370.7	1642.1
s-top	5634.8	1411.6
MJ_{μ}	3219.8	25.3
MJ_{e}	0.0	304.3
$t\bar{t}$	15 718.8	7016.4
VV	893.3	335.4
Total Background	69925.9 ± 270.26	19676.4 ± 132.27
Total Signal	28.6 ± 11.76	19.6 ± 8.04
Data	69 956.0	19 649.0

Table F.5 *Number of events obtained after performing the unconditional fit in the 1-lepton 2- and 3-jet 1 b-tag regions, with a Signal of* $m_H = 125$ GeV. *The uncertainties are the full postfit errors including all NPs with priors, floating normalisations, and correlations.*

Sample	2L-tag				
p_{T}^{V} region (GeV)	$p_{\mathrm{T}}^{V} < 90$	$90 < p_{\mathrm{T}}^{V} < 120$	$120 < p_{\mathrm{T}}^{V} < 160$	$160 < p_{\mathrm{T}}^{V} < 200$	$p_{\mathrm{T}}^{V} > 200$
	2-jet				
$Z + l$	54.8	19.9	10.3	1.7	0.7
$Z + cl$	35.0	11.9	5.8	1.1	0.4
$Z + hf$	81.8	28.7	12.7	2.4	1.0
$W + l$	866.7	242.6	186.7	53.7	45.7
$W + cl$	1562.3	410.0	300.1	76.7	48.9
$W + hf$	1085.2	288.5	231.6	67.6	55.5
s-top	639.9	153.1	113.6	27.7	17.3
MJ_{μ}	456.6	54.2	5.1	1.0	0.0
MJ_{e}	0.0	0.0	42.3	9.0	1.9
$t\bar{t}$	996.2	384.8	416.9	109.3	60.0
VV	68.8	22.1	25.0	10.0	11.5
Total Background	5847.3 ± 68.90	1615.6 ± 26.10	1350.1 ± 21.23	360.3 ± 7.27	242.8 ± 8.58
Total Signal	16.6 ± 6.65	5.8 ± 2.33	8.4 ± 3.37	3.8 ± 1.55	4.8 ± 1.91
Data	5891.0	1570.0	1355.0	375.0	239.0
	3-jet				
$Z + l$	24.5	6.8	3.7	0.8	0.5
$Z + cl$	17.9	4.6	2.5	0.6	0.3
$Z + hf$	41.8	9.7	6.8	1.2	0.7
$W + l$	327.2	80.4	73.5	24.1	23.8
$W + cl$	549.7	126.6	110.7	34.1	27.4
$W + hf$	408.0	104.3	105.5	39.9	44.4
s-top	406.5	98.3	86.8	30.0	24.5
MJ_{μ}	138.4	10.8	1.2	0.0	0.0
MJ_{e}	0.0	0.0	12.3	2.5	0.8
$t\bar{t}$	1974.8	553.0	566.2	173.0	110.0
VV	31.3	9.6	11.5	5.2	5.4
Total Background	3919.9 ± 43.70	1004.0 ± 15.94	980.8 ± 16.1	311.1 ± 6.79	237.8 ± 8.72
Total Signal	4.8 ± 1.96	1.8 ± 0.75	2.9 ± 1.16	1.8 ± 0.71	2.3 ± 0.93
Data	3882.0	1004.0	1051.0	313.0	238.0

Table F.6 *Number of events obtained after performing the unconditional fit in the 1-lepton 2- and 3-jet 2L b-tag regions, with a Signal of* $m_H = 125$ GeV. *The uncertainties are the full postfit errors including all NPs with priors, floating normalisations, and correlations.*

Sample	2M-tag				
p_{T}^V region (GeV)	$p_{\mathrm{T}}^V < 90$	$90 < p_{\mathrm{T}}^V < 120$	$120 < p_{\mathrm{T}}^V < 160$	$160 < p_{\mathrm{T}}^V < 200$	$p_{\mathrm{T}}^V > 200$
	2-jet				
$Z + l$	4.3	1.5	0.7	0.1	0.0
$Z + cl$	7.3	2.4	1.2	0.2	0.0
$Z + hf$	66.4	23.4	10.1	1.5	0.7
$W + l$	65.5	17.8	12.9	3.5	3.0
$W + cl$	262.6	69.5	49.7	11.5	7.3
$W + hf$	735.0	208.4	164.2	46.3	41.4
s-top	597.3	138.0	97.7	20.8	11.5
MJ_μ	204.1	16.0	1.2	0.0	0.0
MJ_e	0.0	0.0	24.9	4.6	1.2
$t\bar{t}$	1236.0	470.6	456.6	85.4	34.7
VV	46.5	15.2	18.6	7.3	6.9
Total Background	3224.9 ± 50.34	962.8 ± 17.46	837.7 ± 17.73	181.3 ± 4.96	106.7 ± 5.56
Total Signal	27.1 ± 10.76	9.0 ± 3.57	13.7 ± 5.40	6.7 ± 2.62	7.9 ± 3.12
Data	3230.0	984.0	868.0	199.0	116.0
	3-jet				
$Z + l$	1.6	0.5	0.3	0.1	0.0
$Z + cl$	3.0	0.8	0.4	0.1	0.0
$Z + hf$	29.9	8.0	5.0	1.2	0.5
$W + l$	22.6	5.8	5.1	1.5	1.3
$W + cl$	92.2	22.4	18.0	5.4	4.2
$W + hf$	261.8	67.9	80.0	26.8	31.5
s-top	397.6	89.0	71.3	21.0	17.0
MJ_μ	70.6	12.3	0.0	0.6	0.0
MJ_e	0.0	0.0	6.0	2.2	0.4
$t\bar{t}$	2593.3	673.4	587.6	126.4	62.9
VV	14.5	4.9	6.3	2.8	3.3
Total Background	3487.1 ± 37.91	885.0 ± 16.48	779.9 ± 14.91	188.0 ± 4.80	121.2 ± 5.69
Total Signal	7.3 ± 2.99	3.0 ± 1.22	4.8 ± 1.93	3.0 ± 1.20	3.6 ± 1.45
Data	3519.0	936.0	780.0	181.0	123.0

Table F.7 *Number of events obtained after performing the unconditional fit in the 1-lepton 2- and 3-jet 2M b-tag regions, with a Signal of* $m_H = 125$ GeV. *The uncertainties are the full postfit errors including all NPs with priors, floating normalisations, and correlations.*

Sample	2T-tag				
p_{T}^V region (GeV)	$p_{\mathrm{T}}^V < 90$	$90 < p_{\mathrm{T}}^V < 120$	$120 < p_{\mathrm{T}}^V < 160$	$160 < p_{\mathrm{T}}^V < 200$	$p_{\mathrm{T}}^V > 200$
	2-jet				
$Z + l$	0.1	0.0	0.0	0.0	0.0
$Z + cl$	0.4	0.1	0.1	0.0	0.0
$Z + hf$	43.4	15.3	6.8	1.7	0.5
$W + l$	2.0	0.4	0.3	0.1	0.1
$W + cl$	13.8	3.4	2.3	0.5	0.3
$W + hf$	407.6	116.0	106.3	30.0	27.6
s-top	486.2	106.5	72.9	13.3	4.7
MJ_μ	136.6	12.8	0.0	0.0	0.0
MJ_e	0.0	0.0	11.6	2.0	1.1
$t\bar{t}$	1172.9	431.8	377.9	44.5	10.3
VV	36.9	12.4	15.4	6.8	6.1
Total Background	2299.9 ± 49.76	698.7 ± 15.50	593.6 ± 15.17	98.6 ± 3.61	50.7 ± 4.16
Total Signal	30.1 ± 11.82	10.0 ± 3.91	15.6 ± 6.08	8.3 ± 3.24	8.7 ± 3.37
Data	2364.0	700.0	591.0	112.0	59.0
	3-jet				
$Z + l$	0.0	0.0	0.0	0.0	0.0
$Z + cl$	0.2	0.0	0.0	0.0	0.0
$Z + hf$	22.2	6.1	3.7	0.9	0.4
$W + l$	0.7	0.2	0.1	0.0	0.0
$W + cl$	4.5	1.1	1.0	0.3	0.1
$W + hf$	142.7	41.6	48.5	19.7	21.9
s-top	336.0	71.7	50.5	10.5	7.3
MJ_μ	35.5	1.9	0.0	0.3	0.0
MJ_e	0.0	0.0	2.4	0.7	0.3
$t\bar{t}$	2638.0	642.7	485.4	61.3	15.6
VV	9.2	3.4	5.8	2.8	2.7
Total Background	3188.8 ± 42.55	768.7 ± 15.65	597.5 ± 14.79	96.5 ± 3.77	48.3 ± 4.08
Total Signal	8.3 ± 3.37	3.3 ± 1.32	5.6 ± 2.20	3.4 ± 1.35	4.1 ± 1.63
Data	3161.0	779.0	590.0	101.0	53.0

Table F.8 *Number of events obtained after performing the unconditional fit in the 1-lepton 2- and 3-jet 2T b-tag regions, with a Signal of* $m_H = 125$ GeV. *The uncertainties are the full postfit errors including all NPs with priors, floating normalisations, and correlations.*

Appendix G

Substructure Variable Definitions

Additional definitions are given here for substructure variables used in Chapter 7.

G.1 Thrust

The jet direction in the centre-of-mass frame defines the thrust axis [226–229]. This is calculated by finding the direction of the unit vector $\hat{n}_{\mathrm{T}}$ which maximises the sum of the momenta of all of the jet constituents: $T = \max \frac{\sum_i |\vec{p}_i \cdot \hat{n}_{\mathrm{T}}|}{\sum_i |\vec{p}_i|}$, which here is in the longitudinal direction.

The thrust major axis is perpendicular to the thrust axis and is found by maximising the following equation, given the restriction $\hat{n}_{maj} \cdot \hat{n}_{\mathrm{T}} = 0$.

$$T_{maj} = \max \frac{\sum_i |\vec{p}_i \cdot \hat{n}_{maj}|}{\sum_i |\vec{p}_i|}, \tag{G.1}$$

The change in definition is such that the projected energy of the most energetic jet onto the major axis in maximised.

The thrust minor axis T_{min} is perpendicular to both the thrust axis and the thrust major axis. It is found by maximising the following equation subject to the restrictions $\hat{n}_{min} \cdot \hat{n}_{maj} = 0$ and $\hat{n}_{min} \cdot \hat{n}_{\mathrm{T}} = 0$.

$$T_{min} = \frac{\sum_i |\vec{p}_i \cdot \hat{n}_{min}|}{\sum_i |\vec{p}_i|}, \tag{G.2}$$

where $\vec{p}_i$ is the four vector of jet constituent i, and $\hat{T}$ is the thrust axis. For a balanced two prong decay this should have a value of T_{min} close to 0, which indicates that the constituents are highly directional. Increasing values of T_{min} indicate a more isotropic distribution.

The second axis, the "Major" axis, is taken to be perpendicular to the Thrust axis; but with the requirement that the projected energy of the most energetic jet onto the Major axis in is maximized [4,5]. Meaning if I took the dot product between the Major axis and the direction of the most energetic jet, this dot product would always be maximum (but still keep the Major axis and the Thrust axis perpendicular). This additional requirement needs to be specified so that the Major axis is unique (there are an infinite number of perpendicular directions to a given direction).

The third axis, called the "Minor" axis, is then perpendicular to these two. However, it turns out that energy flow along this direction is very close to the minimum energy flow along any axis [4,5].

G.2 Fox-Wolfram Moments

The Fox-Wolfram Moments [230], H_l, are defined as

$$H_l = \sum_{i,j} \frac{|\vec{p_i}||\vec{p_j}|}{E^2} P_l(\cos\theta_{ij}), \tag{G.3}$$

where θ_{ij} is the opening angle between two energy clusters i and j (the jet constituents), E is the total jet energy in the jet rest frame, and $P_l(x)$ are the Legendre polynomials.

Each energy cluster is assumed to be a massless pseudoparticle, such that $H_0 = 1$. Back-to-back subjets in the rest frame give $H_1 = 0$, $H_l \approx 1$ for even l and $H_l \approx 0$ for odd l. In the studies here, the ratio of the second and zeroth-order moments is used. In the text this is refered to as FoxWolfram20.

G.3 Dipolarity

Dipolarity [260] is a measure of the colour flow between two hard cores in a jet. Dipolarity takes as input a jet J, and the two subjets, j_1 and j_2. These are used to define the dipolarity D_{12} as

$$\mathrm{D}_{12} = \frac{1}{\Delta R_{12}^2} \sum_{i \in J} \frac{p_{\mathrm{T}i}}{p_{\mathrm{T}j}} \Delta R_{J,i}^2, \tag{G.4}$$

Where ΔR_{12} is the angle between the two subjets j_1 and j_2, and $\Delta R_{a_1,i}$ is the angle between subjet j_i and the centre of the jet J.

Dipolarity is at a minimum when the majority of the radiation from J is between j_1 and j_2, and is at a maximum whenever a large amount of radiation is elsewhere. The ΔR weighting ensures that D_{12} receives large contributions from any softer radiation which is away from j_1 and j_2, reflecting the colour configuration of J. Colour singlets that decay into two jets are expected to have a smaller dipolarity than coloured objects.

Appendix H

Boosted Boson Tagging Decision Functions and Statistics

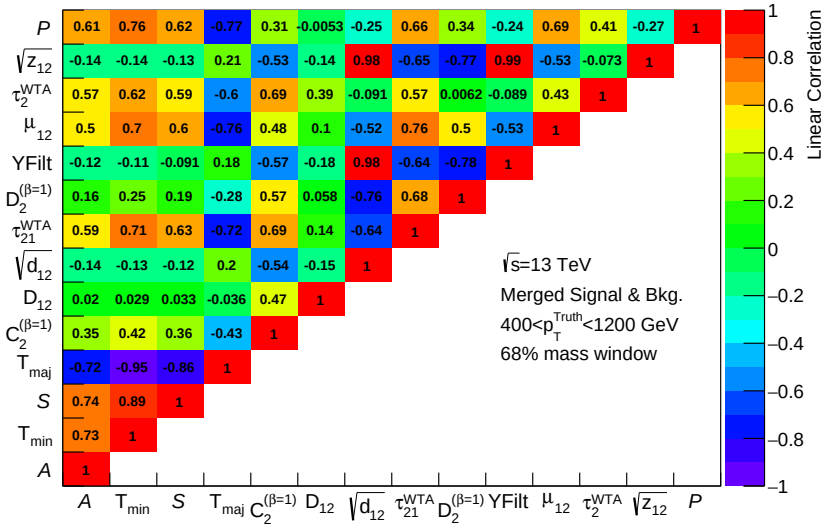


Figure H.1 *The matrix plots show the linear correlation coefficients between the class label (signal or background) and the possible feature inputs for the BDT and DNNs. There is a 68% mass window cut on the groomed jet mass. Here the correlations are shown for combined signal and background datasets where the leading ungroomed C/A jet has* $400 < p_T^{Truth} < 1200$ GeV. *The correlation coefficients for signal and background are shown separately in Section 7.7.1.2.*

H.1 Training Folds Statistics

$C_2^{(\beta=1)}$

Fold	Sig+Bkg Mean	Sig+Bkg Std	Sig Mean	Sig Std	Bkg Mean	Bkg Std
Full	0.0955	0.0338	0.0671	0.0269	0.1042	0.0308
Train cv 009	0.0950	0.0337	0.0676	0.0260	0.1044	0.0308
Valid cv 009	0.0953	0.0339	0.0678	0.0262	0.1047	0.0310
Train cv 007	0.0950	0.0338	0.0675	0.0260	0.1044	0.0309
Valid cv 007	0.0953	0.0338	0.0679	0.0262	0.1046	0.0309
Train cv 003	0.0950	0.0338	0.0676	0.0261	0.1044	0.0309
Valid cv 003	0.0952	0.0337	0.0677	0.0258	0.1046	0.0308
Train cv 000	0.0950	0.0338	0.0675	0.0260	0.1044	0.0308
Valid cv 000	0.0952	0.0338	0.0681	0.0262	0.1044	0.0310
Train cv 002	0.0951	0.0338	0.0677	0.0261	0.1044	0.0308
Valid cv 002	0.0950	0.0338	0.0675	0.0259	0.1044	0.0309
Train cv 005	0.0951	0.0338	0.0676	0.0261	0.1045	0.0309
Valid cv 005	0.0950	0.0337	0.0677	0.0261	0.1044	0.0309
Train cv 006	0.0951	0.0337	0.0676	0.0260	0.1044	0.0308
Valid cv 006	0.0951	0.0338	0.0677	0.0261	0.1045	0.0309
Train cv 001	0.0950	0.0338	0.0676	0.0261	0.1044	0.0309
Valid cv 001	0.0952	0.0337	0.0678	0.0260	0.1046	0.0308
Train cv 008	0.0951	0.0338	0.0676	0.0261	0.1045	0.0309
Valid cv 008	0.0949	0.0337	0.0676	0.0260	0.1042	0.0308
Train cv 004	0.0951	0.0338	0.0677	0.0260	0.1044	0.0309
Valid cv 004	0.0950	0.0337	0.0675	0.0261	0.1045	0.0307

Table H.1 *Statistics for $C_2^{(\beta=1)}$ in the complete dataset after event selection, and the training and validation samples used for the boosted decision trees. For the neural network these are standardised such that they have mean of 0 and a standard deviation of 1.*

$D_2^{(\beta=1)}$

Fold	Sig+Bkg Mean	Sig+Bkg Std	Sig Mean	Sig Std	Bkg Mean	Bkg Std
Full	1.840	0.7006	1.297	0.5344	2.007	0.6595
Train cv 009	1.823	0.717	1.298	0.5177	2.003	0.6866
Valid cv 009	1.828	0.719	1.302	0.5270	2.008	0.6866
Train cv 007	1.825	0.718	1.298	0.5196	2.005	0.6866
Valid cv 007	1.824	0.717	1.300	0.5211	2.003	0.6866
Train cv 003	1.825	0.719	1.297	0.5196	2.006	0.6882
Valid cv 003	1.823	0.713	1.303	0.5213	2.000	0.6817
Train cv 000	1.826	0.718	1.298	0.5204	2.006	0.6868
Valid cv 000	1.822	0.715	1.302	0.5188	1.999	0.6858
Train cv 002	1.824	0.716	1.299	0.5199	2.003	0.6852
Valid cv 002	1.827	0.721	1.299	0.5203	2.007	0.6907
Train cv 005	1.825	0.718	1.299	0.5201	2.005	0.6872
Valid cv 005	1.822	0.716	1.298	0.5196	2.001	0.6846
Train cv 006	1.825	0.717	1.298	0.5203	2.004	0.6862
Valid cv 006	1.825	0.718	1.301	0.5192	2.004	0.6879
Train cv 001	1.824	0.718	1.298	0.5209	2.004	0.6871
Valid cv 001	1.826	0.716	1.300	0.5172	2.005	0.6851
Train cv 008	1.826	0.718	1.301	0.5213	2.006	0.6869
Valid cv 008	1.819	0.716	1.292	0.5162	1.999	0.6856
Train cv 004	1.825	0.717	1.301	0.5224	2.004	0.6866
Valid cv 004	1.823	0.717	1.292	0.5128	2.004	0.6867

Table H.2 *Statistics for $D_2^{(\beta=1)}$ in the complete dataset after event selection, and the training and validation samples used for the boosted decision trees. For the neural network these are standardised such that they have mean of 0 and a standard deviation of 1.*

$\sqrt{d_{12}}$

Fold	Sig+Bkg Mean	Sig+Bkg Std	Sig Mean	Sig Std	Bkg Mean	Bkg Std
Full	38.86	14.84	49.6110	14.57	35.54	13.26
Train cv 009	39.25	15.12	49.66	14.48	35.70	13.62
Valid cv 009	39.17	15.15	49.67	14.70	35.58	13.55
Train cv 007	39.23	15.14	49.67	14.54	35.67	13.61
Valid cv 007	39.23	15.08	49.62	14.51	35.68	13.55
Train cv 003	39.20	15.12	49.63	14.51	35.65	13.61
Valid cv 003	39.32	15.13	49.76	14.61	35.75	13.57
Train cv 000	39.24	15.13	49.69	14.53	35.67	13.61
Valid cv 000	39.22	15.10	49.56	14.54	35.68	13.58
Train cv 002	39.23	15.13	49.67	14.55	35.67	13.60
Valid cv 002	39.23	15.10	49.62	14.46	35.68	13.60
Train cv 005	39.22	15.11	49.62	14.51	35.67	13.60
Valid cv 005	39.27	15.16	49.77	14.59	35.68	13.60
Train cv 006	39.22	15.12	49.64	14.52	35.67	13.59
Valid cv 006	39.27	15.15	49.72	14.58	35.69	13.62
Train cv 001	39.26	15.13	49.66	14.55	35.71	13.61
Valid cv 001	39.15	15.10	49.66	14.46	35.56	13.56
Train cv 008	39.21	15.13	49.63	14.56	35.65	13.61
Valid cv 008	39.31	15.09	49.76	14.45	35.74	13.58
Train cv 004	39.25	15.12	49.65	14.53	35.69	13.60
Valid cv 004	39.20	15.14	49.71	14.55	35.61	13.58

Table H.3 *Statistics for $\sqrt{d_{12}}$ in the complete dataset after event selection, and the training and validation samples used for the boosted decision trees. For the neural network these are standardised such that they have mean of 0 and a standard deviation of 1.*

τ_{21}^{WTA}

Fold	Sig+Bkg Mean	Sig+Bkg Std	Sig Mean	Sig Std	Bkg Mean	Bkg Std
Full	0.4796	0.1744	0.3286	0.1524	0.5263	0.1531
Train cv 009	0.4657	0.1711	0.3271	0.1461	0.5130	0.1522
Valid cv 009	0.4669	0.1711	0.3283	0.1471	0.5143	0.1519
Train cv 007	0.4659	0.1711	0.3270	0.1460	0.5133	0.1522
Valid cv 007	0.4663	0.1710	0.3285	0.1474	0.5133	0.1521
Train cv 003	0.4658	0.1711	0.3268	0.1458	0.5133	0.1522
Valid cv 003	0.4666	0.1709	0.3293	0.1479	0.5134	0.1519
Train cv 000	0.4660	0.1713	0.3270	0.1463	0.5135	0.1523
Valid cv 000	0.4660	0.1705	0.3288	0.1465	0.5129	0.1517
Train cv 002	0.4659	0.1711	0.3274	0.1463	0.5132	0.1523
Valid cv 002	0.4663	0.1710	0.3275	0.1465	0.5137	0.1518
Train cv 005	0.4661	0.1710	0.3276	0.1466	0.5134	0.1520
Valid cv 005	0.4655	0.1712	0.3269	0.1458	0.5129	0.1525
Train cv 006	0.4660	0.1711	0.3269	0.1462	0.5134	0.1521
Valid cv 006	0.4661	0.1710	0.3289	0.1469	0.5129	0.1524
Train cv 001	0.4658	0.1712	0.3271	0.1467	0.5131	0.1522
Valid cv 001	0.4666	0.1706	0.3285	0.1454	0.5138	0.1520
Train cv 008	0.4660	0.1710	0.3273	0.1461	0.5134	0.1521
Valid cv 008	0.4659	0.1712	0.3277	0.1471	0.5132	0.1523
Train cv 004	0.4658	0.1708	0.3280	0.1464	0.5129	0.1520
Valid cv 004	0.4664	0.1718	0.3258	0.1463	0.5145	0.1524

Table H.4 *Statistics for τ_{21}^{WTA} in the complete dataset after event selection, and the training and validation samples used for the boosted decision trees. For the neural network these are standardised such that they have mean of 0 and a standard deviation of 1.*

P

Fold	Sig+Bkg Mean	Sig+Bkg Std	Sig Mean	Sig Std	Bkg Mean	Bkg Std
Full	0.5614	0.2622	0.4036	0.2276	0.6103	0.2528
Train cv 009	0.5293	0.2488	0.3978	0.2175	0.5743	0.2429
Valid cv 009	0.5299	0.2488	0.3984	0.2170	0.5748	0.2430
Train cv 007	0.5292	0.2488	0.3973	0.2169	0.5742	0.2430
Valid cv 007	0.5304	0.2489	0.3997	0.2187	0.5750	0.2427
Train cv 003	0.5288	0.2488	0.3970	0.2172	0.5739	0.2429
Valid cv 003	0.5314	0.2489	0.4006	0.2179	0.5760	0.2430
Train cv 000	0.5294	0.2490	0.3976	0.2173	0.5744	0.2431
Valid cv 000	0.5297	0.2484	0.3989	0.2176	0.5743	0.2424
Train cv 002	0.5292	0.2487	0.3980	0.2178	0.5739	0.2427
Valid cv 002	0.5305	0.2492	0.3978	0.2160	0.5758	0.2436
Train cv 005	0.5296	0.2488	0.3972	0.2173	0.5748	0.2427
Valid cv 005	0.5291	0.2489	0.4002	0.2176	0.5731	0.2435
Train cv 006	0.5295	0.2489	0.3980	0.2175	0.5745	0.2429
Valid cv 006	0.5293	0.2487	0.3977	0.2168	0.5742	0.2429
Train cv 001	0.5293	0.2488	0.3978	0.2173	0.5742	0.2429
Valid cv 001	0.5299	0.2488	0.3983	0.2176	0.5749	0.2428
Train cv 008	0.5294	0.2488	0.3983	0.2174	0.5742	0.2430
Valid cv 008	0.5297	0.2489	0.3967	0.2171	0.5751	0.2428
Train cv 004	0.5294	0.2486	0.3986	0.2173	0.5741	0.2429
Valid cv 004	0.5297	0.2494	0.3960	0.2176	0.5754	0.2431

Table H.5 *Statistics for planar flow in the complete dataset after event selection, and the training and validation samples used for the boosted decision trees. For the neural network these are standardised such that they have mean of 0 and a standard deviation of 1.*

A

Fold	Sig+Bkg Mean	Sig+Bkg Std	Sig Mean	Sig Std	Bkg Mean	Bkg Std
Full	0.0539	0.0562	0.0249	0.0311	0.0629	0.0592
Train cv 009	0.0433	0.0423	0.0233	0.0266	0.0502	0.0445
Valid cv 009	0.0436	0.0425	0.0230	0.0263	0.0506	0.0447
Train cv 007	0.0433	0.0424	0.0231	0.0265	0.0502	0.0446
Valid cv 007	0.0436	0.0422	0.0235	0.0265	0.0505	0.0443
Train cv 003	0.0434	0.0424	0.0230	0.0264	0.0503	0.0446
Valid cv 003	0.0435	0.0423	0.0237	0.0269	0.0503	0.0444
Train cv 000	0.0434	0.0424	0.0233	0.0267	0.0503	0.0445
Valid cv 000	0.0434	0.0424	0.0231	0.0259	0.0504	0.0446
Train cv 002	0.0434	0.0423	0.0232	0.0264	0.0502	0.0444
Valid cv 002	0.0435	0.0427	0.0232	0.0268	0.0505	0.0449
Train cv 005	0.0434	0.0424	0.0232	0.0264	0.0504	0.0445
Valid cv 005	0.0433	0.0423	0.0233	0.0268	0.0501	0.0444
Train cv 006	0.0434	0.0423	0.0232	0.0265	0.0503	0.0445
Valid cv 006	0.0434	0.0425	0.0233	0.0264	0.0503	0.0447
Train cv 001	0.0434	0.0424	0.0231	0.0265	0.0503	0.0445
Valid cv 001	0.0435	0.0424	0.0234	0.0266	0.0504	0.0445
Train cv 008	0.0434	0.0423	0.0232	0.0265	0.0502	0.0445
Valid cv 008	0.0436	0.0425	0.0231	0.0265	0.0505	0.0447
Train cv 004	0.0433	0.0423	0.0233	0.0265	0.0502	0.0444
Valid cv 004	0.0436	0.0427	0.0230	0.0264	0.0506	0.0448

Table H.6 *Statistics for aplanarity in the complete dataset after event selection, and the training and validation samples used for the boosted decision trees. For the neural network these are standardised such that they have mean of 0 and a standard deviation of 1.*

Sample	Signal	Background	Total
Full	57525	185902	243427
Train cv 009	40478	118540	159018
Valid cv 009	13493	39514	53007
Train cv 007	40478	118540	159018
Valid cv 007	13493	39514	53007
Train cv 003	40478	118540	159018
Valid cv 003	13493	39514	53007
Train cv 000	40478	118540	159018
Valid cv 000	13493	39514	53007
Train cv 002	40478	118540	159018
Valid cv 002	13493	39514	53007
Train cv 005	40478	118540	159018
Valid cv 005	13493	39514	53007
Train cv 006	40478	118540	159018
Valid cv 006	13493	39514	53007
Train cv 001	40478	118540	159018
Valid cv 001	13493	39514	53007
Train cv 008	40478	118540	159018
Valid cv 008	13493	39514	53007
Train cv 004	40478	118540	159018
Valid cv 004	13493	39514	53007

Table H.7 *Number of events in the complete dataset after event selection, and the training and validation samples used for the boosted decision tree training. Stratified 10-fold cross validation is used here in the region* $400 < p_T < 1200$ *GeV.*

$C_2^{(\beta=1)}$

Fold	Sig+Bkg Mean	Sig+Bkg Std	Sig Mean	Sig Std	Bkg Mean	Bkg Std
Full	0.0000	1.000	-0.1113	0.9845	0.8352	1.000
Train cv 000	0.000	1.000	-0.1053	0.9938	0.8366	1.000
Valid cv 000	0.001	1.002	-0.1341	0.9476	0.0398	0.7890
Train cv 001	0.000	1.000	-0.1145	0.9914	0.9234	1.034
Valid cv 001	0.001	1.000	-0.0790	0.9640	0.1792	0.8369
Train cv 002	0.000	1.000	-0.1125	0.9796	0.8411	1.002
Valid cv 002	0.006	0.9972	-0.0778	1.011	1.017	1.059
Train cv 003	0.000	1.000	-0.1146	0.9869	0.8344	1.001
Valid cv 003	0.002	0.9971	-0.1016	0.9690	-0.0622	0.7571
Train cv 004	0.000	1.000	-0.0873	0.9780	0.7104	0.9735
Valid cv 004	-0.010	1.004	-0.1275	1.041	1.046	1.066

Table H.8 *Statistics for $C_2^{(\beta=1)}$ in the complete dataset after event selection, and the training and validation samples used for the deep neural networks. For the neural network these are standardised such that they have mean of 0 and a standard deviation of 1.*

$D_2^{(\beta=1)}$

Fold	Sig+Bkg Mean	Sig+Bkg Std	Sig Mean	Sig Std	Bkg Mean	Bkg Std
Full	0.0000	1.000	-0.9575	0.6272	-0.1721	0.8049
Train cv 000	0.0000	1.000	-0.9557	0.6281	-0.1721	0.8032
Valid cv 000	0.0020	0.9978	-0.9608	0.6220	-0.1696	0.8098
Train cv 001	0.0000	1.000	-0.9586	0.6316	-0.1692	0.8050
Valid cv 001	-0.0040	1.001	-0.9587	0.6107	-0.1872	0.8053
Train cv 002	0.0000	1.000	-0.9561	0.6303	-0.1682	0.8081
Valid cv 002	-0.0073	0.9931	-0.9639	0.6101	-0.1931	0.7862
Train cv 003	0.0000	1.000	-0.9573	0.6284	-0.1749	0.8020
Valid cv 003	0.0025	0.9950	-0.9513	0.6191	-0.1571	0.8122
Train cv 004	0.0000	1.000	-0.9600	0.6174	-0.1758	0.8060
Valid cv 004	0.0068	1.013	-0.9530	0.6722	-0.1523	0.8105

Table H.9 *Statistics for $D_2^{(\beta=1)}$ in the complete dataset after event selection, and the training and validation samples used for the deep neural networks. For the neural network these are standardised such that they have mean of 0 and a standard deviation of 1.*

$\sqrt{d_{12}}$

Fold	Sig+Bkg Mean	Sig+Bkg Std	Sig Mean	Sig Std	Bkg Mean	Bkg Std
Full	0.0000	1.000	0.6807	0.9433	-0.1435	0.9042
Train cv 000	0.0000	1.000	0.6823	0.9473	-0.1476	0.9014
Valid cv 000	-0.0025	1.000	0.6719	0.9270	-0.1290	0.9155
Train cv 001	-0.0000	1.000	0.6771	0.9423	-0.1447	0.9028
Valid cv 001	0.0061	1.004	0.7031	0.9503	-0.1332	0.9127
Train cv 002	0.0000	1.000	0.6734	0.9435	-0.1433	0.9075
Valid cv 002	0.0035	0.9971	0.7119	0.9391	-0.1404	0.8884
Train cv 003	0.0000	1.000	0.6790	0.9426	-0.1414	0.9031
Valid cv 003	-0.0002	0.9943	0.6835	0.9409	-0.1512	0.9032
Train cv 004	0.0000	1.000	0.6918	0.9407	-0.1403	0.9060
Valid cv 004	-0.0070	1.005	0.6328	0.9565	-0.1638	0.9005

Table H.10 *Statistics for $\sqrt{d_{12}}$ in the complete dataset after event selection, and the training and validation samples used for the deep neural networks. For the neural network these are standardised such that they have mean of 0 and a standard deviation of 1.*

τ_{21}^{WTA}

Fold	Sig+Bkg Mean	Sig+Bkg Std	Sig Mean	Sig Std	Bkg Mean	Bkg Std
Full	0.0000	1.0000	-0.9446	0.8179	-0.1060	0.8525
Train cv 000	0.0000	1.000	-0.9392	0.8264	-0.1064	0.8529
Valid cv 000	-0.0030	0.9997	-0.9649	0.7813	-0.1620	0.8078
Train cv 001	0.0000	1.000	-0.9479	0.8198	-0.1043	0.8480
Valid cv 001	-0.0017	0.9989	-0.9295	0.8105	-0.1688	0.8252
Train cv 002	0.0000	1.000	-0.9472	0.8113	-0.1000	0.8537
Valid cv 002	0.0072	0.9987	-0.9215	0.8433	-0.1775	0.8047
Train cv 003	0.0000	1.000	-0.9477	0.8153	-0.1054	0.8547
Valid cv 003	0.0019	0.9983	-0.9196	0.8302	-0.2079	0.7873
Train cv 004	0.0000	1.000	-0.9374	0.8174	-0.1686	0.8105
Valid cv 004	-0.0044	1.004	-0.9663	0.8273	-0.0251	0.8669

Table H.11 *Statistics for τ_{21}^{WTA} in the complete dataset after event selection, and the training and validation samples used for the deep neural networks. For the neural network these are standardised such that they have mean of 0 and a standard deviation of 1.*

P

Fold	Sig+Bkg Mean	Sig+Bkg Std	Sig Mean	Sig Std	Bkg Mean	Bkg Std
Full	0.000	1.000	-0.9735	0.7723	-0.2879	0.9815
Train cv 000	0.0000	1.000	-0.9704	0.7748	-0.2882	0.9817
Valid cv 000	0.0015	1.000	-0.9855	0.7630	-0.2854	0.9815
Train cv 001	0.0000	1.000	-0.9771	0.7756	-0.2868	0.9790
Valid cv 001	-0.0002	0.9993	-0.9595	0.7584	-0.2926	0.9907
Train cv 002	0.0000	1.000	-0.9806	0.7602	-0.2857	0.9825
Valid cv 002	0.0062	0.9940	-0.9320	0.8134	-0.2890	0.9721
Train cv 003	0.0000	1.000	-0.9734	0.7745	-0.2856	0.9829
Valid cv 003	-0.0024	1.002	-0.9786	0.7653	-0.3005	0.9778
Train cv 004	-0.0000	1.000	-0.9659	0.7762	-0.2934	0.9815
Valid cv 004	-0.0051	1.004	-1.0128	0.7587	-0.2718	0.9846

Table H.12 *Statistics for planar flow in the complete dataset after event selection, and the training and validation samples used for the deep neural networks. For the neural network these are standardised such that they have mean of 0 and a standard deviation of 1.*

A

Fold	Sig+Bkg Mean	Sig+Bkg Std	Sig Mean	Sig Std	Bkg Mean	Bkg Std
Full	0.0000	1.000	-0.5998	0.5160	0.0244	1.039
Train cv 000	0.0000	1.000	-0.5955	0.5221	0.0196	1.033
Valid cv 000	-0.0080	0.9891	-0.6187	0.4833	0.0355	1.045
Train cv 001	0.0000	1.000	-0.6019	0.5167	0.0218	1.035
Valid cv 001	-0.0017	1.002	-0.5941	0.5140	0.0328	1.058
Train cv 002	0.0000	1.000	-0.6012	0.5140	0.0308	1.047
Valid cv 002	0.0098	1.003	-0.5860	0.5251	0.0090	1.010
Train cv 003	0.0000	1.000	-0.6022	0.5090	0.0280	1.043
Valid cv 003	0.0062	1.004	-0.5862	0.5448	0.0161	1.028
Train cv 004	0.0000	1.000	-0.5981	0.5178	0.0218	1.037
Valid cv 004	-0.0063	1.002	-0.6142	0.5095	0.0286	1.050

Table H.13 *Statistics for aplanarity in the complete dataset after event selection, and the training and validation samples used for the deep neural networks. For the neural network these are standardised such that they have mean of 0 and a standard deviation of 1.*

Sample	Signal	Background	Total
Full	53971	158054	212025
Train cv 000	43176	126443	169619
Valid cv 000	10795	31611	42406
Train cv 001	43177	126443	169620
Valid cv 001	10794	31611	42405
Train cv 002	43177	126443	169620
Valid cv 002	10794	31611	42405
Train cv 003	43177	126443	169620
Valid cv 003	10794	31611	42405
Train cv 004	43177	126444	169621
Valid cv 004	10794	31610	42404

Table H.14 *Number of events in the complete dataset after event selection, and the training and validation samples for the deep neural network. Stratified 5-fold validation is used here in the* $400 < p_T < 1200$ *GeV range.*

Appendix I

BDT Results Check

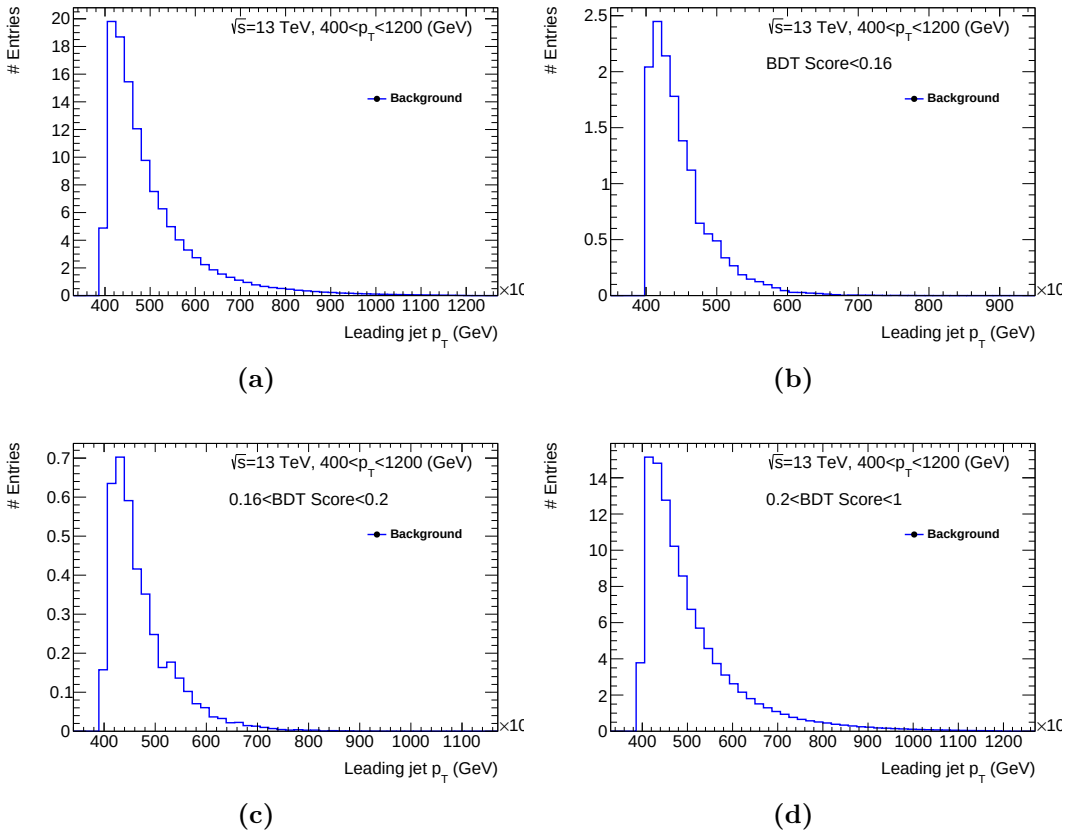


Figure I.1 *Shown here are the leading groomed jet p_T distributions in the background for different values of the BDT output for configuration 31 in Table 7.7. The full background p_T spectrum is shown in (a). In (b), a cut of BDT score$\leq$ 0.16 is applied. In plots (c) and (d) the p_T spectrum is shown for $0.16 < BDT\ Score \leq 0.2$ and $0.2 < BDT\ Score \leq 1$, respectively.*

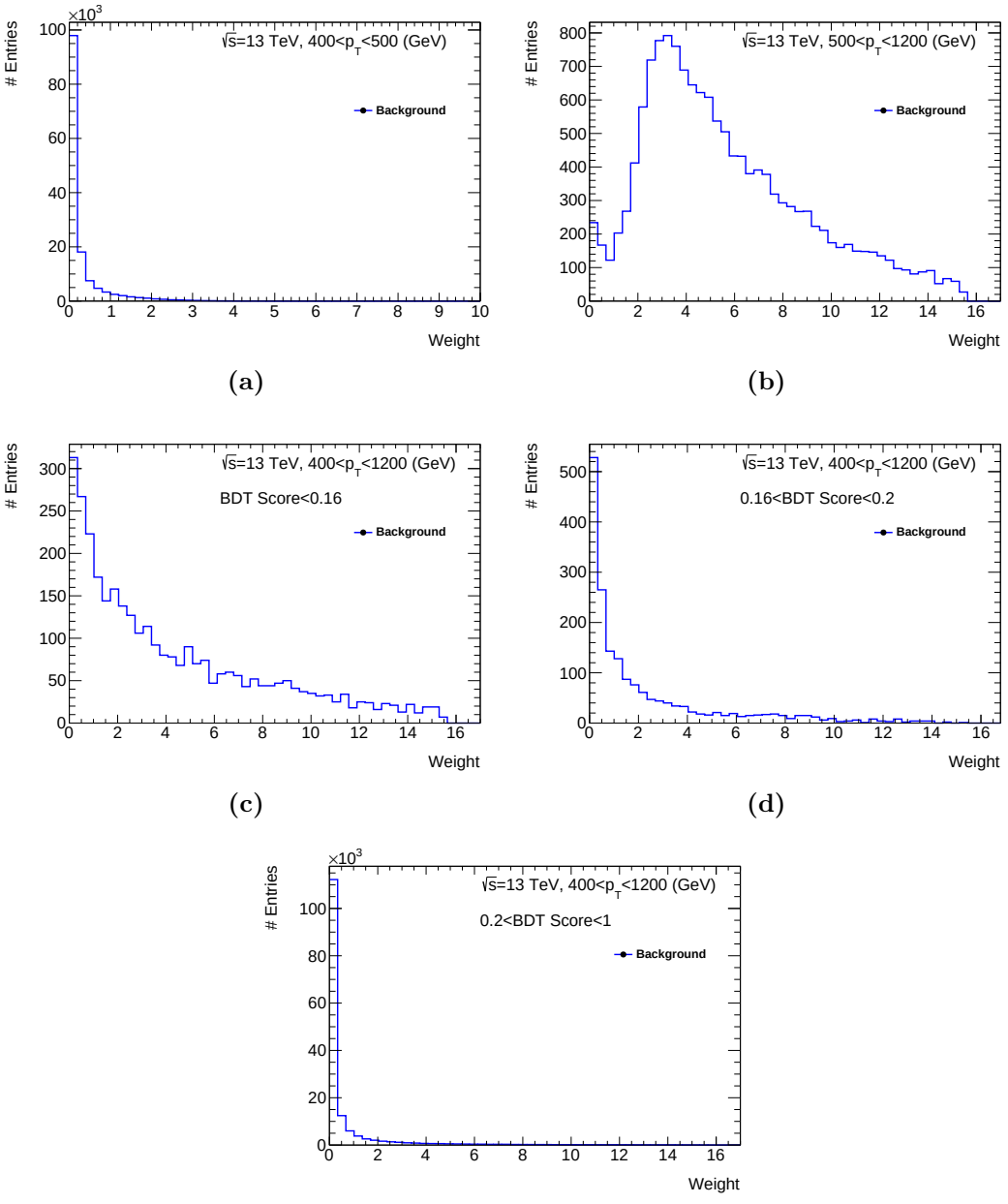
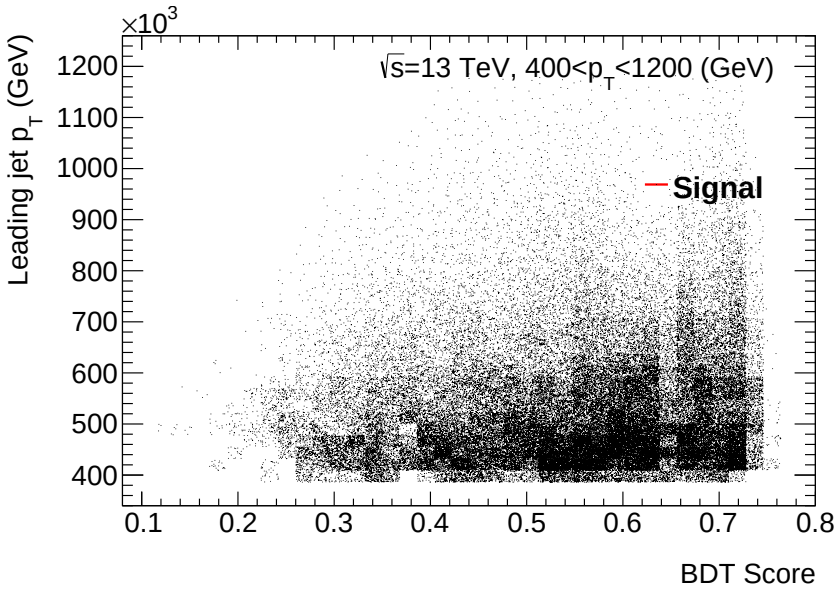
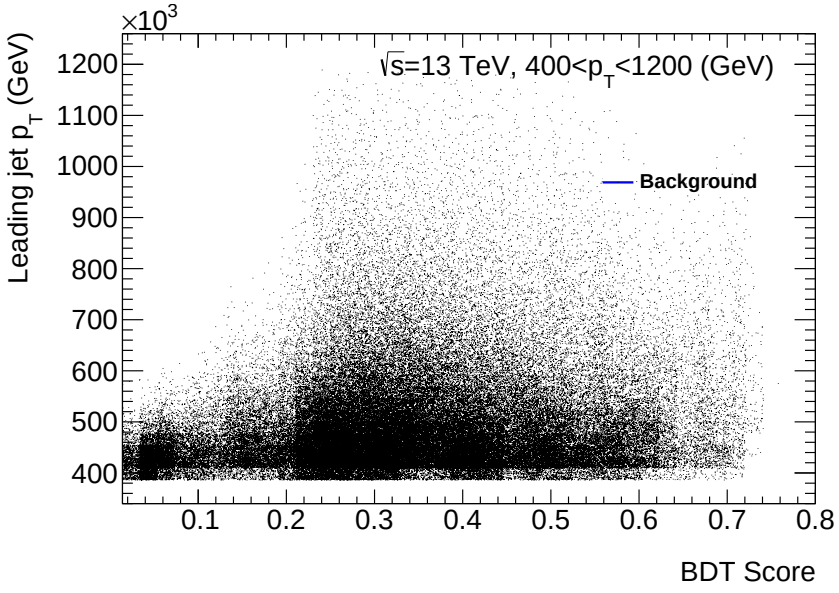


Figure I.2 *Shown here are the weight distributions in the background for different values of the BDT output for configuration 31 in Table 7.7. The weight distribution in (a) and (b) are for events where the leading groomed jet has* $p_T < 500$ *and* $p_T > 500$ GeV*, respectively. In (c), a cut of BDT score*≤ 0.16 *is applied. In plots (c) and (d) the weight distribution is shown for* $0.16 < BDT\ Score \leq 0.2$ *and* $0.2 < BDT\ Score \leq 1$*, respectively.*



(a)



(b)

Figure I.3 *The relationship between the assigned BDT class probability, using configuration 31 from Table 7.7, and the leading groomed jet p_T. Plot (a) shows the signal distribution, and (b) shows the background distribution.*

Bibliography

[1] ATLAS Collaboration, *Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector*, JHEP **01** (2015) p. 069, arXiv: 1409.6212 [hep-ex].

[2] ATLAS Collaboration, *Identification of boosted, hadronically decaying W bosons and comparisons with ATLAS data taken at $\sqrt{s} = 8$ TeV*, Eur. Phys. J. **C76** (2016) p. 154, arXiv: 1510.05821 [hep-ex].

[3] R. Farazmand, *Poorly Drawn Lines: Boundaries*, 2016, URL: {http://poorlydrawnlines.com/comic/boundaries/}.

[4] D. Adams, *The Restaurant at the End of the Universe*, The Hitchhiker's Guide to the Galaxy, Pan Macmillan, 1980.

[5] R. Feynman, R. Leighton, and M. Sands, ""Atoms in Motion: Matter is made of atoms"", *The Feynman Lectures on Physics*, Second, vol. 1, Boston: Addison-Wesley, 1963, chap. 1 p. 1.

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

[7] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (21 1967) p. 1264, URL: http://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[8] A. Salam and J. C. Ward, *Electromagnetic and weak interactions*, Physics Letters **13** (Nov. 1964) p. 168.

[9] L. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation **3** (2008) S08001, URL: http://stacks.iop.org/1748-0221/3/i=08/a=S08001.

[10] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation **3** (2008) S08003, URL: http://stacks.iop.org/1748-0221/3/i=08/a=S08003.

[11] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (9 1964) p. 321, URL: http://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[12] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12** (1964) p. 132.

[13] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) p. 508.

[14] P. W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (1966) p. 1156.

[15] CMS Collaboration, *The CMS experiment at the CERN LHC*, Journal of Instrumentation **3** (2008) S08004, URL: http://stacks.iop.org/1748-0221/3/i=08/a=S08004.

[16] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Physics Letters B **716** (2012) p. 1 , ISSN: 0370-2693, URL: http://www.sciencedirect.com/science/article/pii/S037026931200857X.

[17] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. **B716** (2012) p. 30, arXiv: 1207.7235 [hep-ex].

[18] ATLAS Collaboration, *Evidence for the spin-0 nature of the Higgs boson using ATLAS data*, Phys. Lett. **B726** (2013) p. 120, arXiv: 1307.1432 [hep-ex].

[19] D. Pappadopulo et al., *Heavy Vector Triplets: Bridging Theory and Data*, JHEP **09** (2014) p. 060, arXiv: 1402.4431 [hep-ph].

[20] Particle Data Group, *Review of Particle Physics*, Chin. Phys. **C38** (2014) p. 090001.

[21] G. E. Hinton, S. Osindero, and Y.-W. Teh, *A Fast Learning Algorithm for Deep Belief Nets*, Neural Comput. **18** (July 2006) p. 1527, ISSN: 0899-7667, URL: http://dx.doi.org/10.1162/neco.2006.18.7.1527.

[22] G. E. Hinton and R. R. Salakhutdinov, *Reducing the Dimensionality of Data with Neural Networks*, Science **313** (2006) p. 504, ISSN: 0036-8075, eprint: http://science.sciencemag.org/content/313/5786/504.full.pdf, URL: http://science.sciencemag.org/content/313/5786/504.

[23] A. Ng, *CS294a Deep Learning and Unsupervised Feature Learning: Lecture Notes*, Stanford University, California, USA, 2011, URL: https://web.stanford.edu/class/cs294a/.

[24] L. Breiman et al., *Classification and Regression Trees*, CA: Wadsworth International Group (1984).

[25] Y. Freund and R. E. Schapire, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, Journal of Computer and System Sciences **55** (1997) p. 119 , ISSN: 0022-0000, URL: http://www.sciencedirect.com/science/article/pii/S002200009791504X.

[26] L. Carroll, *Alice's Adventures in Wonderland*, 1865.

[27] W. Disney, *Alice in Wonderland*, 1951.

[28] E. Noether, *Invariante Variationsprobleme*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse **1918** (1918) p. 235, URL: http://eudml.org/doc/59024.

[29] N. Byers, "E. Noether's discovery of the deep connection between symmetries and conservation laws", *Symposium on the Heritage of Emmy Noether Ramat-Gan, Israel, December 2-4, 1996*, 1998, arXiv: physics/9807044 [physics], URL: http://alice.cern.ch/format/showfull?sysnb=0289470.

[30] F. Halzen and A. D. Martin, *Quarks and Leptons*, Wiley, 1985.

[31] J. Griffiths, *Introduction to Elementary Particle Physics*, Wiley VCH, Aug. 2008.

[32] M. Thomson, *Modern Particle Physics*, New York: Cambridge University Press, 2013, ISBN: 9781107034266, URL: http://www-spires.fnal.gov/spires/find/books/www?cl=QC793.2.T46::2013.

[33] F. J. Dyson, *The Radiation Theories of Tomonaga, Schwinger, and Feynman*, Phys. Rev. **75** (3 1949) p. 486, URL: http://link.aps.org/doi/10.1103/PhysRev.75.486.

[34] G. 't Hooft, *Renormalization of Massless Yang-Mills Fields*, Nucl. Phys. **B33** (1971) p. 173.

[35] G. 't Hooft, *Renormalizable Lagrangians for Massive Yang-Mills Fields*, Nucl. Phys. **B35** (1971) p. 167.

[36] E. Fermi, *Versuch einer Theorie der β-Strahlen. I*, Zeitschrift fur Physik **88** (Mar. 1934) p. 161.

[37] F. L. Wilson, *Fermi's Theory of Beta Decay*, American Journal of Physics **36** (Dec. 1968) p. 1150.

[38] T. D. Lee and C. N. Yang, *Parity Nonconservation and a Two-Component Theory of the Neutrino*, Phys. Rev. **105** (5 1957) p. 1671, URL: http://link.aps.org/doi/10.1103/PhysRev.105.1671.

[39] C. S. Wu et al., *Experimental Test of Parity Conservation in Beta Decay*, Phys. Rev. **105** (4 1957) p. 1413, URL: http://link.aps.org/doi/10.1103/PhysRev.105.1413.

[40] SLD Electroweak Group, DELPHI, ALEPH, SLD, SLD Heavy Flavour Group, OPAL, LEP Electroweak Working Group, L3, *Precision electroweak measurements on the Z resonance*, Phys. Rept. **427** (2006) p. 257, arXiv: hep-ex/0509008 [hep-ex].

[41] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. **C680519** (1968) p. 367.

[42] J. Howard and C. Hays, "Measurements of the Higgs Boson in the $H \to \tau\tau$ Decay Channel", Presented 27 Mar 2015, PhD thesis: Oxford U., 2015, URL: https://cds.cern.ch/record/2025641.

[43] Super-Kamiokande Collaboration, *Evidence for Oscillation of Atmospheric Neutrinos*, Phys. Rev. Lett. **81** (8 1998) p. 1562, URL: http://link.aps.org/doi/10.1103/PhysRevLett.81.1562.

[44] SNO Collaboration, *Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **87** (7 2001) p. 071301, URL: http://link.aps.org/doi/10.1103/PhysRevLett.87.071301.

[45] SNO Collaboration, *Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **89** (1 2002) p. 011301, URL: http://link.aps.org/doi/10.1103/PhysRevLett.89.011301.

[46] Z. Maki, M. Nakagawa, and S. Sakata, *Remarks on the Unified Model of Elementary Particles*, Progress of Theoretical Physics **28** (Nov. 1962) p. 870.

[47] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (12 1963) p. 531, URL: http://link.aps.org/doi/10.1103/PhysRevLett.10.531.

[48] M. Kobayashi and T. Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Progress of Theoretical Physics **49** (1973) p. 652, eprint: http://ptp.oxfordjournals.org/content/49/2/652.full.pdf+html, URL: http://ptp.oxfordjournals.org/content/49/2/652.abstract.

[49] N. Jarosik et al., *Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Sky Maps, Systematic Errors, and Basic Results*, Astrophysical Journal Supplement Series **192**, 14 (Feb. 2011) p. 14, arXiv: 1001.4744.

[50] Planck Collaboration, *Planck 2013 results. I. Overview of products and scientific results*, Astron. Astrophys. **571** (2014) A1, arXiv: 1303.5062 [astro-ph.CO].

[51] ALEPH, DELPHI, L3, OPAL Collaborations, LEP Working Group for Higgs Boson Searches, *Search for the Standard Model Higgs boson at LEP*, Physics Letters B **565** (2003) p. 61 , ISSN: 0370-2693, URL: http://www.sciencedirect.com/science/article/pii/S0370269303006142.

[52] CDF Collaboration and D0 Collaboration, *Higgs boson studies at the Tevatron*, Phys. Rev. D **88** (5 2013) p. 052014, URL: http://link.aps.org/doi/10.1103/PhysRevD.88.052014.

[53] J. Baglio and A. Djouadi, *Higgs production at the lHC*, JHEP **1103** (2011) p. 055, arXiv: 1012.0530 [hep-ph].

[54] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*, tech. rep. arXiv:1307.1347. CERN-2013-004, 2013, URL: https://cds.cern.ch/record/1559921.

[55] ATLAS and CMS Collaborations, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at* $\sqrt{s}$ *= 7 and 8 TeV*, tech. rep. ATLAS-CONF-2015-044, CERN, 2015, URL: https://cds.cern.ch/record/2052552.

[56] ATLAS Collaboration, *Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector*, JHEP **04** (2015) p. 117, arXiv: 1501.04943 [hep-ex].

[57] CDF Collaboration and D0 Collaboration, *Evidence for a Particle Produced in Association with Weak Bosons and Decaying to a Bottom-Antibottom Quark Pair in Higgs Boson Searches at the Tevatron*, Phys. Rev. Lett. **109** (7 2012) p. 071804, URL: http://link.aps.org/doi/10.1103/PhysRevLett.109.071804.

[58] ATLAS Collaboration, *Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC*, Physics Letters B **726** (2013) p. 88 , ISSN: 0370-2693, URL: http://www.sciencedirect.com/science/article/pii/S0370269313006369.

[59] ATLAS Collaboration, *Observation and measurement of Higgs boson decays to WW* with the ATLAS detector*, Phys. Rev. **D92** (2015) p. 012006, arXiv: 1412.2641 [hep-ex].

[60] UA1 Collaboration, *Experimental observation of isolated large transverse energy electrons with associated missing energy at* $\sqrt{s}$*=540 GeV*, Physics Letters B **122** (1983) p. 103 , ISSN: 0370-2693, URL: http://www.sciencedirect.com/science/article/pii/0370269383911772.

[61] UA2 Collaboration, *Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider*, Physics Letters B **122** (1983) p. 476 , ISSN: 0370-2693, URL: http://www.sciencedirect.com/science/article/pii/0370269383916052.

[62] M. H. Seymour, *Searches for new particles using cone and cluster jet algorithms: a comparative study*, Zeitschrift für Physik C Particles and Fields **62** (1994) p. 127, ISSN: 1431-5858, URL: http://dx.doi.org/10.1007/BF01559532.

[63] Paradox Wikis, Meneth, *Skylines Wiki: Monuments*, http://www.skylineswiki.com/Monuments, Accessed: 23/05/2016.

[64] Colossal Order, *Cities: Skylines*, Steam, 2015.

[65] M. Lamont, *Status of the LHC*, Journal of Physics: Conference Series **455** (2013) p. 012001, URL: http://stacks.iop.org/1742-6596/455/i=1/a=012001.

[66] LHCb Collaboration, *The LHCb Detector at the LHC*, Journal of Instrumentation **3** (2008) S08005, URL: http://stacks.iop.org/1748-0221/3/i=08/a=S08005.

[67] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, Journal of Instrumentation **3** (2008) S08002, URL: http://stacks.iop.org/1748-0221/3/i=08/a=S08002.

[68] TOTEM Collaboration, *The TOTEM Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation **3** (2008) S08007, URL: http://stacks.iop.org/1748-0221/3/i=08/a=S08007.

[69] MoEDAL Collaboration, *Technical Design Report of the MoEDAL Experiment*, tech. rep. CERN-LHCC-2009-006. MoEDAL-TDR-001, CERN, 2009, URL: https://cds.cern.ch/record/1181486.

[70] C Lefevre, "LHC: the guide (English version). Guide du LHC (version anglaise)", 2009, URL: https://cds.cern.ch/record/1165534.

[71] J. Pequenao, "Computer generated image of the whole ATLAS detector", 2008, URL: https://cds.cern.ch/record/1095924.

[72] Benjamin Wynne, Personal communication.

[73] J. Pequenao, "Computer generated image of the ATLAS inner detector", 2008, URL: https://cds.cern.ch/record/1095926.

[74] S. M. Gibson, *Monitoring radiation damage in the ATLAS Silicon Tracker*, tech. rep. ATL-INDET-PROC-2011-014, CERN, 2011, URL: http://cds.cern.ch/record/1384156.

[75] A. L. Schorlemmer, *Monitoring radiation damage in the ATLAS Pixel Detector*, Journal of Instrumentation **8** (2013) p. C01045, URL: http://stacks.iop.org/1748-0221/8/i=01/a=C01045.

[76] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, (2010), URL: http://cds.cern.ch/record/1291633, ATLAS Insertable B-Layer Technical Design Report Addendum, (2012), ATLAS-TDR-19-ADD-1, http://cds.cern.ch/record/1451888.

[77] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor tracker*, JINST **9** (2014) P08009, arXiv: 1404.7473 [hep-ex].

[78] ATLAS TRT Collaboration, *The ATLAS TRT Barrel Detector*, Journal of Instrumentation **3** (2008) P02014, URL: http://stacks.iop.org/1748-0221/3/i=02/a=P02014.

[79] J. Pequenao, "Computer Generated image of the ATLAS calorimeter", 2008, URL: https://cds.cern.ch/record/1095927.

[80] B. Rossi, *High-energy Particles*, Prentice-Hall physics series, New York, 1952, URL: https://books.google.co.uk/books?id=wLQmAAAAMAAJ.

[81] J. Goodson, *ATLAS magnet system*, http://www.jetgoodson.com/images/thesisImages/magnetSystems.png, [Online: accessed 3 October 2016], 2016.

[82] ATLAS Collaboration, *2015 start-up trigger menu and initial performance assessment of the ATLAS trigger using Run-2 data*, tech. rep. ATL-DAQ-PUB-2016-001, CERN, 2016, URL: https://cds.cern.ch/record/2136007.

[83] ATLAS Collaboration, "Trigger Operation Public Results", Accessed: 13/05/2016, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TriggerOperationPublicResults.

[84] I. Bird, *Computing for the Large Hadron Collider*, Annual Review of Nuclear and Particle Science **61** (2011) p. 99, eprint: http://dx.doi.org/10.1146/annurev-nucl-102010-130059, URL: http://dx.doi.org/10.1146/annurev-nucl-102010-130059.

[85] J. C. Collins and D. E. Soper, *Parton distribution and decay functions*, Nuclear Physics B **194** (1982) p. 445 , ISSN: 0550-3213, URL: http://www.sciencedirect.com/science/article/pii/0550321382900219.

[86] G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, Nuclear Physics B **126** (1977) p. 298 , ISSN: 0550-3213, URL: http://www.sciencedirect.com/science/article/pii/0550321377903844.

[87] V. N. Gribov and L. N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, Sov. J. Nucl. Phys. **15** (1972) p. 438, [Yad. Fiz.15,781(1972)].

[88] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics.*, Sov. Phys. JETP **46** (1977) p. 641, [Zh. Eksp. Teor. Fiz.73,1216(1977)].

[89] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. **D82** (2010) p. 074024, arXiv: 1007.2241 [hep-ph].

[90] H1 and ZEUS Collaborations, *Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data*, The European Physical Journal C **75** (2015) p. 1, URL: http://dx.doi.org/10.1140/epjc/s10052-015-3710-4.

[91] H1 and ZEUS Collaborations, *Combined measurement and QCD analysis of the inclusive $e^{\pm}p$ scattering cross sections at HERA*, Journal of High Energy Physics **2010** (2010) p. 1, URL: http://dx.doi.org/10.1007/JHEP01(2010)109.

[92] A. D. Martin et al., *Parton distributions for the LHC*, Eur. Phys. J. **C63** (2009) p. 189, arXiv: 0901.0002 [hep-ph].

[93] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. **B867** (2013) p. 244, arXiv: 1207.1303 [hep-ph].

[94] B. Andersson et al., *Parton Fragmentation and String Dynamics*, Phys. Rept. **97** (1983) p. 31.

[95] B. Webber, *A QCD model for jet fragmentation including soft gluon interference*, Nuclear Physics B **238** (1984) p. 492 , ISSN: 0550-3213, URL: http://www.sciencedirect.com/science/article/pii/055032138490333X.

[96] T. Sjöstrand and M. van Zijl, *A multiple-interaction model for the event structure in hadron collisions*, Phys. Rev. D **36** (7 1987) p. 2019, URL: http://link.aps.org/doi/10.1103/PhysRevD.36.2019.

[97] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) p. 026, arXiv: hep-ph/0603175 [hep-ph].

[98] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) p. 852, arXiv: 0710.3820 [hep-ph].

[99] G. Corcella et al., *HERWIG 6.5 release note*, (2002), arXiv: hep-ph/0210213 [hep-ph].

[100] M. Bahr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. **C58** (2008) p. 639, arXiv: 0803.0883 [hep-ph].

[101] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) p. 007, arXiv: 0811.4622 [hep-ph].

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

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

[104] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) p. 043, arXiv: 1002.2581 [hep-ph].

[105] B. P. Kersevan and E. Richter-Was, *The Monte Carlo event generator AcerMC version 2.0 with interfaces to PYTHIA 6.2 and HERWIG 6.5*, (2004), arXiv: hep-ph/0405247 [hep-ph].

[106] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) p. 079, arXiv: 1405.0301 [hep-ph].

[107] J. M. Campbell and R. Ellis, *MCFM for the Tevatron and the LHC*, Nucl. Phys. Proc. Suppl. **205-206** (2010) p. 10, arXiv: 1007.3492 [hep-ph].

[108] S. Agostinelli et al., *GEANT4: A Simulation Toolkit*, Nucl. Instrum. Meth. **A 506** (2003) p. 250.

[109] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J **C 70** (2010) p. 823, arXiv: 1005.4568 [physics.ins-det].

[110] T Cornelissen et al., *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, tech. rep. ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002, CERN, 2007, URL: https://cds.cern.ch/record/1020106.

[111] ATLAS Collaboration, *Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment*, tech. rep. ATLAS-CONF-2012-042, 2012, URL: https://cds.cern.ch/record/1435196.

[112] ATLAS Collaboration, *Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics*, (2009), arXiv: 0901.0512 [hep-ex].

[113] T Cornelissen et al., *The new ATLAS track reconstruction (NEWT)*, Journal of Physics: Conference Series **119** (2008) p. 032014, URL: http://stacks.iop.org/1742-6596/119/i=3/a=032014.

[114] ATLAS Collaboration, *Luminosity Determination in pp Collisions at $\sqrt{s} = 7$ TeV Using the ATLAS Detector at the LHC*, Eur. Phys. J. **C71** (2011) p. 1630, arXiv: 1101.2185 [hep-ex].

[115] ATLAS Collaboration, *ATLAS Luminosity Results Run 2*, Accessed: 05/07/2016, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[116] ATLAS Collaboration, *Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data*, accepted by Eur. Phys. J. (2014), arXiv: 1404.2240 [hep-ex].

[117] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data*, tech. rep. ATLAS-CONF-2014-032, CERN, 2014, URL: https://cds.cern.ch/record/1706245.

[118] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, Eur. Phys. J. **C74** (2014) p. 3071, arXiv: 1407.5063 [hep-ex].

[119] R Nicolaidou et al., *Muon identification procedure for the ATLAS detector at the LHC using Muonboy reconstruction package and tests of its performance using cosmic rays and single beam data*, Journal of Physics: Conference Series **219** (2010) p. 032052, URL: http://stacks.iop.org/1742-6596/219/i=3/a=032052.

[120] ATLAS Collaboration, *Muon reconstruction efficiency and momentum resolution of the ATLAS experiment in proton-proton collisions at $\sqrt{s}$=7 TeV in 2010*, accepted by Eur. Phys. J. (2014), arXiv: 1404.4562 [hep-ex].

[121] ATLAS Collaboration, "First 2.36 TeV Collision Events recorded by the ATLAS experiment, December 8th and 14th, 2009", 2009, URL: https://cds.cern.ch/record/1228300.

[122] G. P. Salam, *Towards Jetography*, Eur. Phys. J. **C67** (2010) p. 637, arXiv: 0906.1833 [hep-ph].

[123] W. Lampl et al., *Calorimeter clustering algorithms: Description and performance*, (2008), URL: http://cds.cern.ch/record/1099735.

[124] J.-R. Vlimant, "Measurement of cross section of production of pairs of top quarks /anti-signal in proton/anti-proton collisions at the square root of 1.96 TeV", PhD thesis: Paris U., VI-VII, 2005, URL: http://tel.ccsd.cnrs.fr/tel-00011503.

[125] P. Speckmayer, T. Carli, and C. W. Fabjan, "Energy Measurement of Hadrons with the CERN ATLAS Calorimeter", Presented on 18 Jun 2008, PhD thesis: Vienna, Tech. U., 2008, URL: https://cds.cern.ch/record/1112036.

[126] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, Phys. Rev. **D48** (1993) p. 3160, arXiv: hep-ph/9305266 [hep-ph].

[127] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm*, JHEP **04** (2008) p. 063, arXiv: 0802.1189 [hep-ph].

[128] Y. L. Dokshitzer et al., *Better jet clustering algorithms*, JHEP **08** (1997) p. 001, arXiv: hep-ph/9707323 [hep-ph].

[129] G. P. Salam and G. Soyez, *A Practical Seedless Infrared-Safe Cone Jet Algorithm*, JHEP **05** (2007) p. 086, arXiv: 0704.0292 [hep-ph].

[130] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at* $\sqrt{s} = 7$ *TeV*, Eur. Phys. J. **C 73** (2013) p. 2304, arXiv: 1112.6426 [hep-ex].

[131] ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton–proton collisions at* $\sqrt{s} = 7$ *TeV with the ATLAS detector*, submitted to Eur. Phys. J. (2014), arXiv: 1406.0076 [hep-ex].

[132] M. Cacciari, G. P. Salam, and G. Soyez, *The Catchment Area of Jets*, JHEP **04** (2008) p. 005, arXiv: 0802.1188 [hep-ph].

[133] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Phys. Lett. **B 659** (2008) p. 119, arXiv: 0707.1378 [hep-ph].

[134] ATLAS Collaboration, *Jet energy resolution in proton-proton collisions at* $\sqrt{s} = 7$ *TeV recorded in 2010 with the ATLAS detector*, Eur. Phys. J. **C73** (2013) p. 2306, arXiv: 1210.6210 [hep-ex].

[135] ATLAS Collaboration, *Pile-up subtraction and suppression for jets in ATLAS*, tech. rep. ATLAS-CONF-2013-083, CERN, 2013, URL: https://cds.cern.ch/record/1570994.

[136] ATLAS Collaboration, *Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data*, tech. rep. ATLAS-CONF-2011-102, CERN, 2011, URL: https://cds.cern.ch/record/1369219.

[137] ATLAS Collaboration, *Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment*, (2014), URL: http://cds.cern.ch/record/1664335.

[138] ATLAS Collaboration, *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*, (2014), URL: http://cdsweb.cern.ch/record/1741020.

[139] D0 Collaboration, http://www-d0.fnal.gov/Run2Physics/top/singletop_observation/singletop_observation_updated.html, Accessed: 05/07/2016.

[140] ATLAS Collaboration, *Performance of Missing Transverse Momentum Reconstruction in Proton-Proton Collisions at 7 TeV with ATLAS*, Eur. Phys. J. **C 72** (2012) p. 1844, arXiv: 1108.5602 [hep-ex].

[141] ATLAS Collaboration, *Performance of Missing Transverse Momentum Reconstruction in ATLAS studied in Proton-Proton Collisions recorded in 2012 at 8 TeV*, tech. rep. ATLAS-CONF-2013-082, CERN, 2013, URL: https://cds.cern.ch/record/1570993.

[142] Biography.com Editors, *B.B. King Biography*, http://www.biography.com/people/bb-king-9364839, Accessed: 05/07/2016.

[143] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07** (2002) p. 012, arXiv: hep-ph/0201195 [hep-ph].

[144] ATLAS Collaboration, *ATLAS tunes of PYTHIA6 and PYTHIA8 for MC11*, (2011), URL: http://cdsweb.cern.ch/record/1363300.

[145] ATLAS Collaboration, *New ATLAS event generator tunes to 2010 data*, (2011), URL: http://cdsweb.cern.ch/record/1345343.

[146] P. Golonka and Z. Was, *PHOTOS Monte Carlo: A Precision tool for QED corrections in Z and W decays*, Eur. Phys. J. **C 45** (2006) p. 97, arXiv: hep-ph/0506026 [hep-ph].

[147] G. Luisoni et al., *$HW^{\pm}$/HZ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO*, JHEP **10** (2013) p. 083, arXiv: 1306.2542 [hep-ph].

[148] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. **D 82** (2010) p. 074024, arXiv: 1007.2241 [hep-ph].

[149] C. Englert, M. McCullough, and M. Spannowsky, *Gluon-initiated associated production boosts Higgs physics*, Phys. Rev. **D 89** (2014) p. 013013, arXiv: 1310.4828 [hep-ph].

[150] J. Ohnemus and W. J. Stirling, *Order α_s corrections to the differential cross-section for the WH intermediate mass Higgs signal*, Phys. Rev. **D 47** (1993) p. 2722.

[151] H. Baer, B. Bailey, and J. F. Owens, *$O(\alpha_s)$ Monte Carlo approach to W+Higgs associated production at hadron supercolliders*, Phys. Rev. **D 47** (1993) p. 2730.

[152] O. Brein, A. Djouadi, and R. Harlander, *NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders*, Phys. Lett. **B 579** (2004) p. 149, arXiv: hep-ph/0307206 [hep-ph].

[153] M. Ciccolini, S. Dittmaier, and M. Krämer, *Electroweak radiative corrections to associated WH and ZH production at hadron colliders*, Phys. Rev. **D 68** (2003) p. 073003, arXiv: hep-ph/0306234 [hep-ph].

[154] A. Denner et al., *EW corrections to Higgs strahlung at the Tevatron and the LHC with HAWK*, PoS **EPS-HEP2011** (2011) p. 235, arXiv: 1112.5258 [hep-ph].

[155] L. Altenkamp et al., *Gluon-induced Higgs-strahlung at next-to-leading order QCD*, JHEP **02** (2013) p. 078, arXiv: 1211.5015 [hep-ph].

[156] A. Djouadi, J. Kalinowski, and M. Spira, *HDECAY: A program for Higgs boson decays in the Standard Model and its supersymmetric extension*, Comput. Phys. Commun. **108** (1998) p. 56, arXiv: hep-ph/9704448 [hep-ph].

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

[158] M. Czakon, P. Fiedler, and A. Mitov, *The total top quark pair production cross-section at hadron colliders through $O(\alpha_S^4)$*, Phys. Rev. Lett. **110** (2013) p. 252004, arXiv: 1303.6254 [hep-ph].

[159] P. Nason and G. Zanderighi, *W^+W^- , WZ and ZZ production in the POWHEG-BOX-V2*, Eur. Phys. J. **C 74** (2014) p. 2702, arXiv: 1311.1365 [hep-ph].

[160] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, Phys. Rev. **D83** (2011) p. 091503, arXiv: 1103.2792 [hep-ph].

[161] N. Kidonakis, *NNLL resummation for s-channel single top quark production*, Phys. Rev. **D 81** (2010) p. 054028, arXiv: 1001.5034 [hep-ph].

[162] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-*, Phys. Rev. **D 82** (2010) p. 054018, arXiv: 1005.4451 [hep-ph].

[163] S. Hoeche et al., *QCD matrix elements + parton showers: The NLO case*, JHEP **04** (2013) p. 027, arXiv: 1207.5030 [hep-ph].

[164] ATLAS Collaboration, *Summary plots from the ATLAS Standard Model physics group*, Accessed: 30/06/2016, URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/.

[165] F. James and M. Roos, *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*, Comput. Phys. Commun. **10** (1975) p. 343.

[166] ATLAS Collaboration, *Search for the Standard Model Higgs boson in produced in association with a vector boson and decaying to bottom quarks with the ATLAS detector*, tech. rep. ATLAS-CONF-2012-161, 2012, URL: https://cds.cern.ch/record/1493625.

[167] ATLAS Collaboration, *Improved luminosity determination in pp collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector at the LHC*, Eur. Phys. J. **C 73** (2013) p. 2518, arXiv: 1302.4393 [hep-ex].

[168] N. Kidonakis, "Differential and total cross sections for top pair and single top production", *Proceedings, 20th International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2012)*, 2012 p. 831, arXiv: 1205.3453 [hep-ph], URL: https://inspirehep.net/record/1114754/files/arXiv:1205.3453.pdf.

[169] S. Dittmaier et al., *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, (2011), arXiv: 1101.0593 [hep-ph].

[170] A. L. Read, *Presentation of search results: the CL s technique*, Journal of Physics G: Nuclear and Particle Physics **28** (2002) p. 2693, URL: http://stacks.iop.org/0954-3899/28/i=10/a=313.

[171] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. **C71** (2011) p. 1554, [Erratum: Eur. Phys. J. **C73**, 2501 (2013)], arXiv: 1007.1727 [physics.data-an].

[172] G. Bohm and G. Zech, *Introduction to statistics and data analysis for physicists*, DESY, 2010, ISBN: 9783935702416, URL: https://books.google.co.uk/books?id=aQBHcgAACAAJ.

[173] CMS Collaboration, *Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks*, Phys. Rev. **D89** (2014) p. 012003, arXiv: 1310.3687 [hep-ex].

[174] ATLAS and CMS Collaborations, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV*, JHEP **08** (2016) p. 045, arXiv: 1606.02266 [hep-ex].

[175] ATLAS Collaboration, *Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment*, Eur. Phys. J. **C76** (2016) p. 6, arXiv: 1507.04548 [hep-ex].

[176] *Search for the Standard Model Higgs boson produced in association with a vector boson and decaying to a $b\bar{b}$ pair in pp collisions at 13 TeV using the ATLAS detector*, tech. rep. ATLAS-CONF-2016-091, CERN, 2016, URL: http://cds.cern.ch/record/2206813.

[177] ATLAS Collaboration, *Letter of Intent for the Phase-II Upgrade of the ATLAS Experiment*, tech. rep. CERN-LHCC-2012-022. LHCC-I-023, CERN, 2012, URL: https://cds.cern.ch/record/1502664.

[178] ATLAS Collaboration, *Prospects for the study of the Higgs boson in the VH(bb) channel at HL-LHC*, tech. rep. ATL-PHYS-PUB-2014-011, CERN, 2014, URL: https://cds.cern.ch/record/1740962.

[179] L. G. Almeida et al., *Playing Tag with ANN: Boosted Top Identification with Pattern Recognition*, JHEP **07** (2015) p. 086, arXiv: 1501.05968 [hep-ph].

[180] P. Baldi, P. Sadowski, and D. Whiteson, *Searching for Exotic Particles in High-Energy Physics with Deep Learning*, Nature Commun. **5** (2014) p. 4308, arXiv: 1402.4735 [hep-ph].

[181] S. Raschka, *Python Machine Learning*, Packt Publishing, 2015, ISBN: 9781783555147, URL: https://books.google.co.uk/books?id=GOVOCwAAQBAJ.

[182] O. Behnke et al., *Data Analysis in High Energy Physics: A Practical Guide to Statistical Methods*, Wiley, 2013, ISBN: 9783527653430, URL: https://books.google.co.uk/books?id=NWSoAAAAQBAJ.

[183] A. Ng, *CS229 Machine Learning: Lecture Notes*, Stanford University, 2014, URL: http://cs229.stanford.edu/.

[184] Y. Bengio et al., "Greedy Layer-Wise Training of Deep Networks", 2007 p. 153, URL: http://www.iro.umontreal.ca/~lisa/pointeurs/BengioNips2006All.pdf.

[185] Y. Bengio, *Learning deep architectures for AI*, Foundations and Trends in Machine Learning **2** (2009) p. 1, Also published as a book. Now Publishers, 2009.

[186] F. Rosenblatt, *Principles of Neurodynamics*, Spartan Book, 1962.

[187] C. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, 1st ed. 2006. Corr. 2nd printing 2011, Springer, Oct. 2007, ISBN: 0387310738, URL: http://www.worldcat.org/isbn/0387310738.

[188] Y. LeCun et al., "Neural Networks: Tricks of the Trade", Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, chap. Efficient BackProp p. 9, ISBN: 978-3-540-49430-0, URL: http://dx.doi.org/10.1007/3-540-49430-8_2.

[189] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1*, ed. by D. E. Rumelhart, J. L. McClelland, and C. PDP Research Group, Cambridge, MA, USA: MIT Press, 1986, chap. Learning Internal Representations by Error Propagation p. 318, ISBN: 0-262-68053-X, URL: http://dl.acm.org/citation.cfm?id=104279.104293.

[190] P. Baldi, *Autoencoders, Unsupervised Learning, and Deep Architectures.*, Journal of Machine Learning Research (2012), Proceedings of ICML Workshop on Unsupervised and Transfer Learning.

[191] L. de Oliveira, *AGILEPack: Algorithms for Generalized Inference, Learning, and Extraction Package*, https://github.com/lukedeo/AGILEPack, 2015.

[192] de Oliveira, Luke, Private communication.

[193] D. Erhan et al., *Why Does Unsupervised Pre-training Help Deep Learning?*, J. Mach. Learn. Res. **11** (Mar. 2010) p. 625, ISSN: 1532-4435, URL: http://dl.acm.org/citation.cfm?id=1756006.1756025.

[194] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Corrected, Springer, Aug. 2003, ISBN: 0387952845.

[195] R. E. Schapire, *The strength of weak learnability*, Machine Learning **5** (1990) p. 197, ISSN: 1573-0565, URL: http://dx.doi.org/10.1007/BF00116037.

[196] L. Breiman, *Random Forests*, Mach. Learn. **45** (Oct. 2001) p. 5, ISSN: 0885-6125, URL: http://dx.doi.org/10.1023/A:1010933404324.

[197] G. Louppe et al., "Understanding variable importances in forests of randomized trees", *Advances in Neural Information Processing Systems 26*, ed. by C. Burges et al., 2013 p. 431, URL: http://media.nips.cc/nipsbooks/nipspapers/paper_files/nips26/281.pdf.

[198] B. Gregorutti, B. Michel, and P. Saint-Pierre, *Correlation and variable importance in random forests*, ArXiv e-prints (Oct. 2013), arXiv: 1310.5726 [stat.ME].

[199] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, Journal of Machine Learning Research **12** (2011) p. 2825.

[200] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Second, Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann, June 2005, ISBN: 0120884070.

[201] R. Brun and F. Rademakers, *ROOT — An object oriented data analysis framework*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **389** (1997) p. 81 , New Computing Techniques in Physics Research V, ISSN: 0168-9002, URL: http://www.sciencedirect.com/science/article/pii/S016890029700048X.

[202] N. Dawe et al., *root_numpy: 4.3.0*, June 2015, URL: http://dx.doi.org/10.5281/zenodo.18750.

[203] F. Pérez and B. E. Granger, *IPython: a System for Interactive Scientific Computing*, Computing in Science and Engineering **9** (May 2007) p. 21, ISSN: 1521-9615, URL: http://ipython.org.

[204] W. McKinney, "Data Structures for Statistical Computing in Python", *Proceedings of the 9th Python in Science Conference*, ed. by S. van der Walt and J. Millman, 2010 p. 51.

[205] J. D. Hunter, *Matplotlib: A 2D Graphics Environment*, Computing in Science and Engineering **9** (2007) p. 90, URL: http://scitation.aip.org/content/aip/journal/cise/9/3/10.1109/MCSE.2007.55.

[206] S. v. d. Walt, S. C. Colbert, and G. Varoquaux, *The NumPy Array: A Structure for Efficient Numerical Computation*, Computing in Science and Engineering **13** (2011) p. 22, URL: http://scitation.aip.org/content/aip/journal/cise/13/2/10.1109/MCSE.2011.37.

[207] J Pequenao, "Simulated production of a Higgs event in ATLAS. This track is an example of simulated data modeled for the ATLAS detector on the Large Hadron Collider (LHC) at CERN, which will begin taking data in 2008.", ATLAS Collection., 2008, URL: https://cds.cern.ch/record/1102948.

[208] J. M. Butterworth et al., *Jet substructure as a new Higgs search channel at the LHC*, Phys. Rev. Lett. **100** (2008) p. 242001, arXiv: 0802.2470 [hep-ph].

[209] ATLAS Collaboration, *Performance of large-R jets and jet substructure reconstruction with the ATLAS detector*, tech. rep. ATLAS-CONF-2012-065, 2012, URL: https://cds.cern.ch/record/1459530.

[210] ATLAS Collaboration, *Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector*, JHEP **09** (2013) p. 076, arXiv: 1306.4945 [hep-ex].

[211] ATLAS Collaboration, *Performance of W Boson Boosted Identification with the ATLAS Detector*, tech. rep. ATL-PHYS-PUB-2014-004, CERN, 2014, URL: https://cds.cern.ch/record/1690048.

[212] D. Krohn, J. Thaler, and L.-T. Wang, *Jet Trimming*, JHEP **02** (2010) p. 084, arXiv: 0912.1342 [hep-ph].

[213] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Techniques for improved heavy particle searches with jet substructure*, Physical Review D - Particles, Fields, Gravitation and Cosmology **80** (Mar. 2009) p. 4, arXiv: 0903.5081, URL: http://arxiv.org/abs/0903.5081.

[214] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Techniques for improved heavy particle searches with jet substructure*, Phys. Rev. **D80** (2009) p. 051501, arXiv: 0903.5081 [hep-ph].

[215] CMS Collaboration, *Identification techniques for highly boosted W bosons that decay into hadrons*, JHEP **12** (2014) p. 017, arXiv: 1410.4227 [hep-ex].

[216] CMS Collaboration, *Search for new resonances decaying via WZ to leptons in proton-proton collisions at $\sqrt{s}$ = 8 TeV*, Phys. Lett. **B740** (2015) p. 83, arXiv: 1407.3476 [hep-ex].

[217] L. G. Almeida et al., *Substructure of high-p_T jets at the LHC*, Phys. Rev. **D79** (2009) p. 074017, arXiv: 0807.0234 [hep-ph].

[218] A. J. Larkoski, G. P. Salam, and J. Thaler, *Energy Correlation Functions for Jet Substructure*, JHEP **06** (2013) p. 108, arXiv: 1305.0007 [hep-ph].

[219] A. J. Larkoski, I. Moult, and D. Neill, *Power Counting to Better Jet Observables*, JHEP **12** (2014) p. 009, arXiv: 1409.6298 [hep-ph].

[220] A. J. Larkoski, I. Moult, and D. Neill, *Analytic Boosted Boson Discrimination*, JHEP **05** (2016) p. 117, arXiv: 1507.03018 [hep-ph].

[221] J. M. Butterworth, B. E. Cox, and J. R. Forshaw, *WW scattering at the CERN LHC*, Phys. Rev. **D65** (2002) p. 096014, arXiv: hep-ph/0201098 [hep-ph].

[222] J. Thaler and L.-T. Wang, *Strategies to Identify Boosted Tops*, JHEP **07** (2008) p. 092, arXiv: 0806.0023 [hep-ph].

[223] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, JHEP **03** (2011) p. 015, arXiv: 1011.2268 [hep-ph].

[224] J. Thaler and K. Van Tilburg, *Maximizing Boosted Top Identification by Minimizing N-subjettiness*, JHEP **02** (2012) p. 093, arXiv: 1108.2701 [hep-ph].

[225] A. J. Larkoski, D. Neill, and J. Thaler, *Jet Shapes with the Broadening Axis*, JHEP **04** (2014) p. 017, arXiv: 1401.2158 [hep-ph].

[226] E. Leader and E. Predazzi, *An Introduction to Gauge Theories and Modern Particle Physics*, vol. 2, Cambridge Books Online, Cambridge University Press, 1996, ISBN: 9780511622601, URL: http://dx.doi.org/10.1017/CBO9780511622601.

[227] ATLAS Collaboration, *Measurement of the cross-section of high transverse momentum vector bosons reconstructed as single jets and studies of jet substructure in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector*, New J. Phys. **16** (2014) p. 113013, arXiv: 1407.0800 [hep-ex].

[228] P. Duinker, *Review of e^+e^- physics at PETRA*, Rev. Mod. Phys. **54** (2 1982) p. 325, URL: http://link.aps.org/doi/10.1103/RevModPhys.54.325.

[229] D. P. Barber et al., *Discovery of Three-Jet Events and a Test of Quantum Chromodynamics at PETRA*, Phys. Rev. Lett. **43** (12 1979) p. 830, URL: http://link.aps.org/doi/10.1103/PhysRevLett.43.830.

[230] G. C. Fox and S. Wolfram, *Event shapes in e^+e^- annihilation*, Nuclear Physics B **157** (Oct. 1979) p. 543.

[231] ATLAS Collaboration, *Studies of the impact and mitigation of pile-up on large-R and groomed jets in ATLAS at $\sqrt{s}$ = 7 TeV*, tech. rep. ATLAS-CONF-2012-066, CERN, 2012, URL: https://cds.cern.ch/record/1459531.

[232] ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, tech. rep. ATL-PHYS-PUB-2012-003, CERN, 2012, URL: https://cds.cern.ch/record/1474107.

[233] G. Watt and R. S. Thorne, *Study of Monte Carlo approach to experimental uncertainty propagation with MSTW 2008 PDFs*, JHEP **08** (2012) p. 052, arXiv: 1205.4024 [hep-ph].

[234] ATLAS Collaboration, *ATLAS Run 1 Pythia8 tunes*, tech. rep. ATL-PHYS-PUB-2014-021, CERN, 2014, URL: https://cds.cern.ch/record/1966419.

[235] D. J. Lange, *The EvtGen particle decay simulation package*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **462** (2001) p. 152 , BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines, ISSN: 0168-9002, URL: http://www.sciencedirect.com/science/article/pii/S0168900201000894.

[236] ATLAS Collaboration, *Search for high-mass diboson resonances with boson-tagged jets in proton-proton collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector*, accepted by JHEP (2015), arXiv: 1506.00962 [hep-ex].

[237] ATLAS Collaboration, *Search for resonances with boson-tagged jets in 3.2 fb^{-1} of pp collisions $\sqrt{s}$ =13 TeV collected with the ATLAS detector*, tech. rep. ATLAS-CONF-2015-073, CERN, 2015, URL: https://cds.cern.ch/record/2114845.

[238] ATLAS Collaboration, *Search for resonances with boson-tagged jets in 15.5 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV collected with the ATLAS detector*, tech. rep. ATLAS-CONF-2016-055, CERN, 2016, URL: https://cds.cern.ch/record/2206137.

[239] R. Chislett, "Studies of hadronic decays of high transverse momentum W and Z bosons with the ATLAS detector at the LHC", Presented August 21 2014, PhD thesis: University College London, 2014, URL: {https://www.hep.ucl.ac.uk/theses/Chislett_thesis.pdf}.

[240] ATLAS Collaboration, *Light-quark and gluon jet discrimination in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Eur. Phys. J. **C74** (2014) p. 3023, arXiv: 1405.6583 [hep-ex].

[241] ESPNCricInfo Ltd., *The scoreboard at the end of the Timeless Test, South Africa v England, 5th Test, Durban, March 14, 1939.*, http://www.espncricinfo.com/ci/content/image/index.html?object=62657, [Online: accessed 1 August 2016], 1939.

[242] NVIDIA Corporation, *What is GPU Computing?*, 2013, URL: http://www.nvidia.com/object/what-is-gpu-computing.html.

[243] NVIDIA Corporation, *CUDA*, 2013, URL: http://www.nvidia.com/object/cuda_home_new.html.

[244] K. Group, *OpenCL*, 2013, URL: https://www.khronos.org/opencl/.

[245] J. S. And and E. Kandrot, *CUDA by Example*, Addison-Wesley, 2011, ISBN: 9780131387683.

[246] J Luitjens and S Rennich, *CUDA Warps and Occupancy*, (2011), URL: http://on-demand.gputechconf.com/gtc-express/2011/presentations/cuda_webinars_WarpsAndOccupancy.pdf.

[247] D. K. Hwu and Wen-mei, *Programming Massively Parallel Processors : A Hands-on Approach*, Burlington, MA, USA: Morgan Kaufmann, 2010, ISBN: 9780123814722.

[248] J. Kider, *NVIDIA Fermi Architecture*, (2011).

[249] NVIDIA Corporation, *Whitepaper NVIDIA's Next Generation CUDA Compute Architecture:* () p. 1, URL: www.nvidia.com.

[250] NVIDIA Corporation, *Kepler GK110 Whitepaper*.

[251] NVIDIA Corporation, *Datasheet: NVIDIA Kepler Next-Generation Cuda Compute Architecture*, (2009).

[252] R Hauser, *The ATLAS Data Acquisition and High Level Trigger Systems: Experience and Upgrade Plans*, tech. rep. ATL-DAQ-PROC-2012-073, CERN, 2012, URL: https://cds.cern.ch/record/1497132.

[253] J. Howard, *GPU-Based Data Preparation in the ATLAS Level 2 Trigger*, Edinburgh, 2011, URL: https://indico.cern.ch/event/141309/contributions/1369476/attachments/126040/179013/FutureComputing-Howard-17-June-2011.pdf.

[254] D. Emeliyanov, *Future computing in particle physics : a report from a workshop, Edinburgh*, (2011), URL: http://www.ppd.stfc.ac.uk/PPD/resources/pdf/ppd_seminar_111005_talk_dmitry_emeliyanov.PDF.

[255] NVIDIA Corporation, *Tuning CUDA Applications for Kepler*, (2012), URL: www.nvidia.com.

[256] NVIDIA Corporation, *Nvidia Visual Profiler*, 2013, URL: https://developer.nvidia.com/nvidia-visual-profiler.

[257] "Optimizing Application Performance with CUDA Profiling Tools Why Profile ?", *GPU Technology Conference*, 2012, URL: http://on-demand.gputechconf.com/gtc/2012/presentations/S0419B-GTC2012-Profiling-Profiling-Tools.pdf.

[258] V. Volkov, *Better Performance at Lower Occupancy*, (2010).

[259] NVIDIA Corporation, *CUDA Dynamic Parallelism*, 2012, URL: http://developer.download.nvidia.com/assets/cuda/files/CUDADownloads/TechBrief_Dynamic_Parallelism_in_CUDA.pdf.

[260] A. Hook, M. Jankowiak, and J. G. Wacker, *Jet Dipolarity: Top Tagging with Color Flow*, JHEP **04** (2012) p. 007, arXiv: 1102.1012 [hep-ph].