

A search for the Higgs boson decaying to two  $b$ -quarks in  
association with a dileptonically decaying top-quark pair  
with the ATLAS detector

Ian Allan Connelly

Department of Physics  
Royal Holloway, University of London



A thesis submitted to the University of London for the  
Degree of Doctor of Philosophy

May 16, 2016

---

## DECLARATION

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

Ian Allan Connelly

## Abstract

Details of a search for Higgs production in association with a top-quark pair, where the Higgs decays to two  $b$ -quarks, are presented using  $20.3\text{ fb}^{-1}$  of  $\sqrt{s} = 8\text{ TeV}$  data from the ATLAS detector at the LHC. Whilst the existence of the Higgs boson has been established, there has yet to be a direct observation of the Higgs boson coupling to top quarks. This search aims to provide a measurement of the top-Higgs coupling to probe the nature of the Standard Model of particle physics. Dileptonic top-quark pair events are used to measure the  $b$ -tagging efficiency on data and to search for this Higgs production mechanism. To separate the sought after signal events from the large amount of background events, the application of a technique known as the matrix-element method is presented. An investigation into the impact on the discovery significance using this technique is carried out, resulting in an 8% improvement in sensitivity relative to the nominal analysis.

## Acknowledgements

There are many I would like to thank for getting me through my PhD relatively unscathed.

First and foremost, I would like to thank my supervisor Glen Cowan, and my pseudo-supervisor Pedro Teixeira-Dias, for sharing their extensive knowledge and always being available to provide advice and direct my work.

I would like to thank my mother and father, Alison and Graham Connelly, my sister, Siân Connelly, and grandmother, Patricia Connelly, who have given me their unconditional support and love through my years.

Joshua Thomas-Wilsker for being my mutual code duck and partner in (physics) crime as we explored the world. Veronique Boisvert and Michele Facci Giannelli for their support on many Top and Higgs related discussions. Ricardo Gonçalo who played a huge role in shaping my work during my stay at CERN and taught me how to think like an experimentalist. Olaf Nackenhorst and Austin Bayes for teaching me the ways of the matrix element method and whose work paved the way to my own result. The ATLAS HSG5 and HSG8  $H \rightarrow b\bar{b}$  subgroups, without all of whom in the spirit of collaboration, we would not have any results.

I would like to thank the rest of the particle physics group at Royal Holloway for being amazing colleagues and providing many productive discussions.

I would also like to make a special mention to John Beddow, Duncan Leggat, Simon Wright, Mark Smith, Zara Grout and Ruth Sandbach for their friendship and support as we lived and breathed physics together, both at CERN and back in the UK.

Finally, thank you to everyone else that I have met during my four years of postgraduate study who have shared experiences with me, supported me and helped shape my future.

This work is dedicated to the memory of my late grandfather,

Dennis Allan Connelly.

He saw me start this journey but cannot see me finish.

---

# Contents

<b>1</b>	<b>Introduction</b>	<b>18</b>
<b>2</b>	<b>Theoretical Motivation</b>	<b>21</b>
2.1	Introduction . . . . .	22
2.2	Local Gauge Invariance . . . . .	23
2.3	The Fundamental Particles of the Standard Model . . . . .	24
2.4	The Fundamental Forces of the Standard Model . . . . .	26
2.4.1	Quantum Electrodynamics . . . . .	26
2.4.2	Quantum Chromodynamics . . . . .	27
2.4.3	The Weak Force . . . . .	27
2.5	Spontaneous Symmetry Breaking . . . . .	30
2.5.1	Spontaneous symmetry breaking of a global $U(1)$ symmetry . . . . .	30
2.5.2	Spontaneous symmetry breaking of a local $U(1)$ symmetry . . . . .	32
2.6	The Higgs Mechanism and Electroweak Unification . . . . .	34
2.6.1	Yukawa Coupling . . . . .	36
2.7	The Top Yukawa Coupling and $t\bar{t}H$ Production . . . . .	39
2.7.1	Vacuum Stability . . . . .	42
2.7.2	Higgs Mass Corrections . . . . .	46
2.8	Recent Higgs Results . . . . .	47
<b>3</b>	<b>The LHC and ATLAS Detector</b>	<b>52</b>
3.1	Overview of the LHC Complex . . . . .	52
3.2	ATLAS - A Toroidal LHC Apparatus . . . . .	55
3.2.1	The Inner Detector . . . . .	56

3.2.2	The Calorimetry System . . . . .	59
3.2.3	Muon Spectrometer System . . . . .	62
3.2.4	The Magnet System . . . . .	65
3.2.5	The Trigger and Data Acquisition System . . . . .	65
3.3	Event Simulation . . . . .	68
3.3.1	Event Generation . . . . .	68
3.3.2	Detector Simulation . . . . .	69
<b>4</b>	<b>Object Reconstruction</b>	<b>70</b>
4.1	Vertexing and Tracking . . . . .	71
4.2	Electron Reconstruction and Identification . . . . .	72
4.3	Muons . . . . .	77
4.4	Hadronic Jets . . . . .	79
4.5	<i>B</i> -tagging Hadronic Jets . . . . .	85
4.6	Overlap Removal . . . . .	90
4.6.1	Electrons and Jets . . . . .	90
4.6.2	Jets and Electrons . . . . .	91
4.6.3	Jets and Muons . . . . .	91
4.6.4	Muons and Electrons . . . . .	91
<b>5</b>	<b><i>B</i>-Tagging Calibration in ATLAS</b>	<b>92</b>
5.1	<i>B</i> -Tagging in ATLAS . . . . .	93
5.1.1	Calibration of Algorithms . . . . .	93
5.2	Measurement of <i>B</i> -Tagging Efficiencies in 8 TeV Data . . . . .	94
5.2.1	Object Selection . . . . .	95
5.2.2	Event Selection . . . . .	98
5.2.3	Simulation Samples . . . . .	100
5.2.4	$ee$ and $\mu\mu$ Events . . . . .	101
5.2.5	$e\mu$ Events . . . . .	106
5.2.6	Event Yields . . . . .	111
5.2.7	Calibration Method . . . . .	112
5.2.8	Systematic Uncertainties . . . . .	116

5.2.9	Results . . . . .	124
<b>6</b>	<b>Search for dileptonic <math>t\bar{t}H (H \rightarrow b\bar{b})</math> production</b>	<b>129</b>
6.1	Analysis Outline . . . . .	133
6.2	Object and Event Selection . . . . .	134
6.2.1	Lepton Selection . . . . .	135
6.2.2	Jet Selection . . . . .	135
6.2.3	Object Overlap Removal . . . . .	135
6.3	Event Selection . . . . .	136
6.3.1	Multilepton Veto . . . . .	137
6.4	Monte Carlo Simulation Samples . . . . .	138
6.5	Heavy Flavour Modelling . . . . .	139
6.5.1	Heavy flavour classification of $t\bar{t}+j$ ets . . . . .	139
6.5.2	Top $p_T$ reweighting . . . . .	141
6.5.3	NLO $t\bar{t}+b\bar{b}$ reweighting . . . . .	142
6.5.4	Tag Rate Function . . . . .	144
6.6	Event Classification . . . . .	146
6.7	Neural Network Classifier . . . . .	149
6.8	Analysis Distributions . . . . .	153
6.9	Results . . . . .	154
6.10	Outlook . . . . .	158
<b>7</b>	<b>The Matrix Element Method</b>	<b>160</b>
7.1	Mathematical Description . . . . .	161
7.2	Implementation . . . . .	163
7.3	Integration Approximations . . . . .	166
7.3.1	Helicity . . . . .	167
7.3.2	Permutations . . . . .	168
7.3.3	Kinematic Assumptions . . . . .	169
7.3.4	Integration Ranges . . . . .	172
7.4	Phase Space Volume Element . . . . .	173
7.4.1	Integration Strategy - Integrated Neutrino Solution . . . . .	176

7.4.2	Integration Strategy - Analytic Neutrino Solution . . . . .	177
7.4.3	Discussion of Integration Strategies . . . . .	178
7.5	Statistical Measures . . . . .	178
7.5.1	Separation by Non-Overlapping Area . . . . .	179
7.5.2	Discovery Significance . . . . .	179
7.5.3	NeuroBayes Significance . . . . .	180
7.6	Expected Performance . . . . .	181
7.7	Validation of Methods . . . . .	185
7.8	Truth Matching Efficiency . . . . .	188
7.9	Calculation Timing . . . . .	194
7.10	NeuroBayes Training . . . . .	195
7.11	Simultaneous Fit Results . . . . .	196
7.12	Application of Analysis to the 8 TeV Dataset . . . . .	201
7.13	Post-Fit Validation . . . . .	202
7.14	$t\bar{t}H$ Candidate Event Displays . . . . .	207
<b>8</b>	<b>Future Directions</b>	<b>210</b>
8.1	Run 2 Extrapolation . . . . .	211
8.2	Further Matrix Element Method Variables . . . . .	213
8.2.1	Inter-Jet Particle Flow . . . . .	213
8.2.2	Event Quality . . . . .	215
8.3	Final Comment . . . . .	216
<b>Bibliography</b>		<b>217</b>

---

# List of Figures

2.1	The interaction vertex of QED. . . . .	26
2.2	The interaction vertices of QCD. . . . .	28
2.3	The interaction vertices of the weak force. . . . .	29
2.4	Potential well in a $\phi^4$ theory. . . . .	32
2.5	The interaction vertices of the Higgs boson in the Standard Model. . . . .	38
2.6	Leading-order Feynman diagrams of the $t\bar{t}H$ and $t\bar{t}+b\bar{b}$ processes [13, 46]. . . . .	41
2.7	Two of the lowest order Feynman diagrams for a Higgs boson being produced through gluon-gluon fusion and for a Higgs boson decaying to two photons. . . . .	42
2.8	Higgs production cross-sections at 8 TeV [41]. . . . .	43
2.9	The stability of the Higgs potential as a joint function of the Higgs mass and the top mass [49]. . . . .	44
2.10	The stability of the Higgs potential as a joint function of the quartic Higgs coupling and the top Yukawa coupling at the Planck scale [49]. . . . .	45
2.11	The running of the quartic Higgs coupling for variations of top Yukawa couplings calculated at the top mass [39]. . . . .	46
2.12	Virtual correction to the Higgs mass through a top-loop. . . . .	46
2.13	Higgs mass phase space allowed by precision electroweak fits and experimental bounds in July 2011 [54]. . . . .	48
2.14	Higgs mass measurements in the $H \rightarrow 4l$ channel and the $H \rightarrow \gamma\gamma$ channel with the ATLAS detector using the Run 1 dataset [57]. . . . .	50
2.15	The observed linear correlation between particle masses and their coupling to the Higgs boson using the kappa framework as determined by a fit to the Standard Model expectation [47, 58, 59] . . . . .	51

3.1	Diagram detailing the CERN accelerator complex which feeds the LHC as well as additional experiments [60]. . . . .	54
3.2	Diagram of the ATLAS detector [61]. . . . .	55
3.3	Diagram of the ATLAS inner detector [61]. . . . .	57
3.4	A comparison of the fraction of high threshold hits in the TRT for pions and electrons using $\sqrt{s} = 7$ TeV 2010 data [62]. . . . .	59
3.5	Diagram of the ATLAS calorimeter system [61]. . . . .	60
3.6	Example of the segmentation of the ECAL in the barrel region [61]. . . . .	61
3.7	Diagram of the ATLAS muon spectrometer system [61]. . . . .	63
3.8	Cross-section of the muon spectrometer in the bending plane [61]. . . . .	63
3.9	Illustration of the magnet system windings in the ATLAS magnet system [61]. .	66
4.1	Luminosity recorded by the ATLAS detector as a function of the mean number of interactions per bunch-crossing at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV [77, 78]. . . .	72
4.2	The average probability of a high-threshold hit in the TRT measured in test-beam studies [46]. . . . .	73
4.3	Electron identification efficiency ROC curve [62]. . . . .	74
4.4	Electron energy resolution as a function of energy [61]. . . . .	75
4.5	Summary of the 7 and 8 TeV electron identification efficiencies [79]. . . . .	76
4.6	The muon $p_T$ resolution as a function of $ \eta $ [61]. . . . .	78
4.7	The muon reconstruction efficiencies as a function of $\eta$ [80]. . . . .	79
4.8	Illustration of the variation in jet definitions through the application of different reconstruction algorithms [86]. . . . .	81
4.9	A summary of the jet energy scale correction procedure used by the ATLAS experiment [88]. . . . .	83
4.10	Differential distributions of jet $p_T$ with respect to pile-up sensitive variables [90].	84
4.11	Jet energy resolution as a function of $p_T$ [91]. . . . .	85
4.12	Illustration of the formation of a secondary vertex from a $B$ -hadron decaying in-flight within a reconstructed jet. . . . .	86
4.13	The secondary vertex radial resolution for $B$ -hadrons [61]. . . . .	87
4.14	An illustration of the weak decay chain of a $B$ -hadron [94]. . . . .	89
4.15	The MV1 $b$ -tagging efficiency ROC curve in simulated $t\bar{t}$ events [95]. . . . .	90

5.1	Control plots of the invariant mass spectrum in the $ee$ and $\mu\mu$ channels. . . . .	103
5.2	Control plots highlighting mismodelling in the $ee$ and $\mu\mu$ channels. . . . .	104
5.3	Control plots investigating mismodelling in the $ee$ and $\mu\mu$ channels outside the $Z$ peak window. . . . .	104
5.4	Control plots highlighting mismodelling in the $ee$ and $\mu\mu$ channels inside the $Z$ peak window. . . . .	105
5.5	Control plots showing agreement between data and prediction in the $e\mu$ channel. .	106
5.6	Control plots showing agreement between data and prediction in the $e\mu$ channel. .	107
5.7	Jet kinematic distributions showing a slight mismodelling in the jet $p_T$ spectra in the $e\mu$ channel. . . . .	108
5.8	Lepton kinematic distributions showing agreement between data and prediction in the $e\mu$ channel. . . . .	109
5.9	Control plots showing agreement between data and prediction in the $e\mu$ channel for events with exactly two and three jets. . . . .	110
5.10	An example of the jet $p_T$ mismodelling in $e\mu$ events with exactly two and three jets. . . . .	110
5.11	Flavour fractions as a function of $p_T$ and $ \eta $ . . . . .	113
5.12	Predicted flavour composition of jets in data. . . . .	120
5.13	Distributions of the $b$ -tagging efficiency in data and in simulated $t\bar{t}$ events and the derived scale factor corrections to be applied to simulated data using LC jets with a JVF cut applied. . . . .	125
5.14	Statistical combination of multiple $b$ -tagging calibration methods [118]. . . . .	126
6.1	A leading-order $t\bar{t}H$ Feynman diagram. . . . .	130
6.2	The production cross-section multiplied by branching fraction for potential 8 TeV Higgs search channels [41]. . . . .	131
6.3	Summary of the top-pair branching fractions. . . . .	131
6.4	Top differential cross-section measured with respect to the top $p_T$ and the $t\bar{t}$ system $p_T$ [128]. . . . .	142
6.5	Prefit distributions from the single lepton $t\bar{t}H (H \rightarrow b\bar{b})$ analysis showing poor modelling of the number of selected jets in an event [13]. . . . .	143

6.6	Prefit distributions from the single lepton $t\bar{t}H (H \rightarrow b\bar{b})$ analysis showing poor modelling of the hadronic $H_T$ variable [13]. . . . .	144
6.7	The relative expected number of $t\bar{t}+b\bar{b}$ events split into categories dependent on the nature of additional $B$ -hadrons [13]. . . . .	145
6.8	The background compositions for each analysis region [13]. . . . .	147
6.9	Distribution information produced by NeuroBayes analysis macro for the $\Delta\eta_{jj}^{\max}$ variable in the $\geq 4$ jets, $\geq 4$ $b$ -jets region [135]. . . . .	152
6.10	The discrimination power between signal and background processes after training a neural network in the 3 jets, 3 $b$ -jets, $\geq 4$ jets, 3 $b$ -jets and $\geq 4$ jets, $\geq 4$ $b$ -jets analysis regions [13]. . . . .	153
6.11	The prefit distributions for each analysis region [13]. . . . .	155
6.12	Illustration of the test statistic distribution and the different areas associated with an observation when comparing to two hypotheses. . . . .	157
7.1	CTEQ6LL parton density function for $Q^2 = 500$ GeV $^2$ [155]. . . . .	165
7.2	A factorised component of the full transfer function $\mathcal{W}(\mathbf{x} \mathbf{y})$ parametrised for a $b$ -jet matched to a particle jet with a true energy of 100 GeV extracted from the KLFitter package [156]. . . . .	167
7.3	Illustration of an $s$ -channel $2 \rightarrow 2$ process via a virtual intermediate particle. . . . .	175
7.4	Overlay of the negative $\log_{10}$ summed likelihood for the signal and background hypothesis in the $\geq 4$ jets, $\geq 4$ $b$ -jets region. . . . .	182
7.5	Overlay of $D_1^{\text{Int}}$ in the $\geq 4$ jets, $\geq 4$ $b$ -jets region. . . . .	183
7.6	Overlay of $D_1^{\text{Ana}}$ in the $\geq 4$ jets, $\geq 4$ $b$ -jets region. . . . .	183
7.7	Overlay of the invariant mass of the non- $t\bar{t}$ $b$ -jets taken from the maximum signal hypothesis likelihood for the signal and background hypothesis in the $\geq 4$ jets, $\geq 4$ $b$ -jets region. . . . .	184
7.8	Comparison between data and prediction for $D_1$ in the $\geq 4$ jets, 3 $b$ -jets region. . . . .	186
7.9	Comparison between data and prediction for the invariant mass of the non- $t\bar{t}$ $b$ -jets taken from the maximum signal hypothesis likelihood in the $\geq 4$ jets, 3 $b$ -jets region. . . . .	187

7.10 Comparison between data and prediction for the invariant mass of the non- $t\bar{t}$ $b$ -jets taken from the maximum background hypothesis likelihood in the $\geq$ 4 jets, 3 $b$ -jets region. . . . .	187
7.11 Comparison between data and prediction for $D_1$ in a signal-depleted $\geq$ 4 jets, $\geq$ 4 $b$ -jets region. . . . .	188
7.12 The normalised Higgs candidate invariant mass comparing the truth-matched jet-pair with the maximum signal likelihood Higgs candidate jet-pair. . . . .	192
7.13 The normalised Higgs candidate invariant mass comparing the truth-matched jet-pair with the maximum signal hypothesis likelihood Higgs candidate jet-pair for events which were incorrectly matched. . . . .	193
7.14 The correlation between the true Higgs mass and the reconstructed Higgs mass using the maximum signal hypothesis likelihood. . . . .	193
7.15 Neural network output after the nominal training. . . . .	203
7.16 Neural network output after the nominal training with the inclusion of $D_1^{\text{Ana}}$ . . .	204
7.17 Neural network output after the nominal training with the inclusion of $D_1^{\text{Int}}$ . . .	204
7.18 $D_1^{\text{Ana}}$ after the nominal training. . . . .	205
7.19 $D_1^{\text{Int}}$ after the nominal training. . . . .	205
7.20 $D_1^{\text{Ana}}$ after the nominal training with the inclusion of $D_1^{\text{Ana}}$ . . . . .	206
7.21 $D_1^{\text{Int}}$ after the nominal training with the inclusion of $D_1^{\text{Int}}$ . . . . .	206
7.22 A $t\bar{t}H$ candidate data event demonstrating the Mercedes-like event topology. .	208
7.23 A $t\bar{t}H$ candidate data event demonstrating the back-to-back event topology. .	209
8.1 Projected expected discovery significance as a function of 13 TeV luminosity. .	212
8.2 An illustration of the colour connection between two hard partons and the pro- duction of $q\bar{q}$ -pairs as the hard partons travel away from each other based on the description in [69]. . . . .	214

## List of Tables

2.1	Table summarising the properties of the fermions and bosons [30]. . . . .	36
2.2	Standard Model cross-sections for $t\bar{t}H$ , $t\bar{t}$ and $t\bar{t}+b\bar{b}$ at 8 TeV and the branching ration of $H \rightarrow b\bar{b}$ [40, 41, 44, 45]. . . . .	40
3.1	Design requirements for the ATLAS detector [61]. . . . .	56
3.2	Summary of event rates and processing times in the ATLAS TDAQ system comparing the design expectations and the rates during the 2012 data taking period. . . . .	65
4.1	Summary of the sequential recombination algorithms used to construct jets. . . . .	82
5.1	Summary of the electron selection criteria. . . . .	96
5.2	Summary of the muon selection criteria. . . . .	97
5.3	Summary of the jet selection criteria. . . . .	98
5.4	Summary of the Monte Carlo generator and parton showering programs used to simulate the signal and main background samples for dileptonic top-pair decays. . . . .	101
5.5	Predicted and data yields in the $e\mu$ channel with EM jets. . . . .	111
5.6	Predicted and data yields in the $e\mu$ channel with LC jets. . . . .	111
5.7	Change in the predicted diboson yields in the $e\mu$ channel from changing the event generator. . . . .	112
5.8	The $b$ -tagging discriminant thresholds of each $b$ -tagging working point calibrated. . . . .	115
5.9	The dominant systematic uncertainties in the 2 jet region for the MV1 70% working point using LC jets with a JVF cut. . . . .	127
5.10	The dominant systematic uncertainties in the 3 jet region for the MV1 70% working point using LC jets with a JVF cut. . . . .	128

6.1	Expected number of $t\bar{t}H$ signal events at 8 TeV [41]. . . . .	132
6.2	Summary of lepton selection criteria. . . . .	135
6.3	Summary of jet selection criteria. . . . .	136
6.4	Summary of the baseline simulation samples used in the nominal analysis. . . .	140
6.5	Normalisation correction factors applied to rescale the expected event yield using TRF to the expected event yield using direct tagging for signal and $t\bar{t}+j$ jets backgrounds in all analysis regions. . . . .	146
6.6	Summary of the discriminating variable used in the simultaneous fit for each analysis region. . . . .	147
6.7	Prefit event yields for signal, backgrounds and data in each analysis category [13].	148
6.8	A summary of all the event variables which are used in the neural network in one or more analysis regions [13]. . . . .	150
6.9	The variables rankings provided by NeuroBayes for each of the top ten selected variables in each region where a neural network is trained [13]. . . . .	151
6.10	The observed and median expected background-only 95% confidence level upper limits on $\sigma(t\bar{t}H)$ relative to the Standard Model expectations with the 68% and 95% interval bands around the background-only expectation [13]. . . . .	158
7.1	$B$ -jet transfer function parameters used in Equation (7.4) to produce Figure 7.2.	166
7.2	The association between detector-level and parton-level objects for each permutation after all approximations have been applied. . . . .	170
7.3	The truth matching efficiency for the permutation which maximises the signal likelihood calculation in the region with exactly four $b$ -jets with no additional jets in the $t\bar{t}H$ sample. . . . .	190
7.4	The calculation efficiency for the two integration methods on truth-matched events.	191
7.5	The calculation efficiency for the two integration methods separated into $t\bar{t}H$ , $t\bar{t}+b\bar{b}$ , $t\bar{t}+c\bar{c}$ and $t\bar{t}+LF$ events in the $\geq 4$ jets, $\geq 4$ $b$ -jets region. . . . .	191
7.6	Summary of the Gaussian fits to the distributions in Figure 7.12-7.13. . . . .	194
7.7	Summary of the mean processing time of the matrix element method on events in the $\geq 4$ jets, $\geq 4$ $b$ -jets region selected using direct $b$ -tagging. . . . .	194
7.8	Neural network input variables ranked by Neurobayes with and without the inclusion of $D_1^{\text{Int}}$ . . . . .	197

7.9 Neural network input variables ranked by Neurobayes with and without the inclusion of $D_1^{\text{Ana}}$ . . . . .	198
7.10 Summary of the NeuroBayes significance for the complete set of standardised input variables used by NeuroBayes to train the neural network classifier. . . . .	199
7.11 A summary of all systematic uncertainties included in the construction of the analysis likelihood function which is maximised in a profile fit to data. . . . .	200
7.12 A comparison of the median expected upper exclusion limit, where the systematic uncertainties are profiled using the observed data in order to test the expected sensitivity of the analysis. . . . .	201
7.13 The median expected discovery significance in the three different training categories. . . . .	201
7.14 The expected and observed median upper exclusion limit where the sources of systematic error are profiled using the observed data. . . . .	202
7.15 Summary of the best-fit signal strength values in the unconditional likelihood fit for the three different neural network training categories. . . . .	202
7.16 The discovery significance and $p$ -value for the null hypothesis using the observed data. . . . .	202

# Chapter 1

## Introduction

---

*“Science demands patience.”*

— Arthur C. Clarke

---

The work in this thesis has been carried out as part of the ATLAS Collaboration during Run 1 of the LHC (2010-2012). The 8 TeV centre-of-mass energy dataset recorded during 2012 has been used for all analyses and results presented herein.

The broad aim of this work has been to identify proton-proton collisions resulting in the production of a Higgs boson decaying to two  $b$ -quarks. The focus of this thesis has been to observe the  $t\bar{t}H$  production mode where the top-quark pair decays dileptonically in order to measure the top Yukawa coupling.

The experiments located around the LHC tunnel have been extremely successful in rigorously testing the Standard Model of particle physics [1–3] at the highest centre-of-mass energies ever seen in a hadron collider. During Run 1 of the LHC, a number of high precision Standard Model cross-section measurements have been made by the ATLAS, CMS and LHCb Collaborations and a number of discoveries have been made, such as new bound quarks states [4, 5] and jet-quenching in a quark-gluon plasma [6, 7]. However, arguably the most important discovery has been of the Higgs boson [8, 9]. The existence of this particle, the missing piece of the Standard Model, has now been established well beyond the  $5\sigma$  significance required by the particle physics community to declare a discovery. The particle has been observed to be a boson

in agreement with a zero spin and a positive parity hypothesis [10, 11]. However, there are still measurements to be made of the coupling strength to other particles in order to truly understand the nature of this newly-discovered particle. In particular, the coupling strength of the Higgs boson to the top quark.

The work presented in this thesis will be organised in the following way:

- Chapter 2 introduces the theory behind the highly successful Standard Model of particle physics. The particle phenomenology which the theory describes is discussed with a focus on the role of spontaneous symmetry breaking. This leads into a discussion about the top Yukawa coupling, which is predicted to have the largest coupling strength, but is difficult to measure directly. A measurement of this coupling is important to guide the future theories which will build on the Standard Model to give predictions up to the Planck scale.
- Chapter 3 provides a concise description of the ATLAS detector, explaining the design specifications of the components which make up this 7000 tonnes device. The analyses presented in this thesis make use of all the detector systems.
- Chapter 4 explores how the ATLAS detector reconstructs particles into physics objects for use in analyses. Electrons, muons and hadronic jets are all described as they will be used in the analyses presented in this thesis.
- Chapter 5 describes an analysis designed to measure the  $b$ -tagging efficiency in the full 8 TeV dataset. This analysis was entirely carried out by the author to gain authorship within the collaboration. This work resulted in an ATLAS internal note [12]. During the course of 2012 and 2013, these efficiency measurements were used by multiple physics analyses searching for events involving  $b$ -quark production. Importantly, this included searching for the Higgs boson decaying to two  $b$ -quarks.
- Chapter 6 details the published ATLAS search for dileptonic  $t\bar{t}H$  production [13]. The author’s role in this analysis was to aid in the convergence and validation of the object and event selection within the analysis group and to carry out a study into the potential data overlap with another  $t\bar{t}H$  search. This analysis was not yet at the point where it had Standard Model sensitivity, therefore a 95% CL upper limit was set on the signal strength

of 6.7 times the Standard Model cross-section expectation. The motivation behind this analysis is to be able to measure directly the top Yukawa coupling. As detailed in Chapter 2, such a measurement would provide new information on the nature of the Standard Model and provide further evidence for the fermionic couplings of the Higgs boson.

- Chapter 7 describes an analysis technique known as the matrix element method. This technique has been used successfully, but not particularly widely, in high-energy particle physics analyses. The work presented in this chapter was carried out entirely by the author. An existing matrix element method tool was developed to perform a calculation specific to the  $t\bar{t}H$  analysis presented in Chapter 6. The purpose of this work was to attempt to improve the separation between signal and background events in the nominal analysis. Two different integration methods were studied to calculate a final likelihood-based variable. The methods were validated using data in a control region and the study was expanded to include a subset of systematic uncertainties. The effect on the sensitivity of the analysis was evaluated and found to be improved. The final conclusion drawn by this work was an improvement in the expected discovery significance in the  $t\bar{t}H$  analysis presented in Chapter 6 of 8%.
- Chapter 8 takes the final results from the  $t\bar{t}H$  analysis and the matrix element method studies and extrapolates them to the incoming 13 TeV dataset which is being collected during Run 2 of the LHC. Projections for the amount of data required to observe the  $t\bar{t}H$  process are presented with the inclusion of the matrix element method. Some final thoughts are provided on how the matrix element method results can be expanded upon to provide additional information to the  $t\bar{t}H$  analysis.

The work presented in this thesis is all in standard units. However, the majority of figures which are referenced are presented in natural units where  $c = \hbar = 1$ .

## Chapter 2

# Theoretical Motivation

---

*“The colours red, blue and green are real. The colour yellow is a mystical experience shared by everybody.”*

— Tom Stoppard, Rosencrantz and Guildenstern Are Dead

---

This chapter details the current status of the Standard Model of particle physics and includes details on the relationship between the top quark and the Higgs boson. The Standard Model is a highly successful theory which is the cornerstone of modern particle physics and has provided countless theoretical predictions that have been proven by experimental observations. This chapter has been written using information in Reference [14–18] to guide the correct mathematical descriptions.

An introduction to the Standard Model is presented in Section 2.1 where some historical details are provided. A short discussion on local gauge invariance is given in Section 2.2. This is followed by a description of the fundamental particles and the fundamental forces which are described by the Standard Model in Section 2.3-2.4. An explanation of spontaneous symmetry breaking is provided through worked examples in Section 2.5 in order to aid an understanding of the phenomenology of the Higgs mechanism and its role in the Standard Model which are presented in Section 2.6. Some focus is provided on the importance of coupling between the top quark and the Higgs boson in Section 2.7, the measurement of which is the desired outcome from an observation of the  $t\bar{t}H$  process. This chapter finally concludes with a summary of recent

Higgs physics results from the LHC in Section 2.8.

## 2.1 Introduction

The Standard Model of particle physics is the theoretical model which predicts cross-sections of elementary particle interactions, predicts particle decay rates, provides a classification of all observed particles and a mechanism to generate massive particles [1–3]. It is the culmination of decades of work which started in the 1930s with the conception of quantum mechanics and subsequently quantum field theory.

A theoretical model is expressed in a compact form using the Lagrangian density. In classical field theory, the Lagrangian density is the kinetic energy minus the potential energy in a system:  $\mathcal{L} = \mathcal{T} - \mathcal{V}$ . In quantum field theory, a system is comprised of a number of quantum fields. Each field has an associated free-field Lagrangian (associated with the kinetic term,  $\mathcal{T}$ ) and coupled fields have an interaction Lagrangian (associated with the potential term,  $\mathcal{V}$ ). The sum of these terms constructs the final model:  $\mathcal{L} = \sum_i (\mathcal{L}_i^{\text{free}} + \sum_{j \neq i} \mathcal{L}_{ij}^{\text{int}})$ .

The Standard Model is a unified, self-consistent, Lorentz invariant quantum field theory which has been constructed mathematically to transform under local gauge symmetries. Invariance under local gauge symmetries ensures that the theory is fully renormalisable, which allows for predictions to be made with perturbation theory. Transforming consistently under non-Abelian local gauge symmetries requires particles in the theory to be massless which does not agree with observations. The theory therefore provides a mechanism to give mass terms to vector fields and associated mass terms to fermion fields. The mechanism for generating mass terms is known as spontaneous symmetry breaking and it resolves the problem of imposing a local gauge symmetry with massive particles. The driving motivation behind the use of symmetry groups to describe physical processes is the combined assumption and desire that the physical world we observe is governed by simple symmetries in Nature.

Observations of the physical world have led us to believe that Nature has consistent laws regardless of the position in space. It has been shown that imposing a theory be Lorentz invariant and locally gauge invariant naturally gives rise to renormalisability [19, 20]. This allows the theory to provide predictions at all orders of perturbation theory, and with the correct selection of gauge symmetries, one can make predictions which are consistent with data. The imposition

of a symmetry called local gauge invariance gives rise to gauge bosons which mediate interactions between fields carrying the gauge charge [21]. The theory of electromagnetism was well-established before it was observed to exhibit local gauge invariance. Once this was known, it became a natural extension to try to create theories which described other fundamental forces with this particular behaviour and therefore predict the existence of new intermediate gauge bosons.

The use of group theory was borne out of the observations of patterns in the known particles. In the mid-1960s, patterns in hadronic particles led to the introduction of the quark model (which at the time comprised of three light quarks - up, down and strange) by Gell-Mann and Zweig [22, 23]. Patterns seen in the lepton families motivated the inclusion of a fourth quark (charm) and subsequent work by Greenberg, Han and Nambu [24, 25] introduced the concept of a colour charge (analogous at the time to the well-known electric charge).

Quantum electrodynamics (QED), which was conceptualised in the 1930s and developed by Feynman [26], had been established as a successful theory of electromagnetism, but in the 1950s and 1960s, theorists were using the concept of local gauge invariance to attempt to create theories which could explain experimental data. QED can be described as a  $U(1)$  gauge theory, but it was not conceptualised with this feature in mind.

The modern form of the Standard Model follows from the work of Yang and Mills [21] who in 1954 created a gauge theory based on the  $SU(2)$  symmetry group of isotropic spin conservation. One by-product of this theory was the production of self-interactions of gauge bosons, owing to the non-Abelian nature of the symmetry group. In particular it was the symmetry group used to formulate the theory which controlled the behaviour of the interactions.

The final form of the Standard Model is described as an  $SU(3) \otimes SU_L(2) \otimes U_Y(1)$  gauge theory, where  $SU(3)$  is the symmetry group of the strong nuclear force and  $SU_L(2) \otimes U_Y(1)$  is the symmetry group describing the electroweak interactions.

## 2.2 Local Gauge Invariance

The Standard Model is built upon the premise of local gauge invariance. Any field which transforms under a gauge symmetry carries an associated gauge charge. This charge can be thought of as fixed throughout space-time, in which case the theory is said to have a global gauge sym-

metry. If the charge is allowed to vary depending on the position of the field in space-time then the theory is said have a local gauge symmetry.

The Lagrangian of a free real massless scalar field which transforms under a  $U(1)$  gauge symmetry can be written as

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)^2, \quad (2.1)$$

where the field transforms as  $\phi \rightarrow \phi e^{ie\alpha}$ . It is trivial to observe that this Lagrangian is therefore invariant under a global phase change.

Mathematically, the transformation from a global gauge symmetry to a local gauge symmetry is described by allowing the gauge charge  $\alpha$  to become  $\alpha(x_\mu)$ . This space-time dependence becomes relevant in the kinetic term of the Lagrangian as the transformed field will have an additional product which will be affected by the partial derivative. Without alteration, the Lagrangian will not be locally gauge invariant which is an undesirable feature of a theory.

In order to keep a Lagrangian invariant, the partial derivative is transformed into a covariant derivative. The covariant derivative acts to cancel out the additional terms generated by differentiating the gauge charge. The covariant derivative has to introduce a new vector boson field to the Lagrangian in order to keep the system invariant. This transformation is written as:  $\partial_\mu \rightarrow \mathcal{D}_\mu = (\partial_\mu - ieA_\mu)$ , where  $A_\mu$  is a vector field. The new vector field also needs to transform covariantly in order to ensure the Lagrangian remains invariant. This is given as  $A_\mu \rightarrow A'_\mu = A_\mu + ie\partial_\mu(\alpha(x_\mu))$ .

One can therefore observe that requiring a system to be invariant under a local gauge symmetry naturally gives rise to a gauge boson in the theory, for which it is not possible to write a mass term without explicitly breaking the gauge symmetry.

## 2.3 The Fundamental Particles of the Standard Model

The fundamental particles observed in Nature can all be grouped into two categories: fermions and bosons. The fermions have half-integer spin and the bosons have integer spin. There are three generations of particles observed in the fermionic sector. Each generation has the same fundamental properties with the exception that their masses are different. A summary of the particle properties can be found in Table 2.1.

A fermion field can be decomposed into left-handed and right-handed chirality eigenstates

using the projection operators,  $P_{R/L} = (1 \pm \gamma_5)/2$ . In the massless limit, these states coincide with helicity eigenstates which are defined as the projection of the spin of a particle onto its momentum vector. A massless right-handed particle will have its spin aligned with its momenta and a massless left-handed particle will have its spin anti-aligned with its momenta.

The chiral nature of the weak force which will be described in Section 2.4.3 leads to a natural categorisation of the lepton family which consists of a left-handed doublet,  $(e v)_L$ , and a right-handed singlet,  $e_R$ . There are no right-handed neutrinos in the Standard Model.

Within the group of fermions, there are those which interact with gluons and those that do not. These groups are termed quarks and leptons respectively. Each lepton generation contains a particle which carries integer electric charge (electron) and one which is electrically neutral (neutrino). Each quark generation contains an up-type and down-type quark. The up-type has an electric charge of  $+\frac{2}{3}$  and the down-type has an electric charge of  $-\frac{1}{3}$ . The quarks also carry an additional charge known as colour. Colour has a property known as confinement which prevents the observation of a coloured particle. Instead, bound colourless states of quarks known as hadrons are formed. Hadrons can be mesons, which are composed of a quark anti-quark pair ( $q\bar{q}$ ), and baryons which are composed of a quark or anti-quark triplet ( $qqq$  or  $\bar{q}\bar{q}\bar{q}$ ).

The fermions can be concisely written using their weak isospin multiplets,

$$\begin{aligned} & \left( \begin{array}{c} e \\ v_e \end{array} \right)_L, \quad \left( \begin{array}{c} \mu \\ v_\mu \end{array} \right)_L, \quad \left( \begin{array}{c} \tau \\ v_\tau \end{array} \right)_L, \\ & \left( \begin{array}{c} u \\ d' \end{array} \right)_L, \quad \left( \begin{array}{c} c \\ s' \end{array} \right)_L, \quad \left( \begin{array}{c} t \\ b' \end{array} \right)_L, \\ & e_R, \quad \mu_R, \quad \tau_R, \\ & u_R, \quad c_R, \quad t_R, \\ & d_R, \quad s_R, \quad b_R, \end{aligned}$$

where each quark listed is a colour triplet of the form  $q = (q_r \ q_g \ q_b)$  and  $d', s'$  and  $b'$  are rotational combinations of  $d$ ,  $s$  and  $b$  using the CKM matrix which is discussed in Section 2.4.3.

The group of bosons consists of five gauge bosons and one scalar boson. The gauge bosons are the force carriers of the Standard Model and they mediate between particles which carry their particular gauge charge. The gauge bosons consist of three weak bosons ( $W^\pm, Z^0$ ), the photon ( $\gamma$ ) and the gluon ( $g$ ). The additional scalar boson is the Higgs boson which interacts with

any massive particle.

## 2.4 The Fundamental Forces of the Standard Model

The Standard Model unifies three out of the four known forces: the electromagnetic force, the strong nuclear force and the weak nuclear force. Gravity is described successfully by the theory of general relativity but this has so far resisted attempts to be described successfully in the language of quantum field theory. There are quantum field theories which postulate the existence of a gravitational gauge boson, the graviton, but such a particle has not been observed. Without an observation to give direction to possible theories, there is no way of currently knowing how to combine the theories of general relativity and quantum field theory into a single unified theory.

### 2.4.1 Quantum Electrodynamics

The electromagnetic force is mediated by massless photons which propagate over an infinite range. The photon can interact with any particle which carries an electric charge and it conserves the flavour of a particle. The lowest order interaction vertex of quantum electrodynamics (QED) is between a photon and two electrically charged fermions, as shown in Figure 2.1. In perturbation theory the strength of the QED interaction is dictated by the parameter  $\alpha$  which runs with the interaction energy,  $q^2$ . For  $q^2 \ll m_{W,Z}^2$ ,  $\alpha \approx \frac{1}{137}$  and for  $q^2 \approx m_Z^2$ ,  $\alpha \approx \frac{1}{127}$ . Each QED vertex contributes an order of  $\alpha$  to a cross-section calculation in perturbation theory. The dominant contribution to a QED cross-section calculation comes from the diagram with the lowest order of  $\alpha$ .

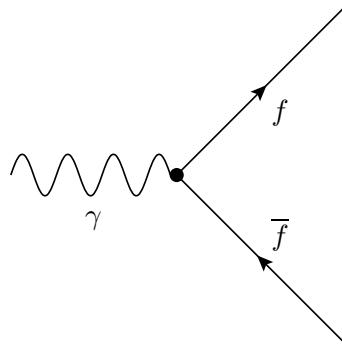


Figure 2.1: The interaction vertex of QED.

### 2.4.2 Quantum Chromodynamics

The strong nuclear force is described by the theory of quantum chromodynamics (QCD). The force is mediated by gluons which interact with any particle carrying colour charge. QCD is very different from the other two forces because of the phenomenology of the colour charge. Colour is the  $SU(3)$  gauge charge which comes in three flavours: red, green and blue. There are eight gluons which exist in QCD, each carrying a different colour and anti-colour charge, or a linear combination of colour anti-colour pairings. As  $SU(3)$  is a non-Abelian group, the fundamental interactions include self-coupling of the gluons, as shown in Figure 2.2.

QCD gives rise to two phenomena - colour confinement and asymptotic freedom.

Colour confinement describes how the energy in the gluon field increases as two coloured objects are separated. As they travel away from each other, the energy in the field will become large enough to produce a quark anti-quark pair. This showering produces particles which are collimated in the direction of flight of the original coloured objects. Once the energy in the gluon field reduces below a threshold, non-perturbative hadronisation effects take place which ensure there are only bound colourless states in the observable final state.

Asymptotic freedom refers to the running coupling of the strong force,  $\alpha_s(r)$ . At small distances between coloured objects, the strong force becomes very weak. This allows the treatment of high energy collisions to consider quarks as individual fields interacting, rather than taking into account the gluon field connecting all coloured objects at the point of production. As the energy of the interacting system evolves, the strong force becomes more dominant. This makes the calculation of QCD processes difficult as a perturbative expansion is not a valid treatment. Higher order processes can have a large contribution to the total cross-section due to the running of  $\alpha_s$  with energy.

### 2.4.3 The Weak Force

The weak nuclear force acts on very short length and time scales. The weak nature of the force is due to the large masses of the intermediate bosons as the weak coupling constant is comparable to the electromagnetic coupling at the weak scale. The weak bosons interact with left-handed fermions (right-handed anti-fermions) carrying a weak isospin quantum number and has two types of interactions known as charged current interactions and neutral current interactions.

The charged current interaction is mediated by the  $W$ -bosons and enables flavour changing

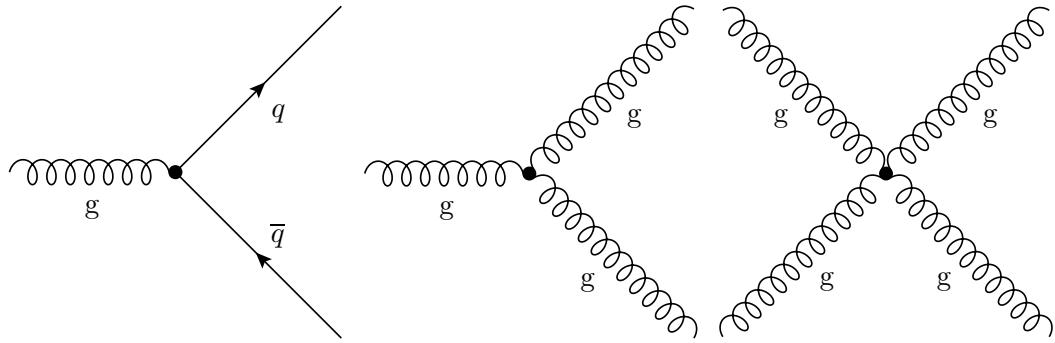


Figure 2.2: The interaction vertices of QCD.

interactions between two fields. The charged current interaction is responsible for radioactive beta decays in atoms as this process enables neutrons to decay to protons through  $d \rightarrow u$  decays. The neutral current interaction is mediated by the Z-boson which acts much like the photon. There are no flavour changing neutral currents allowed without requiring higher order loop processes. The requirement for higher order perturbation expansions means that these processes are highly suppressed and termed rare electroweak decays. The fundamental interaction vertices are shown in Figure 2.3. The weak force is described by a non-Abelian gauge symmetry group which permits self-interactions of the gauge bosons.

It was discovered in 1956 by Wu [27] that the radioactive decay of supercooled Cobalt-60 in a magnetic field violated parity symmetry by studying the angular distributions of electrons emitted by the decays of two different spin states of Cobalt-60 which were aligned parallel and anti-parallel to a magnetic field. This parity violation occurs because the weak bosons only interact with specific particle states: left-handed fermions and right-handed anti-fermions. For the case of massless neutrinos, this means there are no interactions with right-handed neutrinos or with left-handed anti-neutrinos as there is no Lorentz transformation possible which will allow the handedness of a neutrino to be inverted.

Whilst parity violation occurs in the lepton sector of the Standard Model, the combined symmetry of charge-parity (CP) is conserved. This is not the case in the quark sector due to the mixing of flavour and mass eigenstates. The emission of a virtual W-boson allows an up-type quark to change into any down-type quark and vice-versa. The probabilities are measured through flavour physics and expressed in the Cabibbo-Kobayashi-Maskawa (CKM) matrix [28–

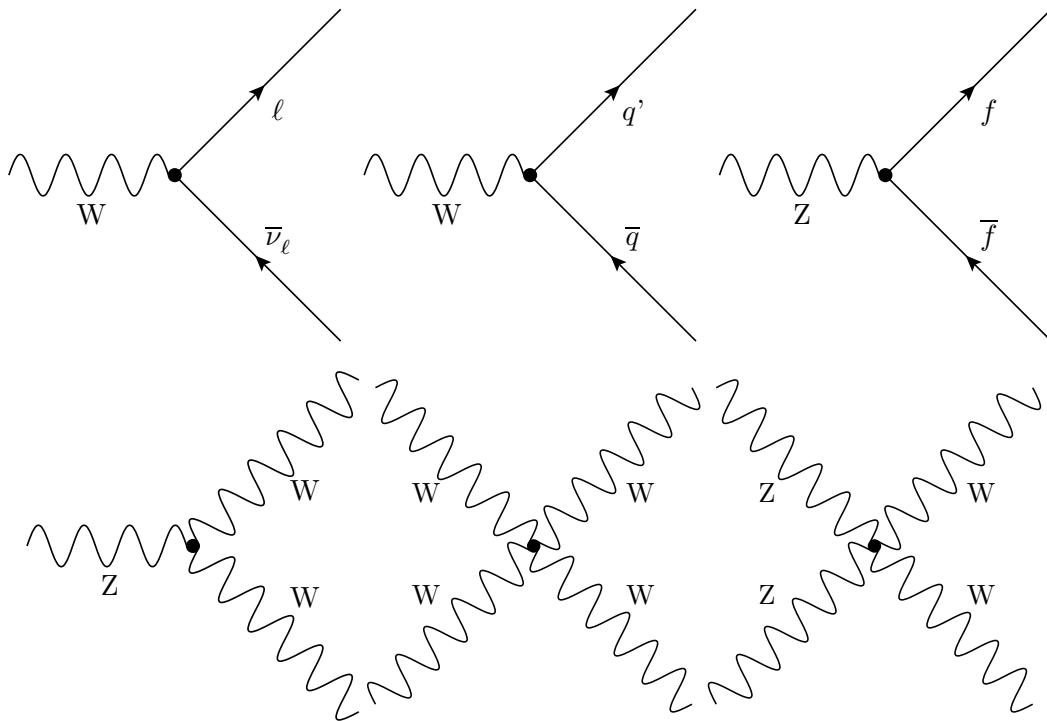


Figure 2.3: The interaction vertices of the weak force.

30],

$$\begin{aligned}
 V_{\text{CKM}} &= \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix}, \\
 &= \begin{pmatrix} 0.97427 \pm 0.00014 & 0.22536 \pm 0.00061 & 0.00355 \pm 0.00015 \\ 0.22522 \pm 0.00061 & 0.97343 \pm 0.00015 & 0.0414 \pm 0.0012 \\ 0.00886^{+0.00033}_{-0.00032} & 0.0405^{+0.0011}_{-0.0012} & 0.99914 \pm 0.00005 \end{pmatrix}, \quad (2.2)
 \end{aligned}$$

where the matrix gives the magnitude of the complex elements. The components of the CKM matrix can be parametrised in a number of different ways, but in each case, a  $3 \times 3$  unitary matrix is required to have 4 variables: 3 mixing angles and 1 complex phase. If this complex phase is non-zero then CP-violation is permitted in the quark sector and this has been observed to occur.

## 2.5 Spontaneous Symmetry Breaking

It was shown by Goldstone that when a global symmetry of a system is spontaneously broken, such that the system is no longer invariant under such a symmetry transformation, massless bosons (known as Goldstone bosons) will be produced in the theory [31]. The discovery of what would become known colloquially as the Higgs mechanism was actually the work of six theorists who developed the work on spontaneous symmetry breaking in the theory of superconductivity [32, 33]. Those theorists are Brout and Englert [34], Higgs [35–37], and Guralnik, Hagan and Kibble [38].

The production of Goldstone bosons was originally a limitation in the application of spontaneous symmetry breaking to particle physics because it was assumed in the 1960s that any massless particles should already have been discovered. It was Higgs who realised that if a local gauge symmetry was broken instead of a global one, the Goldstone bosons would not need to manifest themselves physically as new particles. Instead they could be gauged away (with an appropriate unitary transformation) as additional degrees of freedom in the gauge fields of the theory.

The key discovery here was that the massless gauge bosons associated with the  $SU(2)$  gauge symmetry would become massive vector bosons and the scalar field which was added to break the local gauge symmetry would generate a physical mass term in the Lagrangian. This implies that if spontaneous symmetry breaking of a local gauge symmetry was responsible for generating massive vector bosons, there would be a new massive scalar boson generated in the theory.

In order to better visualise the Higgs mechanism which breaks the electroweak gauge symmetry  $SU_L(2) \otimes U_Y(1)$  to the electromagnetic gauge symmetry  $U_{EM}(1)$ , two examples of spontaneous symmetry breaking are presented for the simpler case of breaking a  $U(1)$  gauge symmetry. The example is first presented using a global gauge symmetry in Section 2.5.1 which is subsequently followed by breaking a local gauge symmetry in Section 2.5.2. These examples provide the form of the potential term in the Lagrangian which has become known as the Higgs potential.

### 2.5.1 Spontaneous symmetry breaking of a global $U(1)$ symmetry

The Lagrangian of a complex scalar field in a potential is defined as

$$\mathcal{L} = \frac{(\partial_\mu \phi)^2}{2} - \frac{\mu^2 \phi^2}{2} - \frac{\lambda \phi^4}{4}, \quad (2.3)$$

where potential term is defined as  $V = \frac{\mu^2\phi^2}{2} + \frac{\lambda\phi^4}{4}$ . If  $\mu^2 < 0$  and  $\lambda > 0$  then this potential has the form shown in Figure 2.4 where the ground-state of the field takes on a non-zero value at the minima. This Lagrangian will be invariant under a  $U(1)$  transformation,  $\phi \rightarrow e^{ie\alpha}\phi$ , as the Lagrangian only has a dependence on squared factors of  $\phi$  for which the phase factors cancel out. A complex field can be expressed as the sum of two real fields in the form  $a + ib$ , where  $a$  and  $b$  are real numbers. Using this, we can express  $\phi = \phi_1 + i\phi_2$  where  $\phi_1$  and  $\phi_2$  are real fields and  $|\phi|^2 = \phi^*\phi = (\phi_1 - i\phi_2)(\phi_1 + i\phi_2) = \phi_1^2 + \phi_2^2$ .

The value of the field at the minima of the potential is known as the vacuum expectation value and it can be calculated by minimising the Lagrangian with respect to the field,

$$|\phi_{\min}|^2 = \phi_1^2 + \phi_2^2 = \frac{-\mu^2}{\lambda} = v^2. \quad (2.4)$$

The point at  $\phi = 0$  is only a minima if  $\mu^2 > 0$  which would mean the Lagrangian contains a real massive scalar field. It should be noted that in the case that  $\mu^2 < 0$ , the vacuum expectation value is real.

The Lagrangian in Equation (2.3) can be rewritten explicitly in terms of real fields,

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi_1)^2 + \frac{1}{2}(\partial_\mu\phi_2)^2 - \frac{1}{2}\mu^2(\phi_1^2 + \phi_2^2) - \frac{1}{4}(\phi_1^2 + \phi_2^2)^2, \quad (2.5)$$

where  $(\partial_\mu\phi)^2 = (\partial_\mu\phi^*)(\partial^\mu\phi) = \partial_\mu(\phi_1 - i\phi_2)\partial^\mu(\phi_1 + i\phi_2)$  is used to simplify the partial derivative terms.

This Lagrangian gives a continuum of positions where the field can acquire a vacuum expectation value and break the symmetry, as  $\phi_1$  and  $\phi_2$  are only constrained to have the sum of their squares add up to be  $v^2$ . One of the simplest choices is to make the field real and select  $\phi_1 = v$  and  $\phi_2 = 0$ .

It is possible to perform a perturbative expansion about the chosen ground-state using

$$\phi_{\min} \rightarrow \phi(x) = \frac{1}{\sqrt{2}}(v + \eta(x) + i\rho(x)) \quad (2.6)$$

This expansion uses two real fields (or equivalently one complex field) to expand about the potential in the  $(\phi_1, \phi_2)$  plane. These values can be substituted back into the Lagrangian to

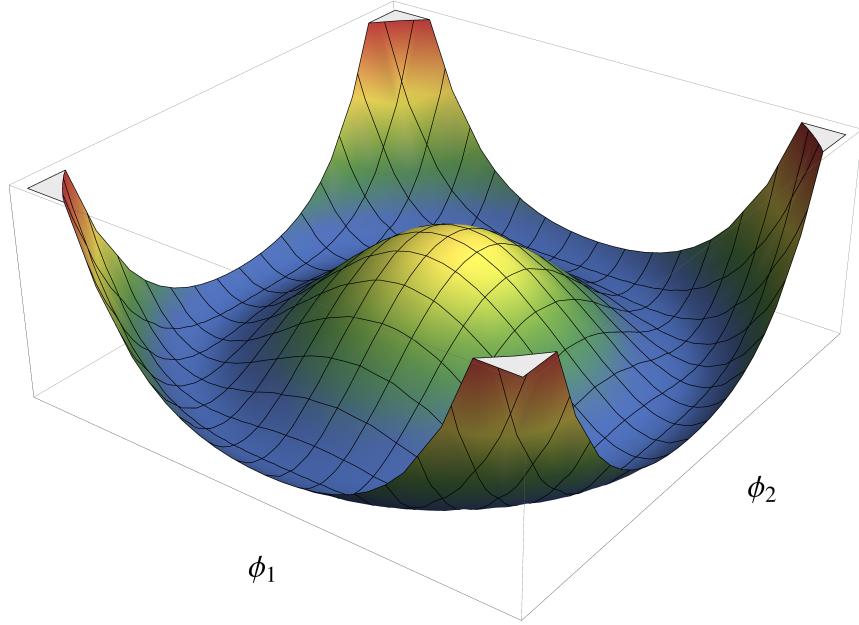


Figure 2.4: The potential well in a  $\phi^4$  theory where  $V(\phi) = \frac{\mu^2\phi^2}{2} + \frac{\lambda\phi^4}{4}$  with  $\mu^2 < 0$  and  $\lambda > 0$ . The point at  $\phi = 0$  is no longer the ground state (which it would be if  $\mu^2 > 0$ ). Instead the system will select one of the points in the minima when the symmetry is spontaneously broken where  $\phi_1^2 + \phi_2^2 = v^2$ .

examine the system close to the minima:

$$\mathcal{L}' = \frac{1}{2}(\partial_\mu \eta)^2 + \frac{1}{2}(\partial_\mu \rho)^2 + \mu^2 \eta^2 + O(\eta^3) + O(\rho^3) + \text{constants.} \quad (2.7)$$

This new Lagrangian describes the system close to the ground-state after spontaneous symmetry breaking has occurred. Whilst the Lagrangian of the system possesses a gauge symmetry, the chosen ground-state after spontaneous symmetry breaking does not and perturbations about the minima are not gauge invariant.

By examining the form of the new Lagrangian, the scalar field,  $\eta$ , gains a kinetic and a real mass term of  $\sqrt{-2\mu^2}$ . The scalar field,  $\rho$  has a kinetic term but does not have a mass term. Therefore, breaking the symmetry of the system has introduced a massive and a massless real scalar field. This massless field is a Goldstone boson which has manifested from the breaking of a global continuous symmetry [31].

### 2.5.2 Spontaneous symmetry breaking of a local $U(1)$ symmetry

The Higgs Model provides a framework for spontaneous symmetry breaking of a local gauge symmetry with an additional gauge transformation to reinterpret the Lagrangian. The Lagrangian

for a scalar field interacting with a massless vector field is

$$\mathcal{L} = (\partial_\mu + ieA_\mu)\phi^*(\partial^\mu - ieA^\mu)\phi - \mu^2\phi^*\phi - \lambda(\phi^*\phi)^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \quad (2.8)$$

Here the term  $(1/4)F_{\mu\nu}F^{\mu\nu}$  is introduced as the kinetic term for the massless vector field,  $A_\mu$ .  $F_{\mu\nu}$  is the field strength tensor which is in this contracted form in the Lagrangian is invariant under a  $U(1)$  phase transformation. In the case of  $\mu^2 > 0$ , this Lagrangian is the QED Lagrangian for a charged scalar field. However, as before, if  $\mu^2$  is less than zero, a potential with a non-zero vacuum expectation value is produced and spontaneous symmetry breaking can occur.

It is possible to express the field at the minima as a perturbative expansion where  $\phi(x)_{\min} \rightarrow \phi(x) = v + \eta(x) + i\rho(x)$ . However, instead of having a linear expression, one can choose to write the perturbation as an exponential function expressing the system in unitary gauge:

$$\phi(x) \rightarrow \phi'(x) = \frac{1}{\sqrt{2}}(v + \eta(x))e^{\frac{i\rho(x)}{v}}. \quad (2.9)$$

Using the exponential expansion ( $e^{ax} = 1 + ax + \frac{(ax)^2}{2!} + \frac{(ax)^3}{3!} + \dots$ ) then we recover the linear expression for the expansion about the minima to first order. However, by choosing to apply a gauge transformation to the scalar field, the vector field needs to be written in a covariant form:

$$A_\mu \rightarrow A'_\mu = A_\mu + \frac{1}{e}\partial_\mu\left(\frac{\rho(x)}{v}\right). \quad (2.10)$$

There is a free choice in how to express the broken symmetry, and one can simplify the mathematics by setting  $\rho(x) = 0$  so that we get an entirely real field  $\phi(x) = (1/\sqrt{2})(v + H(x))$  where we are using  $\eta(x) = H(x)$  to explicitly state that this will play the role of the Higgs field.

The Lagrangian can be re-expressed using these fields expanded to describe the system close to the ground-state,

$$\mathcal{L}' = \frac{1}{2}(\partial_\mu H)^2 - \lambda v^2 H^2 + \frac{1}{2}e^2 v^2 A_\mu A^\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \text{additional terms}. \quad (2.11)$$

The Goldstone bosons which manifest themselves when a global symmetry is spontaneously broken do not appear in this Lagrangian. This is because they have become part of the vector field. By absorbing the Goldstone mode into the description of the vector field under the unitary gauge transformation, the vector field has gained an additional degree of freedom and the La-

grangian contains a mass term. Including such a term initially would have explicitly broken the local gauge symmetry of the Lagrangian. The mass of this vector boson is linearly dependent on the vacuum expectation value,  $m_{A_\mu} = ev$ . In addition to giving mass to the vector field, the scalar field introduced to the theory gains a mass term.

## 2.6 The Higgs Mechanism and Electroweak Unification

In the Standard Model, the Higgs mechanism is used to break the  $SU_L(2) \otimes U_Y(1)$  gauge symmetry to the  $U_{EM}(1)$  gauge symmetry observed in Nature, where Y refers to hypercharge and EM refers to electric charge. In this model, there are three massless  $SU_L(2)$  gauge fields and one massless  $U_Y(1)$  gauge field. The fields corresponding to the  $SU(2)$  group are  $W_\mu^{1,2,3}$  and the field corresponding to the  $U(1)$  group is  $B_\mu$ .

The formalism of spontaneous symmetry breaking presented in Section 2.5 can now be extended to describe the breaking of an  $SU(2) \otimes U(1)$  local gauge symmetry with a complex scalar doublet using a potential of the form  $V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2$ .

To initiate the symmetry breaking, a complex scalar doublet is introduced to the Lagrangian which transforms as a singlet under  $SU(3)$  and  $U_Y(1)$  and as a doublet under  $SU_L(2)$ ,  $\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$ . When spontaneous symmetry breaking occurs, the vacuum expectation value of the scalar field becomes  $v$  and an expansion can be performed about the chosen ground-state. The location of the ground-state of  $\Phi$  is arbitrary as was described in the previous examples. The standard choice in describing the Higgs mechanism is for  $\phi^+ = 0$  and for  $\Im(\phi^0) = 0$  so that only the real part of  $\phi^0$  takes on the non-zero vacuum expectation value,  $\Re(\phi^0) = v$ . The covariant derivative of the ground-state is then defined as

$$D_\mu \Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \partial_\mu H \end{pmatrix} - \frac{i}{2\sqrt{2}} \begin{pmatrix} g_2 W_\mu^3 + g_1 B_\mu & g_2 W_\mu^1 - ig_2 W_\mu^2 \\ g_2 W_\mu^1 + ig_2 W_\mu^2 & -g_2 W_\mu^3 + g_1 B_\mu \end{pmatrix} \begin{pmatrix} 0 \\ v + H \end{pmatrix}. \quad (2.12)$$

The expansion of the kinetic term for the scalar field,  $D_\mu \Phi D^\mu \Phi$  is defined as

$$\begin{aligned} -(D_\mu \Phi D^\mu \Phi) = & -\frac{1}{2} \partial_\mu H \partial^\mu H \\ & - \frac{1}{8} (v + H)^2 g_2^2 (W_\mu^1 - iW_\mu^2)(W^{1\mu} + iW^{2\mu}) \\ & - \frac{1}{8} (v + H)^2 (-g_2 W^{3\mu} + g_1 B^\mu)(-g_2 W_\mu^3 + g_1 B_\mu). \end{aligned} \quad (2.13)$$

This expansion shows the unification of the electroweak force and the generation of the vector field mass terms. The scalar potential introduced in Section 2.5 can also be expanded to generate the self-interaction and mass terms of the Higgs field:

$$\begin{aligned} -V &= -\frac{\lambda}{4}[(v+H)^2 - \frac{\mu^2}{\lambda}]^2 \\ &= -\frac{\lambda}{4}(2vH + H^2)^2 \\ &= -(\lambda v^2 H^2 + \lambda v H^3 + \frac{\lambda}{4} H^4). \end{aligned} \quad (2.14)$$

Studying the expansion of these two equations reveals the generated mass terms. The expansion of  $(v+H)^2$  in Equation (2.13) gives three terms,  $v^2 + H^2 + 2vH$ , which provide mass terms and couplings between the vector field and the Higgs field. The mass and the self-couplings of the Higgs are given by the expansion of the potential in Equation (2.14). Using the general form of a real scalar field, the mass of the Higgs field is given by  $m_H^2 = 2\lambda v^2 = -2\mu^2$ .

The fields  $W_\mu^1$  and  $W_\mu^2$  appear in the Lagrangian with exactly the same mass  $m_W^2 = \frac{1}{4}v^2 g_2^2$ . These two fields are combined into a conjugated pair of complex fields,  $W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2)$ . This new complex doublet is invariant under the remaining  $U_{\text{EM}}(1)$  symmetry resulting in two charged massive vector fields,  $W^+$  and the  $W^-$ .

The remaining vector fields appear in the combination  $-g_1 B_\mu + g_2 W_\mu^3$  which can be normalised by a factor of  $\sqrt{g_1^2 + g_2^2}$ . This field can be expressed using the Weinberg or weak mixing angle,  $\theta_W$ , as  $Z_\mu = W_\mu^3 \cos \theta_W - B_\mu \sin \theta_W$ . Written in this form, a mass term can be extracted for the  $Z$  field of  $m_Z = \frac{1}{2}v\sqrt{g_1^2 + g_2^2}$ .

An additional field needs to be defined using a combination of  $W_\mu^3$  and  $B_\mu$  which is orthogonal to  $Z_\mu$ . This can be defined as  $A_\mu = W_\mu^3 \sin \theta_W + B_\mu \cos \theta_W$ . This field does not appear with a Higgs coupling term which means there cannot be a mass term generated for this field through spontaneous symmetry breaking. This field therefore represents the unbroken symmetry associated with  $U_{\text{EM}}(1)$  and  $A_\mu$  is the massless photon. In this way, the weak and electromagnetic forces are unified through the Higgs mechanism to create the electroweak force.

The electromagnetic charge,  $e$ , can be extracted from the Higgs mechanism in terms of  $g_1$  and  $g_2$  as  $e = g_1 g_2 / \sqrt{g_1^2 + g_2^2}$ . The mass of the  $W$  and the  $Z$  bosons are coupled through the expression,  $m_W = m_Z \cos \theta_W$ . These expressions can be combined to calculate the vacuum expectation value using the expression  $v = 2m_W \sin \theta_W / e$  which is calculated to be approximately

246 GeV.

A summary of  $SU_L(2) \otimes U_Y(1)$  couplings are provided in Table 2.1 for all the left- and right-handed fermions. The  $U_{EM}(1)$  coupling and the masses are also shown for the fermions and bosons of the Standard Model.

Table 2.1: Table summarising the properties of the fermions and bosons [30]. The fermion listings include the  $SU_L(2)$  and  $U_Y(1)$  couplings for the left and right handed fermions. The rows are split into the  $SU_L(2)$  doublets. After symmetry breaking, the electric charge is defined as  $Q_{EM} = Y + T^3$  and is identical for both left- and right-handed fermions.

Fermion	Mass ( $\text{GeV}/c^2$ )	$Q_{EM}$	$Y_L$	$T_L^3$	$Y_R$	$T_R^3$
e	0.000511	-1	$-\frac{1}{2}$	$-\frac{1}{2}$	-1	0
$\nu_e$	$\approx 0$	0	$-\frac{1}{2}$	$\frac{1}{2}$	0	0
$\mu$	0.105	-1	$-\frac{1}{2}$	$-\frac{1}{2}$	-1	0
$\nu_\mu$	$\approx 0$	0	$-\frac{1}{2}$	$\frac{1}{2}$	0	0
$\tau$	1.777	-1	$-\frac{1}{2}$	$-\frac{1}{2}$	-1	0
$\nu_\tau$	$\approx 0$	0	$-\frac{1}{2}$	$\frac{1}{2}$	0	0
$u$	0.0023	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{2}{3}$	0
$d$	0.0048	$-\frac{1}{3}$	$\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{1}{3}$	0
$c$	1.275	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{2}{3}$	0
$s$	0.095	$-\frac{1}{3}$	$\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{1}{3}$	0
$t$	173.3	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{2}{3}$	0
$b$	4.18	$-\frac{1}{3}$	$\frac{1}{6}$	$-\frac{1}{2}$	$-\frac{1}{3}$	0

Boson	Mass ( $\text{GeV}/c^2$ )	$Q_{EM}$
Gluon	0	0
Photon	0	0
$W^+$	80.4	1
$W^-$	80.4	-1
Z	91.2	0
Higgs	125.1	0

### 2.6.1 Yukawa Coupling

The Higgs mechanism is only introduced to create mass terms for the massive vector bosons associated with the weak force. A priori, there is no reason to assume that this field also creates

mass terms for the other massive particles in the Standard Model Lagrangian. The reason that typical Dirac mass terms cannot be written down in the Lagrangian is because of the weak interaction. The parity violating coupling means that a right-handed fermion field transforms differently to a left-handed fermion field and the mass term,

$$m\bar{\Psi}\Psi = m(\bar{\Psi}_L\Psi_R + \bar{\Psi}_R\Psi_L), \quad (2.15)$$

is not gauge invariant.

However, it is possible to create mass terms for fermions through a Yukawa coupling. These couplings have the form  $\bar{\Psi}\phi\Psi$  and are invariant under  $SU_L(2) \otimes U_Y(1)$ . The Yukawa term allows the coupling of left- and right-handed fermion fields and is written for the first generation of leptons as

$$\begin{aligned} \bar{\Psi}_L\phi\Psi_R &= -\frac{1}{\sqrt{2}}Y_e(\bar{\nu}_L, e_L) \begin{pmatrix} 0 \\ v+H \end{pmatrix} e_R, \\ &= -\frac{1}{\sqrt{2}}Y_e\bar{e}_L(v+H)e_R. \end{aligned} \quad (2.16)$$

Expanding the resultant terms gives a mass term of the form  $Y_e v \bar{e}_L e_R$ , where  $Y_e$  is the electron Yukawa coupling. The Yukawa coupling is independent for every fermion field and linearly dependent on mass. Using the form of the fermionic mass term, the mass of a fermion is given by  $m_f = \frac{1}{\sqrt{2}}Y_f v$ . The remaining coupling between the fermionic and the Higgs fields has the interesting feature that the strength of the interaction is determined by the Yukawa coupling. Hence, the Higgs preferentially decays to the heaviest particles kinematically permitted.

All these factors combined result in the Higgs boson coupling to every known massive particle, with a strength determined by the mass of the particle being coupled to. The fundamental Higgs interactions are shown in Figure 2.5.

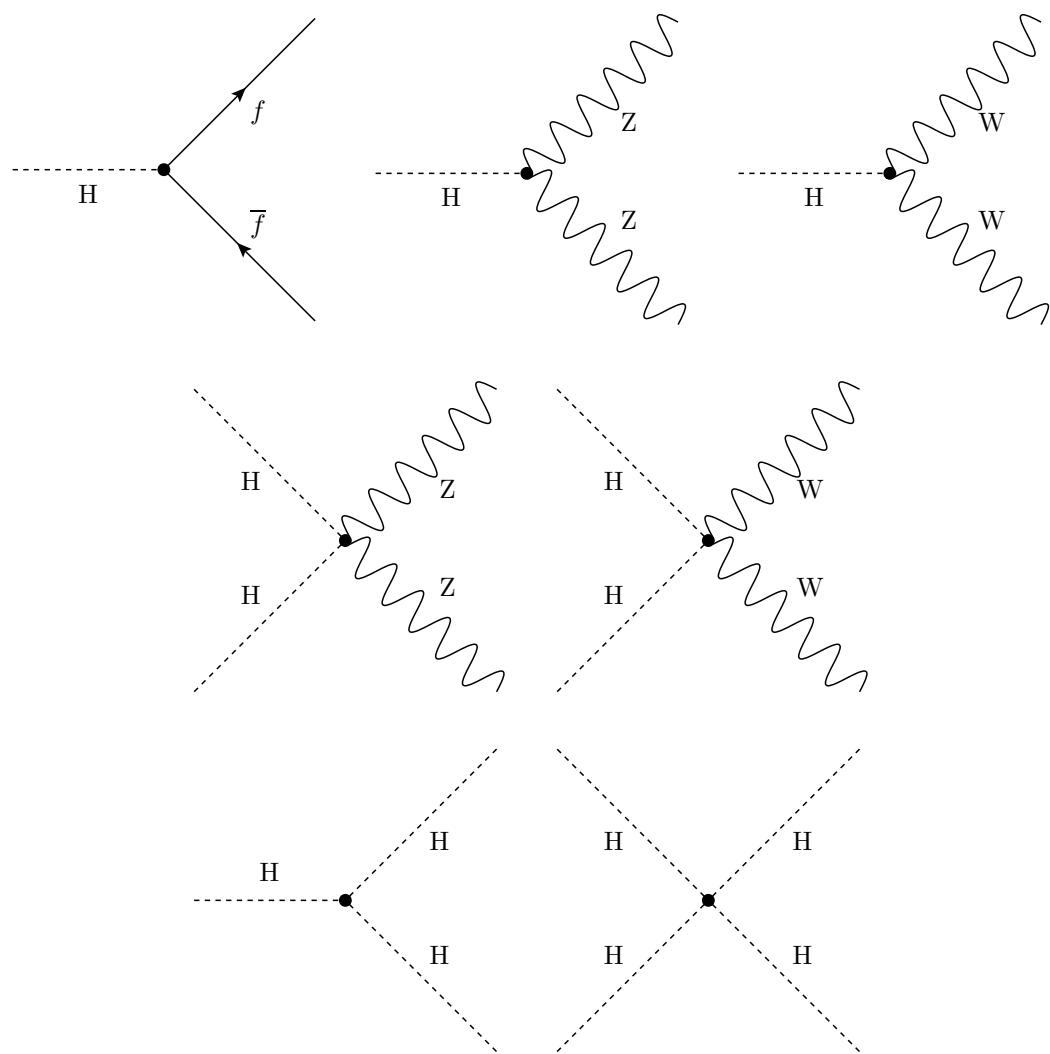


Figure 2.5: The interaction vertices of the Higgs boson in the Standard Model.

## 2.7 The Top Yukawa Coupling and $t\bar{t}H$ Production

The physics of the top-quark and the Higgs boson are tightly coupled to one another. A measurement of the top Yukawa coupling,  $Y_t$ , should be motivated as a key experimental observable at the LHC which will shed light on the energy scale for new physics models [39].

The top quark, with a pole mass of approximately  $173.3 \text{ GeV}/c^2$  [30], is the most massive particle known. The top quark is unusual as it decays via a weak interaction (typically  $t \rightarrow W b$ ) before it can form any stable hadronic states due to its large mass. This means the method of estimating the mass differs from lighter quarks where hadronic masses can be measured in flavour physics analyses. In fact there is a large amount of ambiguity surrounding the definition of the top-quark mass. Direct studies can be carried out, but due to colour confinement the mass measured will correspond to the mass of a simulated particle convolved with hadronisation model effects, rather than the pole mass. Cross-section measurements provide a mass measurement within a well-defined renormalisation scheme, but these measurements are not yet as precise as the direct ones. One can see from the vacuum expectation value and the top mass measurement that the top Yukawa coupling must be of order unity. From a naturalness perspective, this may appear sensible but it raises questions as to why the other Yukawa couplings need to be much smaller.

Measuring the cross-section of the Higgs produced in association with a top-quark pair will provide a direct measurement of the top Yukawa coupling. This is because the cross-section,  $\sigma$ , is proportional to the squared transition matrix element,  $|\mathcal{M}|^2$ . The transition matrix element is proportional to the product of couplings for each vertex in the Feynman diagram. It can therefore be observed that  $\sigma_{t\bar{t}H} \propto |Y_t|^2$  and hence if one assumes that there is no contribution to  $\sigma_{t\bar{t}H}$  from new physics processes, a measurement of  $\sigma_{t\bar{t}H}$  can be interpreted as a measurement of  $Y_t$ .

The  $t\bar{t}H$  production mechanism is a rare process when compared to total Higgs production. As shown in Figure 2.8,  $t\bar{t}H$  production accounts for 5.8% of the total  $pp \rightarrow H$  production cross-section. To be able to identify this process, an analysis needs to have a high signal efficiency. The  $t\bar{t}H$  production process is known to next-to-leading order (NLO) in QCD [40, 41]. Example diagrams of leading-order (LO)  $t\bar{t}H$  production are shown in Figure 2.6a. The decay rate of the Higgs decaying into two  $b$ -quarks is known to next-to-next-to-next-to-leading order (NNNLO) in QCD and NLO in electroweak [40]. The important background of inclusive  $t\bar{t}$  production is known at next-to-next-to-leading order (NNLO) in QCD [42] including resummation of next-

to-next-to-leading logarithmic (NNLL) soft gluon terms [43]. The dominant background is a top-quark pair produced in association with two additional  $b$ -quarks. The additional high- $p_T$   $b$ -quarks are produced from both initial and final state gluon or electroweak (EW) radiation. This process is known only to NLO in QCD. Example diagrams of LO QCD and EW processes are shown in Figure 2.6b-2.6c. A summary of the theoretical predictions and uncertainties related to the  $t\bar{t}H$  search is presented in Table 2.2.

Table 2.2: Predictions from the Standard Model for cross-sections of  $t\bar{t}H$ ,  $t\bar{t}$  and  $t\bar{t}+b\bar{b}$  at 8 TeV and branching ratio of  $H \rightarrow b\bar{b}$ . The values presented here use  $m_H = 125$  GeV and  $m_t = 172.5$  GeV [40, 41, 44, 45]. The uncertainty presented include dependencies on scale, PDF set and  $\alpha_S$  but do not contain any mass-related uncertainties.

	Central Value	Uncertainty (+)	Uncertainty (-)
$\sigma_{t\bar{t}H}$	129 fb	5%	6%
$\sigma_{t\bar{t}}$	253 pb	13%	15%
$\sigma_{t\bar{t}b\bar{b}}$	560 fb	29%	24%
$\text{BR}(H \rightarrow b\bar{b})$	57.7%	3.21%	3.27%

The top Yukawa coupling can be indirectly constrained from processes which require the Higgs boson to couple to a virtual top-quark loop at the lowest order of production. The production of a Higgs boson through gluon-gluon fusion ( $ggF$ ) and the decay of the Higgs boson to two photons both have top-quark loops contributing at the lowest order as shown in Figure 2.7. The cross-section calculations for these processes contain a squared top-Yukawa term and an interference term. The kappa-framework, which is used to create a combined likelihood function from multiple Higgs searches, provides a method to estimate the scaling of the Yukawa terms through linear scaling of the Standard Model predictions [47]. The kappa-framework take into account the Higgs couplings at both the production and the decay vertices and can further constrain their values using the total Higgs width. In these coupling fits, the interference term can help determine the sign of the top Yukawa coupling. These indirect measurements are subject to interpretation as they are loop processes which can be influenced by unknown particles which may exist at higher energies.

If one assumes that only Standard Model particles contribute to the cross-section and decay rate calculations then with the current best measurements from ATLAS (which include measurements from direct  $t\bar{t}H$  searches), the top Yukawa coupling is measured to be  $Y_t = 0.98 \times Y_t^{\text{SM}} {}^{+0.21}_{-0.18}$  [48]. The ATLAS and CMS Collaborations have combined their results to improve the error on the Higgs coupling measurements. Combining these results which in-

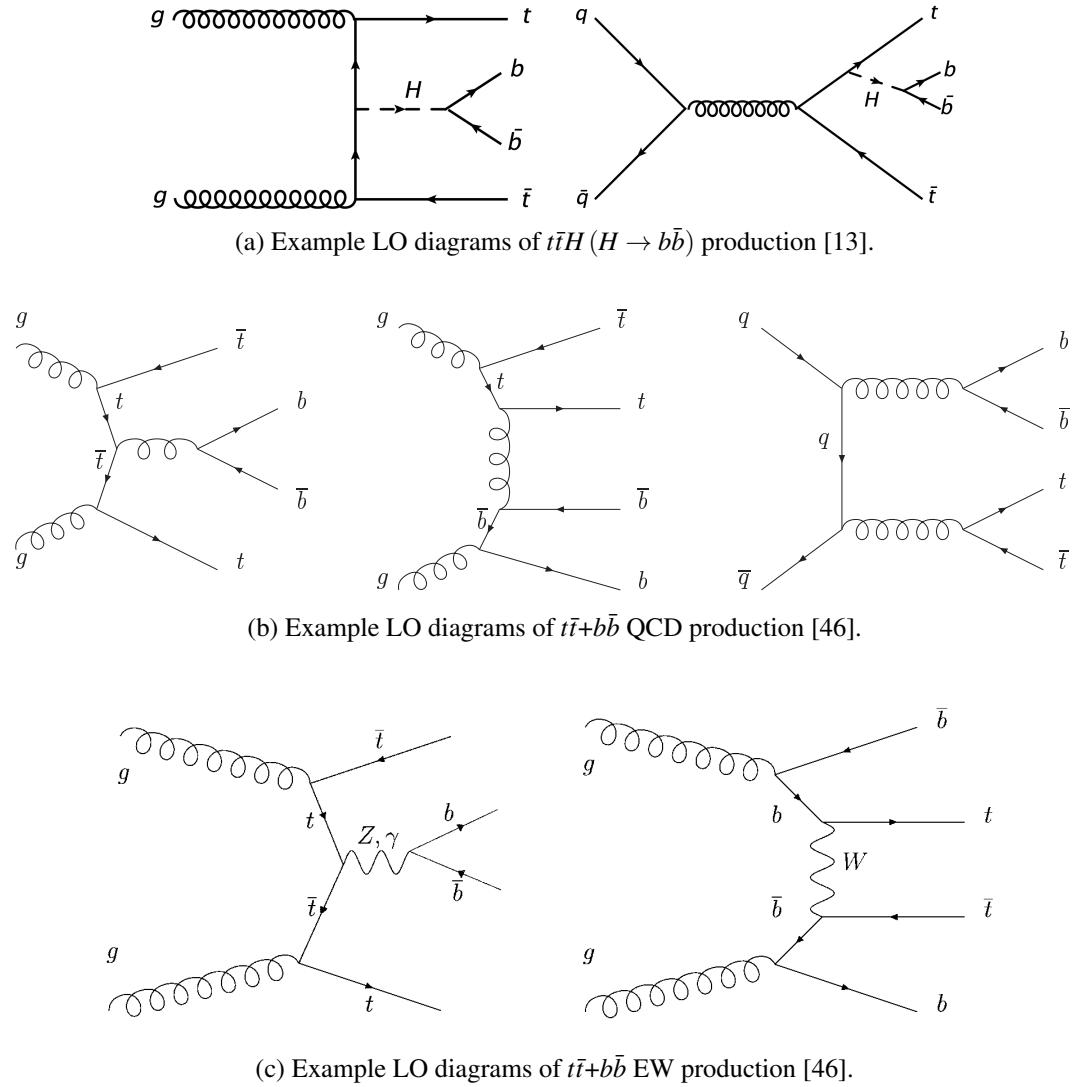


Figure 2.6: Leading-order Feynman diagrams of the signal process ( $t\bar{t}H$ ), and the dominant background process ( $t\bar{t}+b\bar{b}$ ), in the  $t\bar{t}H$  ( $H \rightarrow b\bar{b}$ ) analysis. For the background process, an understanding of QCD and EW mediated processes are both important as the QCD leading-order cross-section is only approximately a factor of 10 higher than the EW cross-section.

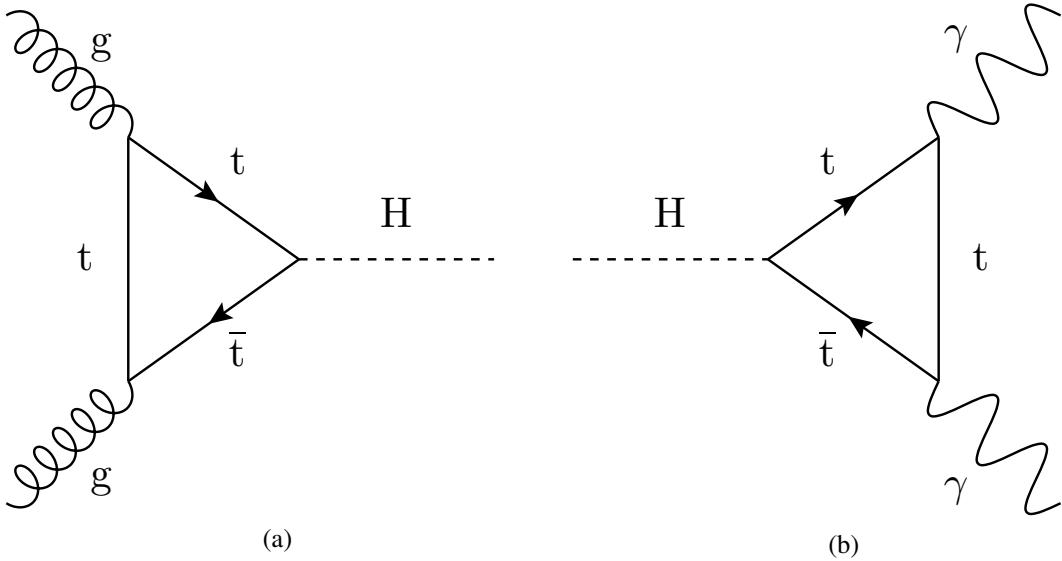


Figure 2.7: Two of the lowest order Feynman diagrams for a Higgs boson being produced through gluon-gluon fusion (Figure 2.7a) and for a Higgs boson decaying to two photons (Figure 2.7b). These processes have a contribution from a top-quark loop leading to a dependence on the top Yukawa coupling. However it is possible for heavy new physics particles to exist in this virtual loop, but the relative size of their contribution to the production and decay rates would be unknown. This means that any indirect measurement of the top Yukawa coupling will be dependent on assumptions made about physics at high energy scales.

clude the most recent  $t\bar{t}H$  published analyses, the top Yukawa coupling is measured to be  $Y_t = 0.89 \times Y_t^{\text{SM}} {}^{+0.14}_{-0.13}$  [48]. A clear observation of  $t\bar{t}H$  will improve the total error on  $\sigma_{t\bar{t}H}$  which is directly proportional to the square of the top Yukawa coupling scale factor. This will in turn improve the total error on the top Yukawa coupling measurement.

### 2.7.1 Vacuum Stability

The stability of the Universe is critically dependent on the value of the top and Higgs masses. The quartic Higgs coupling,  $\lambda$  in Equation (2.14), has a running dependence on energy scale,  $\lambda(\mu)$ . If this parameter changes sign at high energies, it can indicate there is an additional minima elsewhere in the potential and the electroweak minima is a meta-stable point. The running of  $\lambda$  has a dependence on  $m_H$ ,  $m_t$  and  $\alpha_S$ . Given the current measured values for these parameters, limits can be set requiring that the electroweak vacuum is the only minima in the potential. If these conditions are not satisfied it indicates that there is a limiting energy scale after which the Standard Model is no longer valid and new physics must become manifest. The theoretical predictions at NNLO on the value of the Higgs mass for the Universe to remain stable up to the Planck scale ( $m_{\text{Pl}} = 1.22 \times 10^{19} \text{ GeV}/c^2$ ) is given as

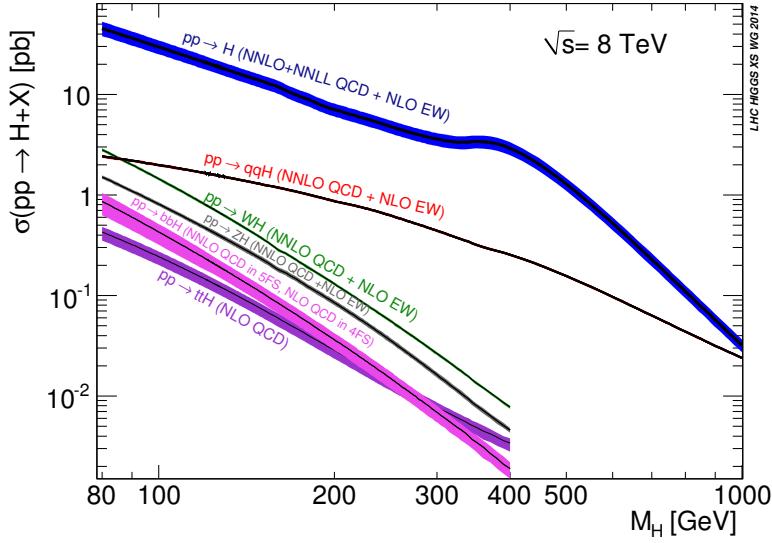


Figure 2.8: Higgs production cross-sections at 8 TeV [41]. For a Standard Model Higgs mass of 125  $\text{GeV}/c^2$ , the nominal  $t\bar{t}H$  production cross-section is  $\sigma_{t\bar{t}H} = 129.3 \text{ fb}$  and the nominal total Higgs production cross-section is  $\sigma_H = 22.3 \text{ pb}$ .

$$\begin{aligned}
 m_H \geq & 129.6 \text{ GeV}/c^2 + 2.0 \times (m_t^{\text{pole}} - 173.34 \text{ GeV}/c^2) \\
 & - 0.5 \text{ GeV}/c^2 \times \left( \frac{\alpha_s(m_Z) - 0.1184}{0.0007} \right) \pm 0.3 \text{ GeV}/c^2,
 \end{aligned} \tag{2.17}$$

which leads to a constraint on the Higgs mass of  $m_H > (129.6 \pm 1.5) \text{ GeV}/c^2$  for the vacuum to be stable up to the Planck scale [49]. This stability condition can be presented on a plane of allowed Higgs and top pole mass values, as shown in Figure 2.9.

The vacuum stability problem can be expressed in a form with a dependence on  $\lambda$  and  $Y_t$ , as shown in Figure 2.10. Both forms are complementary as the top mass and the top Yukawa coupling are only linked at leading order. Therefore the measurement of the top Yukawa coupling through the  $t\bar{t}H$  channel should be viewed as an important contribution to determining the status of the vacuum state. The impact of small variations of  $Y_t$  on the behaviour of  $\lambda$  as the energy scale is increased is shown in Figure 2.11. It can be seen that the value of  $Y_t$  determines the energy scale at which the Higgs potential changes shape. However, both the measurement of the top mass and the top Yukawa coupling are subject to issues of interpretation coming from the comparison of simulated data sample values with theoretical pole values.

Under the assumption of the Standard Model, the current world average variables indicate

that the Universe is in a meta-stable state. This implies that at higher energy scales, there is an additional minima which the Higgs field could fall into, changing the vacuum expectation value. A better understanding of the top Yukawa coupling will help make a stronger statement about the energy range at which the Standard Model breaks down and a new physics model would need to be valid.

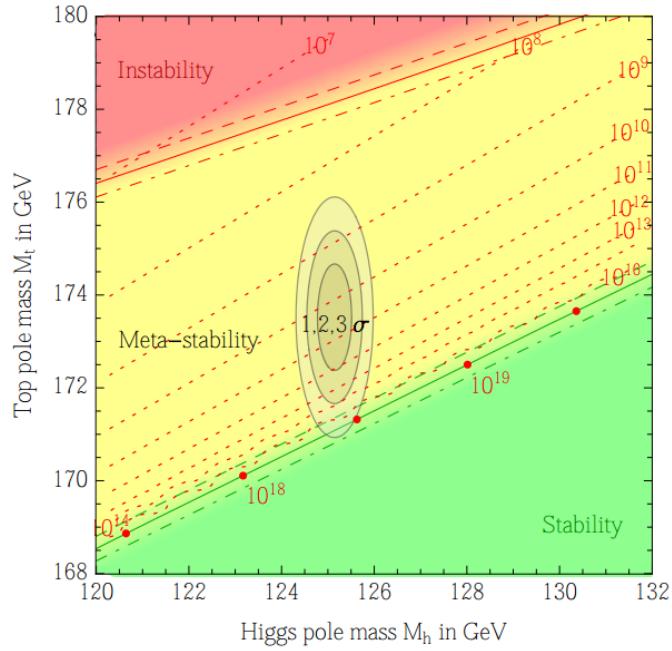


Figure 2.9: The stability of the Higgs potential as a joint function of the Higgs mass and the top mass. The Higgs potential is currently measured to sit in a meta-stable region. The plot shows the uncertainty region about the measured Higgs and top pole masses [49].

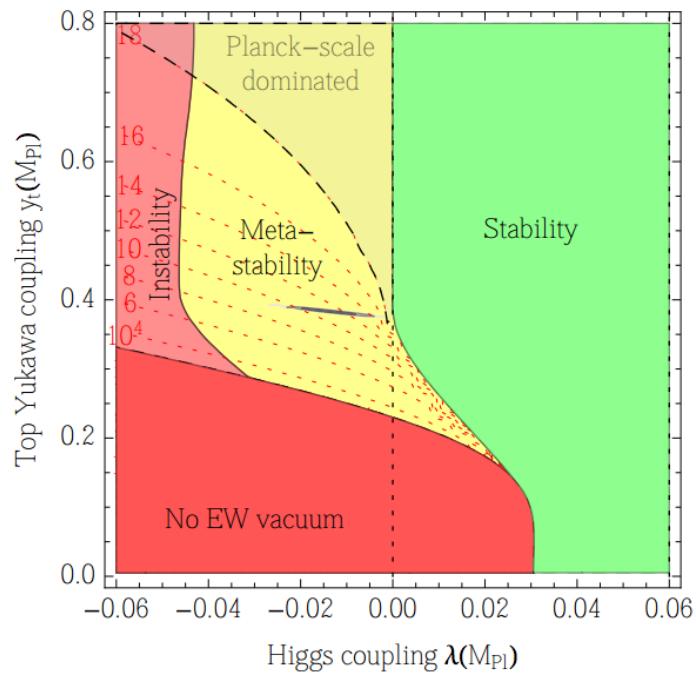


Figure 2.10: The stability of the Higgs potential as a joint function of the quartic Higgs coupling and the top Yukawa coupling. The values of the running quartic Higgs coupling and the top Yukawa coupling, both calculated at the Planck scale under that assumption of the Standard Model, places the Higgs potential in the meta-stable region [49].

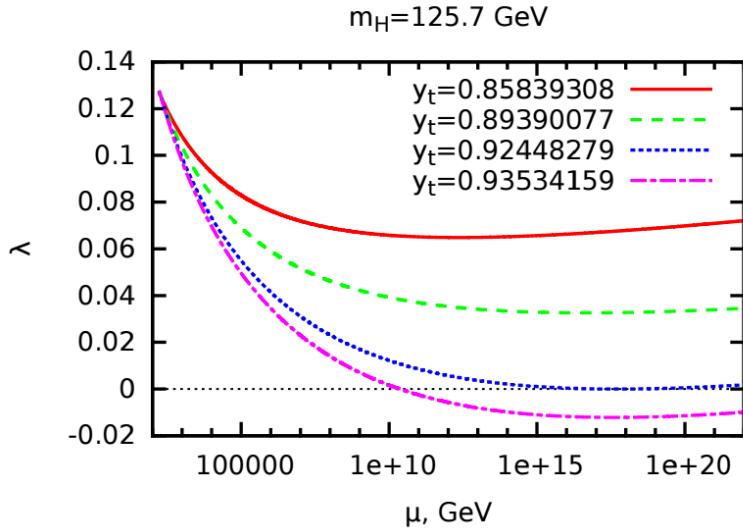


Figure 2.11: Running of the quartic Higgs coupling for different top Yukawa couplings calculated at the top mass [39]. It can be seen here that a precise measurement of  $Y_t$  is required to determine the behaviour of the vacuum up to the Planck scale.

### 2.7.2 Higgs Mass Corrections

The value of the Higgs mass itself is dependent on corrections which come from virtual loop processes. The dominant contributions come from the strongest couplings, which means from the top-quark loop as shown in Figure 2.12, and from the couplings to the  $W$  and  $Z$  bosons.

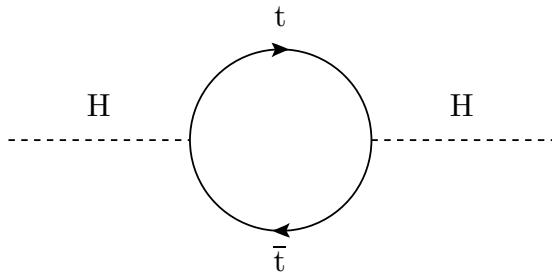


Figure 2.12: Virtual correction to the Higgs mass through a top-loop.

Corrections to the Higgs mass have contributions from a single fermion loop which is defined as

$$\Delta m_H^2 = -\frac{|Y_f|^2}{8\pi} \Lambda_{UV}^2 + \dots, \quad (2.18)$$

where  $\Lambda_{UV}$  indicates the energy scale cut-off which the correction is quadratically dependent on [50]. If the Standard Model is valid up to the Planck scale then this correction will be  $10^{30}$

times larger than the observed Higgs mass when considering the effect from the top quark. A direct observation of the  $t\bar{t}H$  process will enable an improved measurement on the top Yukawa coupling which in turn will provide information about the size of the quantum corrections to the Higgs mass.

A generalised effective quantum field theory can be used to describe the mass of an observed particle by providing the true pole mass of the particle and applying higher-order correction factors to it. This formulation provides the measured mass of the Higgs particle as

$$m_H^2 = (m_H^2)_0 + \Delta m_H^2 + \dots \quad (2.19)$$

The scale of the correction term,  $\Delta m_H^2$  calculated in Equation (2.18), means that either  $(m_H^2)_0$  needs to be orders of magnitude larger than the observed mass which questions why such a tiny factor remains afterwards, or a new physics theory must step in to protect the small pole mass from quadratic divergences (the classic example being supersymmetry).

## 2.8 Recent Higgs Results

Run 1 of the LHC has been an astounding success from the perspective of Higgs physics, with the discovery of a scalar boson within the allowed mass region for a light Higgs particle. To ascertain whether this particle truly is the Higgs boson responsible for the weak vector boson masses and the mass of the fermions, the spin, parity and couplings need to be precisely measured.

Experimental bounds had been placed on the Higgs mass from searches at LEP and the Tevatron indicating that a low mass Higgs would be in the region  $115 < m_H < 158 \text{ GeV}/c^2$ . Precision electroweak fits suggested that  $m_H = 96_{-24}^{+31} \text{ GeV}/c^2$ , as shown in Figure 2.13. The initial announcement on July 4, 2012 of the observation of a Higgs-like particle was mostly based on the two golden decay modes for a light Higgs with decay rates in agreement with the Standard Model. The discovery was announced jointly by the ATLAS and CMS Collaborations and both results combined the  $\sqrt{s} = 7 \text{ TeV}$  and  $\sqrt{s} = 8 \text{ TeV}$  datasets [8, 9]. Those decay modes were  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ^* \rightarrow llll$ . Whilst the diphoton decay has a small branching ratio, the channel is experimentally clean. The Higgs signal appears as a bump over an exponentially decaying background which allows the use of function fitting to identify data which is not in agreement with the background-only model. The four lepton channel is also clean due to the

low ZZ cross-section and the ability to trigger and identify the full leptonic final state.

Using the full 2011 and 2012 datasets, the measurements from the ATLAS and CMS Collaborations were combined to find a measured Higgs mass of  $125.09 \pm 0.24 \text{ GeV}/c^2$  [51–53]. The results from the ATLAS experiment are shown in Figure 2.14 where one can clearly see a Higgs signal above the expected Standard Model backgrounds.

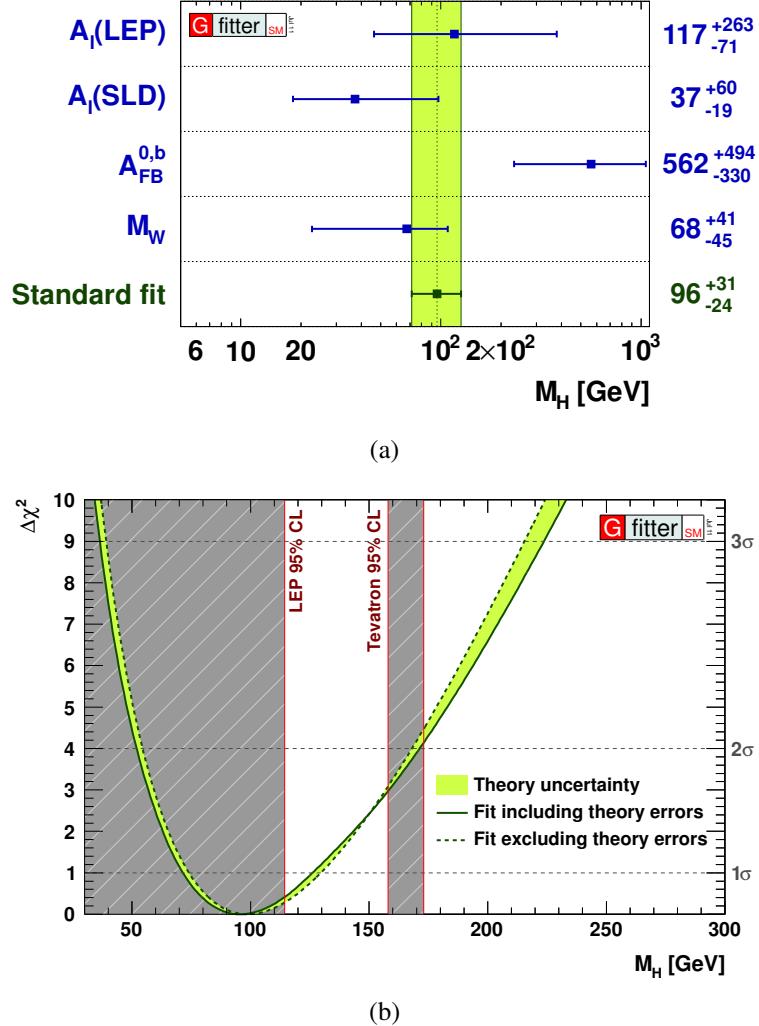


Figure 2.13: Available Higgs mass phase space as allowed by precision electroweak fits (Figure 2.13a) and combined with experimental bounds (Figure 2.13b) in July 2011 [54].

To fully establish the spin-parity quantum numbers ( $J^P$ ) of the observed particle, a number of hypothesis tests were carried out. The coupling (through a top or  $W$  loop) to two photons provides evidence that the particle is not spin-1 through the Landau-Yang theorem [55, 56]. This theorem states that a massive vector boson is unable to decay into two massless vector particles

as the decay amplitude is antisymmetric under the exchange of the two outgoing particles. Analyses by the ATLAS and CMS Collaborations in the  $H \rightarrow \gamma\gamma$ ,  $H \rightarrow WW \rightarrow l\nu l\nu$  and  $H \rightarrow ZZ \rightarrow 4l$  channels all converged to clearly favour the  $0^+$  hypothesis and reject multiple alternate hypotheses for  $0^-$ ,  $1^+$ ,  $1^-$  and  $2^+$  [10, 11, 30].

The current status of the Higgs boson is that it agrees very well with the Standard Model hypothesis and is interpreted in this way. Whilst improved coupling measurements are required to ensure this is the case, combined fits of the Run 1 ATLAS analyses can already probe the nature of the Higgs coupling to fermions and bosons. This is best seen when presented in Figure 2.15, where the Standard Model Higgs predicts a linear relationship between the coupling and mass.

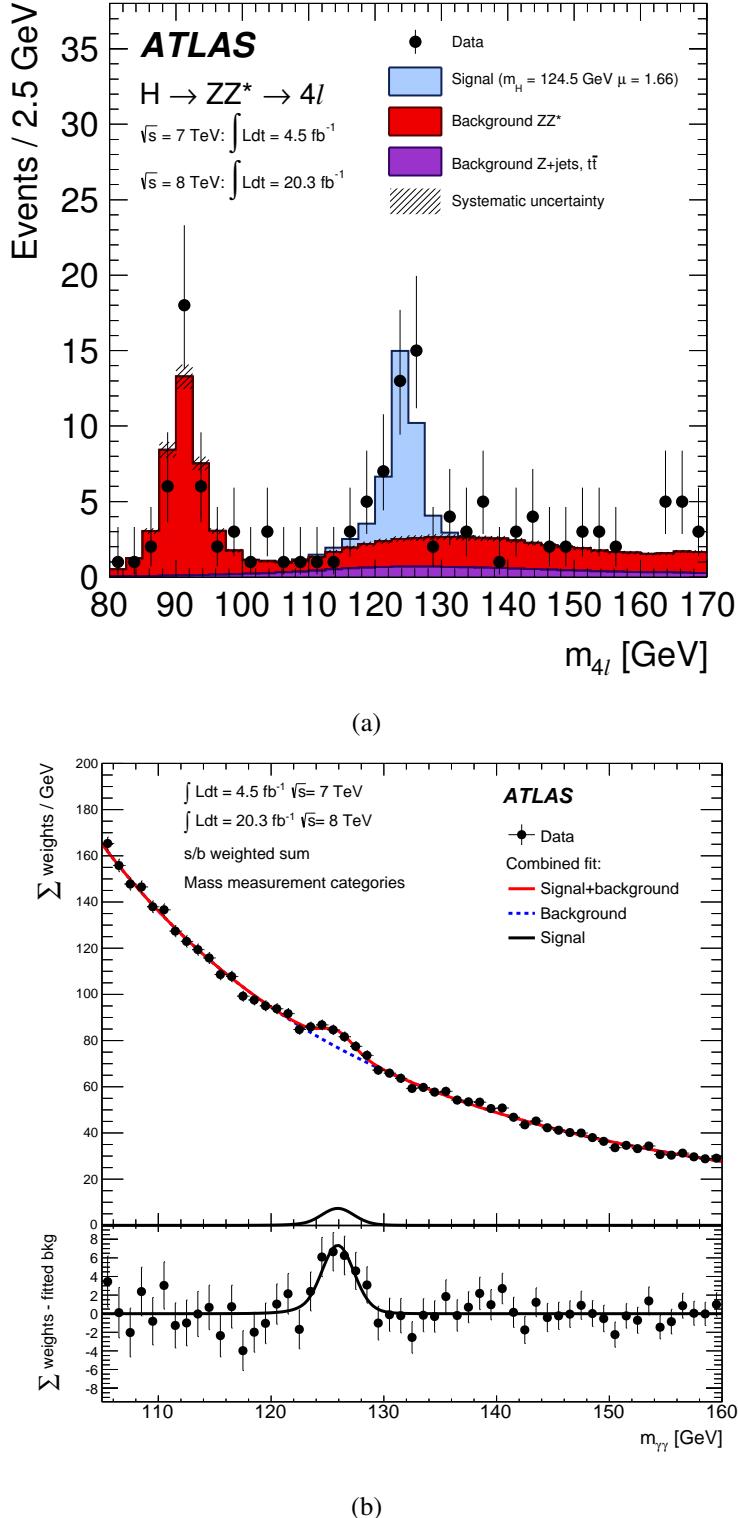


Figure 2.14: Higgs mass measurements in the  $H \rightarrow 4l$  channel (Figure 2.14a) and the  $H \rightarrow \gamma\gamma$  channel (Figure 2.14b) with the ATLAS detector using the Run 1 dataset. In both distributions, a clear signal can be identified in agreement with the signal hypothesis of a Higgs particle with a mass close to 125  $\text{GeV}/c^2$  [57].

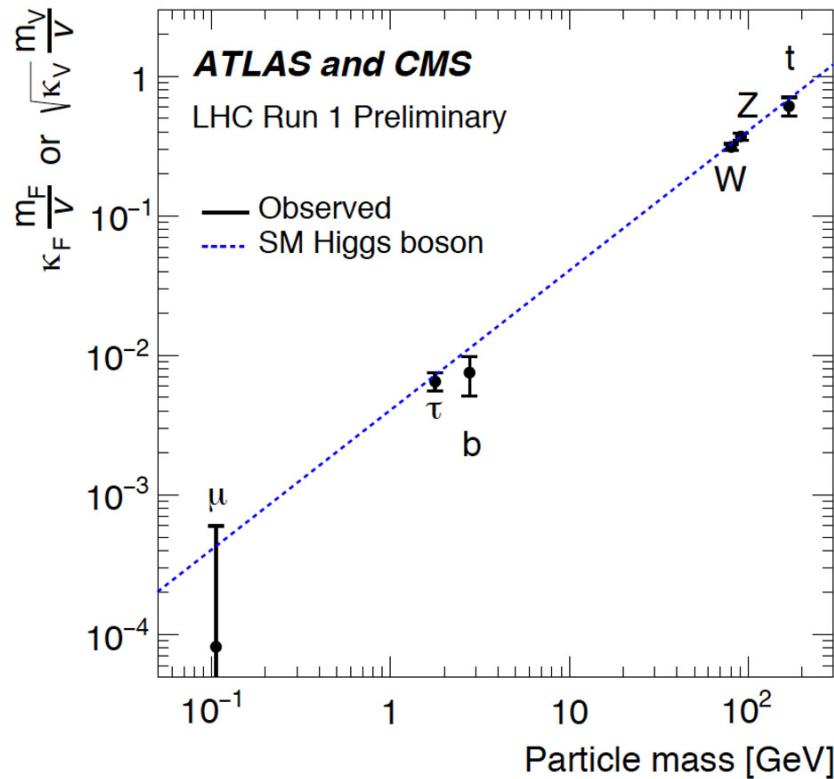


Figure 2.15: The mass of the Standard Model particles shown against the best fit coupling to the Higgs boson using a combined fit to the current published ATLAS and CMS Higgs analyses assuming only Standard Model contributions [47, 58, 59]. The scale factors  $\kappa_F$  and  $\kappa_V$  are fitted parameters which are applied to the Standard Model expectation of the Higgs coupling. A value of 1 for a  $\kappa$  parameter would indicate an observation consistent with the Standard Model. Here the expectation of a linear correspondence between the normalised Higgs coupling and the particle mass is shown to exist within error. The top Yukawa coupling is primarily constrained by the  $gg \rightarrow H$  and  $H \rightarrow \gamma\gamma$  loops but also includes the most recent  $t\bar{t}H$  analyses by ATLAS and CMS.

# Chapter 3

## The LHC and ATLAS Detector

---

*“One can state, without exaggeration, that the observation of, and the search for, similarities and differences are the basis of all human knowledge.”*

— Alfred Nobel

---

This chapter provides information on the accelerator complex at CERN used to deliver proton-proton collisions to the ATLAS detector and details the detector technologies implemented. The operation of the LHC complex at CERN is described in Section 3.1. This is followed by a summary of the ATLAS detector in Section 3.2, with a detailed description of the subsystems which make up the ATLAS detector in Section 3.2.1-3.2.5. A brief discussion on the simulation of events is given in Section 3.3.

### 3.1 Overview of the LHC Complex

The Large Hadron Collider (LHC) is part of an accelerator complex located at CERN, 8 kilometres from Geneva, Switzerland. The layout of the accelerator complex is detailed in Figure 3.1. The LHC is situated 100 metres underground, has a circumference of 27 kilometres and is the highest energy particle collider in the world. It ran at a maximum centre-of-mass energy of 8 TeV during Run 1 and is currently running at 13 TeV in Run 2. To achieve these centre-of-mass energy, a series of accelerators illustrated in Figure 3.1 are used to accelerate two beams

of proton bunches to an energy of half the centre-of-mass collision energy.

Linac 2 takes protons from a hydrogen gas source and accelerates them up to an energy of 50 MeV. These protons are injected into the Proton Synchrotron (PS) Booster, where they are further accelerated to 1.4 GeV. The beam then enters the PS, which accelerates the beam up to 25 GeV, after which it is injected into the Super Proton Synchrotron (SPS). This accelerator is already a discovery machine in its own right, having been the particle accelerator which supplied the proton-anti-proton collisions to the UA1 and UA2 detectors in the 1970s and 1980s, which led to the discovery of the  $W$  and  $Z$  bosons. The SPS accelerates the beam up to 450 GeV, after which it is injected in both the clockwise and anti-clockwise directions into the LHC. Superconducting dipole magnets are used to bend the proton beams around the LHC with a peak field strength of 8.3 T. Each dipole magnet is 15 m in length and contains eight radio frequency cavities which are used to accelerate the beam up to 6.5 TeV. Each radio frequency cavity provides an accelerating field of 5 MV/m. At interaction points situated around the LHC, the beam is focused down with a  $\beta^*$  of 0.55 m in nominal running conditions.

In the first part of Run 1 of the LHC, protons beams were collided at a centre-of-mass energy of 7 TeV, accumulating around  $5 \text{ fb}^{-1}$  of data. The centre-of-mass energy was then increased up to 8 TeV for the latter part of Run 1, delivering over  $20 \text{ fb}^{-1}$  of data.

The LHC has been designed to produce centre of mass collisions at 14 TeV. In order to achieve this design collision energy, the accelerator was shut down on the 16<sup>th</sup> of February 2013 for two years to allow upgrades to be made to the accelerators and to the experiment detectors. The upgrades focused primarily on ensuring the superconducting magnets can handle the high currents required to provide the strong magnetic field and on upgrading detector components so that they can handle the increased particle flux and radiation damage which will be delivered during Run 2. On the 3<sup>rd</sup> of June 2015, the LHC delivered stable proton beams accelerated up to 6.5 TeV and successfully provided collisions at a centre-of-mass energy of 13 TeV.

The maximum instantaneous luminosity designed for the LHC was to be  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ . The goal in 2015 is to reach an instantaneous luminosity of  $1.3 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  with 25 ns bunch spacing filling 2800 out of a maximum of 2808 bunches.

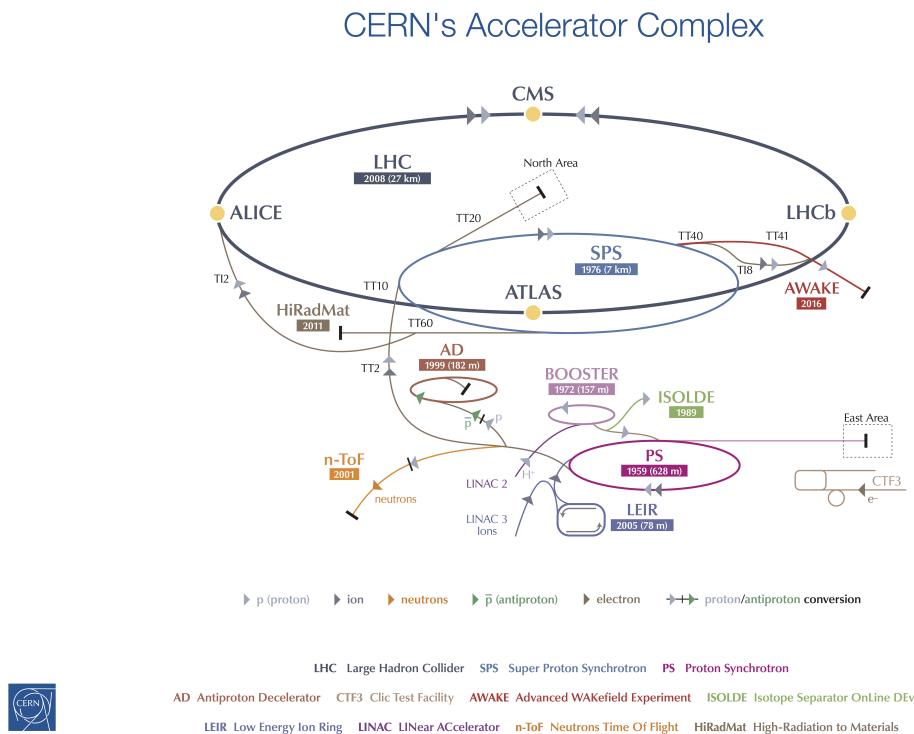


Figure 3.1: Diagram detailing the CERN accelerator complex which feeds the LHC as well as additional experiments [60].

## 3.2 ATLAS - A Toroidal LHC Apparatus

The ATLAS experiment is one of the four main experiments, located at Point 1 of the LHC complex. It is a general purpose detector designed to provide an inclusive picture of proton-proton collisions at the LHC. A well-defined coordinate system is used to describe the detector and the physics objects which are identified. The  $z$ -axis is defined as going along the beam-pipe, where the positive  $z$ -direction is clockwise around the LHC. The  $x$ - $y$  plane is transverse to the beam pipe, where the  $x$ -direction is defined as going radially inwards to the centre of the LHC and the  $y$ -direction is defined as pointing upwards. The azimuthal angle,  $\phi$ , is defined between  $-\pi$  and  $\pi$  clockwise around the  $z$ -axis and the polar angle,  $\theta$ , is the opening angle defined from the positive  $z$ -direction between 0 and  $\pi$ . The transverse plane can also be described as the  $r$ - $\phi$  plane where  $r$  is the radial distance from the beam line. The pseudorapidity is defined as  $\eta = -\ln(\tan(\frac{\theta}{2}))$  and used to describe angles relative to the  $z$ -axis. The direction of reconstructed objects are defined by a measurement of  $(\eta, \phi)$  and a parameterisation of the distance in this space is defined as  $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ .

The ATLAS detector with its subsystems can be seen in Figure 3.2. The detector is designed to be hermetic and provide  $4\pi$  coverage over a large range of  $\eta$ . This allows for the measurement of missing transverse momenta to enable the inference of invisible particles, such as neutrinos or dark matter.

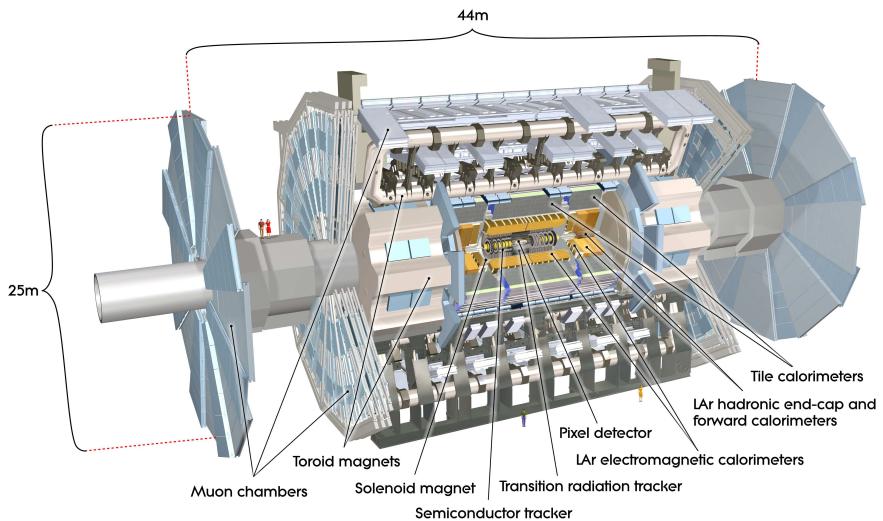


Figure 3.2: Diagram of the ATLAS detector [61].

The ATLAS detector can be described as a series of subsystems which occupy space radially

outwards from the beam pipe. The inner detector (described in Section 3.2.1) is comprised of the pixel detector, semi-conductor tracker (SCT) and the transition radiation tracker (TRT). The electromagnetic calorimeter and the hadronic calorimeter (described in Section 3.2.2) are located outside the inner detector. On the outside of the detector are the muon spectrometers (described in Section 3.2.3). The detector is split into three sections: the barrel and the two end-caps. Typically the barrel region contains the central physics region of  $|\eta| < 2.5$ , though due to the fixed length of the barrel region, the inner detectors can extend beyond this range in the barrel region.

A summary of the required measurement resolutions which have motivated the design of the ATLAS detector are presented in Table 3.1 .

Table 3.1: Design requirements for the ATLAS detector systems where  $p_T$  is provided in  $\text{GeV}/c$  and  $E$  is provided in  $\text{GeV}$  [61].

Detector	Intrinsic Resolution	Coverage	Trigger
Inner Detector	$\sigma_{p_T}/p_T = 0.05\% p_T \oplus 1\%$	$ \eta  < 2.5$	n/a
EM Calorimeter	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$	$ \eta  < 3.2$	$ \eta  < 2.5$
Hadronic Calorimeter	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$	$ \eta  < 3.2$	$ \eta  < 3.2$
Muon Spectrometer	$\sigma_{p_T}/p_T = 10\% (\text{for } p_T = 1 \text{ TeV}/c)$	$ \eta  < 2.7$	$ \eta  < 2.4$

The ATLAS detector contains two magnetic fields (described in Section 3.2.4); a uniform field along the  $z$ -axis within the inner detector generated by a solenoid and a field in the  $r\text{-}\phi$  plane in the remaining subsystems generated by multiple toroids. These fields bend charged particles, allowing for measurements of charged particle momenta through the curvature of their track. A charged particle will be bent by the solenoid in the transverse plane and will be bent by the toroid in the  $r\text{-}z$  plane. During the 8 TeV data-taking period, the LHC delivered 20 million proton-proton bunch-crossings per second. This rate of raw data is too large to allow every single event to be recorded. A multi-level data acquisition system (described in Section 3.2.5) is therefore used to reduce the number of events recorded to a manageable rate by assessing whether an event contains some form of physics which is interesting enough to be stored.

### 3.2.1 The Inner Detector

The Inner Detector (ID) is designed to provide tracking information within  $|\eta| \leq 2.5$  and to identify the vertices of multiple interactions in the busy environment at the collision point, where

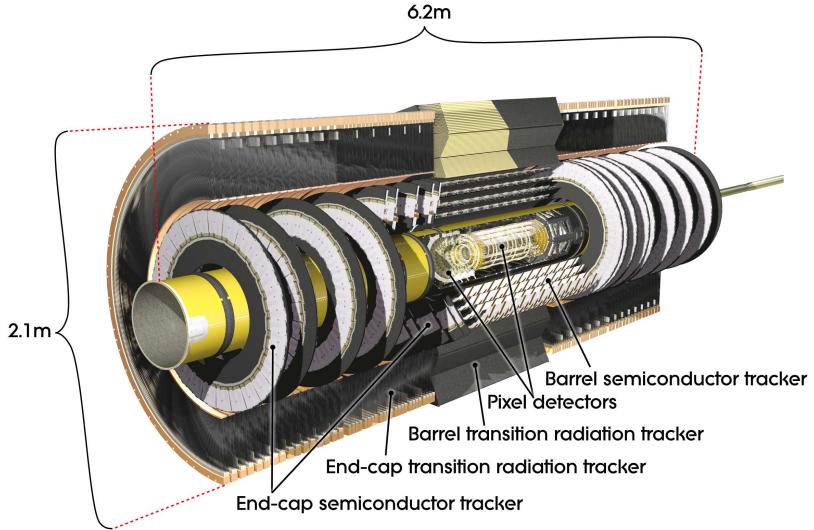


Figure 3.3: Diagram of the ATLAS inner detector [61].

each bunch-crossing can produce on the order of  $10^3$  tracks. The analyses presented in subsequent chapters require excellent lepton momentum resolution and efficient identification of primary and secondary vertices to reconstruct dileptonic top-pair events in data.

To achieve this with the ID, there are three subsystems: the pixel detector, the semi-conductor tracker and the transition radiation tracker, as shown in Figure 3.3. These three subsystems combine to provide independent and complementary information allowing for a momentum resolution of a curved charged track in the ID of  $\sigma_{p_T}/p_T = 0.05\% p_T \oplus 1\%$ , for  $p_T$  in  $\text{GeV}/c$ .

Track and vertex reconstruction requires accurate measurements of parameters relating to the position of the track relative to the primary vertex. The variable  $d_0$  is defined as the distance between the beam-line and the point of closest approach of a track to the beam-line. The  $z$ -component of the point of closest approach used to calculate  $d_0$  is defined as  $z_0$ . For tracks with large momenta, the resolution in these track parameters is dominated by the intrinsic resolution of the detector whereas at low momenta the resolution is dominated by multiple scattering inside the detector. In the barrel region where the amount of material in the inner detector is minimised, the intrinsic resolution in measuring  $d_0$  and  $z_0 \sin \theta$  is  $10\mu\text{m}$  and  $91\mu\text{m}$  respectively [46].

**The Pixel Detector** The pixel detector provides high resolution information about the location of a charged particle very close to the beam line. The pixel detector is made from oxygenated n-type silicon wafers with readout pixels on an n<sup>+</sup>-implanted side. There are three concentric

cylindrical layers in the barrel located within 206 mm the beam-pipe and three disks located in each of the end-caps, perpendicular to the beam-pipe, totalling 80 million installed pixels. Each pixel is identical and has a size in the  $r\text{-}\phi$  plane of  $50\mu\text{m}$  and in  $z$  of  $400\mu\text{m}$ . This provides a resolution in the  $r\text{-}\phi$  plane of  $10\mu\text{m}$  and in  $z$  of  $115\mu\text{m}$  [61].

**The Semi-Conductor Tracker** The SCT detector is located 255 mm radially outward from the beam line and consists of strips of silicon detector which are mounted in pairs with a small angle of 40 mrad between the two strips. These stereo strips have one strip aligned with the beam pipe, and the other is angled slightly off-axis. Each layer therefore provides two independent hits. This allows a measurement of  $(r,\phi)$  without requiring individual pixel hits, thus reducing the number of readouts otherwise required. The SCT consists of four cylindrical layers which surround the components of the pixel detector in the barrel region and nine disks in the end-caps. In the barrel region, the design of the SCT provides an intrinsic resolution of  $17\mu\text{m}$  in the  $r\text{-}\phi$  plane and an intrinsic resolution of  $580\mu\text{m}$  in  $z$  [61].

**The Transition Radiation Tracker** The outermost component of the ID is the TRT whose purpose is to provide continuous tracking at a large radius where the volume would be too costly to instrument with silicon and provide charged particle identification. The TRT is a drift tube chamber built up out of 4 mm diameter straw tubes. Each straw has a negatively charged cathode surface surrounding a grounded anode wire. The tubes are sealed containing a gaseous mixture of 70% Xenon, 27% Carbon Dioxide and 3% Oxygen. This gas will be ionised when a charged particle travels through it. The ionised electrons will drift in the electric field to the anode. The drift radius is measured from the time taken to readout the signal, given a drift velocity, and multiple TRT hits combined with the silicon hits reduces ambiguity in the fitted track. There are 73 straw planes in the barrel region and 160 straw planes in the end-cap region. The straws are aligned with the  $z$ -axis in the barrel region and radially aligned in the end-cap regions. Particles with  $p_{\text{T}} > 0.5 \text{ GeV}/c$  generate approximately 36 hits in the TRT, reducing down to 22 hits in the transition region.

Transition radiation is generated when a charged particle travels between materials with different dielectric constants. The space between straw tube layers contains a  $15\mu\text{m}$  thick polypropylene foil. These foil layers allow the generation of transition radiation which propagates into the gas causing excitation and additional ionisation. Transition radiation from elec-

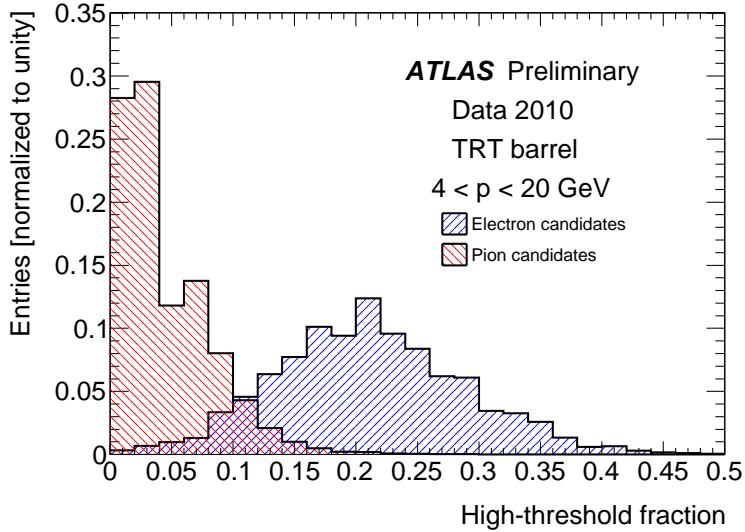


Figure 3.4: A comparison of the fraction of high threshold hits in the TRT for pions and electrons using  $\sqrt{s} = 7 \text{ TeV}$  2010 data [62].

trons typically deposits 8-10 keV into the TRT. Minimum ionising particles such as pions typically deposit 2 keV from transition radiation. The energy of transition radiation photons depends on the Lorentz factor ( $\gamma = E/m$ ) of the charged particle. This means for two different particles with a given energy, the lighter particle will generate more transition radiation than the heavier one. This allows for discrimination between electrons and minimum ionising particles by counting the number of high threshold hits on a TRT track, where the readout threshold is typically around 6-7 keV for pion-electron discrimination. An example of the separation provided by looking at the number of high threshold hits in data is presented in Figure 3.4.

### 3.2.2 The Calorimetry System

The calorimeter system starts at a radius of 1385 mm from the beam line and extends out to a radius of 4230 mm in the barrel region. The calorimeter system is designed to contain electromagnetic and hadronic showers so that there is limited punch-through into the muon system. Materials can be characterised by the average distance a high-energy particle will travel before having its energy reduced to  $1/e$  of the original energy due to material interactions. The radiation length,  $X_0$ , characterises the distance a particle will travel whilst losing energy electromagnetically. The interaction length,  $\lambda_0$ , characterises the distance a particle will travel whilst losing energy via nuclear interactions. Due to the stochastic nature of the energy deposition, the

number of radiation or interaction lengths needs to be large enough contain the entire particle shower and reduce the probability of punch-through to a negligible level. In the barrel region, the electromagnetic calorimeter has a thickness in radiation lengths between  $22\text{-}33 X_0$ . The combined electromagnetic and hadronic calorimeters contribute an active calorimeter system in interaction lengths of approximately  $9.4 \lambda_0$ . Studies have shown that this system provides a good energy resolution of hadronic jets with energies up to 1 TeV. Both electromagnetic and hadronic calorimeters in the ATLAS detector are sampling calorimeters. The energy of a particle is measured by distributing its energy into a series of low energy particles, referred to as a shower, by having it pass through a dense material. The shower evolves quickly and enters a medium which is selected to either ionise, such as the liquid argon calorimeter, or scintillate, such as the hadronic tile calorimeter, and produce electronic signals. In both cases an estimation of the energy of the particle is extracted by summing the contributions from the active layers.

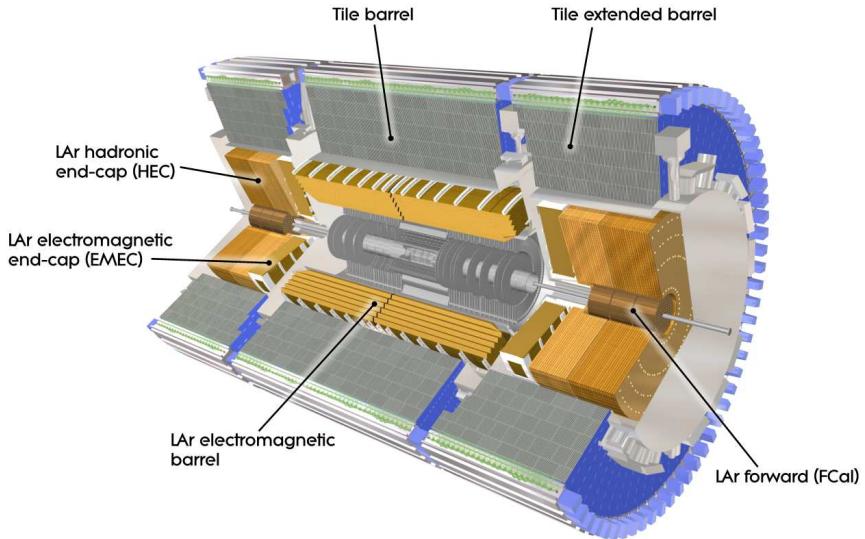


Figure 3.5: Diagram of the ATLAS calorimeter system [61].

**The Electromagnetic Calorimeter** The electromagnetic calorimeter (ECAL) is divided into barrel and end-cap regions. The barrel ECAL is housed in the same cryogenic unit as the solenoid which surrounds the inner detector in order to minimise material interactions between the inner detector and the ECAL. In the barrel region there is approximately  $2\text{-}4 X_0$  of material between the beam line and the ECAL, and around  $0.66 X_0$  of that material is located between the ID and the ECAL. The ECAL is a lead liquid-argon (LAr) detector, which is designed to give

full azimuthal coverage without any uninstrumented regions due to detector geometry. This azimuthal angle coverage is achieved through the use of an accordion structure of the lead absorber and the electrodes. An illustration of the segmentation in the ECAL layers and the accordion structure is shown in Figure 3.6.

Within the pseudorapidity range of  $|\eta| < 2.5$ , the LAr calorimeter provides three calorimetry layers located in both the barrel and in the overlap region between the barrel and end-caps. This design was primarily motivated by the search for  $H \rightarrow \gamma\gamma$  which utilises photon pointing information provided by the calorimeter to identify the vertex which the two photons pointed back to. The  $\eta$  measurement of a photon is taken from a combination of the first and second layers and the  $\phi$  measurement is taken from the second layer.

The ECAL also contains an active LAr presampler layer within  $|\eta| < 1.8$ . The purpose of this layer is to provide an estimation of energy losses which have occurred prior to reaching the ECAL.

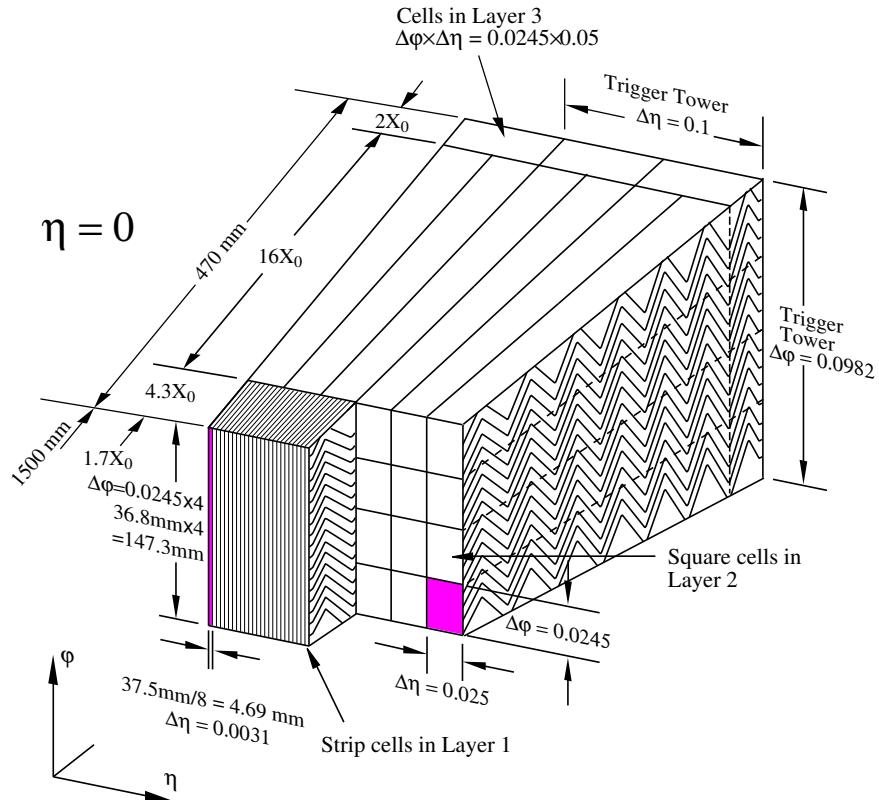


Figure 3.6: Example of the segmentation of the ECAL in the barrel region [61].

**The Hadronic Calorimeter** The hadronic calorimeter (HCAL) is installed outside of the ECAL and uses two different technologies. The barrel region (within  $|\eta| < 1.0$ ) with an extended barrel between  $0.8 < |\eta| < 1.7$  is instrumented using a sampling tile calorimeter with steel absorbers and scintillating tiles. As showering particles pass through the plastic layers, they scintillate emitting a flash of blue light. Wavelength shifting fibres are located at the edge of the tiles and are used to extract signals by shifting them to a lower wavelength as they enter the fibre and guiding it out into photomultiplier tubes.

In the end-cap regions, the same LAr calorimetry technology is employed as in the electromagnetic calorimeter except using copper absorber plates instead of lead. Copper has a longer radiation length than lead, but a slightly smaller hadronic interaction length which is preferable in a hadronic calorimeter.

### 3.2.3 Muon Spectrometer System

The muon detectors are situated in the outermost part of the ATLAS detector as shown in Figure 3.7. Energy lost by a charged particle as a result of bremsstrahlung is proportional to  $E/m^2$  which means it is important for electrons but not particularly for muons. The Bethe-Bloch formula shows that the energy loss due to ionisation goes as  $1/\beta^2$  where  $\beta \equiv v/c$ . For the energies involved at the LHC, muons produced are typically minimally ionising and will pass through the inner detector and calorimeters without depositing very much energy. In order to get a precise momentum measurement of a muon, it is much more accurate to measure the curvature of the track from bending in a magnetic field rather than using a calorimeter. The muon spectrometer has been designed to achieve a  $p_T$  resolution of 10% for tracks with a  $p_T$  of 1 TeV/c.

In the outer part of the detector, the ATLAS detector has a magnetic field generated by large toroids as shown in Figure 3.7. There are eight coils in the barrel toroid which provides bending in the region  $|\eta| < 1.4$  and two end-cap toroids installed in-line with the central solenoid in the region  $1.6 < |\eta| < 2.7$ . The region between the two toroid ( $1.4 < |\eta| < 1.6$ ) is called the transition region, and bending is provided from a combination of both magnetic fields.

Measurements are made using monitored drift tubes (MDTs) and cathode strip chambers (CSCs). Resistive plate chambers (RPCs) and thin gap chambers (TGCs) are installed to provide fast triggering and location information. The muon system has a coverage out to  $|\eta| < 2.7$  and the location of the components are shown in the cross-section of the subsystem in Figure 3.8.

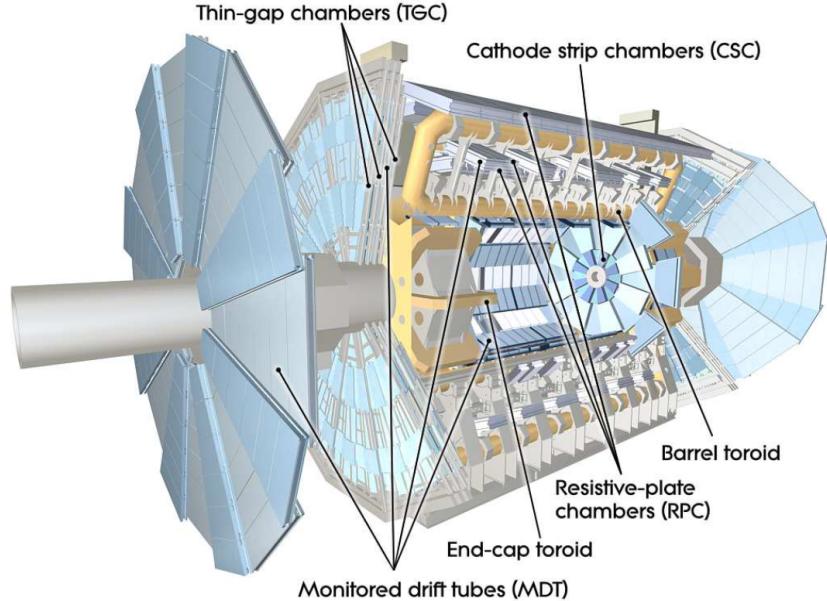


Figure 3.7: Diagram of the ATLAS muon spectrometer system [61].

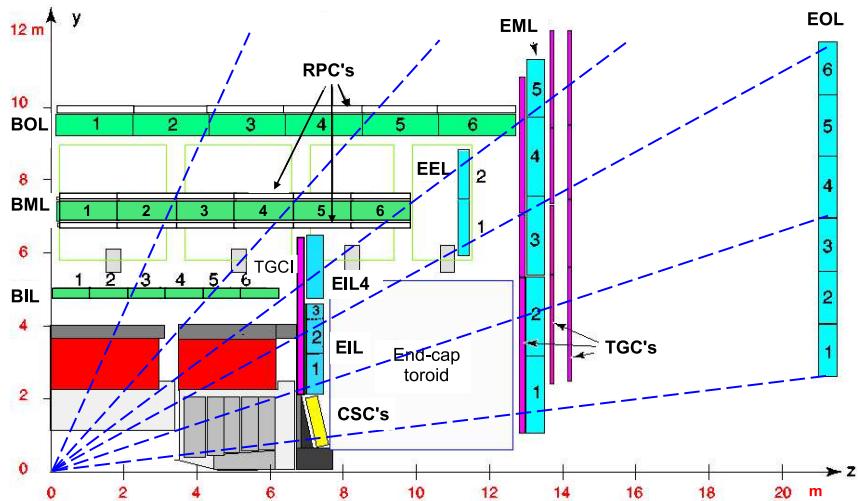


Figure 3.8: Cross-section of the muon detectors in the bending plane. The components are labelled according to their position in the barrel (B) and end-cap (E) regions, and whether they are in the inner (I), middle (M) or outer (O) layer. Only the large (L) component of each subsystem is shown. An extra (E) layer is located in the end-cap region to provide additional coverage where the inner layer does not project out to the outer layer [61].

**Monitored Drift Tubes** MDTs in the muon spectrometer provide instrumentation across the range of  $|\eta| < 2.7$  except for the innermost layer of the end-cap region, where the CSC is installed to cope with higher flux of charged particles. There are eight MDT chambers in the barrel region which are located between the magnet toroids. Each chamber contains three separate layers, within each are between three and eight drift tube layers. The gas inside the drift tubes is 97% Argon and 3% Carbon Dioxide held at a pressure of 3 bar. This gas combination should be stable over the lifetime of the muon detector with next to no degradation of the anode wires.

**Cathode Strip Chambers** The forward region close to the beam line will experience the highest flux of charged particles in the muon system. To cope with this, CSCs are installed at the first inner layer of the end-cap region between  $2.0 < |\eta| < 2.7$ . The CSCs are multi-wire proportional chambers with anode wires that are aligned in the radial direction between two cathodes. The cathodes are strip layers which are orientated with one layer perpendicular to the wires to measure a precision coordinate and the other layer parallel to the wires to measure the transverse coordinate. There are four layers of CSCs to resolve multiple track ambiguities. Combined they provide a measurement in  $(\eta, \phi)$  with a resolution in  $r$  of  $40\mu\text{m}$  and in  $\phi$  of 4 mm [61].

**Trigger Chambers** Triggering information is provided by RPCs and TGCs out to  $|\eta| < 2.4$ . RPCs are used in the barrel region and TGCs are used in the end-cap region as indicated in Figure 3.8. The trigger chambers are multi-purposed providing bunch crossing identification and  $(\eta, \phi)$  position information in addition to triggering and  $p_T$  threshold information. The RPCs have a resolution in  $z$  of 10 mm and in  $\phi$  of 10 mm [61]. The TGCs have a resolution in  $R$  of 2-6 mm and a resolution in  $\phi$  of 3-7 mm [61].

RPCs consist of two charged plates separated by a 2 mm gap. An electric field of 4.9 keV/mm accelerates the ionisation caused by charged particles as they pass between the two plates, generating an electron avalanche. Strips are mounted on the backs of the plates and are orientated orthogonally to each other to provide a measurement in  $(\eta, \phi)$  from a single unit.

TGCs are multi-wire proportional chambers similar in design to the CSCs. They provide an additional measurement in the  $\phi$  direction which complements the existing detectors in the end-cap as well as providing triggering information.

### 3.2.4 The Magnet System

The ATLAS detector contains four superconducting magnets – a barrel solenoid, a barrel toroid and two end-cap toroids. These are shown in Figure 3.9.

The inner detector sits inside a solenoid which generates a uniform magnetic field of 2 T. The field is aligned in the inner detector with the beam line and the solenoid has been constructed to minimise the material thickness, as it sits between the inner detector and the calorimeter. The thickness in the barrel region is approximately 0.66 radiation lengths [61]. The tile calorimeters in the barrel and the end-cap calorimeters provide an iron yoke for the return flux of the barrel solenoid.

The barrel toroid system consists of eight racetrack coils symmetrically installed in  $\phi$ . These toroids generate a magnetic field in the radial direction surrounding the inner detector and calorimetry system with a peak field strength of 4 T and an average field strength of around 0.5 T.

The end-cap toroid system provides additional bending power for the muon system in the forward region. The end-cap toroids generate a radial magnetic field with a peak field-strength of 4 T and an average field strength of around 1 T.

### 3.2.5 The Trigger and Data Acquisition System

The ATLAS detector uses a three-tier triggering system to select events to be stored offline. Events used in physics analyses typically contain a high- $p_T$  object which provides a signature to select events in data. Triggering is required due to the high rate of collisions at the LHC coupled with the limited write-out speed and storage space available to the experiment. A summary of the TDAQ system in 2012 is shown in Table 3.2.

Table 3.2: Summary of event rates and processing times in the ATLAS TDAQ system comparing the design expectations and the rates during the 2012 data taking period. The Level 1 trigger is a hardware trigger which can only make simplified trigger decisions. Level 2 and the Event Filter (EF) are software triggers capable of making more complex trigger decisions based on partial and full reconstruction of the triggered region of interest.

	Collision Rate	Level 1	Level 2	EF
Design	40 MHz	75 kHz ( $< 2.5\mu s$ )	4 kHz (40 ms)	300 Hz (4 s)
8 TeV Data	20 MHz	70 kHz ( $< 2.5\mu s$ )	6.5 kHz (75 ms)	1 kHz (1 s)

The level 1 (L1) trigger is the first stage of the trigger chain and is a hardware trigger. The

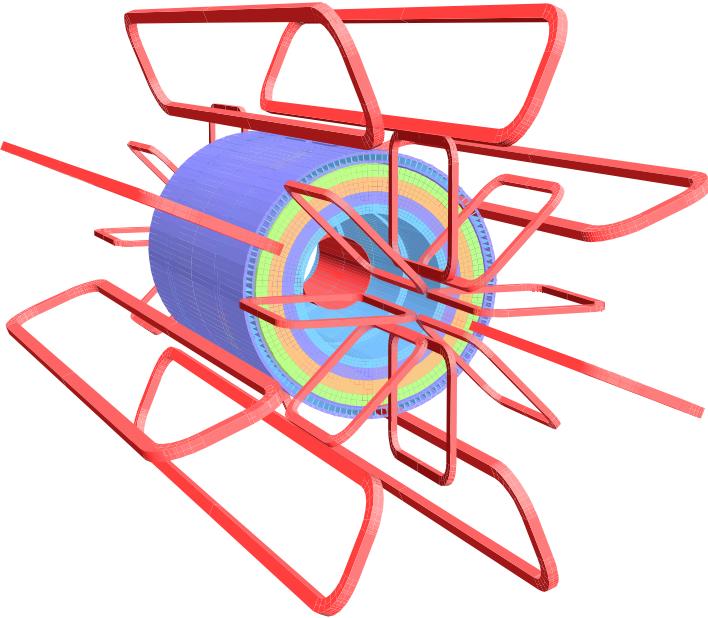


Figure 3.9: Illustration of the magnet system windings within ATLAS. The tile calorimeter, indicated with four layers, contains the central solenoid winding on the inside and the iron yoke on the outside for the return flux. The central solenoid is surrounded by the toroidal windings in the barrel and end-cap regions [61].

trigger decisions made at this level can not be highly complex as they have a limited time with which to identify bunch-crossings of interest using predefined algorithms. Whilst making the L1 decision, all the data recorded from the subsystems needs to be stored in pipeline memories. These are integrated circuits located close to the detector. With LHC design conditions, bunch-crossings will occur every 25 ns. The L1 trigger can only make its decision using the calorimeter system and the muon system. These systems are merged into much coarser resolution for making triggering decisions. For instance, the highly granular calorimeter cells are merged from around 200,000 individual cells to 7168 trigger towers [63]. This grouping of calorimeter cells is highlighted in Figure 3.6 for the barrel ECAL. The  $(\eta, \phi)$  measurements for a trigger tower is recorded if it passes a trigger hypothesis. There is an intrinsic delay between a collision taking place and the detector systems recording data. As such the L1 system is designed to make a decision within  $2.5\mu\text{s}$  of a bunch-crossing and reduce the incoming collision rate of 40 MHz down to a readout rate of 75 kHz.

The subsequent trigger systems are collectively referred to as the high level trigger (HLT) as they are software triggers. The HLT comprises of the level 2 (L2) trigger and the event filter (EF). Once an event passes the L1 trigger, a conversion is made from analogue to digital data and

the region of interest (RoI) which fired the trigger is passed to the L2 trigger. This information is passed to the L2 trigger from the read-out system (ROS). The ROS is a buffer system which implements a push-pull architecture to pass data fragments from detector systems to the trigger system. The L2 trigger requests the data located in the RoI and reconstructs part of the event. The L2 trigger has access to the full granularity of the detector subsystems and the tracking information from the inner detector, both of which are not used in the L1 decision. Trigger chains are seeded by the L1 triggers which then controls how the L2 trigger constructs objects in different detector regions and what trigger hypotheses to test. This system is able to process decisions for multiple bunch-crossings in parallel using a computing farm, where each decision is designed to take 40ms. The triggers applied at the L2 stage have to reduce the event rate of 75 kHz down to a rate of 4 kHz.

The final part of the HLT is the EF which uses the entire event to make a decision. When an event passes a L2 trigger, the full event is extracted out of the ROS and reconstructed offline. This part of the trigger system has the longest allocated time due to the reconstruction of the event and takes up to 4 s. The EF will test hypotheses directly related to the L2 triggers which have passed. If an event passes an EF trigger, then a full trigger chain will have been satisfied from L1 to L2 to EF and this event will be recorded and written to tape. The maximum rate which can be written out from the EF in design conditions is 300 Hz.

In the 2012 LHC runs, the LHC did not run at design specifications. As well as running with a lower centre-of-mass energy, the rate at which events were delivered to ATLAS was reduced by a factor of two. This meant interactions occurred every 50 ns rather than every 25 ns. However, the LHC ran with more protons per bunch than the design specification. For 8 TeV data, there were approximately  $1.7 \times 10^{11}$  protons per bunch rather than planned amount of  $1.1 \times 10^{11}$ . This increased the mean number of collisions per bunch-crossing from 19 to around 40. The average event size therefore increased which caused the processing time and the required bandwidth to increase. However whilst the TDAQ system was operating beyond the design specifications, the system remained reliable throughout Run 1.

### 3.3 Event Simulation

The simulation of physics processes is necessary to make quantitative statements about the nature of the data collected by the ATLAS detector. It is important to model the entire physics process which starts with a partonic interaction from a proton-proton collision and ends with electronic signals being read out by the various detector subsystems. This simulation can be split into two steps; event generation and simulation of particle interactions with the detector.

#### 3.3.1 Event Generation

The simulation of partonic interactions is provided by Monte Carlo generators. There are various stages in the calculation and evolution of an interaction. The hard scattering cross-section of an interaction is provided by a fixed-order transition matrix element which needs to be numerically integrated over the final state phase-space. Given the parton density functions of a proton, the fraction of proton momentum carried by the interacting partons needs to also be taken into account in the integration. Each point sampled in the phase-space provides the parton momenta of a simulated event which can be used to model the process. The output of an event is provided in the standardised HepMC data format which allows a common interface with all event generators [64, 65].

Coloured particles produced by the hard interaction will produce QCD radiation as the energy scale of the process evolves. This is modelled with the parton shower where the DGLAP equations [66–68] are used to evolve the parton splitting probabilities as the energy scale changes from the hard interaction down to energies around 1 GeV. At this point, non-perturbative models are used to simulate the hadronisation of coloured partons into colourless objects using phenomenological models such as the Lund string model [69] or the cluster model [70]. Event generators store the list of stable particles which are then used by the detector simulation. Any particle with  $c\tau < 10\text{mm}$  is decayed by the event generator and interactions with the ATLAS detector are ignored [64].

A final consideration in event generation is to simulate the proton remnants and additional parton interactions which make up the underlying event and result in additional QCD radiation and pile-up vertices. These events can be overlaid in the detector simulation step.

### 3.3.2 Detector Simulation

The detector is simulated using GEANT4 [71] which models the interaction of particles with materials and the response of the detector to such interactions. GEANT4 contains the configuration of the entire ATLAS detector including technical details such as the location of wiring and known misalignments. GEANT4 models the decays of particles as they interact with the detector material and records the energy depositions in the relevant detector components. The ATLAS simulation chain allows multiple types of events to be simulated independently and then overlaid before the final digitisation is simulated. This step calculates the detector signal after summing the contributions from the hard scatter, minimum bias and pile-up events [64]. These classes of event make up the constituents of a single event in data which, when read out by the detector, will be indistinguishable. The final output of the simulation is in the same data format as provided by the detector. This allows the same reconstruction software to be processed on simulation and data.

The ATLAS experiment uses two different forms of detector simulation; a full simulation and a fast simulation. The full simulation uses GEANT4 to model all the detector subsystems which leads to a significant simulation time per event (on the order of 2000 s for a  $t\bar{t}$  event [64]). The fast simulation used by ATLAS is known as ATLFAST II [72, 73]. The inner detector and the muon spectrometers are identical in ATLFAST II and full simulation, but the simulation of the calorimeter differs. The full simulation uses GEANT4 to model the entire calorimeter system whereas the ATLFAST II simulation models the calorimeter with a parameterisation of the calorimeter response and particle showers based on the distribution of active and inactive material in the detector. This results in the ATLFAST II simulation being an order of magnitude faster to simulate an event compared to the full simulation [74].

## Chapter 4

# Object Reconstruction

---

*“The true delight is in the finding out rather than in the knowing.”*

— Issac Asimov

---

Physics processes at the LHC produce particles which are reconstructed by the ATLAS detector into objects. The term object is used to define a final state which has been processed by the detector software to match up with a particular particle hypothesis. Analyses use objects which are broadly categorised as photons, electrons, muons, taus, jets and missing transverse energy. Out of these categories, electrons, muons, jets and missing transverse energy will be used in the subsequent analyses which will be presented in this thesis.

The particles which produce these physics objects interact with material in the detector in different ways, allowing for their identification and reconstruction. Charged particles will leave vertex and tracking information in the inner detector. Electrons, photons and low momentum hadrons will deposit most of their energy in the electromagnetic calorimeter. More energetic hadrons will deposit the remainder of their energy in the hadronic calorimeter. Muons will pass through the inner detector and calorimeters without losing much energy and will subsequently leave tracking information in the spectrometers located on the outside of the ATLAS detector. Finally, the complete coverage in azimuthal angle allows for an inference of particles which have not interacted at all as they traversed the detector by looking for momenta imbalance in the transverse plane resulting in a measurement of the missing transverse energy.

This chapter will provide an overview of how electrons, muons and jets are reconstructed and identified in the ATLAS detector.

## 4.1 Vertexing and Tracking

The inner detector (see Section 3.2.1) provides tracking and vertexing information for charged particles passing through the ATLAS detector.

In order to reconstruct charged tracks, clusters of hits are first used to produce track seeds. Space-points in the three pixel layers and the first layer of the SCT are used to create track seeds. These track seeds are extended out into the SCT in order to form track candidates. The track candidates are fitted using a Gaussian-sum filter which is an extension of the Kalman filter. Each track is allowed to propagate between silicon layers within an associated error. When the track is found to pass through a material layer, multiple subsequent track components are allowed to propagate to the next silicon layer. This forking of the filter allocates a weight to each subsequent track which is propagated through the algorithm and is updated by checking compatibility with the track hypothesis which has been fitted so far. Track forks with weights below a threshold are cut away and the best-fit track using this filter is kept. This method is run for every track candidate allowing for removal of ambiguities. Fakes are cut away by requiring quality cuts on track candidates such as limiting the number of silicon layers where a track passes without leaving a charge reading. The tracks are then extended into the TRT where drift-circle information is used to track the passage through this detector. Once a track is identified through the full inner detector, it is refitted using information from all the inner detector subsystems.

The primary vertex is the term given to the interaction vertex in a bunch-crossing which results in the largest amount of outgoing momentum in the transverse plane. The mean number of interactions per bunch-crossing increased from 9.1 in the 7 TeV run to 20.7 in the 8 TeV run as shown in Figure 4.1. The correct identification of the primary vertex is important as it is used for corrections in high-level reconstruction algorithms. Identifying the primary vertex requires reconstructing all vertices from the reconstructed tracks and then isolating the primary vertex. The  $z$ -position of each track is computed relative to the point of closest approach of the beam spot centre and a seed is identified from the mode of this distribution. An adaptive vertex fitting algorithm [75] is applied which looks at the track around the seed. A  $\chi^2$  fitting procedure is

applied which weights tracks based on their relative position to the seed. Tracks incompatible with the hypothesis of originating from the tested seed by more than seven standard deviations are used to seed new vertices. This procedure is repeated until all tracks are associated with a vertex and refitted [76]. The primary vertex is subsequently identified as the vertex which has the largest sum of squared transverse momenta tracks originating from it, which is written as

$$\text{Primary Vertex} = \max \left( \sum_{\text{trk}} p_{\perp}^2 \right) \quad \forall \text{ vertices.} \quad (4.1)$$

Studies using simulated data with pile-up conditions showed that the primary vertex in  $t\bar{t}$  events is correctly reconstructed 99% of time using this method [46].

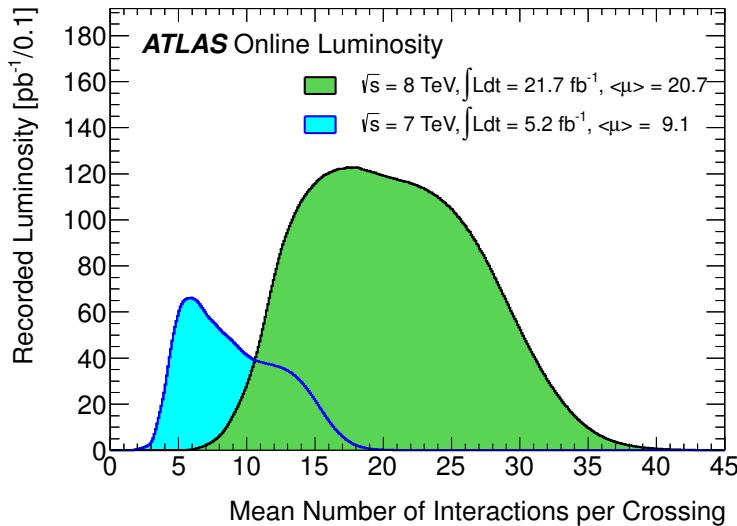


Figure 4.1: Luminosity recorded by the ATLAS detector as a function of the mean number of interactions per bunch-crossing at  $\sqrt{s} = 7 \text{ TeV}$  and  $\sqrt{s} = 8 \text{ TeV}$  [77, 78].

## 4.2 Electron Reconstruction and Identification

Electrons are identified by a track which deposits all its energy in the electromagnetic calorimeter. Due to their low mass, electrons are subject to large energy losses through bremsstrahlung. Around 20-50% of the energy will be lost in this manner in the inner detector before leaving the SCT and this fractional loss is independent of the initial energy. It is important to include a model of bremsstrahlung when identifying and fitting electron tracks, as the radiated photons will have a non-trivial influence on the track momenta. However, it is also important not to ap-

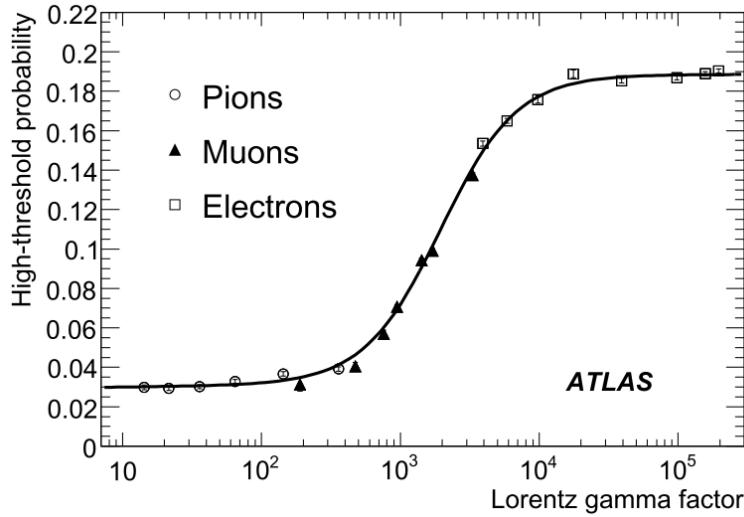


Figure 4.2: The average probability of a high-threshold hit for electron, muons and pions with energies in the range of 2-350 GeV measured in test-beam studies [46]. Given the number of TRT hits will be approximately 20-35 depending on the detector region, the balance of probabilities would indicate that electrons will leave more high threshold hits than pions.

ply a general bremsstrahlung correction to all tracks as this will reduce the overall reconstruction efficiency.

As discussed in Section 3.2.1, the TRT is able to provide electron identification information using high threshold hits coming from the transition radiation emitted when an electron passes through the foil surrounding a straw tube. Electrons with energies greater than 10 GeV should all produce the same level of high threshold readouts as saturation will occur [46]. The amount of transition radiation emitted is dependent on the Lorentz gamma factor of the particle. As this factor will naturally differ between electrons and heavier charged hadrons such as pions, it provides a good method of discrimination and identification of electrons. The probability of a high-threshold hit in the barrel TRT is shown in Figure 4.2 as a function of the Lorentz gamma factor for electrons, pions and muons. The receiver operating characteristic (ROC) curve is shown in Figure 4.3 for identifying electrons correctly by varying the number of high threshold hits required in the TRT.

The measurement of energy deposits in the electromagnetic calorimeter needs to account for three main points of energy loss: out-of-cluster energy, loss of energy before entering the calorimeter and longitudinal energy leakage. All three types of energy loss can be parametrised

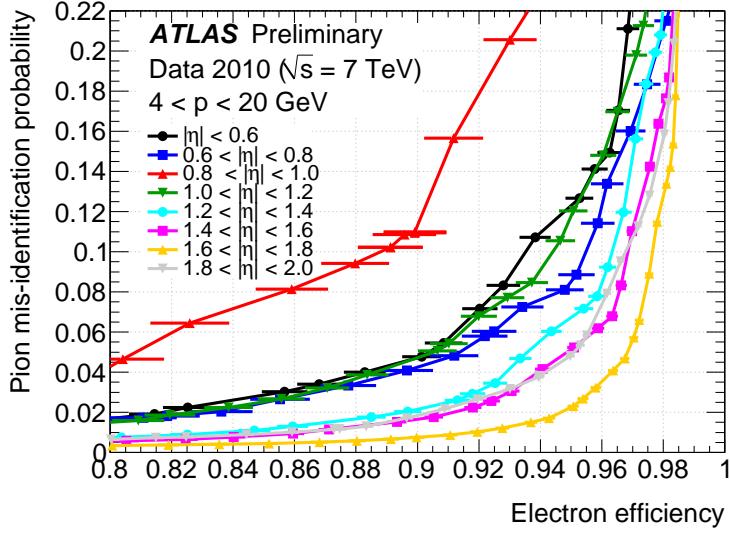


Figure 4.3: The ROC curve showing the relationship between electron identification efficiency and pion misidentification efficiency measured using  $\sqrt{s} = 7$  TeV data by varying the number of high threshold hits in the TRT [62].

using the longitudinal barycentre or shower depth,

$$X = \frac{\sum_{i=1}^3 E_i X_i + E_{\text{PS}} X_{\text{PS}}}{\sum_{i=1}^3 E_i + E_{\text{PS}}}, \quad (4.2)$$

where  $X_i$  is the longitudinal shower depth of each calorimeter layer and PS corresponds to the presampler layer. The energy loss from out-of-cluster energy is linear in  $X$  for a given  $\eta$ . The energy lost before entering the calorimeter is linear in  $X$  for a given  $\eta$  in the region of  $|\eta| < 1.8$  where the presampler is located. The longitudinal energy lost out the back of the calorimeter is calculated as a fraction of the energy measured and is linear in  $\ln(X)$  for a given  $\eta$ . These factors all combine to give a measurement of electron energy with a resolution as shown in Figure 4.4 which follows the form  $\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus 0.17\%$  where the stochastic term ( $a$ ) is fitted between 10-15% for the pseudorapidity range shown [61].

Electron candidates are classified using quality variables which indicate how well reconstructed the objects are. These identification classifications combine calorimeter information such as shower shape variables and tracking information such as the number of high threshold hits in the TRT. A set of standard identification variables are defined based on the rejection of fake electrons. These variables classify electrons into loose, medium and tight, where increasing

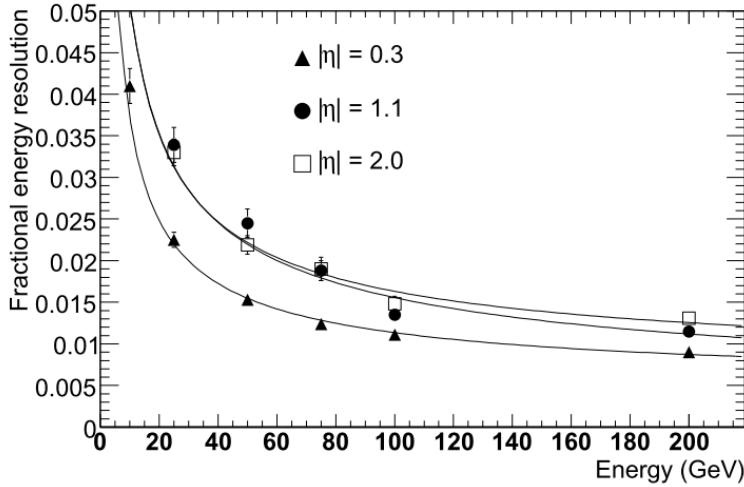


Figure 4.4: Electron energy resolution for fixed points in  $\eta$  as a function of electron energy [61].

tightness indicates an increased electron purity. Different classifications are provided to allow analyses to optimise their signal selections.

These quality conditions, and their associated names, have evolved through the data taking periods to improve the selection efficiencies. The cut-based quality conditions were updated for the 8 TeV run as the increased pile-up conditions meant that some variables with a dependence on the number of vertices became less effective. A further update to the electron quality criteria came from the move away from a cut-based selection to a multivariate likelihood-based one. The improvement with likelihood particle identification variables meant that working points could be defined which either had the same signal efficiency with improved background rejection, or improved signal efficiency and the same background rejection with respect to the cut-based variables. A summary of the efficiencies measured on  $Z \rightarrow ee$  events in data is presented in Figure 4.5 as a function of the number of primary vertices. The distribution in Figure 4.5a shows the efficiency improvement in the cut-based identification variable due to the improved definitions in the 8 TeV dataset compared to the 7 TeV dataset. The distribution in Figure 4.5b then shows how the likelihood-based variables can be tuned to match the performance of the cut-based variables.

Isolation variables are used to reduce backgrounds from non-prompt electrons. The isolation can be track-based or calorimeter based. The variable  $p_T\text{-cone}N$  is a track isolation variable where  $N$  is defined using a cone of radius  $\Delta R < 0.N$ . All tracks found within the cone radius have their transverse momentum summed and a ratio is calculated between electron transverse

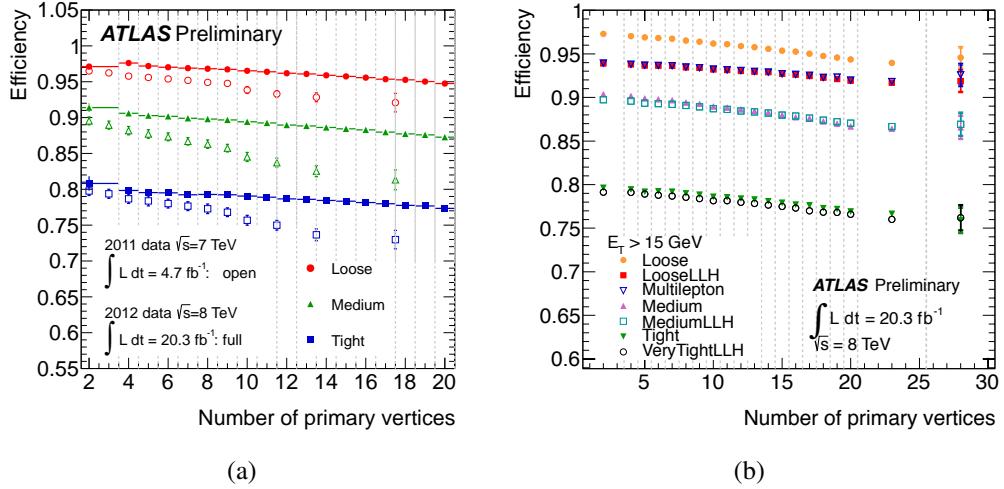


Figure 4.5: Summary of the change in efficiency as a function of the number of reconstructed vertices between 7 and 8 TeV data is shown in Figure 4.5a and a comparison of the cut-based and likelihood-based (LLH) electron identification methods showing their identification efficiencies for 8 TeV analyses is shown in Figure 4.5b as a function of the number of primary vertices [79].

momenta and this summed value. The tracks considered by this variable have quality cuts applied to ensure they come from the same vertex as the electron. The variable  $E_T\text{-cone}M$  is a similar variable which focuses on the isolation of calorimeter energy deposits where  $M$  is defined using a cone of radius  $\Delta R < 0.4$  around the cluster barycentre. A region in the middle of the cone is excluded from this calculation to remove the contribution from the electron. The sum of energies in the calorimeter cells inside the allowed cone region is calculated and used to measure the amount of additional activity around the electron.

These isolation variables measure an absolute amount of activity around an electron. An electron is considered isolated if the activity is less than a threshold amount. Cutting on these variables directly sets the threshold as an absolute amount of activity. Alternatively, the amount of activity can be defined as a relative fraction of the  $p_T$  of the electron and a threshold amount can be fractional instead. It is also possible to measure an isolation efficiency as a function of  $p_T$  such that the isolation threshold varies across  $p_T$  to give a flat efficiency when cutting on these variables.

## 4.3 Muons

The ATLAS detector is designed to reconstruct muons using the large muon spectrometers in the outermost regions of the detector. Electrons and hadrons will mostly be stopped by the calorimeters. Occasionally, it is possible for very energetic particles and hadronic jets to punch-through the calorimeter into the muon spectrometer and fake a muon.

The identification and reconstruction of a muon comes from both the muon spectrometer and the inner detector in order to achieve the best  $p_T$  resolution. The large magnetic field generated by the ATLAS toroids ensures a good  $p_T$  resolution between 2.5-5% for a 100 GeV/c muon and of 10% for a 1 TeV/c muon. In Run 1, ATLAS made use of two different reconstruction algorithms called Muid and Staco to combine standalone muons measured by the muon spectrometer with tagged muons using inner detector and calorimeter information.

The two algorithms differ in the way they search for a pattern of hits in the muon spectrometer and reconstruct the standalone muon. Both algorithms make use of segments which are straight lines connecting hits in the muon spectrometer. The Muid algorithm uses a Hough transform to identify a maximum in parameter space which corresponds to a track fit through the muon spectrometer. The Staco algorithm uses the outermost segment and moves radially inwards, iteratively fitting the segments to create a track. Both reconstruction algorithms include estimations of energy loss in the calorimeters after extrapolating the track through the inner detector to the primary vertex. The tracks found in the muon spectrometer can then be combined and matched to existing tracks in the inner detector in a region of  $|\eta| < 2.5$ . Figure 4.6 shows the power of combining both the muon spectrometer and inner detector tracks to measure the transverse momenta of muons. A summary of reconstruction efficiencies for different selections as a function of pseudorapidity is shown in Figure 4.7 for the Muid reconstruction chain.

The reconstruction efficiency can be improved by using object quality variables similar to the electrons. The variable  $\chi^2_{\text{match}}$  provides a measurement of how well the standalone muon and the inner detector tracks are in agreement with the hypothesis that they are from the same muon. A requirement on the number of hits in the muon spectrometer region also aids the reconstruction efficiency.

Due to the detector construction, the region  $|\eta| < 0.04 - 0.08$  is not instrumented by the muon spectrometer to allow service access to the detector systems. Therefore, tracks which pass through this region suffer from large inefficiencies. These can be recovered through the use of

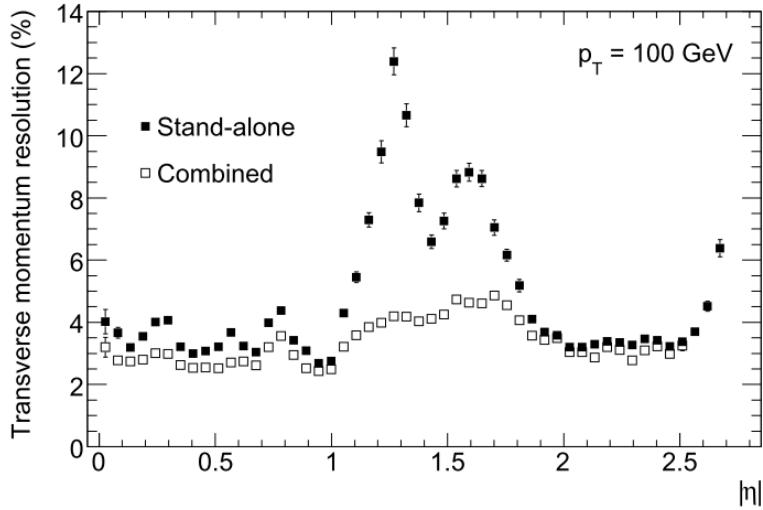


Figure 4.6: The  $p_T$  resolution for 100 GeV/c muons, showing the impact of combining the inner detector tracks with the standalone muon spectrometer tracks [61]. A clear improvement can be identified in the region  $1.1 < |\eta| < 1.7$  which includes the region between the barrel and end-cap regions of the muon spectrometer. Measurements in this region suffer from reduced muon stations to reconstruct hits and reduced bending power in the magnetic field transition.

calorimetric muons, though the reconstruction efficiency is lower than elsewhere in the detector and they are subject to increased errors as can be seen in Figure 4.7.

For all analyses described in this thesis, muons are selected using the tight identification classifier. The tight set of Muid muons contains all combined muons which were found using the Muid algorithm; muons found by an algorithm which searches outside the inner tracker acceptance ( $|\eta| > 2.5$ ) and have at least three hits in the muons stations; and muons found by an algorithm which identifies tracks in the inner detector and extrapolates them to the muons spectrometer where at least two hits, or less than six holes, are required in the muon stations. The muons found by these alternative algorithms are only selected if they have no overlap with the Muid selection. The use of these other algorithms enhances the muon reconstruction efficiency at low  $p_T$ ,  $|\eta|$  close to 0 and  $|\eta| > 2.5$ .

The isolation of muons can be defined with the same type of variable as electrons. Track-based isolation variables use the inner detector track measurement to calculate the activity around the muon within a cone. Calorimeter-based isolation extrapolates the muon track into the calorimeter system and identifies the calorimeter deposits located around the track within the defined radius. The variable can be defined the same way as the electron variable, where an inner cone is removed from the sum, or alternatively an estimate of the transverse energy lost by

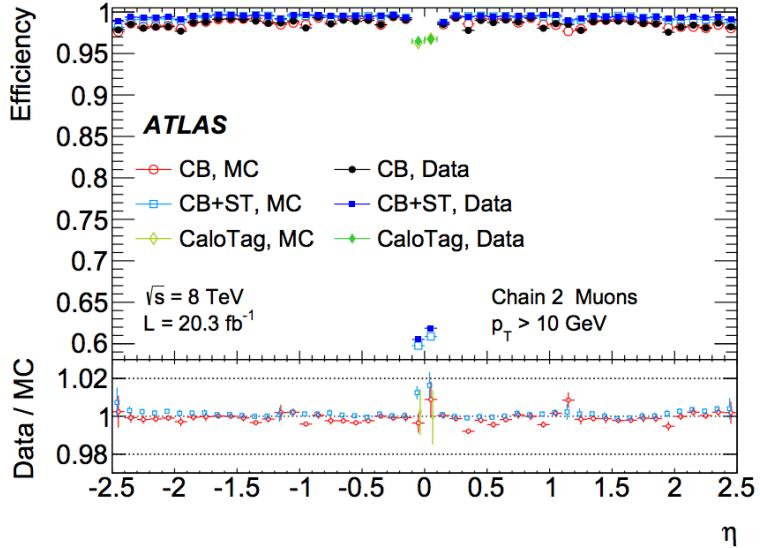


Figure 4.7: Reconstruction efficiencies measured in simulation and data using  $Z \rightarrow \mu\mu$  events [80]. The Muid muon chain is presented here for combined (CB), segment-tagged or combined (CB+ST) and calorimeter tagged (CaloTag) muons. Both the efficiency and the data-simulation agreement are very good for muon reconstruction.

a muon as it traverses through the calorimeter system can be subtracted from the isolation cone measurement.

An alternative isolation variable to fixed cone isolation is called mini isolation [81]. This is a track-based isolation which uses a  $p_T$  dependent cone size. This variable has been shown to have good performance under high pile-up conditions and for muons from highly boosted objects. The variable is defined as  $\text{mini-isolation}_N$  where a cone is defined as having a radius of  $\Delta R < N/p_T$ , where both  $N$  and  $p_T$  are provided in  $\text{GeV}/c$ . A typical value of  $N$  is  $10 \text{ GeV}/c$ . The track momenta are then summed the same way as the normal  $p_T$ -cone variable and required to be below an absolute or relative threshold.

## 4.4 Hadronic Jets

Hard QCD interactions result in a range of softer partons being produced through parton showering. These partons are coloured by definition, but the energies measured in the detector are deposited by collimated collections of colourless hadrons. It is not possible to reconstruct individual coloured objects such as a pair of light quarks produced by the decay of a  $W$  boson. Instead, hadronic jets are used to link experimental observations of hadronic activity to theoret-

ical predictions of quarks and gluons.

Jets are build from an algorithm that takes the calorimeter cell deposits from colourless hadrons and clusters them together to produce a single object. As jets are used to define the hadronic final state of an event, there are theoretical and experimental constraints imposed on their definition. In order to be insensitive to higher order QCD perturbation effects in the parton shower, the boundaries of jets are required to be infrared and collinear safe. This means that soft hadrons which form around around the hard partons are clustered into a single jet, rather than clustering into multiple low  $p_T$  jets and that collinear splitting of partons does not change the number of jets.

The formula for general sequential recombination algorithms uses the following variables,

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \quad (4.3)$$

$$d_{iB} = k_{ti}^{2p}, \quad (4.4)$$

where  $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$  and  $k_{ti}$  and  $y_i$  are the transverse momenta and rapidity of particle  $i$  and  $R$  is the jet radius parameter.

For each particle  $i$ , all values of  $d_{ij}$  with respect to the other particles and the distance with respect to the beam axis,  $d_{iB}$ , are calculated. The smallest value of these variables is defined as  $d_{\min}$ . If this value is calculated from  $d_{ij}$  then the two particles  $i$  and  $j$  are merged, summing their four-momenta. If the value comes from  $d_{iB}$  then the particle  $i$  is defined as a final jet and removed from the list of particles. The algorithm is then iterated until all particles are clustered into a set of jets.

The behaviour of the algorithm is determined by the value of parameter  $p$ . The standard values of this variable are shown in Table 4.1 along with the name and reference for the recombination algorithm. The different behaviour of the recombination algorithms is exemplified in Figure 4.8 where different conclusions are drawn for the same set of calorimeter clusters depending on the value of  $p$ .

ATLAS makes use of the anti- $k_t$  jet reconstruction algorithm with a radius parameter of  $R = 0.4$  for most physics searches. Analyses searching for highly boosted final states typically use the Cambridge-Aachen algorithm with an increased jet radius parameter to identify fat jets formed from two high- $p_T$  objects in close proximity in the  $\eta$ - $\phi$  plane. Substructure techniques

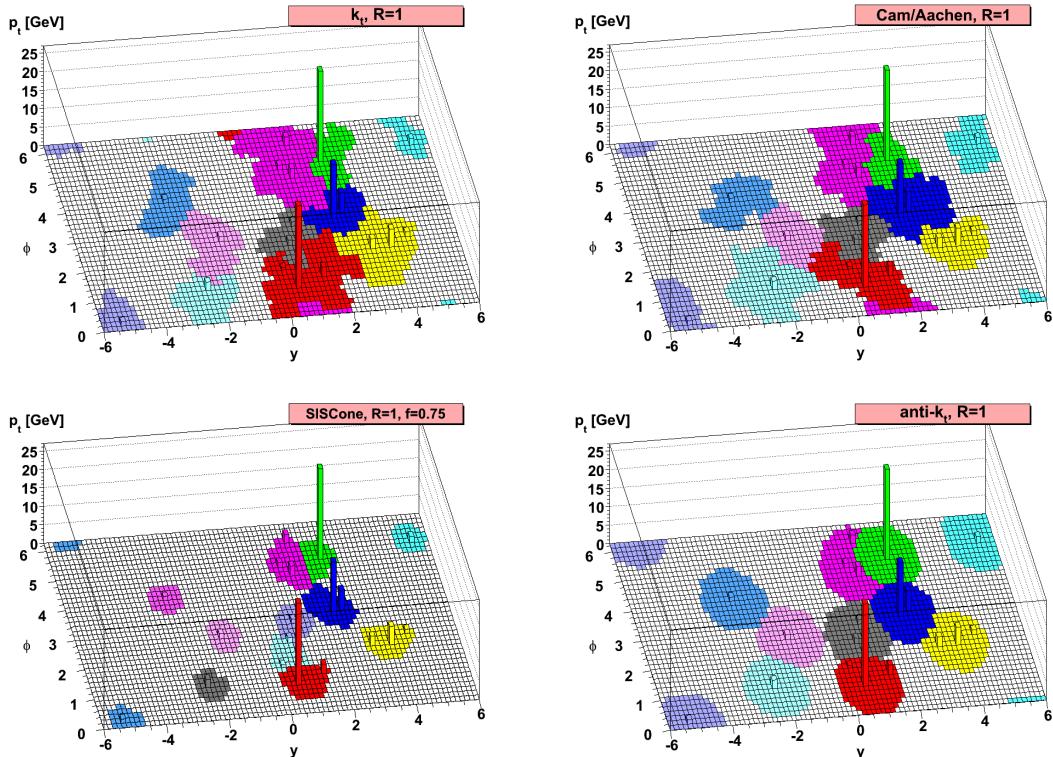


Figure 4.8: Illustration of the variation in jet definitions through the application of different reconstruction algorithms. The anti- $k_t$  algorithm has the desirable behaviour that isolated high- $p_T$  objects will cluster closer to an idealised circle in  $y\phi$  and will be influenced less by softer radiation which is not close to the final state object [86]. A well-defined structure makes it easier to apply general calibration corrections to the reconstructed energies of jets. SiSCone is an alternative infra-red safe jet reconstruction algorithm provided for comparison [87].

Table 4.1: Summary of the sequential recombination algorithms used to construct jet objects by varying the parameter,  $p$ , which is shown in Equation (4.4).

$p$	Algorithm	Reference
1	$k_t$	[82, 83]
0	Cambridge-Aachen	[84, 85]
-1	anti- $k_t$	[86]

are then applied to further study these fat jets and extract smaller radius jet information. These algorithms have the benefit that they can be applied at reconstruction level on calorimeter cells or at truth level either on partons or stable hadrons and exhibit the same behaviour. This aids comparisons between theory predictions and experimental observations.

The calorimeter response in the ATLAS detector is different for energy deposited by an electromagnetic object and energy deposited by a hadronic object. The non-compensating nature of the ATLAS calorimetry system means that the signal for an electromagnetic object is greater than that of a hadronic object for the same energy deposit. Calorimeter cells at this point are said to be at the electromagnetic scale, or EM scale. It is possible to correct the weighting of the energy cells by looking at the shower depth of an energy deposit to correct the response for hadronic objects. Typically a hadronic object will travel further into the calorimeter system than an electromagnetic object because the hadronic interaction length is greater than the electromagnetic radiation length. This correction is called local hadronic cell weighting and calorimeter clusters with this correction are said to be provided at the LC scale. Jets can be reconstructed in the ATLAS detector at either of these energy scales and the granularity of the ATLAS calorimetry system allows for a good measurement of the jet direction.

The reconstruction described for jets does not make use of the inner detector as the recombination of clusters only uses calorimeter information. A number of corrections are therefore required to ensure that the detector response is correctly calibrated and the jet directions are correctly defined with respect to the primary vertex. These are applied to jets after they have been defined at the EM or LC scale as shown in Figure 4.9 to correct their energy to that of truth jets defined at particle level.

A correction is applied to correct the effect of pile-up on the reconstructed jet energies. There are two different types of effects which are associated with pile-up vertices. The first is called in-time pile-up which corresponds to the influence of particles produced from additional proton-proton collisions in a single event which deposit energy in the calorimeters. The second

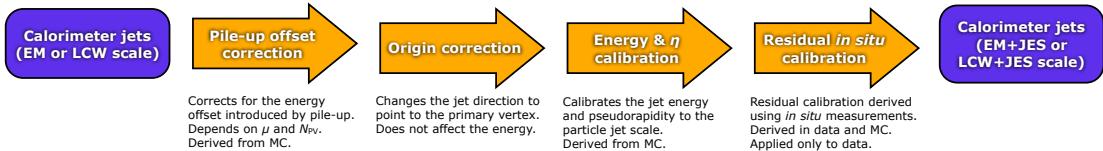


Figure 4.9: A summary of the jet energy scale correction procedure used by the ATLAS experiment [88].

is called out-of-time pile-up which corresponds to the energy deposited by collisions before and after the bunch-crossing in question. A correction is applied based on the jet area as defined in Reference [89]. A measurement of the per jet energy density is sensitive to the impact of in-time pile-up and can be used to derive a correction. The out-of-time pile-up is left largely unaffected and a residual correction is derived from simulated data as a function of the number of vertices in an event and the mean number of interactions per bunch-crossing. The differential  $p_T$  distributions with respect to the number of pile-up vertices and to the average number of interactions per bunch-crossing is shown in Figure 4.10. After this, a small correction is made to the direction of all jets to ensure that they point back to the primary vertex of the event.

A subsequent adjustment is made to correct measured jet energies to the true particle jet energies. The particle jets used as a baseline are composed of all stable hadronic states which are clustered using the same algorithms as the calorimeter clustering. The energy response,

$$\mathcal{R} = \frac{E_{\text{jet}}}{E_{\text{jet}}^{\text{truth}}}, \quad (4.5)$$

is measured in bins of pseudorapidity and true jet energy. The mean of a Gaussian fit is used to calculate  $\langle \mathcal{R} \rangle$  which is used to correct the calorimeter response for different jet energy scales to the particle level energy.

A correction is derived to correct the detector response in data. The correction is measured in-situ using events with a well-defined EM object and one or more jets which should balance in the transverse plane. A measurement of the effective missing transverse energy provides information about the difference in detector response between electromagnetic and hadronic energy deposits. Events containing a photon and jets provide calibration in the range of  $40 < p_T < 400$  GeV/c and events with a Z-boson and jets provide a more accurate calibration in a limited region of  $p_T < 100$  GeV/c.

A summary of the methods used to estimate the jet energy resolution in the 8 TeV dataset can be found in Reference [91]. These measurements found that the jet energy resolution is well

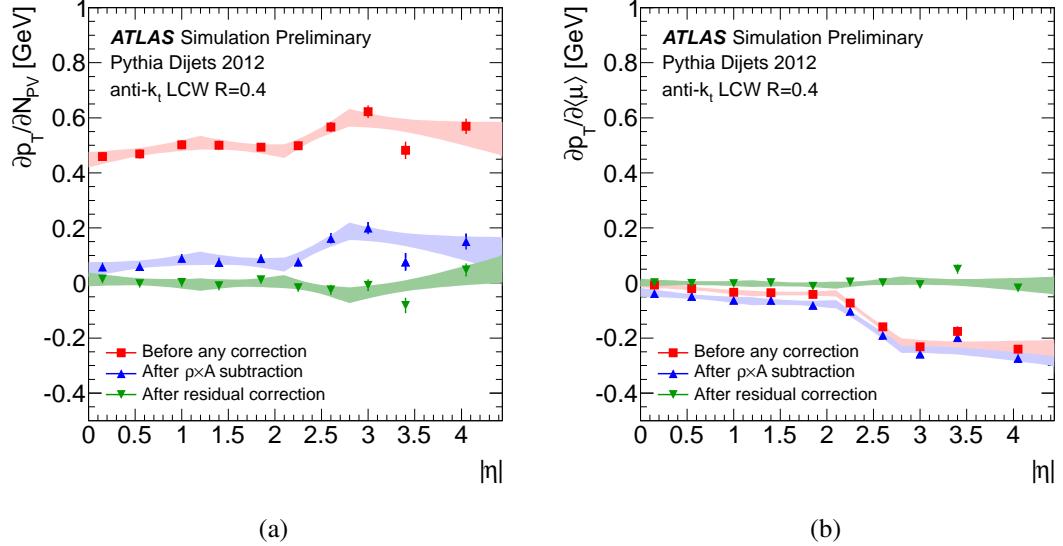


Figure 4.10: Differential distributions of jet  $p_T$  given the number of reconstructed vertices (Figure 4.10a) and given the mean number of interactions per bunch-crossing (Figure 4.10b) as a function of pseudo-rapidity [90]. The dependence of the  $p_T$  on these two pile-up sensitive values is shown to reduce when applying the jet area correction and subsequently applying a residual correction. This dependence was derived from comparing reconstructed jets with truth jets, for a truth-jet  $p_T$  of 25 GeV/c.

modelled in simulated data and as such, no correction is applied in contrast to the jet energy scale. The jet energy resolution has been measured as a function of transverse momentum and has been observed to go from 14% for low- $p_T$  jets down to 5% for high- $p_T$  jets as seen in Figure 4.11.

In contrast to the lepton isolation variables presented in this chapter, jet isolation is not so well-defined. By definition jets are constructed from multiple energy deposits which are clustered together which makes it difficult to use the amount of activity outside the object to define isolation. However, it is possible to reduce the impact on an analysis from jets originating from pile-up vertices. A variable known as jet vertex fraction (JVF) is defined to estimate from which vertex a jet originated [92]. JVF calculates the fraction of charged tracks from each vertex which are matched to the jet and is defined as

$$\text{JVF}(\text{jet}_i, \text{vtx}_j) = \frac{\sum_k p_T(\text{trk}_k^{\text{jet}_i}, \text{vtx}_j)}{\sum_n \sum_l p_T(\text{trk}_l^{\text{jet}_i}, \text{vtx}_n)}. \quad (4.6)$$

If this variable is calculated with respect to the primary vertex it is possible to identify jets which most likely originate from another interaction in the bunch-crossing. If this variable is used in a

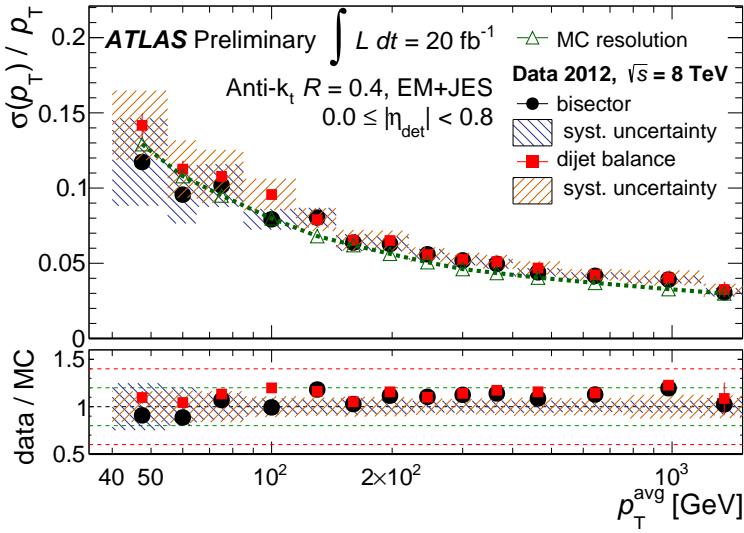


Figure 4.11: Jet energy resolution as a function of  $p_T$  measured in the 8 TeV dataset [91]. The central detector region  $|\eta_{\text{det}}| < 0.8$  is shown here. The uncertainties exhibit the same behaviour with a slightly reduced scale at larger values of  $|\eta|$ .

jet selection, it is only considered for low- $p_T$  jets in the central region with  $p_T < 50 \text{ GeV}/c$  and  $|\eta| < 2.4$ . The  $p_T$  threshold is used because 99% of pile-up jets are in this low  $p_T$  region and the  $|\eta|$  requirement ensures the jets are within the tracker acceptance.

## 4.5 *B*-tagging Hadronic Jets

The search for  $t\bar{t}H$  production is presented in Chapter 6 where the Higgs boson decays to two  $b$ -quarks. However, coloured objects are never physically observed due to QCD confinement. Instead jets are reconstructed in the ATLAS detector from colourless hadrons which are produced through hadronisation of soft partons. Reconstruction of jets has just been discussed but it is relevant to provide a discussion into the methods for identifying jets which originated from a  $b$ -quark. These jets are referred to as  $b$ -jets and a detailed study calibrating the identification efficiencies using data collected with the ATLAS detector is presented in Chapter 5.

The identification of  $b$ -jets is known as  $b$ -tagging and the algorithms designed to identify them are known as taggers. The ability to  $b$ -tag a jet is dependent on finding traits which make hadrons formed with  $b$ -quarks unique compared to other hadrons. An illustration showing the structure of a  $b$ -jet is shown in Figure 4.12. Whilst  $b$ -quarks are the heaviest quarks to form hadronic states, their weak decays are Cabibbo suppressed. That is to say that probability of a  $b$ -quark decaying via the weak interaction is reduced as the decay needs to transition between

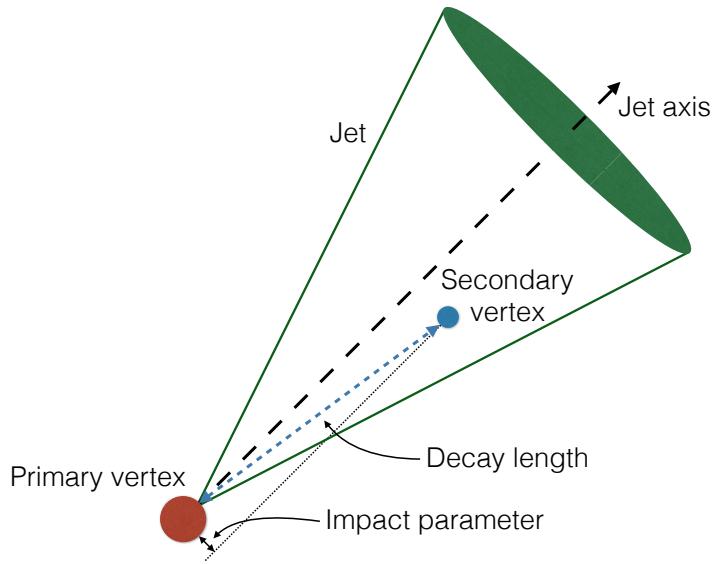


Figure 4.12: Illustration of the formation of a secondary vertex from a  $B$ -hadron decaying in-flight within a reconstructed jet.

two quark generations. This results in a lifetime which allows it to travel on the order of millimetres from the interaction point once the Lorentz factor is taken into account<sup>1</sup>. This decay length is enough to travel away from the primary vertex, but still decay within the pixel layers forming a displaced (or secondary) vertex. Hadrons containing charm quarks also exhibit a similar decay length<sup>2</sup> which can produce a secondary vertex, so this alone is not sufficient to identify  $b$ -jets. As the  $b$ -quark forms the heaviest hadronic states, the invariant mass of the secondary vertex will be larger than typical secondary vertices from other long-lived particles. In addition to this lifetime information, it is also possible to identify a  $b$ -jet through the semi-leptonic decay of a  $B$ -hadron inside a jet where a soft muon is produced.

Generally, the most powerful  $b$ -tagging algorithms make use of a combination of lifetime, decay length and secondary vertex measurements. The design resolution of the radial position of a reconstructed secondary vertex is shown in Figure 4.13. The highest resolution is within the layers of the pixel detectors, and as such it is important to ensure that this detector can cope with high track multiplicities in order to isolate secondary vertices for use in  $b$ -tagging.

A  $b$ -tagging efficiency is defined as the number of true  $b$ -jets correctly identified by a  $b$ -tagging algorithm out of the total number of true  $b$ -jets. Mathematically, this has the form

<sup>1</sup>The mean lifetime of a  $B^+$  hadron is  $1.638 \pm 0.004$  ps which gives a  $c\tau$  of 0.5 mm [30].

<sup>2</sup>The mean lifetime of a  $D^\pm$  hadron is  $1.040 \pm 0.007$  ps which gives a  $c\tau$  of 0.3 mm [30].

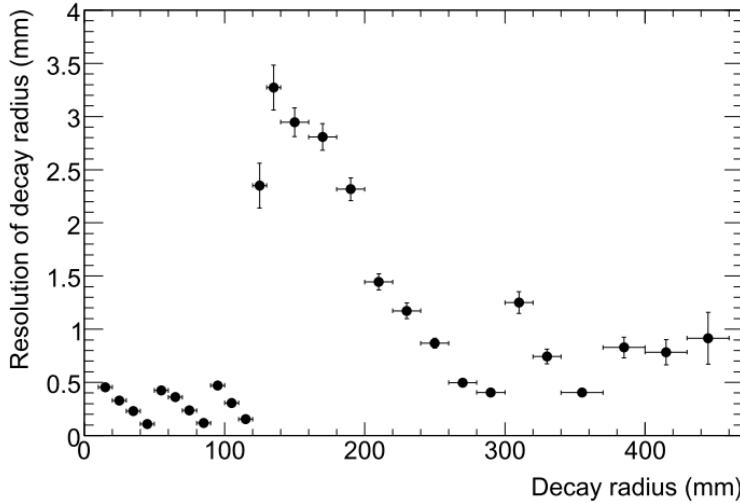


Figure 4.13: The resolution of reconstructing the radial position of a secondary vertex of  $K_S^0 \rightarrow \pi^+\pi^-$  in events containing a  $B$ -hadron decay [61]. The impact of the three layers of pixel detectors can clearly be identified.

$P(\text{tagged as a } b\text{-jet} | b\text{-jet})$ . A non  $b$ -jet which is tagged as a  $b$ -jet is referred to as a mistagged jet. Jets are categorised by their truth information in the context of  $b$ -tagging efficiencies. A spatial matching is performed at the parton level between quarks and the jet axis. If a quark is found to be within  $\Delta R < 0.3$  of a jet, the jet is labelled based on the flavour of the hadron. The matching algorithm hierarchy goes from  $b$ -quarks to  $c$ -quarks to  $\tau$  leptons to light quarks and gluons. If a match is successful then the jet is assigned that flavour and the algorithm stops iterating. For the discussion in this chapter, jets which are not successfully matched to a heavy flavour quark ( $b$  or  $c$ ) are defined to be light flavour jets.

Multiple  $b$ -tagging algorithms exist for reconstructing  $b$ -jets with the ATLAS detector. They all make use of tracks which have been associated to reconstructed jets and pass a selection to reduce the impact of long-lived particles and fake tracks on the  $b$ -tagging efficiency [93].

**Impact Parameter Tagging** The IP3D tagger uses the impact parameters  $d_0$  and  $z_0$  calculated with respect to the primary vertex of selected charged tracks associated to a jet. The signed significance is calculated as the value of one of these variables divided by the error on that measurement. The two-dimensional histogram of these two signed impact parameter significances is used to derive probability distribution functions for different jet flavour hypotheses. Charged tracks coming from the decay of a  $B$ -hadron will have larger impact parameters than charged tracks coming from the primary vertex due to the displaced nature of the secondary vertex. The

IP3D tagger constructs a log-likelihood ratio to discriminate between  $b$ -jets and light jets using the values and correlations of these impact parameters.

**Secondary Vertex Tagging** The SV taggers reconstruct an inclusive secondary vertex. An inclusive secondary vertex applies the assumption that there is only one displaced vertex to be identified. More complex vertex fitting can account for the exclusive weak decay chain of a  $B$ -hadron which results in tertiary vertices. The inclusive reconstruction of a secondary vertex constructs all possible track pairs and examines the quality of the vertex from their intersection. The invariant mass of the track pairs is used to remove tracks which are from  $K_S^0$  and  $\Lambda^0$  decays and  $\gamma$  conversions. The vertex must also be sufficiently displaced from the primary vertex to be considered. All tracks which pass this selection are then used to reconstruct a single vertex. If the vertex fit shows a poor  $\chi^2$  then the track with the largest contribution is removed and the process is iterated until a good vertex fit is found.

The SV0 tagging algorithm uses the signed flight length significance as the discriminating variable. This variable is the signed flight length from the primary vertex to the secondary vertex divided by the error on the flight length measurement. The SV1 tagging algorithm evaluates a log-likelihood ratio combining multiple discriminating properties of the secondary vertex of a  $b$ -jet: the invariant mass; the ratio of energies from tracks associated to the secondary vertex with tracks associated with the jet; the number of two-track vertices found and the  $\Delta R$  between the jet direction and the  $B$ -hadron flight direction.

**Multivariate Taggers** A  $B$ -hadron will often decay weakly to a charm hadron whilst it is in flight. This charm hadron will travel before decaying and can be identified by a tertiary vertex located inside a  $b$ -jet. The JetFitter algorithm uses this weak decay chain of a  $B$ -hadron to a charm hadron and applies a Kalman filter to identify a displaced vertex from the charm decay on the flight path of the  $B$ -hadron [94]. An illustration of the decay process is shown in Figure 4.14. The assumption that the tertiary vertex will be situated along the flight path of the  $B$ -hadron is based on the short lifetime of the charm hadron and which leads to a small lateral displacement which will be on the same order as the tracking resolution [94].

The JetFitterNN algorithm combines information relating to the topological fit along with the invariant mass of all charged particles in the decay chain, the ratio of energies of those particles with the total energy of all charged tracks associated with the jet and the signed flight

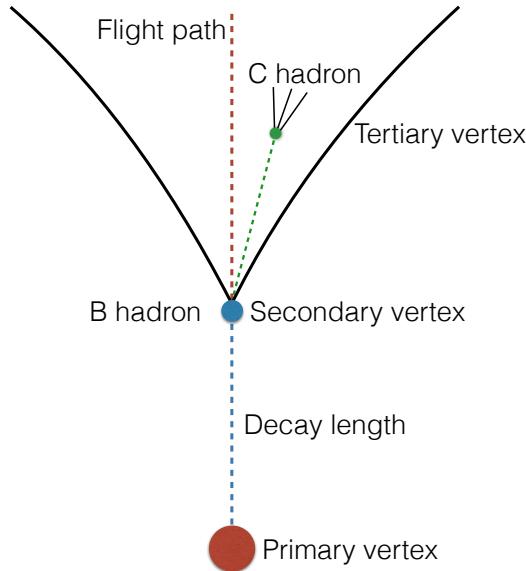


Figure 4.14: An illustration of the weak decay chain from a primary vertex of a  $B$ -hadron to a charm hadron. The lateral displacement is of the same order as the tracking resolution [94].

length significance of the secondary vertex. These variables are combined in a neural network with three output nodes corresponding to  $b$ ,  $c$  and light hypotheses. A log-likelihood ratio is then constructed between the weights of two chosen hypotheses. The JetFitterCOMBNN algorithm includes the output from the IP3D tagging algorithm as an additional neural network input. The final discriminating variable is constructed from the  $b$ -jet likelihood and the light jet likelihood. It is possible to construct a charm tagging variable known as JetFitterCOMBCharm from the outputs of the JetFitterCOMBNN algorithm by constructing likelihood ratios from the neural network outputs with the charm jet likelihood with respect to the other jet hypotheses.

The most commonly used multivariate classifier is the MV1 tagger [93]. This  $b$ -tagging algorithm is a neural network which takes as input the output from IP3D, SV1 and JetFitter-COMBNN. The MV1 tagger is trained to separate  $b$ -jets and light jets and the performance of this tagging algorithm can be seen in Figure 4.15. An alternative training is performed for the MV1c tagging algorithm, which is trained against both light and charm jets to improve the charm jet rejection. Operating points are defined using the measured  $b$ -tagging efficiency on simulated  $t\bar{t}$  events. For a tagging probability with MV1 of  $P(b\text{-tag} | b\text{-jet}) = 70\%$ , the mistag rate for light jets is  $P(b\text{-tag} | \text{light jet}) < 1\%$ .

In the algorithms which make use of multivariate classifiers, care is taken to ensure that trained discrimination does not unintentionally come from the  $p_T$  and  $\eta$  differences between

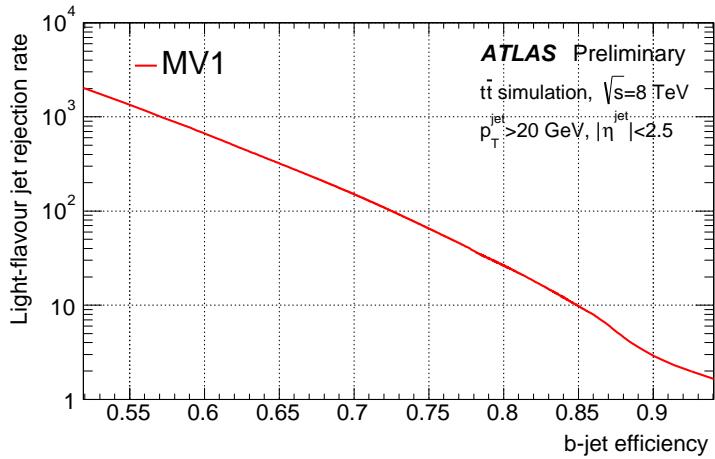


Figure 4.15: The MV1  $b$ -tagging efficiency versus light-jet rejection ROC curve calculated from simulated top-quark pair events produced at  $\sqrt{s} = 8$  TeV [95].

$b$ -jets and light jets. To remove this dependence, the input distributions are reweighted to be flat in coarse 2D bins in  $p_T$  and  $\eta$ .

## 4.6 Overlap Removal

The algorithms used to reconstruct the objects described in this chapter do not use the detector signals in a mutually exclusive way. This can result in the same detector components being used to reconstruct different objects. Physics analyses are therefore required to remove objects if it is thought that there is potential overlap. Overlap removal is performed using a spatial matching between objects and a hierarchy is applied to determine the order in which objects are kept and removed. The ordering of the following sections follows the typical application of overlap removal requirements in top physics analyses during Run 1.

### 4.6.1 Electrons and Jets

The jet reconstruction algorithm discussed in Section 4.4 only uses the calorimeter information to create jet objects. The lack of inner detector information means that electromagnetic calorimeter deposits associated with an electron can be included in the jet reconstruction. To avoid using jets which are likely to be reconstructed from electron energy deposits, jets which are found to be within  $\Delta R < 0.2$  are removed from the selection.

### 4.6.2 Jets and Electrons

There are circumstances which can conspire to provide an instance when a hadronic object could be reconstructed as an electron. Hadronic jets are constructed using the calorimeter energy clusters of charged and neutral hadrons. There will be a low probability that a collimated collection of hadrons will be comprised of a single charged hadron and a number of neutral hadrons. Typically the hadrons in jets are pions, and neutral pions will decay to two photons. In this situation, a single charged track will end up pointing to a large electromagnetic energy deposit, and will be reconstructed as an electron. To avoid using electrons which may have been reconstructed in this way, electrons which are found within  $\Delta R < 0.4$  are removed from the selection.

### 4.6.3 Jets and Muons

Heavy flavour hadrons (hadrons containing charm or bottom quarks) can often decay via the weak interaction through the emission of a  $W$  boson. In approximately 11% of  $W$  decays, a muon and neutrino will be produced. Often the decays of heavy flavour hadrons produce low- $p_T$  muons, but there will be occasions when the decay of a hadron will produce a muon with enough momentum to pass through to the muon spectrometer. In this situation, as the decay will have occurred within the pixel layers of the inner detector, the muon could be reconstructed as a prompt object. To avoid using muons which have been produced through a hadronic decay, muons found within  $\Delta R < 0.4$  are removed from the selection.

### 4.6.4 Muons and Electrons

Muons can occasionally radiate a hard photon via bremsstrahlung. This will result in an energy deposit which appears to share the same inner detector track as the muon and can therefore be reconstructed as an electron. In order to prevent a bias entering the analysis through the reconstruction of muons in the presence of hard photons, the entire event can be vetoed. This technique is used in the  $b$ -tagging analysis discussed in Chapter 5 but is not used in the  $t\bar{t}H$  analysis discussed in Chapter 6.

## Chapter 5

# B-Tagging Calibration in ATLAS

---

*“All science is either physics or stamp collecting.”*

— Ernest Rutherford

---

The  $b$ -tagging algorithms presented in Chapter 4 are used in many ATLAS physics analyses. It is possible to measure the efficiency of a  $b$ -tagging algorithm by counting the number of true  $b$ -jets that are  $b$ -tagged. However, this calculation can only be performed on simulated data where one has access to the truth record of a generated event. To calculate the  $b$ -tagging efficiency in data requires a selection of events with a high purity of true  $b$ -jets which are accurately modelled. A  $b$ -tagging calibration is then defined as calculating the factor required to correct the  $b$ -tagging efficiency measured on simulated data to the  $b$ -tagging efficiency measured using real data.

This chapter will provide a short introduction on why calibration of these algorithms is required and what methods have been established to calculate the  $b$ -tagging efficiency on data in Section 5.1. This will be followed in Section 5.2 by a detailed description of the  $b$ -tagging calibration carried out by the author to achieve ATLAS authorship<sup>1</sup>. The work presented in this section was completed on behalf of the ATLAS Flavour Tagging Combined Performance Group using the full 8 TeV dataset. The entire analysis and final calibration results presented are the

---

<sup>1</sup>In order to gain authorship within the ATLAS Collaboration, a member needs to carry out work which benefits the collaboration as a whole. The work is known as service work or an authorship task and takes a year to complete.

author's own work.

## 5.1 *B*-Tagging in ATLAS

The identification of jets formed through the hadronisation of  $b$ -quarks is important for a number of physics analyses. For example, for a Higgs mass of  $125\text{ GeV}/c^2$ , the dominant decay process is  $H \rightarrow b\bar{b}$  and identification of this process will provide one of the main fermionic coupling measurements. This measurement is important for determining the properties of the Higgs particle and comparing to the Standard Model expectations. In addition, in many supersymmetric models, the decay of a supersymmetric particle to the lightest stable supersymmetric particle will often produce  $b$ -quarks along the chain. Supersymmetry is one of the leading theories for physics beyond the Standard Model and the search will be influenced by  $b$ -tagging efficiencies. The top quark has a branching ratio to  $b$ -quarks of  $\sim 100\%$  via the weak interaction. Any analysis investigating the properties of the top quark will be heavily dependent on the identification efficiency of any flavour tagging algorithms. This efficiency needs to be quantified in data in order to improve comparisons between prediction and data.

### 5.1.1 Calibration of Algorithms

The performance of a  $b$ -tagging algorithm can be different in simulation and data due to inherent shortcomings in the detector and physics simulation. The measurements required to correct this performance in simulation constitutes the calibration. Calculating the  $b$ -tagging efficiency on data presents a tricky task as *a priori* there is no knowledge about the true flavour of a reconstructed jet in data. Therefore one needs to make a selection on data to enhance the true  $b$ -jet purity. Once the  $b$ -tagging efficiency is measured in data, the correction to simulated true  $b$ -jets is provided as a  $p_T$  dependent scale factor constructed from the ratio of the efficiency in data over the efficiency in simulation. Additional measurements are made to quantify the mistag rate associated with  $b$ -tagging algorithms which comes from  $b$ -tagging jets which are not true  $b$ -jets, in order to fully correct the modelling of  $b$ -tagging in simulated data.

There were three methods implemented using the 7 TeV dataset collected by the ATLAS detector in 2011 to calibrate the  $b$ -tagging algorithms. Two of these calibration are muon-based methods known as the  $p_T^{\text{rel}}$  method [96] and the System8 method [96]. Both these methods had

been previously used at the Tevatron to measure  $b$ -tagging efficiencies in data [97]. The third ATLAS method was a newly introduced one which used top-quark pair events in data [98]. The  $t\bar{t}$  production cross-section is a factor of 25-35 times greater at the LHC compared to the Tevatron. This increased production provides a statistically competitive sample of top-pair events to use for  $b$ -tagging calibrations at the LHC. These events are desirable for  $b$ -tagging calibrations as they avoid a dependency on the semi-leptonic decay of  $B$ -hadrons inside the  $b$ -jets which is used to tag events with the muon-based calibration methods. This dependency was thought to introduce a slight bias to the calibration as a number of 7 TeV analyses which used the data to constrain systematic uncertainties found that the data was better described when the  $b$ -tagging scale factors were adjusted within their errors.

## 5.2 Measurement of $B$ -Tagging Efficiencies in 8 TeV Data

The work presented in this section details the calibration of four different  $b$ -tagging algorithms using the full 2012 dataset collected by the ATLAS detector. This analysis uses a calibration method to measure the  $b$ -tagging efficiency in data using top-quark pair events based on the calibration method used with the 7 TeV dataset [98]. These events are selected using a dileptonic kinematic selection. Dileptonic top-quark pair events provide a clean sample of inclusive  $b$ -jets which can be used to measure the  $b$ -tagging efficiency in data.

The numbers derived with this calibration method were the recommendation for all physics analyses which used  $b$ -tagging and presented results at the Hadron Collider Physics Symposium in 2012 and the Rencontres de Moriond conference in 2013. For the full 8 TeV dataset, these numbers were combined with additional calibration methods for improved precision and remained in the group’s recommendations for various working points and jet collections until early 2014 after which a more precise calibration method using a combinatorial likelihood approach became available [99]. The motivation behind carrying out this work was to ensure that a full set of  $b$ -tagging calibrations were available for the  $H \rightarrow b\bar{b}$  analyses pushing for results with the full set of 8 TeV data, such as the search for  $t\bar{t}H$  production presented in Chapter 6.

Details of the object and event selection used to select dileptonic top-pair events in data are presented in Section 5.2.1-5.2.2. A summary of the simulated samples used to provide predictions in this analysis is given in Section 5.2.3. Events are separated depending on the number

of electrons and muons in the final state. Control plots for these different dileptonic states are provided in Section 5.2.4-5.2.5 with discussions about the quality of the data modelling. A detailed discussion about the events which passed the final selection and the calibration method used in this analysis is presented in Section 5.2.6-5.2.7. This is followed by details regarding all the systematic uncertainties considered within this analysis in Section 5.2.8. Finally a full set of results are presented in Section 5.2.9 to correct the  $b$ -tagging efficiency in simulated data to that measured in real data for the MV1  $b$ -tagging algorithm operating at the 70% working point.

### 5.2.1 Object Selection

The physics objects used in this analysis are electrons, muons and jets. Each has a set of isolation criteria, simulation corrections, scale factors and kinematic selection criteria to best isolate the signal sample in data.

The object selection used in this analysis followed the Top Physics Working Group (TopWG) recommendations which were implemented in the TOPROOTCORE framework [81]. Some differences were introduced in this analysis because of technical issues using files produced by the TopWG and the Standard Model Physics Working Group (SMWG). The framework was adapted to use the files produced by the SMWG and the recommendations from the physics performance groups for Moriond 2013 analyses were implemented independently.

The main difference between the selection recommended by the TopWG and the one implemented in this analysis is the muon isolation requirements. The TopWG recommends the use of mini-isolation (see Section 4.3) but this variable was not calculated, nor was it calculable, in the files produced by the SMWG. A previous recommendation from the TopWG existed using isolation cones, so this recommendation was followed instead. An additional tightening of the lepton isolation was applied offline during checks of the analysis and it was seen to improve the agreement with data in the transverse momenta spectrum of the selected jets used in the calibration.

**Electron Selection** The standard selection for a good reconstructed electron is defined using the requirements in Table 5.1. All reconstructed electrons follow the baseline recommendations and calibrations from the Egamma Combined Performance Group.

To remove jets reconstructed from electron calorimeter energy deposits, any jet found within

$\Delta R < 0.2$  of an electron is removed from the jet selection. Following this removal, any electron that is found within  $\Delta R < 0.4$  of a jet is removed from the electron selection to ensure only prompt electrons are used in the analysis.

Two single electron triggers were used in this analysis with different transverse momenta thresholds. The lowest  $p_T$  threshold is at 24 GeV/ $c$  with trigger isolation requirements. This trigger has a high efficiency for low  $p_T$  electrons, but suffers at high  $p_T$  due to the isolation requirement. A higher threshold trigger is included without an isolation requirement to recover this efficiency with a threshold of 60 GeV/ $c$ .

Table 5.1: Summary of the selection criteria for identifying good electrons for which definitions can be found in Section 4.2.

Criterion	Requirements
Definition	Cluster-based reconstruction
Transverse momenta	$p_T \geq 25$ GeV/ $c$
Pseudorapidity	$ \eta  \leq 2.47$
Isolation	Isolation efficiency at 90% for $p_T$ -cone30 Isolation efficiency at 90% for $E_T$ -cone20 $p_T$ -cone30/ $p_T \leq 4\%$ $E_T$ -cone20/ $E_T \leq 7\%$
Identification	Tight
Additional	Reject the crack region $1.37 <  \eta  < 1.52$ $ Unbiased z_0  < 2$ mm

**Muon Selection** Muons in the ATLAS detector can be identified from hits in the inner tracker and in the muon spectrometer. Combined muons are used in this analysis which are identified with the Muid algorithm to combine inner detector and muon spectrometer tracks.

A good muon is defined using the criteria shown in Table 5.2 along with the standard recommendations and calibrations from the Muon Combined Performance Group.

To remove muons which may have been produced from a hadronic decay, any muon found within  $\Delta R < 0.4$  of a good jet is removed. If an electron and a muon are found to share the same track then the entire event is vetoed.

Single muon triggers are used in this analysis to select events in data. Two triggers are used with similar  $p_T$  thresholds to the electron triggers for the same reasoning. A low threshold trigger with isolation requirements is used with a threshold of 24 GeV/ $c$  and a higher threshold trigger without isolation is used with a threshold of 36 GeV/ $c$ .

Table 5.2: Summary of the selection criteria for identifying good muons for which definitions can be found in Section 4.3.

Criterion	Requirements
Definition	Combined Muid
Transverse momenta	$p_T \geq 25 \text{ GeV}/c$
Pseudorapidity	$ \eta  \leq 2.5$
Isolation	$p_{T\text{-cone}30} \leq 2.5 \text{ GeV}/c$ $E_{T\text{-cone}20} \leq 4.0 \text{ GeV}/c$ $p_{T\text{-cone}30}/p_T \leq 4\%$ $E_{T\text{-cone}20}/E_T \leq 7\%$
Identification	Tight
Additional	$ \text{Unbiased } z_0  < 2 \text{ mm}$

**Jet Selection** The standard jet type used in the ATLAS Collaboration is calorimeter anti- $k_t$  jets with a distance parameter of 0.4, as discussed in Section 4.4. Calorimeter jets are built from energy deposits in topological calorimeter clusters (topo-clusters). The calorimeters in the ATLAS detector are non-compensating which means that detector response differs for electromagnetic and hadronic interactions. Two different energy scales can be therefore be used to construct jets in ATLAS. Jets reconstructed at the EM scale uses the raw energy measured with the calorimeter clusters. An additional jet energy calibration can be applied using local cluster weighting (LC) which forms topo-clusters at the LC scale. This topo-cluster calibration uses energy corrections depending on whether the topo-cluster is classified as electromagnetic or hadronic. The jet reconstruction algorithm is run on the calorimeter clusters at either of these energy scales. After the jet finding is complete, a set of jet energy scale (JES) calibrations are applied to create jets at the EM+JES and LC+JES scale (herein referred to as EM and LC jets).

There is not a one-to-one correspondence between EM jets and LC jets and each collection requires an independent set of  $b$ -tagging baselines and efficiency calibrations to be derived. LC jets are designed to improve the resolution of jet energy measurements, but their use varies between analysis groups. Hence it is important to measure the  $b$ -tagging efficiency for both jet collections.

A good jet is defined using the criteria shown in Table 5.3 along with recommendations and calibrations from the JetETMiss Combined Performance Group. The  $b$ -tagging efficiencies are measured for two different JVF working points for each of the jet collections in order to provide a range of calibrations. In this chapter, the two groups of JVF selections will be referred to as *no JVF* or *JVF cut applied*, each of which refers to the pile-up suppression criteria shown in Table

### 5.3.

All plots and results presented in this chapter use LC jets with a JVF cut applied as these jets are consistent with those used in the  $t\bar{t}H$  analysis presented in Chapter 6.

Table 5.3: Summary of the selection criteria for identifying good jets.

Criterion	Requirements
Definition	anti- $k_t$ ( $R = 0.4$ )
Transverse momenta	$p_T \geq 20 \text{ GeV}/c$
Pseudorapidity	$ \eta  \leq 2.5$
Pile-up suppression	no JVF or $\text{JVF} > 0.5$ for $p_T \leq 50 \text{ GeV}/c$ and $ \eta  \leq 2.4$

## 5.2.2 Event Selection

A number of event level requirements are used to select good dileptonic top-quark pair events. Events are required to have at least two jets and exactly two opposite sign leptons. Events are classified by the number and type of leptons selected. The same-flavour and different-flavour events have different background contributions so they are treated separately during the event selection. A summary of the specific event selection criteria is presented in this section.

**Primary Vertex** The primary vertex in an event is defined within ATLAS as the vertex with the highest sum of transverse track momenta originating from the vertex. A requirement is then made in this analysis to only accept events which have at least four tracks coming from this vertex.

**Good Runs List** The Luminosity Working Group provide a list of luminosity blocks for each data taking period where the detector was functioning correctly. There can be cases where detector subsystems go offline which results in events which cannot be correctly or fully reconstructed. Only events which pass the good runs list (GRL) are taken from data and this list defines the maximum luminosity of good physics data that can be used. For the full 2012 dataset, the GRL used in this analysis corresponds to an integrated luminosity of  $20.34 \text{ fb}^{-1}$ .

**Triggers** A logical OR is implemented between the four triggers listed in the lepton selections. Events with two electrons are required to have fired the electron trigger and events with two

muons are required to have fired the muon trigger. A spatial matching of  $\Delta R < 0.15$  is required between a selected lepton and the lepton trigger object.

The ATLAS detector records data in streams according to the fired triggers. The data used in this analysis are the combination of the EGAMMA and MUON streams. It is possible for an event to end up in both streams if both the electron and muon triggers have fired. This is likely to occur in dilepton events where both an electron and muon are produced. To avoid having duplication of events in data, a selection is performed to prevent the overlap of events getting used twice. Any event in the EGAMMA stream will have fired the electron trigger and all events are used. Events in the MUON stream are only selected if the muon trigger and not the electron trigger has been fired.

**Jet Cleaning** Recommendations for rejecting events which have problematic jets are provided by the JetETMiss Combined Performance Group. An event is rejected if a selected jet has been classified as BadLooseMinus. This is the loosest level of jet cleaning. These jets typically have a problem with the calorimeter measurement, where there may be a large fraction of energy deposited in the electromagnetic calorimeter or in a single calorimeter layer. These cuts protect the analysis from events being used which appear to have some form of non-physical energy spike in a calorimeter or from cosmic muons interacting with the calorimeter.

**Pile-up** In order to get a good description of the mean number of interactions per bunch-crossing ( $\langle \mu \rangle$ ), a reweighting procedure is applied to each simulated sample based on the  $\langle \mu \rangle$  distribution measured in data. It is the recommendation of the Inner Detector Tracking Performance Group that the best description between data and simulation is given by a 1.11 scaling of the  $\langle \mu \rangle$  value when using the 8 TeV ATLAS parton shower tunes. This means that for a given  $\langle \mu \rangle$  in simulation, the number of pile-up vertices is best described by a value of  $1.11 \times \langle \mu \rangle$  in data.

**Dilepton Channel Selections** The missing transverse momentum is defined as the magnitude of the vector sum of the transverse momenta of all objects in an event. In the  $ee$  and  $\mu\mu$  channel, the missing transverse momentum is required to be greater than 60 GeV/ $c$ . This cut is designed to remove Z+jet events and events with non-prompt leptons which are expected to have low amounts of real  $E_T^{\text{miss}}$ .

Events which may have originated from a resonance decay are rejected using dileptonic

invariant mass selections in  $ee$  or  $\mu\mu$  events. To remove events which may have originated from a hadronic resonance, the dileptonic invariant mass is required to be greater than  $15 \text{ GeV}/c^2$ . In addition, to remove events which may have originated from a Z-boson decaying to two leptons, events are rejected if they have a dileptonic invariant mass within a  $\pm 9 \text{ GeV}/c^2$  window around the Z mass of  $91 \text{ GeV}/c^2$ .

$H_T$  is defined as the scalar sum of the lepton and jet transverse momenta. In  $e\mu$  events,  $H_T$  is required to be greater than  $130 \text{ GeV}/c$  which efficiently removes Z+jets background events.

### 5.2.3 Simulation Samples

The data and simulated data samples used in this analysis were fully reprocessed after data taking finished at the end of 2012. This reprocessing fixed issues identified through the data taking period such as detector alignment and calibration corrections. A summary of the simulation samples used for the nominal prediction is shown in Table 5.4. During the course of this analysis, there were two changes to the baseline simulation samples.

The  $t\bar{t}$  sample was changed after extensive study by the TopWG from MC@NLO with HERWIG to POWHEG with PYTHIA [100–103]. It was identified that MC@NLO does not provide a good description of jet multiplicities beyond next-to-leading order. POWHEG with PYTHIA was seen to provide a better description of jet kinematics and higher order jet multiplicities.

Initial  $b$ -tagging efficiency measurements were made using a diboson sample simulated with HERWIG [102]. It became apparent that this generator underestimated the number of high- $p_T$  jets. As HERWIG is a leading-order generator, events contributing to the diboson background required two jets to be produced through the parton shower rather than from the matrix element. The use of a multi-leg leading-order generator, such as ALPGEN [104], enables the calculation of matrix elements with additional hard partons. It was seen that switching to samples generated with ALPGEN increased the diboson contribution by a non-negligible amount. This change is presented in Section 5.2.6 in Table 5.7. Changing this sample impacted efficiencies measured in the high jet- $p_T$  region and an additional systematic uncertainty was included to account for this change.

**Non-Prompt Lepton Estimation** Whilst the main backgrounds are accounted for by simulated samples, some events will be selected in data which are difficult or inefficient to model

Table 5.4: Summary of the Monte Carlo generator and parton showering programs used to simulate the signal and main background samples for dileptonic top-pair decays.

Process	Generator	Hadronisation	PS Tune	PDF
$t\bar{t}$	POWHEG [101]	PYTHIA 6 [105]	Perugia2011C [106]	CT10 [107]
Single top (wt)	MC@NLO [100]	HERWIG [102]	AUET2 [108]	CT10 [107]
Z+jets	ALPGEN [104]	HERWIG [102]	AUET2 [108]	CTEQ6L1 [109]
Diboson	ALPGEN [104]	HERWIG [102]	AUET2 [108]	CTEQ6L1 [109]

with simulated data. These are primarily from  $W$ +jets, single top ( $s$  and  $t$  channel) and single lepton top-pair decays. Events from these sources are referred to as fake events in this analysis as they do not produce two prompt leptons. A data-driven method is used to evaluate the number of non-prompt dilepton events which are selected in data. A fake event will typically originate from a jet being misidentified as an electron or from misidentified lepton charge. Muons are harder to fake due to the requirement of hits in the muon spectrometer, but can be falsely identified in areas of the detector where the reconstruction and identification efficiency is lower. Truth matching is applied on the simulated samples to ensure events are only selected with true prompt leptons in order to remove double counting of fake events when estimating them from data.

To estimate the contribution from fake leptons, this data-driven method inverts the opposite-sign lepton requirement on data and makes the assumption that the source of non-prompt dilepton production is insensitive to the charge of the misidentified leptons. The number and shape of same-sign events selected in data is used to model the contribution of fake events in the nominal analysis.

#### 5.2.4 $ee$ and $\mu\mu$ Events

An excess of simulated events was seen in the dilepton invariant mass distribution in the  $Z$ -boson mass peak compared to the data. In this region, the Z+jets sample is the dominant contribution. This issue was only seen in the  $ee$  and  $\mu\mu$  channel and can be seen in the invariant mass distribution in Figure 5.1. The discrepancy in the  $Z$  peak is seen prior to cutting on the missing transverse energy in the event selection. Following this cut, the discrepancy remained. A clear slope in the  $E_T^{\text{miss}}$  distribution can be seen in the comparison between data and prediction in Figure 5.2. One attempt at resolving this issue was to derive a dilepton  $p_T$  reweighting in the  $Z$ -boson mass window before the  $E_T^{\text{miss}}$  cut (which can be interpreted as reweighting the  $Z$ -boson  $p_T$  spectrum), but it was seen that doing this did not remove the apparent mismodelling of the

Z+jets sample. In an earlier version of this analysis which used  $13.8 \text{ fb}^{-1}$  of data, this discrepancy was resolved by normalising the Z+jets sample with a scale factor in the Z-boson mass peak which was derived after the  $E_T^{\text{miss}}$  cut.

A further breakdown of this problem can be seen in Figure 5.3 where the  $E_T^{\text{miss}}$  distribution is provided for events outside the Z peak window and in Figure 5.4 where the  $E_T^{\text{miss}}$  distribution is provided for events inside the Z peak window. The slope in the  $E_T^{\text{miss}}$  distribution can be observed to be coming from inside the Z window, where one may wish to derive a correction for the Z+jets sample. When this analysis was performed, there was no explanation for the poor agreement between data and simulation in a region dominated by on-shell Z decays.

As this problem was not fully understood or resolved using the full 2012 dataset, it was decided to remove the  $ee$  and  $\mu\mu$  channels from the calibration. The impact on the analysis is small. The  $e\mu$  channel drives the performance of the dilepton calibration, as it has a larger amount of data and a higher  $b$ -jet purity, owing to the reduced contribution from background processes. In particular, the Z+jets process has a minimal contribution to the  $e\mu$  channel, which makes this channel less sensitive to any potential problems in the modelling of this background.

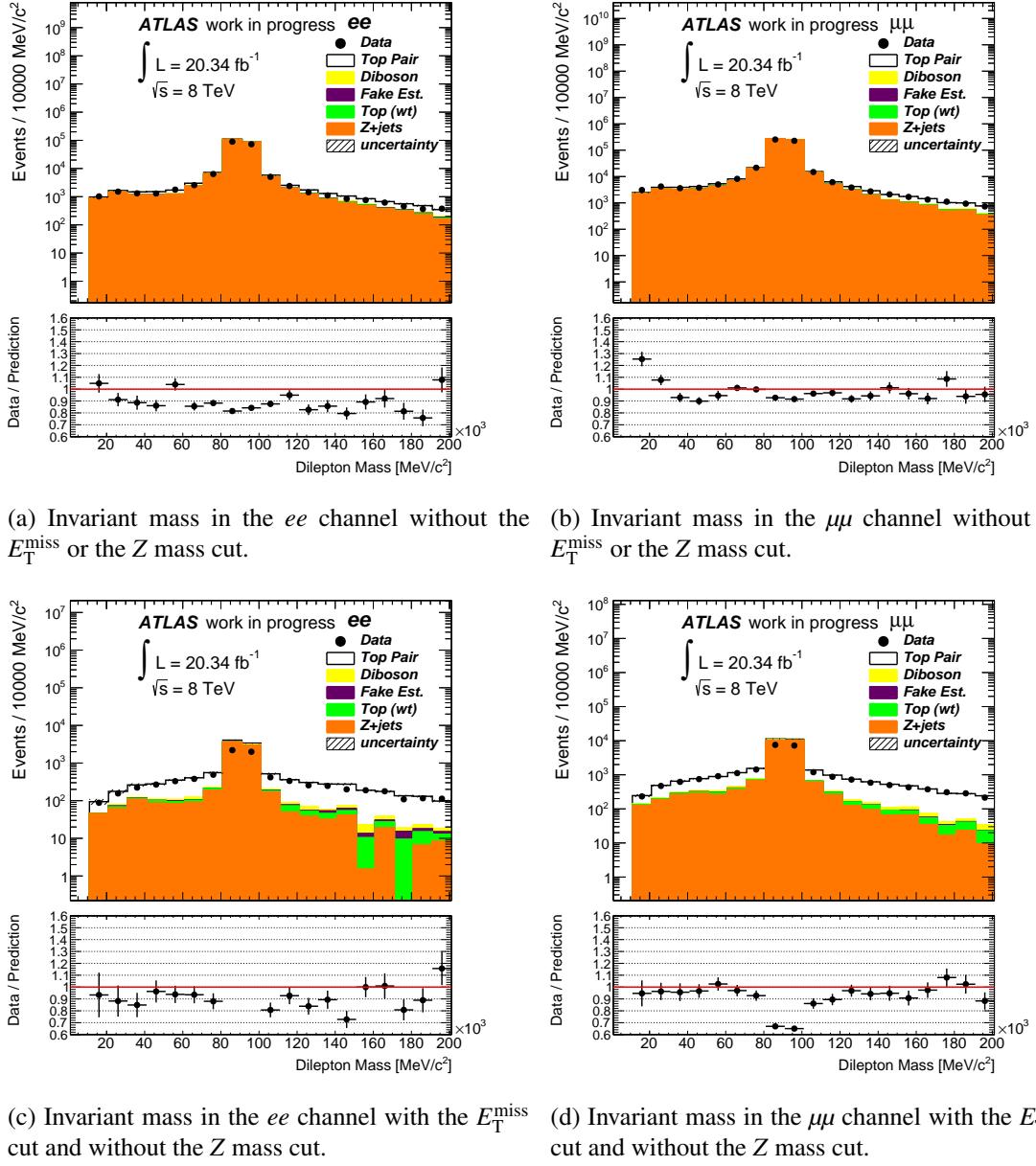
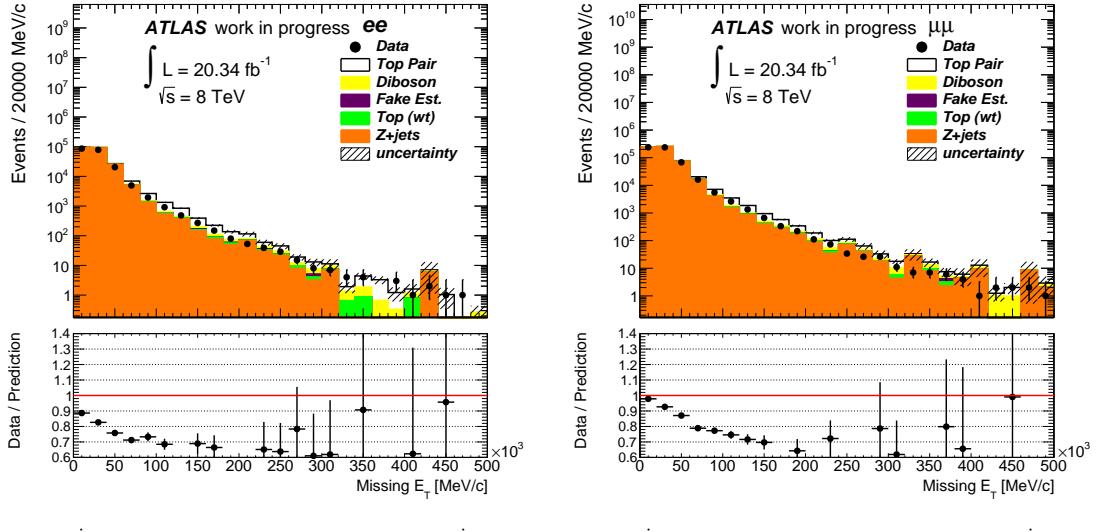
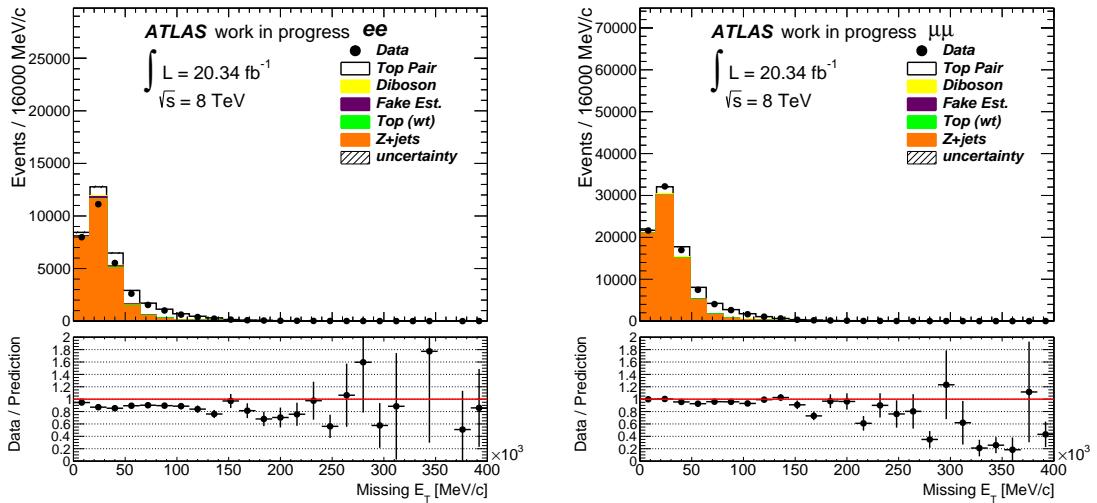


Figure 5.1: Control plots of the invariant mass spectrum in the  $ee$  and  $\mu\mu$  channel before and after the  $E_T^{\text{miss}}$  cut, to show the impact in the  $Z$ +jets dominated region. The selected events use LC jets with a JVF cut applied.



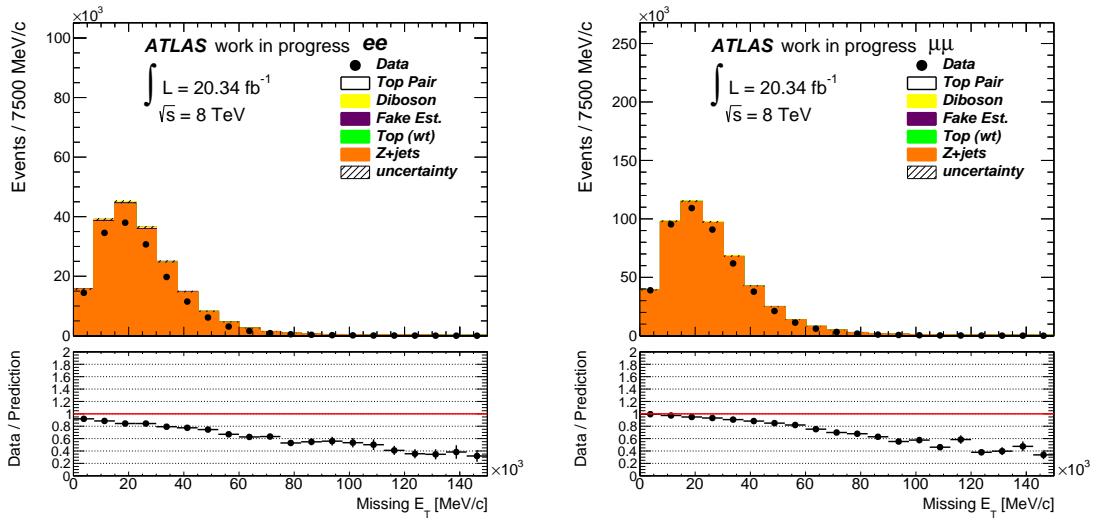
(a)  $E_T^{\text{miss}}$  in the  $ee$  channel without the  $E_T^{\text{miss}}$  or the  $Z$  mass cut.  
(b)  $E_T^{\text{miss}}$  in the  $\mu\mu$  channel without the  $E_T^{\text{miss}}$  or the  $Z$  mass cut.

Figure 5.2: Control plots highlighting mismodelling within the  $ee$  and  $\mu\mu$  channels, where a clear slope can be observed in the  $E_T^{\text{miss}}$  distribution. The selected events use LC jets with a JVF cut applied.



(a)  $E_T^{\text{miss}}$  in the  $ee$  channel without the  $E_T^{\text{miss}}$  cut, outside the  $Z$  peak window.  
(b)  $E_T^{\text{miss}}$  in the  $\mu\mu$  channel without the  $E_T^{\text{miss}}$  cut, outside the  $Z$  peak window.

Figure 5.3: Control plots investigating the mismodelling within the  $ee$  and  $\mu\mu$  channels outside the  $Z$  peak window show a flat comparison between data and simulation. The selected events use LC jets with a JVF cut applied.



(a)  $E_T^{\text{miss}}$  in the  $ee$  channel without the  $E_T^{\text{miss}}$  cut, (b)  $E_T^{\text{miss}}$  in the  $\mu\mu$  channel without the  $E_T^{\text{miss}}$  cut, inside the  $Z$  peak window.

Figure 5.4: Control plots highlighting mismodelling within the  $ee$  and  $\mu\mu$  channels, where a slope can be observed in the  $E_T^{\text{miss}}$  distribution for events inside the  $Z$  peak window. The selected events use LC jets with a JVF cut applied.

### 5.2.5 $e\mu$ Events

In the final calibration analysis, the  $e\mu$  channel is separated into events with exactly two or three jets. If a true top-pair event contains more than two jets, the implication is that there is initial or final state QCD radiation, or hadronic contamination from pile-up vertices. When using the two leading jets in an event for the calibration, the probability of using a non  $b$ -jet increases with jet multiplicity which in turn decreases the  $b$ -jet purity. By splitting the jet multiplicities into exclusive selections, an improved  $b$ -jet purity can be obtained which is beneficial to the calibration analysis. The three jet selection is used in order to increase the amount of data in the calibration analysis without diluting the overall  $b$ -jet purity too much. The two selections are kept independent to allow for a later statistical combination.

The distributions presented in Figure 5.5-5.8 show the level of agreement in the  $e\mu$  channel when looking inclusively within the jet multiplicities. The distributions presented in Figure 5.9 show these events separated out into those with exactly two jets and exactly three jets. In all these distributions there is good agreement between data and simulation except for the jet  $p_T$  distributions shown in Figure 5.10 where there are signs of mismodelling which is further discussed in Section 5.2.8.

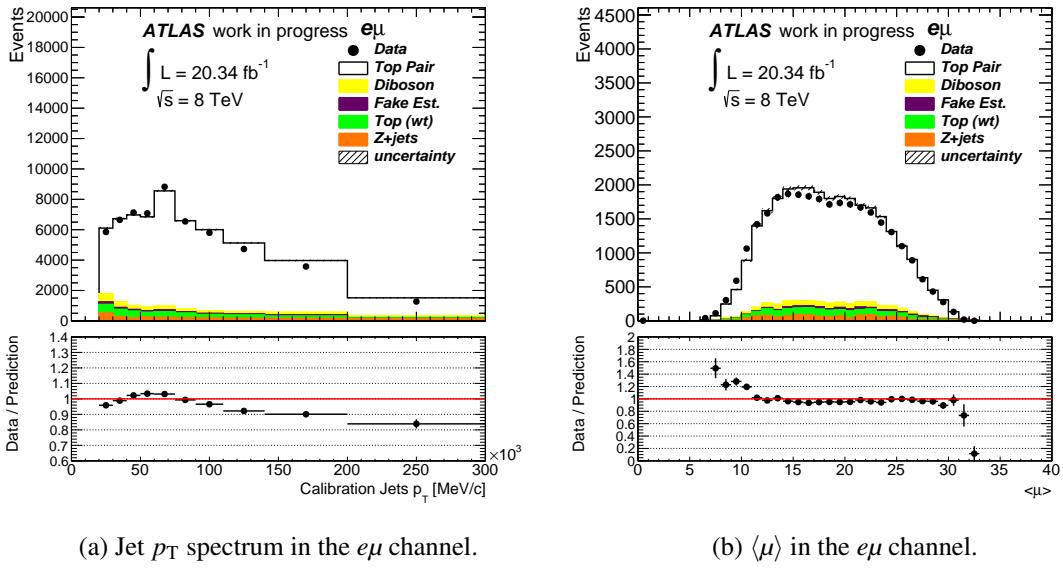


Figure 5.5: Control plots showing the agreement between data and prediction in the  $e\mu$  channel integrated across all jet multiplicities with the full analysis selection applied. The selected events use LC jets with a JVF cut applied.

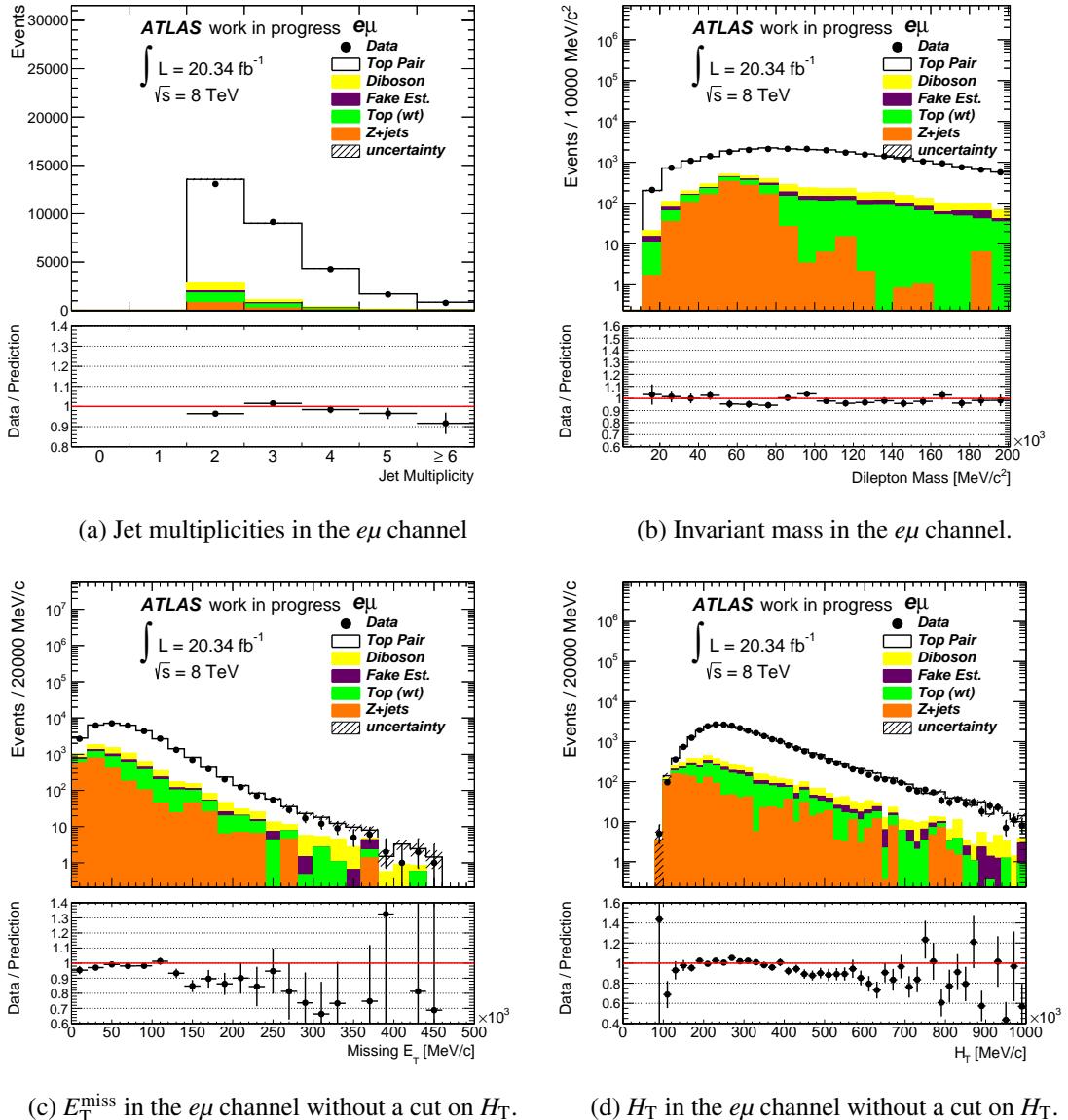


Figure 5.6: Control plots showing the agreement between data and prediction in the  $e\mu$  channel integrated across all jet multiplicities. The distributions have all the analysis cuts applied except for those which are directly correlated with the  $H_T$  cut which are shown with this cut removed. The selected events use LC jets with a JVF cut applied.

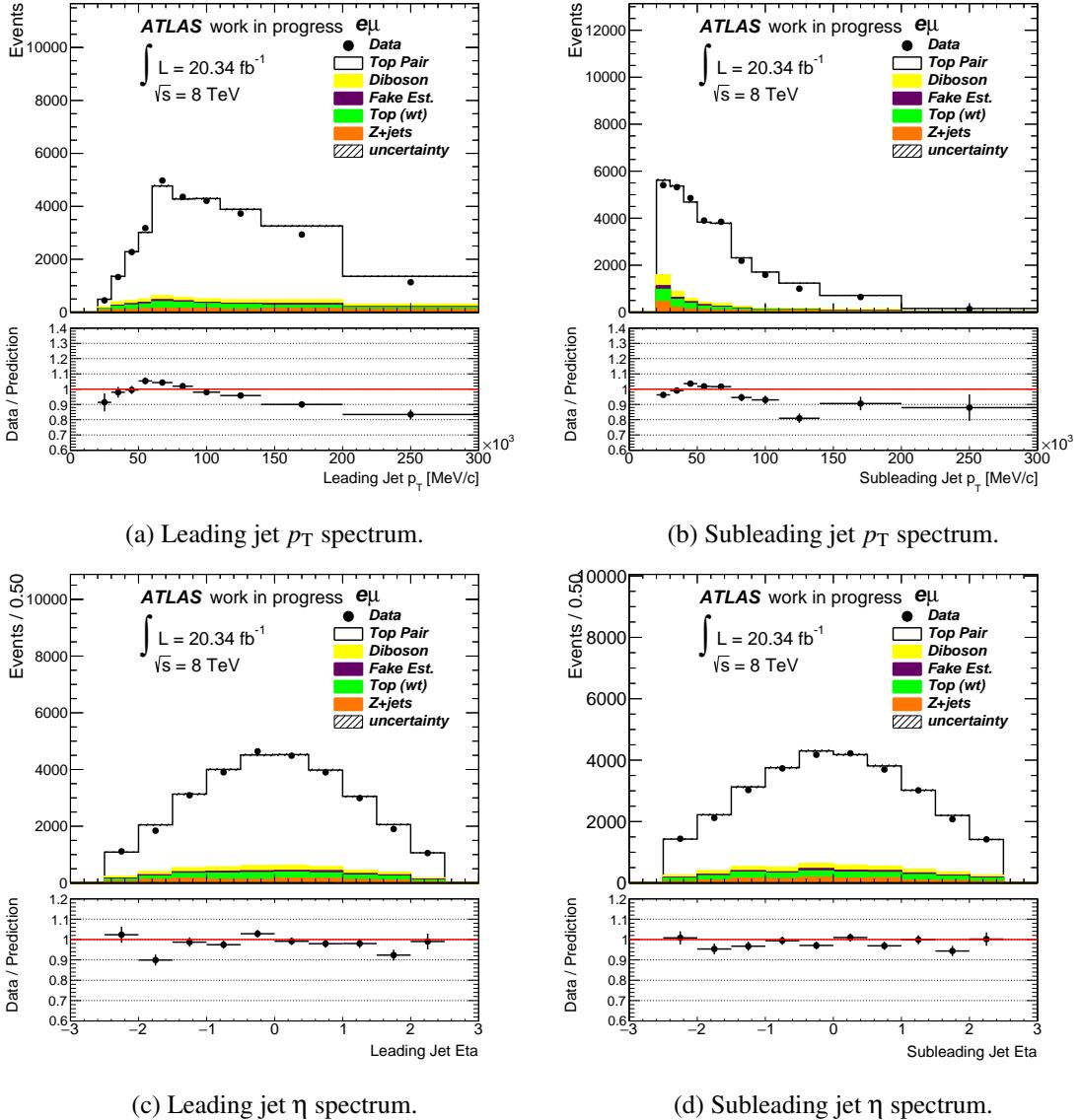


Figure 5.7: Jet kinematic control plots showing the agreement between data and prediction in the  $e\mu$  channel integrated across all jet multiplicities with the full analysis selection applied. Some mismodelling in the jet  $p_T$  spectra can be observed. The selected events use LC jets with a JVF cut applied.

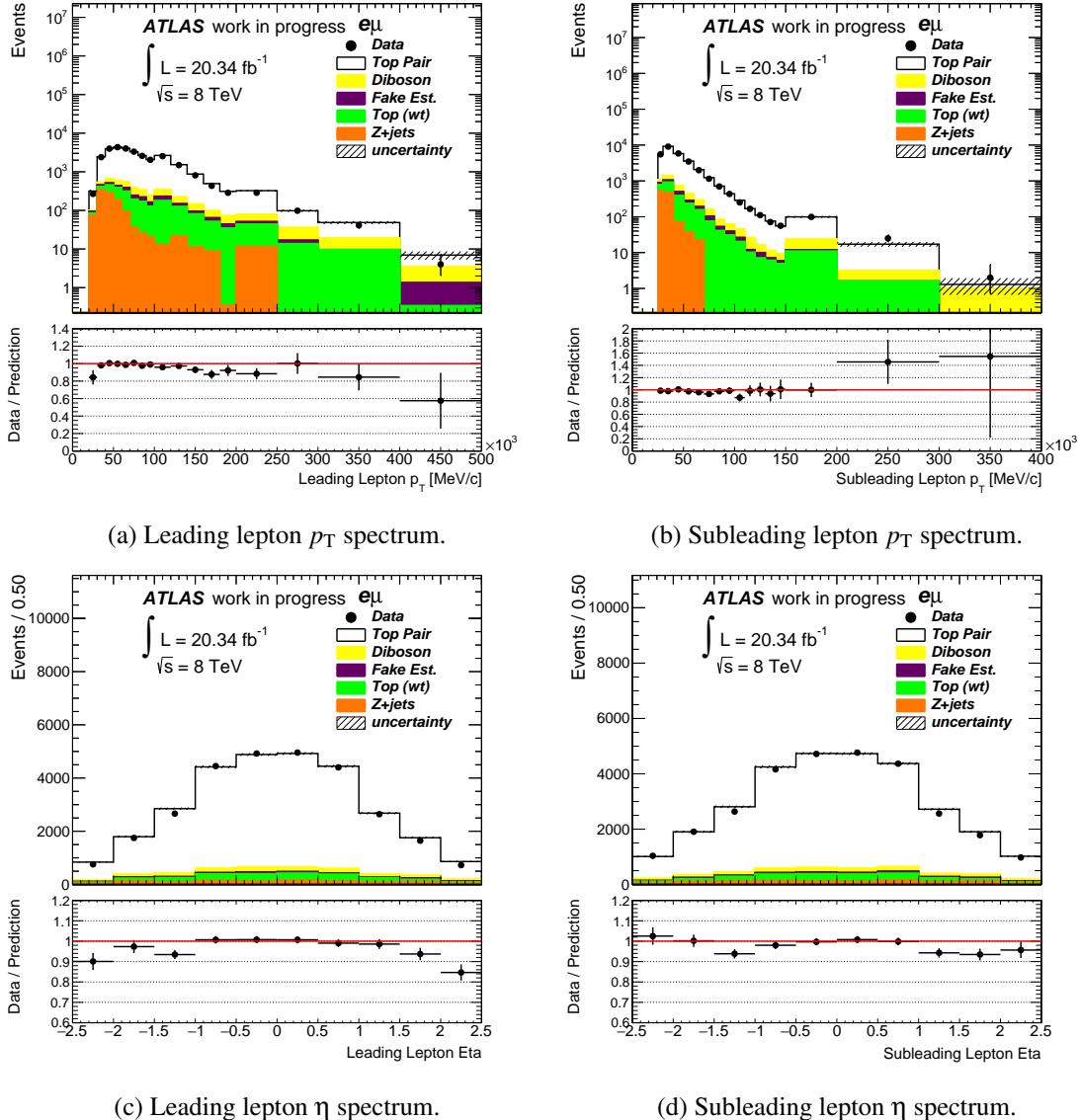


Figure 5.8: Lepton kinematic control plots showing the agreement between data and prediction in the  $e\mu$  channel integrated across all jet multiplicities with the full analysis selection applied. The selected events use LC jets with a JVF cut applied.

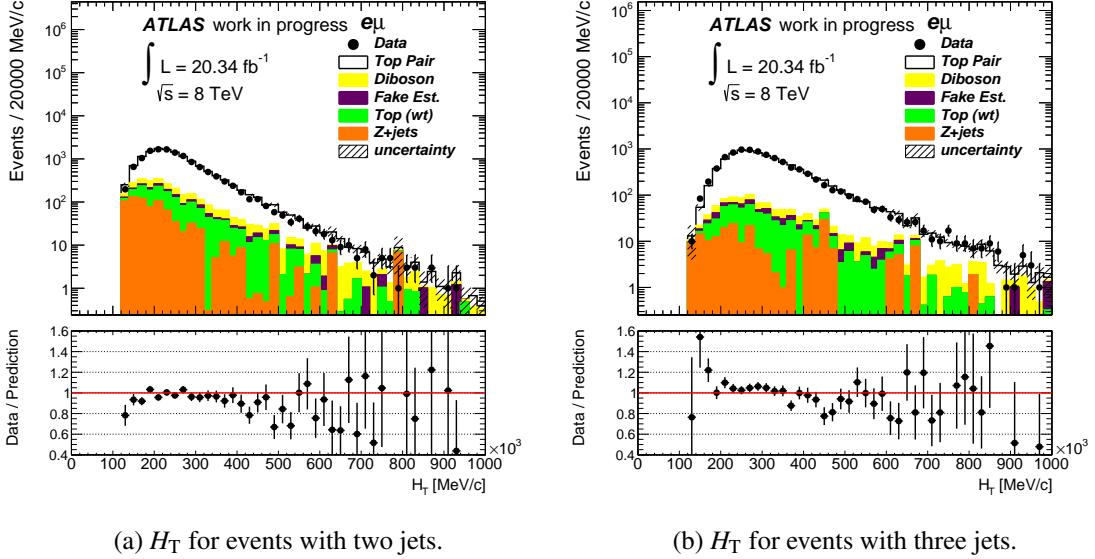


Figure 5.9: Event level control plots showing the agreement between data and prediction in the  $e\mu$  channel after the full selection has been applied and events have been separated into those with exactly two jets and exactly three jets. The selected events use LC jets with a JVF cut applied.

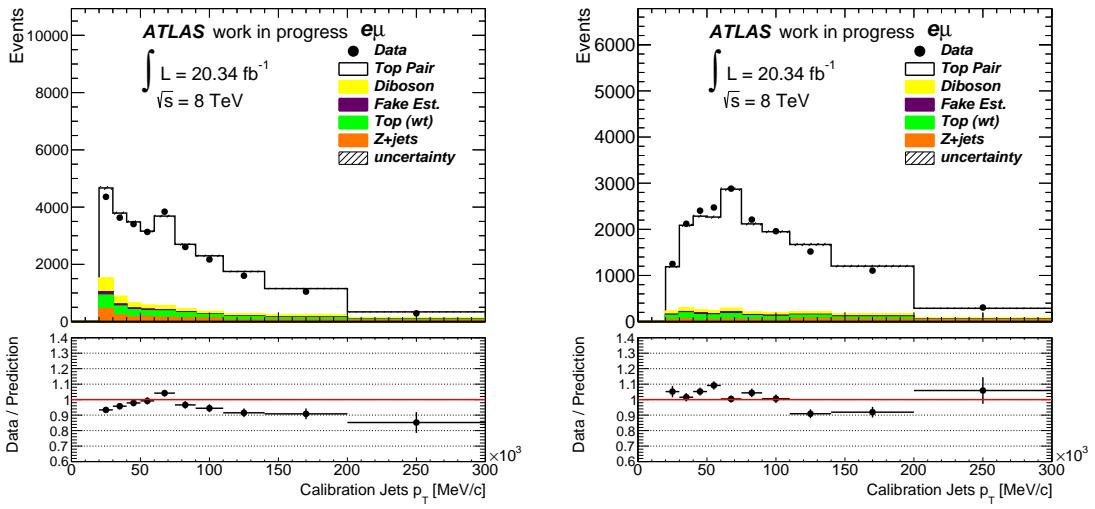


Figure 5.10: A comparison of the jet  $p_T$  spectra for LC jets with a JVF cut applied which enter into the  $b$ -tagging calibration in  $e\mu$  channel for two and three jet events. The shape in the ratio plots is thought to be originating from a mismodelling of the  $b$ -jet  $p_T$  spectrum. An additional systematic uncertainty is included to account for this.

### 5.2.6 Event Yields

There are four independent event selections used for this calibration analysis. The two different jet energy scales (EM and LC) result in two different collections of jets per event. The jets in these collections are not identical and have to be analysed separately. For each of the two jet collections, two different JVF working points are used (none and 0.5) as discussed in Section 5.2.1. A JVF cut will change the number of jets selected which in turn affects the overlap removal with selected jets. The final event yields after all cuts have been applied are shown in Table 5.5-5.6 for each of these selections. The yields are the weighted number of events which are normalised to the data luminosity of  $20.34 \text{ fb}^{-1}$ .

The change in the number of predicted diboson events from changing event generators was quite substantial. A summary of this change is presented in Table 5.7 and is used to motivate the inclusion of a systematic uncertainty to cover the fact that this change happened towards the end of the analysis and the impact was larger than expected.

Table 5.5: Summary of the expected number of events and the number of events from data in the 2 and 3 jet final state cases of the  $e\mu$  dilepton channel after all cuts for EM jets. The errors on the yields are statistical only.

	No JVF		JVF > 0.5	
	2 Jets	3 Jets	2 Jets	3 Jets
$t\bar{t}$	$9128.6 \pm 52.4$	$8236.4 \pm 49.5$	$10897.4 \pm 56.8$	$7834.0 \pm 48.2$
Z+jets	$590.4 \pm 55.7$	$399.3 \pm 43.9$	$683.9 \pm 60.0$	$254.2 \pm 36.0$
Single top	$984.0 \pm 21.9$	$541.4 \pm 16.4$	$972.4 \pm 21.8$	$399.5 \pm 14.5$
Diboson	$887.7 \pm 18.1$	$450.4 \pm 12.4$	$785.0 \pm 16.9$	$294.8 \pm 10.0$
Fakes	$257.0 \pm 16.0$	$150.0 \pm 12.2$	$240.0 \pm 15.5$	$141.0 \pm 11.9$
Total Prediction	$11848.0 \pm 83.1$	$9777.5 \pm 70.3$	$13579.0 \pm 88.5$	$8923.7 \pm 63.8$
Data	$11831 \pm 108.8$	$9660 \pm 98.3$	$13237 \pm 115.1$	$8942 \pm 94.6$

Table 5.6: Summary of the expected number of events and the number of events from data in the 2 and 3 jet final state cases of the  $e\mu$  dilepton channel after all cuts for LC jets. The errors on the yields are statistical only.

	No JVF		JVF > 0.5	
	2 Jets	3 Jets	2 Jets	3 Jets
$t\bar{t}$	$8575.8 \pm 55.8$	$8275.4 \pm 54.2$	$10736.8 \pm 61.7$	$7928.3 \pm 53.1$
Z+jets	$658.9 \pm 59.5$	$429.1 \pm 46.9$	$773.9 \pm 65.5$	$221.7 \pm 33.0$
Single top	$1003.1 \pm 22.0$	$559.4 \pm 16.5$	$988.9 \pm 21.9$	$416.5 \pm 14.7$
Diboson	$926.5 \pm 18.6$	$510.4 \pm 13.3$	$810.8 \pm 17.3$	$301.7 \pm 10.1$
Fakes	$256.0 \pm 16.0$	$162.0 \pm 12.7$	$244.0 \pm 15.6$	$138.0 \pm 11.7$
Total Prediction	$11420.0 \pm 88.0$	$9936.3 \pm 75.9$	$13554.0 \pm 95.5$	$9005.8 \pm 66.0$
Data	$11272 \pm 106.2$	$9905 \pm 99.5$	$13071 \pm 114.3$	$9145 \pm 95.6$

Table 5.7: Change in the event yields for  $e\mu$  events caused by switching diboson event generators from HERWIG to ALPGEN. The increase in the number of events in the JVF selection is due to the increase in the number of high- $p_T$  jets. These jets will automatically pass the JVF cut, allowing for an increased number of events to migrate from the no JVF selection into the JVF selection.

EM	No JVF		JVF > 0.5	
	2 Jets	3 Jets	2 Jets	3 Jets
Diboson (HERWIG)	$613.9 \pm 12.9$	$225.9 \pm 7.7$	$442.4 \pm 11.2$	$96.3 \pm 5.2$
Diboson (ALPGEN)	$887.7 \pm 18.1$	$450.4 \pm 12.4$	$785.0 \pm 16.9$	$294.8 \pm 9.9$
LC	2 Jets	3 Jets	2 Jets	3 Jets
Diboson (HERWIG)	$671.3 \pm 13.6$	$260.6 \pm 8.2$	$461.6 \pm 11.4$	$101.2 \pm 5.3$
Diboson (ALPGEN)	$926.5 \pm 18.6$	$510.4 \pm 13.3$	$810.8 \pm 17.3$	$301.7 \pm 10.1$

### 5.2.7 Calibration Method

The extraction of  $b$ -tagging efficiencies from data uses the assumption that the number of jets which are tagged in data must come from a number of true  $b$ -jets which are correctly tagged (with some tagging efficiency,  $\epsilon_b$ ) and a number of non  $b$ -jets which are incorrectly tagged (with some mistagging efficiency;  $\epsilon_{\text{light}}$ ,  $\epsilon_{\text{charm}}$ ,  $\epsilon_{\text{tau}}$  for light quark or gluon jets, charm jets and  $\tau$  lepton jets respectively). The method requires knowledge of the flavour fractions in the  $t\bar{t}$  and background samples (given as  $f_b$ ,  $f_{\text{charm}}$ ,  $f_{\text{light}}$  and  $f_{\text{tau}}$ ) in order to extract the  $b$ -tagging efficiency from data. The contribution of non-prompt leptons passing the event selection is small but is included in the formulation of the calibration method. The  $b$ -tagging efficiency for these events ( $\epsilon_{\text{np}}$ ) is the fraction of  $b$ -tagged jets in the same-sign data selection.

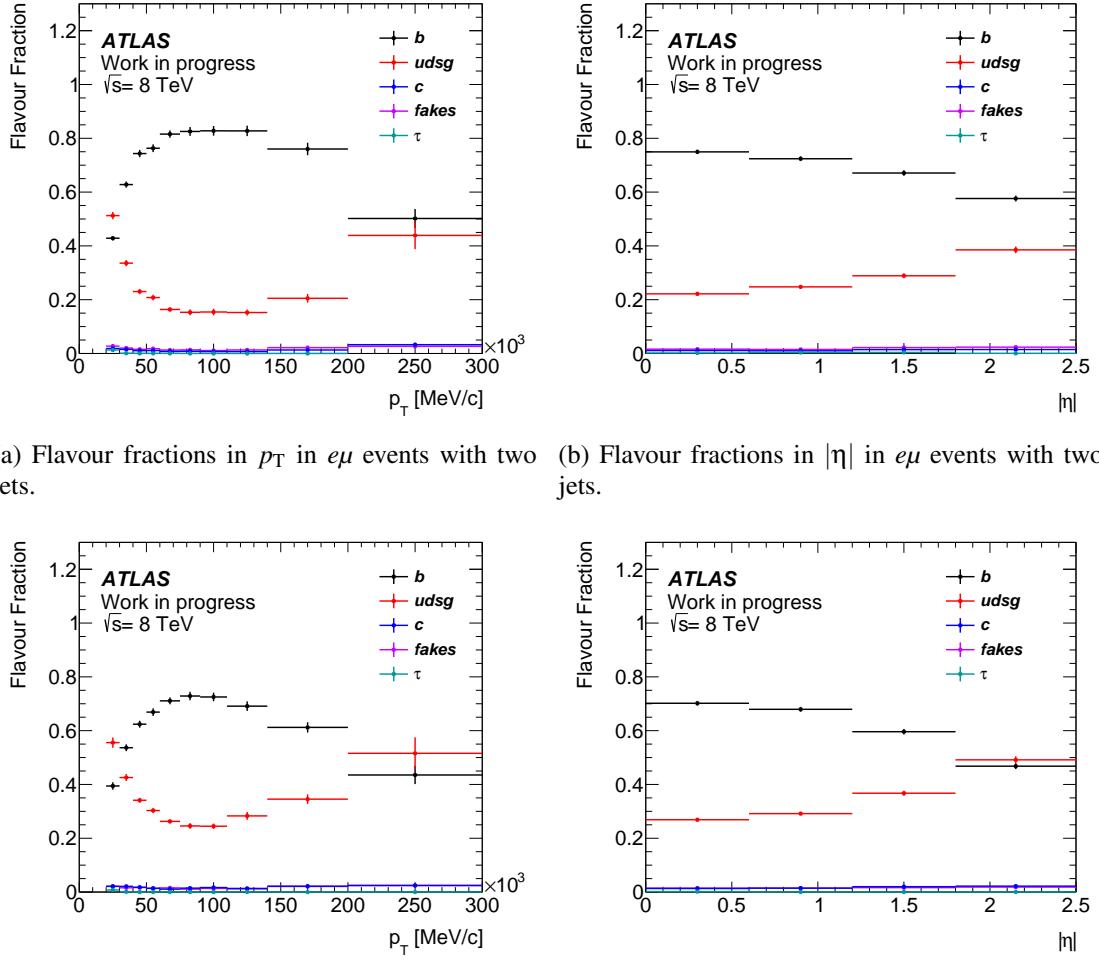
The fraction of jets tagged in data is equated to the predicted flavour fractions while taking into account the tagging efficiencies. Given the expected number of events and the flavour composition of the jets used in the calibration method, the fraction of  $b$ -tagged jets in data is given as:

$$f_{\text{tagged}}^{\text{data}} = f_b \epsilon_b + f_{\text{light}} \epsilon_{\text{light}} + f_{\text{charm}} \epsilon_{\text{charm}} + f_{\text{tau}} \epsilon_{\text{tau}} + f_{\text{np}} \epsilon_{\text{np}}. \quad (5.1)$$

If the simulation is thought to correctly model the flavour composition of the jets in data then this equation can be rewritten to solve for  $\epsilon_b$ ,

$$\epsilon_b = \frac{f_{\text{tagged}}^{\text{data}} - (f_{\text{light}} \epsilon_{\text{light}} + f_{\text{charm}} \epsilon_{\text{charm}} + f_{\text{tau}} \epsilon_{\text{tau}} + f_{\text{np}} \epsilon_{\text{np}})}{f_b^{\text{MC}}}. \quad (5.2)$$

The calibration is performed in histogrammed bins of jet  $p_T$ . The  $p_T$ -binning used in this calibration is 20-30, 30-40, 40-50, 50-60, 60-75, 75-90, 90-110, 110-140, 140-200 and 200-300 GeV/ $c$ . Equation (5.2) is applied to each bin independently to solve the  $b$ -tagging efficiency



(a) Flavour fractions in  $p_T$  in  $e\mu$  events with two jets. (b) Flavour fractions in  $|\eta|$  in  $e\mu$  events with two jets.

(c) Flavour fractions in  $p_T$  in  $e\mu$  events with three jets. (d) Flavour fractions in  $|\eta|$  in  $e\mu$  events with three jets.

Figure 5.11: The  $p_T$  and  $\eta$  flavour fractions for LC jets with a JVF cut applied, separated into the two and three jet events.

in data within that jet- $p_T$  range. The mistag efficiencies are taken from simulation and are corrected to match the mistag efficiencies measured in data. Binomial uncertainties are used for all efficiencies which enter into the calculations.

Examples of the flavour fractions which are used in the measurement of the  $b$ -tagging efficiency are shown in Figure 5.11 as a function of  $p_T$  and  $\eta$ . The calibration is carried out in bins of jet  $p_T$  but a check is made on the flavour fractions in  $\eta$  to check whether integrating over  $\eta$  is acceptable. There does not appear to be any difference in the  $\eta$  dependence of the flavour fractions compared to the 2011 calibration [98] and therefore the calibration is performed inclusively in jet- $\eta$ .

The technical implementation of this calculation uses histograms of jet  $p_T$  which are split according to the true jet-flavour. The truth labelling used on the jet flavour comes from a spatial matching at the quark level, rather than the hadron level. A jet is labelled a true  $b$ -jet if a  $b$ -quark is found within  $\Delta R < 0.3$ . If a jet fails this matching criteria, it is repeated checking for a  $c$ -quark and then a tau lepton. Failing all of these matching criteria results in the jet being labelled as a light-jet. This method of truth labelling is consistent with all flavour tagging calibrations and algorithms within the ATLAS Collaboration as described in Section 4.5.

In the 2011 implementation of this calibration [98], the two jets with the highest  $p_T$  in each event were used in the calculation. This analysis exhibited a lower  $b$ -jet fraction than one would assume for pure  $t\bar{t}$  events due to the high proportion of non  $b$ -jets which have a large  $p_T$  as the jet multiplicity increases. Using the  $e\mu$  channel on its own increases the  $b$ -jet purity compared to the combined dilepton analysis, and separating events into two and three jet samples further improves this.

It was observed that in three jet events there was a reduction in the  $b$ -jet purity when the leading jet had a  $p_T$  greater than 200 GeV/ $c$ . In these events, it was seen that if the third  $p_T$ -ordered jet had a  $p_T$  greater than 50 GeV/ $c$ , then they had a higher  $b$ -jet purity than the leading jet. It was therefore decided that in three jet events, the leading and subleading jets would be used in the calibration unless the leading jet  $p_T$  was greater than 200 GeV/ $c$  and the third  $p_T$ -ordered jet had a  $p_T$  greater than 50 GeV/ $c$ , in which case the subleading and third jet in the event would be used in the calibration. It was seen that the  $b$ -jet purity was relatively constant for the subleading jet over the range of leading jet  $p_T$  versus third jet  $p_T$ .

The aim of this analysis is to derive scale factors to be applied to physics analyses using  $b$ -tagging. These factors correct the  $b$ -tagging efficiency in simulation to that which is measured in data. To derive these scale factors, a baseline simulation  $b$ -tagging efficiency needs to be defined. This is calculated using true  $b$ -jets from the top-pair sample used in the calibration. These events are used as they provide a large sample of inclusive  $b$ -jets. This should prevent a bias being derived in the scale factor relating to the decay process of the  $B$ -hadron. The scale factor is defined as

$$\text{SF} = \frac{\epsilon^{\text{data}}}{\epsilon_{t\bar{t}}^{\text{MC}}} . \quad (5.3)$$

The  $b$ -tagging efficiency calibration is carried out for a number of different  $b$ -tagging algorithms and working points which are listed in Table 5.8. Details on the  $b$ -tagging algorithms can

be found in Section 4.5.

Table 5.8: Table listing the  $b$ -tagging efficiencies evaluated in  $t\bar{t}$  events and the discriminant threshold of each  $b$ -tagging working point used in the calibration. Details on the  $b$ -tagging algorithms can be found in Section 4.5. The JetFitterCOMBCharm calibration examined three working points tuned for charm jet selection. As such, the results of this calibration are mistagging efficiencies of the charm tagger. The two values listed are the log-likelihood ratios from the JetFitterCOMBNN algorithm for the ratio of charm versus bottom and for charm versus light. Working points are derived separately for each jet collection, as the efficiency of a single cut on the  $b$ -tagging discriminant varies between them.

Tagger	Efficiency	LC Jets	EM Jets
MV1	30 %	0.994450	0.992670537
	50 %	0.993981	0.992515446
	60 %	0.9827	0.9867
	70 %	0.7892	0.8119
	75 %	0.6073	0.6065
	80 %	0.3511	0.3900
	90 %	0.0617	0.0616
MV1c	30 %	0.9853	0.9848
	50 %	0.9195	0.9237
	57 %	0.8641	0.8674
	70 %	0.7068	0.7028
	80 %	0.4051	0.4050
	90 %	0.0836	0.0822
SV0	50 %	5.650	5.700
JetFitterCOMBCharm	n/a	-1.0, n/a	-1.0, n/a
	n/a	-1.0, -0.82	-1.0, -0.82
	n/a	-1.0, 1.0	-1.0, 1.0

### 5.2.8 Systematic Uncertainties

The impact of various sources of systematic error on the  $b$ -tagging efficiency and subsequent simulation scale factors have been assessed.

The systematic uncertainties considered in this analysis adjust the  $b$ -tagging efficiency measured in data in each  $p_T$  bin by assigning a relative uncertainty. Each source of systematic uncertainty changes the predicted yield in each  $p_T$  bin by either adjusting the statistical weight of each event or by comparing the difference between the nominal and an alternative simulated data sample.

Flat scale factors related to theoretical cross-section uncertainties adjust the overall normalisation of an individual sample. Reweighting scale factors adjust the event weight based on object information such as identification efficiency uncertainties. Systematic uncertainties related to the object reconstruction change the predicted yields by varying the object energy scale and resolution which causes bin migrations. Some systematic uncertainties, such as  $t\bar{t}$  modelling uncertainties, require the substitution of different samples to evaluate the impact of different modelling parameters.

Each systematic uncertainty is treated as uncorrelated in this analysis. The effect of a systematic uncertainty is evaluated by measuring the shift in the calculated  $b$ -tagging efficiency,

$$\Delta\epsilon = \epsilon_{\text{syst}}^{\text{data}} - \epsilon_{\text{nominal}}^{\text{data}}, \quad (5.4)$$

and by calculating the associated shifted scale factor,

$$\Delta\text{SF} = \frac{\epsilon_{\text{syst}}^{\text{data}}}{\epsilon_{\text{syst}}^{t\bar{t}}} - \frac{\epsilon_{\text{nominal}}^{\text{data}}}{\epsilon_{\text{nominal}}^{t\bar{t}}}. \quad (5.5)$$

**Lepton systematics** Both electrons and muons have a number of efficiencies and scale factors which are calculated by the Combined Performance Groups to improve predictions compared to data. The efficiency of a lepton selection is the product of four different factors: reconstruction, identification, isolation and trigger.

Electron reconstruction and isolation scale factors are derived from  $Z \rightarrow ee$  and  $W \rightarrow e\nu$  selections. The energy scale and resolution is compared between data and simulation using  $Z \rightarrow ee$  and  $J/\psi \rightarrow ee$  events [110, 111].

Muon reconstruction efficiencies are measured on data and simulation using a tag-and-probe

method with  $Z \rightarrow \mu\mu$  events [112], but no measurements of isolation scale factors were available. Muon  $p_T$  scale and resolution efficiencies are measured from data using the  $Z$  mass peak after momentum smearing is applied to simulated muons. There are scale corrections for both the inner detector and the muon spectrometer.

**Jet energy scale (JES)** A number of corrections are applied to jets to correct the measured energy of simulated jets to data [113]. A correction is applied to account for the effect of multiple interactions per bunch-crossing. Corrections relating to in-time pile-up are applied. In-time pile-up refers to multiple interactions per bunch-crossing which causes particles not originating from the primary vertex to contribute to the measured energy in the calorimeter clusters. In addition, there are corrections for out-of-time pile-up from previous or subsequent bunch crossings. These events can affect the measured jet energies since the detector response time is comparable to the time between collisions. Corrections are derived from simulated data and provided in bins of transverse momenta and pseudorapidity.

The detector response is defined as

$$\mathcal{R} = \frac{E_{\text{jet}}^{\text{EM/LC}}}{E_{\text{jet}}^{\text{truth}}}. \quad (5.6)$$

The jet energy scale is corrected using the inverse of this expression.

An *in situ* correction is applied to jets in data, using corrections which are measured using  $\gamma + \text{jets}$  and  $Z + \text{jets}$  events where the  $p_T$  of the system is assumed to be balanced. A correction is derived comparing the  $p_T$  balance in data and in simulation using

$$\delta^{insitu} = \frac{\left\langle \frac{p_T^{\text{jet}}}{p_T^{\text{ref}}} \right\rangle_{\text{data}}}{\left\langle \frac{p_T^{\text{jet}}}{p_T^{\text{ref}}} \right\rangle_{\text{MC}}}, \quad (5.7)$$

where “ref” refers to the  $Z$  boson or photon used to balance the system.

These calibrations and corrections to the jet energy and transverse momenta provide a number of systematic uncertainties which are evaluated, based on the assumptions and methods used to derive them.

- Baseline *in situ* - There are 47 systematics relating to the *in situ* measurements. For this analysis, a reduced set of systematics is used, which have been identified using an

eigenvector decomposition of the covariance matrix to create six effective systematic uncertainties. These six uncertainties are composed of the 5 effective uncertainties which have the largest eigenvalues, giving the greatest contributions to the total uncertainty, and an additional uncertainty accounting for the remainder, to conserve the total uncertainty. This reduction is found to accurately describe the total covariance matrix to the percent level.

- Eta-Intercalibration - When applying the *in situ* corrections, an intercalibration is applied to remove pseudorapidity dependence on the corrections. Two systematic uncertainties are included to account for this, relating to the modelling and the statistical methods used.
- Flavour Composition - The jet flavour composition used to derive the *in situ* calibrations may not accurately reflect the composition of jets in a physics analysis. To account for this, a variation is introduced to adjust the gluon fraction of the samples used in the *in situ* calibrations.
- Flavour Response - In addition to the flavour composition of samples used in the *in situ* calibrations, an uncertainty is also introduced on the calorimeter response, which differs for gluon and quark jets. A flat gluon fraction of  $50\% \pm 50\%$  is used for the flavour related uncertainties.
- $B$ -Jet JES Uncertainty - The calorimeter response can be different for  $b$ -jets compared to light quark and gluon jets which the calibrations are derived from. This uncertainty is applied only to true  $b$ -jets and takes the place of the flavour response and composition uncertainties for these jets.
- Pile-up Corrections - The jet energy corrections are affected by the amount of hadronic activity in the event, which varies depending on the number of interactions per bunch-crossing. An offset is applied to recalculate the energy-momentum vector of a jet to that expected in a zero pile-up environment. There are offsets related to the active area of the jet in the calorimeter, the number of primary vertices and the transverse momenta of the jet. Systematic variations are provided to study how shifts in these variables affect the jet energy scale.
- Non-Closure - A systematic uncertainty is provided relating to the simulation and recon-

struction used to derive the jet energy scale corrections. This impact of this uncertainty depends on the samples used in an analysis. In this analysis a non-negligible effect was observed so the uncertainty was included.

- **High- $p_T$  Single Particle** - Particularly high  $p_T$  jets can be affected by the measurement of a high- $p_T$  hadron in the jet. To account for potential mismodelling, a systematic is provided to propagate the uncertainties on a high- $p_T$  hadron to the jet.

**Jet energy resolution** The jet energy resolution has been measured in data to be well reproduced in simulation using a bisector method [114]. The current recommendations is to refrain from applying any smearing relating to the jet energy resolution in the nominal case, and evaluate a systematic uncertainty through smearing the jet energy to degrade the resolution and check the impact on an analysis with this method.

**Jet reconstruction efficiency** An estimation of the calorimeter jet reconstruction efficiency was made using a tag-and-probe method on jets reconstructed from inner detector tracks. An efficiency was measured by evaluating the number of probe jets which matched a calorimeter jet. The inefficiency of reconstruction is applied as a systematic uncertainty by randomly removing a jet with  $p_T$  less than 30 GeV/ $c$  [81].

**Jet vertex fraction (JVF)** The use of ghost matching tracks in the reprocessed 2012 dataset introduced a change in the reconstruction of jet objects. This procedure propagates the location of tracks into the jet reconstruction algorithm without affecting the properties of the final reconstructed jets [115]. JVF is known to be an effective way of rejecting jets from pile-up vertices, but it is not a well-modelled variable in simulation. The recommendation for analyses being presented at Rencontres de Moriond in 2013 was to apply a systematic uncertainty relating to JVF. This uncertainty is evaluated by varying the JVF cut value depending on whether the jet is a hard-interaction jet or a pile-up jet. The effect of this systematic is to cause event migrations between jet multiplicities. This analysis is particularly sensitive to these uncertainties as the calibrations are derived in exclusive jet regions.

**$B$ -jet reweighting** During the course of this analysis, a shape was identified in the jet  $p_T$  distributions which did not match the data. This shape can be seen in Figure 5.10. In order

to account for the uncertainty associated with this issue, an additional systematic uncertainty was derived to reweight  $b$ -jets in simulation to fit the jet  $p_T$  distribution in data. This shape is thought to originate from a mismodelling of the top quark which results in a harder  $p_T$  spectrum of simulated  $b$ -jets. This systematic uncertainty causes a change in the relative fractions which enter into Equation (5.2) and has the largest impact in the region where the  $b$ -jet fraction is not the dominant contribution. This typically corresponds to the lowest and highest  $p_T$  bins. A reweighting histogram is derived from Figure 5.12, where the  $b$ -jet component is reweighted to match the prediction to data. A separate reweighting histogram is derived for each of the eight calibration selections used and the shift in the  $b$ -tagging efficiency is calculated after reweighting to assess the impact of this uncertainty.

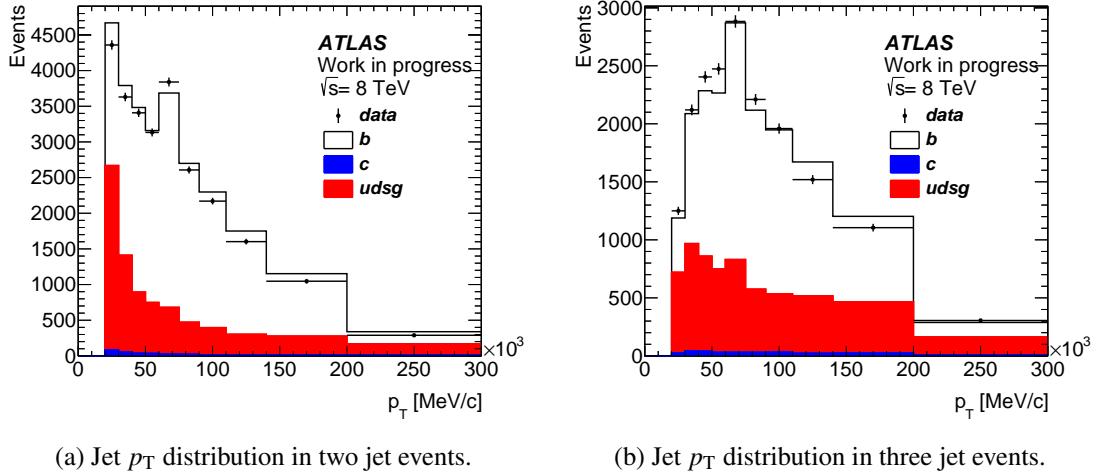


Figure 5.12: Flavour composition compared to data of LC jets with a JVF cut greater than 0.5 used in the calibration for two and three jet events.

**Luminosity** The GRL used in this analysis comes with an estimated uncertainty of 2.8%. This systematic is propagated through the analysis chain by changing the luminosity consistently for all predictions and is seen to have a negligible impact on the analysis.

**Mistag rate** In the derivation of the  $b$ -tagging scale factors, a mistag rate is taken from simulation for the non  $b$ -jets and are corrected to measured mistag rates in data [95]. These mistag rate corrections have an associated systematic uncertainty which is the combination of the various systematic uncertainties considered in the derivation of the correction factors. This uncertainty adjusts the event weight relating to the efficiency of incorrectly  $b$ -tagging a non  $b$ -jet which

propagates into the measurement of  $b$ -tagging efficiencies.

**Theoretical cross-sections** Each of the simulated backgrounds is subject to an uncertainty on the theoretical production cross-section, with the exception of the non-prompt lepton estimation, which is taken directly from data and reflects the large uncertainty in the same-sign selection method. The overall normalisation of the sample is scaled up and down within these theoretical uncertainties to model different sample compositions.

- Top-pair -  $\begin{array}{l} +9.34\% \\ -10.17\% \end{array}$ .
- Z+jets - 4% for inclusive Z normalisation uncertainty plus an additional 24% uncertainty added in quadrature per additional jet that is not generated by the matrix element and is required for the signal selection.
- Single top (Wt) -  $\pm 6.85\%$ .
- Diboson - WW :  $\pm 5\%$  inclusive normalisation uncertainty plus an additional 24% uncertainty added in quadrature per additional jet that is not generated by the matrix element and is required for the signal selection.
- Diboson - WZ and ZZ :  $\pm 5\%$  inclusive normalisation uncertainty. No additional uncertainty is added as the dominant contribution of high- $p_T$  jets is expected to come from the hard partons in the lowest order hard process matrix element.
- Non-prompt lepton estimation -  $\pm 50\%$ .

**Diboson shape uncertainty** The event generator was switched from HERWIG to ALPGEN after it was seen that there was a large difference in the expected number of events passing the selection coupled with the expectation that ALPGEN will provide a more trustworthy estimation of high- $p_T$  jets. An additional systematic was included to cover this change by calculating the efficiencies and scale factors using HERWIG for the diboson sample and assigning half the difference to the nominal as a symmetric systematic uncertainty.

**$t\bar{t}$  - Initial and final state radiation (IFSR)** To evaluate the effect of a different amount of initial and final state radiation coming from the signal sample, two samples produced by ACE-ERMC are used which adjust the value of the strong coupling strength. One sample is set up to

increase the amount of IFSR and the other sample is set up to reduce the amount of IFSR. The values used for the strong coupling strength to generate these events are not excluded by data. A systematic is then evaluated by looking at the difference in calculated efficiencies and calculated scale factors between the sample with increase IFSR and the sample with reduced IFSR. This difference is divided by two and taken as a symmetric systematic uncertainty.

**$t\bar{t}$  - Event generator comparison** The recommendation to compare top-pair event generators comes from the TopWG. The appraisal of this uncertainty allows for variations in the flavour fractions in this analysis to be evaluated. A comparison between the nominal sample and two samples with different event generators is made. The nominal sample is POWHEG+PYTHIA, but for this study a POWHEG+HERWIG sample is used, to factorise out the impact of the hadronisation model. The two samples compared with POWHEG+HERWIG are ALPGEN+HERWIG and MC@NLO+HERWIG. At the time of evaluation, the only POWHEG+HERWIG sample available was a fast simulation sample using ATLFEST II [72]. The MC@NLO sample was also produced with ATLFEST II but the ALPGEN sample was produced with the full detector simulation. In principle, there is should be no difference between the  $b$ -tagging efficiencies using fast or full simulation samples as the inner detector simulation is identical. The  $b$ -tagging efficiency is evaluated in each  $p_T$  bin and the largest difference between POWHEG and the other two generators is taken as the systematic uncertainty related to the modelling of the flavour fractions.

**$t\bar{t}$  - Parton shower** To evaluate the effect of different parton showers, two ATLFEST II samples were compared - POWHEG+PYTHIA and POWHEG+HERWIG. Any differences seen are related to the parton shower, and is taken as a symmetric uncertainty.

**Merged binning** Cross-checks were carried out to investigate the potential impact of statistical fluctuations in systematic uncertainties. In order to evaluate this, diagnostic plots were produced to compare the shape of systematic uncertainties. Whilst a number of systematics appeared to show some variations, it was decided to merge histogram bins for systematic uncertainties which showed clear deviations from an inferred shape due to fluctuating numbers of events. Only systematic uncertainties which had a non-negligible impact on the total systematic uncertainty were rebinned. In addition, rebinning was only carried out for systematic uncertainties which appeared to show the same behaviour in all eight analysis selections for the 70% operating point

of MV1. The resulting systematics which were rebinned were the six effective JES components, the flavour composition and response JES components, the eta-intercalibration JES components and the jet energy resolution. These systematics were rebinned by recalculating the *b*-tagging efficiencies and scale factors in a reduced number of jet  $p_{\text{T}}$  bins. These were 20-30, 30-60, 60-90, 90-140, 140-200 and 200-300 GeV/*c*. The systematic uncertainty was then applied to each of the bins in the full  $p_{\text{T}}$  binning for each reduced  $p_{\text{T}}$  bin.

### 5.2.9 Results

The  $b$ -tagging efficiencies for four flavour tagging algorithms (MV1, MV1c, SV0 and JetFitter-COMBCharm) were measured using  $20.34 \text{ fb}^{-1}$  of  $\sqrt{s} = 8 \text{ TeV}$  data. The results are presented as  $b$ -tagging scale factors calculated with respect to the efficiency measured on true  $b$ -jets in simulated top-pair events. A range of working points were calibrated, from 30% up to 90% efficiency for multiple jet collections and jet selections. This work was originally documented in an ATLAS Internal Note [12]. The calibration used  $e\mu$  dileptonic top-pair events which were separated out into two exclusive measurements for two and three jet events.

The most commonly used  $b$ -tagging algorithm in Run 1 physics analyses is MV1 with a 70% working point for jets with a JVF cut applied. The scale factors for these working points have been calculated to be consistent with unity within the total error. These results are shown in Figure 5.13. The total uncertainty varies from  $\sim 5\%$  in the central jet  $p_T$  calibration bins, and extends up to  $\sim 20\text{-}25\%$  in the lowest and highest  $p_T$  bins. These extrema have the most sensitivity to variations in the relative flavour fractions from systematic uncertainties.

The dominating systematic uncertainties were the jet energy scale uncertainties which are sensitive to the flavour of the jets, jet energy resolution and systematic uncertainties associated with the modelling of the  $t\bar{t}$  and diboson samples. The size of the dominant systematic uncertainties are summarised in Table 5.9-5.10.

To improve the statistical precision of the  $b$ -tagging calibration, these results were combined with additional calibration analyses which used mutually exclusive events in data. The results were kept in their two and three jet categories to allow a better treatment of the correlation of systematic uncertainties when combined with other  $b$ -tagging calibration analyses. An example of such a combination is shown in Figure 5.14, where the three jet calibration presented in this chapter has been combined with a top-pair dilepton tag-and-probe analysis which used exclusively two-jet events [116] and the System8 muon-based calibration method [117].

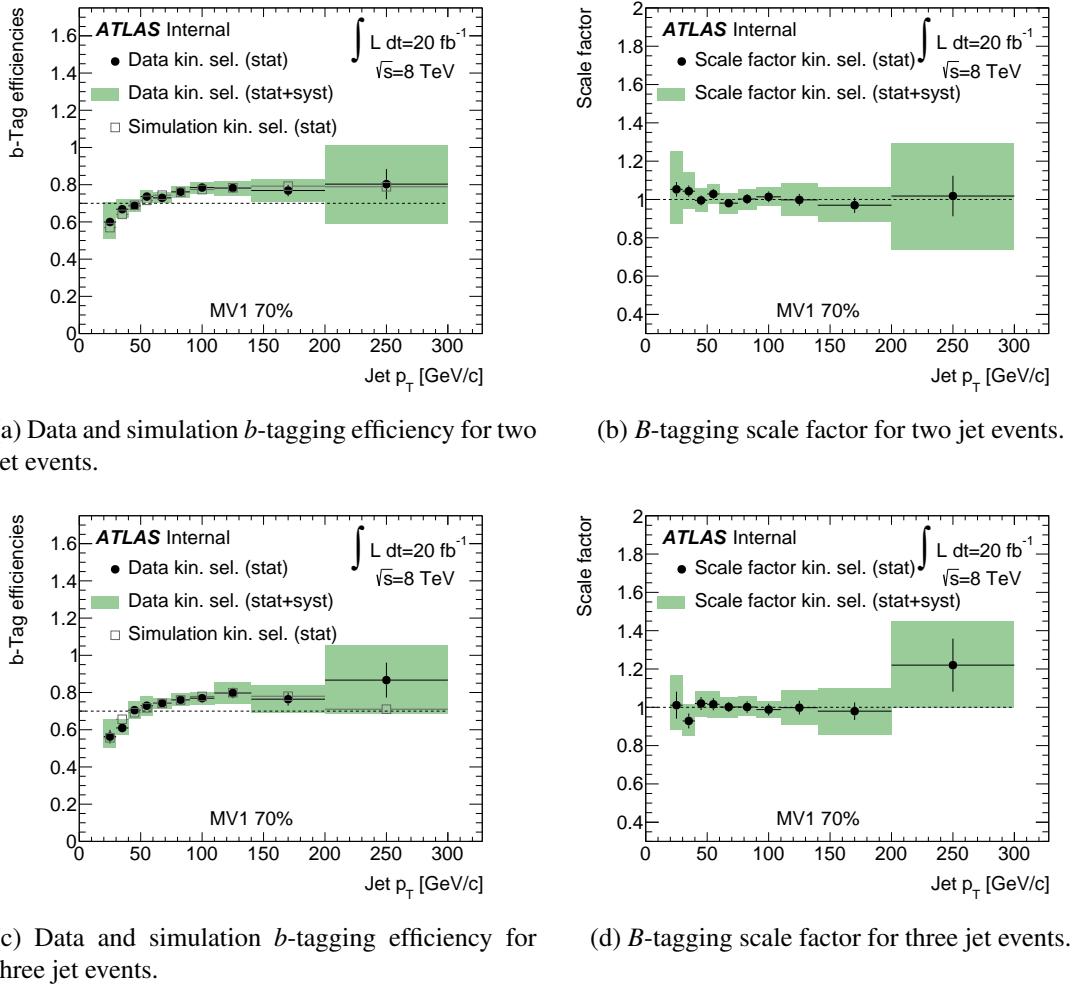


Figure 5.13: Distributions of the  $b$ -tagging efficiency in data and in simulated  $t\bar{t}$  events and the derived scale factor corrections to be applied to simulated data using LC jets with a JVF cut applied.

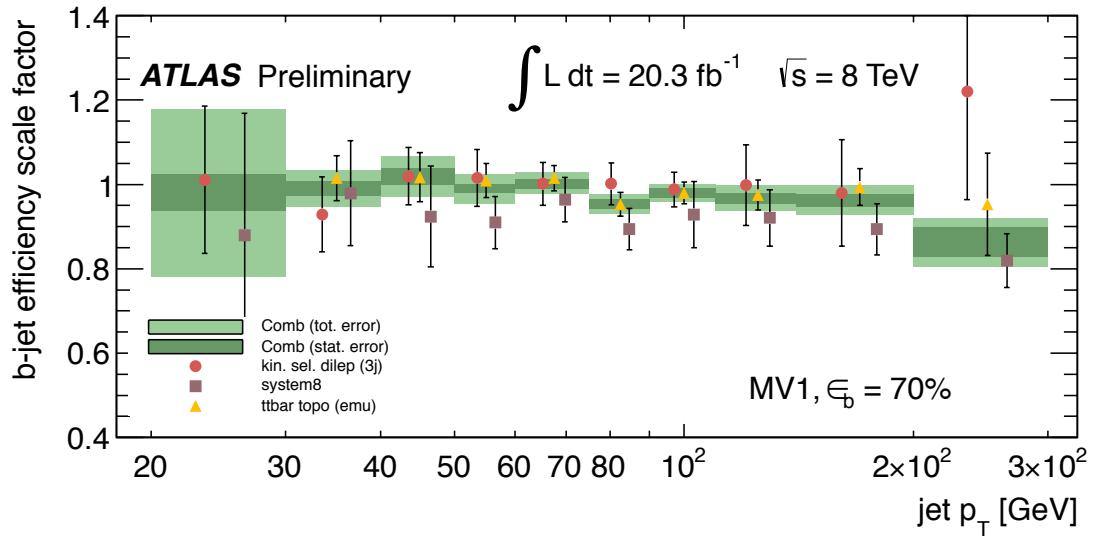


Figure 5.14: A statistical combination of  $b$ -tagging scale factors for use with the full 8 TeV dataset using LC jets with a JVF cut applied [118]. The three-jet calibration presented in this chapter has been included in this combination as “kin. sel. dilep (3j)”. The errors on the points in this plot are the combined statistical and systematic uncertainties.

Table 5.9: The dominant systematic uncertainties in the 2 jet region for the MV1 70% working point using LC jets with a JVFF cut. The uncertainty values are presented as a relative uncertainty (%) on the nominal scale factor. Uncertainties which contribute less than a 2% relative uncertainty in all jet  $p_T$  bins have been removed from this table to only highlight the size of the dominant systematics. The jet energy scale (JES) is presented as a single uncertainty which envelops the combination of all the JES components. The total systematic uncertainty shown takes into account all systematic uncertainties presented in Section 5.2.8.

Jet $p_T$ [GeV/ $c$ ]	20-30	30-40	40-50	50-60	60-75	75-90	90-110	110-140	140-200	200-300
Nominal Scale Factor	1.05	1.04	0.995	1.03	0.981	1.00	1.01	0.998	0.970	1.02
Statistical Uncertainty	3.6	2.9	2.6	2.6	2.2	2.6	2.8	3.1	4.2	10.4
Diboson Herwig	-0.8	-1.3	-1.0	-0.8	-1.0	-1.3	-1.0	-1.4	-2.7	-6.6
JES Down	-11.0	-4.2	-2.4	-2.1	-1.5	-0.8	0.3	0.3	2.2	2.0
JES Up	11.9	3.7	4.7	0.3	1.8	-0.1	-0.0	0.7	-0.3	-4.9
JVFF Down	10.9	0.8	2.0	0.4	-0.1	0.5	-0.9	-0.0	-0.3	0.8
JVFF Up	-1.4	-1.0	1.0	-0.1	-0.4	-0.2	-0.7	0.1	0.4	0.7
$B$ -jet $p_T$ Reweighting	10.1	2.6	0.7	0.3	-0.9	0.7	1.2	1.9	3.2	20.2
Diboson XS Down	-3.4	-2.5	-1.7	-1.4	-1.4	-1.4	-1.2	-1.5	-2.6	-5.7
Diboson XS Up	3.4	2.5	1.7	1.4	1.4	1.4	1.2	1.5	2.5	5.7
$t\bar{t}$ XS Down	2.6	1.7	1.2	1.3	1.0	1.1	1.0	1.0	1.5	3.1
$t\bar{t}$ XS Up	-2.0	-1.3	-0.9	-1.0	-0.8	-0.8	-0.8	-0.8	-1.1	-2.4
$t\bar{t}$ IFSR	3.5	4.1	3.2	2.4	2.4	2.3	1.8	3.0	3.6	7.1
$t\bar{t}$ Generator	-6.4	-5.0	-2.5	-1.0	-2.8	-2.8	-1.5	-4.6	-2.1	5.3
$t\bar{t}$ Parton Shower	1.2	3.0	0.5	1.1	2.2	1.3	1.6	4.6	4.3	6.2
Jet Energy Resolution	5.3	0.8	0.8	0.8	0.1	0.1	-1.3	-1.3	0.4	0.8
Total Systematic Uncertainty (+)	18.6	8.9	5.9	4.3	4.9	4.5	3.8	7.9	8.3	25.2
Total Systematic Uncertainty (-)	16.8	8.5	5.1	4.0	4.9	4.6	3.9	7.9	8.0	25.3

Table 5.10: The dominant systematic uncertainties in the 3 jet region for the MV1 70% working point using LC jets with a JVF cut. The uncertainty values are presented as a relative uncertainty (%) on the nominal scale factor. Uncertainties which contribute less than a 2% relative uncertainty in all jet  $p_T$  bins have been removed from this table to only highlight the size of the dominant systematics. The jet energy scale (JES) is presented as a single uncertainty which envelopes the combination of all the JES components. The total systematic uncertainty shown takes into account all systematic uncertainties presented in Section 5.2.8.

	Jet $p_T$ [GeV/ $c$ ]	20-30	30-40	40-50	50-60	60-75	75-90	90-110	110-140	140-200	200-300
Nominal Scale Factor	1.01	0.929	1.02	1.02	1.00	1.00	0.988	0.998	0.980	0.980	1.22
Statistical Uncertainty	7.0	4.2	3.3	3.1	2.7	3.0	3.1	3.6	4.7	11.4	
Diboson Herwig	-1.3	-0.9	-0.8	-0.9	-0.9	-0.9	-1.1	-1.3	-1.9	-3.8	
JES Down	-12.3	-1.8	-1.7	0.4	0.5	2.4	-0.1	5.2	4.9	8.2	
JES Up	10.1	7.9	3.4	0.4	0.6	-0.9	-0.9	-2.0	-4.0	-13.7	
JVF Down	6.6	1.3	-1.8	-0.1	-2.0	-1.1	-0.4	0.2	0.6	-0.6	
JVF Up	-2.2	-0.9	0.2	1.0	-0.5	0.0	-0.0	0.9	0.6	-2.2	
$B$ -jet $p_T$ Reweighting	-6.8	-1.3	-2.8	-3.8	-0.2	-1.5	-0.2	4.5	5.7	-6.6	
Diboson XS Down	-2.0	-1.2	-0.9	-0.9	-0.9	-0.8	-0.9	-1.1	-1.5	-2.8	
Diboson XS Up	2.0	1.2	0.9	0.9	0.9	0.8	0.9	1.1	1.5	2.8	
$t\bar{t}$ IFSR	1.2	4.3	1.9	2.1	2.6	2.4	1.8	4.0	4.9	4.5	
$t\bar{t}$ Generator	-0.5	-3.9	-2.3	1.1	-2.6	-0.7	1.0	-4.5	-6.0	5.8	
$t\bar{t}$ Parton Shower	6.9	1.5	-1.2	-2.4	0.9	0.2	-0.4	1.8	4.1	-5.3	
Jet Energy Resolution	0.1	2.3	2.3	2.3	-0.4	-0.4	0.8	0.8	2.8	-5.9	
Total Systematic Uncertainty (+)	13.6	8.4	5.5	6.2	4.4	4.4	2.8	8.3	11.3	15.0	
Total Systematic Uncertainty (-)	11.0	7.0	5.5	6.0	4.6	3.6	2.9	8.1	11.4	14.2	

## Chapter 6

# Search for dileptonic $t\bar{t}H$ ( $H \rightarrow b\bar{b}$ ) production

---

*“Absence of evidence is not evidence of absence.”*

— Carl Sagan

---

As detailed in Chapter 2, the search for  $t\bar{t}H$  production provides one of the few ways to measure directly the top Yukawa coupling.

This chapter will detail the search for  $t\bar{t}H$  production with a dileptonically decaying top-pair and a Higgs boson decaying to two  $b$ -quarks, which has been published using the full 8 TeV ATLAS dataset [13]. The author’s contribution has been in the validation and convergence of the event selection with an independent code framework and checking the data overlap with a separate  $t\bar{t}H$  search. This analysis does not include the use of the  $t\bar{t}H$  matrix element method which is described in Chapter 7.

A measurement of the top Yukawa coupling requires a process in which the Higgs boson couples to at least one top quark. In the gluon-gluon fusion Higgs production mechanism and in the diphoton Higgs decay, the top quark couples to the Higgs via a virtual loop. As a virtual loop process has access to energies higher than the real interaction energy, they are sensitive to effects from undiscovered physics processes which become manifest at higher energies. The loop could thus contain particles other than a top quarks which makes it difficult to constrain

the strength of the top Yukawa coupling in a model independent fashion.

The production of a Higgs boson in association with a pair of top quarks is a tree level diagram at leading order which means there are no loop dependencies as shown in Figure 6.1. Identifying and measuring this rate of production will provide a direct measurement of the top Yukawa coupling.

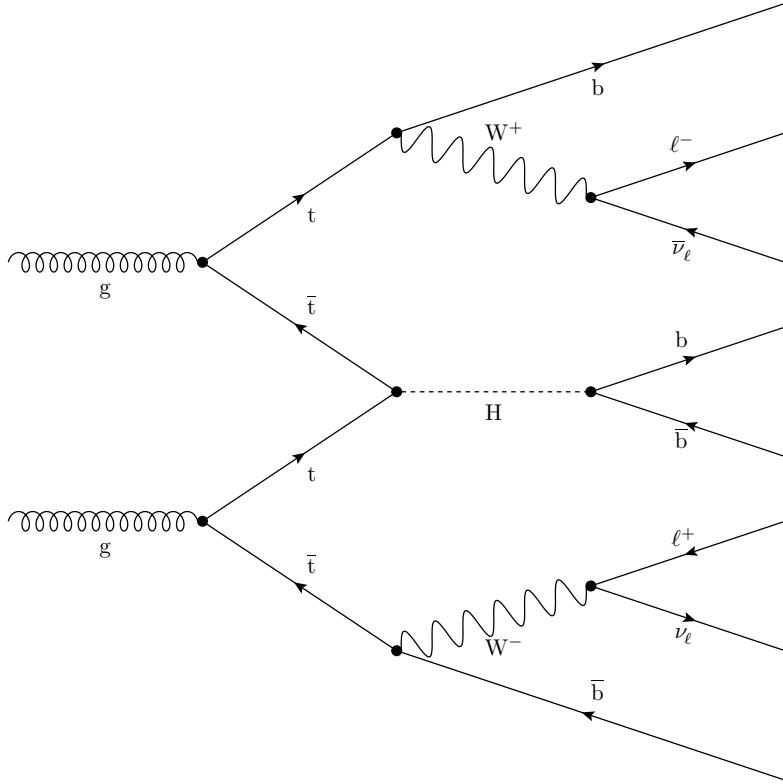


Figure 6.1: A leading-order  $t\bar{t}H$  Feynman diagram.

The total production cross-section at the LHC for a Higgs boson with a mass of  $125 \text{ GeV}/c^2$  is  $22.3 \text{ pb}$  but only  $0.129 \text{ pb}$  of this comes from the  $t\bar{t}H$  production cross-section [41]. Hence the rate of production for a process which gives a direct measurement of the top Yukawa coupling is orders of magnitude smaller than the search channels which can provide an indirect measurement, as can be seen in Figure 6.2. Furthermore, the associated single-top production which can also provide a direct measurement of the top Yukawa coupling has a cross-section which is a factor of seven smaller. Whilst the Higgs branching ratio to two  $b$ -quarks is the largest for a Standard Model Higgs boson with a mass of  $125 \text{ GeV}/c^2$ , one would only expect to produce around 1500  $t\bar{t}H$  events with a  $H \rightarrow b\bar{b}$  decay in  $20 \text{ fb}^{-1}$  of 8 TeV data.

Each top quark decays to  $bW$  with a branching ratio of  $\approx 100\%$ . As such, the decay signature

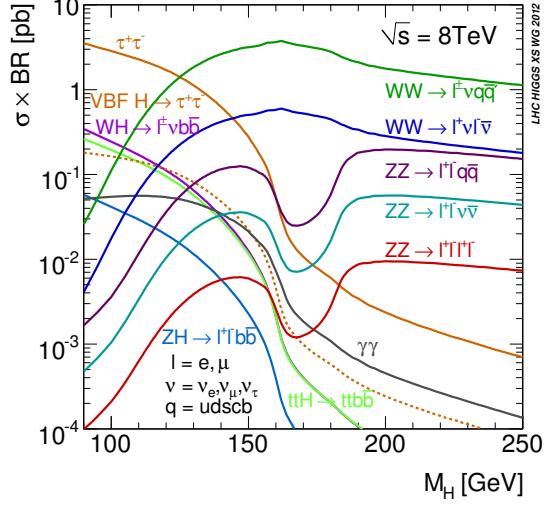


Figure 6.2: The production cross-section multiplied by branching fraction for potential 8 TeV Higgs search channels [41]. The inclusive  $t\bar{t}$  decay is shown in the  $t\bar{t}H$  ( $H \rightarrow b\bar{b}$ ) contribution to this plot.

of a top-pair can be identified using the  $W$ -boson decay. A summary of the top-pair branching ratios is shown in Figure 6.3. Taking into account the dileptonic branching ratio, the number of expected signal events is reduced substantially to around 74, as calculated in Table 6.1.

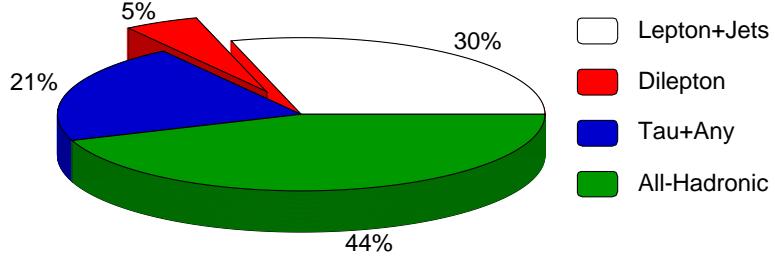


Figure 6.3: Summary of top-pair branching fractions, where lepton refers to electron or muon.

In a physics search, the important figure of merit is the significance with which one may reject the background-only hypothesis. Given a test-statistic,  $t$ , the background-only hypothesis will follow some predicted distribution,  $f(t|bkg)$ . The data will result in an observed value of the test-statistic,  $t_{obs}$ . To quantify whether one would reject the background-only hypothesis, one would compute the  $p$ -value to establish the probability that the hypothesis would produce

Table 6.1: Expected number of  $t\bar{t}H$  signal events in  $20 \text{ fb}^{-1}$  of 8 TeV data, calculated from theoretical cross-sections and branching ratios [41].

$m_H$	$125 \text{ GeV}/c^2$
Total Higgs production [pb]	22.299
$t\bar{t}H$ production [pb]	0.1293
$\text{BR}(H \rightarrow b\bar{b})$	57.7 %
$\text{BR}(t\bar{t} \rightarrow b\ell\nu b\ell\nu)$	4.9 %
$t\bar{t}H \rightarrow b\ell\nu b\ell\nu bb$ [pb]	$3.684 \times 10^{-3}$
Events in data ( $\mathcal{L} = 20 \text{ fb}^{-1}$ )	74

data at least as extreme as the data observed. The calculation of the  $p$ -value is defined as

$$p = \int_{t_{\text{obs}}}^{\infty} f(t|\text{bkg}) dt. \quad (6.1)$$

From the  $p$ -value, one can extract the significance of the observation,  $Z$ , which is related to the  $p$ -value by  $Z = \Phi^{-1}(1 - p)$ , where  $\Phi^{-1}$  is the inverse function of the standard normal distribution. To make a discovery requires a rejection of the background-only hypothesis at the threshold of  $Z = 5$ . It is possible to quantify the discovery potential of an analysis by calculating the expected discovery significance. The expected discovery significance is a function of the number of expected signal events ( $s$ ) and the number of expected background events ( $b$ ). For a single-bin Poisson counting experiment, the expected discovery significance is approximately equal to  $s/\sqrt{b}$ .

Assuming that the search for dileptonic  $t\bar{t}H$  production is a statistically limited analysis, we can make the physically motivated assumption that the dominant background process will be top-pair production with two additional  $b$ -jets. The NLO  $t\bar{t}+b\bar{b}$  cross-section predicts around 550 of these background events in  $20 \text{ fb}^{-1}$  of 8 TeV data. The optimisation of an analysis aims to improve the expected discovery significance but even in this simplified example, the expected discovery significance only just reaches  $3\sigma$  and there are additional backgrounds to consider and systematic uncertainties to take into account. This therefore highlights the difficulty this analysis faces in trying to make a discovery.

An outline of the analysis will be presented in Section 6.1 motivating the object definition, event selection and simulation samples discussed in Section 6.2-6.4. The key feature of this analysis is the heavy flavour modelling used to control the dominant  $t\bar{t}+\text{jets}$  backgrounds. A number of techniques used are presented in Section 6.5. Information on the event classification

is presented in Section 6.6. Details of the neural network used to separate the signal process from the dominant background processes are presented in Section 6.7. The final analysis distributions which provide the expected number of events used in the analysis likelihood function are shown in Section 6.8. This is followed by a description of the fitting procedure used to constrain systematic uncertainties with data and the final analysis results in Section 6.9, where upper limits are set on the signal strength parameter. A closing comment on the outlook of this analysis for Run 2 is presented in Section 6.10 which provides additional motivation for Chapter 7.

## 6.1 Analysis Outline

This analysis applies a profiled likelihood fitting procedure to infer information regarding the contribution from the signal process in data. Events are selected by passing a dileptonic  $t\bar{t}$  event selection which has been optimised to enhance the contribution from the  $t\bar{t}H$  signal process. Events with at least two  $b$ -jets are categorised based on the number of reconstructed jets and the number of  $b$ -tagged jets. In this analysis events are categorised into one of six regions based on whether there are exactly two, exactly three or at least four jets with exactly two, exactly three or at least four  $b$ -jets.

Each category contributes a distribution to the final combined likelihood where the number of events in each bin of the distribution is a discrete random variable,  $n_i$ , which is Poisson distributed about an expected mean  $v_i(\mu) = \mu s_i + b_i$ . The expected number of signal events ( $s_i$ ) and the total number of background events ( $b_i$ ) in each bin are predicted by simulated data samples. When making an observation,  $n_i$  is measured from data, but when examining the sensitivity of the analysis,  $n_i$  can be estimated from a pseudo-dataset. The parameter of interest is the signal strength  $\mu$  which is used to scale the contribution of the signal process when testing different hypotheses. The two hypotheses of interest are the background-only hypothesis where  $\mu = 0$  and the signal-plus-background hypothesis where  $\mu = 1$ . The signal model considered when  $\mu = 1$  is the Standard Model prediction.

The total error on the number of events in the region where this analysis has the most sensitivity is driven by the dominant backgrounds and is on the order of 40%. This motivates the fitting procedure which uses the data in the other regions to reduce this error down to around 5%. As these expected yields are fitted to the observed data, they require the best possible modelling.

For categories with a low signal to background ratio, the distribution of the summed transverse momenta for leptons and jets is used in the likelihood. Categories with a large signal to background ratio use a multivariate classifier optimised for each individual region to separate signal and background events.

The final likelihood function,

$$\mathcal{L}(\mu, \boldsymbol{\theta}) = \prod_{c \in \text{cat}} \left( \prod_{i \in \text{bin}} \text{Pois}(n_i | v_i(\mu, \boldsymbol{\theta})) \right) \prod_{s \in \text{sys}} f(m_s | \theta_s), \quad (6.2)$$

contains a number of nuisance parameters ( $\boldsymbol{\theta}$ ) which reflect the uncertainty in the analysis model. The expected number of signal and background events is parameterised as a function of  $\boldsymbol{\theta}$  such that  $v_i(\mu) \rightarrow v_i(\mu, \boldsymbol{\theta}) = \mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta})$ . The nuisance parameters contain all sources of error considered by the analysis. By fitting the model to the observed data, the overall impact of the nuisance parameters is constrained and the analysis model is improved. The nuisance parameter associated to each source of systematic error,  $\theta_s$ , is constrained by an idealised auxiliary measurement  $m_s$  using the constraint term  $f(m_s | \theta_s)$  [119, 120].

The following sections will detail the relevant analysis components required to achieve a likelihood which can discriminate between signal and background processes by successfully modelling the background-only hypothesis.

## 6.2 Object and Event Selection

The object and event selections evolved from those recommended by the ATLAS Collaboration’s Top Physics Working Group [81]. This analysis requires sensitivity to jet multiplicities larger than a standard top analysis which is optimised to enrich the two  $b$ -jet region with top-pair events only. Details of the object definitions used in this section can be found in Chapter 4.

The signature for the  $t\bar{t}H$  signal process is two leptons (any combination of electrons and muons), four  $b$ -tagged jets and real missing transverse energy ( $E_T^{\text{miss}}$ ) from the two neutrinos. It becomes a natural distinction to perform the analysis in different regions of jet and  $b$ -tagged jet multiplicities as different heavy-flavour backgrounds have varying contributions in these regions. This in turn allows for constraints to be extracted from the comparisons between data and prediction.

### 6.2.1 Lepton Selection

A summary of lepton object requirements for this analysis is presented in Table 6.2. Single lepton triggers which are not prescaled have been used in this analysis to select events with at least one electron or muon from the full 8 TeV dataset. For both electrons and muons, a logical OR is made between two triggers to improve the efficiency for different  $p_T$  ranges. The lower  $p_T$  triggers have isolation requirements, which reduces the efficiency at high  $p_T$ . This is recovered by using a higher threshold trigger without an isolation requirement. The  $p_T$  thresholds for the electron triggers are 24 or 60 GeV/c and for the muon triggers are 24 or 36 GeV/c. The triggers force a minimum lepton  $p_T$  in the object definition, but acceptance is increased by allowing the subleading lepton  $p_T$  to be less than the trigger thresholds. Trigger matching is applied to ensure that one of the leptons used in the analysis fired the trigger. The spatial matching requirement is that a selected lepton has to be within  $\Delta R < 0.15$  of the lepton which fired the trigger for the event to be used.

Table 6.2: Summary of lepton object selections for which definitions can be found in Section 4.2-4.3.

Criterion	Electron	Muon
Leading $p_T$	$\geq 25 \text{ GeV}/c$	$\geq 25 \text{ GeV}/c$
Subleading $p_T$	$\geq 15 \text{ GeV}/c$	$\geq 15 \text{ GeV}/c$
$ \eta $	$\leq 2.47$	$\leq 2.5$
Track isolation	$p_{T\text{-cone}}/p_T < 12\%$	Mini-isolation $10/p_T < 5\%$
Quality	MediumLLH	Tight Muid
$z_0$	$< 2 \text{ mm}$	$< 2 \text{ mm}$

### 6.2.2 Jet Selection

A summary of jet object requirements for this analysis is presented in Table 6.3. This analysis requires the ability to tag jets which have originated from a  $b$ -quark. This requirement motivates the jet selection to ensure that any jet considered has the potential to be  $b$ -tagged. The hadronic jets used have their jet energy scale calibrated with local cluster reweighting as described in Section 4.4.

### 6.2.3 Object Overlap Removal

There are cases where the object reconstruction algorithms presented in Chapter 4 may identify multiple objects from the same source. Overlap removal is performed to remove objects which

Table 6.3: Summary of jet object selections for which definitions can be found in Section 4.4-4.5.

Criterion	Jet
Algorithm	anti- $k_t$ ( $R = 0.4$ )
$p_T$	$\geq 25 \text{ GeV}/c$
$ \eta $	$\leq 2.4$
$b$ -tagging algorithm	MV1
$b$ -tagging operating point	70%
Jet vertex fraction	$\geq 0.5$

have been incorrectly reconstructed as detailed in Section 4.6.

During jet reconstruction, a true electron is treated as just another calorimeter energy deposit. This means it is possible that an electron will be reconstructed as a jet. If a jet is found to be within  $\Delta R < 0.2$  of an electron, this jet is removed from consideration. After this, if an electron is found to be within  $\Delta R < 0.4$  of a jet, the electron is removed from consideration. This is to remove the effect of jet radiation being incorrectly identified as a prompt electron.

Heavy flavour hadrons inside a jet can decay semi-leptonically and produce a muon. In order to prevent these muons being reconstructed as a prompt muon produced at the primary vertex, muons are required to be separated from jets by  $\Delta R > 0.4$ .

### 6.3 Event Selection

An event selection is performed to separate  $t\bar{t}H$  signal events from all background processes. After this selection, events are separated out into control and signal regions based on the number of jets and the number of  $b$ -jets in an event.

The primary vertex of an event is required to have at least five reconstructed tracks to pass the early stage of the event selection. The analysis requires two opposite-sign leptons with  $p_T$  greater than  $15 \text{ GeV}/c$ . At least one lepton is required to have a  $p_T$  greater than  $25 \text{ GeV}/c$  in order to be above the trigger threshold. At least one of those leptons is required to be trigger matched. Simulated leptons are required to be truth matched to allow fake estimations to be obtained with events which fail the truth matching. Selected events are required to have at least two jets. Events which contain a badly reconstructed jet or in which an electron and muon share the same track are vetoed. After this selection is satisfied, events are categorised according to the lepton flavour,  $ee$ ,  $\mu\mu$  and  $e\mu$ , for subsequent event selections.

In the  $ee$  and  $\mu\mu$  channels, a number of requirements are made on the invariant mass of the

dilepton pair. Contamination from leptonic  $Z$  decays is reduced by removing events within a window  $\pm 8 \text{ GeV}/c^2$  of the  $Z$  mass of  $91 \text{ GeV}/c^2$ . If an event has exactly two  $b$ -jets, the dilepton invariant mass is required to be greater than  $60 \text{ GeV}/c^2$  to remove a region of phase-space where the  $Z + \text{jets}$  background is poorly modelled. For events with at least three  $b$ -jets, the invariant mass of the dilepton pair is required to be greater than  $15 \text{ GeV}/c^2$  to reduce the contribution from hadronic resonance decays.

In the  $e\mu$  channel, the scalar sum of the transverse momenta of the leptons and jets is required to be greater than  $130 \text{ GeV}/c$ . This cut is an effective way to remove non- $t\bar{t}$  events.

After the full event selection is applied, the dominant background contribution comes from  $t\bar{t} + \text{jets}$  events. The events from each dilepton channel are combined together and categorised by the number of jets and the number of  $b$ -jets to provide six event categories: 2 jets, 2  $b$ -jets; 3 jets, 2  $b$ -jets; 3 jets, 3  $b$ -jets;  $\geq 4$  jets, 2  $b$ -jets;  $\geq 4$  jets, 3  $b$ -jets and  $\geq 4$  jets,  $\geq 4$   $b$ -jets.

### 6.3.1 Multilepton Veto

The top Yukawa coupling can be measured from any process with an interaction between a top quark and a Higgs boson. Alternative decays of a Higgs boson produced with an associated top-pair can also provide a measurement. However, to facilitate improving the statistical uncertainty on the top Yukawa coupling by combining the results from multiple analyses, one needs to ensure that there is no overlap in the events selected from data.

This problem can be alleviated if all analyses use the same object definitions. However, this is not always the optimal choice for an analysis. The single-lepton analysis, which is complementary to this analysis, requires a single lepton with  $p_T$  greater than  $25 \text{ GeV}/c$ . However, this dilepton analysis has a lower threshold for the subleading lepton. If the single lepton analysis does not check whether an event also has an associated low  $p_T$  lepton, there can be unintentional overlap in data where the same event can be used with different assumptions applied.

Similarly, the decay of the Higgs boson to two  $W$  bosons can produce events which have the same lepton topology as this analysis if the lepton definitions vary. The multilepton veto was designed to run over the standard dilepton analysis and to check whether an event has any additional leptons with a  $p_T$  below  $15 \text{ GeV}/c$ , which could result in the event being selected by a multilepton analysis.

The lepton definitions in the multilepton analysis have many differences compared to the

dilepton analysis. A minimum  $p_T$  of 10 GeV/ $c$  is required rather than 15 GeV/ $c$ . The lepton quality and isolation requirements are much tighter than in the dilepton analysis. The multilepton analysis also implements more complex overlap removal requirements as the low  $p_T$  thresholds will make it harder to distinguish between lepton types. An electron found within  $\Delta R < 0.1$  of a muon is removed. If an electron is then found within  $\Delta R < 0.1$  of another electron, the lowest  $p_T$  electron is removed. If an electron is found within  $\Delta R < 0.3$  of a jet, then the jet is removed from the selection. Finally a muon is removed if it is found within  $\Delta R < 0.04 + 10/p_T^\mu$  of a jet, where the  $p_T$  of the muon is provided in GeV/ $c$ .

A tool was developed to run in parallel with the nominal analysis to determine if an event which passed the dilepton selection in data would also have been selected by a multilepton analysis. This study showed that the level of overlap between the two analyses was negligible. If they will be combined in the future, the simplest choice will be to remove overlapping events from data in the dilepton analysis without adjusting the prediction.

## 6.4 Monte Carlo Simulation Samples

The description of the analysis in Section 6.1 explained that a good background prediction is required to ensure that the likelihood function can be successfully fitted to the data. These predictions are extracted from simulated data which use event generators to model the hard parton interaction and then apply parton showering and hadronisation techniques to model the QCD effects. The selection of simulation samples with particular event generator and hadronisation options has been driven by the high jet multiplicity modelling of different generators. The number of jets in the signal and background processes can be difficult for leading-order generators to model. For instance, a leading-order event generator will have difficulty producing samples of  $t\bar{t}+jets$  which reproduce distributions in data when the jet multiplicity is greater than two. The accuracy of such a sample would be dependent on the parton shower being able to correctly model objects which can be produced by  $N^{(n)}LO$  matrix elements. In general, leading-order generators are unable to provide good agreement with data, unless they provide a multi-leg leading-order calculation where higher-order tree level diagrams are calculated.

A summary of the nominal analysis samples are provided in Table 6.4 with details about the event generator and parton shower model. An inclusive signal sample is generated using

POWHEG-BOX+HELAC-NLO (POWHEL). This allows for consistency across all  $t\bar{t}H$  channels and also enables  $t\bar{t}H$  events which are not  $H \rightarrow b\bar{b}$  to be categorised as signal events if they would pass the selection in data. An estimation of fake events comes directly from combining the simulated background samples with some additional negligible background samples and taking events where the lepton fails to be truth matched but passes the rest of the selection. The use of Monte Carlo to calculate the fake estimation allows the inclusion of the high Monte Carlo statistics which are much larger than any other data-driven methods and appears to provide a good description when compared with data.

## 6.5 Heavy Flavour Modelling

In this analysis, heavy flavour jets refer to  $b$ -quark and  $c$ -quark initiated jets and heavy flavour backgrounds refer to the  $t\bar{t}+b\bar{b}$  and  $t\bar{t}+c\bar{c}$  processes. This analysis is particularly sensitive to the modelling of  $t\bar{t}+\text{jets}$  events in order to correctly model the background-only hypothesis. In the regions where one would expect the highest contribution from the signal process, the background is dominated by  $t\bar{t}$  events with additional heavy flavour jets. Understanding the  $t\bar{t}+\text{jets}$  background processes becomes the highest priority in this analysis in order to remove the possibility that a poorly modelled background process is misinterpreted as a signal contribution. A number of techniques to address the modelling of  $t\bar{t}+\text{jets}$  events are presented in this section.

### 6.5.1 Heavy flavour classification of $t\bar{t}+\text{jets}$

The production of a dilepton top-pair event will produce two  $b$ -jets. However, QCD radiation will result in additional reconstructed jets of varying flavours which will migrate into the signal regions of this analysis. It is important to provide a classification of these events with respect to the additional jets so that their overall normalisation can be correctly extracted from data.

Parton level information is used to identify  $b$ - and  $c$ -quarks which were not produced from the decay of a  $W$  boson produced by the decay of a top quark. These partons are spatially matched to particle jets to categorise an event. Particle jets are reconstructed at truth level from stable particles using the anti- $k_t$  algorithm with a radius parameter of 0.4. They are required to have  $p_T$  greater than 15 GeV/c and lie within  $|\eta|$  less than 2.5. A particle jet is considered matched if a parton lies within  $\Delta R < 0.4$ . Events with at least one  $b$ -quark matched to a particle

Table 6.4: Summary of the baseline simulation samples used in the nominal analysis.

Sample	Event Generator	Order	Normalisation	Parton Shower	PS Tune	PDF
$tH$	POWHEG [101, 121–124]	NLO	Pythia 8 [103]	AU2 [125]	CT10 [107]	
$t\bar{t}$	POWHEG-Box [101, 121, 122]	NLO	NNLO+NNLL	Pythia 6 [105]	Perugia2011C [106]	
Single Top	PowHEG-Box [101, 121, 122]	NLO	aNNLO	Pythia 6 [105]	Perugia2011C [106]	
$Z+\text{jets}$	ALPGEN [104]	Multileg-LO	NLO	Pythia 6 [105]	AUET2 [108]	CTEQ6L1 [109]
Diboson	ALPGEN [104]	Multileg-LO	NLO	HERWIG [102]	AUET2 [108]	CTEQ6L1 [109]
$t\bar{t}V$	MADGRAPH 5 [126]	LO	NLO	Pythia 6 [105]	AUET2B [127]	CTEQ6L1 [109]

jet are categorised as  $t\bar{t}+b\bar{b}$ . Events without a matched  $b$ -quark but with at least one matched  $c$ -quark are categorised as  $t\bar{t}+c\bar{c}$ . Any remaining events are categorised as  $t\bar{t}$ +light jet events.

### 6.5.2 Top $p_{\mathrm{T}}$ reweighting

The distribution of the number of jets in this analysis has been notoriously difficult to model due to the high multiplicities being dealt with. These typically fall into a region where leading-order and even next-to-leading order generators struggle to model data in a satisfactory way. Many studies within the ATLAS Collaboration have converged on the opinion that the poor modelling of the number of jets in an event is tied to the modelling of the top quark. The distribution of  $H_{\mathrm{T}}$  in various analysis regions generally saw a slope where the low  $p_{\mathrm{T}}$  region was underestimated and the high  $p_{\mathrm{T}}$  region was overestimated. This analysis makes use of an artificial neural network to discriminate between signal and background which makes it sensitive to variables which are correlated to the modelling of the top quark. These features are therefore undesirable and need to be corrected.

A top differential cross-section measurement was made using 7 TeV data with the ATLAS detector [128]. The differential cross-sections were made with respect to different unfolded parton level variables and different simulated samples were compared with data. The two variables of interest to this analysis were the top  $p_{\mathrm{T}}$  and the  $t\bar{t}$ -system  $p_{\mathrm{T}}$ . Each of these showed that the simulated top quarks were consistently produced too hard with respect to the data. These distributions are shown in Figure 6.4.

A sequential reweighting is used to correct the top and  $t\bar{t}$ -system modelling. Reweighting the  $t\bar{t}$ -system  $p_{\mathrm{T}}$  has been observed to improve agreement with data in the distribution of the number of jets and reweighting the top  $p_{\mathrm{T}}$  has been observed to improve event level kinematic distributions. The ratio of measured differential cross-sections in simulation and in data is used to derive a reweighting which first corrects the transverse momenta of the  $t\bar{t}$ -system and then corrects the residual difference in the measured cross-sections with respect to the transverse momenta of the top quark. The impact of this correction can be seen in Figure 6.5-6.6 which are taken from the single lepton  $t\bar{t}H (H \rightarrow b\bar{b})$  analysis where a slope can be seen prominently before applying this correction.

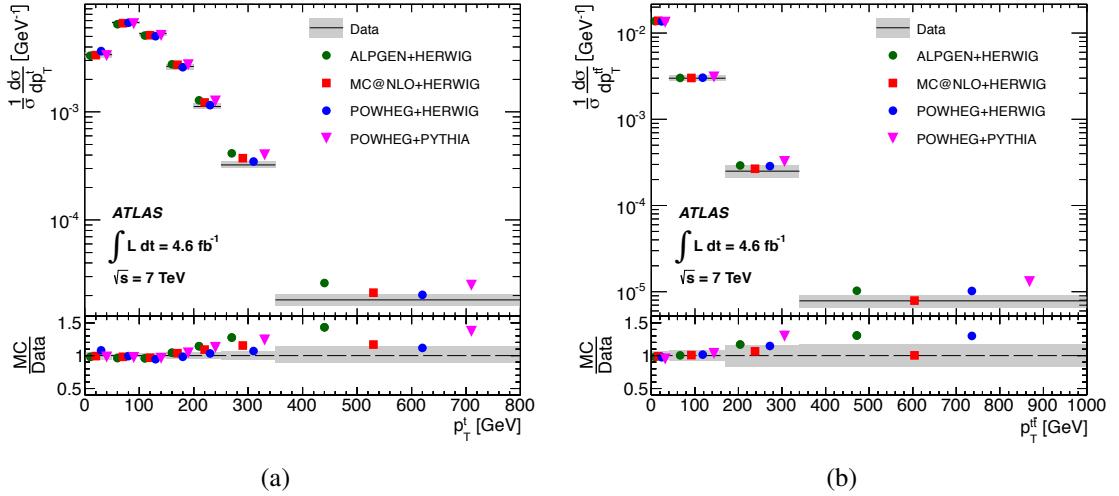


Figure 6.4: Top differential cross-section measured with respect to the top  $p_T$  (Figure 6.4a) and with respect to the  $t\bar{t}$  system  $p_T$  (Figure 6.4b) [128]. The baseline  $t\bar{t}$  sample used in this analysis corresponds to the POWHEG+PYTHIA label on these plots and it can be seen that the generated top kinematics are harder than data.

### 6.5.3 NLO $t\bar{t}+b\bar{b}$ reweighting

It was presented in Chapter 2 that the inclusive  $t\bar{t}$  cross-section is known to NNLO but the  $t\bar{t}+b\bar{b}$  cross-section is only known to NLO. The  $t\bar{t}+b\bar{b}$  background is an important background which requires the best possible modelling. The  $t\bar{t}+b\bar{b}$  sample is generated as part of an inclusive  $t\bar{t}$  sample generated by POWHEG, but the additional heavy flavour component of this sample is generated through the parton showering, not with a matrix element. SHERPA is an alternative NLO event generator which has recently managed to produce  $t\bar{t}+b\bar{b}$  samples with the use of the OPENLOOPS package (SHERPAOL) [45, 129, 130]. This is currently the only model where the  $t\bar{t}+b\bar{b}$  events are generated entirely from the matrix element at NLO. To assess the modelling of the  $t\bar{t}+b\bar{b}$  sample with POWHEG, a comparison was made between three different generators - POWHEG, MADGRAPH 5 and SHERPAOL.

An event is classified based on the number of heavy flavour particle jets it contains and the number of heavy flavour hadrons. These objects are spatially matched to reconstructed jets using the algorithm described in Section 6.5.1. Only heavy flavour hadrons which have not been produced by the decay of a top quark are considered. Gluon splitting to a  $b\bar{b}$  pair can result in two  $B$ -hadrons being matched to a particle jet. If a sample contains at least one jet with a double  $B$ -hadron matching then it is labelled with  $B$ . If a single hadron is matched then it is labelled with  $b$ . An event is classified based on the number of successful matchings found.

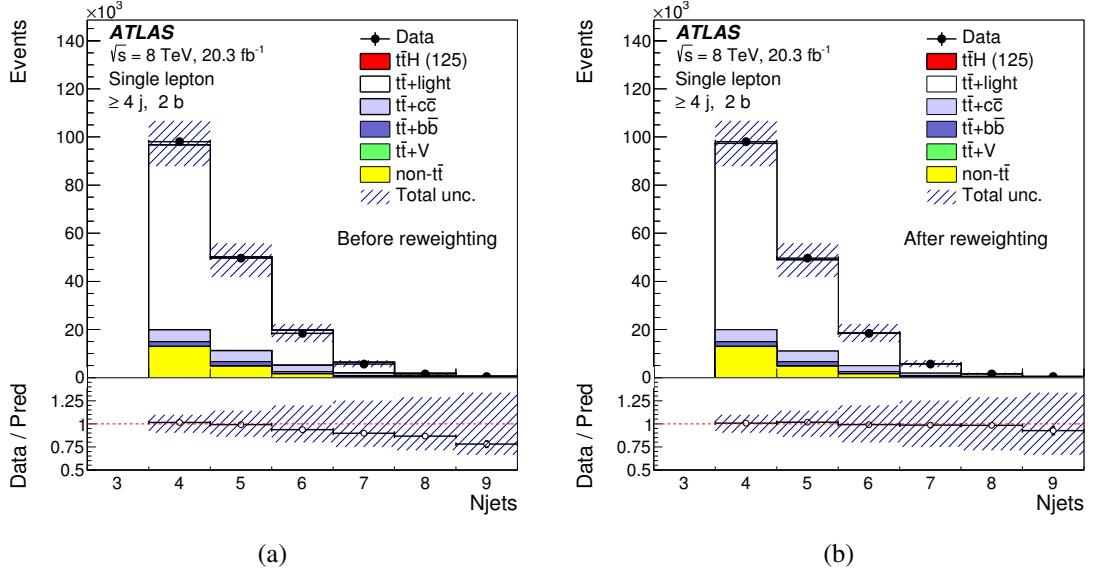


Figure 6.5: Prefit distributions from the single lepton  $t\bar{t}H (H \rightarrow b\bar{b})$  analysis showing poor modelling of the number of selected jets in an event [13]. An improvement can be observed in the data-prediction ratio after applying the sequential top  $p_T$  reweighting.

This heavy flavour labelling can be applied on all  $t\bar{t}$  samples and allows an independent comparison of different event generators and the modelling of additional heavy flavour hadrons. PowHEG generates  $t\bar{t}$  at NLO but is dependent on the parton shower to model additional heavy flavour hadrons. MADGRAPH 5 can model a  $t\bar{t}+b\bar{b}$  final state with a leading-order matrix element. The comparison presented in Figure 6.7 shows that PowHEG and MADGRAPH 5 both agree with the amount of additional heavy flavour jets despite their different generation procedures. However, the comparison shows that PowHEG generally underestimates the amount when compared to SHERPAOL once an event is classified with merged gluon splittings or more than two additional matched  $B$ -hadrons.

As SHERPAOL represents an improved theoretical model over PowHEG, the  $t\bar{t}+b\bar{b}$  sample is reweighted to match this prediction. Two reweighting procedures are applied to the  $t\bar{t}+b\bar{b}$  sample to improve the modelling. The relative cross-sections of  $t\bar{t}+b\bar{b}$  in each of the heavy flavour classifications are corrected to the amounts predicted by SHERPAOL, whilst preserving the total  $t\bar{t}+b\bar{b}$  cross-section. It was also observed that top  $p_T$  distribution in the  $t\bar{t}+b\bar{b}$  sample after being reweighted to data matched very closely to the prediction from SHERPAOL. Therefore, a kinematic reweighting is applied to the  $t\bar{t}+b\bar{b}$  sample to match the SHERPAOL prediction instead of applying the data reweighting described in Section 6.5.2. This allows the shape of the  $t\bar{t}+b\bar{b}$

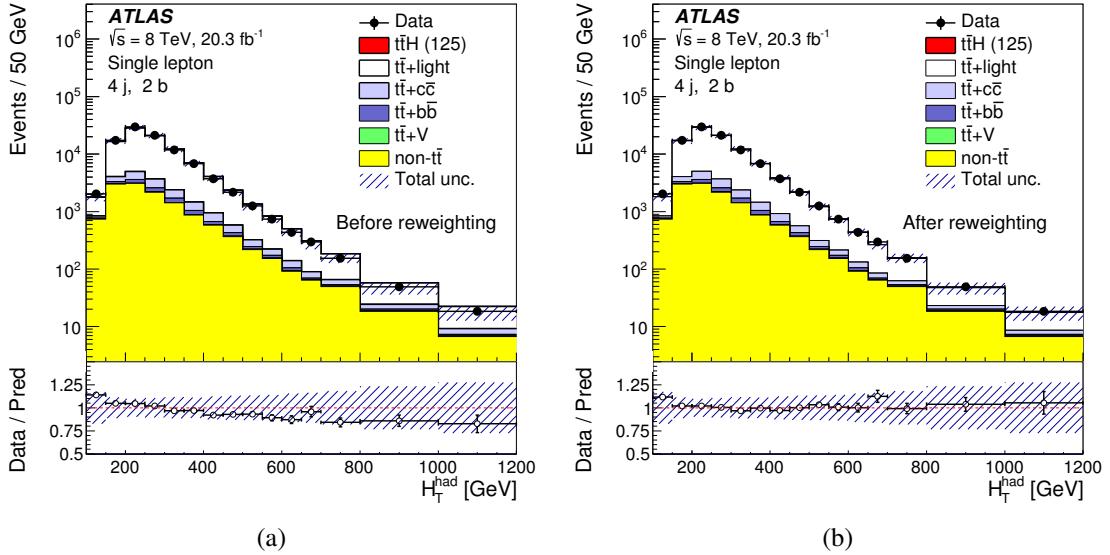


Figure 6.6: Prefit distributions from the single lepton  $t\bar{t}H (H \rightarrow b\bar{b})$  analysis showing poor modelling of the hadronic  $H_T$  variable [13]. An improvement can be observed in the data-prediction ratio after applying the sequential top  $p_T$  reweighting.

sample to be dependent on a theoretical model which can then be varied to provide theoretically motivated uncertainties.

#### 6.5.4 Tag Rate Function

Whilst the simulation samples have a large number of generated events, requiring a large number of jets to be  $b$ -tagged quickly reduces the acceptance of simulated events. This can be seen with a simple example considering an event with four true  $b$ -jets which are tagged with an efficiency of 70%. The combined probability of tagging all four (when treating the efficiency as the probability of tagging) will be reduce the sample size to 24% of its original size. The drop in the number of selected events from  $b$ -tagging is undesirable when using the shape of distributions in the simultaneous fit as statistical fluctuations can cause instability in the fit. The method employed in this analysis to overcome this problem is the tag rate function (TRF) [131]. This method allows the analysis to use the full set of events available prior to  $b$ -tagging, and then weight the events according to the probability that they would have passed the  $b$ -tagging cut.

Mathematically, the problem is formulated as,

$$P(1|N) = \sum_i^N (P(i) \times \prod_{i \neq j}^N (1 - P(j))), \quad (6.3)$$

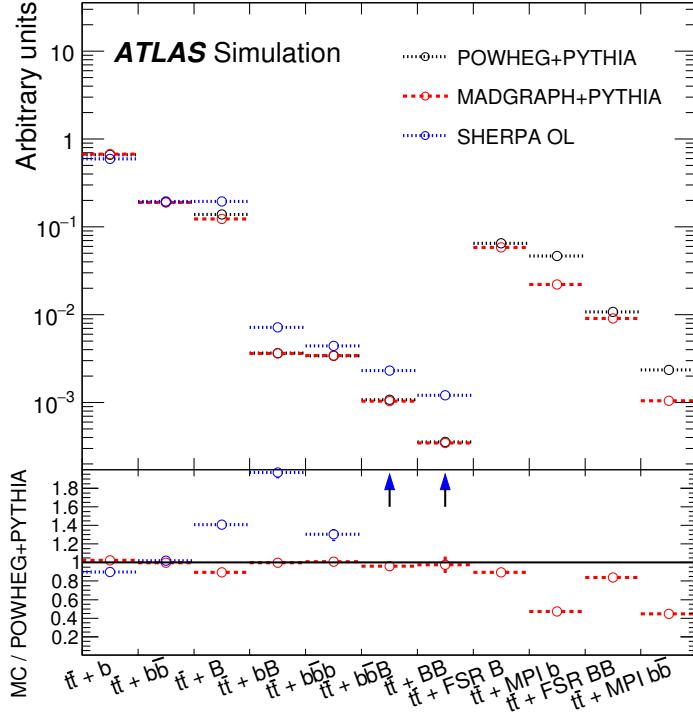


Figure 6.7: A summary of the relative predicted number of  $t\bar{t}+b\bar{b}$  split into a fine classification of different heavy flavour categories dependent on the nature of the additional  $B$ -hadrons in an event [13].

where  $P(i)$  is the probability of the  $i^{\text{th}}$  jet being  $b$ -tagged and the probability that one out of  $N$  jets is tagged is  $P(1|N)$ . The probability that at least one jet is  $b$ -tagged is given as

$$P(1_{>}|N) = 1 - \prod_i^N (1 - P(i)), \quad (6.4)$$

which is 1 minus the probability that no jets are tagged. Given this formulation, one can calculate the probability of higher  $b$ -jet multiplicities as

$$P(2_{>}|N) = P(1_{>}|N) - P(1|N). \quad (6.5)$$

With this method, an event with  $N$  jets will contribute to distributions in all regions between 0 and  $N$   $b$ -jets. An event will have the largest weighted contribution in the region where it would most likely have been selected if a direct tagging method had been applied. Closure tests have been performed by comparing the yields using direct tagging with yields using TRF in a given  $b$ -jet region. A scale factor is applied to the signal and the dominant backgrounds to correct

the yields to those calculated when applying a direct  $b$ -tagging cut. A summary of these scale factors derived from the yields in the two  $b$ -tagging methods is provided in Table 6.5, where a level of non-closure can be observed for the  $t\bar{t}$ +LF background in the  $\geq 4$  jets,  $\geq 4$   $b$ -jets region. The total contribution of this background is very small, so the impact of this correction factor is minimal.

Table 6.5: Normalisation correction factors applied to rescale the expected event yield using TRF to the expected event yield using direct tagging for signal and  $t\bar{t}$ +jets backgrounds in all analysis regions.

Region	$t\bar{t}H$	$t\bar{t}+b\bar{b}$	$t\bar{t}+c\bar{c}$	$t\bar{t}+\text{LF}$
$2j, 2b$	$1.02 \pm 0.02$	$1.02 \pm 0.01$	$1.00 \pm 0.009$	$1.01 \pm 0.001$
$3j, 2b$	$1.00 \pm 0.01$	$1.01 \pm 0.01$	$1.00 \pm 0.006$	$1.01 \pm 0.002$
$3j, 3b$	$1.06 \pm 0.02$	$1.04 \pm 0.02$	$1.04 \pm 0.02$	$1.03 \pm 0.02$
$\geq 4j, 2b$	$1.00 \pm 0.008$	$1.01 \pm 0.006$	$1.01 \pm 0.006$	$1.01 \pm 0.002$
$\geq 4j, 3b$	$1.03 \pm 0.008$	$1.02 \pm 0.01$	$1.02 \pm 0.01$	$0.99 \pm 0.01$
$\geq 4j, \geq 4b$	$1.07 \pm 0.01$	$1.02 \pm 0.03$	$1.07 \pm 0.06$	$1.91 \pm 0.2$

## 6.6 Event Classification

Events are categorised according to the number of jets and  $b$ -jets after passing the event selection. This categorisation creates multiple control regions dominated by different background processes with different relative fractions. A simultaneous fit is performed across all analysis regions which fits both the shape and the number of expected events to data. The distributions used are either the scalar sum of lepton and jet transverse momenta ( $H_T$ ) or a multivariate neural network classifier (NN). The categorisation is shown in Table 6.6 and the background composition of the analysis regions is shown in Figure 6.8.

The prefit event yields for signal, background and data for each analysis region is presented in Table 6.7. These numbers constitute the summed  $s$ ,  $b$  and  $n$  values described in Section 6.1 for each category used to define the analysis likelihood.

The simultaneous fit allows the amount of each background to float across all analysis regions within the associated statistical, theoretical and detector systematic uncertainties. The fit also simultaneously constrains the impact of the systematic uncertainties with the data. This results in an improved estimation of the background composition with reduced errors.

Table 6.6: Summary of the discriminating variable used in the simultaneous fit for each analysis region.  $H_T$  is defined by the scalar sum of the transverse momenta of the jets and leptons in the event and NN refers to a neural network trained in that region.

Region	$2 b$ -jets	$3 b$ -jets	$\geq 4 b$ -jets
2 jets	$H_T$		
3 jets	$H_T$	NN	
$\geq 4$ jets	$H_T$	NN	NN

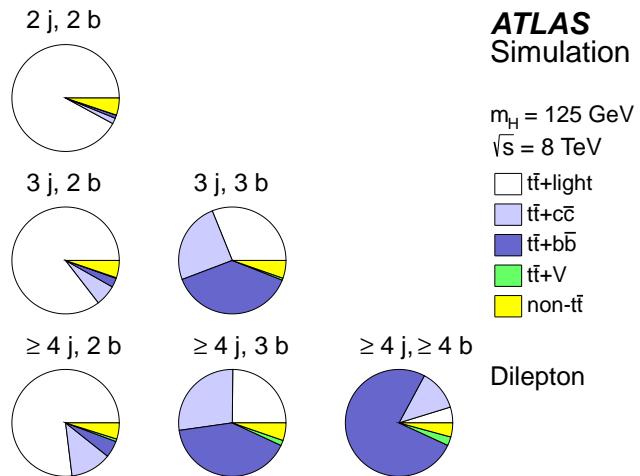


Figure 6.8: The background compositions for each analysis region [13]. A simultaneous fit is performed across these regions in order to constrain systematic uncertainties and different background rates using data. Different regions are dominated by different  $t\bar{t}$ +jet combinations.

Table 6.7: Prefit event yields for signal, backgrounds and data in each analysis category [13]. The quoted uncertainties are the sum in quadrature of the statistical and systematic uncertainties on the yields.

	2 j, 2 b	3 j, 2 b	3 j, 3 b
$t\bar{t}H$ (125)	$1.5 \pm 0.2$	$5.3 \pm 0.5$	$2.2 \pm 0.3$
$t\bar{t}$ +LF	$14000 \pm 1800$	$8100 \pm 880$	$96 \pm 21$
$t\bar{t}+c\bar{c}$	$270 \pm 170$	$600 \pm 320$	$76 \pm 44$
$t\bar{t}+b\bar{b}$	$150 \pm 87$	$260 \pm 130$	$120 \pm 65$
Z+jets	$330 \pm 30$	$190 \pm 49$	$8.2 \pm 3.1$
Single top	$430 \pm 71$	$270 \pm 30$	$7.6 \pm 3.5$
Diboson	$6.8 \pm 2.2$	$4.2 \pm 1.5$	$\leq 0.1 \pm 0.1$
$t\bar{t}V$	$8.4 \pm 2.7$	$21 \pm 6$	$1.9 \pm 0.6$
Lepton misID	$21 \pm 10$	$33 \pm 17$	$0.8 \pm 0.4$
Total	$15000 \pm 1900$	$9500 \pm 1000$	$310 \pm 85$
Data	15296	9996	374
$S/B$	<0.001	0.001	0.006
$S/\sqrt{B}$	0.012	0.053	0.114

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

## 6.7 Neural Network Classifier

This analysis splits the inclusive sample of dilepton  $t\bar{t}$  events into regions dependent on the number of jets and the number of  $b$ -jets identified. This classification ensures there are regions with a high proportion of signal events to background events, as shown in Table 6.7, whilst providing additional regions which are dominated by different  $t\bar{t}$ +jets backgrounds, as shown in Figure 6.8, which are used to constrain systematic uncertainties to data. A signal-rich region is defined by having  $S/B > 1\%$  and  $S/\sqrt{B} > 0.3$ , where  $S$  and  $B$  represent the expected number of signal and background events respectively.

To optimise this analysis to search for the  $t\bar{t}H$  process, a multivariate classifier is used to separate signal and background events in the signal-rich regions. This corresponds to the three regions:  $\geq 4$  jets,  $\geq 4$   $b$ -jets,  $\geq 4$  jets, 3  $b$ -jets and 3 jets, 3  $b$ -jets. A series of kinematic variables have been selected based on their separation power in these three regions. These variables are ranked in each region based on the amount of separation between the signal and background distributions and the top ten are combined by a neural network classifier [132, 133]. The variables and their definitions are shown in Table 6.8 and their ranking within the neural network is shown in Table 6.9.

The neural network is built using the NeuroBayes software package [134, 135]. NeuroBayes has two interesting features which aid the discrimination of the final neural network output. Firstly, NeuroBayes is designed to preprocess input variables prior to providing the variables to the neural network training algorithms. Preprocessing is designed to decorrelate the input variables through transformations of the distributions so that the neural network only has to learn the non-linear correlations between input variables. Secondly, NeuroBayes uses Bayesian-based regularisation on the neural network training in order to constrain the final output and provide physical predictions.

Each input variable is transformed to a normalised unit Gaussian distribution and decorrelated from the other variables by diagonalising the covariance matrix with iterative Jacobian rotations [136]. These distributions are then flattened using the cumulative distribution and binned such that each bin contributes the same statistical power. The flattened distributions are standardised by taking the normalised fraction in each bin and transforming it such that the output

Table 6.8: A summary of all the event variables which are used in the neural network in one or more analysis regions [13].

Variable	Definition
$\Delta\eta_{jj}^{\max}$	Maximum $\Delta\eta$ between two jets.
$m_{bb}^{\min\Delta R}$	Invariant mass of the two $b$ -jets with the minimum $\Delta R$ .
$m_{bb}$	Invariant mass of the two $b$ -jets selected as the Higgs candidate.
$\Delta R_{H\text{lept}}^{\min}$	Minimum $\Delta R$ between the Higgs candidate and the two leptons.
$n_{\text{Higgs}}^{30}$	Number of Higgs candidates within $30 \text{ GeV}/c^2$ of $125 \text{ GeV}/c^2$ .
$\Delta R_{bb}^{\max p_T}$	$\Delta R$ of the $b$ -jet pair with the maximum $p_T$ .
Aplanarity <sub>jet</sub>	1.5 of the second eigenvalue of the jet momentum tensor.
$m_{jj}^{\min}$	Minimum invariant dijet mass.
$\Delta R_{H\text{lept}}^{\max}$	Maximum $\Delta R$ between the Higgs candidate and the two leptons.
$m_{jj}^{\text{closest}}$	The invariant dijet mass which is closest to $125 \text{ GeV}/c^2$ .
$H_T$	The scalar sum of the jet $p_T$ and lepton $p_T$ .
$\Delta R_{bb}^{\max m}$	$\Delta R$ between the pair of $b$ -jets with the maximum invariant mass.
$\Delta R_{lj}^{\min\Delta R}$	$\Delta R$ between the lepton and jet with the minimum $\Delta R$ .
$H_4$	Fifth Fox-Wolfram moment calculated using all jets and leptons.
$p_T^{\text{jet3}}$	$p_T$ of the third $p_T$ -ordered jet.
Centrality	Sum of $p_T$ divided by the sum of $E$ for all jets and leptons.
$m_{jj}^{\max p_T}$	The invariant mass of the dijet pair with the maximum $p_T$ .

distribution has a mean of zero and a variance of one using the transformation,

$$x'_i = \frac{x_i - E[x_i]}{\sqrt{\text{Var}(x_i)}}. \quad (6.6)$$

An example of the preprocessing of variables in NeuroBayes is shown in Figure 6.9 for one of the neural network input variables in the  $\geq 4$  jets,  $\geq 4$   $b$ -jets region.

NeuroBayes uses a three layered neural network with one input layer, one hidden layer and one output layer. It is a forward-feeding network with back-propagation of errors. For the case of classification, the input layer has a number of nodes equal to the number of input variables plus one. The hidden layer has one additional node and the output layer consists of a single output node. The additional node in the input layer is a bias node which contains the target value that the neural network is training to achieve. The inclusion of this node improves the training of the neural network.

The nodes in a neural network are connected to all the nodes in a layer either side of the net-

Table 6.9: The variables rankings provided by NeuroBayes for each of the top ten selected variables in each region where a neural network is trained [13].

Variable	$\geq 4$ jets, $\geq 4$ $b$ -jets	$\geq 4$ jets, 3 $b$ -jets	3 jets, 3 $b$ -jets
$\Delta\eta_{jj}^{\max}$	1	1	1
$m_{bb}^{\min\Delta R}$	2	8	-
$m_{bb}$	3	-	-
$\Delta R_{H\text{lept}}^{\min}$	4	5	-
$n_{\text{Higgs}}^{30}$	5	2	5
$\Delta R_{bb}^{\max p_T}$	6	4	8
Aplanarity <sub>jet</sub>	7	7	-
$m_{jj}^{\min}$	8	3	2
$\Delta R_{H\text{lept}}^{\max}$	9	-	-
$m_{jj}^{\text{closest}}$	10	-	10
$H_T$	-	6	-
$\Delta R_{bb}^{\max m}$	-	9	-
$\Delta R_{lj}^{\min\Delta R}$	-	10	-
$H_4$	-	-	4
$p_T^{\text{jet3}}$	-	-	6
Centrality	-	-	7
$m_{jj}^{\max p_T}$	-	-	10

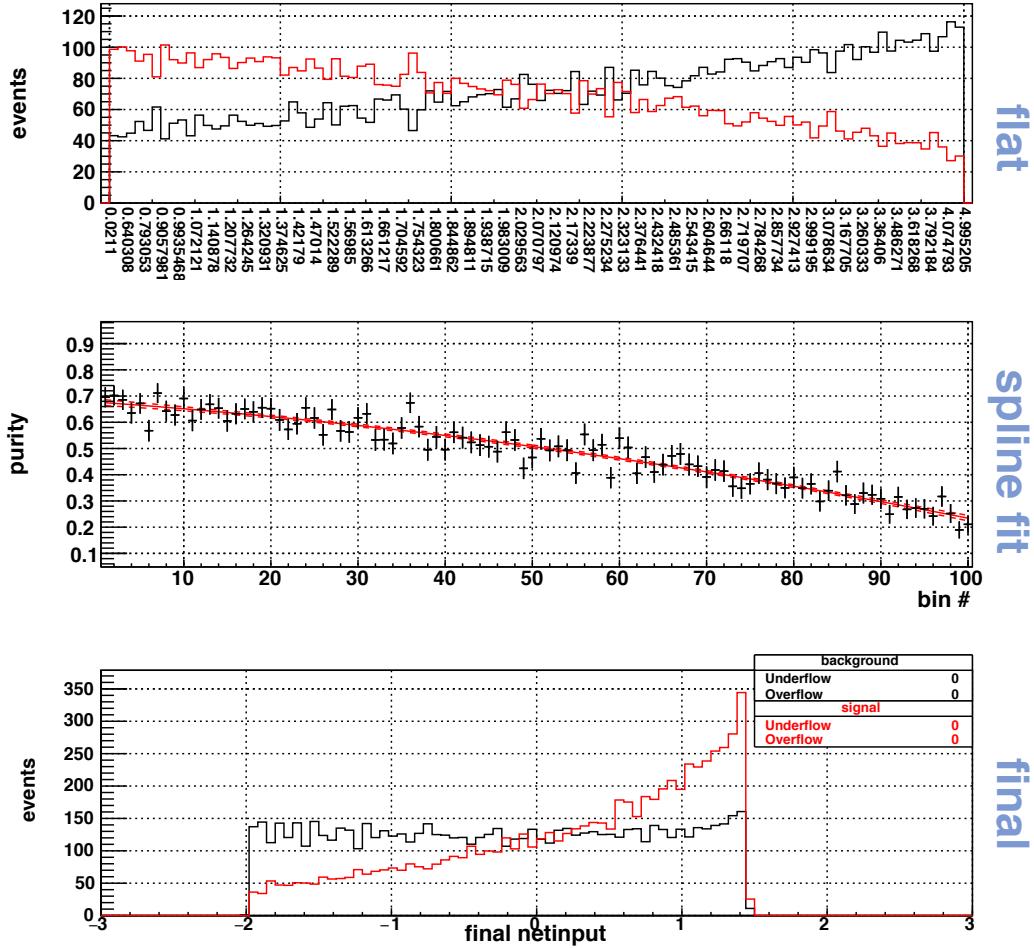


Figure 6.9: Distribution information produced by NeuroBayes analysis macro for the  $\Delta\eta_{jj}^{\max}$  variable in the  $\geq 4$  jets,  $\geq 4$   $b$ -jets region [135]. The distribution labelled flat shows the variable after the pre-processing to normalise, decorrelate, transform to a Gaussian and flatten. The distribution labelled spline fit shows a cubic spline fit through the purity of the flattened distribution, defined by the number of signal events divided by the total signal and background events per bin. The distribution labelled final is the variable after all preprocessing and standardisation. This distribution is input into the first layer of the neural network.

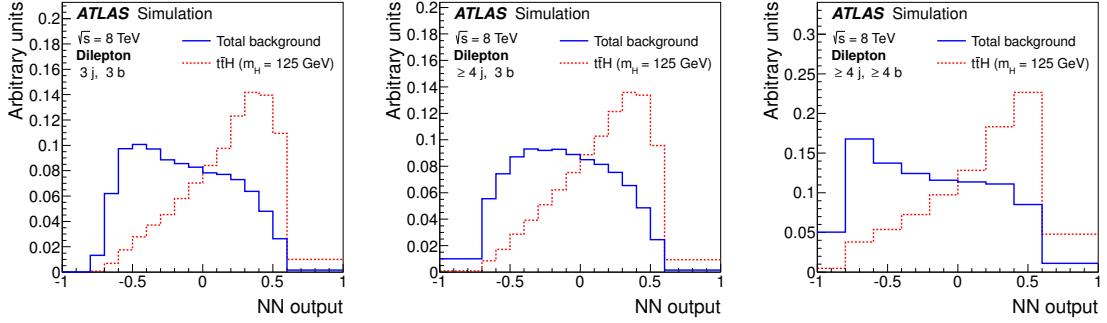


Figure 6.10: The discrimination power between signal and background processes after training a neural network in the 3 jets, 3  $b$ -jets,  $\geq 4$  jets, 3  $b$ -jets and  $\geq 4$  jets,  $\geq 4$   $b$ -jets analysis regions [13].

works structure. Each connection carries a weight and the training of a neural network amounts to the variation of these weights in order to reduce an error function related to the targeted output value. The value of a node in a network in a single iteration has the form,

$$x_j = f\left(\left[\sum_i w_{ij}^{(n)} x_i^{(n-1)}\right] + \mu_j^{(n)}\right), \quad (6.7)$$

where  $n$  is the layer number,  $w_{ij}^{(n)}$  is the weight of a connection between node  $j$  in layer  $(n)$  and node  $i$  in the previous layer,  $(n-1)$ , and  $\mu_j^{(n)}$  is a threshold. The function  $f$  corresponds to a finite mapping from  $[-\infty, +\infty] \rightarrow [-1, 1]$  where NeuroBayes applies a sigmoid function.

The Bayesian aspect is implemented in the regularisation of the network connection weights. The aim is to use the inclusive distributions from the training data to define a prior on the variable which should then be used to direct the update of weights to prevent the neural network providing an output which could be interpreted as unphysical. This regularisation is applied internally by controlling the decay of weights between the input and hidden layer, the bias node and the hidden layer and the hidden layer and the output layer. A more detailed discussion is beyond the scope of this thesis but additional details can be found in Reference [136, 137].

The final separation power of the neural network is larger than the separation of any individual input variable and can be observed in Figure 6.10.

## 6.8 Analysis Distributions

The analysis likelihood described in Section 6.1 requires a distribution for each of the six analysis regions which have a prediction from simulation for the number of signal and background

events in each histogram bin which is then compared to the observed number of events in data. The event corrections described in Section 6.5 and the event weights provided by the ATLAS Combined Performance Groups have ensured that there is good agreement between data and simulation in the regions which have little or no sensitivity to a  $t\bar{t}H$  signal process. The final distributions which are used in the likelihood are shown in Figure 6.11 prior to fitting the likelihood function. Details of the systematic uncertainties applied in this analysis are documented in Reference [13]. Each source of uncertainty provide a new prediction for the number of events from each simulated process which are treated as a relative shift from the nominal expected values. The systematic uncertainties on the simulated data samples have been added in quadrature with the statistical uncertainties and are shown as a hashed area around the nominal prediction. The integral over these distributions correspond to the figures in Table 6.7. However, with these distributions one can observe the individual histogram bins and make the connection between the event selection and the values of  $s_i$ ,  $b_i$  and  $n_i$  which are required to construct the likelihood function.

## 6.9 Results

As a discovery is not possible in this analysis channel without more data, upper limits are set on the  $t\bar{t}H$  production cross-section relative to the Standard Model expectation at a 95% confidence level. A simultaneous fit is performed across all the analysis regions using the distributions indicated in Table 6.6 using the RooStats framework [120]. A likelihood function is created from the product of Poisson likelihoods modelling the expected number of events in each bin of the input histograms with Gaussian or Gamma constraints provided for systematic uncertainties. Distributions are binned to ensure shape information persists into the likelihood evaluation without being artefacts of statistical fluctuations. Systematic uncertainties are evaluated at  $\pm 1\sigma$  with auxiliary measurements and are processed to disentangle normalisation effects from shape effects. Systematic uncertainty histograms are normalised to have the same integral as the nominal histogram and a variation in normalisation from a systematic uncertainty enters into the fit as a pure normalisation shift term. These histograms are then smoothed and any systematic uncertainties which have less than 0.5% variation with respect to the nominal histogram are removed. The final systematic uncertainties are interpolated to provide a continuous parametrisation of the

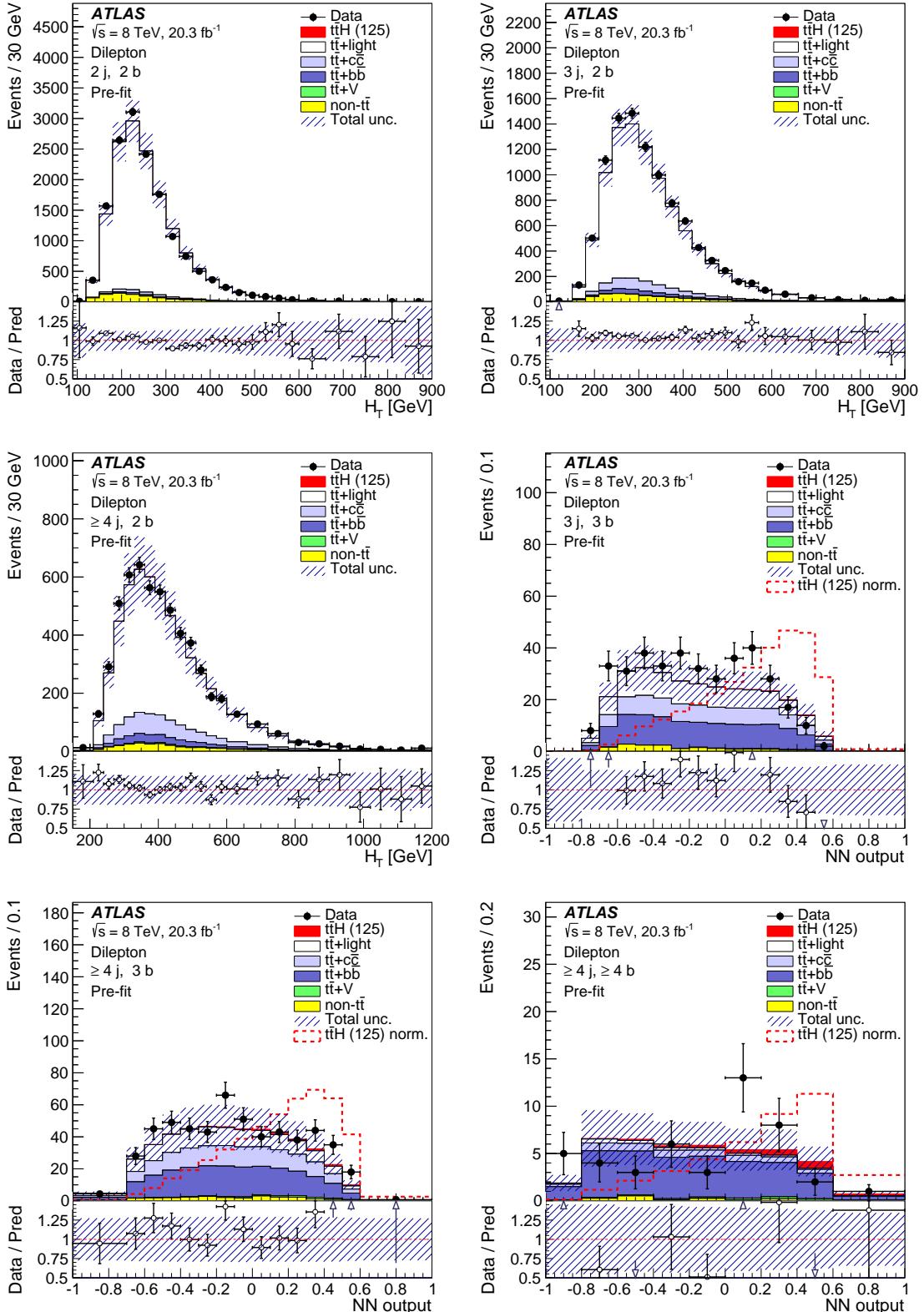


Figure 6.11: The prefit distributions for each analysis region [13]. The signal-rich regions use a neural network classifier trained to separate signal and background events. The remaining control regions use the  $H_T$  distribution to constrain the errors in the signal-rich regions during the simultaneous fit of the full likelihood function. The systematic uncertainties have been added in quadrature with the statistical uncertainties to provide an uncertainty range about the nominal prediction. The statistical error bars shown on the data points are given as  $\sqrt{n_i}$ .

expected number of events as a function of the nuisance parameters ( $s_i(\boldsymbol{\theta})$  and  $b_i(\boldsymbol{\theta})$ ).

The likelihood is described as  $\mathcal{L}(\mu, \boldsymbol{\theta})$  where  $\mu$  is the parameter of interest (signal strength) which quantifies the signal contribution relative to the Standard Model expectation and  $\boldsymbol{\theta}$  represents the set of nuisance parameters which provide uncertainty to the model. This likelihood function is profiled to use the data to fit nuisance parameters to their maximum likelihood estimators.

A constrained version of the profiled likelihood refers to fitting the nuisance parameters to data for a fixed value of  $\mu$ . The profiled set of nuisance parameters is defined as  $\hat{\boldsymbol{\theta}}_\mu$  and corresponds to the values that minimise the difference between data and prediction by varying the number of expected events across all regions. The unconstrained version of the profiled likelihood fits all parameters to data. In this case, both  $\mu$  and  $\boldsymbol{\theta}$  are fitted to the maximum likelihood estimators  $\hat{\mu}$  and  $\hat{\boldsymbol{\theta}}$ . The correlation matrix of the post-fit nuisance parameters allows the data to constrain the impact of systematic uncertainties.

The test statistic used in the analysis to set an upper limit on the parameter of interest is the profiled-likelihood ratio [138],

$$q_\mu = \begin{cases} -2\ln\left(\frac{\mathcal{L}(\mu, \hat{\boldsymbol{\theta}}_\mu)}{\mathcal{L}(\hat{\mu}, \boldsymbol{\theta})}\right) & \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu. \end{cases} \quad (6.8)$$

This variable is dependent on the unconditional likelihood fit which determines  $\hat{\mu}$ . In the case that  $\hat{\mu}$  is greater than the tested  $\mu$  hypothesis, one does not treat the data as less compatible with the hypothesis. The numerator of this variable is the profiled likelihood conditional on the value of  $\mu$  being tested. The denominator is the unconditional maximum likelihood where  $\mu$  and  $\boldsymbol{\theta}$  are simultaneously fitted to data.

An upper limit is set by constructing the distribution of  $q_\mu$  under different  $\mu$  hypotheses. The CLs method is used to identify the values of  $\mu$  which are excluded at a 95% confidence level [139]. In the limit of a large amount of expected background, the distribution of  $f(q_\mu|\mu)$  shown in Figure 6.12 does not vary for different hypotheses of  $\mu$ . However, the observed value of  $q_{\mu,\text{obs}}$  will move in this figure, moving rightwards as data becomes less compatible with the  $\mu$  hypothesis. The CLs method determines the 95% exclusion when the value of  $\text{CL}_\mu$  is less than 5% of  $\text{CL}_b$ . It can be seen that to evaluate the upper limit, an understanding of the distribution of  $q_\mu$  is required under the background-only hypothesis. Expected limits can be set in a similar way,

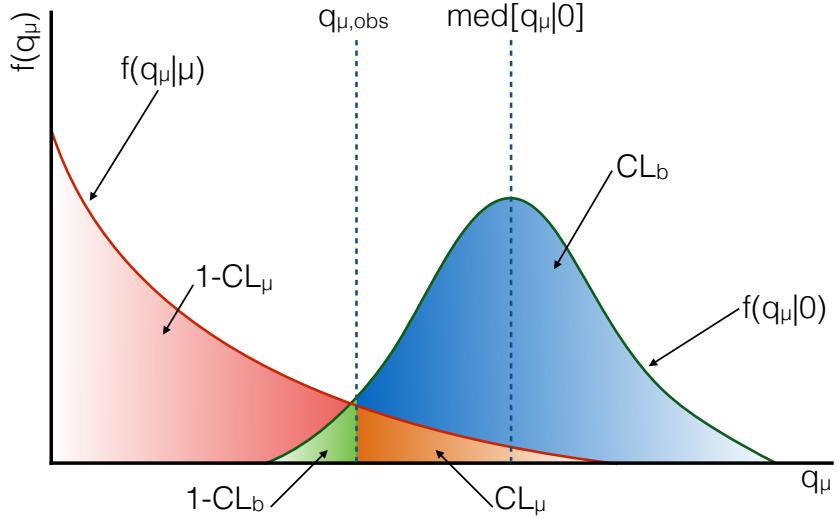


Figure 6.12: Illustration of the test statistic distribution and the different areas associated with an observation when comparing to two hypotheses.

but rather than evaluating on the observed data, one uses the median of  $q_\mu$  under the assumption of the background-only hypothesis.

The expected performance of an analysis can be evaluated with asymptotic approximations using the Asimov dataset which is appropriate for the desired test [138]. If one wishes to evaluate the expected upper limits on the parameter of interest, the dataset is constructed from the background-only hypothesis, where the expected number of events in each bin is equal to the background prediction. In this case the nuisance parameters are profiled to this dataset so no constraints are included from the data. If one wishes to evaluate the expected discovery significance of an analysis then the variable,

$$q_0 = \begin{cases} -2\ln\left(\frac{\mathcal{L}(0, \hat{\theta}_0)}{\mathcal{L}(\hat{\mu}, \hat{\theta})}\right) & \hat{\mu} \geq 0, \\ 0 & \hat{\mu} < 0, \end{cases} \quad (6.9)$$

is calculated using an Asimov dataset which corresponds to the signal-plus-background hypothesis. In this case the number of expected events in each bin is equal to the prediction for a Standard Model signal ( $\mu = 1$ ) plus the background prediction. The discovery significance has a one-to-one mapping with the  $p$ -value using  $Z = \Phi^{-1}(1 - p)$  and it can be shown to approxi-

mately equal  $\sqrt{q_0}$  for this Asimov dataset [138].

Performing an unconditional maximum likelihood fit results in a best-fit value for  $\hat{\mu}$  of  $2.8 \pm 2.0$ . This indicates that the observed discovery significance is approximately  $Z = 1.4$ . An upper limit is determined using  $q_\mu$  and profiling the data. The background-only hypothesis can be rejected if the data contains something which appears to be signal-like. A signal strength of 4.1 is expected to be excluded if there is no signal present in the data and a signal strength of 4.7 is expected to be excluded if a Standard Model Higgs boson exists. The observed exclusion is for a signal strength of 6.7 which indicates an excess of signal-like events in the data. To be sensitive to a Standard Model signal process and therefore potentially discover it, the expected background-only exclusion needs to be sensitive to a signal strength of 1.0. This has not been possible with the Run 1 data. The full set of upper limits is presented in Table 6.10 with the associated 68% and 95% confidence interval bands.

Table 6.10: The observed and median expected background-only 95% confidence level upper limits on  $\sigma(t\bar{t}H)$  relative to the Standard Model expectations with the 68% and 95% interval bands around the background-only expectation [13]. The expected median 95% confidence level upper limit is provided for the Standard Model expectation including the  $t\bar{t}H$  signal process assuming a Higgs mass of 125  $\text{GeV}/c^2$ .

Observed	$-2\sigma$	$-1\sigma$	Median	$+1\sigma$	$+2\sigma$	Median ( $\mu = 1$ )
6.7	2.2	3.0	4.1	5.8	7.7	4.7

## 6.10 Outlook

Whilst it is not possible to discover or refute the existence of the  $t\bar{t}H$  process using the Run 1 data, there is currently an excess of data which could be interpreted as signal-like compared to the amount of data one would expect to see if the Standard Model process exists. The probability of obtaining a result which is at least as signal-like as the one observed if there is no Standard Model  $t\bar{t}H$  process is given by the  $p$ -value. The observed significance of the signal process is 1.4 in the dilepton channel which corresponds to a  $p$ -value of 8%.

The following chapter will document an improved signal discrimination method, known as the matrix element method, which could be added to the analysis in Run 2 and was successfully used in the Run 1 lepton plus jets analysis. In order to observe the signal process in the dilepton channel, this analysis needs to extract all the possible information out of the limited data whilst improving the understanding of the signal and background processes.

The final chapter in this thesis examines the projected sensitivity improvement one could expect in Run 2 by applying this new discrimination technique and also presents a couple additional studies which could be examined with respect to the matrix element technique.

Given how sophisticated the  $t\bar{t}H$  analysis has become within the ATLAS Collaboration through Run 1, it seems that if the  $t\bar{t}H$  production mechanism exists, it will likely be discovered in Run 2 through the combination of top-quark decay channels and Higgs decay channels.

## Chapter 7

# The Matrix Element Method

---

*“The world is full of magical things patiently waiting for our wits to grow sharper.”*

— Bertrand Russell

---

This chapter documents the development and application of the matrix element method analysis technique by the author to the 8 TeV dilepton  $t\bar{t}H$  analysis which has been described in Chapter 6. This technique has been used to good effect in a number of Tevatron analyses to extract a measurement of the top quark mass [140–144]. More recently it has been applied to Higgs searches by the ATLAS and CMS Collaborations [57, 145, 146] and in particular it was applied to the 8 TeV single lepton  $t\bar{t}H (H \rightarrow b\bar{b})$  analysis [13].

As this analysis channel suffers from an overwhelming and irreducible background from  $t\bar{t}+b\bar{b}$ , the aim of applying the matrix element method has been to improve the discrimination between signal and background in the most sensitive analysis regions. The mathematics of the matrix element method are presented in Section 7.1, followed by technical details of the calculation in Section 7.2, integration approximations in Section 7.3 and calculation of the phase space volume element in Section 7.4. A discussion of statistical measures for comparing the performance of different integration strategies is presented in Section 7.5 which are subsequently used to compare the expected performance of the matrix element method in Section 7.6. Validation of the different integration strategies are performed using signal depleted control regions in Sec-

tion 7.7. A study using the matrix element method to perform event reconstruction is presented in Section 7.8 by investigating the efficiency to correctly identify the Higgs candidate decay products. The impact on the  $t\bar{t}H$  analysis presented in Chapter 6 is studied in Section 7.10 by training the neural network in the signal regions with the inclusion of a matrix element method likelihood variable. The expected discovery significance, both with and without the inclusion of the matrix element method, is calculated using a likelihood fit with a subset of systematic uncertainties in Section 7.11. Results are presented in Section 7.12 using the 8 TeV dataset to calculate the observed discovery significance and upper limits on the scaling of the  $t\bar{t}H$  cross-section with respect to the Standard Model expectation. A final validation of the matrix element method is performed by checking the neural network distributions after the fit is performed in Section 7.13. Two events which have been identified in data to be strongly signal-like are shown with their event display in Section 7.14. The final result of this chapter is to recommend that this method be employed in the upcoming 13 TeV  $t\bar{t}H$  analyses.

## 7.1 Mathematical Description

The matrix element method is a tool to link theoretical predictions with experimental observations and construct an optimal test statistic for discrimination between two classes of events.

An event will have a set of observable final state momentum vectors  $\mathbf{x}$  (e.g. jet momenta), and an associated set of unobservable final state partonic momentum vectors relating to the physics process being considered  $\mathbf{y}$  (e.g. parton momenta). Detector and physics effects prevents a direct association between a measured quantity and the truth-level value. These two sets of variables are used to characterise an individual event. The study presented in this chapter uses the  $t\bar{t}H$  final state described in Chapter 6. Therefore the set of observable objects,  $\mathbf{x}$ , contains seven momentum vectors from four  $b$ -jets, two leptons and missing transverse energy which results in 20 independent quantities. The unobservable set  $\mathbf{y}$  contains eight momentum vectors from four  $b$ -quarks, two leptons and two neutrinos which results in 24 independent values.

The joint probability density function for an event,  $\mathbf{x}$ , under the assumption that it was generated by a hypothesised process,  $H$ , is given as  $P(\mathbf{x}|H, \boldsymbol{\alpha}) = \int f(\mathbf{x}, \mathbf{y}|H, \boldsymbol{\alpha}) d\mathbf{y}$ . The hypothesis,  $H$ , will have a set of model parameters  $\boldsymbol{\alpha}$  which may be fixed between different hypotheses or may vary. An example of varying model parameters could be tests for different hypothesised

Higgs masses. However for the purposes of this chapter, the set of model parameters are taken to be the Standard Model expectation values as defined in **MADGRAPH 5** [126] with a Higgs mass of  $125 \text{ GeV}/c^2$  and are therefore constant for the hypotheses considered.

Given an event, the matrix element method determines the probability of the data under a particular hypothesis. Any theoretically described leading-order process can be used as a process hypothesis. The use of higher-order diagrams introduces additional complexity which currently makes their inclusion unfeasible without specialised tools [147, 148]. The Neyman-Pearson lemma explains that the ratio of likelihoods evaluated for two different nested hypotheses will define the optimum discriminating test statistic [149]. One can therefore define a likelihood ratio of the form,

$$t_{\text{LR}} = \frac{P(\mathbf{x}|H_s, \boldsymbol{\alpha})}{P(\mathbf{x}|H_b, \boldsymbol{\alpha})}, \quad (7.1)$$

to create a test statistic to discriminate between a signal hypothesis ( $H_s$ ) and a background hypothesis ( $H_b$ ). A perfect likelihood ratio will approximate the ratio of transition matrix elements. In this analysis, the signal hypothesis was defined as dileptonic  $t\bar{t}H$  ( $H \rightarrow b\bar{b}$ ) production and a background hypothesis was defined as dileptonic  $t\bar{t}+b\bar{b}$  production.

The matrix element method likelihood is a complex function which can be separated into four distinct components: normalisation, parton density functions, differential cross-section and transfer functions. An integration is performed over the final state partonic phase-space which convolves functions from the transition matrix element with conditional probability distribution functions parameterising the detector response for different objects. Energy-momentum conservation between the initial and final state parton momentum-vectors is convolved with parton density functions to give probabilities dependent on the initial parton flavour of the tested hypothesis. The likelihood evaluated using the matrix element method is defined as

$$P(\mathbf{x}|H, \boldsymbol{\alpha}) = \frac{(2\pi)^4}{\sigma_i^{\exp(\boldsymbol{\alpha})}} \int dp_A(\mathbf{y})dp_B(\mathbf{y})f(p_A)f(p_B) \frac{|\mathcal{M}(\mathbf{y}|H, \boldsymbol{\alpha})|^2}{\mathcal{F}} \mathcal{W}(\mathbf{x}|\mathbf{y})d\Phi_N(\mathbf{y}), \quad (7.2)$$

where  $H$  represents the hypothesis being tested.

The integration is performed over the outgoing parton phase-space which is defined by the phase space volume element ( $d\Phi_N(\mathbf{y})$ ). The integration is also performed over the incoming parton phase-space,  $dp_A$  and  $dp_B$ , which will have a dependence on the outgoing parton momentum-vectors as the integration is subject to an implicit delta function,  $\delta^{(4)}(p_A + p_B - \sum_i^N \mathbf{y}_i)$ ,

in order to conserve the energy and momentum of the system. The parton density functions ( $f(p_A), f(p_B)$ ) describe the probability of a production mechanism interaction. In this analysis the production mechanism is required to be  $gg$  or  $q\bar{q}$ . The squared transition matrix element ( $|\mathcal{M}(\mathbf{y}|\boldsymbol{\alpha})|^2$ ) calculates the probability of the outgoing parton momenta being produced by a hypothesised process. A flux factor ( $\mathcal{F}$ ) is associated to the incoming parton momenta and is required to keep the total integrated volume Lorentz invariant. The transfer functions ( $\mathcal{W}(\mathbf{x}|\mathbf{y})$ ) provide the conditional probability of the outgoing reconstructed final state ( $\mathbf{x}$ ) given a parton-level final state ( $\mathbf{y}$ ). This provides the link between the observed physics and the theorised physics which are brought together in a single discriminating variable with the matrix element method. The term  $\sigma_i^{\text{exp}(\boldsymbol{\alpha})}$  is a normalisation factor combining the theoretical cross-section for the process considered with the detector acceptance and event selection efficiency. This term is required to ensure that  $\int P(\mathbf{x}|H, \boldsymbol{\alpha}) d\mathbf{x} = 1$ . However it corresponds to a constant factor which is neglected in the following work as the emphasis is on event-by-event discrimination with the ratio of probabilities, rather than parameter estimation from the interpretation of an experiment likelihood.

## 7.2 Implementation

The calculation of  $P(\mathbf{x}|H, \boldsymbol{\alpha})$  in Equation (7.2) integrates over the parton-level phase-space,  $d\Phi_N(\mathbf{y})$ . This integration is intrinsically complicated due to the high-dimensionality of the integral and needs to be evaluated numerically which can be time-consuming and computationally intensive. The following tools are used to implement the matrix element method calculation in an efficient manner.

The matrix element method implemented in this analysis uses a framework called MEMTOOL, which was originally developed at the University of Göttingen [150]. The framework is written in C++ and implements a modular class structure to allow different processes and process-specific kinematic treatments to be implemented and provides a shared core code base for the Monte Carlo integration.

The integration in MEMTOOL uses the GSL 1.16 version of VEGAS [151–153]. VEGAS is an adaptive importance-sampling numerical integrator which uses previous integration iterations to direct the region of phase-space for the next iteration of integration.

For an integral defined as a function  $f(x)$ , the integrand is divided into  $N$  regions of  $f(x_i)$  each initially with an equal width,  $\Delta x_i$ . After an iteration, each region is divided into  $m_i + 1$  subregions where

$$m_i = \kappa \frac{\overline{f(x_i)} \Delta x_i}{\sum_j f(x_j) \Delta x_j}, \quad (7.3)$$

$\overline{f(x_i)}$  is the contribution to the total integrand in region  $i$  and  $\kappa$  is a large constant which will result in a region being divided up into a maximum of  $\kappa + 1$  subregions. After this subdivision, the subregions are merged back into  $N$  total regions such that each region contains a constant number of subregions. This process is then iterated to perform the integration. This adaptive procedure adjusts the size sampling regions such that the regions which contribute the most to the total integrand have the smallest width to concentrate the area being sampled [151].

The version of VEGAS used provides a measure of uncertainty in the integration from the variance and provides a  $\chi^2$  measure of convergence. Both of these measures are used to determine when to truncate successive integration iterations. The  $\chi^2$  measure calculates the integrand value for an iteration and compares it to the total integrand value for all iterations. This value is expected to be close to 1 if the integration has correctly adapted to the phase-space and is converging [153]. The integration is stopped if the uncertainty in the integration is less than 10% and  $|\chi^2 - 1|$  is less than 0.5. If the uncertainty is larger than 10% then the number of samples integrated is doubled and an additional iteration is run. If the uncertainty is at an acceptable level but the  $\chi^2$  has not converged, then the integration is run for an additional iteration with the same number of samples. A sample size of 2048 has been chosen for this analysis for which studies have shown to be an acceptable balance of quick convergence without requiring excessive doubling of the sample size during run time to reduce the uncertainty.

The transition matrix elements for a hypothesis are generated as standalone C++ code by MADGRAPH 5 [126]. These transition matrix elements are leading-order. State-of-the art developments [147, 148] are taking place to extend the matrix element method to higher-order calculations but these are complicated by the increased number of diagrams and virtual corrections which need to be taken into account when performing numerical integration. There are 10 diagrams considered in the signal hypothesis and 37 background diagrams considered in the background hypothesis. These are a mix of gluon-gluon fusion and quark-antiquark annihilation processes and all are taken into account during the calculation.

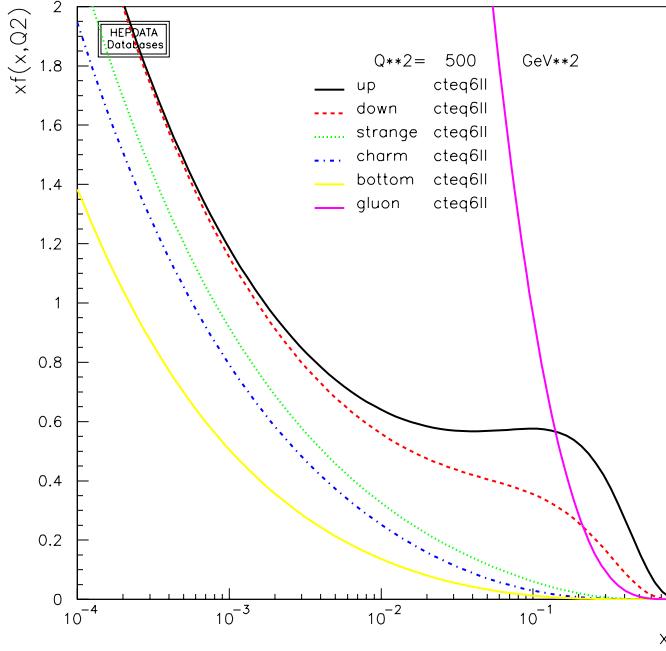


Figure 7.1: CTEQ6LL parton density function for  $Q^2 = 500 \text{ GeV}^2$  [155].

The parton density functions are provided by the LHADPF PDF library [154]. The CTEQ6LL PDF is applied in this analysis which performs a leading-order fit with a leading-order  $\alpha_s$  parameterisation [155]. The quark and gluon distributions from this PDF set are shown in Figure 7.1. The PDFs provide a constraint on the integration by evaluating the probability that the initial state partons can be produced with the energy required by the final state partons.

The detector response functions are derived for use in a kinematic fitting tool called the Kinematic Likelihood Fitter (KLFitter) which is used by top physics analyses in the ATLAS Collaboration to assign probabilities to jet-parton permutations [156]. The functions are derived such that they are independent of the likelihood method being employed which allows them to be used in the matrix element method without any hidden dependence.

Transfer functions are conditional probability density functions, shown in Equation (7.2) as  $\mathcal{W}(\mathbf{x}|\mathbf{y})$ , which parametrise the probability of having an observed kinematic variable,  $x_i$ , given a partonic kinematic value,  $y_i$ . This function can be factorised for each observable kinematic variable as  $\mathcal{W}(\mathbf{x}|\mathbf{y}) = \prod_i^N W(x_i|y_i)$ .

The parton shower and hadronisation effects which affect the jet reconstruction at the correct energy scale have been discussed in Section 4.4. However, despite these corrections, the energy

Table 7.1:  $B$ -jet transfer function parameters used in Equation (7.4) to produce Figure 7.2.

$p_1$	$p_2$	$p_3$	$p_4$	$p_5$
0.030	0.10	0.70	0.11	0.23

of a jet is not exactly the energy of the parton which initiated it. This is not as an important consideration for charged leptons as energy losses from bremsstrahlung and material interactions are taken into account when leptons are reconstructed. This reduces the uncertainty in the estimation of the true lepton energy from the reconstructed lepton. As such, the transfer functions for charged lepton energies can be approximated to be delta functions, but the jet and parton energy components of  $\mathbf{x}$  and  $\mathbf{y}$  need to include a correct measure of the conditional probability distribution.

The detector energy response for a jet is modelled using a double Gaussian as a function of the relative energy shift from truth to reconstructed level,

$$W(\Delta E) = \frac{1}{\sqrt{2\pi}(p_2 + p_3 p_5)} \left( e^{-\frac{(\Delta E - p_1)^2}{2p_2^2}} + p_3 e^{-\frac{(\Delta E - p_4)^2}{2p_5^2}} \right), \quad (7.4)$$

where  $\Delta E = (E_{\text{truth}} - E_{\text{reco}})/E_{\text{truth}}$  and  $p_i$  are fitted parameters with a dependence on the true energy. An example of the parameterisation of the detector response for a  $b$ -jet with a true energy of 100 GeV within the pseudorapidity range of  $0 \leq |\eta| < 0.8$  is shown in Figure 7.2 using the parameter values given in Table 7.1. It can be observed that the energy of a  $b$ -jet will more often provide an underestimation of the initiating parton energy due to effects such as out-of-cone radiation or semi-leptonic  $B$ -hadron decays where energy is lost through a neutrino.

### 7.3 Integration Approximations

The integration required in Equation (7.2) needs to be performed over the allowed phase-space of  $\mathbf{y}$  which results in a 24-dimensional numerical integration. To integrate such a large number of degrees of freedom is an unrealistic task, given the limited availability of computing resources and finite time. In order to reduce the computation time when evaluating the matrix element method likelihood, a number of approximations are implemented. These approximations either reduce the number of times the matrix element method calculation needs to be performed or they reduce the number of integration iterations required for the calculation to converge. The

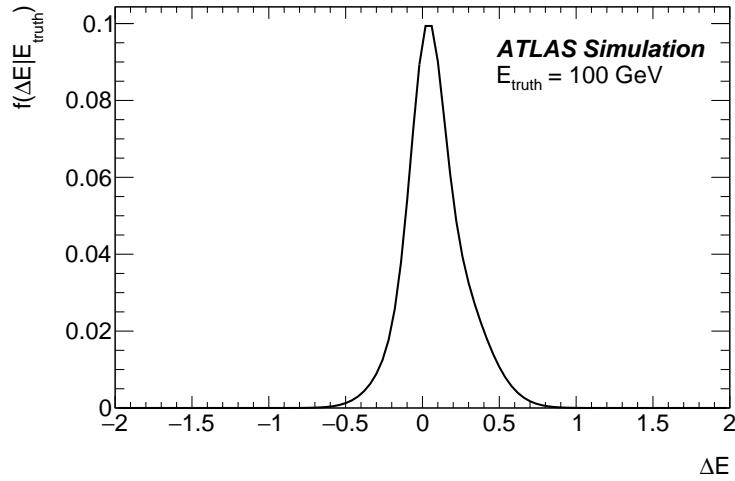


Figure 7.2: A factorised component of the full transfer function  $\mathcal{W}(\mathbf{x}|\mathbf{y})$  parametrised for a  $b$ -jet matched to a particle jet with a true energy of 100 GeV extracted from the KLFitter package [156]. This distribution shows that on average the true parton energy is greater than the reconstructed jet energy. This phenomena can be understood from QCD radiation which is lost outside the jet algorithm cone and from hadronic decays inside the jets which produce neutrinos.

approximations implemented in MEMTOOL include optimised treatment of particle helicities, reduction of possible permutation calculations, reduction of integration variables and limitation of integration ranges.

### 7.3.1 Helicity

For the process being considered, there are two initial state partons and there are eight final state partons. Each of these can have a helicity of  $\{+1, -1\}$  which results in 1024 different helicity combinations. The default matrix element code generated by MADGRAPH 5 evaluates the matrix element for every helicity combination. However, there is redundancy in this calculation for two reasons. There are some helicity combinations which are unphysical, and as such should not be evaluated, and there are helicity combinations which evaluate to exactly the same probability as other helicity combinations, and are therefore invariant.

To reduce the calculation time, the number of helicity states evaluated is reduced by removing these redundant calculations. After removing the unphysical states and the invariant pairings, the calculation also dynamically identifies the dominant integrand helicity states. The states which do not contribute very much are removed from the calculation so as to speed up the calculation. After 100 integration iterations, the helicity state with the maximum contribution

to the likelihood is identified. The other helicity states have their contribution measured relative to this maximum contribution. Any helicity state which contributes less than 10% of the leading helicity state's contribution to the likelihood is deemed to be a sub-leading contribution and is removed from the calculation.

These approximations to the helicity state evaluation reduces the computation time by around a factor of ten without a noticeable reduction in the separation power of the final likelihood.

### 7.3.2 Permutations

The function in Equation (7.2) has a dependence on the ordering of the set of reconstructed objects,  $\mathbf{x}$ . In the terminology used in this chapter, this means that Equation (7.2) is the calculation for a single permutation. To fully evaluate the matrix element method for a single event, all possible orderings are required to be considered within each group of final state objects that can be considered indistinguishable. For example, in the  $t\bar{t}H$  final state, this means that the calculation needs to be evaluated to allow each  $b$ -jet to be associated with each final state  $b$ -quark.

The calculation is made for each permutation and then averaged over to provide a final event likelihood because a priori, the correct assignment of reconstructed objects ( $\mathbf{x}$ ) and partons ( $\mathbf{y}$ ) is unknown. This means there are permutations for jets, neutrinos and leptons if the event has the same-flavour lepton pair. The outgoing partons are mapped to their reconstructed objects as follows:

- 4  $b$ -quarks  $\leftrightarrow$  4  $b$ -jets,
- 2 leptons  $\leftrightarrow$  2 leptons,
- 2 neutrinos  $\leftrightarrow E_T^{\text{miss}}$ .

It is possible to identify invariant permutations where switching two reconstructed objects will not change the final likelihood evaluated for the two permutations. This effect comes from the same region of phase-space dominating the integral which will be identified during the integration. As the system contains two top quarks decaying, there is invariance to be considered by switching the decay products of the two tops. There is also invariance in the decay of the Higgs boson to two  $b$ -quarks.

The maximum number of permutations which could be considered in a dileptonic  $t\bar{t}H$  ( $H \rightarrow b\bar{b}$ ) event is 96. The four jet permutations contribute a factor of 24 and the lepton and neutrino permutations contribute two factors of 2. This number of permutations can be reduced by accounting for the calculation invariance with certain permutations.

The Higgs decay products are an invariant pairing. The leptons can be considered an invariant pairing to prevent the case of permuting all the top decay products such that the same state is being evaluated. The lepton charge is not used in this work to distinguish between the top quark and the anti-top quark. The neutrinos are a special case in the dilepton analysis because individually they are unobservable. The only measurement is from the missing transverse energy which can constrain the summed transverse momenta of the neutrinos. The kinematic integration strategy used will need to treat the neutrinos in such a way that there is no reason to permute their objects. Identifying these invariant pairs reduces the number of evaluated permutations by a factor of 8 which results in 12 permutations coming entirely from jet associations. A summary of the object association between  $\mathbf{x}$  and  $\mathbf{y}$  for each permutation is presented in Table 7.2.

Each permutation will provide a different contribution to the final likelihood. In a similar way to the helicity state, there are diminishing returns for calculating the likelihood for permutations which contribute a small amount to the total likelihood. The physical interpretation of these permutations is that the matrix element or the transfer functions do not find the assignment to be a likely hypothesis. In order to capitalise on this, the permutations are first ranked in the order of their contribution to the total likelihood by using the initial integration warm-up run which is used to prepare the integration grid. Each permutation is evaluated in order from highest contribution to lowest contribution. If after the first set of integration iterations the permutation likelihood is less than a factor of  $10^{-5}$  smaller than the total likelihood for the permutations evaluated at that point, then the integration is stopped for that permutation.

### 7.3.3 Kinematic Assumptions

As has been previously discussed, there are eight final state objects in the  $t\bar{t}H$  leading-order diagram represented by the set  $\mathbf{y}$ . Each parton has four degrees of freedom represented by the energy and three-momentum. Every parton considered has a well-defined mass which removed one degree of freedom in the matrix element method integration which reduces the integration in a 32-dimensional momentum-space to a 24-dimensional phase-space. Whilst it is possible to

Table 7.2: The association between detector-level and parton-level objects for each permutation after all reductions in the text have been applied is summarised for an  $ee$  or  $\mu\mu$  final state. The numbers represent the ordering of each object type within the set of reconstructed objects,  $\mathbf{x}$ . The ordering of the partons in  $\mathbf{y}$  is kept constant and the association with the  $t\bar{t}H$  decay is shown. The neutrinos are not listed as the only reconstructed information about them is the two missing transverse energy components and as such require special treatment which is expanded upon in the next section.

Permutation	Top-decay 1			Top-decay 2			Higgs-decay	
	lepton	neutrino	$b$ -quark	lepton	neutrino	$b$ -quark	$b$ -quark	$b$ -quark
1	1	-	1	2	-	2	3	4
2	1	-	1	2	-	3	2	4
3	1	-	1	2	-	4	3	2
4	1	-	2	2	-	1	3	4
5	1	-	2	2	-	3	1	4
6	1	-	2	2	-	4	3	1
7	1	-	3	2	-	2	1	4
8	1	-	4	2	-	2	3	1
9	1	-	3	2	-	1	2	4
10	1	-	4	2	-	1	3	2
11	1	-	4	2	-	3	2	1
12	1	-	3	2	-	4	1	2

evaluate a high-dimensional integral with numerical integration methods, the time to evaluate on a per-event basis would be too large to make this analysis method feasible. A series of approximations which are motivated by detector performance are made to reduce the number of degrees of freedom to a more manageable size.

It is assumed that the angular resolution of objects measured with the ATLAS detector is small enough that the reconstructed values of  $\eta$  and  $\phi$  are close approximations of the true parton kinematics. This approximation is applied to jets, electrons and muons and corresponds to assuming that the factorised transfer function responsible for this connection is represented by a delta function,

$$W(x_i|y_i) = \delta^{(1)}(x_i - y_i). \quad (7.5)$$

It is assumed that the energy resolution of leptons measured with the ATLAS detector is small enough that the energy is a close approximation of the partonic energy. The lepton energy resolution is typically less than 3% for the  $p_T$  ranges dealt with in this analysis [80, 157]. Lepton energies are treated in the same way as the angular variables by using a delta function as the lepton energy component of the transfer function. Jet energies are not treated with the same approximation as it is known that the detector response varies for different jet flavours and

different jet clustering algorithms.

As the transition matrix elements being used are leading-order, the initial state partons are considered to have zero transverse momenta prior to collision. The initial state momentum is constrained to balance with the final state momentum. This approximation allows the initial state parton momentum-vectors to be dependent on the set of final state parton momentum-vectors,  $\mathbf{y}$ . This is equivalent to integrating over the initial parton momenta and constraining with a delta function,

$$\delta^{(4)}(p_A + p_B + \sum^N y_i). \quad (7.6)$$

The initial state momentum is calculated using the vector sum of the final state momentum vectors and considering the initial state partons in the frame where  $p_T = 0$ . Using the parton momentum fraction Bjorken- $x$  in this frame, the following two relations are derived in order to calculate the  $z$ -momentum distributions between the two initial state partons:

$$x_A + x_B = \frac{2}{\sqrt{s}} \sum^y E_i, \quad (7.7)$$

$$x_A - x_B = \frac{2}{\sqrt{s}} \sum^y p_i^z. \quad (7.8)$$

These momentum fractions are then used to define the momentum-vectors of the initial state partons,

$$p_A = \left( \frac{x_A \sqrt{s}}{2}, 0, 0, \frac{x_A \sqrt{s}}{2} \right), \quad (7.9)$$

$$p_B = \left( \frac{x_B \sqrt{s}}{2}, 0, 0, \frac{-x_B \sqrt{s}}{2} \right). \quad (7.10)$$

There will inevitably be a  $p_T$  imbalance with the reconstructed objects coming from additional radiation in the event beyond a leading-order hypothesis. This will lead to a slow convergence and a less powerful discriminating test statistic when using the initial state parton momentum-vectors defined here. To account for a  $p_T$  imbalance, the initial state partons are boosted into a frame using the total  $p_T$  of the final state parton system so that any  $p_T$  imbalance is reflected with a  $p_T$  imbalance in the initial state partons. The effect of these transformations is to ensure that the implicit delta function described in Equation (7.6) is implemented whilst accounting for the expected  $z$ -momentum of two partons interacting without any transverse mo-

menta.

The Bjorken- $x$  value enters into the matrix element method calculation through parton distribution function as well as the transition matrix element. The value is calculated from the parton energy and the total collision energy,  $x_{\text{Bjorken}} = E_i/E_{\text{total}}$  after initial state momentum-vectors are calculated from the set of final state momentum-vectors, and is used to calculate the interaction probability for a given parton type as shown in Figure 7.1. This results in the initial state momentum-vectors being constrained by the probability of having the required Bjorken- $x$  values. If the parton distribution functions strongly disfavour the required Bjorken- $x$  values then the integration will move away from the region of  $\mathbf{y}$  phase-space.

Finally, constraints can be applied to the neutrinos from the measured missing transverse energy which can be associated with the sum of the transverse neutrino momenta:

$$(E_T^{\text{miss}})_x = \mathbf{v}_x^1 + \mathbf{v}_x^2, \quad (7.11)$$

$$(E_T^{\text{miss}})_y = \mathbf{v}_y^1 + \mathbf{v}_y^2. \quad (7.12)$$

Using these assumptions, the number of degrees of freedom in the integration phase-space is reduced from 24 to 8. This leaves four  $b$ -quark energies and four neutrino momenta components to be integrated over.

### 7.3.4 Integration Ranges

Mathematically, the integration needs to take place over the full available phase-space. However it is possible to limit the integration region to a range which will cover the dominant contribution and as such will approximate the required phase-space. The dominant contributions to the integrand will be from functions inside Equation (7.2) which are peaked for certain values of  $\mathbf{y}$ . These peaks will typically come from the transfer function and from the transition matrix element. Components of the transfer function will be peaked when a phase-space point is favoured by  $W(x_i|y_i)$ . Similarly, the transition matrix element will have Breit-Wigner terms corresponding to virtual particle decays. A top quark will decay to  $Wb$  and the process  $W \rightarrow l\nu$  will take place. The transition matrix element will favour a set of final state parton-level momentum-vectors,  $\mathbf{y}$ , which would be produced if the  $W$ -boson decayed on-shell. This knowledge can be used to define the integration regions for different variables in  $\mathbf{y}$  to aid the convergence of the

integration by removing regions of phase-space which will be known to be disfavoured.

The integration range for a parton energy is adjusted according to its component of the factorised transfer function. For a given energy, it is possible to extract an estimate of the width of the relevant component ( $\sigma_{TF}$ ). The parton integration is then performed within a  $3\sigma_{TF}$  window around this initial value. For integration variables related to an invariant mass, a range is calculated using the transfer function components of the two decay products. This is compared with a user-defined mass range and the most extreme values are taken to define the integration window. This is to optimise the integration region whilst providing the system flexibility to adjust if there are unexpected final state energies which could prevent the integration converging.

In this situation, one could find that the integration phase-space does not cover all the dominant contributions to the integrand which would result in multiple competing regions of phase-space contributing similar amounts. Without the flexibility to ensure that the phase-space is extended to encompass all sources of dominant likelihood contributions, the integration will take longer to converge for some permutations.

## 7.4 Phase Space Volume Element

The phase space volume element is derived from Fermi's Golden Rule [16], which describes the density of final states integrated over. The treatment of this volume is important when transforming variables from the momentum components to more useful quantities which aid in the integration convergence.

The phase space volume element for eight particles in the final state is

$$d\Phi_8(\mathbf{y}) = \prod_{i=1}^{N=8} \frac{d^3 \vec{p}_i}{(2\pi)^3 2E_i}, \quad (7.13)$$

where the set  $\mathbf{y}$  is now explicitly written using the final state parton momentum-vectors. Nominally this expression is considered in the Cartesian coordinate system, where  $d^3 \vec{p} = dp_x dp_y dp_z$ . If one integrates only over the energy of a parton, and assumes the measured angular components are exact, then one can transform this to polar coordinates,

$$\frac{d^3 \vec{p}_i}{(2\pi)^3 2E_i} = \frac{\sqrt{E_i^2 - m_i^2}}{2(2\pi)^3} dE d\Omega, \quad (7.14)$$

where  $d\Omega = \sin\theta d\theta d\phi$ .

One needs to take care if using additional integration variables which are functions of these variables, for instance, invariant mass variables. In this case, one needs to evaluate the determinant of the Jacobian matrix of the initial integration variables (momenta and energies) and the used integration variables (momenta, energy, masses or other variable combinations). The Jacobian matrix,  $J$ , is defined as the partial derivatives of all integration variables,  $\mathbf{f}(\mathbf{y})$ , with respect to the initial variables,  $\mathbf{y}$ ,

$$J = \frac{d\mathbf{f}(\mathbf{y})}{d\mathbf{y}} = \begin{pmatrix} \frac{\partial f_1}{\partial y_1} & \dots & \frac{\partial f_1}{\partial y_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_N}{\partial y_1} & \dots & \frac{\partial f_N}{\partial y_N} \end{pmatrix}. \quad (7.15)$$

The reciprocal of the determinant of this matrix is the multiplicative factor which rescales the integration volume when it undergoes a variable transformation. The Jacobian determinant has been evaluated and simplified symbolically in the following cases using Mathematica 9 [158].

There are cases when the integration over some parameters,  $y_i$ , can be approximated using a narrow width approximation to analytically solve the values of  $y_i$  for a point in the integration phase-space if the partons are produced by the decay of a massive particle. The invariant mass of a particle,  $V$ , decaying to two massless particles,  $p_1$  and  $p_2$  is given by

$$m_V^2 = E_1 E_2 (1 - \cos\theta_{12}), \quad (7.16)$$

where  $E_i$  is the energy of the two decay products and  $\theta_{12}$  is the opening angle between the two decay products. One could therefore decide to only use  $E_1$  as an integration variable and analytically calculate the value of  $E_2$  assuming it was produced by the on-shell decay of an intermediate particle. In this situation, the integration variable and associated factors are removed and a phase-space correction term is used. This term is  $\pi\Gamma M$ , where  $M$  is the mass of a heavy intermediate particle and  $\Gamma$  is the width of the intermediate particle. An illustration of this process is shown in Figure 7.3 where one could consider the production of a  $W$ -boson which decays to a lepton and neutrino.

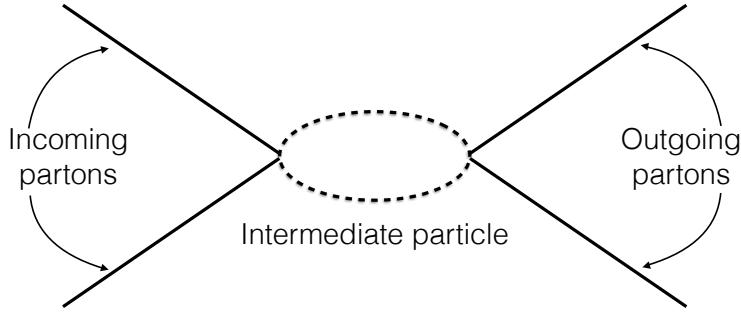


Figure 7.3: Illustration of an  $s$ -channel  $2 \rightarrow 2$  process via a virtual intermediate particle.

This term is derived using the analytic form of the Breit-Wigner function,

$$BW(m^2) = \frac{1}{(m^2 - M^2)^2 + \Gamma^2 M^2}. \quad (7.17)$$

Integrating this function across the entire domain of this function gives a total area of  $\pi/(\Gamma M)$ . The matrix element will have this function associated with the mediator, but when using an analytic solution one selects points at the location  $m = M$  rather than across the whole range of  $m$ . As such, the integral over this function is not evaluated in its entirety as required by the matrix element method likelihood. Instead, a single evaluation is performed with an area of  $1/(\Gamma^2 M^2)$ . A correction factor of  $(\pi \Gamma M)$  is therefore applied to rescale the volume element to the correct size.

VEGAS will perform optimally if the integration variables are aligned with peaked structures within the integration phase-space as this allows the importance sampling to quickly identify the regions which contribute the most to the integrand. These peaks come from the transfer functions (which identify the favoured parton kinematic phase-space, given the reconstructed kinematics) and from peaked probability distributions within the matrix (where Breit-Wigner invariant mass peaks prefer kinematics close to the on-shell decay of intermediate particles). The Breit-Wigner prior motivates the choice to use mass variables within the integration where possible, rather than integrating momentum space. This can clearly be seen in the signal hypothesis calculation, where the matrix element will prefer to constrain itself to a region of phase space where an intermediate Higgs boson decays to two  $b$ -quarks. It has been noted that VEGAS performed

more efficiently when using this transformation, wasting less integration points and converging quicker.

To optimise the integration, two different integration strategies were investigated which applied different methods to select the integration phase-space. The purpose of the study was to identify the benefits of using an analytic solution to calculate the neutrino solutions, rather than relying on full numerical integration. A balance needs to be struck between the time for the calculation to converge with the behaviour of the final likelihood variables.

The initial set of integration variables in the set  $\mathbf{y}$  after all the approximations discussed in Section 7.3 are applied are four  $b$ -quark energies labelled  $E_{j_1}, E_{j_2}, E_{j_3}, E_{j_4}$  and four neutrino momenta components labelled  $p_1^x, p_1^y, p_1^z$  and  $p_2^z$ . The measured missing transverse energy is used to analytically constrain the variables  $p_2^x$  and  $p_2^y$ . The numbering of the energies is such that  $E_{j_i}$  refers to a jet associated with parton,  $i$ . The jets are not explicitly labelled here as they will be permuted through all combinations. One additional note is that  $b$ -quarks 1 and 2 are linked to the top quark decays and  $b$ -quarks 3 and 4 are linked to the Higgs candidate in the signal hypothesis and to the additional  $b\bar{b}$  radiation in the background hypothesis. The jet associated with  $b$ -quark 3 will always be the highest energy jet out of the pair being associated to  $b$ -quarks 3 and 4 in order to aid the integration.

Two different integration strategies will be described in this section which differ in the treatment of the neutrino momentum-vector estimation. A method which purely integrates the momentum-vector components is presented in Section 7.4.1 and an alternative method which uses an analytic neutrino solution is presented in Section 7.4.2.

### 7.4.1 Integration Strategy - Integrated Neutrino Solution

Out of the eight integration variables, the approach in this strategy is to integrate over six of them and analytically solve  $p_1^z$  and  $p_2^z$  using the narrow width approximation for the  $W \rightarrow l\nu$  processes. The integration variables are  $E_{j_1}, E_{j_2}, E_{j_3}, p_{\nu_1}^x, p_{\nu_1}^y$ . The sixth integration variable varies depending on the hypothesis being evaluated. For the signal hypothesis, a transformation is made from the sub-leading energy Higgs-candidate jet,  $E_{j_4}$ , to the dijet invariant mass,  $M_{j_3 j_4}$ . For the background hypothesis the variable is kept as  $E_{j_4}$ .

The phase space volume element is shown in Equation (7.18) and the associated Jacobian factors in Equation (7.19)-(7.20).

$$d\Phi = dE_{j_1} dE_{j_2} dE_{j_3} dM_{j_3 j_4} dp_{v_1}^x dp_{v_1}^y \frac{1}{2(2\pi)^3 E_{v_1}} \frac{1}{2(2\pi)^3 E_{v_2}} (\pi M_W \Gamma_W)^2 \prod_{i=1}^4 \frac{\sqrt{E_{j_i}^2 - m_{j_i}^2}}{2(2\pi)^3}, \quad (7.18)$$

$$\det|J|_{\text{sig}} = \left| \frac{E_{j_3}(1 - \cos(\theta_{34}))}{M_{j_3 j_4}} \right|, \quad (7.19)$$

$$\det|J|_{\text{bkg}} = 1. \quad (7.20)$$

### 7.4.2 Integration Strategy - Analytic Neutrino Solution

The approach in this strategy is to make use of a dilepton kinematic solution for the neutrinos, described in Reference [159]. This method solves the  $t\bar{t}$  neutrino system up to a four-fold ambiguity by assuming the top quarks decay on-shell to  $bW$  and the  $W$ -bosons decay on-shell to  $l\nu$ . The calculation requires the parton momentum-vectors for the  $b$ -quarks associated to the top quark decays and the leptons associated to the  $W$ -boson decays. The calculation then provides a set of neutrino momentum-vector solutions. This analytic solution is used to direct the selection of phase-space points in the integration.

The integrand will not be evaluated if the parton momentum-vectors associated with the  $t\bar{t}$  system do not produce a valid dilepton top-pair neutrino solution at a given phase space point. In the implementation used in this method, the mass of the  $W$ -bosons in the analytic calculation are allowed to float by taking them as integration variables. This removes the narrow width approximation initially assumed by the neutrino calculation and provides some freedom to integration which helps it to converge for events which do not quite conform to the dilepton top-pair kinematics. The Breit-Wigner functions in the transition matrix element will still disfavour a  $W$ -boson decay which is extremely off-shell.

Out of the initial eight integration variables there are six which are used. If the  $W$ -boson masses were not allowed to float in the integration there would be only be integration variables associated to the  $b$ -quarks. Five of the integration variables are  $E_{j_1}, E_{j_2}, E_{j_3}, M_{W_1}, M_{W_2}$ . For the signal hypothesis, the sixth integration variable is the dijet invariant mass  $M_{j_3 j_4}$ . For the background hypothesis,  $E_{j_4}$  is used as the sixth integration variable.

The phase space volume element is shown in Equation (7.21) and the associated Jacobian factors in Equation (7.22)-(7.23).

$$d\Phi = dE_{j_1}dE_{j_2}dE_{j_3}dM_{j_3j_4}dM_{W_1}dM_{W_2} \frac{1}{2(2\pi)^3 E_{v_1}} \frac{1}{2(2\pi)^3 E_{v_2}} (\pi M_{\text{top}} \Gamma_{\text{top}})^2 \prod_{i=1}^4 \frac{\sqrt{E_{j_i}^2 - m_{j_i}^2}}{2(2\pi)^3}, \quad (7.21)$$

$$\det|J|_{\text{sig}} = \left| \frac{E_{j_3}(1 - \cos(\theta_{j_3j_4}))(p_{l_1}^z E_{v_1} - E_{l_1} p_{v_1}^z)(p_{l_2}^z E_{v_2} - E_{l_2} p_{v_2}^z)}{4M_{j_3j_4}M_{W_1}M_{W_2}E_{v_1}E_{v_2}} \right|, \quad (7.22)$$

$$\det|J|_{\text{bkg}} = \left| \frac{(p_{l_1}^z E_{v_1} - E_{l_1} p_{v_1}^z)(p_{l_2}^z E_{v_2} - E_{l_2} p_{v_2}^z)}{4M_{W_1}M_{W_2}E_{v_1}E_{v_2}} \right|. \quad (7.23)$$

### 7.4.3 Discussion of Integration Strategies

Towards the end of this work, it was identified that the use of a narrow width approximation to derive an analytic solution should be interpreted as a transformation of integration variables to the square of the invariant mass used. This means that the Jacobian determinant should include an additional correction to the phase-space volume element which was not implemented. The impact of this on the analysis is expected to be small as the neglected factor will contain a number of terms which are fixed during the integration. The main impact is that the likelihood is no longer a true probability as there will be a level of non-closure. However the values returned from the calculation are not strictly treated as absolute probabilities, and are interpreted as the result of an evaluated function. As such, it is not expected to affect the results of this analysis.

## 7.5 Statistical Measures

The histogram of a variable which offers discrimination between two classes of event, such as  $t_{LR}$  described in Equation (7.1), is constructed such that the distribution from signal events differs to that of background events. The difference between these two distributions can be used to determine if signal was present in the data.

In order to determine whether one method is better than another at discriminating between the signal and background processes, a measure of statistical agreement needs to be quantified using the shapes of distributions. The following calculations use histograms where the number of expected signal events in the  $i^{\text{th}}$  bin is  $s_i$ , and the number of expected background events is  $b_i$ .

### 7.5.1 Separation by Non-Overlapping Area

When examining by eye, one of the simplest ways to quantify the separation between two distributions is visualised by the amount of non-overlapping area. Normalising the distributions to unity such that  $s_i$  and  $b_i$  are fractions, this can be quantified by

$$\text{Sep} = \sum_i^n |s_i - b_i|. \quad (7.24)$$

This measure has been considered but it is not invariant under a non-linear transformation of the test statistic and as such alternative measures are preferred.

### 7.5.2 Discovery Significance

The statistical measure  $q_0$  has been introduced in Section 6.9 as a test of the background-only hypothesis to calculate the discovery significance of the signal-plus-background model. It is possible to use this measure when comparing the per-bin yields under different hypotheses. The comparison of the number of events per histogram bin can be modelled as a product of Poisson likelihoods. The discovery significance using a Poisson model takes into account the total number of expected events in the total distribution to quantify the significance to reject the background-only hypothesis. The Poisson model,

$$\ln(L(\mu)) = \sum_i^N n_i \ln(\mu s_i + b_i) - (\mu s_i + b_i), \quad (7.25)$$

is evaluated on an Asimov dataset where  $n_i$  is the total number of expected signal and background events in the  $i^{\text{th}}$  bin [138]. The discovery significance,  $Z = \sqrt{q_0}$ , is calculated from

$$q_0 = -2\ln\left(\frac{L(0)}{L(\hat{\mu})}\right), \quad (7.26)$$

where the final value is the number of sigma away from the background-only hypothesis that the data would be if the signal process existed. All expected discovery significances calculated in the chapter use an integrated luminosity of  $20.3 \text{ fb}^{-1}$ .

### 7.5.3 NeuroBayes Significance

The use of NeuroBayes [134] in the nominal  $t\bar{t}H$  analysis was discussed in Section 6.7, where input variables are first preprocessed by NeuroBayes to decorrelate and standardise them. NeuroBayes uses its own definition of significance to rank variables and determine if they contribute relevant additional information to the classifier [135]. The premise is to compute the correlation matrix between the input variables and the target variable. The target is defined as +1 for signal events and -1 for background events. These values are used to steer the classification of the neural network.

The significance is calculated using an iterative  $N - 1$  algorithm where the correlation to target is computed for each set of  $N - 1$  variables, removing one variable from the set each time. The variable which affects the correlation to target the least is the variable which contributes the least information to the classifier. This variable is then removed and the iterative procedure starts again until only one variable remains. This variable is deemed to be the most important. The significance of a variable is calculated when it is removed from the set. It is equal to the loss of correlation multiplied by  $\sqrt{n}$ , where  $n$  is the sample size.

When the significance is measured at the point where the variable is removed using this procedure, it is known as *additional significance*. When the significance is measured using the correlation to target of a single variable multiplied by the sample size, without taking into account the other variables in the set, it is known as *individual significance*. When the significance is calculated as the loss of correlation to target between using  $N$  variables and  $N - 1$  variables, it is known as *significance loss*. The additional significance will be equal to the individual significance for the most important variable. The additional significance will be equal to the significance loss for the least important variable.

The significance is measured before the neural network is trained in order to inform which variables contain important information which should be used in the training and which variables do not provide any new information to the classifier. This measure will become relevant when including the matrix element method information into the neural network training to establish if there is an overall gain to using it.

## 7.6 Expected Performance

The matrix element method provides a likelihood for each hypothesis. This likelihood is the summed contribution from all permutations of associating variables in  $\mathbf{x}$  with variables in  $\mathbf{y}$ . It is possible to extract the likelihood for each permutation and the one of most interest is the permutation with the largest likelihood. This permutation is the maximum likelihood solution and should represent the best matching between parton level and detector level objects which can be thought of as kinematic event reconstruction.

From this maximum likelihood permutation it is possible to extract which reconstructed object was paired with which matrix element parton. This allows the identification of a Higgs candidate by extracting the kinematics of the two jets assigned to the Higgs decay product.

As previously noted, the test statistic which will have the maximum discriminating power between two hypotheses is the likelihood ratio. In this analysis, a monotonic function of the likelihood ratio, which gives the same discriminating power, is used as an optimal discriminating variable. This variable is termed  $D_1$  and is defined as

$$D_1 = \frac{\mathcal{L}(\mathbf{x}|t\bar{t}H, \boldsymbol{\alpha})}{\mathcal{L}(\mathbf{x}|t\bar{t}H, \boldsymbol{\alpha}) + \beta \mathcal{L}(\mathbf{x}|t\bar{t}+b\bar{b}, \boldsymbol{\alpha})}, \quad (7.27)$$

where  $\beta$  represents a parameter that is adjusted to minimise the reduction in separation that arises from finite binning. The  $D_1$  variable is constructed to have a finite range and to push background-like events towards zero and signal-like events towards one. In the following results, the value of  $\beta$  for the integrated neutrino strategy is 0.46477 and the value of  $\beta$  for the analytic neutrino strategy is 0.43428.

If an event fails to be classified using the matrix element method, the event is assigned a value of  $D_1 = 0$  during post-processing of the results. This can occur if the event kinematics are extremely disfavoured by the integration method and the calculation cannot be performed. The value of zero is chosen as the momentum-vectors of the event are definitely not signal-like. The standard likelihood ratio is undefined for these events and there is no prior value which they can be assigned. Thus using  $D_1$  provides a value for all events. As the maximum likelihood is also undefined, a permutation is selected at random to calculate any event reconstruction variables as this best represents our understanding of the event. The proportion of events where this effect occurs is presented in Section 7.8 and is typically on the order of 0-2% depending on the

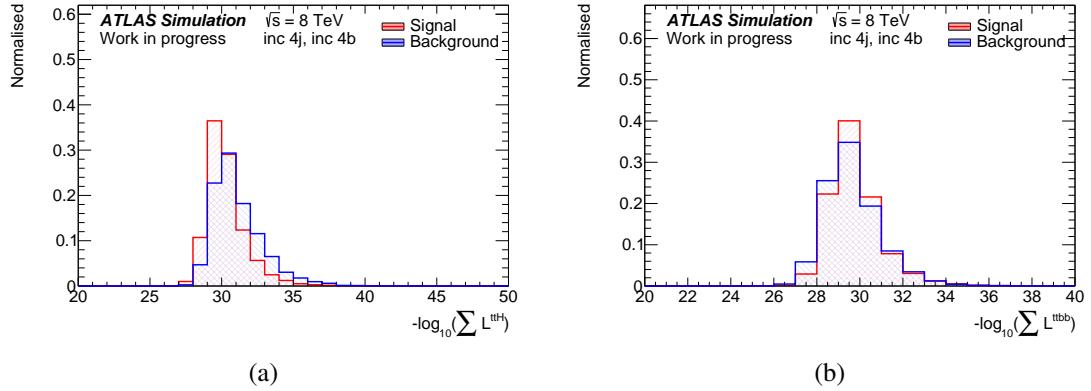


Figure 7.4: Overlay of the negative  $\log_{10}$  summed likelihoods for the signal hypothesis in Figure 7.4a and for the background hypothesis in Figure 7.4b in the region with at least four  $b$ -jets using the integrated neutrino strategy

integration method. The important check is whether the distributions show any mismodelling by assigning these values when they are compared between prediction and data. This is investigated in Section 7.7.

The aim of constructing a variable like  $D_1$  is to create a highly discriminating statistic to use as an input variable in the neural network to improve the final separation. A good separation in the shape between signal and background will improve the overall performance of the analysis in the simultaneous fit. An example of the distributions of the signal and background likelihoods in the region with at least four  $b$ -jets is shown in Figure 7.4 using the integrated neutrino strategy. The separation in these distributions is not very large, but the Neyman-Pearson lemma tells us that this is equivalent to a likelihood ratio against a uniform phase-space hypothesis. The variable  $D_1$  shows a better separation between signal and background events and is shown in Figure 7.5 for the integrated neutrino strategy and in Figure 7.6 for the analytic neutrino strategy.

Using the separation measured by non-overlapping area, the separation in Figure 7.4a is 41% and the separation in Figure 7.4b is 15%. Calculating the expected discovery significance for a Standard Model signal gives  $Z = 0.478$  for Figure 7.4a and  $Z = 0.431$  for Figure 7.4b.

The separation measured for the  $D_1$  likelihood ratios is 75% for Figure 7.5 and 72% for Figure 7.6. The expected discovery significance is  $Z = 0.600$  for Figure 7.5 and  $Z = 0.585$  for Figure 7.6. As expected, using the likelihood ratio vastly improves the discrimination between the signal and background hypotheses which makes it a good candidate statistic for use in the neural network classifier.

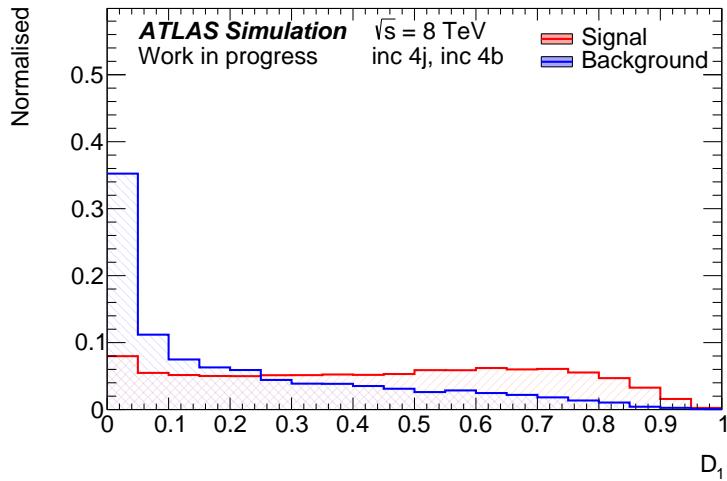


Figure 7.5: Overlay of the  $D_1$  likelihood ratio variable in the region with at least four  $b$ -jets calculated using the integrated neutrino strategy.

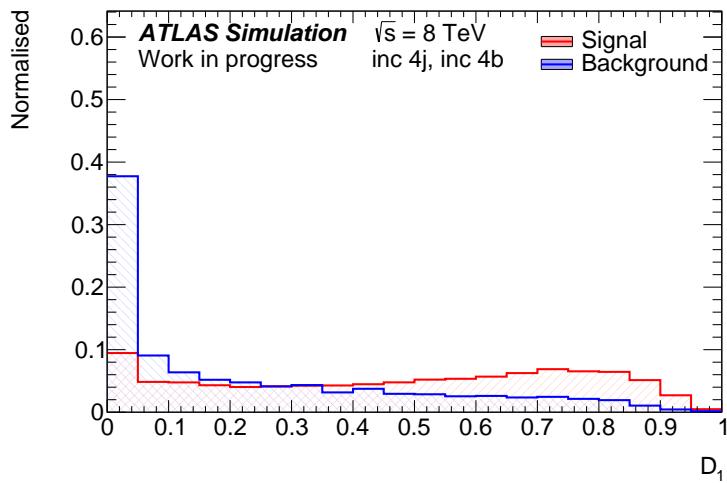


Figure 7.6: Overlay of the  $D_1$  likelihood ratio variable in the region with at least four  $b$ -jets calculated using the analytic neutrino strategy.

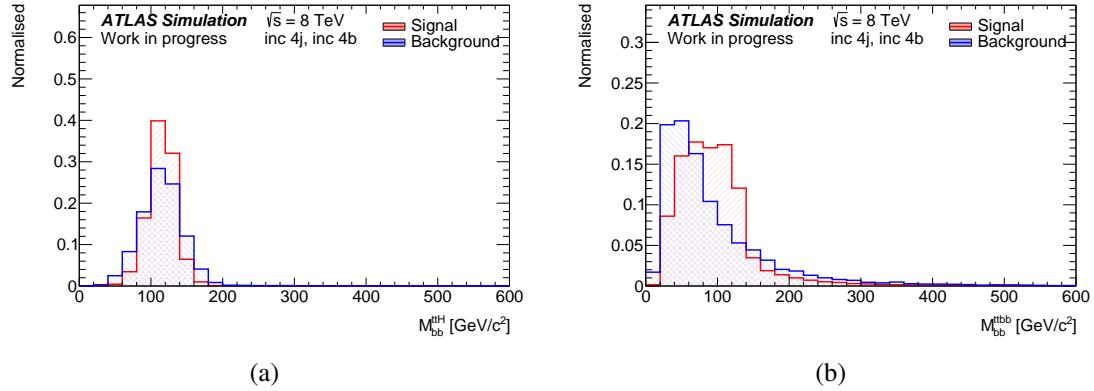


Figure 7.7: Overlay of the invariant mass for the non- $t\bar{t}$   $b$ -jets taken from the permutation with the largest likelihood for the signal hypothesis in Figure 7.7a and for the background hypothesis in Figure 7.7b in the region with at least four  $b$ -jets using the integrated neutrino strategy.

It is also possible to look at the expected shape of the kinematic event reconstruction variables, namely the mass of the Higgs candidate in the signal calculation and the mass of non- $t\bar{t}$   $b$ -jets in the background calculation. The background invariant mass object is less well defined as the leading-order diagrams can either have a gluon decaying to two  $b$ -quarks, in which case the invariant dijet mass will be the virtual gluon mass, or they can be two  $b$ -quarks which have been radiated from two separate vertices. These distributions can be seen in Figure 7.7a for the signal hypothesis and Figure 7.7b for the background hypothesis using the integrated neutrino strategy.

The separation between these distributions for signal events and background events is not as great as has been presented in the  $D_1$  variable. However, the potential to apply event reconstruction in a Higgs analysis can clearly open the door to potential studies involving the angular distribution of the decay products in signal and background event reconstruction and inter-jet particle flow around these  $b$ -jets.

The background hypothesis invariant mass distribution offers better discrimination than the signal hypothesis because true signal events will likely reconstruct the Higgs mass as the invariant mass as the remaining event kinematics will be strongly favoured as the  $t\bar{t}$  candidate. This means signal events should still try and peak around the Higgs candidate mass of 125 GeV/ $c^2$ . Background events will have an invariant mass without any clear peaked region owing to the virtual nature of the production of the non- $t\bar{t}$   $b$ -jets.

## 7.7 Validation of Methods

Both integration strategies presented in Section 7.4 have shown a similar expected performance on simulated data. There are two important metrics which need to be considered when discussing the benefits and limitations of the two different strategies. The first is the computation time taken per event. The method being documented here is very CPU intensive, and with large amounts of simulated data for a nominal selection and hundreds of systematic uncertainty variations, it is important to reduce this as much as possible. The second is the performance of the calculation. The likelihood needs to behave similarly on both data and simulation in order to use the variable in an analysis. A distribution which is not well-modelled can not be used to search for evidence of an unobserved process.

The application of the matrix element method in the signal region makes it difficult to construct a control region to test that the background-only hypothesis is well described, especially for a variable which will have different probability density functions being convoluted depending on the jet flavour. Looking at data in the signal region without any kind of validation can lead to a bias in drawing a conclusion. Confidence needs to be established in the calculation before one can trust the final distribution of the data.

The region with at least four jets and exactly three  $b$ -jets provides a validation region which has a similar background composition as the region with at least four  $b$ -jets. This region was used in the following study to assess the performance of the matrix element method calculations on data. A large fraction of expected events are  $t\bar{t}$  with additional heavy flavour jets but there are also some  $t\bar{t}+LF$  events with which to test the performance of background hypothesis calculation on events which are mistagged. Furthermore, the amount of expected signal is lower than in the region with at least four  $b$ -jets so the region can be used completely unblinded as the background modelling has already been well studied in the nominal analysis.

As the analysis uses a neural network trained to separate signal and background as the discriminating distribution in the region with at least four  $b$ -jets, it is possible to partially unblind this region using an anti-NN cut which reveals data in the signal depleted region. This region is where the amount of expected signal in each bin is less than 2% of the total expected yield. Whilst there are reduced data in this region, it does provide a validation region in which to further compare the performance on data events.

It is important to validate the method on a background-only region to ensure the performance

on data and simulation match. Once the method has been validated, it is possible to unblind the region with at least four  $b$ -jets and look to use the output of the matrix element method to improve the expected discovery significance.

The distributions shown in Figure 7.8-7.10 are comparison plots between data and simulation for the three distributions presented in the previous section in the validation region with at least four jets and exactly three  $b$ -jets. Both calculation methods appears to behave similarly in data and in simulation in this region. There is a slight data excess in this region before the simultaneous fit which can be observed as a flat shift in the ratio plot.

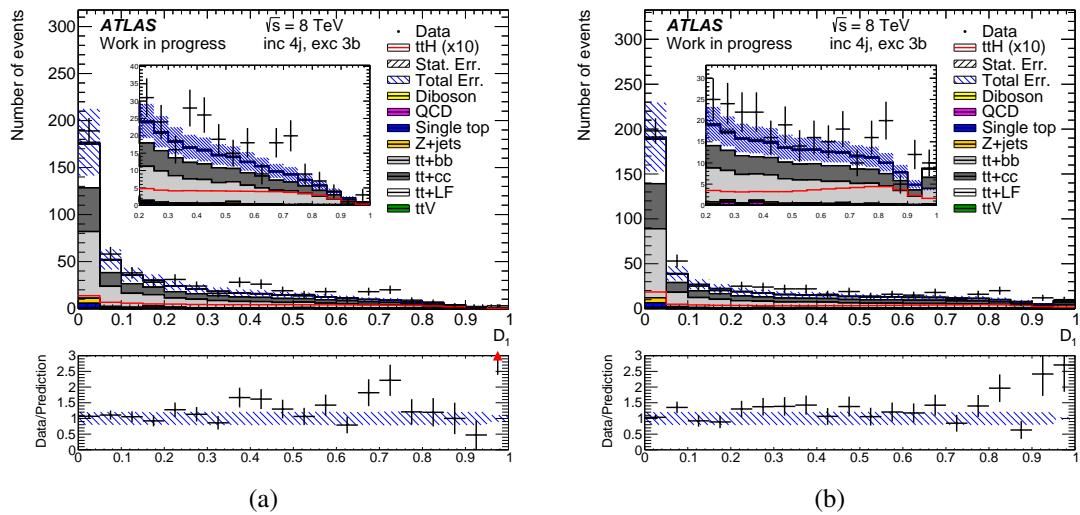


Figure 7.8: Data-simulation comparison for the likelihood ratio variable  $D_1$  calculated using the integrated neutrino solution in Figure 7.8a and using the analytic neutrino solution in Figure 7.8b in the region with at least four jets and exactly three  $b$ -jets. An inlay has been provided in the region  $0.2 \leq D_1 < 1.0$  to allow a closer inspection of the background composition in this region. The total error indicated in this plot is the statistical error combined with the 50% prefit uncertainty on the  $t\bar{t}+b\bar{b}$  background.

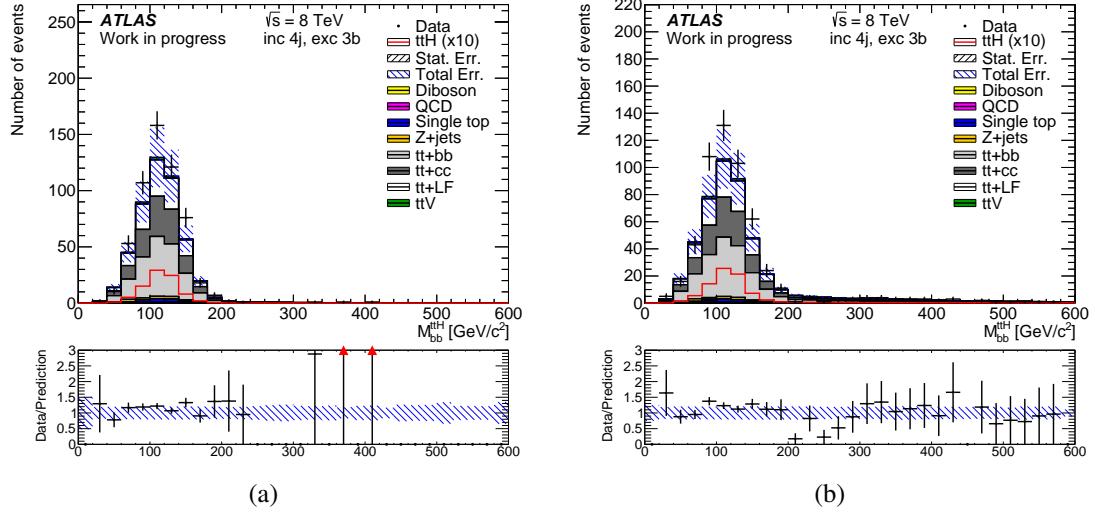


Figure 7.9: Data-simulation comparison for the invariant mass of the non- $t\bar{t}$   $b$ -jets which has the largest contribution to the signal hypothesis likelihood using the integrated neutrino solution in Figure 7.9a and using the analytic neutrino solution in Figure 7.9b in the region with at least four jets and exactly three  $b$ -jets. The total error indicated in this plot is the statistical error combined with the 50% prefit uncertainty on the  $t\bar{t}+b\bar{b}$  background.

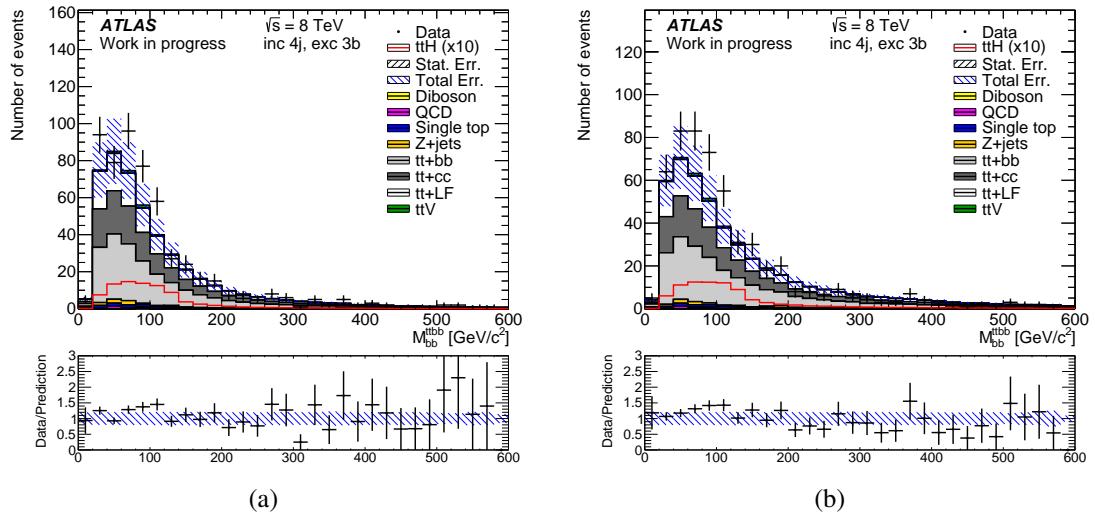


Figure 7.10: Data-simulation comparison for the invariant mass of the non- $t\bar{t}$   $b$ -jets which has the largest contribution to the background hypothesis likelihood using the integrated neutrino solution in Figure 7.10a and using the analytic neutrino solution in Figure 7.10b in the region with at least four jets and exactly three  $b$ -jets. The total error indicated in this plot is the statistical error combined with the 50% prefit uncertainty on the  $t\bar{t}+b\bar{b}$  background.

The distribution for  $D_1$  in the anti-NN region with at least four  $b$ -jets is shown in Figure 7.11. In this region, the limited data precludes any conclusion on the performance of the two methods on data and simulation.

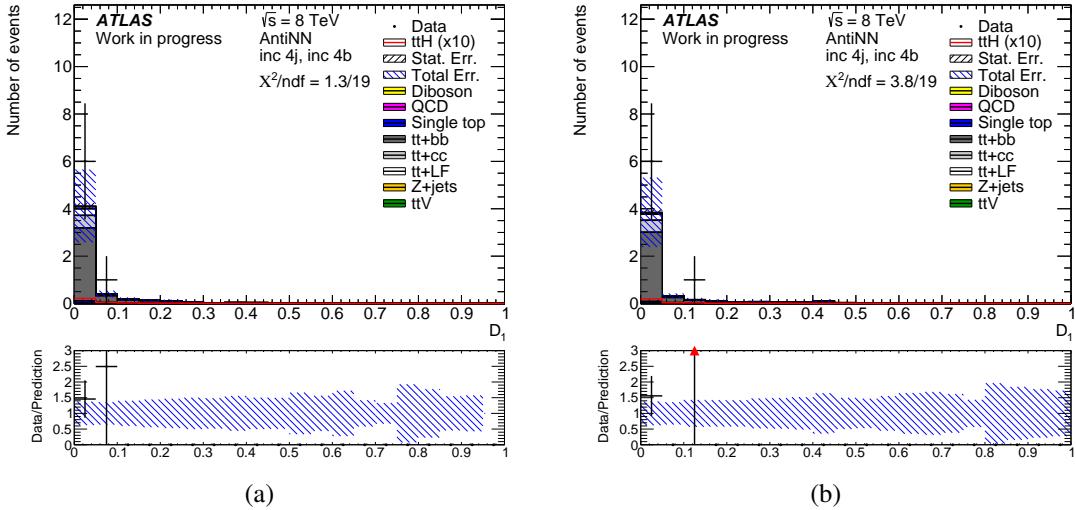


Figure 7.11: Data-simulation comparison for the likelihood ratio statistic  $D_1$  calculated using the integrated neutrino solution in Figure 7.11a and using the analytic neutrino solution in Figure 7.11b in the anti-NN region with at least four  $b$ -jets. The total error indicated in this plot is the statistical error combined with the 50% prefit uncertainty on the  $t\bar{t}+b\bar{b}$  background.

The  $t\bar{t}+b\bar{b}$  background has a large prefit uncertainty of 50% which has been included in these plots as it is one of the dominant systematic uncertainties in the analysis. The search is currently limited by the statistical uncertainties of the data in the  $\geq 4$  jets,  $\geq 4$   $b$ -jets region. It has not been possible to provide the full set of systematic uncertainties in these distributions, however the matrix element method variables are in agreement within the errors included. As such, the performance of the matrix element method in these regions does not appear to cause any concern when evaluated on data and simulation.

## 7.8 Truth Matching Efficiency

As presented in Section 7.6, it is possible to attempt event reconstruction in this  $t\bar{t}H$  search by using the permutation which maximises the signal hypothesis likelihood. In order to quantify the confidence in this method, one needs to know how often the maximum likelihood permutation correctly identifies the true Higgs pairing. A measure of this matching efficiency is dependent

on the definition of a truth matched event and is defined using the following definition:

$$\epsilon_{\text{match}} = \frac{\text{Number of truth-matched events with correct Higgs pairing}}{\text{Number of truth-matched events}}. \quad (7.28)$$

The simulated data samples used in the analysis stores truth information from the event generation about the parton kinematics from the hard interaction. This truth information can be used to spatially match partons to jets after the full simulation chain has taken place. These jets can then be compared to the jets which are assigned to the Higgs candidate in the matrix element method calculation.

An event from the  $t\bar{t}H$  signal sample which has passed the full analysis selection in the region with at least four  $b$ -jets is required to also satisfy the following criteria for this efficiency study:

- Exactly four true  $b$ -jets selected.
- All selected jets pass a  $b$ -tagging cut at the 70% working point.
- The Higgs boson decays to two  $b$ -quarks.
- The  $W$ -bosons produced by the decay of the top quarks decay to electrons or muons.

If these checks are passed, the parton to jet matching uses the following algorithm:

- The  $b$ -quark pair from the Higgs decay is matched to all permutations of the four jets.
- If the  $\Delta R$  between the parton and jet is below a threshold for both pairings then the value  $|\Delta R|^2 = |\Delta R_{\text{jet } i}^{\text{parton } 1}|^2 + |\Delta R_{\text{jet } j}^{\text{parton } 2}|^2$  is stored.
- The pairing with the smallest  $|\Delta R|^2$  is taken to define the jets from the Higgs decay.
- The  $b$ -quarks associated with the top-pair decay are matched to the remaining jets in the event and if a pairing is below the threshold then the event is defined as truth matched.

With events passing this stringent categorisation, the reconstructed matching efficiency can be defined on these subset of truth-matched events. The threshold chosen for the  $\Delta R$  cut is 0.3. This cut is slightly smaller than the jet algorithm cone size of 0.4 which should ensure that the partons lie inside, or close to, the jet which they are matched to. The fraction of events which

have passed this truth-matching strategy in the exclusive 4 jets, 4  $b$ -jets region in the  $t\bar{t}H$  ( $H \rightarrow b\bar{b}$ ) dilepton signal sample was measured to be 6.3%. If the truth matching criteria is loosened to only consider events in which the two true  $b$ -jets from the Higgs decay are accepted, this fraction increases to 8.1%.

The efficiency measurement of interest is the ability to reconstruct the correct Higgs candidate with the matrix element method. However, it is important to check that the jets in an event can also be matched to the top-quark decays as the matrix element method uses the full event kinematics in its calculation. If an event is passed to the matrix element calculation without all the correct jets available then the method is unlikely to correctly reconstruct the event, which is why these events are excluded from the efficiency calculation.

The efficiency to reconstruct the Higgs candidate is calculated by comparing the invariant mass of the two jets associated to the Higgs through the truth matching and with the matrix element method. A difference of up to  $5 \text{ MeV}/c^2$  is allowed to account for any potential floating point errors in the two calculations. The efficiency to reconstruct the Higgs candidate with the matrix element method for both integration methods is shown in Table 7.3.

These results show a very similar performance in the reconstruction efficiency of the two methods. The analytic neutrino solution has a slightly poorer truth matching efficiency which can be understood given the more constraining nature of this method. The calculation will favour permutations which have a clear on-shell  $t\bar{t}$  decay producing the top decay products.

It can also be seen that the constraining nature of the analytic method results in a reduced calculation efficiency. This can be observed in Table 7.4 where the efficiency for returning a likelihood on true  $t\bar{t}H$  events are presented. The calculation efficiencies in  $t\bar{t}H$ ,  $t\bar{t}+b\bar{b}$ ,  $t\bar{t}+c\bar{c}$  and  $t\bar{t}+\text{LF}$  events with at least four  $b$ -jets are presented in Table 7.5. The reconstruction efficiencies of the kinematic integration methods appear to be consistent comparing signal and different background samples.

Table 7.3: The truth matching efficiency for the permutation which maximises the signal likelihood calculation in the region with exactly four  $b$ -jets with no additional jets in the  $t\bar{t}H$  sample. This corresponds to a kinematic event reconstruction and calculates the efficiency of correctly pairing the Higgs candidate.

	Truth Matching Efficiency
Integrated Neutrino Strategy	57.7%
Analytic Neutrino Strategy	56.4%

The invariant mass of the Higgs candidate jets is presented in Figure 7.12 for all truth

Table 7.4: The calculation efficiency for the two integration methods on truth-matched events. This efficiency corresponds to how often the integration method being studied fails to find a region of phase-space in which a likelihood can be evaluated.

	Calculation Efficiency
Integrated Neutrino Strategy	99.9%
Analytic Neutrino Strategy	98.7%

Table 7.5: The calculation efficiency in the  $\geq 4$  jets,  $\geq 4$   $b$ -jets region for  $t\bar{t}H$  signal events and  $t\bar{t}+b\bar{b}$ ,  $t\bar{t}+c\bar{c}$  and  $t\bar{t}+\text{LF}$  background events which are selected using the two  $b$ -tagging methods.

TRF	$t\bar{t}H$	$t\bar{t}+b\bar{b}$	$t\bar{t}+c\bar{c}$	$t\bar{t}+\text{LF}$
Integrated Neutrino Strategy	99.8%	99.8%	99.7%	99.8%
Analytic Neutrino Strategy	97.9%	98.1%	98.1%	98.8%
Direct Tagging	$t\bar{t}H$	$t\bar{t}+b\bar{b}$	$t\bar{t}+c\bar{c}$	$t\bar{t}+\text{LF}$
Integrated Neutrino Strategy	99.9%	99.9%	99.8%	100%
Analytic Neutrino Strategy	98.6%	98.6%	97.9%	100%

matched events using the integrated neutrino strategy. Two distributions are presented; one generated from the two jets matched to the  $b$ -quarks from the Higgs decay at truth level and one of the invariant mass of the two jets which the matrix element method associates with the Higgs in the permutation with the maximum likelihood. The invariant mass of the Higgs candidate in events where the matrix element method does not select the correct pair of  $b$ -jets is shown in Figure 7.13. The correlation between the two invariant mass values is shown in Figure 7.14 for both scenarios. Finally a summary of the Gaussian fits shown in Figure 7.12-7.13 is presented in Table 7.6.

The values presented in Table 7.6 provide some indication that the event reconstruction is not entirely unbiased. The width of the fit is smaller with the matrix element method event reconstruction than with the true jets associated with the Higgs particle. As the jets associated using truth matching represents the best case scenario, this difference in width indicates that there is a bias in the event reconstruction. The reason behind this comes from the integration which takes place. The decay of the Higgs particle is formulated in the matrix element calculation with a relativistic Breit-Wigner function which is centred on the chosen pole mass of 125 GeV/ $c^2$ . Whilst there is a width associated with the Higgs decay which enables off-shell decays, the integration will converge faster and with a larger likelihood if there exists a jet pairing which satisfies the on-shell hypothesis at parton level. This means that if an event was generated such that the Higgs was off-shell, the matrix element method will prefer a jet-pairing which is on-

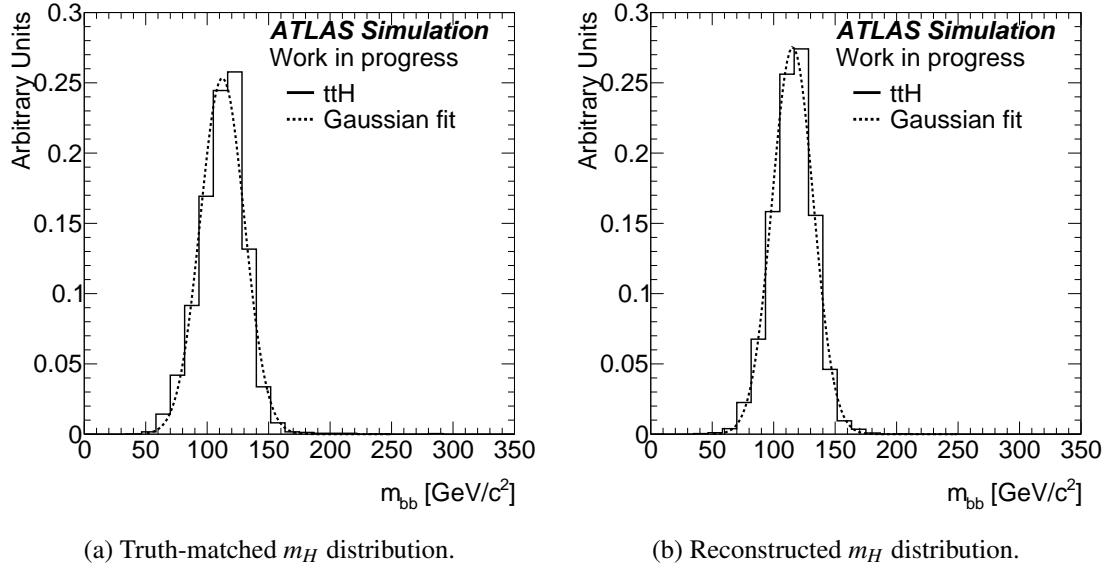


Figure 7.12: The normalised invariant mass distribution for the Higgs candidate is shown here using the same set of signal events which have passed the truth-matching requirement. The truth-matched distribution uses the jets matched to  $b$ -quarks coming from the Higgs decay. The reconstructed distribution uses the jets matched to the Higgs candidate using event reconstruction with the matrix element method using the integrated neutrino strategy. No weights have been applied to the histograms and a Gaussian fit has been applied to the histograms to identify the mean and width of the distributions for shape comparisons.

shell. Events with an incorrect pairing show this behaviour. The misreconstructed events have a mean true invariant mass which is less than the mean true invariant mass of the whole sample. However the matrix element method still favours reconstructing these events with a higher invariant mass closer to the mean of the whole sample because this will favour the hypothesis where the parton invariant mass is closer to  $125 \text{ GeV}/c^2$ . Therefore, the permutation with the maximum likelihood will provide a biased event reconstruction for off-shell events. However as it makes use of the full event kinematics, theoretical predictions and detector response effects, it should be the case that this is the best that can be achieved with reconstructed jets.

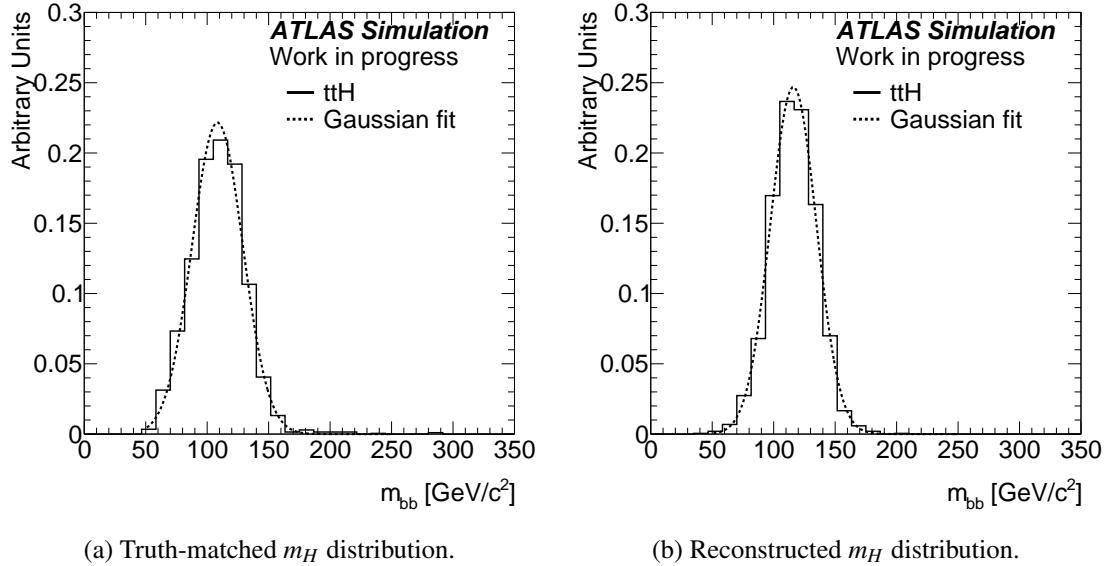


Figure 7.13: The normalised invariant mass distribution for the Higgs candidate is shown here using signal events which have passed the truth-matching requirement but has failed the event reconstruction from the matrix element method, using the integrated neutrino strategy, as the incorrect pair of jets have been assigned to the Higgs candidate. The truth-matched distribution uses the jets matched to  $b$ -quarks coming from the Higgs decay. No weights have been applied to the histograms and a Gaussian fit has been applied to the histograms to identify the mean and width of the distributions for shape comparisons.

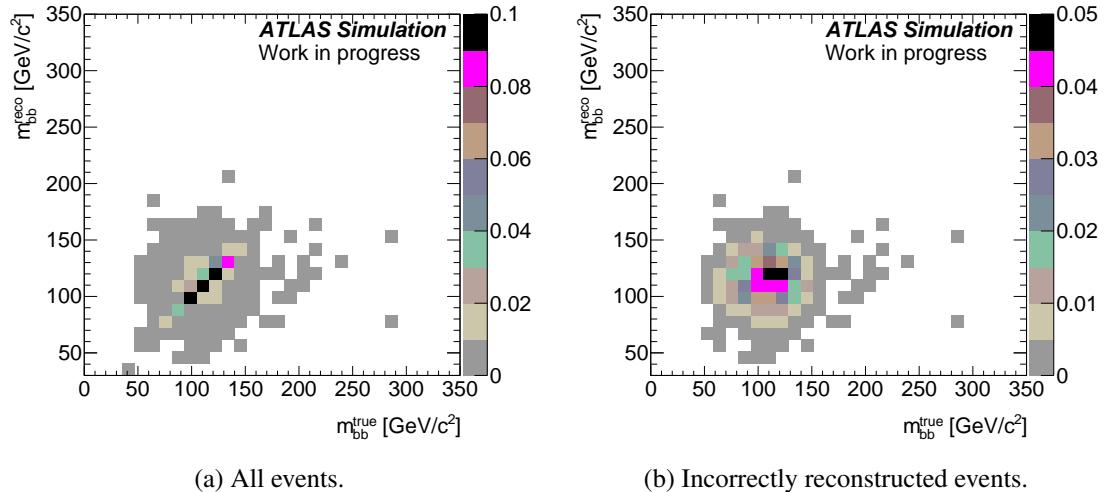


Figure 7.14: The correlation between the true Higgs mass and the reconstructed Higgs mass using the matrix element method with the integrated neutrino strategy is shown for all events and for events which the matrix element method incorrectly assigned jets to the Higgs candidate. If an event has been correctly reconstructed, it will populate the diagonal of the distribution as there will be 100% correlation between the variables.

Table 7.6: Summary of the Gaussian fits to the distributions in Figure 7.12-7.13. The truth matched distribution corresponds to the dijet invariant mass which is truth-matched to the  $b$ -jets associated to the Higgs decay. The event reconstruction distribution corresponds to using the maximum likelihood from matrix element method using the integrated neutrino strategy to reconstruct the event topology and calculating the dijet invariant mass associated with the Higgs candidate. The misidentified events category corresponds to the subset of events where the matrix element method incorrectly pairs the jets for the Higgs candidate. The comparison is then performed between the truth-matched dijet invariant mass and the dijet invariant mass from the matrix element method. When using the matrix element method to reconstruct the event topology, the mean of the distribution does not vary much between correctly and incorrectly finding the Higgs candidate. For the misidentified events, the mean of the truth-matched dijet invariant mass is reduced. The matrix element method will prefer jet pairings where the invariant mass of the Higgs candidate is consistent with  $125 \text{ GeV}/c^2$ .

Matching Criteria	Truth Matched Distribution		Event Reconstruction Distribution	
	Mean [ $\text{GeV}/c^2$ ]	Sigma [ $\text{GeV}/c^2$ ]	Mean [ $\text{GeV}/c^2$ ]	Sigma [ $\text{GeV}/c^2$ ]
All events	112.4	17.9	115.5	16.7
Misidentified events	108.4	20.6	116.1	18.7

## 7.9 Calculation Timing

The performance of the two integration strategies has been shown to be comparable in discriminating between the signal and background hypotheses and in correctly identifying the Higgs candidate in truth-matched events. The remaining measure of the two integration strategies is the integration time per event. An average computing time measured with events in the  $\geq 4$  jets,  $\geq 4$   $b$ -jets region which are selected using direct  $b$ -tagging is presented in Table 7.7. The analytic neutrino method can be observed to run the integration for the background hypothesis marginally faster than the integrated neutrino method. This is primarily due to the fast rejection of permutations which do not conform to the analytic  $t\bar{t}$  kinematic constraints. As there is a slight penalty to pay in the calculation efficiency when using the analytic neutrino method, results on the full analysis are presented in this chapter using both methods to provide a full comparison.

Table 7.7: Summary of the mean processing time of the matrix element method on events in the  $\geq 4$  jets,  $\geq 4$   $b$ -jets region selected using direct  $b$ -tagging.

Neutrino Strategy	Integrated		Analytic	
	$t\bar{t}H$	$t\bar{t}+b\bar{b}$	$t\bar{t}H$	$t\bar{t}+b\bar{b}$
Event Type	40 s	30 s	40 s	30 s
Signal Hypothesis	270 s	170 s	170 s	130 s

## 7.10 NeuroBayes Training

In order to improve the nominal analysis, the matrix element method needs to provide new information which helps to discriminate between the signal and background processes. As the matrix element method should already approach the most discriminating test statistic, one needs to consider the level at which the processes are modelled. The simulated  $t\bar{t}H$  and  $t\bar{t}$ +jets samples used in this analysis all use NLO matrix elements whilst the matrix element method is providing discrimination using LO information only. Therefore in principle, including the matrix element method information into the neural network training will only improve discrimination if there are variables in the training which parametrise higher order physics effects. The main difference between LO and NLO distributions is generally seen in pseudorapidity variables so one would expect an improvement in discrimination if these are included along with the matrix element method variables.

The  $\geq 4$  jets,  $\geq 4$   $b$ -jets region of the analysis offers the most power for discovery. This region has been selected to provide comparisons of the neural network training when the matrix element method information is included. The neural network training is split into two samples to allow cross-training and evaluation on the full sets of simulated data without a special generation of training data. The splitting uses the event number which is unique within simulated samples to split into even and odd selections. The amount of data in each of these samples is approximately equal.

The ranking for the ten nominal input variables plus  $D_1$  is shown in Table 7.8-7.9 for the integrated neutrino strategy and the analytic neutrino strategy respectively. The ranking procedure indicates that the  $D_1$  variable is highly correlated with the target and outperforms the best variable in the nominal selection. The rankings of the ten variables used in the published analysis are included to show how information contained in the  $D_1$  variable changes the importance of information in the nominal variables. One clear example of this is the  $n_{\text{Higgs}}^{30}$  variable which drops from being ranked fifth to being ranked eleventh. As the matrix element method contains information about permutations of jets and how they satisfy the  $t\bar{t}H$  signal hypothesis, it would make sense that this variable is an approximation of the matrix element method and is therefore made redundant. The change in rankings does not provide all the evidence that including the matrix element information improves the analysis as it may be possible that any information in the matrix element method was already parameterised by the different input variables and including

$D_1$  has only provided this information in a more efficient way. Therefore one should compare the values presented in Table 7.10 to look at the overall performance of the neural network. The total significance and correlation to target show an improvement of around 10% with the inclusion of matrix element method information which shows that there is additional information provided to the analysis to improve the classification of signal and background processes.

## 7.11 Simultaneous Fit Results

The combined likelihood function used in the full  $t\bar{t}H$  analysis was described in Section 6.1. The likelihood is built from histograms in each of the analysis categories where each bin is modelled with a Poisson distribution. The likelihood function has the form

$$\mathcal{L}(\mu, \boldsymbol{\theta}) = \prod_{c \in \text{cat}} \left( \prod_{i \in \text{bin}} \text{Pois}(n_i | v_i(\mu, \boldsymbol{\theta})) \right) \prod_{s \in \text{sys}} f(m_s | \theta_s). \quad (7.29)$$

The parameter of interest is the scaling of the Standard Model cross-section expectation,  $\mu$ , and the sources of systematic error are parameterised with  $\boldsymbol{\theta}$ .

The list of potential sources of systematic error in the full analysis is on the order of a hundred. Some of these uncertainties are normalisation uncertainties and some are shape uncertainties. When considering the impact of the matrix element method on a final set of results, it is more instructive if the comparisons are made like-for-like. Some shape systematic uncertainties can alter the reconstructed object kinematics which would require the matrix element method to be recalculated. Unfortunately, this is beyond the time scope of this project. However, systematic uncertainties which affect the global normalisation or adjust the statistical weight of an event can be included in a comparison without a lengthy recalculation.

Three systematic uncertainty categorisations are analysed to observe how the expected discovery significance depends on the inclusion of systematic uncertainties. The *statistical uncertainty* category only includes the data and simulation statistical uncertainties in the fit. As these uncertainties are very small, it provides very little room for the system to vary. The *theoretical cross-section* category includes the statistical uncertainties and floating normalisations constrained within the best theoretical uncertainties for each simulated process. Finally, there is the category which includes *all* the systematics which have been included in this study. This set of systematic uncertainties is much reduced from the standard analysis owing to the limited time

Table 7.8: Neural network input variables listed in order of the additional significance which is used by NeuroBayes to rank variables as a measure of importance using the full set of training data. The ranking of the original ten variables in the  $\geq 4$  jets,  $\geq 4$   $b$ -jets region in the published analysis is provided in order to see the effect of including the matrix element method information in the form of  $D_1$  for the integrated neutrino solution. The significances are provided by the Neurobayes package and a brief explanation is provided in Section 7.5.3.

Variable	Additional Significance	Individual Significance	Significance Loss	Global Correlation	Ranking from [13]
$D_1^{\text{Int}}$	72.44	72.44	33.97	67.9%	n/a
$\Delta\eta_{ij}^{\max}$	23.68	43.53	19.27	49.0%	1
$m_{jj}^{\min}$	15.68	33.98	7.17	71.7%	8
$\Delta R_{\text{Hep}}^{\min}$	12.14	32.23	8.91	56.0%	4
$m_{bb}$	10.34	40.37	8.67	47.1%	3
$\Delta R_{bb}^{\max p_T}$	7.68	36.43	6.93	41.5%	6
$m_{bb}^{\min \Delta R}$	5.59	40.20	5.64	73.4%	2
$\Delta R_{\text{Hep}}^{\max}$	4.63	31.73	4.50	57.8%	9
Aplanarity <sub>jet</sub>	4.19	29.12	3.85	36.9%	7
$m_{jj}^{\text{closest}}$	3.51	31.77	2.40	68.6%	10
$n_{\text{Higgs}}^{30}$	0.66	33.59	0.66	71.7%	5

Table 7.9: Neural network input variables listed in order of the additional significance which is used by NeuroBayes to rank variables as a measure of importance using the full set of training data. The ranking of the original ten variables in the  $\geq 4$  jets,  $\geq 4$   $b$ -jets region in the published analysis is provided in order to see the effect of including the matrix element method information in the form of  $D_1$  for the analytic neutrino solution. The significances are provided by the Neurobayes package and a brief explanation is provided in Section 7.5.3.

Variable	Additional Significance	Individual Significance	Significance Loss	Global Correlation	Ranking from [13]
$D_1^{\text{Ana}}$	70.01	70.01	33.60	63.3%	n/a
$\Delta\eta_{jj}^{\max}$	25.57	43.52	20.68	48.0%	1
$m_{jj}^{\min}$	20.17	33.99	8.66	71.5%	8
$\Delta R_{\text{Hep}}^{\min}$	13.30	32.23	10.11	55.5%	4
$m_{bb}$	10.35	40.36	8.21	47.6%	3
$\Delta R_{bb}^{\max p_T}$	7.71	36.42	6.82	41.6%	6
$m_{bb}^{\min \Delta R}$	7.13	40.20	7.01	73.0%	2
$m_{jj}^{\text{closest}}$	5.48	31.77	3.57	68.3%	10
Aplanarity <sub>jet</sub>	4.75	29.12	4.39	36.6%	7
$\Delta R_{\text{Hep}}^{\max}$	3.78	31.73	3.79	58.1%	9
$n_{\text{Higgs}}^{30}$	0.91	33.58	0.91	71.7%	5

Table 7.10: Summary of the NeuroBayes significance for the complete set of standardised input variables used by NeuroBayes to train the neural network classifier. This value is calculated before any training is performed. A comparison is provided between the set of variables used in the nominal analysis and set which includes the matrix element method information.

Training Category	Total Significance ( $\sigma$ )	Total Correlation to Target (%)
Nominal	72.8	43.8
Nominal+ $D_1^{\text{Int}}$	80.3	48.3
Nominal+ $D_1^{\text{Ana}}$	80.2	48.3

and resources available for this work. The set of systematic uncertainties contains  $b$ -tagging uncertainties, lepton identification, trigger and reconstruction efficiencies, systematic uncertainties relating to model reweighting and a systematic uncertainty to compare the difference between TRF and direct tagging. The same smoothing and pruning procedure used in the nominal analysis is applied here to remove any uncertainties which change the yields by less than 0.5%. The list of uncertainties included in this study is presented in Table 7.11 which corresponds to a handful of the dominant uncertainties in the full analysis. The important systematic uncertainties which are missing from this set are those relating to jet energy scale. However, there are 45 variations which need to be considered to fully evaluate the impact of the jet energy scale uncertainties, each of which requires a completely new matrix element method likelihood calculation.

As detailed in Chapter 6, the analysis uses a simultaneous fit across six regions defined by the jet and  $b$ -jet multiplicities. The matrix element method is introduced as an additional variable in the neural network in the  $\geq 4$  jets, 3  $b$ -jets and the  $\geq 4$  jets,  $\geq 4$   $b$ -jets regions. The other four regions have their input to the fit unchanged. An improved separation between signal and background should improve the expected discovery significance. Results of the median upper exclusion limit are presented in Table 7.12 where data has been profiled to find the best nuisance parameter estimators. The improvement in the expected discovery significance with the associated null hypothesis  $p$ -value for the case where systematic uncertainties are included in the likelihood are presented in Table 7.13.

It is clear that the inclusion of matrix element information in the neural network has improved the expected discovery significance, which means the analysis has become more sensitive to a Standard Model  $t\bar{t}H$  observation for  $\mu = 1$ . With the set of systematic uncertainties included in this analysis, the improvement to upper exclusion limit is around 6.5% by including the matrix element method information into the neural network training. The expected discovery

Table 7.11: A summary of all systematic uncertainties included in the construction of the analysis likelihood function which is maximised in a profile fit to data. Systematic type encodes whether the systematic uncertainty is manifest as a normalisation (N) or a shape (S) effect. The number of variations describes the number of systematic shifts associated with an uncertainty. Typically this value is two, corresponding to  $\pm 1\sigma$  shifts from the nominal. In some cases there is only one variation if the uncertainty represents not applying a particular analysis technique. The  $b$ -tagging uncertainties correspond to changing  $b$ -tagging efficiencies. Multiple systematic sources affect these efficiencies so they are decomposed into orthogonal uncertainty components to allow the data to help constrain their impact. From this list of systematics, only those which contribute more than a 0.5% variation in bin yields in a region per sample are included.

Uncertainty	Systematic Type	Variations
Luminosity	N	2
$t\bar{t}+b\bar{b}$ normalisation	N	2
$t\bar{t}+c\bar{c}$ normalisation	N	2
$t\bar{t}$ normalisation	N	2
Z+jets normalisation	N	2
Single top normalisation	N	2
QCD normalisation	N	2
Diboson normalisation	N	2
$t\bar{t}V$ normalisation	N	2
B-Tagging $b$ -jets	N, S	12
B-Tagging $c$ -jets	N, S	8
B-Tagging light jets	N, S	24
B-Tagging extrapolation	N, S	2
top $p_T$ reweighting	N, S	1
$t\bar{t}$ $p_T$ reweighting	N, S	1
$t\bar{t}+b\bar{b}$ reweighting	N, S	1
Direct tagging	S	1
Z+jets $p_T$ reweighting	N, S	1
Electron ID efficiency	N, S	2
Electron reconstruction efficiency	N, S	2
Electron trigger efficiency	N, S	2
Muon ID efficiency	N, S	2
Muon reconstruction efficiency	N, S	2
Muon trigger efficiency	N, S	2

significance shows an improvement of 8%. This can be considered a non-negligible improvement given the analysis has already been highly optimised to search for the  $t\bar{t}H$  process through the object selection, event selection and the selection of variables used in the neural network. This result shows that there is still high-level information which is not being fully exploited in this analysis. If the selection of variables had provided a complete picture of the leading-order process hypotheses then the neural network should have been able to construct a variable which was as effective as the matrix element method likelihood ratio and the expected discovery sig-

Table 7.12: A comparison of the median expected upper exclusion limit, where the systematic uncertainties are profiled using the observed data in order to test the expected sensitivity of the analysis. The nominal category corresponds to the same training variables as the published analysis. The additional categories include the  $D_1$  variable. An improvement can be seen through the inclusion of the matrix element method variables to the neural network training.

Training Category	Stat. Only	XS Only	All Syst
Nominal	2.44	3.55	3.80
Nominal+ $D_1^{\text{Int}}$	2.35	3.32	3.56
Nominal+ $D_1^{\text{Ana}}$	2.35	3.32	3.55

Table 7.13: The median expected discovery significance for the three different training categories evaluated using  $q_0$  presented in Equation (6.9) with an Asimov dataset corresponding to a  $\mu = 1$  prediction with all systematic uncertainties included in this study. The median  $p$ -value for the null hypothesis associated through the one-to-one mapping to the median significance is also shown.

Training Category	Discovery Significance	$p_0$
Nominal	0.633	0.2634
Nominal+ $D_1^{\text{Int}}$	0.696	0.2464
Nominal+ $D_1^{\text{Ana}}$	0.698	0.2459

nificance would not improve. These results show that the matrix element method should be used in the coming Run 2 ATLAS analyses to improve the sensitivity of the analysis and reduce the total luminosity required to make an observation of  $t\bar{t}H$  production in data.

## 7.12 Application of Analysis to the 8 TeV Dataset

The results presented so far have centred on the expected impact of the matrix element method to the full 8 TeV dilepton  $t\bar{t}H$  ( $H \rightarrow b\bar{b}$ ) analysis. An observed upper-limit can be calculated along with information about the maximum likelihood estimator  $\hat{\mu}$  using the unconditional likelihood fit to data.

Using the full set of systematic uncertainties described in the previous section, the expected and observed 95% confidence level upper limits on the signal process cross-section relative to the Standard Model expectation are shown in Table 7.14. The observed upper limit is less constraining than expected for a Standard Model signal process which is explained by the values of  $\hat{\mu}$  shown in Table 7.15 which reveals the fit has found an excess of data which could be described by an enhancement of the signal model. The associated discovery significance and  $p$ -value for the background-only hypothesis are presented in Table 7.16. The inclusion of the matrix element method information results in a interpretation of the data which is in closer

agreement with the background-only hypothesis compared to the nominal analysis.

Table 7.14: The expected and observed median upper exclusion limit where the sources of systematic error are profiled using the observed data.

Training Category	Expected	Observed
Nominal	3.80	5.61
Nominal+ $D_1^{\text{Int}}$	3.56	5.14
Nominal+ $D_1^{\text{Ana}}$	3.55	4.82

Table 7.15: Summary of the best-fit signal strength values in the unconditional likelihood fit for the three different neural network training categories.

Training Category	$\hat{\mu}$
Nominal	$2.03 \pm 1.95$
Nominal+ $D_1^{\text{Int}}$	$1.71 \pm 1.85$
Nominal+ $D_1^{\text{Ana}}$	$1.33 \pm 1.84$

Table 7.16: The discovery significance and  $p$ -value for the null hypothesis using the observed data.

Training Category	Z	$p_0$
Nominal	1.08	0.140
Nominal+ $D_1^{\text{Int}}$	0.961	0.168
Nominal+ $D_1^{\text{Ana}}$	0.750	0.227

Importantly, these results are consistent with the published analysis which features the complete set of systematic uncertainties. In the published analysis described in Chapter 6, an excess of signal-like events were observed and the observed limit was larger than expected [13]. The inclusion of the matrix element method appears to reduce the results relative to the comparison without the matrix element method. This reinterpretation of the data would indicate that with a greater discrimination between the signal  $t\bar{t}H$  and the background  $t\bar{t}+b\bar{b}$  processes, the data becomes better described by the background hypothesis. However, all the results are compatible within their errors and there is no conclusive difference from the inclusion of the matrix element method with respect to discovering the dilepton  $t\bar{t}H$  ( $H \rightarrow b\bar{b}$ ) signal process in the 8 TeV dataset.

## 7.13 Post-Fit Validation

The expected performance of the analysis has been improved, but it is important to check the behaviour of the distributions of the neural network inputs and neural network outputs after the

fitting procedure has been performed. The normalisation of different samples is adjusted during fit to provide the best comparison between data and simulation within the provided errors. The unconditional likelihood fit allows the signal strength ( $\mu$ ) to float to its best-fit value ( $\hat{\mu}$ ). The measured  $\hat{\mu}$  values for the three training categories are presented in Table 7.15. The absolute error on  $\hat{\mu}$  can be seen to reduce when the matrix element method information has been included in the neural network. This implies that improving the separation between signal and background results in an improved measurement of signal in data.

The postfit neural network output distributions are presented in Figure 7.15-7.17. Constraining correlated systematic uncertainties with the profiled likelihood fit to data reduces the overall error on the final post-fit distributions. A comparison between the distributions output by neural networks which included  $D_1$  in their training and the nominal baseline show very similar behaviour with respect to the post-fit agreement between data and simulation.

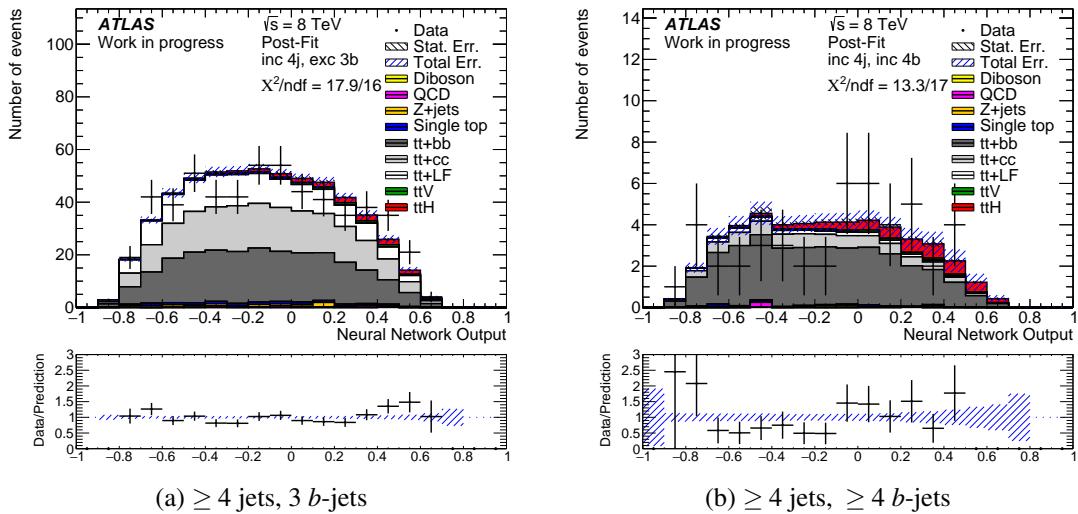
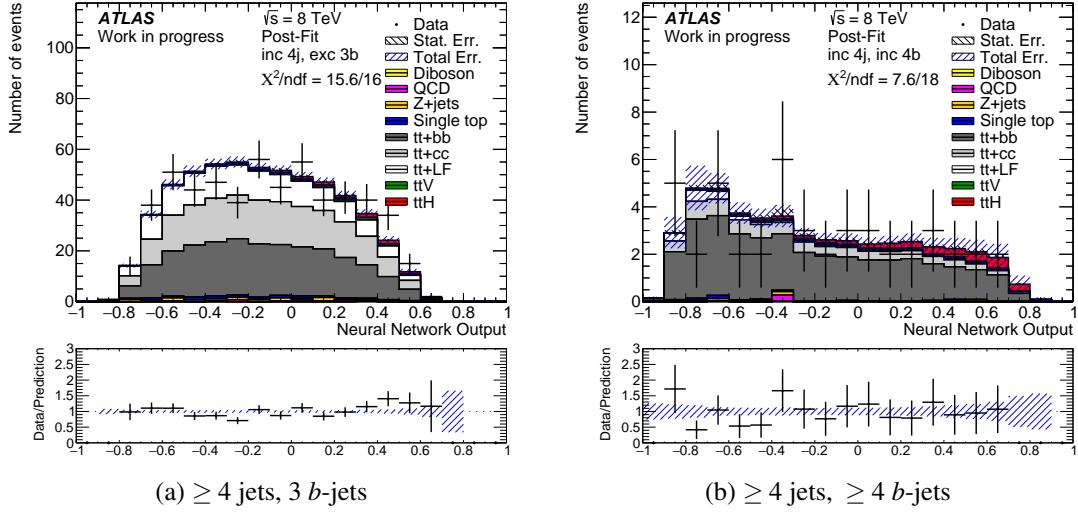
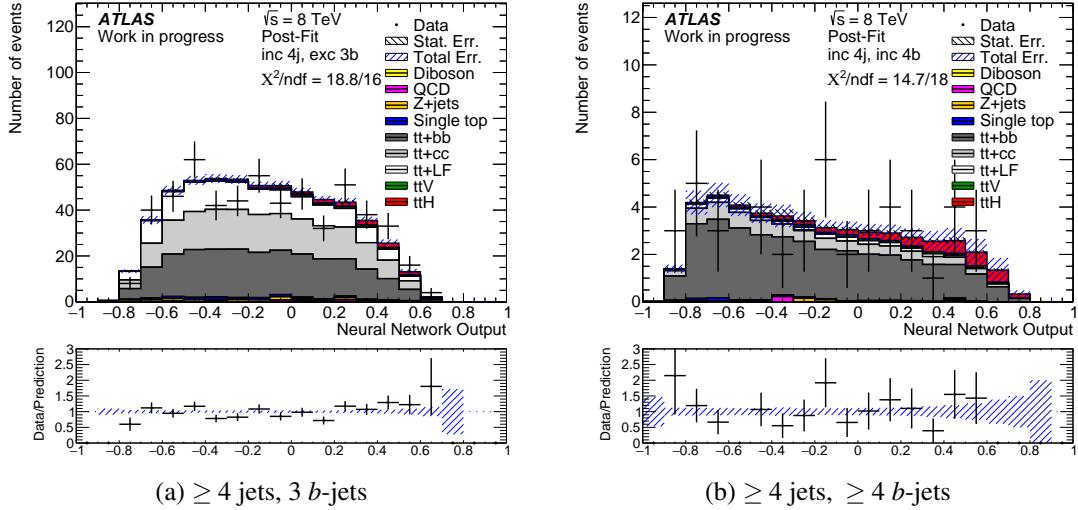
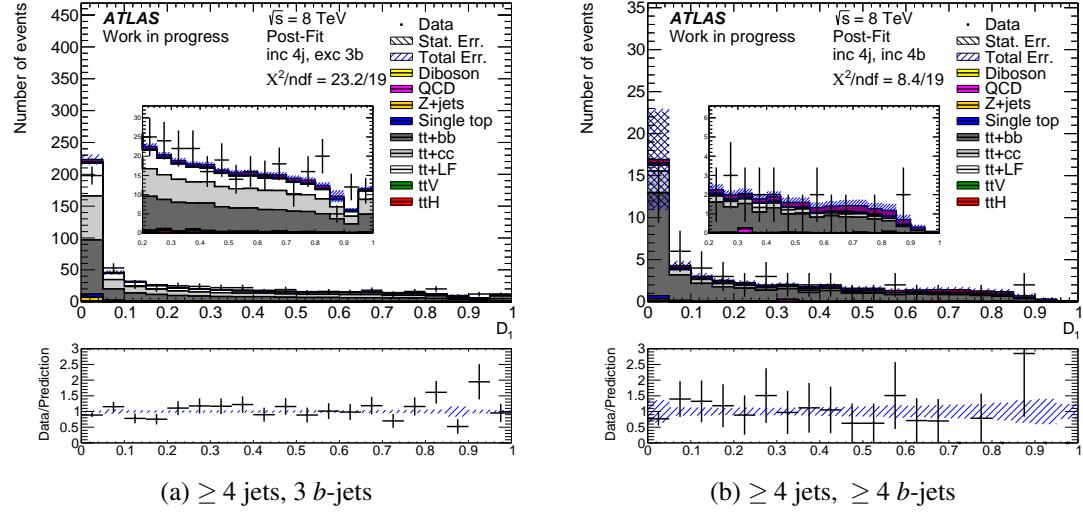
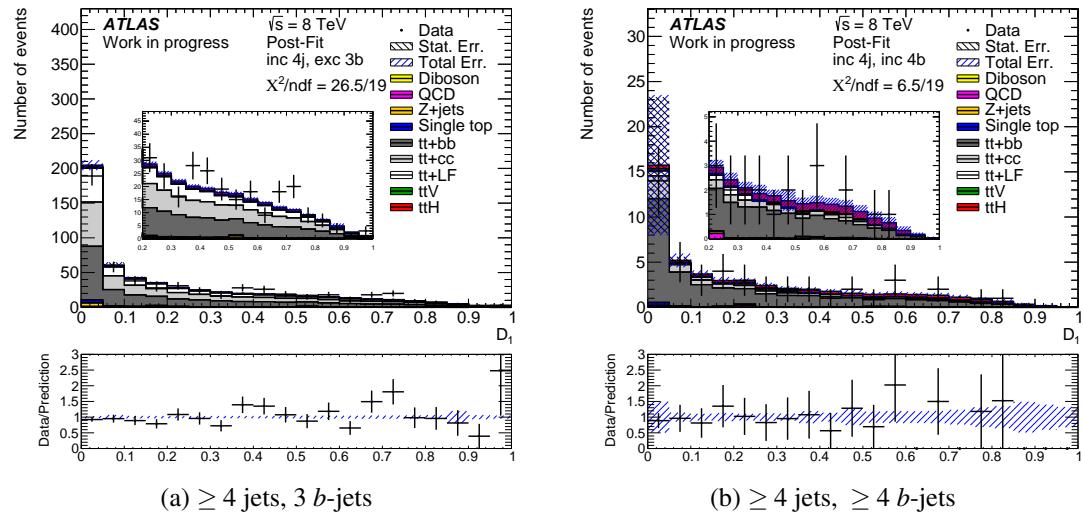


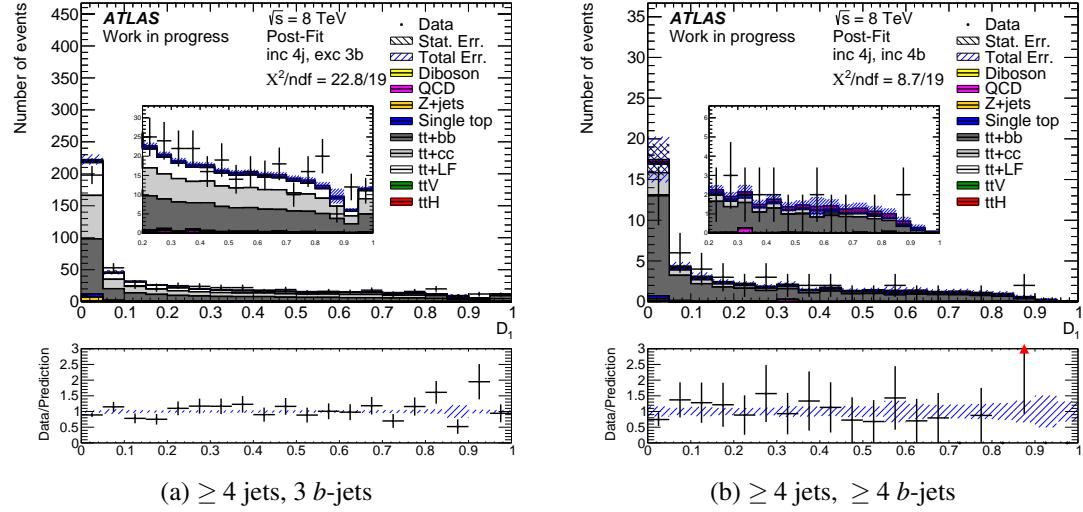
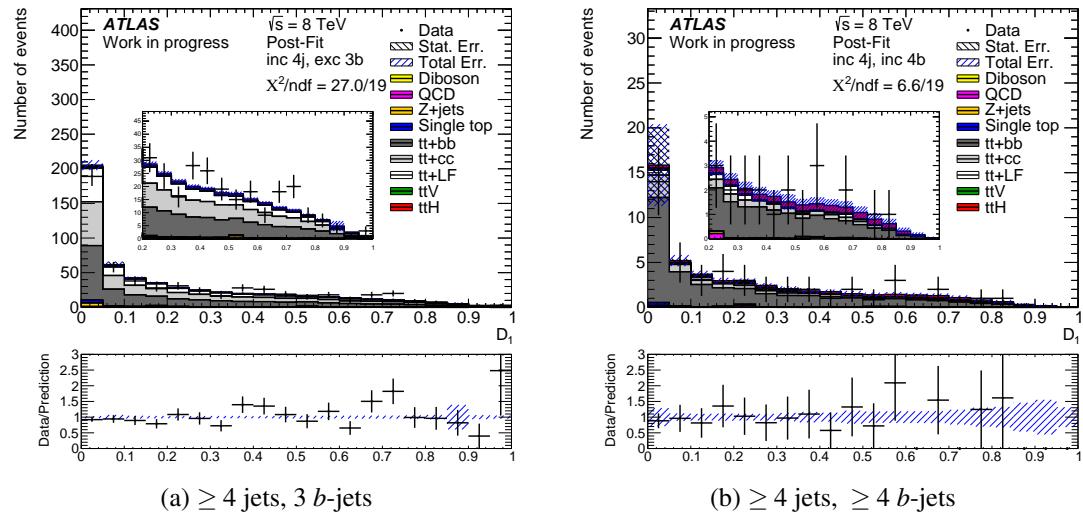
Figure 7.15: Neural network output after the nominal training.

The  $D_1$  distribution can be plotted after the nominal analysis is performed in order to see how the distribution is adjusted when the neural network has no dependence on its shape. The distributions in Figure 7.18-7.19 show this scenario for the two different neutrino strategies. Similarly, the distribution can be plotted after fitting the neural network output which was trained with the  $D_1$  variables. These distributions are presented in Figure 7.20-7.21. The important comparison between these distributions is how the variable behaves after the fit, and whether there is any indication of a bias or slope appearing. Within the data statistical uncertainties

Figure 7.16: Neural network output after the nominal training with the inclusion of  $D_1^{\text{Ana}}$ .Figure 7.17: Neural network output after the nominal training with the inclusion of  $D_1^{\text{Int}}$ .

and the limited simulation systematic uncertainties, the performance and behaviour of the  $D_1$  distribution appears to be in good agreement between data and simulation. There does not appear to be any bias introduced when using this variable in the neural network training.

Figure 7.18:  $D_1^{\text{Ana}}$  after the nominal training.Figure 7.19:  $D_1^{\text{Int}}$  after the nominal training.

Figure 7.20:  $D_1^{\text{Ana}}$  after the nominal training with the inclusion of  $D_1^{\text{Ana}}$ .Figure 7.21:  $D_1^{\text{Int}}$  after the nominal training with the inclusion of  $D_1^{\text{Int}}$ .

## 7.14 $t\bar{t}H$ Candidate Event Displays

It can be a useful exercise to produce event displays of signal candidate events. The two 8 TeV data events shown here were selected based on their high  $D_1$  value indicating that they are signal-like events. The two different event topologies can be observed in the transverse plane. The three-point Mercedes-like topology can be observed in Figure 7.22 and the back-to-back topology can be observed in Figure 7.23. In these displays jets are represented by a white cone, muons are represented by a red line, electrons are represented by a green line and missing transverse energy is represented by a dashed orange line. The top candidates can be identified by pairing the two closest jet-lepton pairs. The remaining jets can be considered the Higgs candidate for the purpose of this observation. It is interesting to see that both topologies have been identified in events which the matrix element method classified as strongly signal-like.

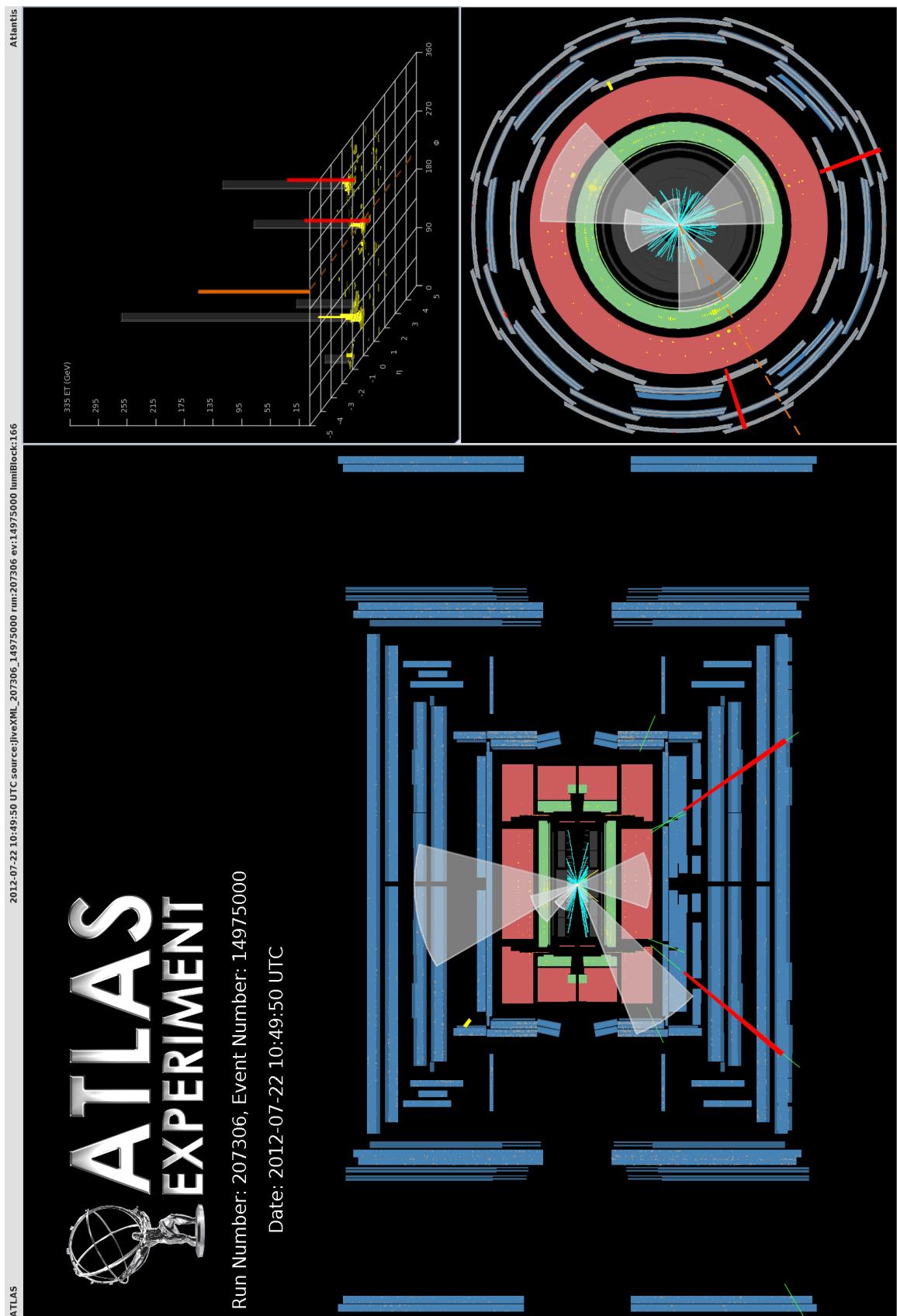


Figure 7.22: A  $t\bar{t}H$  candidate event demonstrating the Mercedes-like event topology.

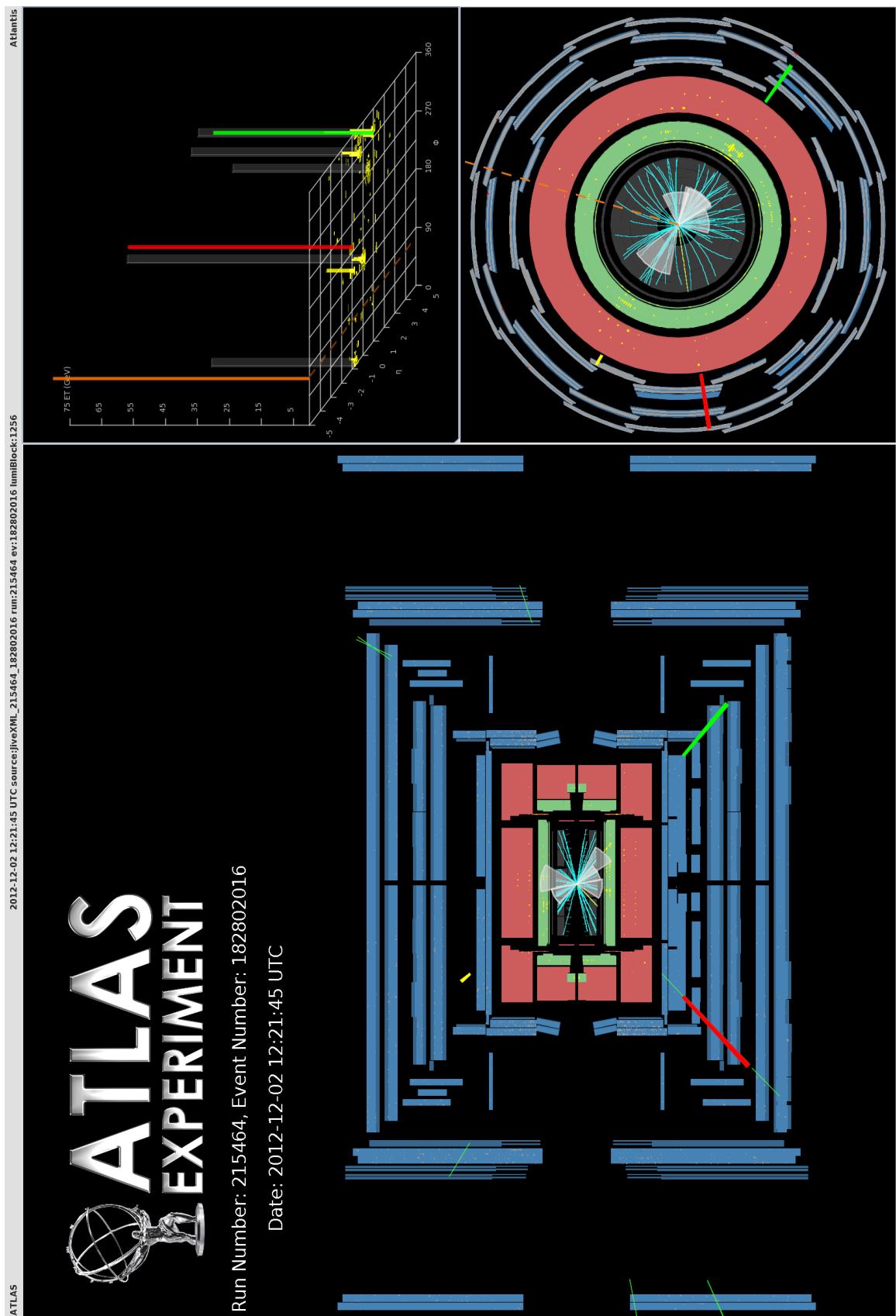


Figure 7.23: A  $t\bar{t}H$  candidate data event demonstrating the back-to-back event topology.

## Chapter 8

# Future Directions

---

*“What a fine persecution – to be kept intrigued without ever quite being enlightened.”*

— Tom Stoppard, Rosencrantz and Guildenstern Are Dead

---

The work presented in this thesis has all been focused on the 8 TeV dataset collected by the ATLAS detector during Run 1. However, the main focus of the majority of physics analyses in the ATLAS Collaboration has now been directed towards the new data being recorded at the increased centre-of-mass energy of 13 TeV which signals the start of Run 2 of the LHC.

Hopefully within this new dataset we will find evidence of the  $t\bar{t}H$  production mechanism and potentially make a direct measurement of the top Yukawa coupling. The matrix element method presented in this thesis has the potential to reduce the error on this measurement by improving the discovery significance so that the analysis becomes more effective at isolating the signal process in data.

This chapter will address some comments which can be made regarding the search for the  $t\bar{t}H$  process in Run 2 data. The expected change in process rates will be discussed in Section 8.1 where the projected impact of the matrix element method will also be included. Further additional improvements to the nominal analysis using matrix element method information will be suggested in Section 8.2. A final closing comment on the anticipated timeline for the Run 2  $t\bar{t}H$  analysis is provided in Section 8.3.

## 8.1 Run 2 Extrapolation

The  $t\bar{t}H$  and matrix element method results presented in this thesis lend themselves well to projections for Run 2 of the LHC. The change of the centre-of-mass energy from 8 TeV to 13 TeV increases the available phase-space for particle interactions. The  $t\bar{t}H$  cross-section is predicted to increase by a factor of 3.9 and the  $t\bar{t}$  inclusive cross-section is predicted to increase by a factor of 3.3. The expected discovery significance should therefore increase by a factor of  $k_{t\bar{t}H}/\sqrt{k_{t\bar{t}}}$ , where  $k_i$  is the cross-section scaling factor for process  $i$ .

The sensitivity of the analysis is heavily dependent on the number of background events in the signal-rich regions. In the  $\geq 4$  jets,  $\geq 4$   $b$ -jets region, the dominant background is  $t\bar{t}+b\bar{b}$ . The NLO cross-section of  $t\bar{t}+b\bar{b}$  is predicted to increase by a factor of 4.7 from 8 TeV to 14 TeV [45, 160], but there are currently no theoretical NLO cross-section predictions made for 13 TeV. As such, the following sensitivity study will be carried out by assuming the total background will scale in the same manner as the inclusive  $t\bar{t}$  cross-section, but it is important to note that the key background,  $t\bar{t}+b\bar{b}$ , is likely to scale at a rate which is closer to that of the signal cross-section.

The discovery significance gives the number of sigma away from the null hypothesis that the data is. For discovery of the Standard Model signal-plus-background model, this value is approximately equal to  $1/\sigma_{\hat{\mu}}$ . The nominal dilepton analysis gives a current expected discovery significance of  $Z = 0.5$ , where  $\sigma_{\hat{\mu}} = 2.0$  [13].

One can therefore make a basic extrapolation out to 13 TeV by taking into account the changing cross-sections using the function,

$$Z(L) = \frac{Z_{8 \text{ TeV}}}{\sqrt{L_{8 \text{ TeV}}}} \times \frac{k_{t\bar{t}H}}{\sqrt{k_{t\bar{t}}}} \times \sqrt{L}. \quad (8.1)$$

This calculation assumes that the inclusive  $t\bar{t}$  process is the dominant background process. The scaling of the expected discovery significance as a function of 13 TeV luminosity can be seen in Figure 8.1 for the nominal analysis and for the 8% improvement provided by including the  $D_1$  matrix element method variable. The amount of luminosity required to reach the  $3\sigma$  evidence level is highlighted on the distribution. For the nominal analysis where the Run 2 detector performance is assumed to be equivalent to Run 1, it will require  $158.6 \text{ fb}^{-1}$  of 13 TeV data, whereas the amount of data required when including the matrix element method information reduces substantially to  $135.9 \text{ fb}^{-1}$ .

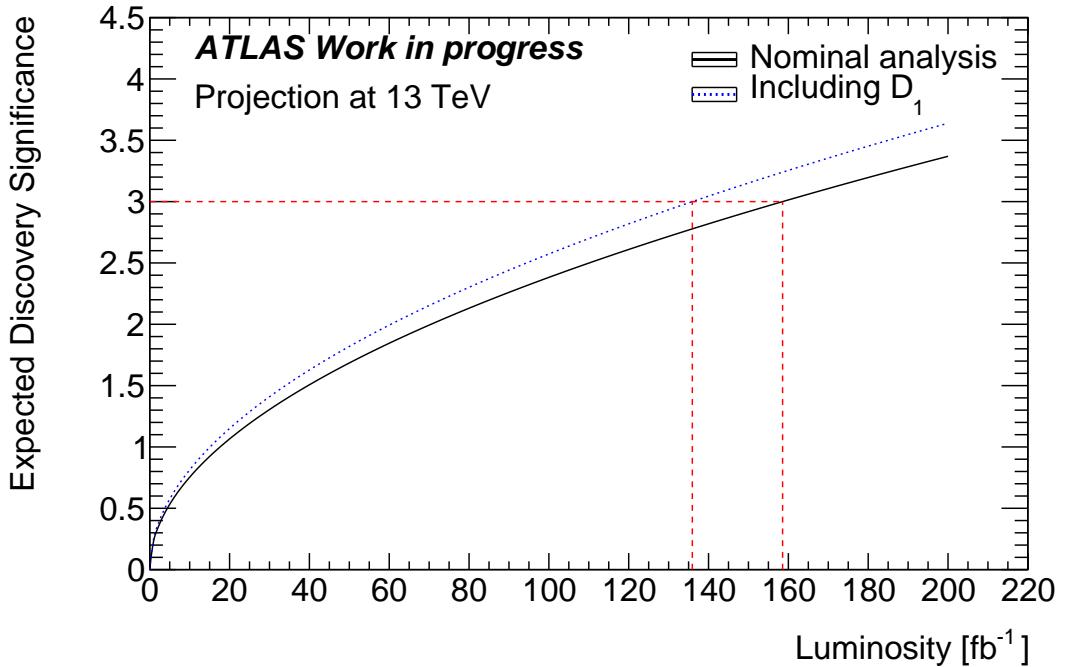


Figure 8.1: Projected expected discovery significance as a function of 13 TeV luminosity.

Requiring around  $20 \text{ fb}^{-1}$  less data for a  $3\sigma$  observation comes with an associated monetary saving. The annual running cost of Run 1 has been estimated to be 1.1 billion CHF [161]. When colliding at 8 TeV in 2012, the power consumption was approximately 650 GWh and around  $20 \text{ fb}^{-1}$  of physics data was recorded by the ATLAS detector. In Run 2 it is expected that the annual integrated luminosity delivered will be  $40 \text{ fb}^{-1}$  and somewhere between 100 and 200  $\text{fb}^{-1}$  of data will be collected in total with an estimated power consumption of 750 GWh [161]. It is therefore possible to extrapolate the running cost of Run 1 to Run 2 and calculate the cost of delivering an additional  $\sim 20 \text{ fb}^{-1}$  of data. Assuming that the running cost is linear with the energy consumption, Run 2 will be expected to be a factor of 1.15 more expensive than Run 1. However, the rate of events will be approximately doubled meaning that the cost per  $\text{fb}^{-1}$  will be reduced in Run 2. Combining these factors together gives the cost of  $20 \text{ fb}^{-1}$  in Run 2 to be approximately 0.65 billion CHF. There are hundreds of physics analyses taking place within the ATLAS Collaboration and many additional experiments which all benefit from the proton-proton collisions delivered by the LHC. Therefore, one should not interpret this cost as representative of the saving one could expect from applying the matrix element method in the dilepton  $t\bar{t}H$  analysis as the LHC has physics potential beyond making this observation. However, it is useful

to illustrate the impact that small improvements can have on an analysis which is limited by the statistical power of the available data. For nominal conditions in Run 2, the amount of data recorded per week will hopefully be on the order of  $1 \text{ fb}^{-1}$  per week. Therefore, being able to reach a physics result with  $20 \text{ fb}^{-1}$  less data amounts to making a publication 20 weeks earlier than would otherwise be possible.

During Run 2 there will be additional  $t\bar{t}H$  analyses looking at different top-quark pair and Higgs decays. Therefore whilst it may be a struggle to find evidence of the dileptonic  $t\bar{t}H$  process where the Higgs decays to two  $b$ -quarks, it would be safe to assume that a combination of analyses will be sensitive enough to find evidence or possibly make a  $5\sigma$  observation of the  $t\bar{t}H$  process by the end of Run 2. Importantly, including the matrix element method into the dilepton analysis and possibly other  $t\bar{t}H$  analyses will substantially reduce the amount of data required to reach that goal.

## 8.2 Further Matrix Element Method Variables

The matrix element method has so far been discussed as a likelihood ratio providing discrimination between two hypotheses. In this analysis it has been between the signal  $t\bar{t}H$  process and the background  $t\bar{t}+b\bar{b}$  process. However, it was also touched upon in Chapter 7 that the matrix element method provides the possibility of event reconstruction of the signal and background hypotheses.

### 8.2.1 Inter-Jet Particle Flow

It was observed that the invariant dijet mass distribution offered better discrimination between signal and background when it was constructed using the permutation which maximised the background hypothesis. The signal hypothesis offered less discrimination due to the jet combinatorics which could make a pairing close to the Higgs mass. However, the signal hypothesis provides a possibility of constructing another high-level variable based on inter-jet particle flow (also referred to as colour flow).

Colour flow has been studied in the past when considering searches for the Higgs decay to two  $b$ -quarks. In an analysis searching for this decay process, the typical background processes will involve QCD interactions to produce the  $b$ -quarks, such as gluon splitting. Whilst this may

lead to different angular distributions due to the spin-0 nature of the Higgs boson compared to the spin-1 nature of the gluon, it will also lead to a difference in the amount of soft particle radiation produced during the hadronisation process.

The Lund string model provides a nice visualisation of this abstract process [69]. When two coloured partons are created in a hadron collider, there will be a coloured string which links the partons together. As the partons travel, the string stretches and splits into collimated  $q\bar{q}$  pairs. These soft partons will travel in the direction of the closest hard parton and will eventually hadronise into colourless particles. An illustration of this process is shown in Figure 8.2.

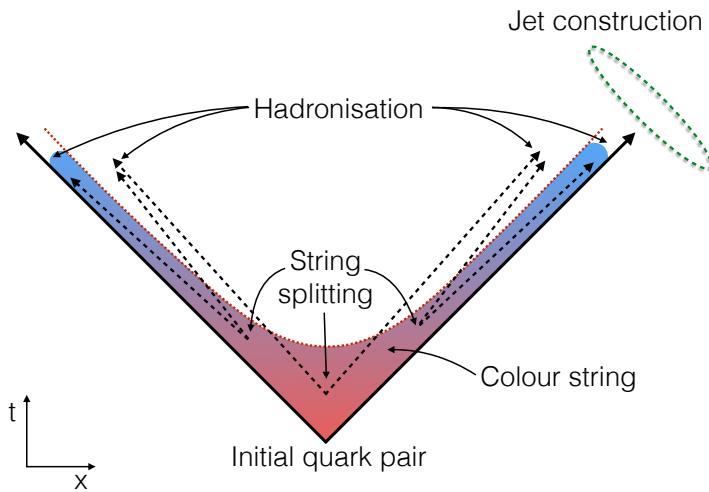


Figure 8.2: An illustration of the colour connection between two hard partons and the production of  $q\bar{q}$ -pairs as the hard partons travel away from each other based on the description in [69]. The potential energy held in the coloured field increases until it splits. These new partons have less energy than the initial partons and they pair up to ensure there are no coloured final states when the energy has been dissipated.

When a  $t\bar{t}+b\bar{b}$  background event is produced inside the ATLAS detector, there are colour strings connecting all the QCD partons in the event. These strings stretch from the initial partons through to the top-quark decay products. Furthermore, the additional  $b$ -quarks produced by this process will also have a colour connection linking back to the top-quarks.

In the  $t\bar{t}H$  signal process, the Higgs is a colourless object. This means that whilst the colour flow topology of the  $t\bar{t}$  production in the signal process will be consistent with the background processes, there will be a difference between the Higgs decay products and the additional  $b$ -quarks in the background processes. The  $b$ -quarks produced by the decay of the Higgs boson

will only be colour connected to each other, not to the other QCD final state objects. This should lead to a difference in the amount of soft particle radiation between the  $b$ -quarks produced by the Higgs and the rest of the event.

The CMS  $VH(H \rightarrow b\bar{b})$  analysis makes use of a colour flow variable called pull angle which looks at the distribution of particles within reconstructed jets to discriminate between signal and background events [162, 163]. However, the proposal here would be to construct a new variable which makes uses of the event reconstruction provided by the matrix element method to identify the Higgs candidate and aid the discrimination.

If the Higgs candidate can be identified with the matrix element method, then one can start to study the distribution of charged tracks relative to this object. If the effect of pile-up vertices is not overwhelming, it should be possible to discriminate between signal and background events even further. This should provide additional discrimination as the matrix element method does not use any event information beyond the leading-order objects.

### 8.2.2 Event Quality

The permutation with the largest likelihood has already been identified to provide a level of event reconstruction. It is possible to consider other permutation likelihoods to get information about the quality of that reconstruction. If the permutation with the second largest likelihood is of a comparable size to the permutation with the largest likelihood, it would imply that both permutations offer a similar level of event reconstruction. However, if the second largest likelihood is much smaller than the largest, that provides information about how much that permutation should be trusted.

An event quality variable relating to the kinematic event reconstruction can therefore be defined as

$$Q_{ij} = \frac{\mathcal{L}_i(\mathbf{x}|\boldsymbol{\alpha})}{\mathcal{L}_j(\mathbf{x}|\boldsymbol{\alpha})}. \quad (8.2)$$

This variable could be extended to compare any two permutation likelihoods. This variable  $Q_{10}$  will compare the subleading permutation likelihood with the leading permutation likelihood. The distribution of  $Q_{10}$  will be between 0 and 1 by construction, with events closer to 0 offering more confidence in the permutation selected for the event reconstruction.

Such a variable could be included into a neural network with additional matrix element variables such as  $D_1$  or invariant dijet mass variables. This variable in itself will not offer any

particular discriminating power between the signal and background processes but should be considered as a weighting variable which a neural network should be able to identify as a relevant variable during the training process.

### 8.3 Final Comment

The detailed  $t\bar{t}H$  analysis performed on the Run 1 dataset should ensure that the analysis group is well-prepared for Run 2. The Run 1 analysis developed a strong understanding of the  $t\bar{t}$ +jets background modelling and the use of the tag rate function, both of which were initially problematic for the analysis. It should be expected that a preliminary Run 2 analysis will be produced with between 10 and 15  $\text{fb}^{-1}$  aiming at ICHEP 2016. Whilst this initial analysis will attempt to replicate the standard analysis in Run 1 with the updates to object definitions and the ATLAS event data model, there will be a large amount of scope for improvements to the Run 1 analysis after this first conference deadline. Topics such as the matrix element method, multivariate reconstruction techniques, new region definitions to enhance the analysis sensitivity and potentially continuous  $b$ -tagging will hopefully all combine to ensure that an analysis with between 15 and 25  $\text{fb}^{-1}$  using the total data collected in 2015 and 2016 will be more sensitive to a Standard Model Higgs produced in association with top-quark pairs and therefore be able to determine if this process does indeed exist. If a discovery of this production mechanism can be made, it will provide proof that the Higgs boson couples to top quarks and a direct measurement of the size of the top Yukawa coupling will be possible.

# Bibliography

All references provided are publicly available except in the circumstance where attribution could only be provided using an internal ATLAS document.

- [1] S. L. Glashow, *Partial symmetries of weak interactions*, Nucl. Phys. 22 (1961) 579–588.
- [2] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. 19 (1967) 1264–1266.
- [3] A. Salam, *Weak and electromagnetic interactions*, in *Elementary Particle Physics: Relativistic Groups and Analyticity*, N. Svartholm, ed. Eighth Nobel Symposium, Lerum, May, 1968.
- [4] The LHCb Collaboration, *Observation of the resonant character of the Z(4430)<sup>-</sup> state*, Phys. Rev. Lett. 112 (2014), no. 22, 222002.
- [5] The LHCb Collaboration, *Observation of J/ψp resonances consistent with pentaquark states in λ\_b^0 → J/ψk<sup>-</sup>p decays*, Phys. Rev. Lett. 115 (2015) 072001.
- [6] The ATLAS Collaboration, *Observation of a centrality-dependent dijet asymmetry in lead-lead collisions at √s<sub>NN</sub> = 2.76 TeV with the ATLAS detector at the LHC*, Phys. Rev. Lett. 105 (Dec, 2010) 252303.
- [7] The CMS Collaboration, *Observation and studies of jet quenching in PbPb collisions at √s<sub>NN</sub> = 2.76 TeV*, Phys. Rev. C 84 (Aug, 2011) 024906.
- [8] The ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B 716 (2012), no. 1, 1 – 29.
- [9] The CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B 716 (2012), no. 1, 30 – 61.

- [10] The ATLAS Collaboration, *Evidence for the spin-0 nature of the Higgs boson using ATLAS data*, Phys. Lett. B 726 (2013) 120–144.
- [11] The CMS Collaboration, *Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV*, Phys. Rev. D 92 (Nov, 2014) 012004.
- [12] I. Connelly, J. Thomas-Wilsker, R. Gonçalo, and M. zur Nedden, *Measurement of the b-tagging scaling factors using top-pair events at  $\sqrt{s} = 8$  TeV using the ATLAS detector*, ATL-COM-PHYS-2013-743, CERN, Geneva, Jun, 2013.
- [13] The ATLAS Collaboration, *Search for the Standard Model Higgs boson produced in association with top quarks and decaying into  $b\bar{b}$  in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector*, Eur. Phys. J. C 75 (2015), no. 7, 349.
- [14] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*. Westview Press, 1995.
- [15] F. Halzen and A. D. Martin, *Quarks & Leptons: An Introductory Course in Modern Particle Physics*. Wiley, 1984.
- [16] M. Thomson, *Modern Particle Physics*. Cambridge University Press, 2013.
- [17] S. Weinberg, *The making of the Standard Model*, in *50 Years of Yang-Mills Theory*, G. ’t Hooft, ed. World Scientific, 2005.
- [18] K. Stelle, *Lecture notes on unification*.  
<http://www.imperial.ac.uk/workspace/theoreticalphysics/Public/MSc/Unification/>, 2010.
- [19] G. ’t Hooft and M. J. G. Veltman, *Regularization and renormalization of gauge fields*, Nucl. Phys. B 44 (1972) 189–213.
- [20] G. ’t Hooft, *Why do we need local gauge invariance in theories with vector particles? An introduction*, NATO Sci. Ser. B 59 (1980) 101.
- [21] C. N. Yang and R. L. Mills, *Conservation of isotopic spin and isotopic gauge invariance*, Phys. Rev. 96 (Oct, 1954) 191–195.

- [22] M. Gell-Mann, *A schematic model of baryons and mesons*, Phys. Lett. 8 214 (1964).
- [23] G. Zweig, *An  $SU_3$  model for strong interaction symmetry and its breaking*, in *Developments in the Quark Theory of Hadrons. Volume 1. 1964 - 1978*, D. Lichtenberg and S. P. Rosen, eds., pp. 22–101. Hadronic Press, 1964.
- [24] O. W. Greenberg, *Spin and unitary-spin independence in a paraquark model of baryons and mesons*, Phys. Rev. Lett. 13 (Nov, 1964) 598–602.
- [25] M. Y. Han and Y. Nambu, *Three triplet model with double  $SU(3)$  symmetry*, Phys. Rev. 139 (1965) B1006–B1010.
- [26] R. P. Feynman, *Space-time approach to quantum electrodynamics*, Phys. Rev. 76 (Sep, 1949) 769–789.
- [27] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, *Experimental test of parity conservation in beta decay*, Phys. Rev. 105 (Feb, 1957) 1413–1415.
- [28] N. Cabibbo, *Unitary symmetry and leptonic decays*, Phys. Rev. Lett. 10 (Jun, 1963) 531–533.
- [29] M. Kobayashi and T. Maskawa,  *$CP$ -Violation in the renormalizable theory of weak interaction*, Progress of Theoretical Physics 49 (1973), no. 2, 652–657.
- [30] J. Beringer, *et. al.* (Particle Data Group), *Review of Particle Physics*, Chin. Phys. C 38 (2014) 090001.
- [31] J. Goldstone, *Field theories with superconductor solutions*, Il Nuovo Cimento 19 (1961), no. 1, 154–164.
- [32] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, *Theory of superconductivity*, Phys. Rev. 108 (Dec, 1957) 1175–1204.
- [33] J. Schwinger, *Gauge invariance and mass*, Phys. Rev. 125 (Jan, 1962) 397–398.
- [34] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. 13 (Aug, 1964) 321–323.

- [35] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. 13 (Oct, 1964) 508–509.
- [36] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. 12 (1964), no. 2, 132–133.
- [37] P. W. Higgs, *Spontaneous symmetry breakdown without massless bosons*, Phys. Rev. 145 (1966), no. 4, 1156–1163.
- [38] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global conservation laws and massless particles*, Phys. Rev. Lett. 13 (Nov, 1964) 585–587.
- [39] F. Bezrukov and M. Shaposhnikov, *Why should we care about the top quark Yukawa coupling?*, J. Exp. Theor. Phys. 120 (2015) 335–343.
- [40] The LHC Higgs Cross Section Working Group Collaboration, *Handbook of LHC Higgs cross sections: 1. Inclusive observables*, arXiv:1101.0593.
- [41] The LHC Higgs Cross Section Working Group Collaboration, *Handbook of LHC Higgs cross sections: 3. Higgs properties*, arXiv:1307.1347.
- [42] M. Czakon, P. Fiedler, and A. Mitov, *Total top-quark pair-production cross section at hadron colliders through  $O(\alpha_s^4)$* , Phys. Rev. Lett. 110 (June, 2013) 252004.
- [43] M. Czakon, A. Mitov, and G. Sterman, *Threshold resummation for top-pair hadroproduction to next-to-next-to-leading log*, Phys. Rev. D 80 (Oct., 2009) 074017.
- [44] M. Czakon and A. Mitov, *Top++: A program for the calculation of the top-pair cross-section at hadron colliders*, Comput. Phys. Commun. 185 (2014) 2930.
- [45] F. Cascioli, P. Maierhöfer, N. Moretti, S. Pozzorini, and F. Siegert, *NLO matching for  $t\bar{t}b\bar{b}$  production with massive b-quarks*, Phys. Lett. B 734 (June, 2014) 210–214.
- [46] The ATLAS Collaboration, *Expected performance of the ATLAS experiment - detector, trigger and physics*, CERN-OPEN-2008-020, CERN, Geneva, Dec, 2009.
- [47] The LHC Higgs Cross Section Working Group Collaboration, *LHC HXSWG interim recommendations to explore the coupling structure of a Higgs-like particle*, arXiv:1209.0040.

- [48] The ATLAS Collaboration, *Measurements of the Higgs boson production and decay rates and coupling strengths using  $pp$  collision data  $\sqrt{s} = 7$  and  $8$  TeV in the ATLAS experiment*, CERN-PH-EP-2015-125, CERN, Geneva, Jul, 2015.
- [49] D. Buttazzo, G. Degrassi, P. P. Giardino, G. F. Giudice, F. Sala, A. Salvio, and A. Strumia, *Investigating the near-criticality of the Higgs boson*, JHEP 12 (Dec., 2013) 89.
- [50] S. P. Martin, *A supersymmetry primer*, Adv. Ser. Direct. High Energy Phys. 21 (2010) 1–153.
- [51] The ATLAS Collaboration, *Combined measurements of the mass and signal strength of the Higgs-like boson with the ATLAS detector using up to  $25\text{ fb}^{-1}$  of proton-proton collision data*, ATLAS-CONF-2013-014, CERN, Geneva, Mar, 2013.
- [52] The CMS Collaboration, *Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the Standard Model predictions using proton collisions at 7 and 8 TeV*, Eur. Phys. J. C 75 (Dec, 2014) 212.
- [53] The ATLAS and CMS Collaborations, *Combined measurement of the Higgs boson mass in  $pp$  collisions at  $\sqrt{s} = 7$  and  $8$  TeV with the ATLAS and CMS experiments*, Phys. Rev. Lett. 114 (May, 2015) 191803.
- [54] M. Baak, M. Goebel, J. Haller, A. Hoecker, D. Ludwig, *et. al.*, *Updated status of the global electroweak fit and constraints on new physics*, Eur. Phys. J. C 72 (2012) 2003.
- [55] L. D. Landau, *On the angular momentum of a two-photon system*, Dokl. Akad. Nauk Ser. Fiz. 60 (1948) 207–209.
- [56] C.-N. Yang, *Selection rules for the dematerialization of a particle into two photons*, Phys. Rev. 77 (1950) 242–245.
- [57] The ATLAS Collaboration, *Measurement of the Higgs boson mass from the  $H \rightarrow \gamma\gamma$  and  $H \rightarrow ZZ^* \rightarrow 4\ell$  channels with the ATLAS detector using  $25\text{ fb}^{-1}$  of  $pp$  collision data*, Phys. Rev. D 90 (2014), no. 5, 052004.

- [58] M. Pieri, *Higgs boson coupling measurements at the LHC - first release of ATLAS + CMS Higgs coupling combination*, in *Proceedings of the 3<sup>rd</sup> Annual Large Hadron Collider Physics Conference (LHCP 2015)*. Geneva, Sept, 2015.
- [59] The 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*, ATLAS-CONF-2015-044, CERN, Geneva, Sep, 2015.
- [60] *The CERN accelerator complex*. OPEN-PHO-ACCEL-2013-056, Oct, 2013.
- [61] The ATLAS Collaboration, *The ATLAS experiment at the CERN Large Hadron Collider*, JINST 08003 (2008).
- [62] The ATLAS Collaboration, *Particle identification performance of the ATLAS transition radiation tracker*, ATLAS-CONF-2011-128, CERN, Geneva, Sep, 2011.
- [63] W. Buttinger, *The ATLAS Level-1 trigger system*, Journal of Physics: Conference Series 396 (2012), no. 1, 012010.
- [64] The ATLAS Collaboration, *The ATLAS simulation infrastructure*, Eur. Phys. J. C 70 (2010) 823–874.
- [65] M. Dobbs and J. B. Hansen, *The HepMC C++ Monte Carlo event record for high energy physics*, Comput. Phys. Commun. 134 (2001) 41–46.
- [66] G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, Nucl. Phys. B 126 (1977) 298.
- [67] 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) 641–653.
- [68] V. N. Gribov and L. N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, Sov. J. Nucl. Phys. 15 (1972) 438–450.
- [69] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, *Parton fragmentation and string dynamics*, Physics Reports 97 (1983), no. 2-3, 31–145.

- [70] B. R. Webber, *A QCD model for jet fragmentation including soft gluon interference*, Nucl. Phys. B 238 (1984) 492.
- [71] S. Agostinelli *et. al.*, *Geant4 – A simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 506 (2003), no. 3, 250 – 303.
- [72] The ATLAS Collaboration, *The ATLAS simulation infrastructure*, Eur. Phys. J. C 70 (May, 2010) 823–874.
- [73] The ATLAS Collaboration, *Fast simulation for ATLAS: Atfast-II and ISF*, ATL-SOFT-PROC-2012-065, CERN, Geneva, Jun, 2012.
- [74] The ATLAS Collaboration, *Performance of the Fast ATLAS Tracking Simulation (FATRAS) and the ATLAS Fast Calorimeter Simulation (FastCaloSim) with single particles*, ATL-SOFT-PUB-2014-001, CERN, Geneva, Mar, 2014.
- [75] R. Frühwirth, W. Waltenberger, and P. Vanlaer, *Adaptive vertex fitting*, CMS-NOTE-2007-008, CERN, Geneva, Mar, 2007.
- [76] The ATLAS Collaboration, *Performance of primary vertex reconstruction in proton-proton collisions at  $\sqrt{s} = 7 \text{ TeV}$  in the ATLAS experiment*, ATLAS-CONF-2010-069, CERN, Geneva, Jul, 2010.
- [77] The ATLAS Collaboration, *Number of interactions per bunch-crossing*. <https://twiki.cern.ch/twiki/bin/view/ATLASPublic/LuminosityPublicResults>. Revision 44, accessed 24 Aug 2015.
- [78] The ATLAS Collaboration, *Luminosity determination in pp collisions at  $\sqrt{s} = 7 \text{ TeV}$  using the ATLAS detector at the LHC*, Eur. Phys. J. C 71 (2011) 1630.
- [79] The ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data*, ATLAS-CONF-2014-032, CERN, Geneva, Jun, 2014.
- [80] The ATLAS Collaboration, *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data*, Eur. Phys. J. C 74 (2014), no. 11, 3130.

- [81] B. Acharya *et. al.*, *Object selection and calibration, background estimations and MC samples for the Winter 2013 top quark analyses with 2012 data*, ATL-COM-PHYS-2013-088, CERN, Geneva, Jan, 2013.
- [82] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, *Longitudinally-invariant  $k_T$ -clustering algorithms for hadron-hadron collisions*, Nucl. Phys. B 406 (1993), no. 1-2, 187–224.
- [83] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, Phys. Rev. D 48 (Oct., 1993) 3160–3166.
- [84] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, *Better jet clustering algorithms*, JHEP 8 (Aug., 1997).
- [85] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, in *Proceedings of Workshop on Monte Carlo Generators for HERA Physics*. Hamburg, Germany, 1999.
- [86] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- $k_T$  jet clustering algorithm*, JHEP 0804 (2008) 063.
- [87] G. P. Salam and G. Soyez, *A practical seedless infrared-safe cone jet algorithm*, JHEP 5 (May, 2007) 86.
- [88] The ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton-proton collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector*, Eur. Phys. J. C75 (2015), no. 1, 17.
- [89] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Phys. Lett. B 659 (2008) 119–126.
- [90] The ATLAS Collaboration, *Pile-up subtraction and suppression for jets in ATLAS*, ATLAS-CONF-2013-083, CERN, Geneva, Aug, 2013.
- [91] The ATLAS Collaboration, *Data-driven determination of the energy scale and resolution of jets reconstructed in the ATLAS calorimeters using dijet and multijet events at  $\sqrt{s} = 8$  TeV*, ATLAS-CONF-2015-017, CERN, Geneva, Apr, 2015.

- [92] D. W. Miller, A. Schwartzman, and D. Su, *Jet-vertex association algorithm*, ATL-COM-PHYS-2008-008, CERN, Geneva, Jan, 2008.
- [93] The ATLAS Collaboration, *Performance of b-jet identification in the ATLAS experiment*. ATL-COM-PHYS-2012-824; to be published, 2015.
- [94] G. Piacquadio and C. Weiser, *A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS*, Journal of Physics: Conference Series 119 (2008), no. 3, 032032.
- [95] The ATLAS Collaboration, *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*, ATLAS-CONF-2014-046, CERN, Geneva, Jul, 2014.
- [96] The ATLAS Collaboration, *Measurement of the b-tag efficiency in a sample of jets containing muons with  $5 \text{ fb}^{-1}$  of data from the ATLAS detector*, ATLAS-CONF-2012-043, CERN, Geneva, Mar, 2012.
- [97] The D0 Collaboration, *b-Jet Identification in the D0 Experiment*, Nucl. Instrum. Meth. A620 (2010) 490–517.
- [98] The ATLAS Collaboration, *Measuring the b-tag efficiency in a top-pair sample with  $4.7 \text{ fb}^{-1}$  of data from the ATLAS detector*, ATLAS-CONF-2012-097, CERN, Geneva, Jul, 2012.
- [99] The ATLAS Collaboration, *Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment*, ATLAS-CONF-2014-004, CERN, Geneva, Feb, 2014.
- [100] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, JHEP 0206 (2002) 029.
- [101] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP 0411 (2004) 040.
- [102] G. Corcella, I. G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M. H. Seymour, and B. R. Webber, *HERWIG 6: An event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, JHEP 01 (2001) 010.

- [103] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. 178 (2008) 852–867.
- [104] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP 07 (2003) 001.
- [105] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 physics and manual*, JHEP 05 (2006) 026.
- [106] P. Z. Skands, *Tuning Monte Carlo generators: The Perugia tunes*, Phys. Rev. D 82 (2010) 074018.
- [107] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin, and C.-P. Yuan, *New parton distributions for collider physics*, Phys. Rev. D 82 (2010) 074024.
- [108] The ATLAS Collaboration, *New ATLAS event generator tunes to 2010 data*, ATL-PHYS-PUB-2011-008, CERN, Geneva, Apr, 2011.
- [109] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky, and W. K. Tung, *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP 07 (2002) 012.
- [110] The ATLAS Collaboration, *Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data*, Eur. Phys. J. C 72 (Oct, 2011) 1909–1954.
- [111] L. Iconomidou-Fayard, K. Lohwasser, T. Serre, and E. Tiouchichine, *Electron efficiency measurements in early 2012 data*, ATL-COM-PHYS-2012-783, CERN, Geneva, Jun, 2012.
- [112] The ATLAS Collaboration, *Muon reconstruction efficiency in reprocessed 2010 LHC proton-proton collision data recorded with the ATLAS detector*, ATLAS-CONF-2011-063, CERN, Geneva, Apr, 2011.
- [113] The ATLAS Collaboration, *Jet energy scale and its systematic uncertainty in proton-proton collisions at  $\sqrt{s} = 7 \text{ TeV}$  with ATLAS 2011 data*, ATLAS-CONF-2013-004, CERN, Geneva, Jan, 2013.

- [114] The ATLAS Collaboration, *Jet energy resolution in proton-proton collisions at  $\sqrt{s} = 7$  TeV recorded in 2010 with the ATLAS detector*, Eur. Phys. J. C 73 (2013), no. 3, 2306.
- [115] M. Cacciari, G. P. Salam, and G. Soyez, *The catchment area of jets*, JHEP 04 (2008) 005.
- [116] R. Hawkings, *Measurement of the b-tagging efficiency of the MV1 algorithm in pp collisions at  $\sqrt{s} = 8$  TeV using  $e\mu$  dilepton  $t\bar{t}$  events*, ATL-COM-PHYS-2013-381, CERN, Geneva, Apr, 2013.
- [117] The ATLAS Collaboration, *Measurement of the b-tag efficiency in a sample of jets containing muons with  $5 \text{ fb}^{-1}$  of data from the ATLAS detector*, ATLAS-CONF-2012-043, CERN, Geneva, Mar, 2012.
- [118] I. Burmeister, I. Connelly, A. Ferretto, F. Filthaut, R. Gonçalo, R. Hawkings, H. Esch, F. Parodi, D. Sidorov, C. Schiavi, J. Thomas-Wilsker, J. Yu, and M. Zur Nedden, *Measurement of the b-tagging performance in the 2012 ATLAS data*, ATL-COM-PHYS-2013-456, CERN, Geneva, Apr, 2013.
- [119] K. Cranmer, G. Lewis, L. Moneta, A. Shibata, and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, CERN-OPEN-2012-016, CERN, Geneva, Jan, 2012.
- [120] L. Moneta, K. Belasco, K. S. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke, and M. Wolf, *The RooStats project*, PoS ACAT2010 (2010) 057.
- [121] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP 06 (2010) 043.
- [122] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP 11 (2007) 070.
- [123] G. Bevilacqua, M. Czakon, M. V. Garzelli, A. van Hameren, A. Kardos, C. G. Papadopoulos, R. Pittau, and M. Worek, *HELAC-NLO*, Comput. Phys. Commun. 184 (2013) 986–997.

- [124] 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) 56–74.
- [125] The ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, ATL-PHYS-PUB-2012-003, CERN, Geneva, Aug, 2012.
- [126] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, *et. al.*, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP 1407 (2014) 079.
- [127] The ATLAS Collaboration, *ATLAS tunes of Pythia 6 and Pythia 8 for MC11*, ATL-PHYS-PUB-2011-009, CERN, Geneva, Jul, 2011.
- [128] The ATLAS Collaboration, *Measurements of normalized differential cross sections for  $t\bar{t}$  production in  $pp$  collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector*, Phys. Rev. D 90 (2014), no. 7, 072004.
- [129] F. Cascioli, P. Maierhöfer, and S. Pozzorini, *Scattering amplitudes with Open Loops*, Phys. Rev. Lett. 108 (Mar., 2012) 111601.
- [130] T. Gleisberg, S. Höche, F. Krauss, M. Schönher, S. Schumann, F. Siegert, and J. Winter, *Event generation with SHERPA 1.1*, JHEP 2 (Feb., 2009) 7.
- [131] A. Shibata and B. Clement, *Tagging rate function b-tagging*, ATL-PHYS-PUB-2007-011, CERN, Geneva, August, 2007.
- [132] C. M. Bishop, *Neural Networks for Pattern Recognition*. Clarendon Press, 1995.
- [133] J. Friedman, *An overview of computational learning and function approximation*, in *From Statistics to Neural Networks. Theory and Pattern Recognition Applications*, V. Cherkassky, J. Friedman, and H. Wechsler, eds. Springer-Verlag, 1994.
- [134] M. Feindt and U. Kerzel, *The NeuroBayes neural network package*, Nucl. Instrum. Meth. A559 (2006) 190–194.
- [135] <phi-t>, *The NeuroBayes user's guide*.  
[http://neurobayes.phi-t.de/nb\\_doc/NeuroBayes-HowTo.pdf](http://neurobayes.phi-t.de/nb_doc/NeuroBayes-HowTo.pdf), 2010.

- [136] M. Feindt, *A neural bayesian estimator for conditional probability densities*, arXiv/physics:0402093.
- [137] D. J. C. MacKay, *A practical bayesian framework for backpropagation networks*, Neural Computation 4 (May, 1992) 448–472.
- [138] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C 71 (Feb., 2011) 1554.
- [139] A. L. Read, *Modified frequentist analysis of search results (The CL(s) method)*, in *Workshop on confidence limits, CERN, Geneva, Switzerland, 17-18 Jan 2000: Proceedings*. 2000.
- [140] The D0 Collaboration, *Measurement of the top quark mass in the dilepton channel*, Phys. Rev. D 60 (Jul, 1999) 052001.
- [141] The D0 Collaboration, *A precision measurement of the mass of the top quark*, Nature (June, 2004) 638–642.
- [142] The D0 Collaboration, *Measurement of the top quark mass in the lepton + jets final state with the matrix element method*, Phys. Rev. D 74 (Nov, 2006) 092005.
- [143] The CDF Collaboration, *Precise measurement of the top-quark mass in the lepton + jets topology at CDF II*, Phys. Rev. Lett. 99 (Oct, 2007) 182002.
- [144] The CDF Collaboration, *Top quark mass measurement in the lepton plus jets channel using a modified matrix element method*, Phys. Rev. D 79 (Apr, 2009) 072001.
- [145] The CMS Collaboration, *Measurement of the properties of a Higgs boson in the four-lepton final state*, Phys. Rev. D 89 (2014), no. 9, 092007.
- [146] The CMS Collaboration, *Search for a Standard Model Higgs boson produced in association with a top-quark pair and decaying to bottom quarks using a matrix element method*, Eur. Phys. J. C 75 (Feb, 2015) 251–292.
- [147] T. Martini and P. Uwer, *Extending the matrix element method beyond the Born approximation: Calculating event weights at next-to-leading order accuracy*, arXiv:1506.08798.

- [148] J. M. Campbell, W. T. Giele, and C. Williams, *The matrix element method at next-to-leading order*, JHEP 11 (2012) 043.
- [149] J. Neyman and E. S. Pearson, *On the problem of the most efficient tests of statistical hypotheses*, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character 231 (1933) 289–337.
- [150] C. Brachem, *Studies for a top quark mass measurement using the matrix element method in the semileptonic channel*, Master’s thesis, Georg-August-Universität Göttingen, Feb, 2012.
- [151] G. P. Lepage, *A new algorithm for adaptive multidimensional integration*, Journal of Computational Physics 27 (1978) 192–203.
- [152] G. P. Lepage, *VEGAS: An adaptive multi-dimensional integration program*, Cornell Preprint CLNS (1980) 80–447.
- [153] M. Galassi *et. al.*, *GNU Scientific Library Reference Manual*.  
<http://www.gnu.org/software/gsl/>.
- [154] M. R. Whalley, D. Bourilkov, and R. C. Group, *The Les Houches Accord PDFs (LHAPDF) and LHAGLUE*, in *HERA and the LHC: A workshop on the implications of HERA for LHC physics. Proceedings, Part B*. 2005.
- [155] J. Pumplin, D. R. Stump, J. Huston, H.-L. Lai, P. Nadolsky, and W.-K. Tung, *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP 7 (July, 2002) 12.
- [156] J. Erdmann, S. Guindon, K. Kroeninger, B. Lemmer, O. Nackenhorst, *et. al.*, *A likelihood-based reconstruction algorithm for top-quark pairs and the KLFitter framework*, Nucl. Instrum. Meth. A748 (2014) 18–25.
- [157] The ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, Eur. Phys. J. C 74 (2014), no. 10, 3071.
- [158] Wolfram Research Inc., *Mathematica*, 2012. Version 9.0.

- [159] L. Sonnenschein, *Analytical solution of  $t\bar{t}$  dilepton equations*, Phys. Rev. D 73 (Mar, 2006) 054015.
- [160] A. Bredenstein, A. Denner, S. Dittmaier, and S. Pozzorini, *NLO QCD corrections to  $pp \rightarrow t\bar{t}b\bar{b} + X$  at the LHC*, Phys. Rev. Lett. 103 (2009) 012002.
- [161] CERN Press Office, *LHC Season 2: facts & figures*.  
<http://press.web.cern.ch/backgrounder/lhc-season-2-facts-figures>, Mar, 2015.
- [162] The 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. D 89 (2014), no. 1, 012003.
- [163] J. Gallicchio and M. D. Schwartz, *Seeing in color: Jet superstructure*, Phys. Rev. Lett. 105 (Jul, 2010) 022001.