The hunt for the Higgs boson: the $WW \to \ell\nu\ell\nu$ final state at the ATLAS detector

by

Shannon R. Walch

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Physics)
in The University of Michigan
2012

Doctoral Committee:

Professor Jianming Qian, Chair
Professor Rudolf P. Thun
Associate Professor Katsuo Kurabaiyashi
Associate Professor Finn Larsen
Assistant Professor Junjie Zhu

© Shannon R. Walch 2012

All Rights Reserved

Dedicated to my parents.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support of many scientists in my collaboration. My gratitude goes to my adviser, Jianming Qian, for his guidance throughout this work. It was a priviledge to be introduced to particle physics research by his example, and I thank him for his help from my first basic questions to finishing this dissertation. I also would like to acknowledge the amazing team of researchers he brought together at the University of Michigan who I was fortunate to work with. I thank Jonas and Gemma for helping me learn the finer points of performing an analysis and looking over my shoulder to point out bugs when I was blind to them. I also thank my fellow students, Hao, Devin, Aaron, and Jonathon, for their ever-willing help in getting code to compile and ROOT to obey.

I am grateful to the University of Michigan Rackham Graduate School and Physics Department for their support during my entire graduate program and particularly my time resident at CERN. I know it cannot be easy to monitor students 4,000 miles away, but I would not have learned nearly as much without the opportunity. Thank you, particularly to Beth and Christina! My thanks also go to Dan, Niels, Nir, and Dinos for their guidance and help with my service work and to the members of Higgs Sub Group 3 for their input during my time with them.

Finally, I would like to thank my family, particularly my parents, for their unwavering support as my research led me across first the country and then an ocean. I couldn't have done it without you. Finally, thank you, Chester, for believing that I could excel at whatever I attempted.

PREFACE

The ATLAS Collaboration has over 3000 members from 38 countries, all of whom contribute to the maintenance and operation of the detector, the understanding of the data, and the study of physics processes. No results could be produced without the combined expertise of this large team. The analysis presented in this dissertation draws upon studies and results produced by many teams over several years and acknowledges this with gratitude.

I helped with generation of samples for analysis and comparisons of cutflows as code for these tasks was being developed and merged into one framework available to the entire $H \to WW \to \ell\nu\ell\nu$ analysis group. The plots and cutflows shown in Chapters V, VII, VIII, and the first part of IX were produced by me using these common tools. My main responsibility was to develop methods for estimation of the Z/γ^*+jets background discussed in Chapter VIII. I developed different methods to cross-check each other and performed additional studies to verify the performance of these methods and estimate their associated systematic uncertainties. The group working on this analysis was large and many members were interested in using these results for their own studies. To facilitate this, I incorporated several of these methods into macros that were embedded into the common framework; I later expanded this code to include calculations for the WW and top control regions as well. These were used in group-wide systematic uncertainty studies, derivation of other backgrounds' estimates, and exploration of multivariate analysis techniques.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

PREFACE iv

LIST OF FIGURES viii

LIST OF TABLES xii

LIST OF ABBREVIATIONS xv

ABSTRACT xvi

CHAPTER

I. Introduction 1

II. The Standard Model, Electroweak Symmetry Breaking, and the Higgs Boson 3

2.1 The Standard Model 4
2.2 Electroweak Symmetry Breaking 7
2.3 Searches for the Higgs Boson 11
2.3.1 Decay Channels of the Higgs Boson 11
2.3.2 Limits from Direct Searches 14
2.3.3 Constraints from Electroweak Precision Measurements 16

III. LHC and ATLAS 18

3.1 The Large Hadron Collider 18
3.1.1 Higgs Boson Production at the LHC 21
3.2 ATLAS Detector 24

3.2.1 Inner Detector . . . 27
3.2.2 Calorimeters . . . 29
3.2.3 Muon Spectrometer . . . 31
3.2.4 Magnet Systems . . . 34
3.2.5 Triggering . . . 35

IV. The ATLAS Monte Carlo and Data Chain . . . 37

4.1 Background Processes . . . 38
4.2 Monte Carlo Generation . . . 41
4.3 Simulation of the ATLAS Detector . . . 44
4.4 Digitization . . . 45
4.5 Object Reconstruction . . . 45
4.5.1 Track and Primary Vertex Reconstruction . . . 46
4.5.2 Electron Reconstruction and Identification . . . 47
4.5.3 Muon Reconstruction and Identification . . . 50
4.5.4 Jet Reconstruction . . . 52
4.5.5 Identification of b-Jets . . . 54
4.5.6 Missing Energy Reconstruction . . . 56

V. Event Selection . . . 58

5.1 Data Selection and Trigger Requirements . . . 59
5.2 Lepton and Jet Selection . . . 62
5.3 Topological and Jet Selections . . . 64

VI. Efficiency Corrections to Monte Carlo Simulation . . . 69

6.1 The Tag and Probe Method . . . 70
6.2 Lepton Trigger Efficiencies . . . 71
6.3 Lepon Reconstruction Efficiencies . . . 72
6.4 Energy and Resolution Smearing . . . 73
6.5 b-Tagging Efficiency . . . 74
6.6 Drell-Yan $p_T^{\ell\ell}$ Reweighting . . . 76
6.7 Systematic Uncertainties . . . 77

VII. The WW, Top, and W+jets Backgrounds . . . 79

7.1 WW Continuum Background . . . 80
7.2 Top Background . . . 82
7.2.1 Top Zero Jet Background . . . 83
7.2.2 Top One and Two Jet Backgrounds . . . 84
7.3 W+jets Background . . . 85

VIII. Z Background Estimation Methods . . . 89

8.1 Scale Factor method 89
8.1.1 Procedure and Results 89
8.1.2 Cross-checks and systematic uncertainties 92
8.2 ABCD method 94
8.2.1 Procedure and Results 96
8.2.2 Region Dependencies 102
8.2.3 ABCD Closure Tests 104
8.2.4 Method Adaptations 106
8.2.5 ABCD Systematic Uncertainties 108
8.3 Ratio of Ratio Method 110
8.3.1 Procedure, Results, and Systematic Uncertainties . 110
8.4 Comparison of Estimation Methods 113

IX. Statistical Methods and Observed Limits 118

9.1 Final Results 118
9.2 Statistical Methods and Limits 126

X. Conclusion 132

BIBLIOGRAPHY 134

LIST OF FIGURES

Figure

2.1 The potential well of the scalar field ϕ with $\mu^2 < 0$ and $\lambda > 0$. 9

2.2 The branching ratios of several Standard Model Higgs boson decay channels. 12

2.3 The expected and observed limits of the Higgs boson's mass for the two Tevatron experiments using up to 10 $\text{fb}^{-1}$of data. 15

2.4 The χ^2 fit of precision electroweak data using measurements from LEP, the Tevatron, and other experiments as a function of Standard Model Higgs mass 17

3.1 The LHC complex with its four collision point detectors, ATLAS, CMS, ALICE, and LHCb, shown relative to the Geneva area. 19

3.2 The CERN accelerator complex showing the chain of accelerators that deliver beams to the LHC. 20

3.3 Diagram of the structure of a hadron collision; the central hard collision and its daughter particles are shown in red, gluon radiation is shown in blue, interactions amongst the recoiling particles is in purple, and final state hadrons are shown in green. 22

3.4 Higgs production cross-sections for the major Higgs boson production processes as a function of Higgs mass for $\sqrt{s} = 7$ TeV. 23

3.5 Feynmann diagrams for the major production modes of the Higgs boson at the LHC. 24

3.6 The ATLAS detector and its major subsystems. 25

3.7 A cross-section of the ATLAS inner detector. 28

3.8 The calorimeters of the ATLAS detector. 29

3.9 The ATLAS Muon Spectrometer. 32

4.1 Tree-level diagrams of the diboson production channels at the LHC. 38

4.2 Tree-level diagrams for WW production from gluon fusion processes at the LHC. 39

4.3 Tree-level diagrams for top quark pair production (a), associated production of top quarks (b), and the t-channel (c) and s-channel (d) production modes for single top quarks. 40

4.4 Diagrams showing the production of vector boson with jet events at the LHC. 41

4.5 Diagram of barrel EM calorimeter module showing different layers and cell dimensions. 48

4.6 Diagram showing the stations crossed by muon tracks with different values of η. 51

4.7 The structure of jets originating from light quarks or gluons compared to that of a b-jet. 55

5.1 The cumulative luminosity delivered to ATLAS (green) and recorded by the detector (yellow) as a function of time for the 2011 data-taking. 60

5.2 $m_{\ell\ell}$ distributions in the ee (left), $e\mu$ (center), and $\mu\mu$ (right) channels after selecting events with two opposite signed leptons, before removal of low mass resonances and the Z peak.l 64

5.3 $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ distributions for the ee (left), $e\mu$ (center), and $\mu\mu$ (right) channels after $m_{\ell\ell}$ selections. 65

5.4 The jet multiplicity distributions for ee (left), $e\mu$ (center), and $\mu\mu$ (right) channels after the $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ requirements. 66

6.1 Distributions of $p_T^{\ell\ell}$ in the $\mu\mu$ zero jet channel in the range $12 < m_{\ell\ell} < 50$ GeV before (left) and after (right) reweighting in the $p_T^{\ell\ell}$ vs $m_{\ell\ell}$ plane. 76

7.1 Transverse mass distributions in the low Higgs mass WW control regions for events with zero jets (left), one jet (center) and at least two jets (right). 81

7.2 Distributions of transverse mass for the top one (left) and two inclusive (right) jet control regions. 85

8.1 $m_{\ell\ell}$v$E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ plane for the $\mu\mu$ 1 jet channel after the jet multiplicity cut; regions C and D are used for calculation of the scale factors. . . 90

8.2 $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ distributions within the Z-window for the $\mu\mu$ zero jet channels after the zero jet requirement (left) and $p_T^{\ell\ell}$ cut (right). 92

8.3 Calculated scale factors and estimates in the $\mu\mu$ zero jet channel after the $p_T^{\ell\ell}$ cut for different $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ thresholds. 93

8.4 The regions used for the ABCD estimate of the Z/γ^*+jets contamination in the signal region. 96

8.5 Ratios between regions B and D in data (points) and Z/γ^*+jets Monte Carlo (histograms) after the $p_T^{\ell\ell}$ cut in the zero jet channel (left) and after the p_T^{Tot} cut in the one jet channel (right). 102

8.6 ABCD estimates in data for the $p_T^{\ell\ell}$ cuts (zero jet channels, left plot) and the p_T^{Tot} cuts (one jet channels, right plot) as a function of lower bound in $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$. 103

8.7 The regions used for data-driven closure tests shown in the $\mu\mu$ one jet channel. 105

9.1 Distribution of transverse mass in the zero jet channel after all selections. 120

9.2 Transverse mass distribution in the one jet channel after all selections. 122

9.3 The original m_T distribution (left) and mapped distribution (right) each in five bins; the distributions shown are for the zero jet channel. 126

9.4 Expected (dashed) and observed (solid) 95% confidence level upper limits on the ratio between calculated and Standard Model cross-sections as a function of m_H. 129

9.5 The expected (dashed) and observed (solid) p-values, or the probability of the Standard Model backgrounds producing the observed distribution in data, as a function of m_H. 130

9.6 Expected (dashed) and observed (solid) 95% confidence level upper limits on the ratio between calculated and Standard Model cross-sections as a function of m_H in each flavor and jet multiplicity channel. 131

LIST OF TABLES

Table

2.1 The names, masses, electric charge, family structure, and interactions of the Standard Model particles. 5

2.2 The mass, charge, and interaction carried at low energy of the Standard Model vector bosons. 6

3.1 Performance goals of the main ATLAS detector subsystems. 25

4.1 Cross-sections of background processes for $\sqrt{s} = 7$ TeV, including branching fractions for processes with specific decay modes; processes without a specific decay mode are inclusive. 43

4.2 Percentage of Monte Carlo generated events generated with different pile-up conditions to correspond to data-taking periods in 2011. . . 44

5.1 Single lepton triggers used for this analysis. 61

7.1 Scale factors derived from the WW control regions. 82

7.2 Scale factors derived from the top one jet control region. 84

7.3 Criteria for definition of anti-identied leptons used in fake lepton method estimation of W+jets background. 86

8.1 Scale factors and scale factor method estimates calculated for the Z/γ^*+jets background in the ee and $\mu\mu$ channels. 91

8.2 Contributions to the systematic uncertainties of the scale factor method in the ee one jet channel after the p_T^{Tot} cut. 95

8.3 Estimates from the ABCD method for the low Higgs mass signal region in the ee and $\mu\mu$ channels. 98

8.4 Observed data events (after background subtraction) after the $p_T^{\ell\ell}$ and $\Delta\phi_{\ell\ell}$ cuts in the ee zero jet channel in each ABCD region. 99

8.5 Estimates from the CDEF method for the low Higgs mass WW control region in the ee and $\mu\mu$ channels. 100

8.6 Estimates from the ABCD method for the intermediate Higgs mass signal region in the ee and $\mu\mu$ 0 and 1 jet channels. 101

8.7 Estimates from the CDEF method for the intermediate Higgs mass WW control region in the ee and $\mu\mu$ 0 and 1 jet channels. 101

8.8 Closure tests comparing A to BC/D in the ee and $\mu\mu$ 0 and 1 jet channels. 104

8.9 Results of the data-driven closure tests from data and Z/γ^*+jets Monte Carlo for the ee and $\mu\mu$ channels using the abcd regions as shown in Figure 8.7. 106

8.10 Results and closure tests for the alternate A'B'CD method in the ee and $\mu\mu$ channels. 107

8.11 ABCD estimate results in the ee 0 jet channel after the $p_T^{\ell\ell}$ cut for each of the object systematic variations. 109

8.12 Region statistical (left three columns) and systematic (right three columns) uncertainties for the ABCD and CDEF methods. 111

8.13 Results from the ratio of ratio estimates of Z/γ^*+jets background for the ee and $\mu\mu$ zero and one jet channels. 112

8.14 The contributing systematic uncertainties for the ration of ratio method estimates after the $p_T^{\ell\ell}$ cut in the $\mu\mu$ zero jet channel. 113

8.15 Comparison of results from all methods of estimating Z/γ^*+jets contamination throughout the cutflow in the ee channels. 114

8.16 Comparison of results from all methods of estimating Z/γ^*+jets contamination throughout the cutflow in the $\mu\mu$ channels. 115

9.1 The expected number of background events and the observed number of data events for preselection and zero jet events. 121

9.2 The expected number of background events and the observed number of data events for the one jet events after the jet multiplicity cut. . 123

9.3 The expected number of background events and the observed number of data events for two jet events after the jet multiplicity cut. 125

LIST OF ABBREVIATIONS

LEP Large Electron Position Collider

PDF Parton Distribution Function

ATLAS A Toroidal LHC ApparatuS

LHC Large Hadron Collider

LEP Large Electron Positron

SCT Semi-Conductor Tracker

TRT Transition Radiation Tracker

RPC Resistive Plate Chambers

TGC Thin Gap Chambers

MDT Monitored Drift Tubes

CSC Cathode Strip Chambers

RoI Region of Interest

QCD Quantum ChromoDynamics

ABSTRACT

The hunt for the Higgs boson: the $WW \to \ell\nu\ell\nu$ final state at the ATLAS detector

by

Shannon R. Walch

Chair: Jianming Qian

Electroweak symmetry breaking stands as the last sector of the standard model of particle physics to be experimentally verified. The electroweak symmetry must be broken to incorporate massive particles into the theory and accurately reflect particle properties. The search for the Higgs boson is the search for the defining prediction of the Higgs mechanism, the simplest method to break electroweak symmetry. This thesis presents results on the search for the standard model Higgs boson in the $WW \to \ell\nu\ell\nu$ final state using 4.7 fb^{-1} of data collected by the ATLAS detector in the 2011 data-taking at the LHC.

The search is performed using a cut-based analysis to define the final signal regions in three lepton flavor channels (ee,$e\mu$,$\mu\mu$) and three jet multiplicities (0, 1, and ≥ 2). Backgrounds are estimated using several Monte Carlo simulation and data-driven techniques. The estimate of the Z+jets background is presented in particular detail, with three independent methods serving as comparisons to each other as well as studies focusing on particular aspects of the estimates. Fits of the final distributions of transverse mass are used for limit-setting in the CL_s method. The mass range of $100 < m_H < 600$ GeV was searched for evidence of the existence of the standard

model Higgs boson. No significant excesses above the prediction of known standard model processes was identified. Limits were then set at the 95% confidence level excluding masses for the Higgs boson in the range $131 < m_H < 260$ GeV.

CHAPTER I

Introduction

Since the discovery of the first fundamental particle in 1895, the knowledge of particle properties and their interactions has been gathered into one mathematical theory called the Standard Model. To date, no experimental results have disproved its predictions, but tests of the Standard Model remain to be completed. The electroweak symmetry of the theory must be broken to incorporate masses and accurately predict the observed interactions between particles, but the simplest method of breaking electroweak symmetry predicts a scalar boson that has not yet been discovered. This particle, called the Higgs boson [1], is the last undiscovered particle in the Standard Model and its discovery or exclusion is necessary to understand the nature of electroweak symmetry breaking.

While the Higgs mechanism with its characteristic Higgs boson is the simplest method to explain the breaking of electroweak symmetry, it is not the only method. Theories of physics beyond the Standard Model frequently alter the nature of electroweak symmetry breaking and produce different distinguishing particles. Supersymmetry predicts a minimum of five scalar and pseudoscalar particles in place of the Higgs boson [2], while extra-dimensional theories require families of vector bosons to perform the same function [3]. Studying the Higgs sector may yield valuable clues about what particle behaviors may exist beyond those observed and incorporated in

the Standard Model, giving extra importance to this search.

The search for the Standard Model Higgs boson has been carried out both directly and indirectly at numerous previous accelerators and detectors, including the experiments at LEP [4] and the Tevatron [5]. Today, the Large Hadron Collider operated by CERN outside Geneva, Switzerland continues the hunt by probing the highest energy collisions ever produced with its experiments. This dissertation reports on the search for the Higgs boson in 4.7 fb^{-1} of data collected from $\sqrt{s} = 7$ TeV proton-proton collisions during 2011. The data were collected by the ATLAS detector, one of the multipurpose detectors located on the LHC's collision points. The focus is on the $H \to WW \to \ell\nu\ell\nu$ decay mode and the analysis uses cut-based techniques. Backgrounds are estimated using a variety of methods relying on both data and Monte Carlo simulation, before the CL_s method of statistical fitting is used to assess the significance of the observed results.

This thesis is structured as follows: The theoretical structure of the Standard Model and motivation for the Higgs mechanism is presented in Chapter II. Chapter III describes the LHC machine and the ATLAS detector, and the impact these apparatuses have on the search for the Higgs boson. The software chain used to produce simulations and reconstruct data objects for analysis is described in Chapter IV. The requirements used to define the final signal regions are presented in Chapter V. As simulation cannot exactly mimic the features of data through the changing conditions at the LHC, efficiency corrections and uncertainties are described in Chapter VI. Chapters VII and VIII present techniques used to estimate the contamination from background processes. The final results are presented in Chapter IX, with the conclusions and outlook of this work discussed in Chapter X.

CHAPTER II

The Standard Model, Electroweak Symmetry Breaking, and the Higgs Boson

The Standard Model of particle physics describes particle interactions and characteristics with a remarkable degree of accuracy; its predictions of the interactions of the twelve fermions are in agreement with every direct experimental test to date. However, it fails to address several outstanding questions about particle properties, namely how the electroweak symmetry is broken and its related question of how to incorporate particle masses into the Standard Model. The simplest method to break the electroweak symmetry is called the Higgs mechanism, and it requires the existence of a scalar neutral particle called the Higgs boson. This particle is considered to be the last missing particle of the Standard Model, and the search for it is the topic of this dissertation. A brief review of the content and history of the Standard Model is covered in Section 2.1. The theory of electroweak symmetry breaking via the Higgs mechanism is covered in Section 2.2, while the current status of searches for the Higgs boson is presented in Section 2.3.

2.1 The Standard Model

The Standard Model is a $SU_C(3) \times SU_L(2) \times U_Y(1)$ gauge theory that seeks to describe the fundamental matter particles and their interactions. The fundamental particles consist of twelve spin-$\frac{1}{2}$ fermions grouped into three families. Each family contains a charged lepton, a neutral lepton, and two quarks; each particle has the same interactions as its correspondent in the other families, but with different masses. The lightest charged lepton is the electron, discovered in 1895 [6]; the muon discovered in 1936 [7, 8] and the tau lepton discovered in 1975 [9] are its correspondents in the second and third families. Each charged lepton is paired with a neutral lepton called a neutrino; neutrinos exist in three flavors, and were discovered according their flavor-specific interactions [10–12]. Quarks, unlike leptons, exist only in bound states with one or two other quarks; these states are called hadrons. The first family quarks, the up and down, and the lighter second family quark, the strange quark, were indirectly observed in deep inelastic scattering experiments in 1969 [13, 14], which validated the quark model of hadron structure. The other three quarks were discovered as experiments reached the energy levels necessary for their production: the charm quark in 1974 [15, 16], the bottom quark in 1977 [17], and the top quark in 1995 [18, 19]. The masses and family structure of these particles is summarized in Table 2.1. Each of the fermions also has a anti-particle partner, a particle with the same mass and spin but opposite values of all other quantum numbers such as electric charge.

The greatest strength of the Standard Model lies in accurately predicting the interactions of all observed particles. The particles interact by exchanging spin-1 vector bosons. The electromagnetic interaction is mediated by the massless photon; the photon's relatively strong coupling to charged fermions at low energies and masslessness means the electromagnetic interaction can propogate long distances. Hence the photon was discovered long before the structure of the Standard Model was created. By comparison, the weak force, responsible for radioactive decays, produces interactions

Symbol (Name)	Mass	Electric Charge	Interactions
First Family			
e (electron)	$0.511 \pm 1.3 \times 10^{-8}$ MeV	-e	Electromagnetic, Weak
ν_e (e neutrino)	< 2 eV	0	Weak
u (up quark)	1.7-3.1 MeV	$\frac{2}{3}$e	Electromagnetic, Strong, Weak
d (down quark)	4.1-5.7 MeV	-$\frac{1}{3}$e	Electromagnetic, Strong, Weak
Second Family			
μ (muon)	$105.7 \pm 4 \times 10^{-6}$ MeV	-e	Electromagnetic, Weak
ν_μ (μ neutrino)	< 2 eV	0	Weak
c (charm quark)	$1.29^{+0.05}_{-0.11}$ GeV	$\frac{2}{3}$e	Electromagnetic, Strong, Weak
s (strange quark)	100^{+30}_{-20} MeV	-$\frac{1}{3}$e	Electromagnetic, Strong, Weak
Third Family			
τ (tau lepton)	1.77 ± 0.16 GeV	-e	Electromagnetic, Weak
ν_τ (τ neutrino)	< 2 eV	0	Weak
t (top quark)	$172.9 \pm 0.6 \pm 0.9$ GeV	$\frac{2}{3}$e	Electromagnetic, Strong, Weak
b (bottom quark)	$4.19^{+0.18}_{-0.06}$ GeV	-$\frac{1}{3}$e	Electromagnetic, Strong, Weak

Table 2.1: The names, masses, electric charge, family structure, and interactions of the Standard Model particles. Natural units (with $c = 1$) are used here and throughout this thesis [20].

at a much lower rate and propogates only short distances. It also is the only force felt by all matter particles. In 1961, it was discovered that at high energies the electromagnetic and weak interactions can be unified into one force [21], which required two charged massive force carriers, one massive neutral force carrier, and one massless neutral force carrier. While the photon was easily identified as the massless, neutral boson, the other force carriers were predicted to be heavy, and thus responsible for the weak force's short range. The first hints of the massive neutral boson, the Z, was seen in neutral current interactions in 1973 [22, 23]. The charged $W^\pm$ and neutral Z bosons were directly observed in 1983 [24–27].

The W bosons are worthy of note as the force carriers involved in flavor-changing interactions; for example, the top quark decays by emitting a W boson and a bottom quark. The Standard Model particles can be grouped into doublets according to these interactions, with each family containing a doublet of a lepton and its neutrino and another doublet of the up- and down-type quarks. However, the helicity, or

Symbol (Name)	charge	mass	interaction
γ (photon)	0	0	electromagnetic
$W^{\pm}$ (W boson)	$\pm$e	80.4 GeV	weak
Z^0 (Z boson)	0	91.2 GeV	weak
g (gluon)	0	0	strong

Table 2.2: The mass, charge, and interaction carried at low energy of the Standard Model vector bosons. [20]

projection of a particle's spin along its momentum, can affect these reactions as angular momentum must be conserved. Particles can be classified as either right- or left-handed based on whether the spin and momentum vectors are parallel or antiparallel, and right-handed charged leptons are not observed to interact with any corresponding neutrinos. This means that while such neutrinos may exist, right-handed leptons exist as $SU_L(2)$ singlets and not doublets.

Quarks, unlike the leptons, also experience the strong force and carry color charge, which exists in three types and three anti-types. This force is carried by the massless gluon, discovered in 1979 [28–30]. While the gluon is massless, it is also self-interacting; hence at low energies the strong force is the strongest of the interactions. This self-interaction of gluons produces confinement of color charges, which means bare quarks and gluons cannot be observed. If colored particles are seperated, it becomes energetically favorable for the strong force to produce quark-antiquark pairs, which join to the existing quarks and produce new hadrons. At sufficient energies, these hadrons will propogate together in a cluster of particles called a jet. Only quark-antiquark pairs (called mesons) and triplets with one of each color (called baryons) carry no net color charge and thus can freely propogate. The theory of these interactions is called Quantum ChromoDynamics (QCD), and this is also generically used to describe such interactions. The properties of all the vector bosons are summaried in Table 2.2, and with the fermions comprise the total particle content of the Standard Model.

2.2 Electroweak Symmetry Breaking

The mathematical framework of the Standard Model is that of a $SU_c(3) \times SU_L(2) \times U_Y(1)$ gauge theory. These groups correspond to local gauge symmetries and the Lagrangian is required to be invariant under transformations in each group. The $SU_c(3)$ portion corresponds to the strong interaction, while the $SU_L(2) \times U_Y(1)$ is the electroweak interaction. Calculations are done using the Lagrangian density, shown here for the first family of particles only:

$$L = \sum_{\psi=\nu_e,e,u,d} i\bar{\psi}\gamma^\mu D_\mu \psi - \frac{1}{2} Tr G_{\mu\nu} G^{\mu\nu} - \frac{1}{4} F^{j\mu\nu} F^j_{\mu\nu} - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} \tag{2.1}$$

where the covariant derivative $D_\mu = \partial_\mu - \frac{ig\sigma^b}{2} A_b - \frac{ifY}{2} B_\mu - ig_s \frac{\lambda^a}{2} G_a$, and the field terms $F^j_{\mu\nu} = \partial_\mu A^j_\nu - \partial_\nu A^j_\mu + g\epsilon^{jk\ell}$, j and b =1,2,3, $a = 1 \ . \ . \ . \ 8$, $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$, and $G_{\mu\nu} = \partial_\mu G_\nu - \partial_\nu G_\mu - ig[G_\mu, G_\nu]$. $\overrightarrow{\sigma}$ are the Pauli matrices, A, B, and G are the generators or vector boson fields, and ψ are the fermion fields. All the gauge symmetries must hold in the Langrangian for the theory to be renormalizable and give valid predictions; this imposes stringent boundaries on the behavior of included fields. For example, to preserve the $U_Y(1)$ symmetry, the transformation of a particle field as $\psi(x) \rightarrow e^{i\theta(x)}\psi(x)$ must be accompanied by a gauge field transformation of the form $A_\mu \rightarrow A_\mu - \partial_\mu \theta(x)$.

Introducing mass terms directly into this Lagrangian violates the $SU_L(2) \times U_Y(1)$ symmetry. Considering first fermion mass terms, which take the form of:

$$L_{mass} = m\bar{\psi}\psi \tag{2.2}$$

As the fermion fields are nominally doublets, this term appears to be a singlet in $SU_L(2)$ space and rotations would be invariant as required. However, this term can be rewritten using the Dirac matrices to describe the helicity states of the included

particles. As $(\gamma^5)^2 = 1$:

$$\begin{aligned} L_{mass} =& m\bar{\psi}\psi \\ =& m\bar{\psi}\frac{1+\gamma^5}{2}\frac{1+\gamma^5}{2}\psi + m\bar{\psi}\frac{1-\gamma^5}{2}\frac{1-\gamma^5}{2}\psi \\ =& m\bar{\psi}_L\psi_R + m\bar{\psi}_R\psi_L \end{aligned} \tag{2.3}$$

As discussed in Section 2.1, right-handed leptons exist as singlets while left-handed leptons are doublets; this makes these terms doublets in $SU_L(2)$ space and not invariant. Hence fermion mass terms break the $SU_L(2)$ gauge symmetry.

Boson mass terms would take the form:

$$L_{mass} = \frac{1}{2}m^2 A^\mu A_\mu \tag{2.4}$$

As previously discussed, the $U_Y(1)$ symmetry imposes the transformation condition of $A_\mu \rightarrow A_\mu - \partial_\mu\theta(x)$ on gauge fields. Applying this transformation to the mass terms yields:

$$\frac{1}{2}m^2 A^\mu A_\mu \rightarrow \frac{1}{2}m^2(A^\mu - \partial^\mu\theta(x))(A_\mu - \partial_\mu\theta(x)) \tag{2.5}$$

$\theta(x)$ is an arbitrary phase, and so the only way to ensure equality of these terms is to require $m = 0$. Mass terms of vector bosons are necessary to break the electroweak symmetry, as at low energies the electromagnetic force's gauge boson is massless while the weak force's carriers are not. A mechanism to incorporate mass terms into the Lagrangian without violating the gauge symmetries is needed.

The simplest way to break the electroweak symmetry and incorporate mass terms is via a mechanism discovered in 1964 by Higgs, Brout and Englert, and Guralnik, Hagen, and Kibble now called the Higgs mechanism [1, 31, 32]. Assume there is an additional scalar field ϕ that transforms as an $SU_L(2)$ doublet. This field has the

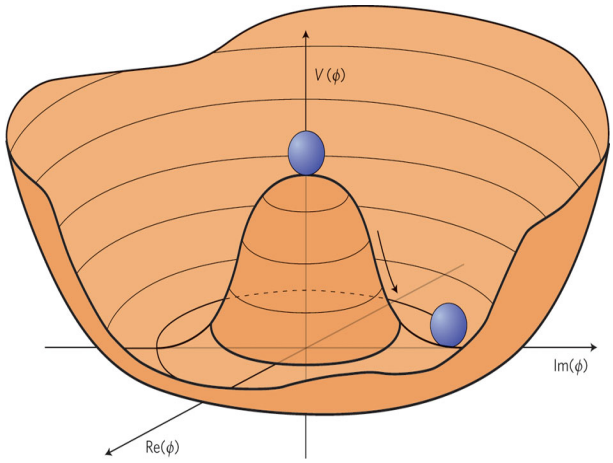

Figure 2.1: The potential well of the scalar field ϕ with $\mu^2 < 0$ and $\lambda > 0$. [33]

potential:

$$V(\phi) = \mu^2\phi^2 + \lambda\phi^4 \tag{2.6}$$

where we require $\mu^2 < 0$ and $\lambda > 0$. This potential has the shape shown in Figure 2.1. To continue calculations with this field, it is necessary to expand around its minimum, but this potential has infinitely many equally viable minima. We choose to take the minimum as $< \phi >= \sqrt{\frac{-\mu^2}{2\lambda}}$; this choice spontaneously breaks the symmetry, as other minima were equally viable options for the vacuum state. We then define a field $\phi' = \phi - < \phi >$ as the deviation from the ground state or vacuum.

Assuming this field has a Yukawa coupling to fermions with strength g_f, terms of the form:

$$L_{int} = g_f\bar{\psi}\psi\phi = g_f\bar{\psi}_R\psi_L\phi + g_f\bar{\psi}_L\psi_R\phi \tag{2.7}$$

are added to the Lagrangian. As both ϕ and ψ_L are $SU_L(2)$ doublets, such terms preserve the symmetry. Expanding in ϕ' yields:

$$L_{int} = g_f\langle\phi\rangle\bar{\psi}_R\psi_L + g_f\bar{\psi}_R\psi_L + g_f\langle\phi\rangle\bar{\psi}_L\psi_R + g_f\bar{\psi}_L\psi_R \tag{2.8}$$

This equation contains mass terms for fermions with $m = g_f\langle\phi\rangle$ and interaction terms between the fermions and the scalar boson that is the quantum of this field. Measured masses are then a result of how strongly fermions couple to this Higgs field; heavy particles are more strongly coupled while light particles interact very weakly with the field. This interaction generates mass terms for fermions without violating the gauge symmetries.

Calculating the kinetic energy of the field ϕ requires evaluating the covariant derivative as:

$$D^\mu\phi D_\mu\phi = (\partial^\mu - \frac{ig\sigma^b}{2}A_b - \frac{ifY}{2}B^\mu)\phi(\partial_\mu - \frac{ig\sigma^b}{2}A_b - \frac{ifY}{2}B_\mu)\phi \tag{2.9}$$

Expanding this yields terms of the form $-f^2\frac{Y^2}{4}\langle\phi\rangle^2 B^\mu B_\mu$, which is a mass term for the $U_Y(1)$ vector field. Rewriting this in full produces three mass terms in total, two that correspond to $SU_L(2)$ generators, and one that is a combination of the final $SU_L(2)$ generator and the $U_Y(1)$ generator.

Rewriting the ϕ field's potential itself in terms of ϕ' yields new Lagrangian terms of:

$$L_\phi = (\mu^2\langle\phi\rangle^2 + \lambda\langle\phi\rangle^4) - 2\mu^2\phi'^2 + 4\lambda\langle\phi\rangle\phi'^3 + \lambda\phi'^4 \tag{2.10}$$

The second term of this expression is a mass term for the ϕ' field with $m_{\phi'} = 2\sqrt{-\mu^2} = 2\sqrt{2\lambda}$. The choice of $\mu^2 < 0$ is now obviously necessary to produce a positive mass at this stage.

Incorporation of this scalar field has thus predicted three massive electroweak vector bosons, two charged bosons with mass around 80 GeV and on neutral one with mass around 94 GeV [21, 34, 35], while leaving one electroweak boson massless. The vacuum expectation value $v = \sqrt{-\mu^2/\lambda}$ is expected to be approximately 246 GeV. Fermion mass terms are allowed but must be measured experimentally. Finally, this

mechanism predicts a scalar, electrically neutral, massive boson, though its mass is not predicted. Inclusion of this mechanism into the unified electroweak interaction by Weinberg, Salam, and Glashow [21,34,35] produced the Standard Model as it is known today, and the discoveries of the two charged W bosons and the neutral Z boson with the predicted masses are some of its strongest supporting experimental evidence. However, observation of the scalar boson, called the Higgs boson, is necessary to prove that this mechanism accurately describes electroweak symmetry breaking.

2.3 Searches for the Higgs Boson

The mass of the Higgs boson is not directly predicted by the electroweak symmetry breaking procedure discussed in Section 2.2, and any other theoretical limits on this parameter are fairly weak. For example, it has been shown that the unitarity of WW scattering processes for highly energetic bosons fails unless the mass of the Higgs boson is less than 1 TeV [36]. However, as the only additional information available at the time indicated the Higgs boson mass must also be greater than 4 GeV, a wide range was left to be covered by direct searches. A discussion of the decay channels of the Higgs boson is included in Section 2.3.1; production channels are specific to the collider in question and will be presented in the following chapter. Experimental results prior to those discussed in this dissertation are covered in Section 2.3.2. Indirect results based on precision measurements of electroweak processes are discussed in Section 2.3.3.

2.3.1 Decay Channels of the Higgs Boson

Most of the information available on the mass of the Higgs boson comes from direct searches for the boson's existance. As no fundamental scalar particles have been observed, a Higgs boson must be unstable and rapidly decay; its predicted decays can be calculated from the couplings between it and other Standard Model

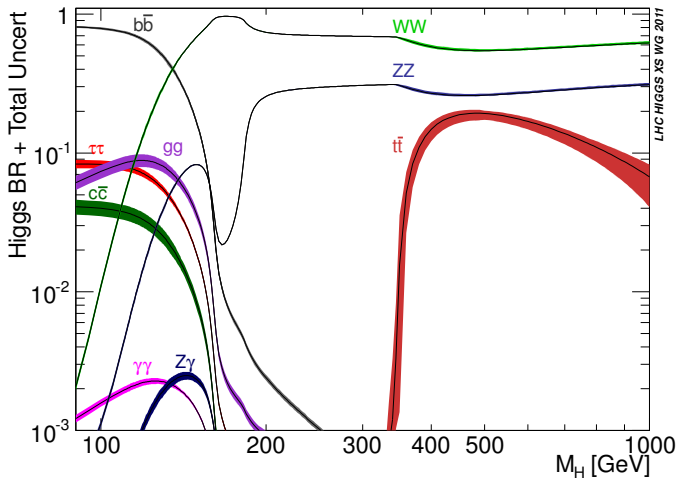

Figure 2.2: The branching ratios of several Standard Model Higgs boson decay channels. [37]

particles. The exact branching ratio of a given decay mode depends on the mass of the Higgs boson, and thus searches for each of these final states are sensitive to different ranges of mass. Figure 2.2 shows the branching ratios of many of the Higgs boson's decays to final state particles for Higgs boson masses between 100 and 1000 GeV.

For the low mass range up to 135 GeV, the predominant decay mode of the Higgs boson is to pairs of b quarks. At colliders with no significant underlying hadronic activity in events, this mode is a feasible search channel. However, the largest colliders in operation, such as the Tevatron and the Large Hadron Collider, produce hadron collisions and high levels of hadronic backgrounds. This makes final states dominated by quark activity difficult to seperate from the background. At such colliders, the decays to pairs of photons or τ particles become the main discovery channels in the low mass range. The diphoton channel has the benefit of the entire final state being measurable; this means the mass of the Higgs boson creating the photon pair can be reconstructed. The search in this channel is for a resonance peak over a smooth background. The τ lepton final state presents some additional challenges as a Higgs boson

decay channel because the τ leptons are not stable. They decay electroweakly producing neutrinos and $W^{(*)}$ bosons, with the bosons then decaying to lepton-neutrino or quark-antiquark pairs. These final states therefore involve measuring missing transverse energy and cannot be used to reconstruct a mass peak from the event, making this final state a search for an excess over the expected backgrounds.

For Higgs masses greater than 135 GeV, the decays of the Higgs boson to diboson final states become the primary discovery channels. The ZZ final state is one of the primary decay channels for the high mass range; its lower cross-section than the WW decay and lower branching fraction to leptonic or semileptonic states, however, mean it is not as powerful a discovery channel for the Higgs boson. But the $ZZ \to \ell\ell\ell\ell$ channel is one of the few final states with no invisible or strongly interacting particles amongst the Higgs boson's decay products. As leptons typically have lower fake rates than photons, this channel is likely to provide the most precise measurement of the properties of the Higgs boson, such as its mass and spin, after discovery.

The decay to pairs of W bosons is the preferred decay mode of the Higgs boson with the highest branching ratio in the high mass range. The W bosons can then decay to leptonic ($\ell\nu\ell\nu$), semi-leptonic ($\ell\nu q\bar{q}$), and hadronic ($q\bar{q}q\bar{q}$) final states, but the hadronic state is difficult to analyze due to high levels of background at modern hadron colliders. Leptonic final states, composed of either electrons or muons, have a far lower background, promoting the $WW \to \ell\nu\ell\nu$ decay channel to the primary discovery channel used for much of the Higgs mass range. As this final state involves neutrinos and therefore missing transverse energy in an event, the strategy of this search is to look for an excess over the number of expected background events. This is the chosen final state for analysis in this dissertation.

2.3.2 Limits from Direct Searches

Since the prediction of the Higgs boson's existence, several experiments at different colliders have searched for this particle. Prior to the start of the Large Hadron Collider experiments to be discussed in Chapter III, experiments at the Large Electron Positron (LEP) Collider operated by CERN and at the Tevatron operated by Fermi National Accelerator Laboratory (Fermilab) had published the most stringent limits on the mass of the Standard Model Higgs boson.

The LEP Collider was operated by CERN between 1989 and 2000, colliding electrons and positrons at center of mass energies between 91 and 209 GeV. After the conclusion of data-taking, the four LEP experiments combined data across all available search channels, making up to 2.46 fb^{-1} available for analysis. The final result of this analysis was that at the 95% confidence level the mass of the Standard Model Higgs boson must be above 114.4 GeV [4].

The Tevatron was a proton-antiproton collider located at Fermi National Accelerator Laboratory. It operated at center of mass energy of 1.96 TeV between 2001 and 2011. The combination of all experimental results from its two detectors CDF and D0 in different decay channels included up to 10 fb^{-1} of data. This excluded the mass ranges of $100 < m_H < 106$ GeV and $147 < m_H < 179$ GeV for the Standard Model Higgs boson [5]. These limits were also set at the 95% confidence level. The Tevatron results are shown graphically in Figure 2.3. This plot shows the expected and observed cross-sections measured in the analysis, normalized to the cross-section of Standard Model Higgs boson production at that mass. If the observed curve dips below one, this indicates that the analysis has observed fewer events then can agree with a combined signal and background prediction, and the mass is said to be excluded. Together with the LEP results, this leaves a small low mass range, between $115 < m_H < 147$ GeV, and the high mass range above $m_H = 179$ GeV to be searched at future experiments.

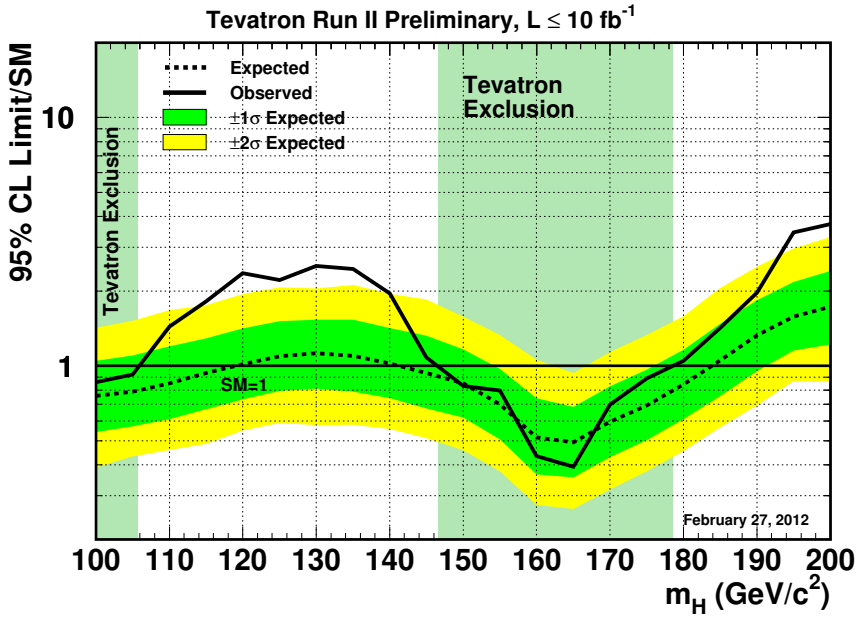

Figure 2.3: The expected and observed limits of the Higgs boson's mass for the two Tevatron experiments using up to 10 $\text{fb}^{-1}$of data. All possible final states are included. Regions where the expected curve dips below one are defined as excluded to the Standard Model Higgs boson at 95% confidence level [5].

2.3.3 Constraints from Electroweak Precision Measurements

Additional information on the likely mass of the Higgs boson can be inferred by looking at precision measurements of other electroweak processes and quantities. Interactions between fermions and the Higgs boson in loops contribute to determination of parameters such as the W boson and top quark masses, and so precise measurements of these parameters yields information about the Higgs boson mass. Such studies are done with measurements from LEP, the Tevatron, and other experiments. Measurements include those made of the Z pole, the mass of the top quark, and the mass and width of the W boson. A χ^2 fit is then done to estimate the mass of the Higgs boson that best agrees with the data. The resulting curve for the measurements listed above is shown in Figure 2.4. The minimum point is at $m_H = 89^{+35}_{-26}$ GeV with a 95% confidence level limit of $m_H < 158$ GeV [38]. This minimum value is lower than that excluded by direct searches, and the fit has been redone including the LEP and Tevatron direct search results. This fit places the most likely value for the Higgs mass at 120^{+12}_{-5} GeV [39]. This gives hope that the mass of the Standard Model Higgs boson is only slightly above the limits of previous experiments, and within the reach of the next generation of effort, discussed in the following chapters.

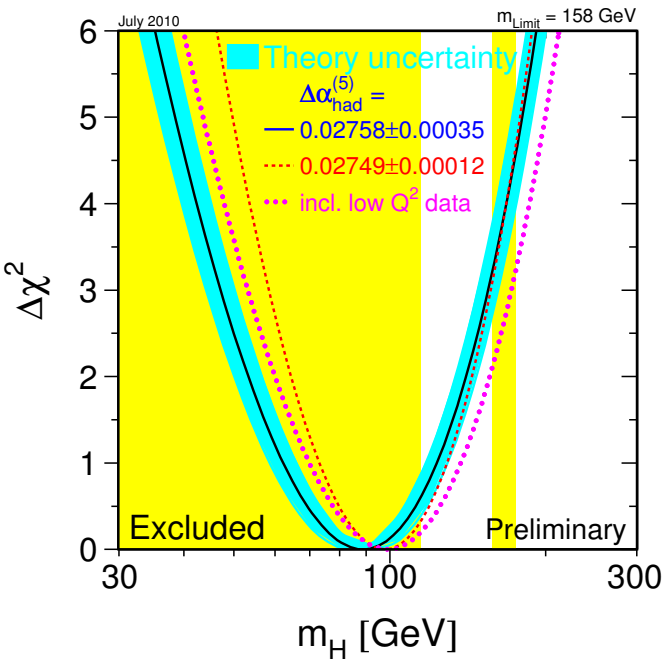

Figure 2.4: The χ^2 fit of precision electroweak data using measurements from LEP, the Tevatron, and other experiments as a function of Standard Model Higgs mass. [38]

CHAPTER III

LHC and ATLAS

The data used for the analysis presented in this dissesrtation were collected at the A Toroidal LHC ApparatuS (ATLAS) detector located on the beamline of the Large Hadron Collider (LHC) at CERN, outside Geneva, Switzerland. ATLAS is one of four collision point detectors operating at the LHC; ATLAS and CMS are multipurpose detectors designed for discovery of new fundamental particles, while LHCb and ALICE focus on B hadron physics and heavy ion collisions, respectively. The lay-out of the CERN accelerator complex is shown in Figure 3.1. Information about the design of the LHC and the expected production of the Higgs boson in its collisions are presented in Section 3.1; details about the ATLAS detector follow in Section 3.2.

3.1 The Large Hadron Collider

The LHC is a particle accelerator operated by CERN and located outside of Geneva, Switzerland. The ring holding the collider is 27 km in circumference and located approximately 100 m underground; it housed the LEP Collider between 1989 and 2000. The LHC also makes use of CERN's existing accelerator chain as shown in Figure 3.2, though new hardware and connection tunnels were necessary to produce and transport beams at the energy and current required by the LHC. While capable

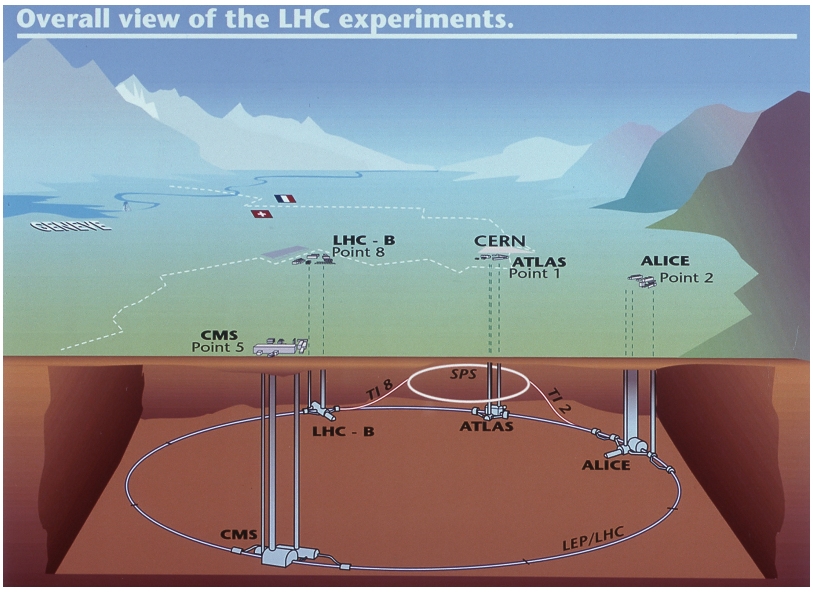

Figure 3.1: The LHC complex with its four collision point detectors, ATLAS, CMS, ALICE, and LHCb, shown relative to the Geneva area. ©1999 CERN [40].

of accelerating multiple types of hadrons, the LHC reaches the highest center of mass energies as a *pp* collider, with a design limit on center of mass energy of 14 TeV [41]. Furthermore, the LHC produces high intensity beams of particles; at full capacity, it contains 2808 proton bunches per beam at a spacing of 25 ns, which corresponds to a beam current of 0.584A and a stored beam energy of 382 MJ. This allows the LHC to achieve an instantaneous luminosity of $10^{34} cm^{-2} s^{-1}$.

The main components of the accelerator are the dipole magnets which produce a magnetic field that bends the particles around the ring. The LHC contains 1232 superconducting dipoles; these are made of niobium-titanium as has been successfully used at the Tevatron and DESY accelerators, but the LHC maintains a magnet temperature of 2K, half of that used at previous colliders. The cooler temperature allows the magnets to produce fields of 8.33 T, but also imposes heat tolerances almost an order of magnitude tighter. The dipoles are cooled using superfluid helium. Several other types of superconducting magnets are also used to focus and correct

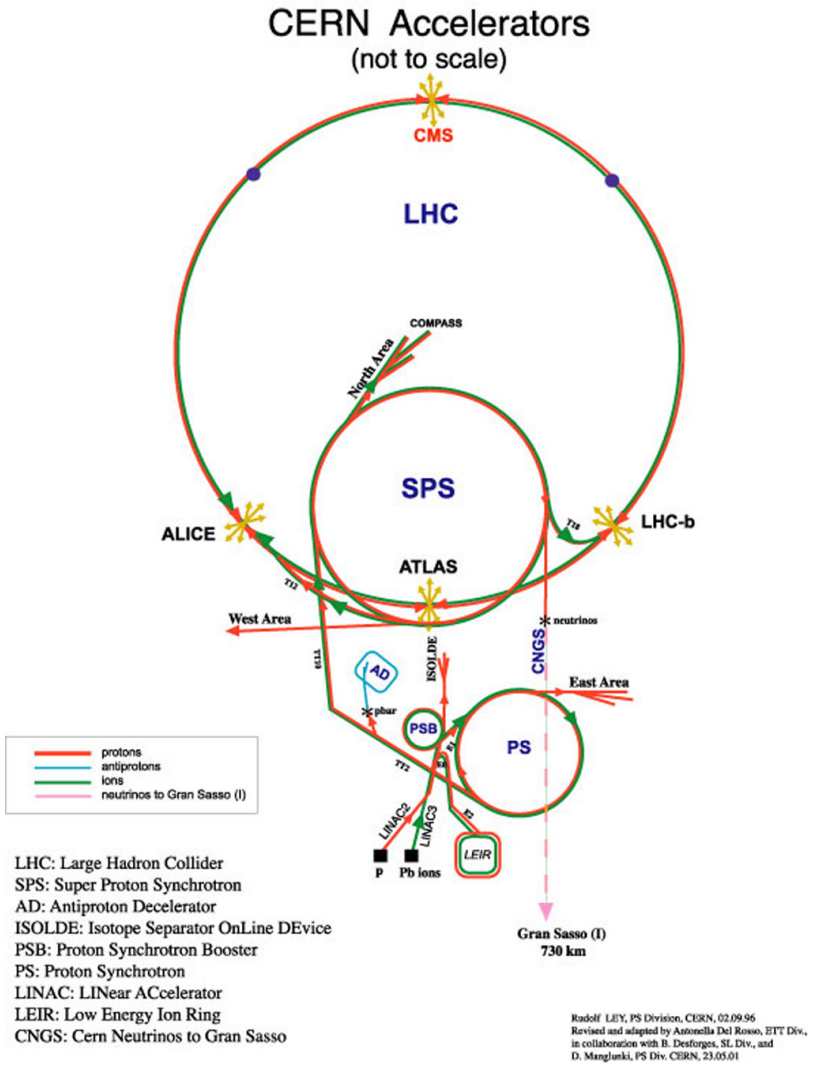

Figure 3.2: The CERN accelerator complex showing the chain of accelerators that deliver beams to the LHC. ©2001 CERN [42]

the beams' position, move the beams to collide, guide beams into the accelerator ring, and dump the beams as necessary. As the magnets can store an additional 600 MJ of energy, several layers of system sensors and protections are in place in case of problems with the beam or magnets. Beams are captured and accelerated by a 400 MHz superconducting cavity system, producing the highest energy particle collisions on Earth.

3.1.1 Higgs Boson Production at the LHC

Hadron colliders such as the LHC are at the forefront of the accessible energy spectrum for particle physics and therefore the primary tools for searching for the Higgs boson. Protons are more suitable than electrons for circular colliders as their higher mass means less energy loss to radiation as the particles accelerate, and circular colliders can be constructed smaller and more cheaply than linear ones. However, protons are not fundamental particles and determining the initial state of a proton collision is therefore not possible. Proton-proton collisions at the LHC can involve the valence up and down quarks of the proton, sea quarks created in quark anti-quark pairs within the particle, or gluons exchanged and radiated by any of these quarks. This makes a wide range of initial interactions available for study, but any given final state can be generated by several different interactions. Furthermore, the initial momentum of the interacting particles cannot be exactly determined; while Parton Distribution Function (PDF)s can be used to describe the range of energies and momenta carried by a proton's constitutents, quantities parallel to the initial direction of motion cannot be calculated. Hence the analysis presented here will focus on transverse quantities which can be completely reconstructed and for which conservation laws hold.

Studies at the LHC must also address the fact that events will likely contain large numbers of particles; the structure of a hard collision between hadrons is shown in

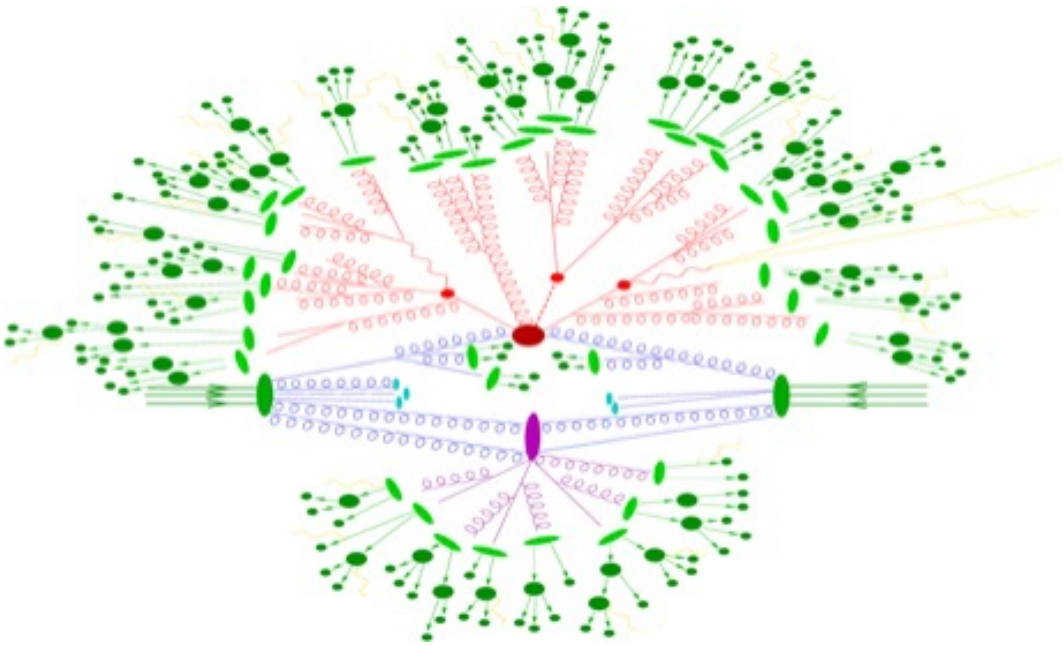

Figure 3.3: Diagram of the structure of a hadron collision; the central hard collision and its daughter particles are shown in red, gluon radiation is shown in blue, interactions amongst the recoiling particles is in purple, and final state hadrons are shown in green.

Figure 3.3. Hard collisions between proton constituents will result in a recoil from the other components of the initial protons, and the recoiling particles will become part of the final event. High energy particles can also radiate additional particles, primarily photons and gluons, both before and after a hard collision; these particles can impact an event's kinematics and therefore must be properly accounted for. The bunch structure of the LHC beam brings the possibility that other protons in the bunch will also collide and produce underlying events to the primary hard collision. This is called pile-up. Additional particles can also be added to an event from beam halo and backgrounds from previous bunch crossings, as well as cosmic rays penetrating the detector from above. All these factors contribute to particle-rich events being produced for analysis.

The Standard Model Higgs boson is expected to be produced in several channels at the LHC [44,45]. The exact ratios between the contributions depends on the center

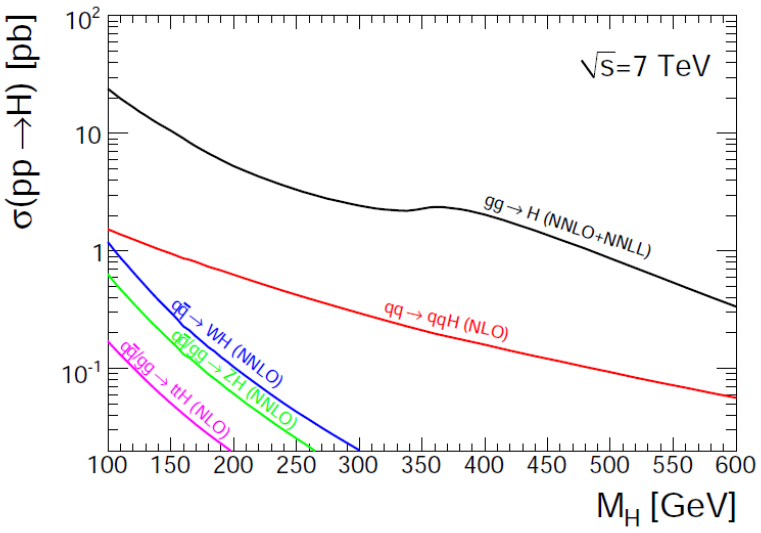

Figure 3.4: Higgs production cross-sections for the major Higgs boson production processes as a function of Higgs mass for $\sqrt{s} = 7$ TeV. © 2010 ATLAS [43].

of mass energy of collisions at the LHC and on the mass of the Higgs boson; Figure 3.4 shows the production cross-sections as a function of Higgs mass for $\sqrt{s} = 7$ TeV, the LHC's operating energy in 2010-2011. Feynman diagrams showing these modes are in Figure 3.5. The dominant contribution to the inclusive production cross-section is from the gluon-gluon fusion interaction, as it relies on primarily strong processes. This process uses the high energy gluons radiated frequently by energetic protons and the heavy quarks accessible at the energy range of the LHC. The Higgs boson's couplings are dependent on fermion mass, which contributes to the higher cross-section of this production channel. The second most dominant channel at the LHC is the vector boson fusion channel. This process has a lower probability due to requiring the production of pairs of massive vector bosons and involving smaller electroweak couplings, but it is also sensitive to couplings between Higgs and vector bosons, which are an important window into the behavior of Standard Model physics and phenomena beyond the Standard Model. This process produces decays with at

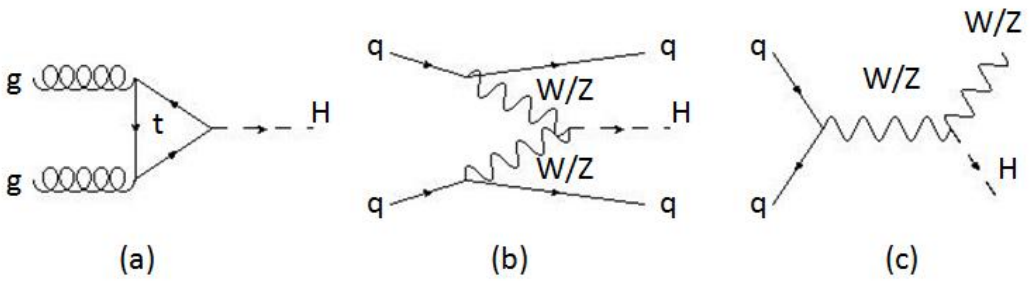

Figure 3.5: Feynmann diagrams for the major production modes of the Higgs boson at the LHC. Gluon-gluon fusion with a top quark loop (a) is the dominant contribution, followed by vector boson fusion (b). The associated WH and ZH processes (c), also called Higgs-strahlung production, are the third most dominant channel.

least two jets, typically in the forward or backward directions of the detector, which alters the topology of such events compared to those produced by gluon-gluon fusion. Production associated with a W or Z boson, called the Higgs-strahlung processes, was the primary production mode explored in the LEP and Tevatron results presented in Section 2.3.2, but this channel is a tertiary contribution to Higgs production at the LHC.

3.2 ATLAS Detector

The ATLAS Detector is one of two multipurpose detectors located at collision points on the LHC beamline and designed for discovery of fundamental particles. Figure 3.6 shows the layout of the detector with its major subsystems. The detector stands 25 m high and 44 m long, and weighs approximately 7000 metric tons. It has 4π coverage with nominal forward-backward symmetry with respect to the interaction point. The high rate of collisions at the LHC necessitates that all subsystems respond quickly to incoming particles and have a high granularity to distinguish between overlapping events. The performance goals of the ATLAS detector are shown in

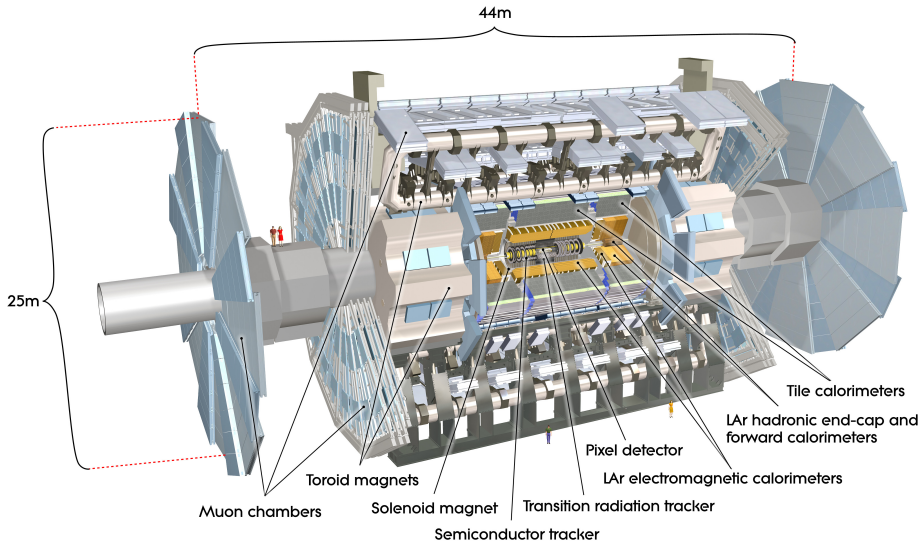

Figure 3.6: The ATLAS detector and its major subsystems. Human figures are shown on the beampipe and standing in front of the detector for scale. © 2008 CERN [46]

Table 3.1.

The search for the Higgs boson was used as a standard for designing the performance requirements of the ATLAS detector [47]. Accurate measurement of lepton transverse momentum was required both of low momentum particles close to the interaction point and muons that can penetrate many materials with minimal interactions. This is necessary in order to be sensitive to all leptonic Higgs decays, particularly if a lighter Higgs boson produces low momentum leptons. Track recon-

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

Table 3.1: Performance goals of the main ATLAS detector subsystems. [47]

struction and vertexing is critical for reconstruction of τ leptons and B hadrons which can be present in decays of the Higgs boson, as well as identifying these particles in background processes. Several of the Higgs decay modes produce neutrinos which escape detection and therefore appear as missing transverse energy in the detector. This requires a wide angular acceptance and a full coverage hadronic calorimeter to measure an event's constituent particles and accurately calculate the missing transverse energy. As the collision rate of the LHC is too high for all information on all events to be captured and stored, a trigger system capable of distinguishing interesting events with both high and low transverse momentum particles is necessary to capture the desired information while providing the first rejection of backgrounds.

ATLAS addresses these requirements with three main subsystems sensitive to all Standard Model particles except neutrinos. Components of the detector are layered around the interaction point. Closest to the interaction is the inner detector, discussed in Section 3.2.1, a tracking and vertexing system immersed in a 2 T solenoidal magnetic field. Outside of the inner detector is the electromagnetic and hadronic calorimeter, presented in Section 3.2.2 which capture and measure the energy of most particles produced in collisions. As muons can consistently escape both of these systems, the outermost layer of the ATLAS detector is a large muon spectrometer, discussed in Section 3.2.3, which provides a second and independent measurement of the transverse momentum and charge of muons. It makes use of an air-core toroidal magnet system to generate bending of muon tracks. The trigger system will be presented in Section 3.2.5.

The ATLAS geometry is determined by the detector's cylindrical shape. The z-axis is defined as the beamline, with the positive x-axis pointing toward the center of the LHC ring and the positive y-axis pointing upwards. The origin is the nominal collision point in the center of the detector. Transverse quantities of interest such as tranverse momentum (p_T) and transverse missing energy (E_T^{miss}) are defined in the

$x - y$ plane. As the detector is nominally forward-backward symmetric, side A is distinguished as points with positive z and side C with negative z. The polar angle θ is the angle from the beamline and the azimuthal angle ϕ is the angle around the beamline. Pseudorapidity ($\eta = -\ln(\tan\theta/2)$) is more commonly used than the polar angle as many processes produce particles with a constant distribution in η. Detector systems are divided into barrel regions at low η, typcially in layers of constant radius, and endcap regions at high η with detector elements in the transverse plane. The division allows for greater coverage of the collision point and tailoring of detector elements for higher precision at low η to measure events resulting from hard scattering and greater radiation hardness at high η with its higher flux of particles.

3.2.1 Inner Detector

Closest to the interaction point within ATLAS is the inner detector, a tracking system designed to provide vertexing and momentum measurements at the interaction point [47]. A cross-section of this detector is shown in Figure 3.7. Precision tracking is provided by the pixel and silicon microstrip trackers in the range $|\eta| < 2.5$. The pixel detector contains three layers of silicon pixels, each layer 250 μm thick and each pixel 50×400 μm in size. These are oxygenated n-type wafers with readout pixels on the n^+-implanted side, as this gives good charge collection even after type inversion; it is expected that the high radiation exposure of the pixel detector will invert its type within the lifetime of the experiment. The pixel detector is segmented in R-ϕ and z with an intrinsic accuracy of 10 μm in R-ϕ and 115 μm in z in both barrel and endcap. It requires 80.4 million readout channels, and provides track and vertex reconstruction within the bunch crossing. This is necessary to accurately tag heavy-flavored jets and to distinguish particles in a given event from backgrounds produced in other interactions.

Outside the pixel detector is the Semi-Conductor Tracker (SCT) [47]. It contains

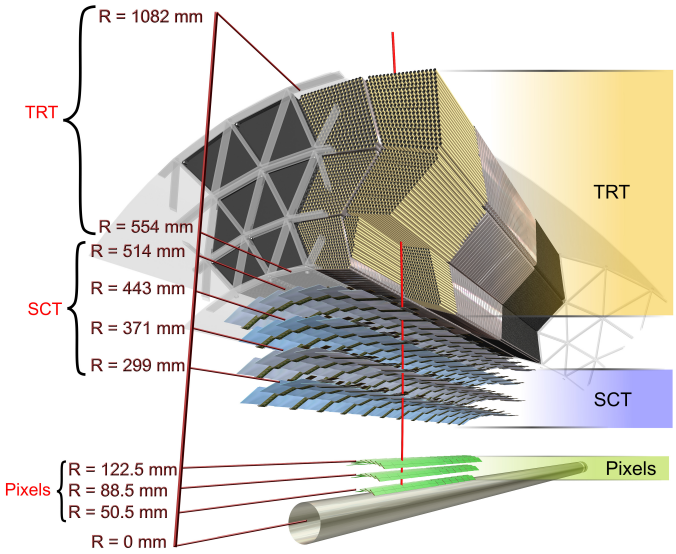

Figure 3.7: A cross-section of the ATLAS inner detector. © 2008 CERN [48]

four layers of paired silicon strips; one strip of each pair is at an angle of 40 mrad to the nominal to measure both R and ϕ coordinates while requiring fewer readout channels than the pixel detector. The SCT has an intrinsic accuracy of 17 μm in R-ϕ and 580 μm in z in both barrel and endcap with 6.3 million readout channels. For cost effectiveness and ease of production, the sensors are classic single-sided p-in-n technology nominally operated at 150V, though this can be increased to between 250-350V within ten years of operation, to accommodate the radiation dosage and deterioration of the detector [49]. Both the silicon detectors are kept in the temperature range between -5 and -10°C to contain annealing from radiation and leakage current.

The largest and outermost of the inner detector trackers is the Transition Radiation Tracker (TRT) [47]. Straw tubes 4mm in diameter and filled with a xenon-based gas mixture provide typically 36 hits per track; this high number of hits over a large radius contributes strongly to momentum measurements of tracks. The tubes have

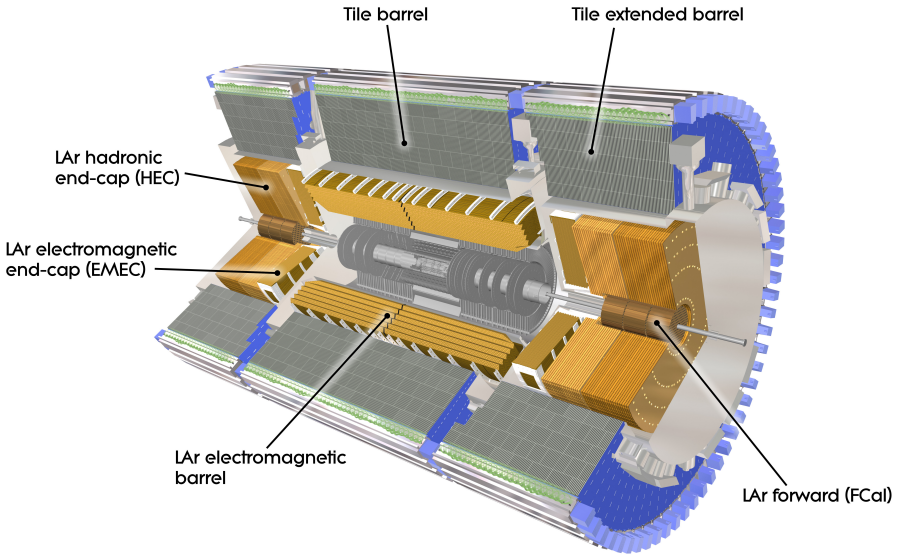

Figure 3.8: The calorimeters of the ATLAS detector. © 2008 CERN [51]

an intrinsic accuracy of 130 μm and provide only R-ϕ information, requiring reconstruction with silicon hits to determine position in z. The TRT has 351,000 readout channels. This tracker also provides electron identification in addition to the information provided by the electromagnetic calorimeter; this is done by placing radiators between drift tubes [50]. When crossing the boundary between radiator and tube, electrons produce photons that are absorbed by the xenon gas and yield a larger signal than that of minimum ionizing charged particles [49]. These photons are collected by the tubes along with electrons produced by passing tracks, and thresholds in the front-end electronics differentiate between the two to provide identification of electrons in addition to the inner detector's primary purposes of track and vertex reconstruction.

3.2.2 Calorimeters

The ATLAS detector has calorimeter systems outside the inner detector to measure particle and jet energies; the lay-out of the major sections of these systems

are shown in Figure 3.8. These must be capable of containing electromagnetic and hadronic showers for accurate measurement and radiation containment. To that end, the Liquid Argon (LAr) electromagnetic calorimeter, the innermost calorimeter of ATLAS, is more than 22 radiation lengths thick in the barrel, or 9.7 interaction lengths (24 and 10 in the endcap); the barrel is segmented into three layers, called the strip, middle and back layers from innermost to outermost, and the endcap is segmented into two layers [47]. The calorimeter has lead absorber plates and liquid argon as the detecting medium; argon was chosen for its linear behavior, stability of response over the expected lifetime of the detector, and intrinsic radiation hardness [49]. The LAr calorimeter has accordian shaped kapton electrodes as this configuration can be mounted without requiring azimuthal cracks for support structures and gives the calorimeter complete ϕ coverage. It is divided into three sections, one barrel and two endcap, each within its own cryostat. To limit the volume of non-active material crossed by particles produced at the interaction point, the calorimeter shares a vacuum vessel with the solenoid magnet of the inner detector, removing the need for an additional vessel and two additional walls. It also features a presampler in the region $|\eta| < 1.8$ to correct for energy lost by electrons and photons passing through the inner detector.

Many decay channels of physics processes such as the Higgs boson require accurate reconstruction of jets and missing transverse energy; the hadronic calorimeters of ATLAS have the largest extension in η to accomodate this need [47]. Hadronic jet measurements in the barrel are provided by the tile calorimeter, a sampling calorimeter with steel absorber plates and scintillating tiles as the active material. Additional steel-scintillator sensors in gaps necessary for support and readout electronics allow for partial energy recovery in these regions [49]. The tile calorimeter is segmented three layers deep in both barrel and endcap, with a total interaction length of 7.4 in each of these regions; the depth is necessary to prevent punch-through to the muon

spectrometer and energy loss in events.

The endcap and forward hadronic calorimeters extend the ATLAS detector's reach to $|\eta| = 4.9$. A copper and liquid argon sampling calorimeter is located behind the electromagnetic endcap calorimeters within the same cryostats; this overlaps with the tile and forward calorimeters to reduce a drop in material density in transition regions [47]. The forward calorimeters provide three additional layers of sensors at the highest η range. The innermost is copper and liquid argon for sensitivity to electromagnetically interacting particles while the outer two use tungsten and liquid argon for sensitivity to hadrons. These extend the calorimeters as close to the beamline as possible for accurate reconstruction of jets and missing transverse energy.

3.2.3 Muon Spectrometer

The outermost and largest subsystem of the ATLAS detector is the muon spectrometer [47, 49]; it is made up of two types of precision tracking chambers and two types of triggering chambers in the magnetic field of three air-core toroidal magnets as shown in Figure 3.9. Its size and position is necessary as energetic muons can penetrate the inner detector and calorimeters with only small energy loss, and the long lever arm produced by tracking muons from the inner detector allows for an accurate momentum measurement. The design goal of the spectrometer was a transverse momentum resolution of 10% for 1 TeV tracks, which corresponds to a resolution of ≤ 50 μm for a track with a sagitta of 500 μm in the bending plane. Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC) serve as triggering sensors for the spectrometer and provide a second coordinate measurement for reconstructed tracks. Much of the precision tracking is done by Monitored Drift Tubes (MDT) along the barrel and in the outer layers of the endcaps; Cathode Strip Chambers (CSC) serve the same purpose in the inner layer of each endcap.

The Monitored Drift Tubes cover the range $|\eta| < 2.7$ except for the range $2 <$

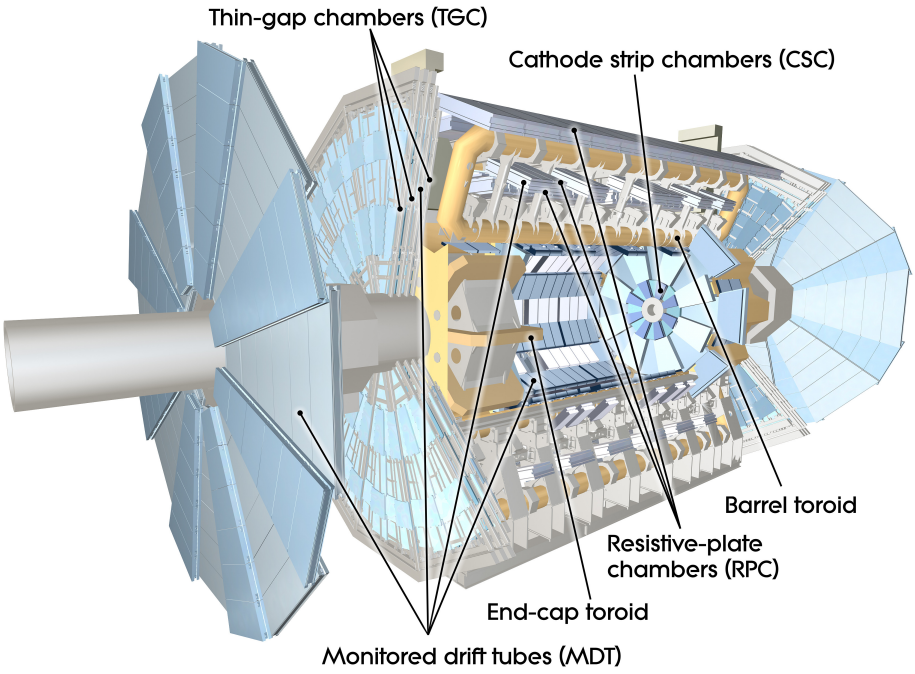

Figure 3.9: The ATLAS Muon Spectrometer. © 2008 CERN [52]

$|\eta| < 2.7$ on the innermost endcap layer for precise measurement of muon tracks. The drift tubes are 30 mm in diameter and contain $ArCO_2$ gas held at a pressure of 3 bar around a central tungsten-rhenium wire. Drift tubes were favored here as the cylindrical geometry with its radial electric field gives tracking that is only weakly dependent on the angle of incidence, while the mechanical separation of the tubes allows easy construction and maintenance over the lifetime of the experiment. Tubes are arranged in layers within chambers, with three to eight layers of tubes per chamber, with the tubes offset between layers to eliminate ambiguity on which side of the wire the particle passed. The system has an intrinsic resolution of 80 μm per tube or 35 μm per chamber across its 1088 chambers and 320000 readout channels. Both barrel and endcap contain three layers of chambers with some overlap between ϕ sectors; tracks through the overlap regions can be used for studies on the alignment of the detector, which is necessary for accurate reconstruction of tracks. The drift tube chambers also have built-in alignment sensors to ensure the alignment remains

within the necessary 30 μm.

Drift tubes cannot handle the high particle fluxes experienced in the high η regions on the inner layers of the endcap, hence Cathode Strip Chambers are used in the range $2.0 < |\eta| < 2.7$ here. These are multiwire proportional chambers with the cathode planes segmented into strips. Strips are arranged in orthogonal directions to measure two coordinates of the track. This granularity, higher than that reached by the MDTs, is necessary for the higher particle fluxes in this region while providing better time resolution. CSCs have a resolution of 40 μm in the bending plane and 5 μm in the transverse direction.

Triggering in the muon spectrometer requires specialized chambers to provide bunch crossing identification, well-defined transverse momentum thresholds, and measurement of coordinates not provided by the precision tracking chambers. In the barrel, Resistive Plate Chambers complement the operation of the MDTs; these are gaseous parallel electrode-plate detectors. As these don't depend on wires, they are less sensitive to misalignment and require comparatively fewer hits to produce a track. The barrel contains three layers of unequally spaced RPCs; the smaller distance between the inner two layers provides a low momentum trigger for muons in the range of 6-9 GeV while the longer lever arm between the inner and outermost layer is better suited for high momentum tracks. In the endcaps, triggering comes from Thin Gap Chambers (TGC), multi-wire proportional chambers with the distance between wire and cathode is smaller than that between wires. This geometry produces strong electric fields that yield good time resolution. In addition to serving the ATLAS trigger system, all triggering chambers are used to provide the coordinate in the non-bending plane, as this is not measured by MDTs and is necessary for track reconstruction before physics analysis can take place.

3.2.4 Magnet Systems

Each of ATLAS's tracking systems, the inner detector and muon spectrometer, requires a magnetic field for momentum measurement [47]. The inner detector is housed in a solenoid producing a nominal 2T axial field; the flux is returned by the steel of the barrel tile calorimeter. The solenoid runs at 4.5K with a current of 7.73 kA. As the solenoid lies within the calorimeter, it has the potential to absorb energy from particles passing through, making reconstruction of their properties difficult. Thus it was desirable to construct a light magnetic system for the inner detector to minimize this effect. The solenoid is 10 cm thick and weighs 5.4 tons while storing 40 MJ of energy at full field, thus satisfying this requirement with an energy to mass ratio of 7.4 kJ/kg and a depth of 0.66 interaction lengths at normal incidence. Furthermore, the solenoid shares a vacuum vessel with the barrel electromagnetic calorimeter and thus removes the need for two additional vessel walls.

Magnetic fields for the muon spectrometer comes from three sets of air-core toroidal magnets, one in the barrel and two in the endcaps. The air-core magnets are lighter and include less material to minimize multiple scattering of muons. In the barrel, the coils are housed in eight individual race-track shaped vacuum vessels; the endcap magnets are connected as large wheels with sixteen segments. Endcap magnets are oriented such that the magnetic field remains perpendicular to muons passing through the transition region, at the cost of the magnetic field not being constant through the entire spectrometer. The field strength is roughly 0.5T in the barrel and 1T in the endcap while the magnets run at 4.6K with a current of 206 kA. Together, the magnet systems allow the ATLAS detector to measure particle momenta from a few hundred MeV to the TeV scale.

3.2.5 Triggering

At its nominal running parameters, the LHC produces events at a rate of 40 MHz, too high to be read completely from the ATLAS detector. The detector employs a triggering system to selectively read out events with interesting parameters for storage and analysis. The trigger operates in three stages; the level 1 trigger provides the first selection of events with minimal information while the level 2 trigger and event filter, collectively called the high level trigger, access more information for more sophisticated decision-making [47].

The level 1 trigger must either pass or fail events within 2.5 μs. It does this by searching for objects with large transverse momentum, missing tranverse energy, or total tranverse energy. Information on events is taken only from a subset of detectors; for lepton-focused analyses such as the one presented here, the primary sources are either the muon trigger chambers or the calorimeters, evaluated in the trigger software with a reduced granularity. If passed, the level 1 trigger identifies a Region of Interest (RoI) within the detector where the interesting object passed and passes information on the coordinates, object type, and threshold passed to the high level trigger. The level 1 trigger can be prescaled to pass a subset of events passing some threshold; this allows study of events with lower energy objects that occur too frequently for all instances to be read out. As currently installed, the level 1 trigger can handle a final rate of up to 75 kHz, though this can be upgraded over the lifetime of the experiment.

The level 2 trigger accesses all information contained within RoIs selected by the level 1 trigger for decision-making; this represents approximately 2% of the data available on each event, and allows for more involved logic. A processing time of up to 40 ms is allowed, with a maximum final rate of 3.5 kHz. The level 2 trigger hands information on passing events on to the event filter, which can use offline analysis procedures such as basic object reconstruction and association of objects to asses an event's properties. The event filter is allowed 4s of processing time and produces a

final event rate of 200 Hz. Events that exceed the processing time thresholds without the trigger system reaching a decision are also read out in a dedicated stream for offline analysis and debugging. Hence all data for analysis has been preselected in some way for interesting properties.

CHAPTER IV

The ATLAS Monte Carlo and Data Chain

Predictions of the expected signal and background events and their properties are necessary for calibration of the ATLAS detector and for any study of particle behavior. Monte Carlo simulations of particle generation and propagation and of detector response are used to generate samples of events, in the same format as data, for these studies. In particular, as the $H \to WW \to \ell\nu\ell\nu$ search is for an excess of events and not a resonance peak, Monte Carlo simulation of the Higgs signal and its major backgrounds is necessary for measurement.

This chapter describes the chain by which simulated physics events are produced and raw event information from the ATLAS detector is reconstructed for analysis. Processes that produce backgrounds to the search for the Higgs boson in the $WW \to \ell\nu\ell\nu$ final state are introduced in Section 4.1. Generation of events via Monte Carlo processes is discussed in Section 4.2. Section 4.3 covers simulation of the ATLAS detector response using GEANT4 and Section 4.4 covers the digitization of these signals into raw event data of the form produced by the detector. Object reconstruction, performed on both data and Monte Carlo events, is discussed in Section 4.5.

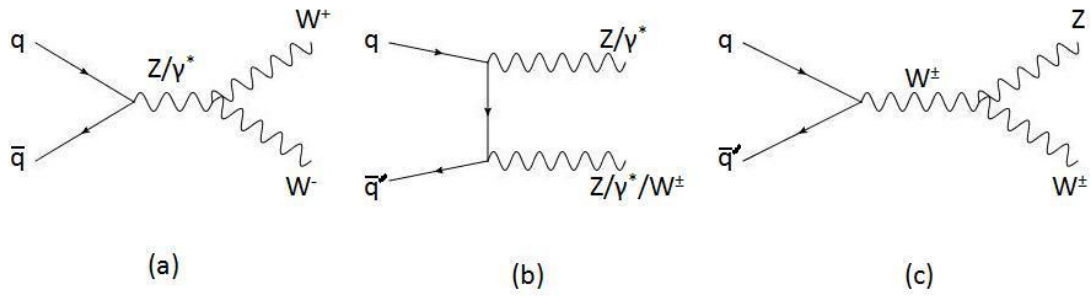

Figure 4.1: Tree-level diagrams of the diboson production channels at the LHC. Contributions to the WW continuum background (a) is a primary background in this analysis, while the other diboson channels (b - c) have slightly different signatures and smaller cross-sections.

4.1 Background Processes

The signature of the Higgs boson in this decay channel is $\ell\ell' + E_{\mathrm{T}}^{\mathrm{miss}}$+jets ($\ell, \ell' = e, \mu$), but this final state is not unique to the decays of the Higgs boson. Several other processes in the Standard Model can produce background to this signal. The irreducible backgrounds produce this exact signature of two leptons with missing transverse energy. Reducible backgrounds have a different final state from the Higgs signal, but occur with a cross-section large enough that the rate at which an object in these events is misidentified to form the signal final state is comparable to the signal production rate.

The primary irreducible background in this analysis is production of pairs of W bosons; it is called the WW continuum as these pairs have no constraint on the momenta of the decay products while the daughter particles of a Higgs boson's decay are restricted by the mass of the parent particle. Diagrams representing contributions to this background are shown in (a) of Figure 4.1 and in Figure 4.2. While the spin and mass of the Higgs boson give events resulting from these decays slightly different kinematics than this background generally, this background will be the dominant source of contamination for low jet multiplicity events.

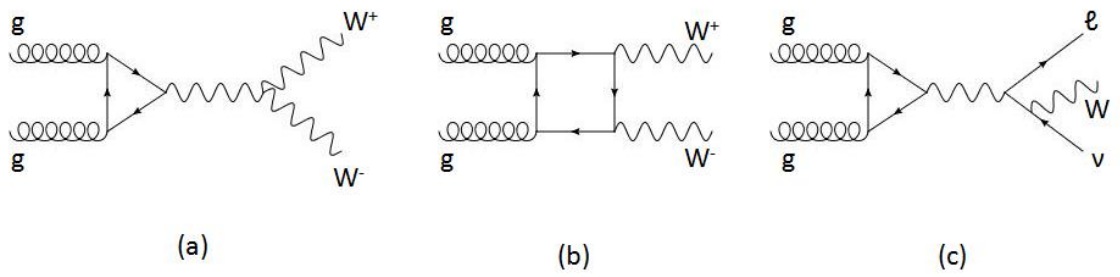

Figure 4.2: Tree-level diagrams for WW production from gluon fusion processes at the LHC.

Other diboson events such as WZ, $W\gamma$, and ZZ contribute to background contamination but are less significant than WW production; diagrams showing production of such events are (b) and (c) of Figure 4.1. Such events mimic the signal only if a lepton has failed to be reconstructed in the detector, as in the case of $WZ \to \ell\nu\ell\ell$, or if a particle has been misidentified, as when a photon is falsely identified as an electron in a $W\gamma$ event. While $ZZ \to \ell\ell\nu\nu$ events would produce the two leptons with missing energy signature, this background is not significant in part because the cross-section for this process is much smaller than other sources of background. Secondly, the visible leptons can be used to reconstruct the known Z mass, making such events easily identifiable.

The other primary source of irreducible background is from the decays of top quarks; the main production channels of such events are shown in Figure 4.3. Top quarks have a lifetime too short to hadronize and decay almost immediately into W bosons and b quarks; thus top quark pair production produces $WW+$ jets events that become background to the Higgs signal if the W bosons decay leptonically. These are the dominant source of background in the high jet multiplicity final states. One distinguishing characteristic of these events is that of containing jets resulting from B hadrons. The decays of the b quark to lighter quarks are suppressed, so B hadrons typically travel a measurable distance before decaying and produce narrow

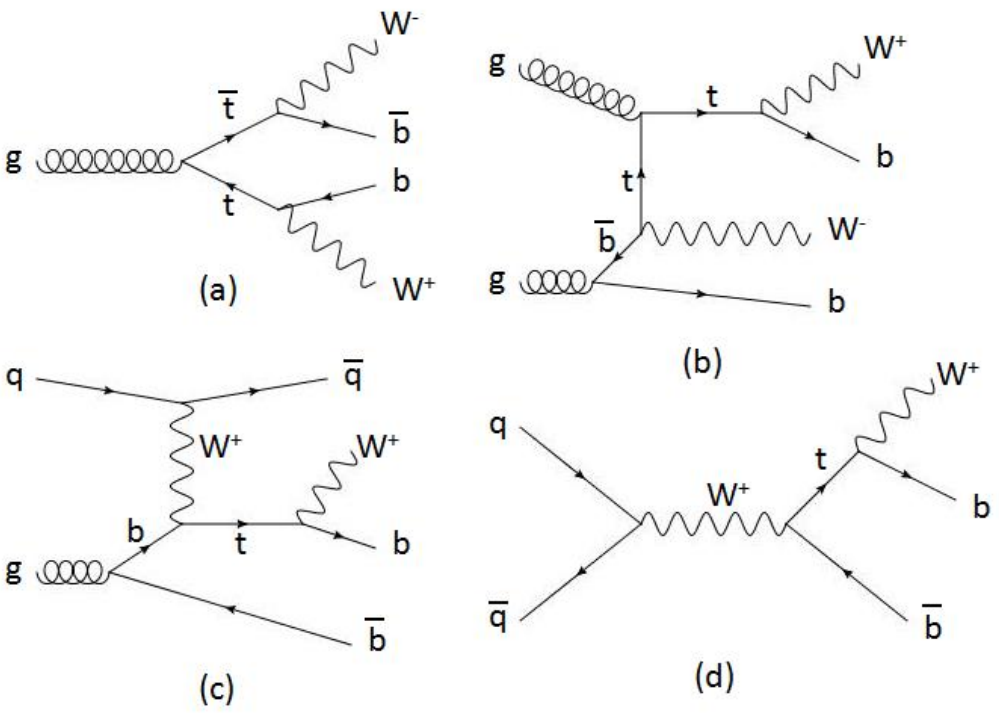

Figure 4.3: Tree-level diagrams for top quark pair production (a), associated production of top quarks (b), and the t-channel (c) and s-channel (d) production modes for single top quarks.

jets removed from the original interaction point. These charateristics can be used to remove such events, but the rate of production of top quarks at the LHC is sufficiently high that the contamination is still present.

The primary source of reducible backgrounds in the $H \to WW \to \ell\nu\ell\nu$ channel is the production of single vector bosons with jets; diagrams showing production channels of such events are in Figure 4.4. The $W \to \ell\nu$+jets process becomes a background when another object in the event, typically a jet, is falsely identified as a lepton. While stringent quality requirements on all leptons in the analysis can eliminate many of these events, the production of W bosons occurs at a rate several orders of magnitude higher than that of Higgs boson production at the LHC. The $Z/\gamma^* \to \ell\ell$+jets events also become a background if another object in the event, such as one of the leptons or a jet, is mismeasured and the event appears to contain missing

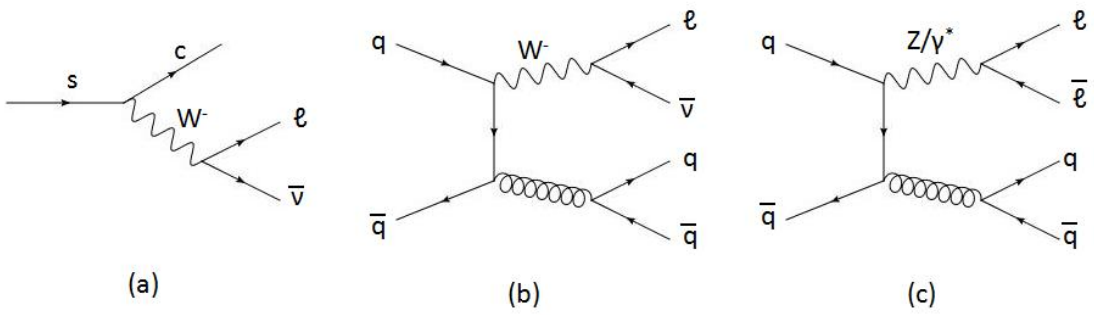

Figure 4.4: Diagrams showing the production of vector boson with jet events at the LHC. Such events can be produced through the decays of heavy quarks (a) or through radiation (b-c).

transverse energy. Again, the small likelihood of such mismeasurement is counterbalanced by the large number of events produced, and these backgrounds must be considered while searching for the Higgs boson.

4.2 Monte Carlo Generation

Generation of simulated events is done via Monte Carlo techniques. Using the known probability distributions and matrix elements describing interactions within the Standard Model, the final state of some specified interaction is modeled, including its kinematics, and its cross-section is calculated using perturbative techniques. As protons are not fundamental particles, PDFs that describe the possible characteristics of the quarks and gluons that could collide are also needed to determine the possible initial states. This modeling results in a set of particles produced from some interaction and the properties of those particles. Initial and final state radiation is also modeled, as are remnants of the initial colliding particles. The next step is to allow the partons to shower, or produce additional quarks via strong interactions. Finally, the strongly interacting particles are grouped into color singlet hadrons and decays of unstable particles are simulated. Additional interactions can then be added

to simulate pileup interactions within the event.

Monte Carlo generators come in two types, classified according to which parts of this process they handle. Matrix-element generators simulate the original hard interaction and calculate its cross-section. Multipurpose generators additionally have algorithms for the showering of particles and hadronization of final state quarks. Typically the output of a matrix-element generator is passed to a multipurpose generator to use its showering and hadronization algorithms. As this can produce overlapping sets of output, matching schemes [53–56] are applied to avoid double-counting of events and over-estimation of the cross-section.

The $H \to WW \to \ell\nu\ell\nu$ decay has been calculated to next to next to leading order for QCD generation processes and next to leading order for electroweak corrections for the gluon-gluon fusion interaction [57–63], vector boson fusion production [64–66], and production associated with W and Z bosons [67, 68]. Uncertainties on these calculations in this analysis are presented in Section 6.7. Decay branching ratios are calculated using HDECAY [69]. Cross sections and their uncertainties [37] and PDF uncertainites [70] have been calculated for the LHC running environment and are used here. Samples simulating the gluon-gluon and vector boson fusion Higgs boson production are generated using POWHEG [71, 72] interfaced to PYTHIA [73] for parton showering and hadronization. The Higgs-strahlung processes are also modeled by PYTHIA. As the simulated transverse momentum of the Higgs boson dictates strongly the properties of the daughter particles, the p_T spectrum of the Higgs boson is reweighted in these samples to match that achieved by theoretical calculations with transverse momentum resummation [74–77].

The irreducible WW continuum and $t\bar{t}$ background is modeled by MC@NLO [78–82] with HERWIG [83] used for parton hadronization. As this generator doesn't include production of WW pairs from gluon-gluon interactions, this contribution is modeled by gg2WW [84] in additional samples. The diboson WZ/ZZ processes are

Process	Generator	cross-section σ (pb) (x BR)
Inclusive $W \to \ell\nu$	ALPGEN	10.5x10^3 [91, 92]
Inclusive $Z/\gamma^* \to \ell\ell(M_{\ell\ell} > 40$ GeV)	ALPGEN	10.7x10^2 [92, 93]
Inclusive $Z/\gamma^* \to \ell\ell$ ($10 < M_{\ell\ell} < 40$ GeV)	ALPGEN	3.9x10^3 [93]
$t\bar{t}$	MC@NLO	164.6 [94]
Single top t-channel	AcerMC	64.2 [95]
Single top Wt	AcerMC	15.6 [95]
Single top s-channel	AcerMC	4.6 [95, 96]
WZ	SHERPA	18.0
ZZ	SHERPA	5.6
$qq/qg \to WW \to \ell\nu\ell\nu(\ell = e, \mu, \tau)$	MC@NLO	4.7
$gg \to WW \to \ell\nu\ell\nu(\ell = e, \mu, \tau)$	gg2WW	0.14
$\gamma W \to \ell\nu(\ell = e, \mu, \tau)$	MADGRAPH	135.4
$b\bar{b}$ (2ℓ filter, $p_T > 10$ GeV)	PYTHIA	4270

Table 4.1: Cross-sections of background processes for $\sqrt{s} = 7$ TeV, including branching fractions for processes with specific decay modes; processes without a specific decay mode are inclusive. Generators used are also noted.

simulated using SHERPA [85]. The $W\gamma^*$ background is modeled using MADGRAPH [86] with two lepton filters set at 15 GeV for the leading lepton and 5 GeV for the subleading lepton. AcerMC [87] is used for the single top in the s, t, and Wt production channels. The single $Z/\gamma^{(*)}$, W, and $W\gamma$ are modeled with ALPGEN [88] and HERWIG with the MLM matching scheme [55] is used for hadronization of the jets. MC@NLO uses the CTEQ6.6 PDF [89], ALPGEN uses CTEQ6L1 [89], and PYTHIA and HERWIG use MRST2007 [90]. The samples, their generators and their cross-sections are summarized in Table 4.1.

The data used for this analysis were taken over the course of 2011, during which time the instantaneous luminosity of collisions increased dramatically at the LHC. In particular, increasing the number of particles within a bunch and the number of colliding bunches resulted in events with more and more pile-up, or secondary interactions within a bunch crossing. To model this, Monte Carlo generated events were combined with additional pile-up events accompanying each hard collision; the percentage of events with a given number of pile-up interactions was chosen to match

Data Periods	Percentage of Data	Integrated Luminosity (pb^{-1})
B-D	3.2%	160
E-H	17.4%	872
I-K	25.8%	1291
L-M	53.5%	2677

Table 4.2: Percentage of Monte Carlo generated events generated with different pile-up conditions to correspond to data-taking periods in 2011.

to the major data-taking periods. Four different configurations, called periods to mirror the data periods that inspired this, were used and are detailed in Table 4.2. The Monte Carlo events are then normalized to the luminosity used within each data period prior to analysis.

4.3 Simulation of the ATLAS Detector

To produce realistic and therefore usable samples from Monte Carlo generation, interactions between the particles and the ATLAS detector must be simulated. This process, called simulation, is performed by GEANT4 [97]; while in wide use amongst the particle physics community, extensive work was undertaken between 2000 and 2003 to develop ATLAS-specific functionality and embed the program in ATLAS's software framework [98]. GEANT4 handles both propogation of particles through the detector and interactions between those particles and the detector's active and inactive elements. It produces hits, including information on the location of the interaction and amount of energy deposited. The program was stress-tested under pile-up conditions both in preparation for actual running conditions at the LHC's full luminosity and to ascertain that the main event's parameters would be treated independently of the number of underlying interactions occuring simultaneously.

Simulation relies on accurate information about the geometry of the ATLAS detector. Information about the size and location of the detector's components is contained in a database embedded in the ATLAS software framework and available for

both simulation and reconstruction processes, so misalignment of detector elements can be treated properly. Originally filled with information based on the design specifications, the ATLAS geometry is updated as alignment studies in subdetector systems or other calibration data become available.

4.4 Digitization

The reponse of the detector to the energy deposits produced during simulation is handled separately in a process called digitization. It handles, for example, propogation of charges in the tracking or liquid argon systems or that of light in the tile calorimeter, as well as the response of the readout electronics to these signals. As such, digitization must be handled separately for each subdetector with its specific technology. The goal is to produce a sample with the same format and basic properties as the data produced by the detector in real collisions, though Monte Carlo generated events add additional simulation-specific information. To ensure a realistic detector response, the raw data objects produced by digitization are compared to data produced in system tests and the digitization is tuned accordingly [99].

4.5 Object Reconstruction

Reconstruction is the process by which detector response information such as drift times or hits, either collision data from the ATLAS detector or Monte Carlo generated events after digitization, are grouped together and used to calculate the properties of a physics object such as a particle. Several levels of processing take place during reconstruction as information within a subsystem is combined and possibly calibrated, and then information from several systems combined to produce an object such as an electron or muon. The $H \to WW \to \ell\nu\ell\nu$ analysis makes use of six types of reconstructed objects: tracks, primary vertices, electrons, muons, jets, and missing

transverse energy. The details of the creation of these objects with information from the detector will be covered in the following sections.

4.5.1 Track and Primary Vertex Reconstruction

A charged particle passing through ATLAS's inner detector leaves behind a series of hits in the detector elements; these hits are reconstructed into a track. Tracks are then used in the reconstruction of leptons and primary vertices, or locations of the original interactions. This section will focus on inner detector tracks and electrons while muons with muon spectrometer tracks will be covered in Section 4.5.3.

Track reconstruction in the inner detector proceeds in three stages [100]. The first step is to process the hits in the silicon detectors to produce three-dimensional space points. For hits in the pixel detector, this is straightforward, as each pixel has a defined location within the detector. Within the SCT, hits in two layers are necessary to find a unique point, and so producing space points introduces noise suppression. Hits within the three layers of the pixel detector and the first SCT layer are used to form seeds, and seed tracks are then extended through the outer layers of the SCT to collect additional hits and form track candidates. A score is then computed for each track, with higher points for tracks with more hits and pixel hits weighted over SCT ones. Tracks with low scores are rejected [101]. This process removes fake tracks and resolves ambiguities of clusters shared by several track candidates. Finally the track is extended into the TRT to collect additional hits and the entire track is refit with the new information. The refitted track's score is compared to the original silicon-hit only score, and hits that degrade the score are marked as outliers. These tracks are then available for use in other reconstruction processes and offline analysis.

Tracks are also used to reconstruct primary vertices, or interaction points within the collision region. This is necessary to resolve underlying events from interesting collisions and other background sources. Here the ATLAS detector has also adopted

multiple algorithms [102], with some using tracks to identify proto-vertices and then iteratively improving calculation of the location while other perform the optimization in the process of associating tracks to the vertex.

4.5.2 Electron Reconstruction and Identification

Electron-related measurements within ATLAS are carried out primarily by the electromagnetic calorimeter and the inner detector. An electron created within a collision should first pass through the inner detector, bending in the ϕ-direction within the magnetic field and leaving hits within the silicon layers and drift tubes, and then pass through the solenoid magnet itself, before creating a shower of particles within the liquid argon calorimeter and being absorbed. Reconstruction of electrons begins with the calorimeter before incorporating tracking information.

Raw data from the EM calorimeter is calibrated for local problems to create cells, and cells are then clustered for further reconstruction. The middle layer of the LAr calorimeter, which is expected to contain 80% of an electron's shower, has cells $\Delta\eta \times \Delta\phi = 0.025{\times}0.025$ as shown in Figure 4.5.2; the strips layer has cells $\Delta\eta \times \Delta\phi = 0.003{\times}0.1$. This fine granularity allows for precision in measuring the angular coordinates of showers and their originating particles, but the energy of the shower is deposited over several cells that must be associated together. For electron and photon reconstruction, this is carried out by a sliding-window algorithm [103]. First, the energy deposited within $\eta \times \phi$ regions of the calorimeter through all layers is grouped to form towers. Second, a window of fixed size in η and ϕ is moved across the grid of towers to locate local maxima and form preclusters. Having located a precluster seed, the precluster position is calculated by finding the energy-weighted η and ϕ barycenters of all cells within a window around the center tower; this window is typically smaller than that used in precluster finding to keep the position calculation less sensitive to noise. Finally, a cluster is formed by collecting cells within a rectangle

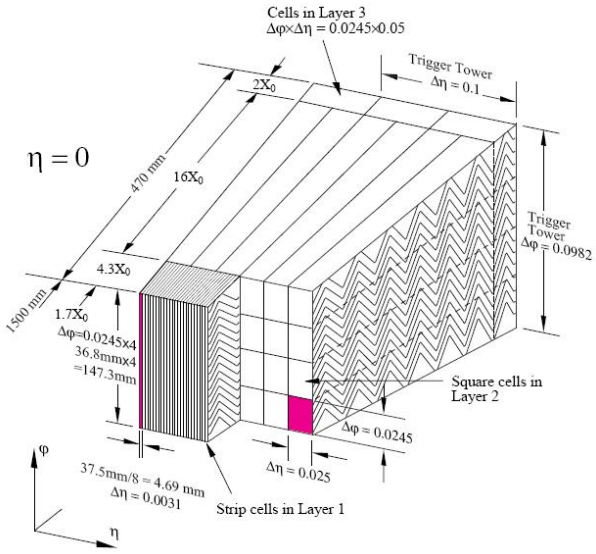

Figure 4.5: Diagram of barrel EM calorimeter module showing different layers and cell dimensions. ©2008 CERN [47]

centered on the calculated seed position. This is evaluated first in the middle layer, then in the strips layer using the middle layer's barycenter, and then finally in the presampler with the strips' barycenter and the back layer with the middle layer's barycenter. The cluster size used is dependent on the particle type and location within the detector. For electrons, the values are 3×7 cells in the barrel and 5×5 in the endcap. The extra width in ϕ in the barrel is necessary as barrel electron cross more material and therefore undergo more interactions while bending in the ϕ-direction, producing smearing. Cells are also smaller in the endcap, to preserve a finer granularity with the higher particle flux found there, so more cells are included to produce a similar sized cluster.

After calorimeter clusters are formed, electron reconstruction continues by matching clusters to tracks within the inner detector. For high transverse momentum electrons ($p_T > 10$ GeV), tracks within the inner detector are extrapolated from the last measurement point in the inner detector to the middle layer of the EM calorimeter. Tracks from photon-conversion pairs are not used. If the track is within $\Delta\eta \times \Delta\phi$

$= 0.05 \times 0.10$ of the cluster, the track is matched to the cluster. Several tracks can be matched to one cluster, in which case the track closest to the cluster is given precedence.

After a track has been matched, the object is considered an electron candidate and the cluster energy is recalculated to produce an electron energy. In addition to the energy deposits in the cluster, the estimated energy deposits in material before the calorimeter, within cells outside the cluster (lateral leakage), and beyond the EM calorimeter (longitudinal leakage) are included.

Electron candidates are produced by three sources: signal electrons from within collisions, background electrons from photon conversions and Dalitz decays, and hadronic jets that showered within the EM calorimeter. A series of quality criteria are used to classify candidates according to their likely source [104]. Three levels are used, called loose, medium, and tight, with an expected rejection of jets of 500, 5000, and 50000, respectively, as calculated from Monte Carlo. These levels are cumulative, such that a tight electron has fulfilled the loose and medium requirements as well. Loose electrons are selected by requirements on the shower shape within the EM calorimeter middle layer and a low level of hadronic leakage. The medium quality adds cuts on the energy deposits within the EM calorimeter strip layer, the quality of the track, and the quality of the match between track and cluster. Tight electrons additionally check the ratio between cluster energy and track momentum, transition radiation recorded by the TRTs, and for a hit in the pixel layer closest to the interaction point (called the b-layer). These last two are powerful for rejecting charged pions and photons converting to electrons. Cuts on these variables are optimized in 10 bins in cluster η and 11 bins in cluster transverse energy to account for the different behavior of electrons passing through different amounts of types of detector material with different energies and produce a clean sample of electrons for analysis.

4.5.3 Muon Reconstruction and Identification

Muon reconstruction and identification are the province of ATLAS's muon spectrometer. The muon spectrometer is capable of reconstructing muon tracks without additional information, but several algorithms for muon reconstruction exist and make use of information from the inner detector and calorimeters as well. This analysis uses STACO combined muons, described below.

Muon track reconstruction begins in the muon spectrometer by identifying regions of activity approximately $\Delta\eta \times \Delta\phi$=0.4×0.4 in size where there is at least one hit in a trigger chamber (either RPC or TGC) in both coordinates. The algorithm than creates pairs from hits within the region of activity of the same or adjacent stations. Extrapolation to other hits in the station is used to resolve any ambiguity in drift circle combination. The result is a straight-line segment loosely required to point to the interaction point; hits within a station are close enough that the bending of the track is neglible and a straight line adequately models the track. The procedure is run in two passes, the first strict fitting requires a second coordinate hit from a trigger chamber while the second tries several possible second coordinates. Finally, segments from the strict fitting are extrapolated to other stations using scans of the estimated momentum and knowledge of the magnetic field of the spectrometer; a schematic of the stations passed by straight tracks at different η values is shown in Figure 4.5.3. Only candidates containing two or more segments are retained. The hits within the segments are then globally refitted and hits that are outliers to the refitted track are excluded. The effects of material crossed by the muon are incorporated using a momentum-dependent parameterization of the expected interactions [105].

This algorithm, called the Muonboy algorithm [106, 107], produces a standalone or muon-spectrometer-only track. Such tracks could be produced by decays of heavy hadrons in the calorimeter, by cosmic rays, or by energetic jets that punch through the calorimeters into the muon spectrometer as well as by collision events. To reconstruct

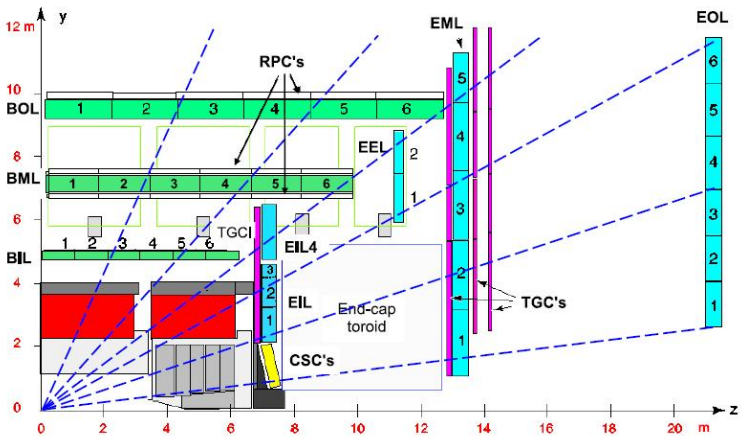

Figure 4.6: Diagram showing the stations crossed by muon tracks with different values of η. In general, tracks cross 2x4 inner, 2x3 middle, and 2x3 outer layers of MDT tubes. ©2008 CERN [47]

muons coming from collisions, standalone tracks are matched with inner detector tracks to form combined muons. The procedure for producing inner detector muon tracks is identical as that for electrons and is covered in Section 4.5.1. To match these tracks, the STACO algorithm uses the χ^2 value of the difference between ID and MS track vectors weighted by the combined covariance matrix to select matching tracks. The combined track is then created as a statistical combination of the two component tracks [108]:

$$T = (C_{ID}^{-1} + C_{MS}^{-1})^{-1}(C_{ID}^{-1}T_{ID} + C_{MS}^{-1}T_{MS}) \tag{4.1}$$

Unlike electrons, muons do not have seperate identification criteria classifying the quality of the muon; this is due to lower expected rates of particles besides muons reaching the muon spectrometer. Identification as muon tracks is implied for all tracks in the muon spectrometer.

4.5.4 Jet Reconstruction

The measurement of groups of hadrons called jets is done primarily by the liquid argon and tile hadronic calorimeters; reconstruction of jets starts by locating clusters of energy within the calorimeters indicating incident particles. While for electron reconstruction this was done with a sliding-window algorithm that produced clusters with a constant size, clusters for jet reconstruction are made using a topological clustering algorithm [103] and are variable in size. Clustering procedes in two steps: cluster making and cluster splitting.

Cluster making begins by producing a list of seeds, or cells in the calorimeter with a large signal-to-noise ratio. The signal in this case is the cell energy, while the noise is the expected root-mean-square of electronic noise with contributing pile-up. The neighbors of these cells are then associated with the seeds in proto-clusters if they have sufficient energy. A low threshold is used as criteria for addition of cells, which provides inherent noise suppresstion to the method, while a threshold between the seed defintion and inclusion thresholds is used to find neighbor seeds. The neighbors of neighbor seeds are considered in the next iteration of clustering. Proto-clusters sharing cells are merged, and when no more neighbors can be added, proto-clusters of sufficient energy are converted to clusters.

Cluster splitting begins by finding local maxima cells with at least 0.5 GeV, a higher energy than any of its neighbors, and at least four neighbors. Maxima in the middle layer EM calorimeter are primarily used, with maxima from the strips layer and hadronic calorimeters added if they do not overlap with the middle layers maxima. The cells found by the cluster-making process are then grouped with the local maxima using the same criteria as for cluster making, but without merging of clusters. Cells shared by multiple local maxima are added to all clusters with a weight dependent on the distance to the central maximum. The newly split clusters and those that didn't contain a viable maximum are then ready for jet reconstruction.

While jets are a common feature of events at the LHC, they are not fundamental objects described by quantum chromodynamic theories. A correspondence is needed between the predicted behavior of partons and observable quantities to allow for meaningful measurements. The theoretically sound links between the two are infrared and collinear safe quantities; this means neither the presence of additional soft particles between a jet's constituents nor the splitting of a constituent particle and its momentum into two collinear particles should have any impact on jet reconstruction [20, 109].

Within ATLAS, infrared and collinear safe jets are obtained using the anti-k_T jet clustering algorithm, a sequential recombination algorithm with negative power (p = -1) [110–112]. It is based upon the equations below:

$$\begin{aligned} d_{i,j} &= min(k_{T,i}^{-2}, k_{T,j}^{-2}) \frac{\Delta_{i,j}^2}{R^2} \\ d_{i,B} &= k_T, i^{-2} \end{aligned} \tag{4.2}$$

where $k_{T,i}$ is the transverse momentum of particle i, $\Delta_{i,j}$ is the angular seperation between particles i and j, R is the radius parameter used, and B references the position of the beamline. For this analysis, a radius of 0.4 is used. The algorithm proceeds by calculating the two distances shown in Equation 4.2 for a pair of objects in the event. If the object-to-object distance is smaller than the object-to-beamline distance, the two objects are combined. If the momentum relative to the beamline is smaller, the object is taken as finished and removed from further iterations. The negative power has the effect that hard particles dominate the behavior of the combination, while soft particles have less impact and will cluster to hard particles more than to other soft objects. Thus a single hard particle surrounded by soft particles would be reconstructed into a perfectly conical jet. Hard particles close to one another will reconstruct as multiple jets with a shared boundary, with the harder particle producing the more

conical jet.

Jet calibration requires additional steps, as the calorimeter response to electromagentically interacting particles is different from its response to hadrons. The default calibration of the calorimeters is an electromagnetic scale calibration taken from test beam data, $Z \to ee$ events, and minimally ionizing particle muons [113]. Calibration corrects the jet energy to hadronic scale and removes detector effects such as inactive material and the response of the electronics. The electromagnetic scale calibration is automatically applied to all calorimeter cells and applies to electron showers as well as hadronic jets. Additionally, a pileup correction is taken from subleading jets in data events passing the level 1 trigger. The momentum of clusters is corrected to point to the primary vertex with the largest sum of tranverse momentum amongst its tracks. Finally, a jet energy scale calibration is taken by comparing data to Monte Carlo simulation and applying a p_T and η dependent correction to data.

4.5.5 Identification of b-Jets

Jets can be produced by either quarks or radiated gluons, and the originating particle determines many properties of the resulting jet. This correlation is particularly strong for jets resulting from B hadrons, or b-jets. The b quark of a B hadron has relatively weak couplings to lighter quarks and therefore these hadrons have a longer lifetime, on the order of 1.5×10^{-12}s, than other heavy hadrons. Such hadrons can propogate about 450 μm from the point of origin, a measureable distance in the ATLAS detector, and produce jets with vertices distinct from the primary vertex, as shown in Figure 4.7. This and other properties of these jets can be used to tag jets as likely originating from b quarks.

Algorithms to remove b-jets are used in this analysis to remove events resulting from decays of top quarks. The IP3D algorithm [114] combines the transverse and longitudinal impact parameter significances of all tracks within a jet to select jets orig-

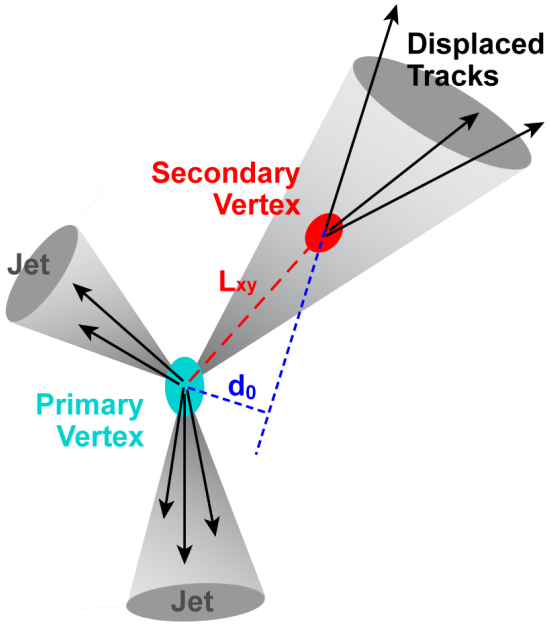

Figure 4.7: The structure of jets originating from light quarks or gluons compared to that of a b-jet. Jets from B hadrons originate at vertices removed from the original interaction and therefore have a larger impact parameter (d_0) and flight length (l_{xy}).

inating away from the primary vertex. The JetFitter algorithm [115] uses a Kalman filter to identify a line through the primary, b-decay, and c-decay vertices, allowing information about the B hadron's flight path and structure of decays within the jet to be used for identification. The output of these two algorithms and additional variables describing the event's topology are than passed to a neural network analyzer to produce a final discriminating variable [116]. Its application in this analysis is described in Section 5.3.

4.5.6 Missing Energy Reconstruction

The incident proton beams of the LHC carry no transverse momentum, and so hard collisions should also contain no net transverse momentum even if some of the constituent particles do. If these events also involve particles that don't interact with the ATLAS detector such as neutrinos, these particles will not be measured and the event's net transverse momentum will be non-zero. The amount of momentum necessary to bring this sum to zero is called the missing transverse energy, or $E_{\mathrm{T}}^{\mathrm{miss}}$. As $H \to WW$ decays always involve neutrinos, missing transverse energy is an important property of the events studied in this analysis. Detector effects such as dead regions, cracks in coverage, and noisy or hot components can also produce fake $E_{\mathrm{T}}^{\mathrm{miss}}$ and contribute to the background.

The missing transverse energy reconstruction used for this analysis is called refined final missing transverse energy and is based on all other reconstructed objects in the event [117]. Unlike algorithms that primarily use calorimeter clusters for calculation, object-based $E_{\mathrm{T}}^{\mathrm{miss}}$ can make use of the best available calibrations for each object and produce a quantity consistent with the other events in the object. It also allows greater sensitivity in analyses that make use of low p_T objects, such as Standard Model Higgs searches.

$E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction begins by assembling a list of topological clusters such as

those used for jet reconstruction and described in Section 4.5.4. This algorithm provides built-in noise suppression to the calculation via the thresholds used on the energy of cells included. These clusters are then compared to high p_T objects such as electrons, τ leptons, photons, and muons; clusters that overlap with these objects are removed from the list and replaced with the reconstructed objects. Isolated high p_T tracks not part of the other reconstructed objects are also included in the sum of visible transverse momentum. Clusters not part of any high p_T objects are then used to reconstruct mini-jets using a clustering radius of 0.2. Mini-jets are calibrated as π^0 hadrons, or electromagnetically calibrated, if the majority of the energy deposit is in the EM calorimeter, and the remaining mini-jets are calibrated as charged pions or hadronically calibrated. For the remaining clusters, the appropriate calibration is estimated. The visible momentum in the transverse direction is calculated from these objects, and its inverse is taken as the the missing transverse energy and used in physics analyses such as the search for the Higgs boson.

CHAPTER V

Event Selection

The process studied in this dissertation is that of the Standard Model Higgs boson decaying via two W bosons to leptons and neutrinos. The leptons considered are either electrons or muons, as τ leptons have lifetimes too short to interact with the ATLAS detector and have different reconstruction and background challenges. The neutrinos will escape detection and produce an event with real missing transverse energy. As the Higgs boson is expected to be neutral, the resulting leptons should have opposite charges. The data analyzed in the search for this signal and the trigger requirements imposed in recording the data are presented in Section 5.1. Requirements on the reconstructed objects included are discussed in Section 5.2. Finally, the topological selections used to differentiate the signal process from other particle interactions that also produce two oppositely charged leptons with missing transverse energy are discussed in Section 5.3.

Five processes contribute the majority of the background to this analysis; these were presented in Section 4.1. The irreducible backgrounds come primarily from WW and top quarks events, while the dominant sources of reducible backgrounds are from W+jets and Z/γ^*+jets events with falsely reconstructed or poorly measured objects. Other diboson production channels contribute a small reducible background that is not large compared to the other sources. Selections used to remove such events from

the final signal sample are discussed in this chapter, while methods to estimate the remaining contamination from these backgrounds will be presented in chapters VII and VIII.

5.1 Data Selection and Trigger Requirements

The data used for this analysis were obtained at the LHC between March 22 and October 30, 2011; the cumulative luminosity recorded by the ATLAS detector is shown in Figure 5.1. Proton-proton collisions were produced with a center of mass energy of 7 TeV. A period of data-taking is called a run, and runs are organized into run periods during which the LHC's operating conditions are constant. The ATLAS detector calculates both the absolute and relative luminosity of the recorded data in each run using several different detectors and methods both during data-taking and during offline analysis [118, 119]. The online measured luminosity is used to break the recorded data into luminosity blocks, with parameters such as data quality being defined per block. The performance of the ATLAS detector is constant during a luminosity block, and in the advent of a change such as a system failure, data in previous luminosity blocks will be properly marked as suitable for analysis. The results of the different luminosity algorithms are compared offline to determine the best luminosity measurement for a luminosity block, and analyses take ranges of blocks for the data sample. The integrated luminosity of an analysis is therefore dependent on the data quality selections. It also depends on the triggers used for an analysis, as many triggers are prescaled to select only a fraction of the events passing the trigger threshold and therefore sample only a fraction of the available data.

The quality of recorded data in ATLAS is described by a system of flags [121]. Flags are generated by comparing distributions produced by a detector system during data-taking with expected distributions. For example, the number of hits in the TRT can be compared to expectation to see if an unusually low or high numbers of tubes

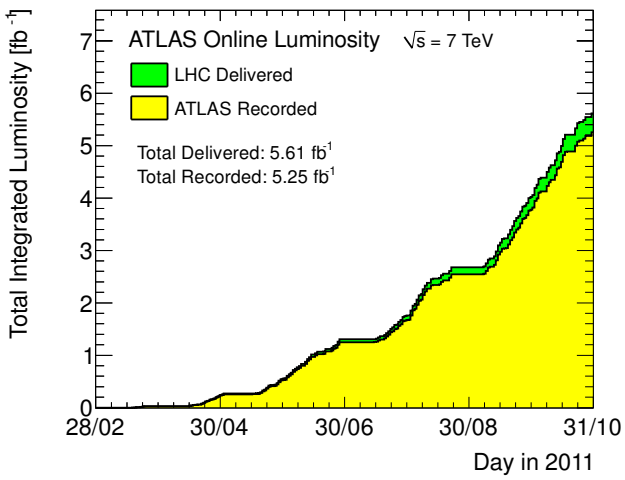

Figure 5.1: The cumulative luminosity delivered to ATLAS (green) and recorded by the detector (yellow) as a function of time for the 2011 data-taking. ©2011 ATLAS [120].

are being activated by collisions. If the distributions are in good agreement, the flags corresponding to that subsystem are set to indicate the data can be used for physics analysis. Flags are also set for systems such as the magnets. Composite flags are defined as logical combinations of detector flags to track the components needed for particular objects; for example, data quality flags for electrons are set by checking the status of inner detector, solenoid magnet, and electromagnetic calorimeter flags. This analysis requires that data be recorded with the magnet and all detector subsystems of ATLAS working well. This is necessary as the $H \to WW \to \ell\nu\ell\nu$ analysis studies missing transverse energy, and $E_{\mathrm{T}}^{\mathrm{miss}}$ can only be accurately reconstructed if information about objects in all parts of the detector is available.

This analysis makes use of unprescaled single lepton triggers with the lowest available momentum thresholds. During the 2011 data-taking, the running conditions and instantaneous luminosity of the LHC changed dramatically, which required changes to the prescales of the available triggers and higher p_T thresholds for triggers to remain unprescaled. The triggers used for this analysis are summarized in Table 5.1 along with the periods they were applied to. The numbers following the particle identifier

Period	e triggers	μ triggers
B-I	EF_e20_medium	EF_mu18_MG
J	EF_e20_medium	EF_mu18_MG_medium
K	EF_e22_medium	EF_mu18_MG_medium
L-M	EF_e22vh_medium1	EF_mu18_MG_medium

Table 5.1: Single lepton triggers used for this analysis. Events in the ee and $\mu\mu$ channels must fire triggers of that flavor; events in the $e\mu$ channel are accepted if they fire either of the two chains.

(e or mu) indicate the p_T threshold used. For electron triggers, the phrase "medium" means medium level identification requirements were imposed and the "vh" means η-dependent p_T thresholds were used. For the muon triggers, the "MG" indicates the reconstruction algorithm used and the phrase "medium" means a higher level 1 trigger threshold was used to start the trigger chain. Events in the ee or $\mu\mu$ channels are required to fire the e or μ triggers respectively, while events in the $e\mu$ channel may fire either flavor triggers. With these requirements, the integrated luminosity of data analyzed in this dissertation is 4.712 fb^{-1}with a 3.7% systematic uncertainty discussed further in Section 6.7.

Additional quality requirements are imposed on the selected data to remove events within luminosity blocks that show evidence of detector problems, cosmic ray activity, or remnants of earlier collisions. This is done by requiring the event contain a primary vertex with three or more associated tracks where a track is required to have $p_T >$ 0.4 GeV. This removes contamination from cosmic rays that penetrate the detector and produce events that appear to have two isolated leptons. The energy distributions of jets in an event are checked as described in [122] to remove events with evidence of poor calorimeter performance. If it is found that a small region or small number of cells in one of the calorimeters contains most of the energy of a "jet," or if the "jet" appears well after or before a collision, the event is rejected. In such cases, the missing transverse energy cannot be trusted to be accurately calculated and the event cannot be analyzed.

5.2 Lepton and Jet Selection

The main feature of the final state studied in this analysis is two oppositely signed isolated leptons, either electrons or muons. Candidate leptons are reconstructed as described in Sections 4.5.2 and 4.5.3, but additional requirements are necessary to ensure well reconstructed leptons with low contamination from fake objects. For electrons, candidates are required to pass the requirement $|\eta| < 2.47$ and must not lie within the range $1.37 < |\eta| < 1.52$; this ensures that the electron's track passed within the inner detector and that the electron did not pass through the barrel and endcap calorimeter cryostats where its energy would be poorly reconstructed. Additional cleaning cuts remove electrons that hit regions of the detector with hardware problems such as dead electronic connections or data acquisition boards. Tight electrons as defined in [104] are used in this analysis. The electron is also required to originate from the primary vertex selected previously by imposing requirements on the electron's impact parameter, or the distance of closest approach between the track and the primary vertex. This is required to be less than 1 mm in the longitudinal direction and the transverse impact parameter significance, or ratio of the impact parameter to its measured uncertainty, must be less than 10. Electrons analyzed have a transverse momentum of at least 15 GeV and are isolated, with little energy deposited around them in the detector. This removes jets that are reconstructed as electrons and electrons produced in hadronic decays. The sum of the energy deposited in a cone of radius 0.3 is calculated using clusters in the calorimeter and using tracks in the inner detector. Both of these sums are normalized to the electron's momentum, and the calorimeter isolation is required to be less than 0.14 and the track isolation is less than 0.13.

Candidate muons are reconstructed via the STACO combined algorithm described in Section 4.5.3. They are required to have $|\eta| < 2.4$; this is smaller than the η range of the muon spectrometer but selects only muons that have inner detector tracks.

Additional cuts are imposed on this inner detector track to ensure it is well measured. To reject cosmic ray muons that may have passed earlier cuts, the impact parameter of the muon is required to be within 1 mm of the primary vertex in the longitudinal direction and have a transverse impact parameter significance of less than 3. Muons must have a transverse momentum of at least 15 GeV. Like the electrons, these high p_T muons use the energy deposited in a cone of 0.3 around the muon in the calorimeter and inner detector normalized to the muon's momentum as discriminating variables; the thresholds are 0.14 and 0.15 for calorimeter and tracking isolation respectively.

This analysis uses anti-k_T jets reconstructed as described in Section 4.5.4. Jets are defined as having at least 25 GeV of transverse energy and $|\eta| < 4.5$. In the range $2.75 < |\eta| < 3.25$, an excess of jets is observed that was found to be due to pile-up in the transition region between the endcap and forward calorimeters. The p_T threshold in this range was raised to 30 GeV to keep this background small. The probability that a jet originated at the primary vertex under study and not from pile-up is also computed by associating tracks between the primary vertex and jet; this method was first used by the D0 collaboration [123] and has been adapted for use in ATLAS [124, 125]. A jet is required to have 75% of the transverse momentum of its constituent tracks originate from the primary vertex.

Overlaps between objects are addressed after all objects are selected. Electrons found to be within a cone of 0.1 of a selected muon are removed as likely produced by radiation from the muon. It is also possible for a calorimeter cluster to be used in the reconstruction of multiple electrons or a jet and an electron. If two electrons are found to be within a cone of 0.1, the one lowest in p_T is removed as being a duplicate. If a jet is within a cone of 0.3 of an electron, it is removed for the same reason.

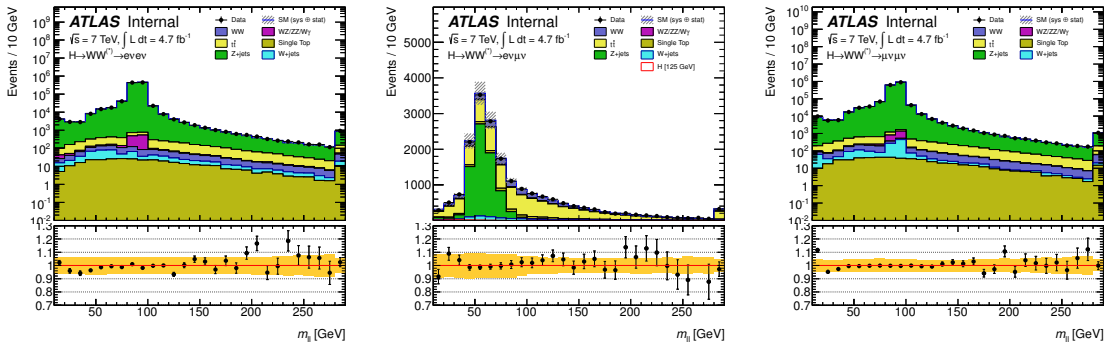

Figure 5.2: $m_{\ell\ell}$ distributions in the ee (left), $e\mu$ (center), and $\mu\mu$ (right) channels after selecting events with two opposite signed leptons, before removal of low mass resonances and the Z peak. All jet multiplicities are combined.

5.3 Topological and Jet Selections

Events for further analysis are selected as having exactly two leptons with the properties defined in the previous section. At least one of these leptons must have a transverse momentum of 25 GeV; this lepton is required to match the object firing the trigger for this event, and the trigger is approximately 100% efficient for leptons with this momenta. The leptons are also required to have opposite charge. Such events are dominated by two-body decays of resonances such as the Z boson and J/ψ particle, as shown in Figure 5.2. To remove these, limits on the dilepton mass ($m_{\ell\ell}$) are imposed around these resonance peaks. To remove the low mass resonances, $m_{\ell\ell}$ must be larger than 12 GeV for the same flavor (ee, $\mu\mu$) channels and 10 GeV for the opposite flavor channel ($e\mu$). The same flavor channels also have an explicit veto of Z boson events by requiring that the dilepton mass of the event is more than 15 GeV away from the mass of the Z boson, taken to be 91.19 GeV [126]. Many of the remaining events from reducible backgrounds are rejected by imposing a cut on the relative missing transverse energy. This is defined as:

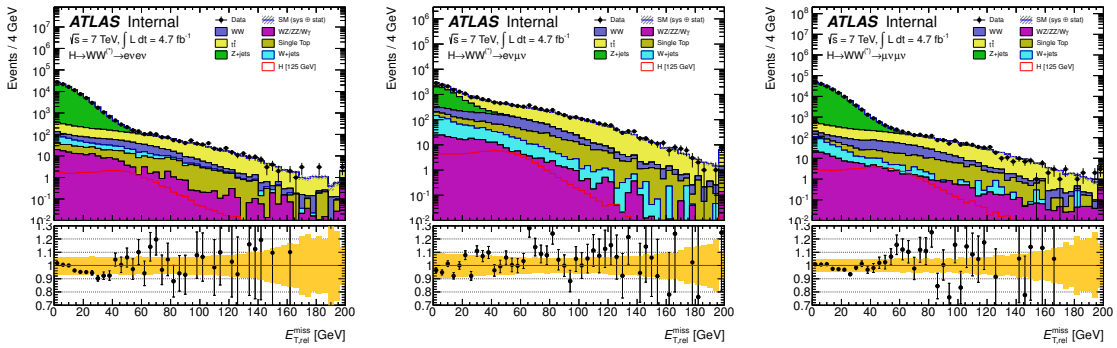

Figure 5.3: $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ distributions for the ee (left), $e\mu$ (center), and $\mu\mu$ (right) channels after $m_{\ell\ell}$ selections. All jet multiplicities are included.

$$E_{\mathrm{T,Rel}}^{\mathrm{miss}} = \begin{cases} E_{\mathrm{T}}^{\mathrm{miss}} & \mathrm{if} \Delta\phi \geq \frac{\pi}{2} \\ E_{\mathrm{T}}^{\mathrm{miss}} \sin\Delta\phi & \mathrm{if} \Delta\phi < \frac{\pi}{2} \end{cases} \tag{5.1}$$

where $\Delta\phi$ is the azimuthal angle between the missing transverse energy and the closest lepton or jet. This decreases the significance of the missing transverse energy if it is parallel to a visible object; $E_{\mathrm{T}}^{\mathrm{miss}}$ close to an object is often produced by mismeasurement of the object, and scaling down the missing energy increases the effectiveness of this cut in removing such events. The relative missing transverse energy is required to be greater than 45 GeV for the same flavor channels as they suffer from more contamination from reducible backgrounds as shown in Figure 5.3. The opposite flavor channel has less contamination from these sources and the threshold used here is 25 GeV to keep signal acceptance high.

The three flavor channels are further divided into three jet multiplicity channels: zero jets, exactly one jet, and two or more jets. The relative abundance of each multiplicity is shown in Figure 5.4. As each jet multiplicity suffers from different backgrounds, additional selections are applied to each jet multiplicity. In the zero jet channel, one additional cut is applied to the dilepton transverse momentum ($p_T^{\ell\ell}$). This is required to be greater than 45 GeV for the same flavor channels and 30 GeV

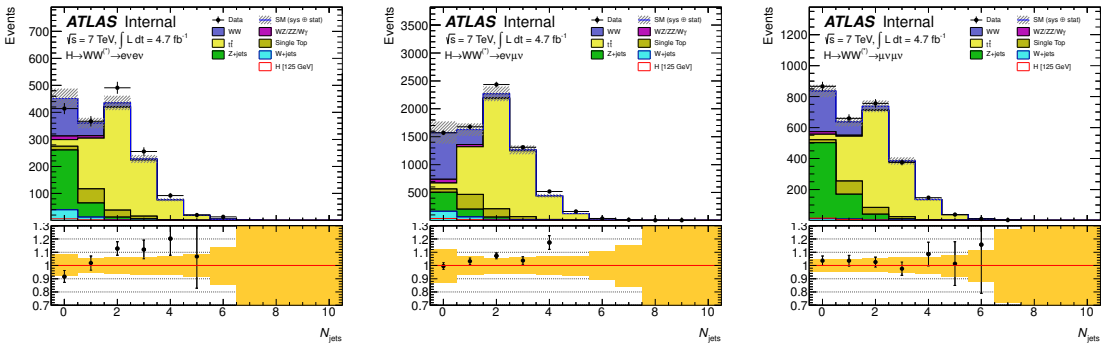

Figure 5.4: The jet multiplicity distributions for ee (left), $e\mu$ (center), and $\mu\mu$ (right) channels after the $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ requirements.

for the $e\mu$ channel. This requirement is aimed at removing Z/γ^*+jets contamination from the analysis, which can be a significant background in the high pile-up conditions at the LHC.

In the one jet channel, a veto is applied to events where the jet has an 80% probability or higher of originating from a B hadron. This probability is calculated by a neural network that works on the output of two jet-tagging algorithms, one that compares the longitudinal and transverse impact parameters of a jet's tracks to check for secondary vertices (called IP3D) and another that uses a Kalman filter to check the event topology for distinct vertices [114–116]; the procedure for this was presented in Section 4.5.5. The output of the neural network is required to be greater than -1.25, with this operating point derived from top events as described in Reference [127] and being 80% efficient at identifying b-jets. This cut removes contamination from top quark decays in this channel. Events in the one jet channel must have total transverse momentum (p_T^{Tot}), or the vector sum of the leptons', jet's, and missing transverse energy's momentum vectors, of less than 30 GeV. This removes events with high levels of soft hadronic energy that failed to be reconstructed as jets. Finally, $Z \to \tau\tau$ events are rejected by reconstructing the mass of the assumed Z boson assuming the neutrinos are collinear with the visible leptons. If this mass

is found to be within 25 GeV of the nominal Z boson mass, the event is rejected. The collinear approximation is not useful in the zero jet channel as the leptons are typically opposite in ϕ and so this veto is not applied there.

Events with two or more jets are handled differently as the signal in this channel is expected to come primarily from vector boson fusion and not from gluon-gluon fusion, the dominant production channel in the zero and one jet channels. Decays from vector boson fusion production of Higgs bosons are expected to have two energetic jets with a clear gap in rapidity between them. To select these, the jets are required to be in opposite hemispheres in η with the difference in η between them satisfying $|\Delta\eta_{jj}| > 3.8$. Additionally, the dijet mass is required to be greater than 500 GeV and the event must not contain a third jet within the central region ($|\eta| < 3.2$) of the detector. Events in the two jet channel must also satisfy the b-tag veto, total transverse momentum cut, and $Z \to \tau\tau$ veto described for the one jet channel.

Additional topological selections are applied to all channels depending on the range of Higgs boson masses under investigation. For masses less than 200 GeV, the spin-0 nature of the Standard Model Higgs boson and limited helicity states available to neutrinos implies the visible leptons will be correlated in direction. This means the azimuthal dilepton opening angle will be small, which implies also that the dilepton mass will be small. These features are not present in the WW continuum events and are therefore important discriminants against this background. For the zero and one jet channels, the dilepton mass is required to be less than 50 GeV; for the two or more jet events, $m_{\ell\ell}$ must be less than 80 GeV. Additionally, the opening angle between leptons ($\Delta\phi_{\ell\ell}$) must be less than 1.8 radians. For heavier masses of the Higgs boson, the decay products receive a boost that weakens the correlation in lepton direction, so to maintain a reasonable efficiency on the signal the cut on opening angle is removed and the dilepton mass cut is raised to $m_{\ell\ell} < 150$ GeV. The dilepton mass requirements previously mentioned to remove resonances are still applied. For very high masses of

above 300 GeV, these requirements on $m_{\ell\ell}$ and $\Delta\phi_{\ell\ell}$ are removed entirely.

The final variable of interest in this analysis is the transverse mass; it is defined as:

$$\sqrt{(E_T^{\ell\ell} + |E_T^{miss}|)^2 - (\overrightarrow{p}_T^{\ell\ell} + \overrightarrow{p}_T^{miss})^2} \tag{5.2}$$

The missing transverse energy in the decay of a Higgs boson is the sum of the two neutrinos' momenta and so the exact mass of the boson cannot be reconstructed. The transverse mass is constructed such that the its upper limit is the mass of the Higgs boson for such events, and distributions of this variable are expected to have a shoulder if the Higgs boson is present. This variable will be used to check agreement between predictions and observed data throughout this analysis. Unlike previous analyses of this type, no cuts are applied to the transverse mass; instead, the transverse mass distributions are fit in the process of determining the significance of the observed data and in setting limits, described in Section 9.2.

CHAPTER VI

Efficiency Corrections to Monte Carlo Simulation

Generation and simulation of physics processes and detector effects are critical components of understanding the data taken by the ATLAS detector at the LHC. However, not all effects can be perfectly modeled. Furthermore, all analyzed data is acquired only after passing trigger and reconstruction requirements that are not perfectly efficient. These efficiencies must be known before the cross-section of a process can be measured. Thus generated Monte Carlo events must be corrected to match the observed properties of data before information about the underlying processes can be extracted. Corrections are applied by assigning a weight to the event, where the weight carries information about how common an event's set of properties are.

This chapter discusses the corrections and weights applied to events to produce better agreement between data and Monte Carlo. Section 6.1 describes the tag and probe method used to determine many of the efficiencies and corrections included in this analysis. Calculation of lepton trigger efficiencies is discussed in Section 6.2 and that of lepton reconstruction efficiencies in Section 6.3. Events passing the b-tag veto are also receive an efficiency correction described in Section 6.5. Additional corrections are also applied to specific properties of simulated events to produce better agreement between data and simulation for specific distributions. Section 6.4 de-

scribes corrections of this type applied to the energy of electrons and momentum of muons. The Z/γ^* Monte Carlo samples receive an additional correction presented in 6.6.

The same effects that produce the need for these corrections also produce the experimental systematic uncertainties in this analysis. As many of the same techniques used for efficiency calculations are used to determine these uncertainties, they are discussed in Section 6.7. This section also presents systematic uncertainties due to signal cross-section calculations.

6.1 The Tag and Probe Method

Before an object can be studied in ATLAS, it must first trigger the data acquisition system and be successfully reconstructed. Measuring the efficiencies of the trigger and reconstruction algorithms is intrinsically difficult as objects that fail these processes are either not recorded or combined with a huge background of fake objects. The most common method to circumvent these difficulties and measure these efficiencies is the tag and probe method [128, 129]. As it is used frequently to derive the efficiencies discussed in later sections, it is introduced here.

The tag and probe method relies on processes that produce two easily identifiable objects. The most common process is that of $Z \to \ell\ell$ with two visible leptons in the final state, but $W \to \ell\nu$ is also used with one lepton and missing transverse energy in the final state. One of these object is selected as the tag object and required to pass stringent quality cuts to reduce contamination from other backgrounds and fake objects. All such events with a good tag object are taken as the initial sample, and the cut of interest is then applied to the second probe object to define a final sample. The efficiency of the cut is then taken as:

$$\epsilon = \frac{N_{final}}{N_{initial}} \tag{6.1}$$

6.2 Lepton Trigger Efficiencies

Electron trigger efficiencies are calculated using the tag and probe method described in Section 6.1 on three samples of data events: $W \to e\nu$, $Z \to ee$, and $J/\psi \to ee$ as described in [104]. Events are selected with triggers on missing transverse energy for W events and single electron triggers for Z and J/ψ decays; the triggering or tag objects are required to match to the triggering Region of Interest with $\Delta R < 0.15$ to ensure the probe object is unbiased by the trigger selection (primarily a concern for the dilepton decays). Additional requirements are applied as follows:

- $W \to e\nu$: Events must have $E_{\mathrm{T}}^{\mathrm{miss}} > 25$ GeV, transverse mass greater than 40 GeV, and $\Delta\phi$ between the missing transverse energy and any jet with momentum greater than 10 GeV of more than 2.5 radians. The probe electron must pass through the inner detector with $|\eta| < 2.47$ and have a transverse energy greater than 15 GeV.

- $Z \to ee$: The tag electron must have transverse energy greater than 20 GeV and pass the tight identification criteria discussed in Section 4.5.2. The probe electron must be of opposite charge and have a transverse energy greater than 15 GeV. The dilepton mass of the event must fall within $80 < m_{ee} < 100$ GeV, close to the measured mass of the Z boson.

- $J/\psi \to ee$: The tag electron must have transverse energy greater than 5 GeV and pass the tight identification criteria discussed in Section4.5.2. The probe electron must be of opposite charge and have a transverse energy greater than 4 GeV. The two electrons must be separated by $\Delta R > 0.1$. The dilepton mass

of the event must fall within $2.5 < m_{ee} < 3.5$ GeV, close to the measured J/ϕ mass.

The efficiency for different triggers can then be calculated by counting the number of probe electrons that pass each trigger; these are calculated for both data and Monte Carlo in different η regions. Scale factors are then calculated by taking the ratio between the data and Monte Carlo efficiencies.

A similar tag and probe method using $Z \to \mu\mu$ events is used to establish the efficiencies and corresponding scale factors for the muon trigger chains. As this analysis involves two leptons, each event generates a trigger scale factor per lepton which are combined to apply to the event weight as:

$$SF_{event} = \frac{[1 - (1 - \epsilon_{MC}^{lead} \times SF^{lead}) \times (1 - \epsilon_{MC}^{sub} \times SF^{sub})]}{[1 - (1 - \epsilon_{MC}^{lead}) \times (1 - \epsilon_{MC}^{sub})]} \tag{6.2}$$

where ϵ_{MC}^{lead}, ϵ_{MC}^{sub} are the per-lepton trigger efficiencies in Monte Carlo for the leading and sub-leading leptons, and SF^{lead}, SF^{sub} are the per-lepton scale factors for the leading and sub-leading leptons.

6.3 Lepon Reconstruction Efficiencies

Calculation of the electron reconstruction efficiences is carried out with the tag and probe method as outlined in Reference [104]. Selection of the tag objects is identical to that used to determine the trigger efficiencies described in Section 6.2. To calculate the reconstruction efficiency, all clusters created by the sliding-window algorithm [103] are taken as probe objects and the number successfully reconstructed is used for the efficiency calculation. The efficiency of an electron being reconstructed with a certain identification level such as medium or tight, where tight is the criteria used in this analysis, is calculated identically to the method used for the trigger efficiency calculation with the identification criteria replacing the checks on the trigger. The

Monte Carlo event's weight is then multiplied by the reconstruction and identification scale factors of its component electrons in the analysis.

As muons do not have separate identification criteria from the reconstruction requirements, only reconstruction efficiencies are calculated. The tag and probe method is again used on $Z \to \mu\mu$ events. The tag muon is required to fire the unprescaled trigger with the lowest momentum threshold for the data-taking period in question and match to the primary vertex of the event [130]. It must have $p_T > 20$ GeV and fall in the range $|\eta| < 2.4$; this is the extent of the trigger range of the muon spectrometer. The tag muon must be a combined muon as described in Section 4.5.3 and be isolated with the sum of the energy deposited by tracks around the muon in question within a cone of 0.4 normalized to the muon's transverse momentum less than 0.2. The event must contain a second muon, which can be either an inner detector track or standalone muon track depending on which efficiency is being calculated. The dilepton mass of the system is required to be within 10 GeV of the nominal Z boson mass. The percentage of probe muons that successfully matches to an inner detector track is used to calculate the reconstruction efficiency and the ratio between these efficiencies in data and Monte Carlo is used to weight the generated events used in the analysis.

6.4 Energy and Resolution Smearing

Additional corrections are applied to events to correct specific distributions that may be poorly modeled in the Monte Carlo or supplement the calibration of specific aspects of the detector. For electrons, all clusters within the electromagnetic calorimeters are automatically calibrated with constants derived using test beam data. This is supplemented by a cluster energy correction derived from $Z \to ee$ and $J/\psi \to ee$ events as described in Reference [104]. The two energy ranges probed allow the linearity of the electromagnetic calorimeter to be checked. For Z decays in the barrel

with a dilepton mass between 80 and 100 GeV, the two electrons are required to be identified as medium and with opposite sign. Pairs are also made with a tight electron passing through the barrel and a loose electron passing through the forward region; as the forward electron does not pass through the inner detector tracking system its charge is unknown and so no charge requirement is imposed. For J/ψ decays, the dilepton mass must be between 2.5 and 3.5 GeV and all electrons are required to be tight. The energy scale corrections α are calculated by minimizing the negative unbinned log-likelihood:

$$-\ln L_{Tot} = \sum_{k=1}^{N_{events}} -\ln L_{i,j}\Big(\frac{m_k}{1+\frac{\alpha_i+\alpha_j}{2}}\Big) \tag{6.3}$$

where m_k is the measured dilepton mass, α_i and α_j are the leptons' energy scale corrections, and $L_{i,j}$ is a probability distribution function describing how compatible the event is with the Z lineshape. The corrections are calculated in different η regions to account for the η-dependence of the material distribution. These are then applied to both data and Monte Carlo events.

For Monte Carlo generated muons, the transverse momentum must be adjusted to match the momentum resolution of Monte Carlo to that of data. The muon resolution was measured using data taken when the toroid magnets in the muon spectrometer were shut down; this produced straight muon tracks that did not have smearing due to the track's curvature. The results are documented in Reference [131]. The corrections are applied to both the muon spectrometer and inner detector track momenta, and the combined track momentum is recalculated and used for analysis.

6.5 b-Tagging Efficiency

Reweighting of the Monte Carlo is also done to account for different efficiencies of the b-tagging algorithm in data and Monte Carlo. This analysis uses a neural net

combination of two b-tagging algorithms as described in Section 5.3. The methods used to derive the working points and their efficiences are described in Reference [127] and have been updated with additional tagging algorithms and the 2011 data. Four methods are used to calculate the efficiency of the b-tagging algorithms. The first uses b-decays that produce a muon and studies the muon momentum transverse to the muon and jet axis. Templates for this distribution are produced for muons resulting from b-decays, c-decays, and lighter quark decays, and a linear combination of these templates is fit to the data both before and after b-tagging so that the b-component can be extracted. This method assumes b-tagging is equally efficient in hadronic decays as in leptonic ones.

A second method selects a sample of $b \rightarrow X\mu D^* \rightarrow X\mu D^0(\rightarrow K^-\pi^+)\pi^+$, as reconstructing the $\mu + D^0$ within a jet provides a very clean sample of b-jet events. The algorithms can then be run on this sample and the efficiency calculated assuming all input events contain b-quarks. The branching ratio for this process is small, however, and the results of this estimation carry large statistical errors.

The final methods rely on selecting a sample of $t\bar{t}$ events. As the branching ratio of $t\bar{t} \rightarrow b\bar{b}W^+W^-$ is close to one, a significant fraction of these decays produce two-lepton events that are naively expected to contain exactly two b jets. The leptons can be used to select a sample of these events and after taking contributions from other backgrounds into account the efficiency of the algorithms can be calculated. It is also possible to require such events to contain at least one b-tagged jet and then use the other jets for the calculation. After calculating these efficiencies in both data and Monte Carlo, the ratio between the two is used to reweight Monte Carlo events in the one and two jet channels in which the b-jet veto is applied.

6.6 Drell-Yan $p_T^{\ell\ell}$ Reweighting

Observed discrepancies between data and Monte Carlo in the low $m_{\ell\ell}$ regions, dominated by Z/γ^* events, are treated by reweighting of these events. Requiring events with $m_{\ell\ell} < 2 \times p_T$ and treating the analysis in exclusive jet bins focuses on distributions for this background that are unlikely to be well normalized with the inclusive K-factor calculations used and produces poorly modeled distributions of dilepton transverse momentum ($p_T^{\ell\ell}$). To address this, the $p_T^{\ell\ell}$ vs $m_{\ell\ell}$ plane is divided into bins; steps are taken every 10 GeV between 10 and 110 GeV in $m_{\ell\ell}$ and with bin edges (0,5,10,20,30,40,50,60) in $p_T^{\ell\ell}$. Contributions from other backgrounds are subtracted from the observed number of events in each region; these contributions are estimated from Monte Carlo. The ratio between data and Z/γ^* Monte Carlo is than taken as a reweighting factor and applied to this background in all distributions. Weights are calculated in the ee and $\mu\mu$ channels for the zero and one jet bins. The effect of this reweighting can be seen in Figure 6.1; these plots show the $p_T^{\ell\ell}$ distributions of the signal region ($12 < m_{\ell\ell} < 50$ GeV) for the $\mu\mu$ zero jet channel before and after this reweighting.

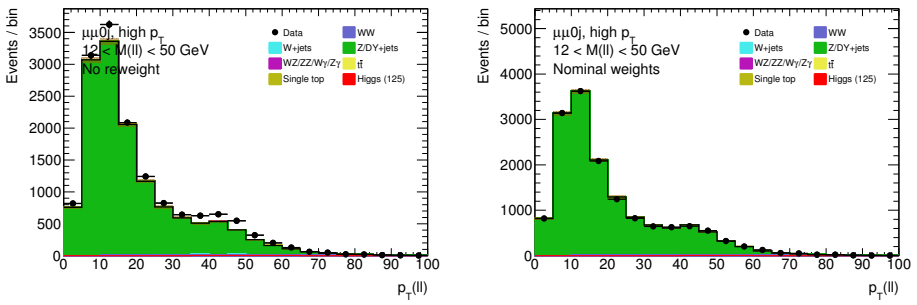

Figure 6.1: Distributions of $p_T^{\ell\ell}$ in the $\mu\mu$ zero jet channel in the range $12 < m_{\ell\ell} < 50$ GeV before (left) and after (right) reweighting in the $p_T^{\ell\ell}$ vs $m_{\ell\ell}$ plane.

6.7 Systematic Uncertainties

Systematic uncertainties enter the analysis through uncertainties on the calcuations of cross-sections, through uncertainties on detector effects, and through uncertainty on the luminosity calculation. The primary sources of theoretical systematic uncertainties come from the calculation of the expected Higgs boson cross-section. The calculations used were discussed in Section 4.2, and included at most next to next to leading order terms. The uncertainty on the gluon-gluon fusion calculation due to the omission of higher-order QCD radiative corrections is between 5 and 10%, with an additional 1% uncertainty due to omission of higher order electroweak corrections [37]. The calculations also use a large top mass approximation and incorporate uncertainties on the measured top and bottom quark masses, which incorporate uncertainties of less than 5%. The vector boson fusion channel has the same sources of uncertainties with a total additional 1-2%. The PDFs used introduce an uncertainty of an additional 8%.

Uncertainty is also introduced by the division of the analysis into bins of jet multiplicity, as the identification of a jet is based on properties that are not fundamental. This introduces a 7-10% uncertainty for the zero jet bin, a 17-21% uncertainty for the one jet bin, and a 65-77% uncertainty for the two or more jet bin, with the range of values covering the entire range of Higgs boson masses assessed in this analysis.

Detector-related sources of systematic uncertainties come from object reconstruction and identification efficiencies, momentum or energy resolutions, momentum or energy scales, and measurement of the luminosity. The luminosity of data included in this analysis is determined by van der Meer scans and its uncertainty is calculated to be 3.7% [132]. Uncertainties on lepton identification and reconstruction are measured using the tag and probe methods described earlier in this chapter. For electron triggering, identification and reconstruction, calculated uncertainties are approximately 1% each, with muon identification and reconstruction also 1% [133]. Uncertainty on

the energy scales of leptons is less than 1%.

The dominant sources of experimental systematic uncertainties are the jet energy scale, missing transverse energy, and pile-up. The jet energy scale uncertainty is determined using both collision data and test beam events; while less than 2.5% for high transverse momentum central jets, this uncertainty increases to 14% for jets with $p_T < 30$ GeV in the forward region [134]. The uncertainty due to pile-up is found to be less than 5% [133]. The uncertainty due to missing transverse energy includes all component object uncertainties in its calculation and is between 1 and 7% with an average of 2.6% [135]. These object systematics are propagated through the event selection and through all background estimation methods as part of the determination of systematic uncertainties on the entire analysis.

CHAPTER VII

The WW, Top, and W+jets Backgrounds

Several Standard Model processes can produce events that mimic the signature of $H \to WW \to \ell\nu\ell\nu$ decays. The WW continuum and $t\bar{t}$ decays can produce events with two oppositely charged leptons and missing transverse energy; this is the exact signature of the $H \to WW \to \ell\nu\ell\nu$ signal process and so these are called irreducible backgrounds. Additional backgrounds come from events such as W+jets or Z/γ^*+jets events where some part of the event has been incorrectly reconstructed, leptons in the case of W+jets and missing transverse energy in the case of Z/γ^*+jets.

Samples of Monte Carlo generated events of these processes are used for analysis and prediction of the contribution, but these samples cannot perfectly recreate the expected contamination. Theoretical uncertainties, for example, are introduced by the order to which the process was calculated in perturbation theory and by the choice of PDF sets used for Monte Carlo generation. The simulation and digitization processes do not perfectly mimic the behavior of particles in the detector due to imperfect knowledge of the detector's material distributions and other technical problems; this can contribute to discrepancies between Monte Carlo and data. Furthermore, it is not possible to produce or analyze samples of Monte Carlo events comparable to the number of data events available, which introduces additional complications. For rare backgrounds, it is possible that the signal region requirements remove all or close to

all the Monte Carlo events, leaving no events to calculate an expected background. Monte Carlo events are weighted by luminosity such that one generated event can equate to several expected events in a certain region; a cut that removes or includes one Monte Carlo event can then change the expected number of background events by several events. Thus Monte Carlo alone is not always a sufficient estimator of the background contamination.

In this analysis, background estimates from Monte Carlo are supplemented or replaced by data-driven estimates if possible; most of the methods used for this are discussed in this chapter. The WW continuum and top backgrounds supplement the Monte Carlo predictions using comparisons between data and Monte Carlo control regions; these are discussed in Sections 7.1 and 7.2. The W+jets background is estimated solely from data using a fake lepton method discussed in Section 7.3. Estimation techniques for the Z/γ^* background and their cross-checks will be covered in the following chapter.

7.1 WW Continuum Background

The WW continuum produces events with the closest resemblance to the signal Higgs boson decays. The primary differences in event kinematics comes from the defined mass of the Higgs boson and its spin-0 nature. To produce two spin-1 W bosons from a spin-0 Higgs boson, the two W bosons must have opposite spins. The neutrinos produced when these bosons decay can only exist in certain helicity states as discussed in Section 2.1, and this produces a small azimuthal angle between the two visible leptons and a small dilepton mass. Cuts on these variables are used to increase the purity of the signal region. As this electroweak process is well understood and has a fairly low production rate at the LHC, Monte Carlo simulation produces a reasonable estimate of the signal region contamination, and is supplemented by a control region of WW events defined in data.

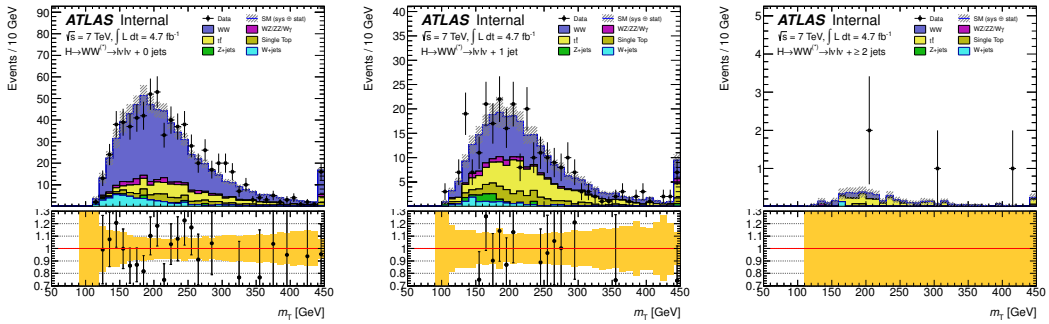

Figure 7.1: Transverse mass distributions in the low Higgs mass WW control regions for events with zero jets (left), one jet (center) and at least two jets (right). All lepton flavors are combined.

The control region is defined by the nominal cutflow selections as described in Chapter V without the cut on dilepton opening angle, $\Delta\phi$. The cut on dilepton mass is also altered. For the search for a low mass Higgs boson ($m_H < 200$ GeV), the signal region is defined by $m_{\ell\ell} < 50$ GeV and the WW control region is defined as $m_{\ell\ell} > 80$ GeV. In the same flavor channels, the Z-window veto makes this limit $m_{\ell\ell} > 106$ GeV. The ratio between the observed data events and expected WW events taken from Monte Carlo is used as a normalization factor to scale the WW Monte Carlo prediction in signal region cutflows and plots; the agreement between data and Monte Carlo is good and is shown for the control regions in Figure 7.1. In setting limits on the mass of the Higgs boson, the observed data in this region is used to constrain the WW background.

Contamination from other background processes in this control region is taken into account by subtracting the expected contamination from the observed number of events before the scale factors are calculated. The expectation is obtained using the same methods as those used for these backgrounds in the signal region. The top backgrounds are scaled by factors taken from the top control regions described in Section 7.2. W+jets contamination is estimated using the fake-factor method described in Section 7.3. For the ee and $\mu\mu$ channels, the Z/γ^* contamination is

Jet Multiplicity	ee	$e\mu$	$\mu\mu$	Inclusive
0 Jet	1.29±0.23	1.04±0.08	0.82±0.14	1.03±0.06
1 Jet	1.69 ±0.5	0.80±0.15	0.77±0.36	0.91±0.13

Table 7.1: Scale factors derived from the WW control regions. Statistical uncertainties only are shown.

estimated using the ABCD method described in Section 8.2.

Results from the control regions are shown in Table 7.1. Scale factors inclusive across all flavor channels are used for limit-setting and normalizing the WW background in final result plots; these agree with one within statistical uncertainties.

When studying heavier possible masses for the Higgs boson, the $m_{\ell\ell}$ range of the signal region increases to 150 GeV, but signal contamination in the range above this limit is still significant. For the intermediate range $200 < m_H < 300$ GeV, the region with $m_{\ell\ell} > 150$ GeV is considered in limit-setting but is not used to constrain the WW background. In the study of heavier masses, no WW control region is defined and the prediction of this background is taken solely from Monte Carlo simulation.

7.2 Top Background

Another important source of background arises from the leptonic decays of $t\bar{t}$ pairs, with the additional feature that such decays are expected to include two b-jets. This background is strongly reduced in the zero jet channel by the zero jet requirement, and estimation of the remaining contamination requires a specialized control region described in Section 7.2.1. In the one and two jet channels, a veto of events containing jets with b-tags suppresses this background, and a control region is defined by reversing this cut; the procedure is described in Section 7.2.2.

7.2.1 Top Zero Jet Background

The top backgrounds contribute in the zero jet channel if the jets produced in the decay are somehow lost, either by being low in transverse momentum, in the forward direction, or failing to be reconstructed. This is done by estimating the number of top candidates before jet multiplicity requirements and the efficiency for top events to pass the jet veto [136].

The efficiency for top events to pass the jet veto is estimated using a control sample of two leptons with missing transverse energy and at least one b-tagged jet. The selections used to define this sample are identical to those used in the nominal analysis described in Chapter V. The b-tagged jet with the highest b-tagging weight is taken as the tagging jet; all other jets in the event more than $\Delta R > 0.1$ away from the tagging jet are used as probing jets. The jet veto efficiency P_1^{Btag} is calculated as the number of events with no probing jets divided by the total number of events in the sample.

The efficiency for top events to be vetoed, or P_2^{est}, is then calculated as:

$$P_2^{est} = P_2^{MC} \frac{(P_1^{Btag,Data})^2}{(P_1^{Btag,MC})^2} \tag{7.1}$$

where P_2^{MC} is the efficiency for top events in Monte Carlo samples to have zero jets and $P_1^{Btag,MC}$ and $P_1^{Btag,Data}$ are the fractions of events in the control regions defined above with no probing jets for Monte Carlo and data events, respectively. The jet veto efficiency is squared to produce the event veto efficiency as the expectation is for top events to produce two b-jets and both must be lost for top events to contribute in the zero jet channel; this is described in more detail in References [137, 138].

The number of top events is estimated from data in a sample of events of two leptons with missing transverse energy and no jet requirements or topological criteria. The contributions from other backgrounds are subtracted using Monte Carlo

Jet Multiplicity	ee	$e\mu$	$\mu\mu$	Inclusive
1 Jet	1.19±0.17	1.03±0.07	1.01±0.12	1.03±0.07

Table 7.2: Scale factors derived from the top one jet control region. Statistical Uncertainties only are shown.

simulations that have been corrected with scale factors when available, such as for the WW continuum described in Section 7.1. This number is then scaled by the event veto efficiency P_2^{est} to find the predicted top contamination in the zero jet channel.

7.2.2 Top One and Two Jet Backgrounds

In the one and two jet channels, the top background is controlled primarily through the b-tagging veto on jets in the events. The remaining contribution is estimated from Monte Carlo normalized to the observed data in a top-dominated control region. In the one jet channel, this region is defined as events with two leptons, missing transverse energy, and exactly one jet which is b-tagged; the b-tagging criteria is the inverse of that used to reject events in the nominal selections. No requirements on dilepton opening angle or dilepton mass are applied. The expected contamination from non-top backgrounds are subtracted from the observed data. As there is no ABCD estimate of the Z/γ^*+jets background available in this region, the Monte Carlo prediction is scaled with a factor taken from the Z-window; derivation of this factor is discussed in detail in Section 8.1. The ratio between the corrected number of observed events and the top Monte Carlo prediction in this region is used as a normalization factor applied to the top Monte Carlo simulation in the final signal region; these are shown in Table 7.2.

In the two jet channel, the same procedure is applied. The control region is defined by selecting events with two leptons, missing transverse energy, and two or more jets, one of which must be b-tagged. Events containing additional central jets with $|\eta| < 3.2$ are vetoed, and no other requirements are imposed. The ratio between

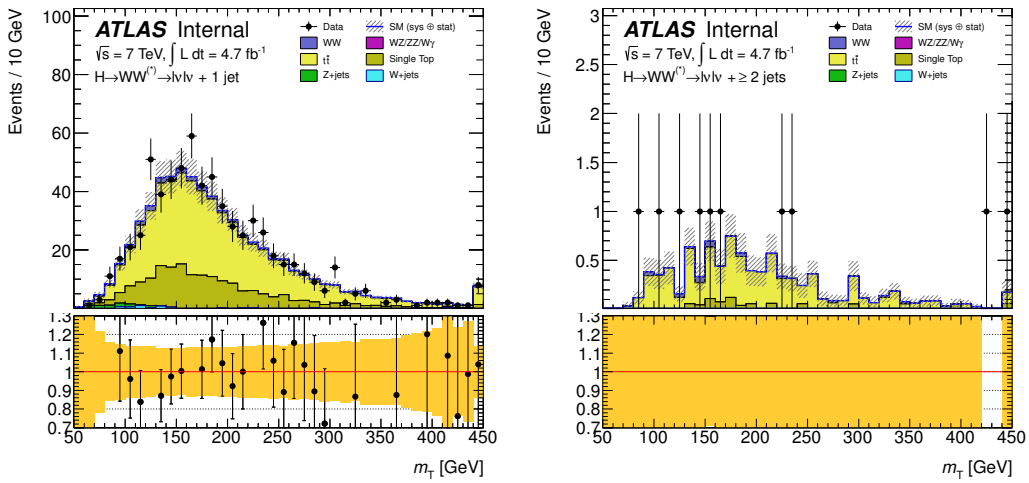

Figure 7.2: Distributions of transverse mass for the top one (left) and two inclusive (right) jet control regions.

data and Monte Carlo in this region is then taken as a scale factor for the final signal region. Figure 7.2 shows distributions of the transverse mass in the top one and two jet control regions.

7.3 W+jets Background

W+jets events constitute a background to the $H \rightarrow WW \rightarrow \ell\nu\ell\nu$ signal in the case where the W boson decays leptonically and a jet or other object in the event is misreconstructed as a lepton, producing an event with two leptons, one of them fake, and real missing transverse energy. The likelihood of this happening is small considering the stringent lepton identification requirements imposed, but must be considered because the production cross-section of jets at the LHC is large compared to the Higgs boson production cross-section. Due to the complexity of jet reconstruction and its similarity to electron reconstruction in relying on shape information within the calorimeter, Monte Carlo generators are unlikely to do an adequate job modeling this background and a data-driven method has been developed to replace it [136, 138].

Anti-identified Electrons	Anti-identified Muons
Same p_T, η ranges as identified electrons	Same p_T, η ranges as identified muons
$N_{hit}(\text{Pixel} + \text{SCT}) \geq 4$	Same ID track requirements as identified muons
$z_0 < 1$ mm	$z_0 < 1$ mm
Normalized calorimeter isolation < 0.3	Normalized calorimeter isolation < 0.3
Normalized track isolation < 0.13	
Identifed electrons removed	Identified muons removed

Table 7.3: Criteria for definition of anti-identied leptons used in fake lepton method estimation of W+jets background. Identified leptons are those selected by the nominal analysis. The isolation variables are defined in Section 5.2.

The method used to estimate the W+jets background uses a sample of events where one lepton has passed the criteria used in the nominal selection and described in Section 5.2 and the second lepton has failed the full set of criteria but passed a second, looser set of cuts. The first is referred to as the identified lepton and the second as the anti-identified lepton. The estimate is then computed as:

$$\begin{aligned} N^{ee}_{ID+fake} &= f_e \times N_{IDe+antiIDe} \\ N^{\mu\mu}_{ID+fake} &= f_\mu \times N_{ID\mu+antiID\mu} \\ N^{e\mu}_{ID+fake} &= f_e \times N_{ID\mu+antiIDe} + f_\mu \times N_{IDe+antiID\mu} \end{aligned} \tag{7.2}$$

where f_e,f_μ are the electron and muon fake factors defined in Equation 7.3. The criteria used to select anti-identified leptons are shown in Table 7.3. These mimic those used in the nominal selection but with the impact parameter significance and isolation cuts removed. The anti-identified leptons must also fail the full identification criteria.

The lepton fake factors are determined using a separate sample of dijet events selected using a photon trigger with an E_T threshold of 20 GeV and not electron or muon triggers like those used for the nominal selections; this imposes no identification criteria on leptons that could bias the sample. Contributions from electroweak

production of W and Z bosons are rejected by removing events with dilepton mass within the Z-window ($76 < m_{\ell\ell} < 106$ GeV) or transverse mass above 30 GeV. The remaining electroweak contribution is subtracted using Monte Carlo. The fake factors are then computed as:

$$f_\ell = \frac{N_{ID\ell}}{N_{antiID\ell}} \tag{7.3}$$

where the identification criteria are those described in Table 7.3. The procedure is then applied to both the final signal region as well as the WW control region for estimating the contamination from this background.

This method is validated using same sign events, as the production of fake leptons should be independent of the charge of the W boson while the same sign channels have significantly less contribution from the WW, top, and Z processes. Several potential sources of bias exist for this method and are evaluated to calculate its systematic uncertainty. Bias from the trigger selection is estimated using samples selected with two different photon triggers; the fractional difference in fake factors is taken as an uncertainty. Bias due to contamination of electroweak events is estimated by varying the cross-section used in the background subtraction by $\pm 20\%$ and contributes a 10% uncertainty to the total. The possibility of differences between the dijets sample used to estimate the fake factors and the W+jets sample where they are applied is evaluated by calculating fake factors in both samples in Monte Carlo; again the fractional difference is used as a systematic uncertainty and is found to be dependent on transverse momentum and between 20-50%. Dependence of data-taking period and therefore pile-up conditions of the fake factors was studied but was found to be negligible and does not contribute to the uncertainty. The estimation of the fake factors was also checked using different identification criteria and using samples of Z+jets events, and the resulting fake factors agreed with the original set within uncertainties. The successful completion of these cross-checks and calculations means

this method is ready for use in this analysis.

CHAPTER VIII

Z Background Estimation Methods

One of the reducible backgrounds to the $H \to WW \to \ell\nu\ell\nu$ analysis comes from events with two real opposite-sign leptons with fake missing transverse energy, the most common source of which is same flavor Z/γ^*+jets events. In early data-taking, this background was estimated using the scale factor method described in Section 8.1. This method calculates a correction factor for the mismodeling of the $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ cut efficiency using events in the Z-window; this factor can be applied to Monte Carlo to correct predictions of this background. As pile-up increased, Monte Carlo simulations were difficult to tune to precisely match data-taking conditions, and more data-driven estimates were explored. The ABCD method discussed in Section 8.2 is now the primary method used for estimating this background. The ratio of ratio method, described in Section 8.3, is also less dependent on Monte Carlo modeling and provides a cross-check to the ABCD method. The results of all methods are compared and discussed in Section 8.4.

8.1 Scale Factor method

8.1.1 Procedure and Results

In earlier versions of this analysis, such as Reference [136], the scale factor method has been used to estimate the contamination from the Z/γ^*+jets background in

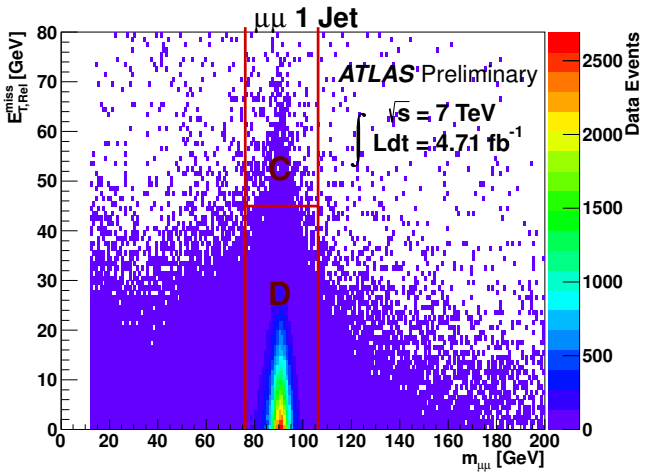

Figure 8.1: $m_{\ell\ell}$v$E^{\text{miss}}_{\text{T,Rel}}$ plane for the $\mu\mu$ 1 jet channel after the jet multiplicity cut; regions C and D are used for calculation of the scale factors.

the final signal region. In this method, a $E^{\text{miss}}_{\text{T,Rel}}$ mismodeling factor is derived by comparing same-flavor data and Monte Carlo events within the Z window rejected by the nominal analysis. Events in this region are classified into a high $E^{\text{miss}}_{\text{T,Rel}}$ region of events with $E^{\text{miss}}_{\text{T,Rel}} > 45$ GeV (labeled region C) or a low $E^{\text{miss}}_{\text{T,Rel}}$ region with all other events (called region D). These regions are shown in the $m_{\ell\ell}$-$E^{\text{miss}}_{\text{T,Rel}}$ plane in Figure 8.1. The cut value used was set to mirror the $E^{\text{miss}}_{\text{T,Rel}}$ threshold used to define the signal region in the nominal cutflow. The efficiency for this cut in the Z window is calculated in both data and Monte Carlo and the ratio of these efficiences is used as a multiplicative factor to correct the Monte Carlo prediciton of Z/γ^*+jets events in the final signal region (region A):

$$A_{cor} = A_{MC} \times \frac{C_{Data}}{C_{Data} + D_{Data}} \frac{C_{MC} + D_{MC}}{C_{MC}} \tag{8.1}$$

Contamination in the Z window from other backgrounds is estimated using the prediction from Monte Carlo simulation and this is subtracted from the observed number of events before the scale factor is calculated. Table 8.1 shows the results from the scale factor method compared to the uncorrected Monte Carlo prediction after each

Cut	Raw MC Estimate	Scale Factor	Scaled Estimate
ee 0 Jet			
0 Jet	25.67±3.80	$1.04\pm0.03^{+0.34}_{-0.38}$	$26.72\pm4.02^{+7.03}_{-5.34}$
$p_T^{\ell\ell} > 45$ GeV	3.57±1.22	$0.89\pm0.31^{+0.32}_{-0.38}$	$3.19\pm1.55^{+1.77}_{-1.23}$
$\Delta\phi_{\ell\ell} < 1.8$	3.57±1.22	$1.56\pm0.99^{+0.53}_{-0.32}$	$5.55\pm4.01^{+3.10}_{-0.62}$
ee 1 Jet			
1 Jet	15.85±2.45	$1.09\pm0.07^{+0.36}_{-0.22}$	$17.21\pm2.88^{+4.75}_{-6.81}$
b-tag veto 1jbtag	14.21±2.29	$1.10\pm0.07^{+0.36}_{-0.22}$	$15.62\pm2.71^{+3.66}_{-5.68}$
$p_T^{Tot} < 30$ GeV	3.02±0.99	$1.10\pm0.17^{+0.44}_{-0.12}$	$3.32\pm1.19^{+1.80}_{-1.31}$
$Z \to \tau\tau$ veto	2.43±0.89	$1.10\pm0.17^{+0.45}_{-0.12}$	$2.66\pm1.06^{+1.81}_{-0.99}$
$\Delta\phi_{\ell\ell} < 1.8$	2.20±0.86	$0.84\pm0.69^{+0.25}_{-0.23}$	$1.85\pm1.68^{+1.30}_{-1.02}$
$\mu\mu$ 0 Jet			
0 Jet	77.31±6.15	$1.07\pm0.02^{+0.35}_{-0.35}$	$82.85\pm6.77^{+12.92}_{-19.23}$
$p_T^{\ell\ell} > 45$ GeV	16.21±2.51	$1.17\pm0.17^{+0.28}_{-0.20}$	$19.00\pm4.08^{+4.13}_{-4.32}$
$\Delta\phi_{\ell\ell} < 1.8$	16.21±2.51	$1.62\pm0.42^{+0.64}_{-0.45}$	$26.25\pm7.93^{+7.98}_{-2.69}$
$\mu\mu$ 1 Jet			
1 Jet	45.53±4.11	$1.08\pm0.05^{+0.38}_{-0.20}$	$49.20\pm4.91^{+14.29}_{-10.49}$
b-tag veto	41.62±3.90	$1.08\pm0.05^{+0.40}_{-0.19}$	$44.88\pm4.63^{+14.30}_{-10.02}$
$p_T^{Tot} < 30$ GeV	11.14±2.16	$0.90\pm0.10^{+0.29}_{-0.11}$	$10.00\pm2.21^{+3.49}_{-1.06}$
$Z \to \tau\tau$ veto	10.55±2.12	$0.90\pm0.10^{+0.28}_{-0.09}$	$9.44\pm2.15^{+3.34}_{-1.04}$
$\Delta\phi_{\ell\ell} < 1.8$	6.25±1.42	$1.13\pm0.46^{+0.45}_{-0.26}$	$7.04\pm3.31^{+4.78}_{-3.02}$

Table 8.1: Scale factors and scale factor method estimates calculated for the Z/γ^*+jets background in the ee and $\mu\mu$ channels. Estimate directly from Monte Carlo simulation is also shown. Symmetric uncertainties are statistical; asymmetric uncertainties are calculated from systematic variation of object parameters.

cut in the ee and $\mu\mu$ channels. Within statistical uncertainties, the scale factors calculated at many steps with the cutflow agree with one, indicating that little rescaling is needed to correct the efficiency of this cut in simulation. Results calculated after the $\Delta\phi_{\ell\ell}$ requirements show the largest deviations. This can be attributed in part to the low efficiency of this cut in the Z-window and resulting low number of events available to make the calculation. The effect is also due to correlations between $\Delta\phi_{\ell\ell}$ and $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ as the phase space for events with collinear leptons that reconstruct to the Z mass is larger for those with jet activity (the likely source of the missing energy) than for those without. The estimates themselves agree with the unscaled predictions within uncertainties.

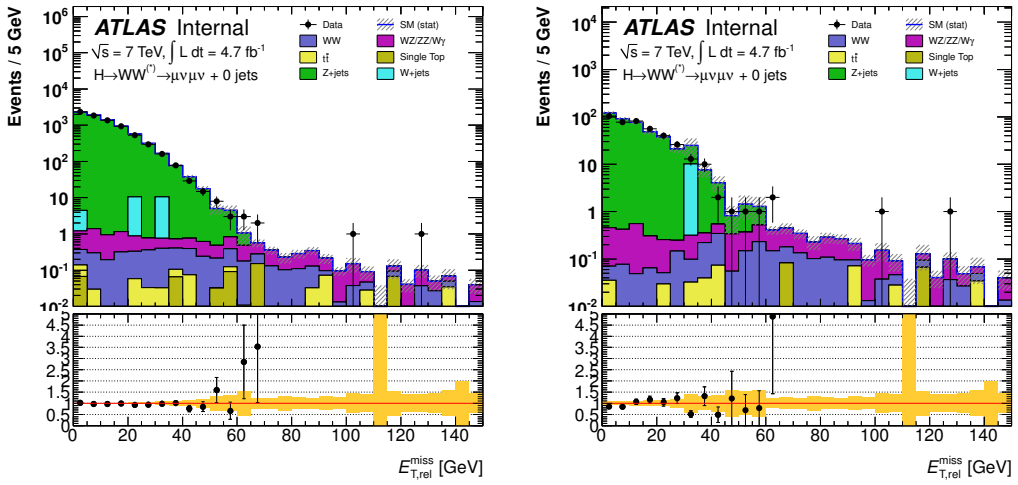

Figure 8.2: $E^{\text{miss}}_{\text{T,Rel}}$ distributions within the Z-window for the $\mu\mu$ zero jet channels after the zero jet requirement (left) and $p_T^{\ell\ell}$ cut (right).

8.1.2 Cross-checks and systematic uncertainties

The goal of the scale factor method is to correct the efficiency of the Z/γ^*+jets Monte Carlo for the $E^{\text{miss}}_{\text{T,Rel}}$ requirement in the signal region. The Z-window is used as a control region of events expected to have little real missing transverse energy to calculate the needed efficiencies; this region provides large statistics for making the calculation with a low percentage of other contaminating processes. Figure 8.2 shows distributions of $E^{\text{miss}}_{\text{T,Rel}}$ within the Z-window. These show agreement between data and Monte Carlo predictions within 15-20%, with excesses of data over Monte Carlo in the high $E^{\text{miss}}_{\text{T,Rel}}$ tail. Discrepancies are expected as modeling of large fake missing transverse energy in Monte Carlo simulations is difficult, as it relies on modeling failures of the detector. Poor normalization of the high $E^{\text{miss}}_{\text{T,Rel}}$ region is therefore expected.

The $E^{\text{miss}}_{\text{T,Rel}}$ threshold of 45 GeV between regions C and D corresponds to the cut imposed in defining the signal region, but this threshold also impacts the stability of this estimate. Figure 8.3 shows the effect of varying this threshold in both the Z-window and in the signal region between 40 and 50 GeV in the $\mu\mu$ zero jet channel

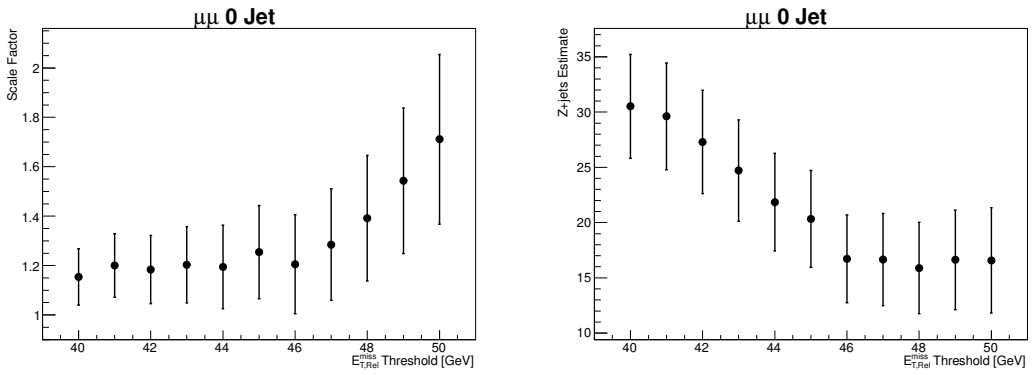

Figure 8.3: Calculated scale factors and estimates in the $\mu\mu$ zero jet channel after the $p_T^{\ell\ell}$ cut for different $E^{\text{miss}}_{\text{T,Rel}}$ thresholds. Scale factors are shown in the left plot and signal region estimates in the right plot.

after the $p_T^{\ell\ell}$ cut. The ee channel shows similar behavior. These plots show that the scale factors derived in the Z-window are not significantly affected by changing the $E^{\text{miss}}_{\text{T,Rel}}$ threshold; this indicates that while the normalization of the Monte Carlo might not match the data, the shape of this tail is well modeled within statistical uncertainties. The estimates themselves can decrease by up to 50%. This is due to the steep decrease in events of this background for higher values of $E^{\text{miss}}_{\text{T,Rel}}$, but the behavior is also affected by the relatively few Monte Carlo events dominating the estimate in the signal region, and puts this method at risk of statistical fluctuations in simulation.

No lower bound is imposed on $E^{\text{miss}}_{\text{T,Rel}}$ within the Z-peak used for calculation. Contributions from QCD can be significant at low missing energy and are not likely to be well modeled or addressed by background subtraction. The large cross-sections for such processes at the LHC means that such events could alter the results of these estimates. To check for this, the lower $E^{\text{miss}}_{\text{T,Rel}}$ bound on region D was varied between 0 and 30 GeV and the scale factors were recalculated. All variations of the resulting scale factors were well within the statistical uncertainties on the nominal estimate, indicating that the shape of the Monte Carlo distribution agrees well with data within

this region.

Systematic uncertainties are dominated by the object systematics described in Section 6.7; these are assessed by varying a parameter in the reconstruction, such as the resolution of jets, by one standard deviation and recalculating the scale factors and estimates with the altered objects. The primary contributions to these uncertainties come from the calorimeter cluster energy scale and jet energy scale. Table 8.2 shows the contributing systematic variations in the ee one jet channel after the p_T^{Tot} requirement. The statistical uncertainty on this scale factor is 15%, while the systematic uncertaintes are between 11% and 40%, and systematic uncertainties appear to be the dominant contribution. For the estimates themselves, the statistical uncertainty contributes 36% while the systematic uncertainties are 40-54%; hence systematic uncertainties dominate in this estimate. Similar results are reached for the other channels as well.

8.2 ABCD method

The ABCD method is a data-driven estimate of the Z/γ^*+jets contamination in the signal and WW control regions. It is carried out in the same flavor channels (ee and $\mu\mu$) for events with either exactly zero or exactly one jet. The procedure used for these estimates is covered in Section 8.2.1; verification of the underlying estimates and closure tests of the method in both data and Monte Carlo are discussed in Sections 8.2.2 and 8.2.3. An adaptation of the method designed to reduce remaining dependence on Monte Carlo simulation for signal region estimates is presented in Section 8.2.4. Systematic uncertainties on the estimates are discussed in Section 8.2.5. Uncertainties quoted in this section are statistical only unless explicitly stated otherwise.

Systematics: *ee* 1jpttot		
Variation	Scale Factor	Estimate
Nominal	1.10±0.17	3.32±1.19
AllClustersDown	1.06±0.16	2.62±1.04
AllClustersUp	0.93±0.13	2.62±0.98
ElecResolutionDown	1.06±0.16	3.20±1.15
ElecResolutionUp	1.10±0.17	3.33±1.20
ElecScaleDown	1.10±0.17	3.33±1.20
ElecScaleUp	1.05±0.16	3.18±1.14
IDLOW	1.10±0.17	3.33±1.20
IDUP	1.10±0.17	3.33±1.20
JERUp	0.79±0.11	3.44±1.17
JESDown	1.18±0.18	2.57±1.08
JESUp	0.85±0.12	4.96±1.65
MSLOW	1.10±0.17	3.33±1.20
MSUP	1.10±0.17	3.33±1.20
MuonScale	1.10±0.17	3.32±1.19
PileUpDown	1.03±0.16	2.51±1.00
PileUpUp	1.02±0.15	3.16±1.13
Upward uncertainty	40%	54%
Downward uncertainty	11%	40%

Table 8.2: Contributions to the systematic uncertainties of the scale factor method in the *ee* one jet channel after the p_T^{Tot} cut.

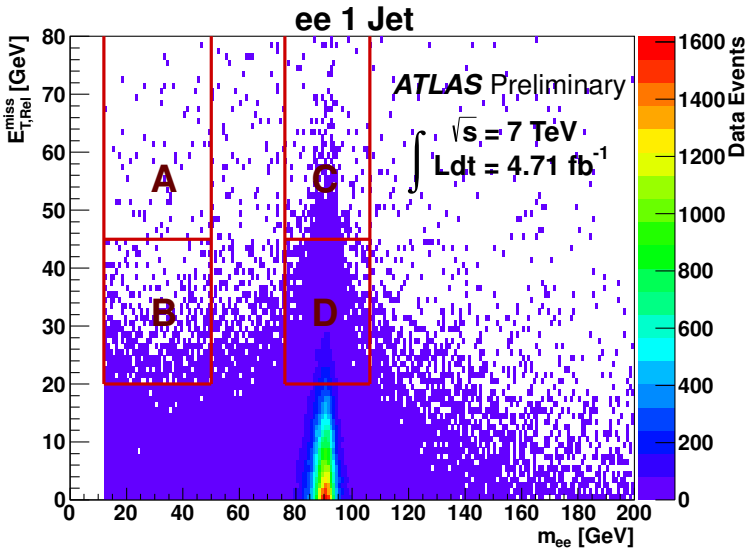

Figure 8.4: The regions used for the ABCD estimate of the Z/γ^*+jets contamination in the signal region. Boundaries used for low Higgs mass analysis are shown.

8.2.1 Procedure and Results

The ABCD method for signal region estimates uses four regions defined in the $m_{\ell\ell}$-$E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ plane as shown in Figure 8.4; the boundaries are:

- A: $12 < m_{\ell\ell} < 50$ GeV, $E_{\mathrm{T,Rel}}^{\mathrm{miss}} > 45$ GeV (signal region);
- B: $12 < m_{\ell\ell} < 50$ GeV, $20 < E_{\mathrm{T,Rel}}^{\mathrm{miss}} < 45$ GeV;
- C: $|m_{\ell\ell} - m_Z| < 15$ GeV, $E_{\mathrm{T,Rel}}^{\mathrm{miss}} > 45$ GeV;
- D: $|m_{\ell\ell} - m_Z| < 15$ GeV, $20 < E_{\mathrm{T,Rel}}^{\mathrm{miss}} < 45$ GeV;

These boundaries are set to match those used in the nominal event selection for searches for Higgs bosons with low mass and avoid overlap with signal and other control regions. Assuming that the ratios of Z/γ^*+jets events between these four regions are constant, the contamination of the signal region by this background can be estimated as:

$$A^{est}_{Z/DY} = B_{obs}\frac{C_{obs}}{D_{obs}} \tag{8.2}$$

Contamination from other backgrounds such as the WW continuum is taken into account by subtracting the expected number of such events from the number of observed events before calculation. The non-Z background estimates are taken from Monte Carlo simulation of these backgrounds. The B, C, and D regions are expected to be dominated by Z/γ^*+jets events with a low level of contamination from other processes. Monte Carlo simulation is only used for this background subtraction; simulation of the Z/γ^*+jets background is not used, making this estimate independent of any mismodeling of fake $E^{\text{miss}}_{\text{T,Rel}}$ in simulation of this background. The ABCD estimates throughout the cutflow are summarized in Table 8.3 for the ee and $\mu\mu$ channels. Comparison with unscaled Monte Carlo predictions for this background shows that the ABCD estimates are typically larger, but statistical uncertainties on both the Monte Carlo predictions and the data-driven estimates are large enough that these deviations are not significant.

The ABCD estimate can be calculated after each cut in the cutflow, but topological cuts with a dependence on either $m_{\ell\ell}$ or $E^{\text{miss}}_{\text{T,Rel}}$ will have different efficiencies in the B, C, and D regions and skew the resulting estimate. Cuts on $\Delta\phi_{\ell\ell}$ are particularly striking in this regard, as detailed in Table 8.4. This cut has a much higher efficiency in region B than in regions C and D, which increases the estimate relative to previous points in the cutflow and leaves very low statistics for calculation. To avoid this, ABCD estimates and statistical uncertainties computed after the $p_T^{\ell\ell}$ and p_T^{Tot} cuts are used for the predictions and to normalize shapes taken from Monte Carlo. This reintroduces the need for simulation of Z/γ^*+jets backgrounds to properly estimate the efficiency of the omitted cuts, the $\Delta\phi_{\ell\ell}$ requirement and $Z \to \tau\tau$ veto, before these estimates can be used for limit-setting.

	ABCD estimate	Est./MC ratio
ee 0 Jet		
0 Jet	24.88±1.06	0.97±0.15
$p_T^{\ell\ell} > 45$ GeV	3.78±1.31	1.06±0.52
$\Delta\phi_{\ell\ell} < 1.8$	6.40±3.91	1.80±1.25
ee 1 Jet		
1 Jet	19.99±1.35	1.26±0.21
b-tag veto	18.09±1.24	1.27±0.22
$p_T^{Tot} < 30$ GeV	3.35±0.49	1.11±0.40
$Z \to \tau\tau$ veto	3.27±0.48	1.35±0.53
$\Delta\phi_{\ell\ell} < 1.8$	2.17±1.73	0.99±0.87
$\mu\mu$ 0 Jet		
0 Jet	69.09±1.79	0.89±0.07
$p_T^{\ell\ell} > 45$ GeV	17.98±2.56	1.11±0.23
$\Delta\phi_{\ell\ell} < 1.8$	29.32±6.75	1.81±0.50
$\mu\mu$ 1 Jet		
1 Jet	54.06±2.37	1.19±0.12
b-tag veto	48.58±2.18	1.17±0.12
$p_T^{Tot} < 30$ GeV	9.45±0.95	0.85±0.18
$Z \to \tau\tau$ veto	9.48±0.95	0.90±0.20
$\Delta\phi_{\ell\ell} < 1.8$	8.69±3.21	1.39±0.60

Table 8.3: Estimates from the ABCD method for the low Higgs mass signal region in the ee and $\mu\mu$ channels. The ratio between the estimate and the raw Monte Carlo prediction are also shown.

Estimates of the Z/γ^*+jets contamination in the WW control regions is also necessary. This is done by defining two additional regions:

- E: $m_{\ell\ell} > 106$ GeV, $E_{\mathrm{T,Rel}}^{\mathrm{miss}} > 45$ GeV (WW contol region);
- F: $m_{\ell\ell} > 106$ GeV, $25 < E_{\mathrm{T,Rel}}^{\mathrm{miss}} < 45$ GeV;

The estimate is then calculated as:

$$E_{Z/DY}^{est} = F_{obs}\frac{C_{obs}}{D_{obs}} \tag{8.3}$$

While referred to as the CDEF method or estimate throughout this dissertation, this nomenclature refers only to the different regions used; the procedure for this estimate

Sample	A	B	C	D
Data after $p_T^{\ell\ell} > 45$ GeV	52.00±7.21	122.88±11.80	36.50±12.13	1186.95±34.88
Data after $\Delta\phi < 1.8$	52.00±7.21	122.98±11.80	13.03±7.80	250.21±16.05
Efficiency	1.00	1.00	0.36	0.21

Table 8.4: Observed data events (after background subtraction) after the $p_T^{\ell\ell}$ and $\Delta\phi_{\ell\ell}$ cuts in the ee zero jet channel in each ABCD region. The final line has the efficiency of the $\Delta\phi_{\ell\ell}$ cut relative to the previous stage. The slight increase in region B is due to contaminating events in Monte Carlo being removed while no observed events were cut.

is the same as that for the ABCD procedure. Background subtraction is performed using the same technique as for the signal region ABCD estimate. Cuts on $\Delta\phi_{\ell\ell}$ are not applied as these cuts are not used to define the WW control region. The resulting estimates and ratios to Monte Carlo predictions are summarized in Table 8.5. Similarly to the ABCD estimates, the CDEF estimate appears to be larger than the unscale Monte Carlo prediction but both carry large statistical uncertainties and deviations from one are therefore not significant.

For the analysis of the intermediate range of Higgs boson masses ($200 < m_H < 300$ GeV), the $m_{\ell\ell} < 50$ GeV requirement is changed to $m_{\ell\ell} < 150$ GeV. This requires adapting the regions used for these estimates. The C and D regions of the Z-peak are unchanged, but the upper bound on regions A and B is taken at $m_{\ell\ell} = 150$ GeV, with the range above that used for regions E and F. The $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ boundaries remain unchanged. The procedure for these estimates remains the same. Results are shown for the signal region in Table 8.6 and for the WW control region in Table 8.7. The signal region estimates for this intermediate mass range are very similar to the estimates made with the low Higgs mass boundaries. The WW control region estimates appear reasonable but are difficult to judge due to the low statistics available at the high $m_{\ell\ell}$ region.

	CDEF estimate	Est./MC ratio
ee 0 Jet		
0 Jet	69.71±2.04	1.02±0.11
$p_T^{\ell\ell} > 45$ GeV	1.48±0.58	0.85±0.69
ee 1 Jet		
1 Jet	15.49±1.13	1.14±0.26
b-tag veto	14.71±1.06	1.15±0.27
$p_T^{Tot} < 30$ GeV	3.03±0.44	1.83±1.10
$Z \to \tau\tau$ veto	2.92±0.43	1.76±1.06
$\mu\mu$ 0 Jet		
0 Jet	125.53±2.72	0.99±0.07
$p_T^{\ell\ell} > 45$ GeV	4.90±0.93	1.94±1.05
$\mu\mu$ 1 Jet		
1 Jet	32.98±1.64	1.43±0.24
b-tag veto	30.87±1.54	1.38±0.23
$p_T^{Tot} < 30$ GeV	5.33±0.56	1.15±0.43
$Z \to \tau\tau$ veto	5.08±0.54	1.24±0.50

Table 8.5: Estimates from the CDEF method for the low Higgs mass WW control region in the ee and $\mu\mu$ channels. The ratio between the estimate and the unscaled Monte Carlo prediction are also shown.

	ABCD estimate	Est./MC ratio
ee 0 Jet		
0 Jet	204.61±4.96	0.99±0.06
$p_T^{\ell\ell} > 45$ GeV	6.03±2.07	1.29±0.60
$\Delta\phi_{\ell\ell} < 1.8$	6.03±2.07	1.29±0.60
ee 1 Jet		
1 Jet	55.10±3.28	1.10±0.13
b-tag veto	51.70±3.07	1.13±0.14
$p_T^{Tot} < 30$ GeV	10.31±1.39	0.97±0.25
$Z \to \tau\tau$ veto	10.30±1.40	1.02±0.27
$\Delta\phi_{\ell\ell} < 1.8$	10.30±1.40	1.02±0.27
$\mu\mu$ 0 Jet		
0 Jet	441.51±7.57	0.98±0.04
$p_T^{\ell\ell} > 45$ GeV	27.92±3.88	1.22±0.24
$\Delta\phi_{\ell\ell} < 1.8$	27.92±3.88	1.22±0.24
$\mu\mu$ 1 Jet		
1 Jet	159.12±6.16	1.06±0.07
b-tag veto	147.84±5.79	1.04±0.07
$p_T^{Tot} < 30$ GeV	27.00±2.58	0.92±0.15
$Z \to \tau\tau$ veto	27.10±2.61	0.96±0.16
$\Delta\phi_{\ell\ell} < 1.8$	27.10±2.61	0.96±0.16

Table 8.6: Estimates from the ABCD method for the intermediate Higgs mass signal region in the ee and $\mu\mu$ 0 and 1 jet channels. The ratio between the estimate and the raw Monte Carlo prediction are also shown.

	CDEF estimate	Est./MC ratio
ee 0 Jet		
0 Jet	14.92±0.75	0.87±0.18
$p_T^{\ell\ell} > 45$ GeV	0.63±0.29	0.36±0.30
ee 1 Jet		
1 Jet	4.00±0.47	1.39±0.65
b-tag veto	3.86±0.44	1.49±0.75
$p_T^{Tot} < 30$ GeV	0.89±0.16	1.00±0.00
$Z \to \tau\tau$ veto	0.83±0.16	1.00±0.00
$\mu\mu$ 0 Jet		
0 Jet	23.03±0.93	1.18±0.22
$p_T^{\ell\ell} > 45$ GeV	1.06±0.42	2.13±2.30
$\mu\mu$ 1 Jet		
1 Jet	8.02±0.67	1.72±0.63
b-tag veto	7.58±0.61	1.63±0.59
$p_T^{Tot} < 30$ GeV	1.33±0.19	1.17±0.84
$Z \to \tau\tau$ veto	1.33±0.19	1.17±0.85

Table 8.7: Estimates from the CDEF method for the intermediate Higgs mass WW control region in the ee and $\mu\mu$ 0 and 1 jet channels. The ratio between the estimate and the raw Monte Carlo prediction are also shown.

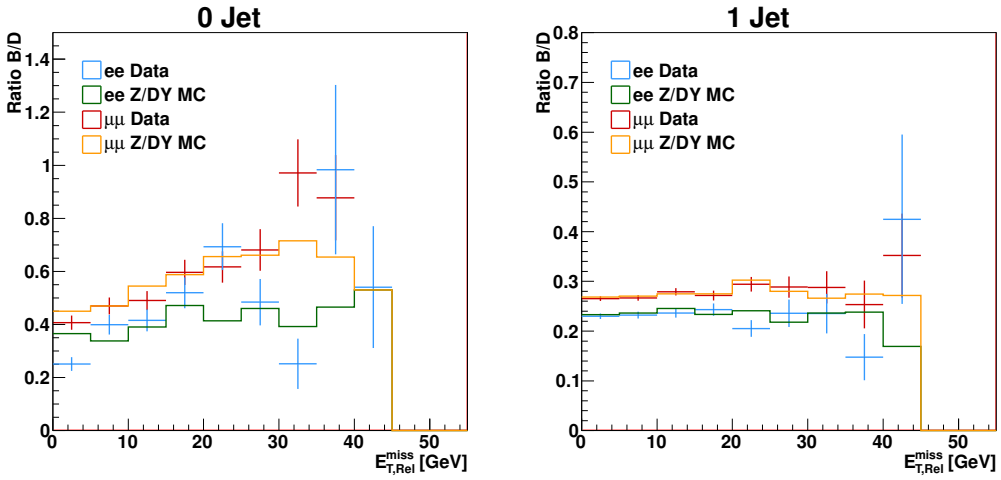

Figure 8.5: Ratios between regions B and D in data (points) and Z/γ^*+jets Monte Carlo (histograms) after the $p_T^{\ell\ell}$ cut in the zero jet channel (left) and after the p_T^{Tot} cut in the one jet channel (right).

8.2.2 Region Dependencies

The ABCD method depends on independence between $m_{\ell\ell}$ and $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ for the Z/γ^*+jets background. To check for dependence on $m_{\ell\ell}$ in the $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ spectra, the ratios between the $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ distributions in regions B and D is calculated using both data and Z/γ^*+jets Monte Carlo; these are presented in Figure 8.5.

The ratio between region B and D indicates the relative abundances of this background, and this information is not important for the ABCD method. The flatness of these distributions is a measure of how similar the $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ shape is between these two regions. The ee and $\mu\mu$ one jet channels both show constant ratios between 0.24 and 0.28 for the range $0 < E_{\mathrm{T,Rel}}^{\mathrm{miss}} < 40$ GeV, indicating good agreement in shape. The excess in the 40-45 GeV range carries large statistical errors and seems to be a fluctuation. The zero jet channels both show an increase in this ratio, indicating that the $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ distribution in region D falls more steeply than that of region B. The agreement between the observed ratio in data and that of the simulated Z/γ^* events is an additional measure of the agreement between data and Monte Carlo in these

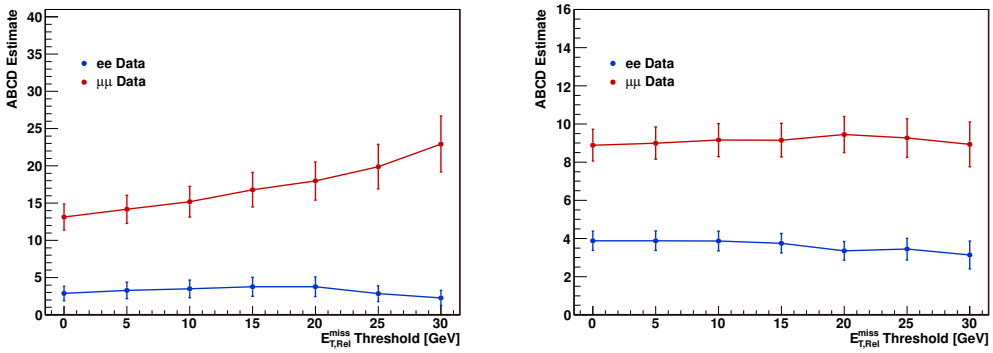

Figure 8.6: ABCD estimates in data for the $p_T^{\ell\ell}$ cuts (zero jet channels, left plot) and the p_T^{Tot} cuts (one jet channels, right plot) as a function of lower bound in $E^{\text{miss}}_{\text{T,Rel}}$.

regions. While the data and Monte Carlo agree well in the one jet channels, the ee zero jet channel has kinks between 20 and 35 GeV and the $\mu\mu$ zero jet channel has an excess between 30 and 40 GeV that are not matched by similar features in the Monte Carlo. As the ABCD estimate relies on the integral of these regions, shape discrepancies do not necessarily impact the estimate. However, simulation is used to assess the underlying assumptions of the ABCD method for evaluating systematic uncertainties, and this discrepancies are used to quantify the agreement between data and Monte Carlo in Section 8.2.5.

The upper $E^{\text{miss}}_{\text{T,Rel}}$ boundary of regions B and D is set by the threshold used to define the signal region. The lower bound was placed to avoid the low $E^{\text{miss}}_{\text{T,Rel}}$ region that would contain contamination from QCD backgrounds that could not be easily estimated or removed. To check the effect of this threshold, ABCD estimates for the signal region were performed using lower bounds between 0 and 35 GeV. The estimates after the $p_T^{\ell\ell}$ cuts in the zero jet channels and the p_T^{Tot} cuts in the one jet channels are shown in Figure 8.6.

The one jet channels and ee zero jet channel show no significant dependence on the lower bound on $E^{\text{miss}}_{\text{T,Rel}}$, but the $\mu\mu$ zero jet channel shows an upward trend. The

Cut	Alpgen		Pythia	
	A_{MC}	$\frac{B_{MC}C_{MC}}{D_{MC}}$	A_{MC}	$\frac{B_{MC}C_{MC}}{D_{MC}}$
ee 0 Jet	25.67±3.80	26.10±0.75	25.19±6.23	24.64±0.89
ee 0J $p_T^{\ell\ell} > 45$ GeV	3.57±1.22	3.58±0.52	4.78±2.24	4.60±0.59
ee 0J $\Delta\phi < 1.8$	3.57±1.22	3.38±1.00	4.78±2.24	7.17±1.28
ee 1 Jet	15.85±2.45	19.49±0.86	9.90±3.07	17.10±0.96
ee 1J btag veto	14.21±2.29	17.88±0.81	8.91±2.90	15.85±0.93
ee 1J p_T^{Tot} GeV	3.02±0.99	3.84±0.33	2.41±1.45	3.96±0.38
ee 1J $Z \to \tau\tau$ veto	2.43±0.89	3.71±0.32	2.41±1.45	3.61±0.34
ee 1J $\Delta\phi$	2.20±0.86	3.26±0.82	2.41±1.45	4.38±0.92
$\mu\mu$ 0 Jet	77.31±6.15	66.44±1.24	89.31±14.10	69.84±2.22
$\mu\mu$ 0J $p_T^{\ell\ell} > 45$ GeV	16.21±2.51	13.05±1.21	16.67±5.93	18.34±1.89
$\mu\mu$ 0J $\Delta\phi < 1.8$	16.21±2.51	15.04±2.56	16.67±5.93	25.47±3.54
$\mu\mu$ 1 Jet	45.53±4.11	50.66±1.50	50.81±10.21	43.73±2.10
$\mu\mu$ 1J btag veto	41.62±3.90	46.24±1.42	48.49±9.95	39.89±2.00
$\mu\mu$ 1J p_T^{Tot} GeV	11.14±2.16	11.30±0.66	12.63±5.04	8.59±0.73
$\mu\mu$ 1J $Z \to \tau\tau$ veto	10.55±2.12	11.28±0.66	12.63±5.04	8.78±0.75
$\mu\mu$ 1J $\Delta\phi$	6.25±1.42	8.26±1.65	4.59±3.24	8.65±1.73

Table 8.8: Closure tests comparing A to BC/D in the ee and $\mu\mu$ 0 and 1 jet channels. These are performed using the nominal ALPGEN and alternative PYTHIA Z/γ^*+jets simulation. Uncertainties are statistical only.

lower bound is therefore kept to avoid skewing the final estimate in this channel.

8.2.3 ABCD Closure Tests

As mentioned previously, the ABCD method relies on independence between $m_{\ell\ell}$ and $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ for the Z/γ^*+jets background. This assumption is checked by performing closure tests comparing BC/D to A in Z/γ^*+jets Monte Carlo. Both the nominal ALPGEN and an alternate set of PYTHIA Monte Carlo samples are considered. Results of these closure tests for the ee and $\mu\mu$ channels are in Table 8.8. Agreement between A and BC/D indicates the method is working well. These results show acceptable agreement within statistical uncertainties with both generators at the $p_T^{\ell\ell}$ and p_T^{Tot} requirements, which are the calculations used in limit-setting.

Additional closure tests were run using only data; these were performed by dividing regions B and D of the nominal low mass estimate each into two sub-regions.

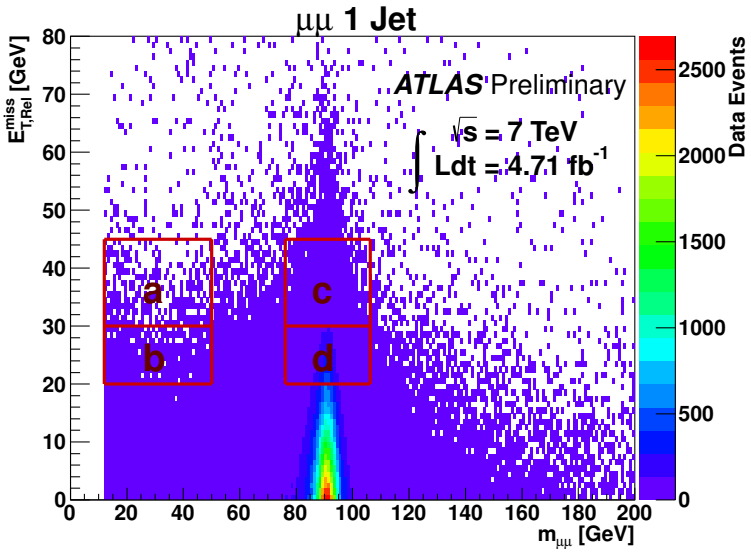

Figure 8.7: The regions used for data-driven closure tests shown in the $\mu\mu$ one jet channel.

These sub-regions are shown in Figure 8.7. Background subtraction on the observed data is done using Monte Carlo as for the nominal estimates.

The results of closure tests for these sub-regions in both data and Z/γ^*+jets Monte Carlo are shown in Table 8.9 for the ee and $\mu\mu$ channels. The first two columns display results for the Z/γ^*+jets Monte Carlo only, for comparison to the data results. These show good agreement throughout much of the cutflow, except for after the $\Delta\phi_{\ell\ell}$ cut; this is due the $m_{\ell\ell}$-dependence of this requirement as discussed in Section 8.2.1.

Closure of this comparison using data is shown in the third and fourth columns; the closure in these channels shows discrepancies after several of the later cuts. This corresponds to the kinks seen in the ratio between $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ distributions in Regions B and D described in Section 8.2.2 and highlights the sensitivity of these estimates to the upper bound on $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$.

Cut	MC a	MC bc/d	Data a	Data bc/d
ee 0 Jet				
0 Jet	239.22±11.46	228.06±6.33	211.96±15.97	212.48±9.15
$p_T^{\ell\ell} > 45$ GeV	27.99±3.18	27.30±2.38	19.57±5.64	36.96±4.47
$\Delta\phi_{\ell\ell} < 1.8$	27.42±3.13	25.80±3.49	19.67±5.64	46.47±7.99
ee 1 Jet				
1 Jet	153.76±7.35	166.56±4.85	153.52±14.14	153.87±7.48
b-tag veto	139.36±7.02	151.45±4.60	127.84±12.72	139.62±7.10
$p_T^{Tot} < 30$ GeV	56.86±4.46	59.48±2.45	45.11±8.62	48.89±3.57
$Z \to \tau\tau$ veto	50.97±4.25	57.79±2.44	41.19±8.39	48.27±3.61
$\Delta\phi_{\ell\ell} < 1.8$	41.97±3.68	52.28±4.46	28.99±7.56	45.86±5.92
$\mu\mu$ 0 Jet				
0 Jet	631.84±18.09	552.12±9.84	606.64±25.79	535.34±13.35
$p_T^{\ell\ell} > 45$ GeV	82.59±5.47	80.64±4.66	131.32±12.23	90.69±7.55
$\Delta\phi_{\ell\ell} < 1.8$	82.59±5.47	89.10±7.88	131.37±12.23	114.69±12.82
$\mu\mu$ 1 Jet				
1 Jet	367.11±11.31	392.82±7.85	373.53±20.18	386.67±12.46
b-tag veto	333.75±10.80	358.28±7.46	337.00±18.83	346.57±11.72
$p_T^{Tot} < 30$ GeV	139.14±7.13	148.82±4.16	131.03±11.86	140.24±6.39
$Z \to \tau\tau$ veto	135.40±7.06	146.60±4.17	128.19±11.73	138.76±6.44
$\Delta\phi_{\ell\ell} < 1.8$	111.87±6.07	133.48±8.20	101.14±10.47	118.07±10.20

Table 8.9: Results of the data-driven closure tests from data and Z/γ^*+jets Monte Carlo for the ee and $\mu\mu$ channels using the abcd regions as shown in Figure 8.7. Observed data events are after subtraction of estimated contaminating events.

8.2.4 Method Adaptations

As was noted in Section 8.2.1, ABCD and CDEF estimates after the $p_T^{\ell\ell}$ and p_T^{Tot} cuts are used for limit-setting as the later topological cuts, notably the $\Delta\phi_{\ell\ell}$ cut, have $m_{\ell\ell}$-dependent efficiencies. The efficiencies for the omitted cuts are then taken from Z/γ^*+jets Monte Carlo. The method can be adapted to avoid this by taking the observed number of events in region B after the final topological cut but using the event counts in regions C and D at an earlier stage. The estimate can then be written as:

$$A'^{est}_{Z/DY} = B'_{obs}\frac{C_{obs}}{D_{obs}} \tag{8.4}$$

Z/DY A after $\Delta\phi$	A'B'CD Closure Test	A'B'CD Estimate	A'B'CD/MC ratio
ee 0 Jet			
3.57 $\pm$1.22	3.56 $\pm$ 0.52	3.78 $\pm$ 1.31	1.06 $\pm$ 0.52
ee 1 Jet			
2.20 $\pm$0.86	3.11 $\pm$ 0.27	2.73 $\pm$ 0.40	1.24 $\pm$ 0.52
$\mu\mu$ 0 Jet			
16.21 $\pm$2.51	13.03 $\pm$ 1.20	17.93 $\pm$ 2.55	1.11 $\pm$ 0.23
$\mu\mu$ 1 Jet			
6.25 $\pm$1.42	9.22 $\pm$ 0.54	7.70 $\pm$ 0.78	1.23 $\pm$ 0.31

Table 8.10: Results and closure tests for the alternate A'B'CD method in the ee and $\mu\mu$ channels. The left column has the unscaled Monte Carlo prediction in the signal region for the Z/γ^* process, while the second column contains the results of this procedure when performed on the Monte Carlo. The third column has the results of this estimate in data, and the fourth shows the ratio between the data prediction and the unscaled Monte Carlo.

Here, the primed quantities are calculated at a different point in the cutflow from the unprimed numbers. This procedure is equivalent to multiplying the nominal estimate by a cut efficiency taken from region B, or to using B as a control region corrected with a $E^{\text{miss}}_{\text{T,Rel}}$-mismodeling factor taken from the Z-window. This also requires looser assumptions on the ratios between regions. The results from this calculation where region B is taken after the $\Delta\phi_{\ell\ell}$ cut in all channels but counts in C and D are taken from after the $p_T^{\ell\ell}$ and p_T^{Tot} cuts as before are shown in Table 8.10.

Closure tests of this method performed with Z/γ^* Monte Carlo show good agreement between the direct prediction and the estimate results within statistical uncertainties. The results of this A'B'CD estimate after the $\Delta\phi_{\ell\ell}$ cut in data in the zero jet channels are almost unchanged from the ABCD results after the $p_T^{\ell\ell}$ cut. Only one additional cut is applied to region B' compared to region B; this is the $\Delta\phi_{\ell\ell}$ cut, and as has been previously noted it is a very high efficiency in the low $m_{\ell\ell}$ regions. The zero jet A'B'CD results are approximately 99% of the ABCD ones, a change too small to be significant. The one jet results for the A'B'CD method show an increase in the ratio between estimate and Monte Carlo when compared to the results of the ABCD

after the p_T^{Tot} cut, but agree better with one than the ABCD results after the $\Delta\phi_{\ell\ell}$ requirement. As the results of the A'B'CD method are completely independent of the Z/γ^*+jets Monte Carlo simulation, these are the inputs used for limit-setting in this dissertation. The A'B'CD method is not used for the WW control region estimates, as the $\Delta\phi_{\ell\ell}$ cut is not needed for those regions.

8.2.5 ABCD Systematic Uncertainties

The main source of systematic uncertainties for this method comes from violation of the assumption of constant ratios between regions. This is assessed using the results of the closure tests in Z/γ^*+jets Monte Carlo. When low available Monte Carlo statistics could produce an artificially good closure, the larger of the non-closure and its uncertainty is used for calculation. Closure tests in the ABCD, CDEF, and ABEF regions are used, with each being treated as a linear combination of uncertainties on the component A and B, C and D, or E and F regions as:

$$\begin{aligned} \Delta_{ABCD} =& \sigma_{AB} + \sigma_{CD} \\ \Delta_{CDEF} =& \sigma_{CD} + \sigma_{EF} \\ \Delta_{ABEF} =& \sigma_{AB} + \sigma_{EF} \end{aligned} \tag{8.5}$$

Using this description of the uncertainties preserves information on the direction of any non-closure. This system of equations is solved to calculate region specific uncertainties. This is necessary to handle correlations between the signal region and WW control region systematic uncertainties.

Closure tests in Monte Carlo can assess the uncertainty of this method only as well as the Monte Carlo accurately reflects the behavior seen in data. Discrepancies have been observed in the regions used in the ABCD method and are discussed in Section 8.2.2. These discrepancies are accounted for by adding an additional 15% un-

Systematics: *ee* 0jptll	
Nominal	3.78±1.31
AllClustersDown	3.75±1.31
AllClustersUp	3.84±1.31
ElecResolutionDown	3.66±1.34
ElecResolutionUp	3.82±1.31
ElecScaleDown	3.86±1.32
ElecScaleUp	3.54±1.33
IDLOW	3.84±1.31
IDUP	3.84±1.31
JERUp	3.84±1.32
JESDown	3.34±1.29
JESUp	4.35±1.34
MSLOW	3.84±1.31
MSUP	3.84±1.31
MuonScale	3.78±1.31
PileUpDown	3.77±1.31
PileUpUp	3.81±1.31
Upward Error = 17%	
Downward Error = 12%	

Table 8.11: ABCD estimate results in the *ee* 0 jet channel after the $p_T^{\ell\ell}$ cut for each of the object systematic variations. The combined systematic uncertainties for this estimate are shown at the bottom.

certainty to the AB and EF region uncertainties in the zero jet channel; no additional uncertainty is included in the C and D uncertainty as this is taken as the standard of comparison. As no significant discrepancies were observed between the data and Monte Carlo in the one jet channels, no additional uncertainty is included there. This prescription is used for both the ABCD and the A'B'CD methods.

As the estimate is primarily data-driven, object systematic uncertainties on simulation as discussed in Section 6.7 enter only through the background subtraction. The effect of this on the resulting estimates is small compared to uncertainties from the closure tests. The largest effect is in the *ee* zero jet channel, as region C has the lowest data statistics which makes this estimate most sensitive to variations in the background subtraction; the results for this estimate are shown in Table 8.11. For

this calculation, the statistical uncertainty is 37% while the uncertainty due to the object variations is 12-17%; thus these variations are not the dominant contribution to this uncertainty. They are incorporated into region systematic uncertainties. The statistical and systematic uncertainties calculated for the low mass selections according to region are shown in Table 8.12 after the jet multiplicity cuts, the $p_T^{\ell\ell}$ cuts, and the p_T^{Tot} cuts. Statistical uncertainties for limit-setting are calculated after the $p_T^{\ell\ell}$ and p_T^{Tot} requirements to match the estimates used for calculating limits. Systematic uncertainties for the zero jet channels are used after the $p_T^{\ell\ell}$ requirements, but those for the one jet channels are calculated after the jet multiplcity cut to ensure adequate statistics to assess the non-closure of the method.

Other possible sources of uncertainty are that of fluctuations in the backgrounds and signal contamination, particularly in the low $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$, low $m_{\ell\ell}$ regions. Uncertainties on the level of background contamination are dependent on interdependent uncertainties on the calculation of the appropriate cross-section; for example, many calculations are inclusive in jet multiplicity while these estimate methods are applied in exclusive jet bins. A proper handling of this requires varying all of the component uncertainties and fitting to the observed data in the B, C, and D regions, and while beyond the scope of this analysis, would be a beneficial extension of the estimate methods.

8.3 Ratio of Ratio Method

8.3.1 Procedure, Results, and Systematic Uncertainties

The ratio of ratio method uses the same regions as the ABCD method, but asserts that the ratio of ratio between regions is equal between data and Monte Carlo. The estimated contamination in the signal region can then be calculated as:

Cut	σ_{AB}^{stat}	σ_{CD}^{stat}	σ_{EF}^{stat}	σ_{AB}^{sys}	σ_{CD}^{sys}	σ_{EF}^{sys}
ee 0 Jet	0.04	0.02	0.02	0.17	-0.07	0.18
ee 0J $p_T^{\ell\ell} > 45$ GeV	0.10	0.33	0.20	0.44	0.67	0.50
ee 1 Jet	0.04	0.05	0.05	0.25	-0.01	0.15
ee 1J $p_T^{Tot} < 30$ GeV	0.06	0.13	0.07	0.37	0.14	0.51
$\mu\mu$ 0 Jet	0.02	0.02	0.02	0.18	0.06	0.15
$\mu\mu$ 0J $p_T^{\ell\ell} > 45$ GeV	0.06	0.13	0.14	0.48	0.22	0.55
$\mu\mu$ 1 Jet	0.03	0.04	0.03	0.16	0.04	0.21
$\mu\mu$ 1J $p_T^{Tot} < 30$ GeV	0.04	0.09	0.05	0.22	0.14	0.46

Table 8.12: Region statistical (left three columns) and systematic (right three columns) uncertainties for the ABCD and CDEF methods. These are calculated from closure test results, agreement between data and Monte Carlo in regions B and D, and object systematic uncertainties.

$$A^{est} = A_{MC}\frac{D_{MC}}{B_{MC}C_{MC}}\frac{C_{obs}B_{obs}}{D_{obs}} \tag{8.6}$$

This automatically corrects for any discrepancy in the ratios between regions, but could potentially introduce bias due to mismodeling in the Monte Carlo simulation. It is also sensitive to low statistics in either data or Monte Carlo control regions. For this reason the ratio of ratio method is kept as a cross-check. The results of the estimation method are summarized for the ee and $\mu\mu$ channels in Table 8.13. The results from the unscaled Monte Carlo Z/γ^*+jets prediction is shown for comparison.

The results from these estimates are in agreement with the Monte Carlo prediction within statistical uncertainties. The zero jet estimate results tend to be similar to the unscaled Monte Carlo, while the results of the ratio of ratio method in the one jet channels tend to be larger than the prediction from simulation. One observation about this method is that it tends to carry large statistical uncertainties, particularly at later stages in the cutflow. This is due to the estimate relying on three regions in data and four in Monte Carlo; while this gives the estimate a sensitivity, low numbers of events in any one of these regions will propagate into a correspondingly large statistical uncertainty on the final estimate.

Cut	Raw MC Estimate	Ratio of Ratio Estimate
ee 0j	25.67±3.80	$24.46\pm3.83^{+2.84}_{-1.10}$
ee 0j $p_T^{\ell\ell} > 45$ GeV	3.57±1.22	$3.76\pm1.91^{+1.21}_{-1.78}$
ee 0j $\Delta\phi_{\ell\ell} < 1.8$	3.57±1.22	$6.75\pm5.12^{+2.35}_{-1.91}$
ee 1j	15.85±2.45	$16.25\pm2.83^{+2.49}_{-4.10}$
ee 1j b-tag veto	14.21±2.29	$14.38\pm2.60^{+1.93}_{-3.18}$
ee 1j $p_T^{Tot} < 30$ GeV	3.02±0.99	$2.64\pm0.97^{+1.48}_{-0.98}$
ee 1j $Z \to \tau\tau$ veto	2.43±0.89	$2.14\pm0.87^{+1.52}_{-0.73}$
ee 1j $\Delta\phi_{\ell\ell} < 1.8$	2.20±0.86	$1.47\pm1.35^{+1.25}_{-0.79}$
$\mu\mu$ 0j	77.31±6.15	$80.40\pm6.89^{+5.31}_{-7.39}$
$\mu\mu$ 0j $p_T^{\ell\ell} > 45$ GeV	16.21±2.51	$22.32\pm5.13^{+4.28}_{-0.90}$
$\mu\mu$ 0j $\Delta\phi_{\ell\ell} < 1.8$	16.21±2.51	$31.60\pm10.28^{+11.38}_{-5.13}$
$\mu\mu$ 1j	45.53±4.11	$48.58\pm5.08^{+11.44}_{-4.43}$
$\mu\mu$ 1j b-tag veto	41.62±3.90	$43.72\pm4.74^{+11.40}_{-4.08}$
$\mu\mu$ 1j $p_T^{Tot} < 30$ GeV	11.14±2.16	$9.32\pm2.10^{+3.09}_{-0.63}$
$\mu\mu$ 1j $Z \to \tau\tau$ veto	10.55±2.12	$8.86\pm2.06^{+2.91}_{-0.63}$
$\mu\mu$ 1j $\Delta\phi_{\ell\ell} > 1.8$	6.25±1.42	$6.57\pm3.14^{+5.37}_{-2.53}$

Table 8.13: Results from the ratio of ratio estimates of Z/γ^*+jets background for the ee and $\mu\mu$ zero and one jet channels. Statistical and asymmetric systematic uncertainties are shown on the estimates, and the unscaled Monte Carlo prediction is shown for comparison.

Systematics: $\mu\mu$ 0jptll	
Nominal	22.32±5.13
AllClustersDown	22.81±5.47
AllClustersUp	20.01±4.46
ElecResolutionDown	22.42±5.15
ElecResolutionUp	22.42±5.15
ElecScaleDown	22.42±5.15
ElecScaleUp	22.42±5.15
IDLOW	22.62±5.18
IDUP	22.84±5.23
JERUp	20.75±4.62
JESDown	21.76±4.67
JESUp	25.25±6.39
MSLOW	22.47±5.14
MSUP	22.81±5.22
MuonScale	22.22±5.09
PileUpDown	21.97±5.21
PileUpUp	21.14±4.74
Upward uncertainty = 0.19 %	
Downward uncertainty = 0.04 %	

Table 8.14: The contributing systematic uncertainties for the ration of ratio method estimates after the $p_T^{\ell\ell}$ cut in the $\mu\mu$ zero jet channel.

The systematic uncertainties on this method are driven by the object systematics described in Section 6.7. Their contributions are evaluated by varying the parameters by one standard deviation and recalculating the estimate. The largest contributions are from the jet energy scale variations. An example table showing all the contributions considered in the systematic uncertainty calculation for after the $p_T^{\ell\ell}$ cut in the $\mu\mu$ zero jet channel is shown in Table 8.14. Due to the reasons discussed previously, the statistical uncertainties are larger than the systematic ones and dominate for this method.

8.4 Comparison of Estimation Methods

Full comparisons of all estimation methods are shown in Table 8.15 for the ee channels and Table 8.16 for the $\mu\mu$ channels. These show statistical (listed first)

Cut	Raw MC Est.	SF Est.	ABCD Est.	Ratio Est
ee 0 Jet	$25.67 \pm 3.80^{+17.79}_{-11.13}$	$26.72 \pm 4.02^{+7.03}_{-5.34}$	$24.88 \pm 1.06^{+0.56}_{-0.56}$	$24.46 \pm 3.83^{+2.84}_{-1.10}$
$p_T^{\ell\ell} > 45$ GeV	$3.57 \pm 1.22^{+2.56}_{-0.56}$	$3.19 \pm 1.55^{+1.77}_{-1.23}$	$3.78 \pm 1.31^{+0.64}_{-0.45}$	$3.76 \pm 1.91^{+1.21}_{-1.78}$
$\Delta\phi_{\ell\ell} < 1.8$	$3.57 \pm 1.22^{+2.56}_{-0.56}$	$5.55 \pm 4.01^{+3.10}_{-0.62}$	$6.40 \pm 3.91^{+3.45}_{-3.24}$	$6.75 \pm 5.12^{+2.35}_{-1.91}$
			$(3.78 \pm 1.31^{+0.64}_{-0.45})$	
ee 1 Jet	$15.85 \pm 2.45^{+10.82}_{-8.18}$	$17.21 \pm 2.88^{+4.75}_{-6.81}$	$19.99 \pm 1.35^{+2.63}_{-2.61}$	$16.25 \pm 2.83^{+2.49}_{-4.10}$
b-tag veto	$14.21 \pm 2.29^{+9.50}_{-6.76}$	$15.62 \pm 2.71^{+3.66}_{-5.68}$	$18.09 \pm 1.24^{+2.53}_{-2.53}$	$14.38 \pm 2.60^{+1.93}_{-3.18}$
$p_T^{Tot} < 30$ GeV	$3.02 \pm 0.99^{+3.12}_{-1.17}$	$3.32 \pm 1.19^{+1.80}_{-1.31}$	$3.35 \pm 0.49^{+0.72}_{-0.72}$	$2.64 \pm 0.97^{+1.48}_{-0.98}$
$Z \to \tau\tau$ veto	$2.43 \pm 0.89^{+3.05}_{-0.88}$	$2.66 \pm 1.06^{+1.81}_{-0.99}$	$3.27 \pm 0.48^{+0.87}_{-0.86}$	$2.14 \pm 0.87^{+1.52}_{-0.73}$
$\Delta\phi_{\ell\ell} < 1.8$	$2.20 \pm 0.86^{+0.92}_{-0.87}$	$1.85 \pm 1.68^{+1.30}_{-1.02}$	$2.17 \pm 1.73^{+0.86}_{-0.77}$	$1.47 \pm 1.35^{+1.25}_{-0.79}$
			$(2.73 \pm 0.40^{+0.87}_{-0.86})$	

Table 8.15: Comparison of results from all methods of estimating Z/γ^*+jets contamination throughout the cutflow in the ee channels. The uncertainties are given as (stat.)$\pm$(sys). The results in parentheses are derived from the A'B'CD method.

and asymmetrical systematic (listed second) uncertainties for predictions from the unscaled Monte Carlo simulations, the scale factor method described in Section 8.1, the ABCD method described in Section 8.2, and the ratio of ratio method described in Section 8.3. For the unscaled, scaled, and ratio of ratio methods, systematic uncertainties are calculated solely from variations of the object systematics described in Section 6.7 in Monte Carlo simulation of all processes; for the ABCD method, systematic uncertainties are calculated from the non-closure of Monte Carlo closure tests added in quadrature with variations of the object systematic uncertainies. The zero jet channels ABCD results also incorporate additional uncertainties derived from observed discrepancies in the modeling of Monte Carlo relative to data described in Section 8.2.5.

These estimates show good agreement amongst the methods in the zero jet channels before the $\Delta\phi_{\ell\ell}$ requirement; as previously discussed, discrepancies after this cut are due to the strong dependence on $m_{\ell\ell}$ of this requirement, driven by the kinematics of requiring leptons close in ϕ but with a dilepton mass close to the Z-peak. In the one jet channels, agreement is good for all the methods throughout the cutflow. The

Cut	Raw MC Est.	SF Est.	ABCD Est.	Ratio Est
$\mu\mu$ 0 Jet	$77.31\pm6.15^{+41.25}_{-32.67}$	$82.85\pm6.77^{+12.92}_{-19.23}$	$69.09\pm1.79^{+11.05}_{-11.05}$	$80.40\pm6.89^{+5.31}_{-7.39}$
$p_T^{\ell\ell} > 45$ GeV	$16.21\pm2.51^{+5.07}_{-5.61}$	$19.00\pm4.08^{+4.13}_{-4.32}$	$17.98\pm2.56^{+4.55}_{-4.45}$	$22.32\pm5.13^{+4.28}_{-0.90}$
$\Delta\phi_{\ell\ell} < 1.8$	$16.21\pm2.51^{+5.07}_{-5.61}$	$26.25\pm7.93^{+7.98}_{-2.69}$	$29.32\pm6.75^{+7.62}_{-7.52}$	$31.60\pm10.28^{+11.38}_{-5.13}$
			$(17.93\pm2.55^{+4.55}_{-4.45})$	
$\mu\mu$ 1 Jet	$45.53\pm4.11^{+24.33}_{-15.95}$	$49.20\pm4.91^{+14.29}_{-10.49}$	$54.06\pm2.37^{+4.90}_{-4.90}$	$48.58\pm5.08^{+11.44}_{-4.43}$
b-tag veto	$41.62\pm3.90^{+23.16}_{-14.71}$	$44.88\pm4.63^{+14.30}_{-10.02}$	$48.58\pm2.18^{+4.37}_{-4.37}$	$43.72\pm4.74^{+11.40}_{-4.08}$
$p_T^{Tot} < 30$ GeV	$11.14\pm2.16^{+2.58}_{-1.69}$	$10.00\pm2.21^{+3.49}_{-1.06}$	$9.45\pm0.95^{+0.30}_{-0.30}$	$9.32\pm2.10^{+3.09}_{-0.63}$
$Z \to \tau\tau$ veto	$10.55\pm2.12^{+2.47}_{-1.44}$	$9.44\pm2.15^{+3.34}_{-1.04}$	$9.48\pm0.95^{+0.72}_{-0.72}$	$8.86\pm2.06^{+2.91}_{-0.63}$
$\Delta\phi_{\ell\ell} < 1.8$	$6.25\pm1.42^{+1.85}_{-1.71}$	$7.04\pm3.31^{+4.78}_{-3.02}$	$8.69\pm3.21^{+2.29}_{-2.26}$	$6.57\pm3.14^{+5.37}_{-2.53}$
			$(7.70\pm0.78^{+0.72}_{-0.72})$	

Table 8.16: Comparison of results from all methods of estimating Z/γ^*+jets contamination throughout the cutflow in the $\mu\mu$ channels. The uncertainties are given as (stat.)$\pm$(sys). The results in parentheses are derived from the A'B'CD method.

estimates involving Monte Carlo, however, have fewer available events to analyze and therefore have correspondingly larger statistical uncertainties.

The unscaled Monte Carlo for this background carries comparable statistical uncertainties to the other methods but has markedly larger systematic uncertainties. As these predictions do not rely directly on data, any mismodeling or uncertainty in sample generation or object reconstruction is uncorrected in these samples and this is reflected in the comparatively larger uncertainties. Monte Carlo simulations do have the benefit of producing information on the parameters of the entire event, thus providing shape information which is necessary for observations of new processes or setting limits. None of the other methods can make shape predictions; they provide an estimate that the shape taken from Monte Carlo is normalized to for use in plots.

The scale factor method incorporates information from the Z-window in both data and Monte Carlo into the estimate. The production of Z bosons is a well understood Standard Model process, and comparing data to Monte Carlo predictions in the Z peak gives a good sense of needed corrections in regions of phase space that aren't as well understood. Furthermore, as this estimate incorporates information from data, systematic uncertainties in object reconstruction partially cancel out. Drawing

information from more regions can introduce sensitivity to low numbers of events; this estimate method shares the low number of generated events in the signal region with the unscaled predictions, but also uses the high $E_{\mathrm{T,Rel}}^{\mathrm{miss}}$ tail of the Z-window which is sparsely populated and hence the estimate carries larger statistical errors. Finally, the scale factor calculated in the Z-window is expected to be universal for this background, and thus can be used to rescale Monte Carlo predictions of the Z/γ^* background for control regions as well as the low, intermediate, and high mass signal regions.

The ratio of ratio method combines the ABCD estimate with Monte Carlo corrections, explicitly building a measure of non-closure into the estimate itself. Hence it carries larger systematic uncertainties than the ABCD method but smaller ones than the Monte Carlo or scale factor predictions. As it relies on four regions in Monte Carlo and three in data, this method is sensitive to statistical fluctuations in any of these areas, which is reflected in its larger statistical uncertainties. This method does assume that any violation of the ABCD method assumptions is accurately modeled in the Monte Carlo; it has no measure of how well the Monte Carlo models the needed distributions in data. If simulation were to mismodel the missing transverse energy, this method could introduce a bias into the estimate that the ABCD method would not have. While the agreement between this and the ABCD method indicate that this is not a problem, for this reason it is used only as a cross-check in this analysis.

The ABCD and A'B'CD methods carry comparable statistical uncertainties but smaller systematic ones to the other estimation methods. This is due to relying primarily on data for calculation, and so systematic errors in energy resolution and other parameters are identical in these estimates and the signal and control regions in which they are applied. Monte Carlo simulations are only used for background subtraction and hence mismodeling of these backgrounds only becomes significant if one of the B, C, or D control regions is expected to have a high level of contamination

from other processes. These methods do rely on strong underlying assumptions about the independence of dilepton mass and missing transverse energy for this background, and the size of systematic uncertainty reflects confidence in these assumptions. While they appear reasonable in the one jet channels, the zero jet $\mu\mu$ jet channel shows evidence that these assumptions do not always hold, and hence this channel carries the largest uncertainties. Due to its many control regions, the ABCD method's predictions cannot be applied outside the regions they were designed for, making this a less flexible estimate than either of the Monte Carlo methods.

The ABCD method's independence of Z/γ^*+jet Monte Carlo simulation was a strength during the course of the 2011 data-taking as beam and collision conditions at the LHC changed rapidly. Changing pile-up conditions required careful tuning of simulation, and this need was reflected in the missing transverse energy distributions. Three major productions of Monte Carlo samples were made in an attempt to tune simulations to the observed data over the course of 2011, each of which produced very different predictions for this background from Monte Carlo. By using only the data, the ABCD methods self-tuned to changing running conditions and provided stable predictions for comparison. As it is able to produce an estimate for this background without using Z/γ^*+jets simulation, the A'B'CD method is used for the statistical analysis in this dissertation.

For the final statistical evaluation of this analysis, transverse mass distributions of each background are necessary, and these the ABCD and A'B'CD methods cannot produce. Transverse mass is dependent on both dilepton mass and missing transverse energy and so cannot be reasonably computed for region A from the B, C, or D regions. Hence the shape of the transverse mass distributions of this background is taken from Monte Carlo and normalized to the data-driven prediction for the statistical analysis discussed in the following chapter.

CHAPTER IX

Statistical Methods and Observed Limits

This chapter presents the final results of the selections described in Chapter V and background esimates described in Chapters VII and VIII. The results of the cutflow are presented in Section 9.1. Finally, the method used to calculate limits and the results of those calculations are discussed in Section 9.2.

9.1 Final Results

A comparison of the observed number of events in data and the expected number of events from the background estimates is shown for the preselection requirements and zero jet events in Table 9.1. Comparisons for the inclusive Z control region and the zero jet Z and WW control regions are also shown. In the preselection of events prior to the missing transverse energy requirement, the expected background agrees well with the observed number of events, with discrepancies of less than 5%. This portion of the analysis is dominated by Z/γ^*+jets events in the same flavor channels, and shows good agreement within the Z-window as well.

The zero jet results are also shown in Table 9.1. As expected, the WW continuum is the dominant source of background contamination in the final signal region, though all other backgrounds are expected to give contributions similar in magnitude to the signal of a Higgs boson with $m_H = 125$ GeV. This mass is chosen as a representative

reference as it is slightly above the limits observed by LEP and within the reach of this analysis. This similarity between the expected signal and contribution between the top, Z/γ^*+jets, and W+jets backgrounds implies that additional work will be needed to decrease this contamination and improve the sensitivity of this analysis. A slight excess of data over the expected background is observed, but it is not statistically significant. The cut on transverse mass is listed for easy comparison at $m_H = 125$ GeV only; this cut is not applied to the data used for setting limits.

Transverse mass (m_T) is the final variable of interest in this analysis; for the Higgs signal, distributions should have a shoulder at the mass of the boson. In earlier versions of this analysis, windows were defined for each possible mass of the Higgs boson under study and applied to define the final signal region. This has been replaced by a fit to this distribution that will be discussed in Section 9.2. The transverse mass distribution for the zero jet channel is shown in Figure 9.1, with the expected shape from a Standard Model Higgs boson with $m_H = 125$ GeV overlaid for comparison. The expected background agrees with the observed data within uncertainties, and no features of note are seen in the distribution.

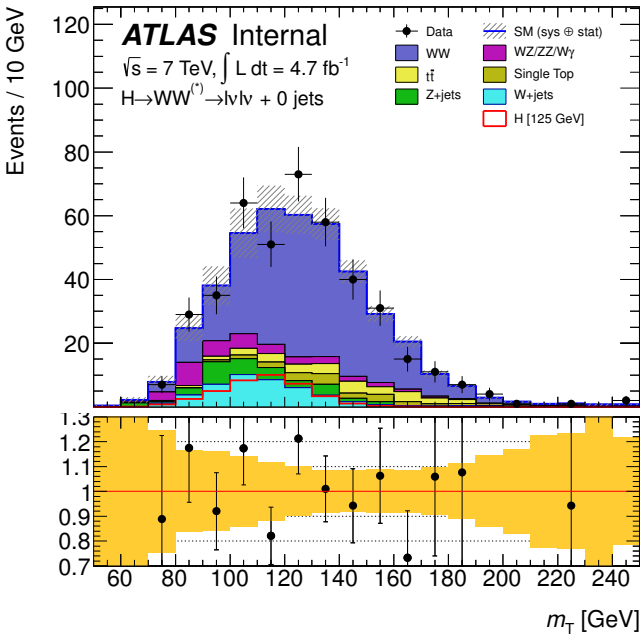

Figure 9.1: Distribution of transverse mass in the zero jet channel after all selections. All flavor channels are combined.

	Signal [125 GeV]	WW	$WZ/ZZ/W\gamma$	$t\bar{t}$	Single Top	Z/γ^*+jets	W+jets (MC)	Total Bkg.	Observed	Data/MC
lepton p_{T}	180.88 ± 0.40	3871.13 ± 9.55	3213.71 ± 20.73	16332.99 ± 27.65	1753.35 ± 11.35	2800841.59 ± 1309.28	1035.58 ± 55.09	2827048.36 ± 1310.98	2823123	1.00 ± 0.00
OS leptons	178.44 ± 0.39	3856.07 ± 9.53	2733.64 ± 15.79	16240.59 ± 27.57	1734.89 ± 11.30	2793671.35 ± 1307.67	694.49 ± 39.78	2818931.03 ± 1308.74	2816240	1.00 ± 0.00
$m_{\ell\ell} > 12, 10$ GeV	175.14 ± 0.39	3848.15 ± 9.52	2716.39 ± 15.68	16212.03 ± 27.55	1731.87 ± 11.29	2790810.80 ± 1307.34	686.23 ± 39.60	2816005.47 ± 1308.41	2806551	1.00 ± 0.00
Z veto (for $ee, \mu\mu$)	173.11 ± 0.39	3421.45 ± 8.97	559.70 ± 8.00	14456.63 ± 26.02	1542.39 ± 10.66	278762.28 ± 404.21	606.76 ± 37.50	299349.21 ± 407.09	298691	1.00 ± 0.00
$E_{\mathrm{T,rel}}^{\mathrm{miss}} > 45, 25$ GeV	90.83 ± 0.28	1829.52 ± 113.71	177.88 ± 4.82	7265.98 ± 494.64	832.79 ± 57.21	1522.96 ± 33.78	251.63 ± 24.96	11880.77 ± 684.85	12231	1.03 ± 0.06
Z control region (incl)	88.27 ± 0.28	2328.82 ± 7.41	2409.16 ± 14.79	9832.11 ± 21.45	1056.32 ± 8.82	2519138.88 ± 1244.38	462.78 ± 32.30	2535228.08 ± 1245.13	2526898	1.00 ± 0.00
0j: jet veto	56.74 ± 0.23	1273.34 ± 79.20	96.62 ± 3.71	172.36 ± 12.07	94.19 ± 6.94	1038.66 ± 27.91	173.94 ± 20.94	2849.12 ± 116.46	2849	1.00 ± 0.04
0j: $m_{\ell\ell} < 50$ GeV	45.22 ± 0.21	312.57 ± 19.60	40.81 ± 2.77	28.70 ± 2.33	19.08 ± 1.77	167.87 ± 9.60	56.42 ± 12.11	625.46 ± 30.72	645	1.03 ± 0.06
0j: $p_{\mathrm{T},\ell\ell}$ >45,30 GeV	40.14 ± 0.19	282.27 ± 17.72	34.79 ± 2.51	27.06 ± 2.21	17.63 ± 1.66	27.96 ± 6.42	43.68 ± 10.60	433.40 ± 28.39	443	1.02 ± 0.08
0j: $\Delta\phi_{\ell\ell} < 1.8$	39.03 ± 0.19	275.68 ± 17.31	33.32 ± 2.43	26.62 ± 2.18	17.49 ± 1.65	27.64 ± 6.42	38.94 ± 10.12	419.69 ± 27.80	429	1.02 ± 0.08
0j: $0.75 \cdot m_{\mathrm{H}} \leq m_{\mathrm{T}} \leq m_{\mathrm{H}}$	25.86 ± 0.15	107.82 ± 6.90	11.66 ± 1.29	6.50 ± 0.75	4.64 ± 0.67	14.05 ± 3.77	25.87 ± 8.45	170.55 ± 13.70	174	1.02 ± 0.11
0j: Z control region	46.59 ± 0.21	1586.83 ± 6.19	602.14 ± 6.29	185.63 ± 2.93	98.88 ± 2.69	1990016.47 ± 1113.62	293.06 ± 27.72	1992783.01 ± 1114.01	1986323	1.00 ± 0.00
0j: WW control region	0.28 ± 0.03	470.72 ± 3.28	25.69 ± 1.46	87.24 ± 1.96	42.46 ± 1.77	7.34 ± 2.17	37.45 ± 9.33	670.90 ± 10.57	697	1.04 ± 0.04

Table 9.1: The expected number of background events and the observed number of data events for preselection and zero jet events. All flavor channels are combined and statistical errors only are included. The expectation from a Standard Model Higgs boson with $m_H = 125$ GeV is shown for comparison.

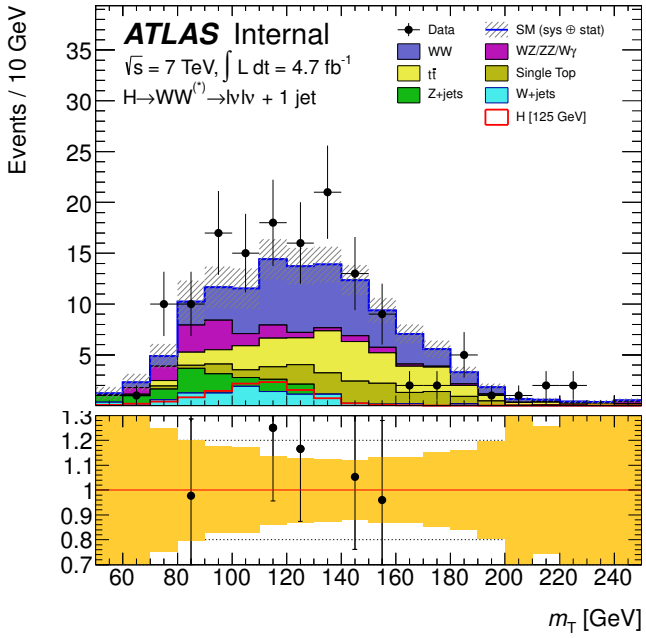

Figure 9.2: Transverse mass distribution in the one jet channel after all selections. All flavor channels are combined.

Results for the one jet channels after the jet multiplicity cut are shown in Table 9.2. All flavor channels are combined. In this channel, the dominant background contributions are predicted to be from top quark decays, though the WW continuum gives almost as large a contamination. The reducible backgrounds also contaminate the signal region with approximately the same number of events as the signal itself. Increasing the purity of the signal region would increase the sensitivity of this analysis if these backgrounds can be removed while maintaining a reasonable signal acceptance. Similar to the results in the zero jet channel, the ratio between the observed data and expected background is consistently greater than one, but the deviations are not statistically significant. The transverse mass distribution in the one jet channel is shown in Figure 9.2; it has a slight excess at low transverse masses that is balanced by deficits in the upper tail of the distribution, bringing the overall ratio close to one. This excess is not significantly greater than the uncertainties on the measurements, and so cannot be ruled out as a statistical fluctuation.

	Signal [125 GeV]	WW	$WZ/ZZ/W\gamma$	$t\bar{t}$	Single Top	Z/γ^*+jets	W+jets (MC)	Total Bkg.	Observed	Data/MC
1j: exactly one jet	22.67 ± 0.13	363.48 ± 53.55	55.76 ± 2.66	1417.77 ± 58.87	429.69 ± 18.59	357.19 ± 16.63	64.24 ± 13.32	2688.12 ± 141.77	2706	1.01 ± 0.06
1j: b-jet veto (25 GeV, 80% eff)	20.89 ± 0.13	337.80 ± 49.77	51.97 ± 2.55	392.01 ± 16.70	132.76 ± 6.33	332.65 ± 15.81	62.23 ± 13.23	1309.41 ± 83.52	1369	1.05 ± 0.07
1j: $p_{\mathrm{T}}^{\mathrm{tot}} < 30$ GeV	14.03 ± 0.10	238.89 ± 35.22	34.05 ± 2.16	171.27 ± 7.61	76.66 ± 3.99	108.11 ± 7.73	46.65 ± 12.90	675.63 ± 52.14	684	1.01 ± 0.09
1j: $Z \to \tau\tau$ veto	13.99 ± 0.10	233.10 ± 34.37	33.59 ± 2.16	164.40 ± 7.33	73.68 ± 3.86	84.74 ± 7.18	46.65 ± 12.90	636.17 ± 50.68	644	1.01 ± 0.09
1j: $m_{\ell\ell} < 50$ GeV	10.93 ± 0.09	51.65 ± 7.69	13.55 ± 1.57	31.23 ± 1.84	16.92 ± 1.34	23.69 ± 3.06	23.07 ± 11.73	160.10 ± 16.50	170	1.06 ± 0.14
1j: $\Delta\phi_{\ell\ell} < 1.8$	10.11 ± 0.09	47.08 ± 7.02	13.12 ± 1.57	29.59 ± 1.77	16.42 ± 1.32	10.02 ± 1.98	22.43 ± 11.71	138.66 ± 15.56	145	1.05 ± 0.15
1j: $0.75 \cdot m_{\mathrm{H}} \leq m_{\mathrm{T}} \leq m_{\mathrm{H}}$	6.30 ± 0.07	16.92 ± 2.57	4.65 ± 0.97	7.44 ± 0.72	4.10 ± 0.59	4.60 ± 1.34	5.87 ± 3.23	43.58 ± 5.17	56	1.28 ± 0.23
1j: Z control region	25.85 ± 0.14	541.61 ± 3.51	817.36 ± 8.80	1546.86 ± 8.46	490.03 ± 6.01	397782.56 ± 489.69	130.28 ± 15.89	401308.70 ± 490.15	401641	1.00 ± 0.00
1j: WW control region	0.12 ± 0.02	127.62 ± 1.68	11.51 ± 1.00	85.41 ± 1.92	32.99 ± 1.56	8.67 ± 1.96	10.40 ± 3.45	276.61 ± 5.07	270	0.98 ± 0.06
1j: Top control region	1.14 ± 0.03	19.49 ± 0.70	1.92 ± 0.47	440.66 ± 4.55	170.86 ± 3.57	6.94 ± 1.68	1.35 ± 0.95	641.21 ± 6.16	676	1.05 ± 0.04

Table 9.2: The expected number of background events and the observed number of data events for the one jet events after the jet multiplicity cut. Statistical errors only are shown. All flavor combinations (ee, $e\mu$, $\mu\mu$) are combined. The expected number of events from a Standard Model Higgs boson with $m_H = 125$ GeV is shown for comparison.

Table 9.3 shows the comparison between observed data and expected background in the two jet channel. This channel has a lower Higgs boson production cross-section than the zero and one jet cases, and thus has significantly lower statistics even with this integrated luminosity. The applied cuts are also stringent to produce a clean final sample. Only one event is observed in data after the $\Delta\phi_{\ell\ell}$ requirement, with an expectation of 1.8 ± 0.26; agreement cannot be accurately assessed at this point. The expected contamination from the WW continuum and top backgrounds is small, and there are insufficient events to assess the background contribution from the Z/γ^*+jets and W+jets backgrounds. Unlike the zero and one jet channels where the dominant backgrounds populate the final signal region more than the expected signal, the two jet requirements appear to select a similar number of Higgs boson decays as the main backgrounds. While desirable, this feature could result from the low number of events in the analysis and will need to be reassessed with more data to see if the purity of the final signal region is sufficient.

	Signal [125 GeV]	WW	$WZ/ZZ/W\gamma$	$t\bar{t}$	Single Top	Z/γ^*+jets	W+jets (MC)	Total Bkg.	Observed	Data/MC
$\geq$ 2j: at least 2 jets	11.42 ± 0.09	142.27 ± 1.73	25.50 ± 1.53	5555.70 ± 16.11	313.60 ± 4.81	120.09 ± 6.68	13.46 ± 2.70	6170.63 ± 18.44	6676	1.08 ± 0.01
$\geq$ 2j: central jet veto	8.95 ± 0.08	113.27 ± 1.55	19.54 ± 1.44	3278.68 ± 12.68	238.38 ± 4.33	89.48 ± 5.94	10.61 ± 2.38	3749.96 ± 15.00	3811	1.02 ± 0.02
$\geq$ 2j: opp. hemispheres	4.85 ± 0.05	52.95 ± 1.05	8.26 ± 0.90	1380.03 ± 8.16	110.79 ± 2.94	35.62 ± 3.62	4.71 ± 1.57	1592.35 ± 9.63	1630	1.02 ± 0.03
$\geq$ 2j: $\Delta\eta_{\rm jj} > 3.8$	2.00 ± 0.02	9.29 ± 0.44	1.09 ± 0.19	68.09 ± 1.66	7.94 ± 0.74	5.71 ± 1.71	0.00 ± 0.00	92.12 ± 2.54	96	1.04 ± 0.11
$\geq$ 2j: $m_{\rm jj} > 500$ GeV	1.41 ± 0.02	4.35 ± 0.29	0.44 ± 0.12	27.15 ± 1.00	2.27 ± 0.37	0.71 ± 0.35	0.00 ± 0.00	34.92 ± 1.16	30	0.86 ± 0.16
$\geq$ 2j: b-jet veto ($p_{\rm T} \geq 25$ GeV, 80% eff)	1.28 ± 0.02	3.92 ± 0.27	0.38 ± 0.11	10.10 ± 0.59	1.01 ± 0.25	0.71 ± 0.35	0.00 ± 0.00	16.12 ± 0.79	13	0.81 ± 0.23
$\geq$ 2j: $p_{\rm T,tot} < 30$ GeV	0.95 ± 0.01	2.50 ± 0.22	0.20 ± 0.08	4.13 ± 0.38	0.60 ± 0.19	0.27 ± 0.20	0.00 ± 0.00	7.71 ± 0.52	6	0.78 ± 0.32
$\geq$ 2j: $Z \to \tau\tau$ veto	0.95 ± 0.01	2.44 ± 0.21	0.21 ± 0.08	4.09 ± 0.38	0.60 ± 0.19	0.16 ± 0.16	0.00 ± 0.00	7.51 ± 0.51	6	0.80 ± 0.33
$\geq$ 2j: $m_{\ell\ell} < 80$ GeV	0.94 ± 0.01	1.08 ± 0.15	0.10 ± 0.06	1.33 ± 0.24	0.35 ± 0.14	0.16 ± 0.16	0.00 ± 0.00	3.01 ± 0.36	2	0.66 ± 0.48
$\geq$ 2j: $\Delta\phi_{\ell\ell} < 1.8$	0.82 ± 0.01	0.75 ± 0.12	0.09 ± 0.06	0.83 ± 0.20	0.14 ± 0.10	0.00 ± 0.00	0.00 ± 0.00	1.81 ± 0.26	1	0.55 ± 0.56
$\geq$ 2j: $0.75 \cdot m_{\rm H} \leq m_{\rm T} \leq m_{\rm H}$	0.48 ± 0.01	0.22 ± 0.07	0.01 ± 0.01	0.18 ± 0.08	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.40 ± 0.11	0	0.00 ± nan
$\geq$ 2j: Z control region	15.83 ± 0.11	200.38 ± 2.08	989.66 ± 10.08	8099.62 ± 19.49	467.41 ± 5.87	131339.86 ± 261.79	39.43 ± 4.71	141136.37 ± 262.83	138934	0.98 ± 0.00
$\geq$ 2j: WW control region	0.00 ± 0.00	1.37 ± 0.16	0.11 ± 0.06	2.76 ± 0.30	0.26 ± 0.12	0.00 ± 0.00	0.00 ± 0.00	4.50 ± 0.36	4	0.89 ± 0.45
$\geq$ 2j: Top control region	0.09 ± 0.00	0.36 ± 0.09	0.01 ± 0.01	8.18 ± 0.55	0.83 ± 0.22	0.00 ± 0.00	0.00 ± 0.00	9.37 ± 0.60	10	1.07 ± 0.34

Table 9.3: The expected number of background events and the observed number of data events for two jet events after the jet multiplicity cut. Statistical errors only are shown. All flavor combinations are included. The expected number of events from a Standard Model Higgs boson with $m_H = 125$ GeV is shown for comparison.

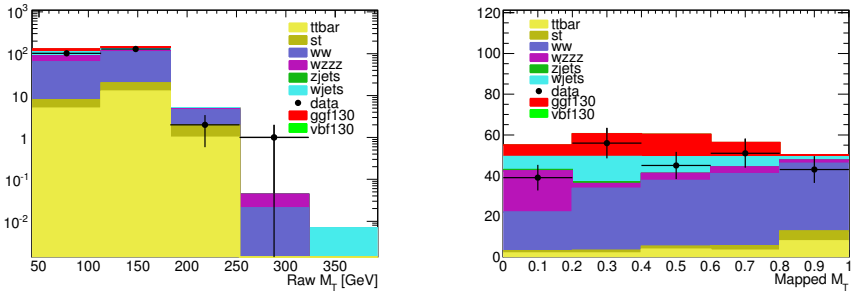

Figure 9.3: The original m_T distribution (left) and mapped distribution (right) each in five bins; the distributions shown are for the zero jet channel. The original distribution suffers from low statistics and empty bins in the high tails that have been removed via the mapping for better fitting.

9.2 Statistical Methods and Limits

As mentioned in Section 9.1, the variable of interest is the transverse mass of the system. No cuts are applied to this quantity; instead a fit of this spectrum to expectation is used for setting limits and testing agreement with the Standard Model Higgs boson prediction. This gives increased sensitivity of the analysis, incorporates shape information into the limit-setting, and allows for parameterization of the results as a function of the Higgs mass, which is necessary prior to a mass measurement.

As shown in Figures 9.1 and 9.2, the transverse mass distributions have a central peak and long tails with low statistics in the various backgrounds. To avoid the uneven distribution of simulated events, the transverse mass is mapped in each flavor and jet multiplicity channel to uniformly distribute the sum of expected background events, which eliminates empty bins. An example of the results of this mapping is shown in Figure 9.3. The zero jet channels are mapped to five bins, the one jet channels to three, and the two jet channel uses no binning due to the small number of events in the final signal region.

Fits are performed using likelihood functions constructed of Poisson probability terms for the signal and control regions in each flavor channel, jet multiplicity, and

m_T bin. Uncertainties are treated as nuisance parameters in the distributions with additional constraints. The likelihood is then:

$$\mathcal{L} = \{ \prod_{k=ee,e\mu,\mu\mu} \prod_{j=0}^{Njets} \prod_{i=1}^{Nbins} P(N_{ijk}|\mu s_{ijk} + \sum_{\ell}^{Nbg} b_{ijk\ell})\} \times \{\prod_{i=1}^{N_\theta} N(\tilde{\theta}|\theta)\} \qquad (9.1)$$

where $P(N_{ijk}|\mu s_{ijk} + \sum_{\ell}^{Nbg} b_{ijk\ell})$ are the Poisson distributions and $N(\tilde{\theta}|\theta)$ are the constraints, either gaussian or Poisson distributions in this analysis. The signal and background expectations are also functions of the nuisance parameters θ, so the signal expectation is written as $s = s_0 \prod \nu(\theta)$ where $\nu(\theta)$ is function interpolating the impact of the uncertainty between the known nominal and $\pm$ one standard deviation measurements. Nuisance parameters are treated in four categories:

- Flat systematic uncertainty: for systematic uncertainties with no effect on the shape of the transverse mass, the nuisance parameter is taken as $\nu_{flat}(\theta) = \kappa^\theta$ with a Gaussian constraint. κ is determined by measuring the uncertainty for $\theta = \pm 1$ and gives the contribution from this nuisance parameter to the signal or background at one standard deviation from the nominal value.

- Shape systematic uncertainty: for systematic uncertainties that do affect the shape, the nuisance parameter is split into a flat component affecting only the normalization and treated as described above, and a shape component. The shape component is taken as $\nu_{shape}(\theta) = 1 + \epsilon\theta$ with a Gaussian constraint.

- Statistical uncertainty: These arise from data-driven background estimation methods and from Monte Carlo predictions. The nuisance parameter is written as $\nu_{stat}(\theta) = \theta$ with a Poisson constraint.

- Data control regions: Here nuisance parameters enter through the expected number of events $\lambda = \mu s + \theta b_{target} + \sum_i^{N_{bg}-1} b_i$ where b_{target} is the background constrained by the control region in question and the summation is over all other

backgrounds under study. The constraint is given by a Poisson distribution.

The test statistic is then constructed as:

$$q_\mu = -2\ln(\frac{\mathcal{L}(\mu, \hat{\theta_\mu})}{\mathcal{L}(\hat{\mu}, \hat{\theta})}) \tag{9.2}$$

Here μ is a parameter that modifies the signal strength such that $\mu = 0$ corresponds to the background-only hypothesis and θ is the set of nuisance parameters. The denominator is the maximum possible likelihood while the numerator is maximized for a given value of μ. Limits are then computed using the CL_s method with a 95% confidence level; this method entails evaluating the test statistic in the signal+background case and the background only case and using the ratio between them ($CL_s = CL_{s+b}/CL_b$) to determine limits [139–141].

No significant excess of observed data with respect to the expectation from the Standard Model background is found over the analyzed mass range. The ratio of the observed upper limit on the cross-section to the expected cross-section of a Standard Model Higgs boson is shown in Figure 9.4; masses for which this ratio is less than one are taken as excluded. The observed exclusion range is $131 < m_H < 260$ GeV, while the expected exclusion is $127 < m_H < 230$ GeV. A slight excess of events is observed in the low mass region, which prevents the observed limit from matching the expected one in this region. The p-value of the observed data, or probability of seeing an excess of this magnitude or larger in the background-only case, is shown in Figure 9.5. The minimum probability within the excess is located at $m_H = 140$ GeV and is 0.2, and it is not statistically significant. A deficit of observed events relative to expectation in the high mass region allows the observed limit to exceed expectation there. Limits calculated in each flavor and jet multiplicity channel are shown in Figure 9.6 and show no pattern of excesses or deficits between the flavor or jet multiplicities. As expected, the zero jet channels contribute most of the sensitivity

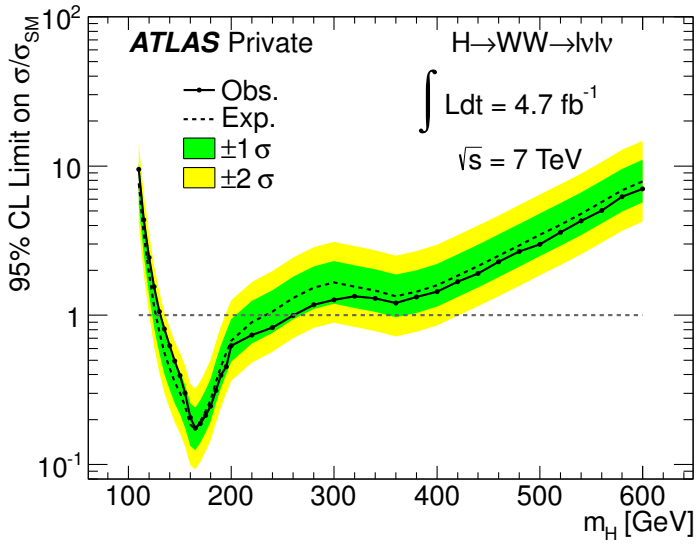

Figure 9.4: Expected (dashed) and observed (solid) 95% confidence level upper limits on the ratio between calculated and Standard Model cross-sections as a function of m_H. All data, flavor channels, and jet multiplicities are combined. The green and yellow regions indicate the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands respectively on the expected limit.

of this search. These new limits add greatly to the range of mass excluded to the Standard Model Higgs boson, but still leave a small window between 115 and 130 GeV and the range above to 260 GeV the next iteration of the search.

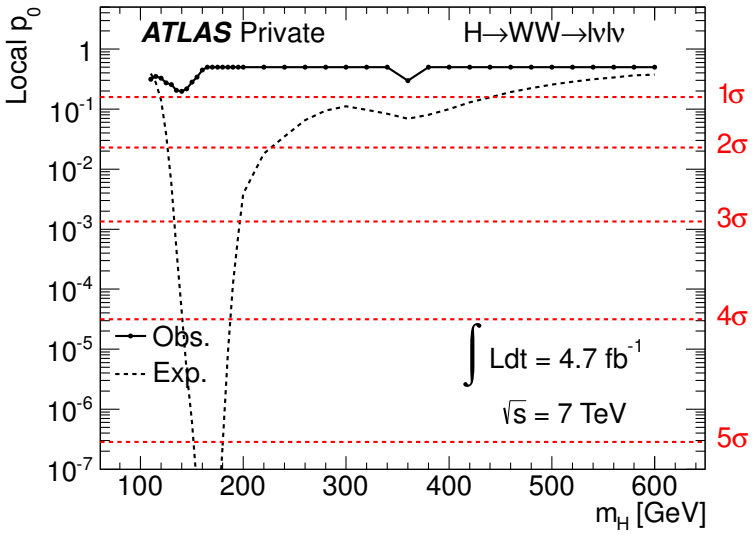

Figure 9.5: The expected (dashed) and observed (solid) p-values, or the probability of the Standard Model backgrounds producing the observed distribution in data, as a function of m_H.

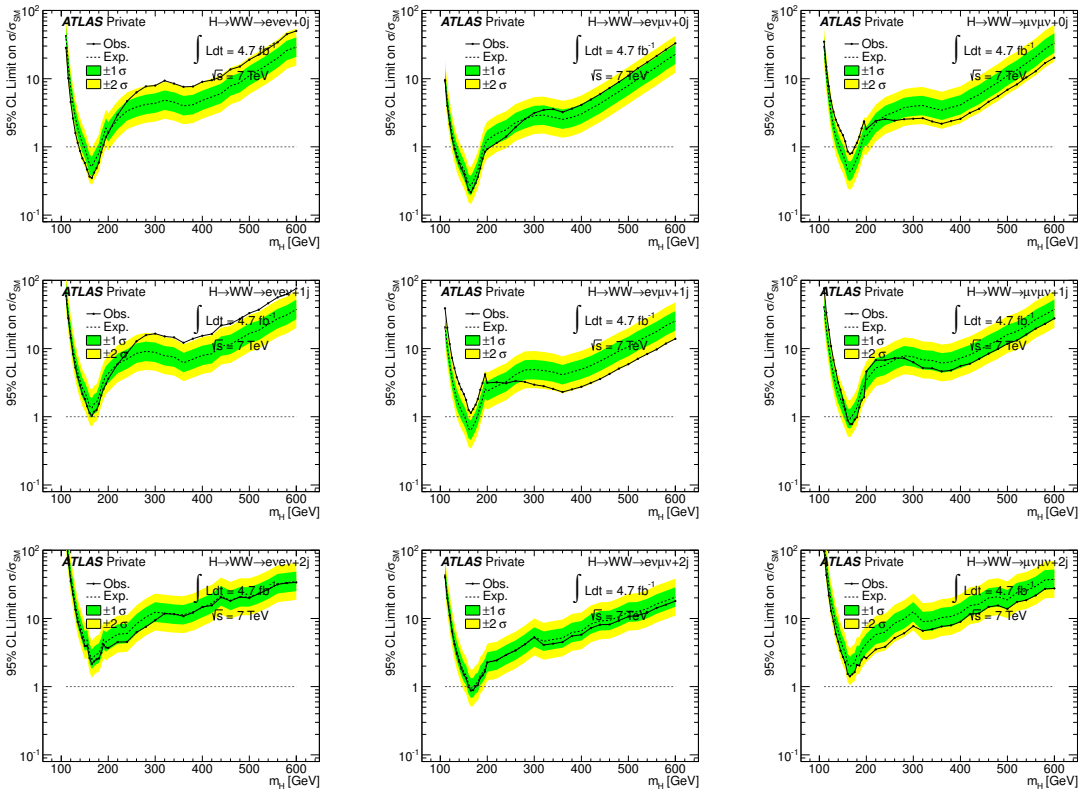

Figure 9.6: Expected (dashed) and observed (solid) 95% confidence level upper limits on the ratio between calculated and Standard Model cross-sections as a function of m_H in each flavor and jet multiplicity channel. The zero jet channels are the top row, one jet channels the center row, and the two jet channels the bottom row; ee results are in the left column, $e\mu$ in the center column, and $\mu\mu$ in the right column.

CHAPTER X

Conclusion

The Standard Model of particle physics has passed numerous tests of its predictions, but it will not be fully tested without proof of the mechanism of electroweak symmetry breaking. Discovery or exclusion of the Higgs boson, signature of the simplest method of breaking the electroweak symmetry, will provide critical information to complete this last portion of the Standard Model. Theoretical motivations and the results of precision electroweak measurements indicate that discovery and exclusion lie within the reach of the LHC and its experiments, with masses below 160 GeV.

This dissertation presents the status of the search for the Standard Model Higgs boson in the $H \to WW \to \ell\nu\ell\nu$ final state. The ATLAS detector recorded over 5 fb^{-1} of data during the 2011 $\sqrt{s} = 7$ TeV data-taking at the LHC, of which 4.7 fb^{-1} are included in this analysis. The final signal region was defined using a cut-based analysis, with control regions, data-driven estimates, and Monte Carlo simulation used to estimate the contamination from various backgrounds. The results presented here exclude the range of $131 < m_H < 260$ GeV as possible masses of a Standard Model Higgs boson at 95% confidence level.

While the results of this analysis have greatly reduced the mass range a Standard Model Higgs boson could occupy, neither discovery nor full exclusion has been achieved. Combined with the results of other channels, however, tantalizing hints

begin to emerge. Analysis of the final state $H \to \gamma\gamma$ with the ATLAS detector over the 2011 dataset observed an excess located at $m_H = 126$ GeV [142]. The local significance was calculated to be 2.8 standard deviations away from the background-only expectation, though when the effect of analyzing large number of mass points and possible statistical fluctuations was included, this decreased to a significance of 1.5 standard deviations. The analysis of the $H \to ZZ^{(*)} \to \ell\ell\ell\ell$ final state observed an excess at $m_H = 125$ GeV [143]. These hints promise that the next phase of the search for the Higgs boson will produce answers to some of the Standard Model's most pressing questions.

As this channel is one of the most sensitive in the remaining low mass range, it will continue to be important as the search continues. Additional work in developing new ways to distinguish background contributions, removing their contamination from the signal regions, and including leptons with lower transverse momentum thresholds will increase the sensitivity of this channel. Collisions at the LHC have begun again with a center of mass energy of 8 TeV, increasing the available data and the reach of this analysis. This year should give a definite answer as to the direction the Standard Model will take in its final test.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Higgs,P., Physical Review Letters **13** (1964) 508–509

[2] Djouadi, A, Phys. Rept. **459** (2008) 1–241

[3] Cheng, H-C; Dobrescu, B; Hill,C, Nucl. Phys. B **589** (2000) 249–268

[4] Barate, R et al, Phys. Lett. B **565** (2003) 61–75

[5] The TEVNPH Working Group, Combined CDF and D0 Search for Standard Model Higgs Boson Production with up to 10 fb^{-1} of Data, FERMILAB-CONF-12-065-E, Fermilab, 2012

[6] Thomson, J.J., Philosophical Magazine **90:S1** (2010) 25–29

[7] Anderson, C; Neddermeyer, S, Phys. Rev. **50** (1936) 263

[8] Street, J; Stevenson, E, Phys. Rev. **52** (1937) 1003

[9] Perl, M et al, Phys. Rev. Lett. **35** (1975) 1489–1492

[10] Cowan, C; Harrison, F; Kruse, H; McGuire, A; Reines, F, Science **124** (1956)

[11] Danby, G et al, Phys. Rev. Lett. **9** (1962)

[12] Kodama, K et al, Phys. Lett. B **504** (2001) 218–224

[13] Bloom, E. et al, Phys. Rev. Lett. **23** (1969) 930–934

[14] Breidenbach, M. et al, Phys. Rev. Lett. **23** (1969) 935–939

[15] Augustin, J. -E. et al, Phys. Rev. Lett. **33** (1974) 1406–1408

[16] Aubert, J. et al, Phys. Rev. Lett. **33** (1974) 1404–1406

[17] Herb, S. W. et al, Phys. Rev. Lett. **39** (1977) 252–255

[18] Abe, F. et al, Phys. Rev. Lett. **74** (1995) 2626–2631

[19] Abachi, S. et al, Phys. Rev. Lett. **74** (1995) 2422–2426

[20] K. Nakamura et al., JPG **37** (2010)

[21] Glashow, S, Nucl. Phys. **22** (1961) 579–588

[22] Hasert, F et al, Phys. Lett. B **46** (1973) 121–124

[23] Hasert, F et al, Phys. Lett. B **46** (1973) 138–140

[24] UA1 Collaboration, Physics Letters B **122** (1983) 103–116

[25] UA2 Collaboration, Physics Letters B **122** (1983) 476–485

[26] Arnison, G et al, Phys. Lett. B **126** (1983) 398–410

[27] Bagnaia, P et al, Phys. Lett. B **129** (1983) 130–140

[28] Brandelik, R et al, Phys. Lett. B **86** (1979) 243–249

[29] Berger, C et al, Phys. Lett. B **86** (1979) 418–425

[30] Barber, D et al, Phys. Rev. Lett. **43** (1979) 830–833

[31] Brout,R. and Englert,F., Physical Review Letters **13** (1964) 321–323

[32] Guralnik,G.; Hagen,C.R.; Kibble,T., Physical Review Letters **13** (1964) 585–587

[33] Alvaverz-Gaume, L and Ellis, J, Eyes on a prize particle, 2011

[34] Weinberg, Steven, Phys. Rev. Lett. **19** (1967) 1264–1266

[35] Salam, A, in *Proc. of the 8th Nobel Symposium on 'Elementary particle theory, relativistic groups and analyticity'*, edited by Svartholm, N (Almqvist & Wiksell, 1968)

[36] Lee, B; Quigg, C; Thacker, H, Phys. Rev. D **16** (1977) 1519–1531

[37] Dittmaier, S; Mariotti, C; Passarino, G; Tanaka, R; et al, Handbook of LHC Higgs Cross Sections, CERN-2011-002, CERN, 2011

[38] The ALEPH, CDF, D0, DELPHI, L3, OPAL, SLC Collaborations, the LEP Electroweak Working Group, the Tevatron Electroweak Working Group, the SLD electroweak and heavy flavor groups, Precision Electroweak Measurements and Constraints on the Standard Model, CERN-PH-EP-2010-095, CERN, 2010

[39] Baak, M et al, arXiv:1107.0975v1 (2011)

[40] CERN. https://cdsweb.cern.ch/record/40525

[41] Evans, L. and Bryant, P, JINST **3** (2008)

[42] CERN. https://cdsweb.cern.ch/record/42384

[43] ATLAS Collaboration, ATLAS Sensitivity Prospects for Higgs Boson Production at the LHC Running at 7, 8, or 9 TeV, ATL-PHYS-PUB-2010-015, CERN, November 2010

[44] Spira, M; Djuoadi, A; Graudenz, D; Zerwas, P.M., arXiv:9504378v1 (1995)

[45] Baglio, J and Djouadi, A, arXiv:1012.0530v4 (2010)

[46] CERN. https://cdsweb.cern.ch/record/1095924

[47] ATLAS Collaboration, Journal of Instrumentation **3** (2008)

[48] CERN. https://cdsweb.cern.ch/record/1095926

[49] ATLAS Collaboration, Detector and physics performance technical design report, ATLAS-TDR-14, CERN, May 1999

[50] ATLAS Collaboration, Inner Detector Technical Design Report Volume 1, ATLAS-TDR-4, CERN, April '997

[51] CERN. https://cdsweb.cern.ch/record/1095927

[52] CERN. https://cdsweb.cern.ch/record/1095929

[53] Catani, S et al, JHEP **11** (2001) 063

[54] Krauss, F, JHEP **08** (2002) 015

[55] Mangano, M; Moretti, M; Pittau, R, Nucl. Phys. B **632** (2002) 343–362

[56] Hoeche, S et al, JHEP **05** (2009) 053

[57] Harlander, R; Kilgore, W, Phys. Rev. Lett. **88** (2002) 201801

[58] Anastasiou, C; Melnikov, K, Nucl. Phys. B **646** (2002) 220

[59] Ravindran,V; Smith, J; van Neervan, W, Nucl. Phys. B **665** (2003) 325

[60] Anastasiou, C; Boughezal, R; Petriello, F, JHEP **04** (2009) 003

[61] de Florian, D; Grazzini, M, Phys. Lett. B **674** (2009) 291

[62] Aglietti, U; Bonciani, R; Degrassi, G; Vicini, A, Phys. Lett. B **595** (2004) 432

[63] Actis, S; Passarino, G; Sturm, C; Uccirati, S, Phys. Lett. B **670** (2008)

[64] Bolzoni, P; Maltoni, F; Moch, S-O; Zaro, M, Phys. Rev. Lett. **105** (2010) 011801

[65] Ciccolini, M; Denner, A; Dittmaier, S, Phys. Rev. Lett. **99** (2007) 161803

[66] Ciccolini, M; Denner, A; Dittmaier, S, Phys. Rev. D **77** (2008) 013002

[67] Brein, O; Djouadi, A; Harlander, R, Phys. Lett. B **579** (2004) 149

[68] Ciccolini, M; Dittmaier, S; Kramer, M, Phys. Rev. D **68** (2003) 073003

[69] Djouadi, A; Kalinowski, J; Spria, M, Comput. Phys. Commun. **108** (1998) 56

[70] Botje, M; Butterworth, J; et al, CERN Note (2011)

[71] Nason, P, JHEP **0411** (2004) 040

[72] Frixione, S; Nason, P; Oleari, C, JHEP **0711** (2007) 070

[73] Sjostrand, T; Mrenna, S; Skands, P, JHEP **0605** (2006) 026

[74] Bozzi, G; Catani, S; de Florian, D; Grazzini, M, Phys. Lett. B **564** (2003)

[75] Bozzi, G; Catani, S; de Florian, D; Grazzini, M, Nucl. Phys. B **737** (2006) 73

[76] Bozzi, G; Catani, S; de Florian, D; Grazzini, M, Nucl. Phys. B **791** (2008) 1

[77] de Florian, D; Ferrera, G; Grazzini, M; Tommasini, D, JHEP **064** (2011)

[78] Frixione, S and Webber, B., JHEP **06** (2002) 029

[79] Frixione, S; Nason, P; Webber, B, JHEP **08** (2003) 007

[80] Frixione, S; Laenen, E; Motylinski, P; Webber, B, JHEP **03** (2006) 092

[81] Fixione, S; Laenen, E; Motylinski, P; Webber, B; White, C, JHEP **07** (2008) 029

[82] Frixione, S; Stoeckli, F; Torrielli, P; Webber, B, JHEP **01** (2011) 053

[83] Carcella, H et al, JHEP **01** (2001) 010

[84] Binoth, T; Ciccolini, M; Kauer, N; Kramer, M, JHEP **12** (2006) 046

[85] Gleisberg, T; Hoeche, S; Drauss, F; Schaelicke, A; Schumann, S; Winter, J, JHEP **02** (2004) 056

[86] Alwall, J and others, JHEP **09** (2007) 028

[87] Kersevan, B and Richter-Was, E, arXiv:0405247 (2004)

[88] Mangano, M et al, JHEP **07** (2003)

[89] Nadolsky, P et al., Phys. Rev. D **78** (2008) 013004

[90] Sherstnev, A and Thorne, R, Eur. Phys. J. C **55** (2008) 553

[91] Melnikov, K and Petriello, F, Phys. Rev. Lett. **96** (2006) 231803

[92] Melnikov, K and Petriello, F, Phys. Rev. D **74** (2006) 114017

[93] Catani, S et al., Phys. Rev. Lett. **103** (2009) 082001

[94] Kidonakis, N, PoS: ICHEP (2010)

[95] Kidonakis, N, Phys. Rev. D **83** (2011) 091503

[96] Kidonakis, N, Phys. Rev. D **81** (2010) 054028

[97] Agostinelli, S et al., Nucl. Instrum. Meth. **A506** (2003) 250

[98] Aad, G et al., Eur. Phys. J. **C70** (2010) 823

[99] Aad, G et al., ATLAS Computing: Technical Design Report, CERN-LHCC-2005-022, CERN, 2005

[100] The ATLAS Collaboration, The Expected Performance of the Inner Detector, CERN-OPEN-2008-020, CERN, December 2008

[101] Cornelissen, T; Elsing, M; Fleischmann, S; Gavrilenko, I; Liebig, W; Moyse, E; Salzburger, A, Concepts, Design, and Implementation of the ATLAS New Tracking (NEWT), ATL-SOFT-PUB-2007-007, CERN, December 2007

[102] Piacquadio, G; Prokofiev, K; Wildauer, A, J. Phys.: Conf. Ser. **119** (2008)

[103] Lampl, W; Laplace, S; Lelas, D; Loch, P; Ma, H; Menke, S; Rajagopalan, S; Rousseau, D; Snyder, S; Unal, G, Calorimeter Clustering Algorithms: Description and Performance, ATL-LARG-PUB-2008-002, ATL-COM-LARG-2008-003, CERN, Geneva, Apr 2008

[104] Aad, G et al, Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data, , CERN, 2011

[105] The ATLAS Collaboration, Muon identification and measurements, CERN/LHCC 99-14, CERN, May 1999

[106] Hassani, S; Chevalier, L; Lancwith tailon, E; Laporte, J.-F.; Nicolaidou, R; Ouraou, A, NIM A **572** (2007) 77–79

[107] Nicolaidou, R; Chevalier, L; Hassani, S; Laporte, J-F; Le Menedeu, E; Ouraou, A, J. Phys.: Conf. Ser. **219** (2010)

[108] The ATLAS Collaboration, Muon Reconstruction and Identification: Studies with Simulated Monte Carlo Samples, CERN-OPEN-2008-020, CERN, December 2008

[109] The ATLAS Collaboration, Jet Reconstruction Performance, CERN-OPEN-2008-020, CERN, December 2008

[110] Cacciari, M; Salam, G; Soyez, G, JHEP **04** (2008) 063

[111] Cacciari, M and Salam, G, Phys. Lett. **B641** (2006) 57–61

[112] Cacciari, M; Salam, G; Soyez, G

[113] Doug Schouten, In Situ Measurements of Jet Energy Scale in ATLAS, ATL-PHYS-SLIDE-2011-175, CERN, April 2011

[114] The ATLAS Collaboration, b-Tagging Performance, CERN-OPEN-2008-020, CERN, December 2008

[115] Piacquadio, G; Weiser, Christian, J. Phys.: Conf. Ser. **119** (2008)

[116] The ATLAS Collaboration, Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data, ATLAS-CONF-2011-102, CERN, July 2011

[117] The ATLAS Collaboration, Measurement of Missing Travsverse Energy, CERN-OPEN-2008-020, CERN, December 2008

[118] Ask, S; Malon, D; Pauly, T; Shapiro, M, Report from the Luminosity Task Force, ATL-GEN-PUB-2006-002, CERN, July 2006

[119] The ATLAS Collaboration, Eur. Phys. J. C **71** (2011)

[120] ATLAS

[121] Corso-Radu, A; Hadavand, H; Illchenko, Y; Kolos, S; Okawa, H; Slagle, K; Taffard, A, Data Quality Monitoring Framework for the ATLAS experiment; Performance achieved with colliding beams at the LHC, ATL-DAQ-PROC-2011-007, CERN, January 2011

[122] Aad, G et al, arXiv:1112.6426v1 (2011)

[123] Abazov, V.M. and others, arXiv:0612040v1 (2006)

[124] Miller, D; Schwartzman, A; Su, D, Jet-Vertex Association Alogirthm, ATL-COM-PHYS-2008-008, CERN, January 2008

[125] Miller, D; Schwartzman, A; Su, D, Pile-up jet energy scale corrections using the jet-vertex fraction method, ATL-COM-PHYS-2009-180, CERN, September 2009

[126] K. Nakamura et al., JPG **37** (2010)

[127] The ATLAS Collaboration, Calibrating the b-Tag Efficiency and Mistag Rate in 35 pb^{-1} of Data with the ATLAS Detector, ATLAS-CONF-2011-089, CERN, June 2011

[128] Acosta, D. and others, arXiv:040678v2 (2004)

[129] Abazov, V.M. and others, Phys. Rev. D **76** (2007)

[130] The ATLAS Collaboration, Muon reconstruction efficiency in reprocessed 2010 LHC proton-proton collision data recorded with the ATLAS detector, ATLAS-CONF-2011-063, CERN, March 2011

[131] The ATLAS Collaboration, ATLAS Muon Momentum Resolution in the First Pass Reconstruction of the 2010 $p-p$ Collision Ddata at $\sqrt{s} = 7$ TeV, ATLAS-CONF-2011-046, CERN, March 2011

[132] The ATLAS Collaboration, Luminosity Determination in pp Collisions at $\sqrt{s} =$ 7 TeV using the ATLAS Detector, ATLAS-CONF-2011-116, CERN, 2011

[133] The ATLAS Collaboration, Search for the Standard Model Higgs boson in the $H \to WW^* \to \ell\nu\ell\nu$ decay mode with 4.7 fb^{-1} of ATLAS data at $\sqrt{s} = 7$ TeV, ATLAS-CONF-2012-012, CERN, March 2012

[134] Aad, G et al, Jet energy measurement with the aTLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV, CERN-PH-EP-2011-191, CERN Note, 2011

[135] Aad, G et al, Eur. Phys. J. C **72** (2012) 1844

[136] The ATLAS Collaboration, Higgs Boson Searches using the $H \to WW^{(*)} \to \ell\nu\ell\nu$ Decay Mode with the ATLAS Detector at 7 TeV, ATLAS-CONF-2011-005, CERN, March 2011

[137] Mellado, B; Ruan, X; Zhang, Z, Phys. Rev. **D84** (2011)

[138] The ATLAS Collaboration, Prospects for Higgs boson searches using the $H \to WW^{(*)} \to \ell\nu\ell\nu$ decay mode with the ATLAS detector at $\sqrt{s} = 10$ TeV, ATL-COM-PHYS-2010-185, CERN, June 2010

[139] Read, A, Modified Frequentist Analysis of Search Results (The CL_s Method, CERN-2000-005, CERN, 2000

[140] Read, A, J. Phys. G **28** (2002) 2693–2704

[141] Cowen, G; Cranmer, K; Gross, E; Vitells, O, Eur. Phys. J. C **71** (2011) 1554

[142] The ATLAS Collaboration, Search for the Standard Model Higgs boson in the diphoton decay channel with 4.9 fb^{-1}of ATLAS data at $\sqrt{s} = 7$ TeV, ATLAS-CONF-2011-161, CERN, 2011

[143] The ATLAS Collaboration, Search for the Standard Model Higgs boson in the decay channel $H \to ZZ^{(*)} \to 4\ell$, ATLAS-CONF-2011-162, CERN, 2011