

Search for a fermiophobic charged Higgs boson in proton-proton collisions with the ATLAS detector

Zhelun Li

Department of Physics

McGill University, Montreal

October 2023

A thesis submitted to McGill University in partial fulfillment of the
requirements of the degree of

Doctor of Philosophy

©Zhelun Li, 2023

Dedication

This thesis is dedicated to my grandparents: **Shutang Li** and **Xiulan Zhou**.

I further dedicate this thesis to my fellow countrymen. May you live in a **democratic, free, and prosperous** society.

Abstract

During the run-2 phase of the Large Hadron Collider, the ATLAS detector has recorded an integrated luminosity of 140.1 fb^{-1} of proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$. The new run conditions have improved the sensitivity of the search for many beyond-the-Standard-Model processes, including the production of charged Higgs bosons. This thesis presents the search for low-mass charged Higgs bosons, which would have invariant masses between 110 GeV and 200 GeV, and could be decaying into $W(\ell\nu)\gamma$ final state particles. The signal process of charged Higgs production is simulated using the Georgi-Machacek model. Standard Model background processes are estimated using a combination of Monte Carlo simulations and data-driven methods. The W -mass constraint is used to infer the neutrino's four-momentum, which is a crucial component in the determination of the invariant mass of the charged Higgs boson. Ultimately, the upper limits of charged Higgs production cross section times the $W(\ell)\gamma$ branching ratio, $\sigma \times \text{BR}$, are presented separately for different Higgs mass points.

Abrégé

Lors de la phase 2 du Grand Collisionneur de Hadrons, le détecteur ATLAS a enregistré une luminosité intégrée de $140,1 \text{ fb}^{-1}$ de collisions proton-proton à une énergie dans le centre de masse de $\sqrt{s} = 13\text{TeV}$. Les nouvelles conditions de prises de données ont amélioré la sensibilité de la recherche de nombreux processus d'au-delà du modèle standard, y compris la production de bosons de Higgs chargés. Cette thèse présente la recherche de bosons de Higgs chargés de faible masse, qui auraient des masses invariantes comprises entre 110 GeV et 200 GeV, et se désintégreraient en particules d'état final $W(\ell\nu)\gamma$. Le processus de signal de la production de Higgs chargé est simulé à l'aide du modèle Georgi-Machacek. Les processus de bruit de fond du Modèle Standard sont estimés à l'aide d'une combinaison de simulations Monte Carlo et de méthodes basées sur les données elles-mêmes. La contrainte de masse W est utilisée pour déduire le quadrivecteur d'impulsion du neutrino, qui est un élément crucial dans la détermination de la masse invariante du boson de Higgs chargé. En fin de compte, les limites supérieures de la section efficace de production de Higgs chargé multipliée par le rapport de branchement $W(\ell)\gamma$, $\sigma \times \text{BR}$, sont présentées séparément pour différents points de masse du Higgs.

Acknowledgements

The list of people I am indebted to for this thesis could not be easily summarized in just one small section. The following paragraphs are my attempts to include all mentors, colleagues, and acquaintances during my Ph.D. program. However, there are so many more names that I must have left out. Please forgive me in that case.

First and foremost, I would like to thank my supervisors: Professor Francois Corriveau and Professor Andreas Warburton, for giving me this fantastic opportunity to study at McGill, a wonderful place to conduct academic research. Throughout my Ph.D. program, they have offered a great deal of support in research, personal life, and career development. It would be an understatement to say that I would not have been the person I am today without them.

I wish to thank all of my friends in Montreal. My friends in Montreal are a mélange of excellently cultivated youngsters with various characteristics: Chloe YQ Guo, Caden Meng Cao, Yezhou Wang, Ai Liu, Chenyi Zhou, Xinyu Pan, Cheryl Huiyi Wang, and many more. I hope they will one day understand how much I wish to spend more time with all of them together. Montreal will never be the same for me without them.

My colleagues at McGill are simply the best co-workers one could ever imagine. I wish to express my deepest gratitude to John McGowan, to whom I will be forever indebted. The course of my Ph.D. program is challenging, especially during the COVID era. I simply cannot imagine what it would have been like if John was not there. Thank you for

the numerous helpful discussions, basketball games, and also hearing me whining about my life for four straight years. I would also like to thank Professor Brigitte Vachon, Dr. Shreya Saha, Alessandro Ambler, Dr. Xingguo Li, and Dr. Raphael Hulsken for their help and support.

I would like to take this opportunity to thank Justin Tianyu Yang, Tingyu Zhang, Dr. Miaoran Lu, Dr. Yousen Zhang, and Shuhui Huang for their companionship during my stay at CERN. The one-year stay in Europe would not have been one of the best years of my life without them. I would also like to thank members of the ATLAS Japan team: Haurka Asada, Aoto Tanaka, Jiaqi Zang, Marin Furukawa, Yuichiro Hayashi, and many more wonderful people. Thank you for your kindest support.

I also want to thank two magnificent researchers: Dr. Chase Shimmin and Professor Ben Carlson. They have been encouraging and helpful in my exploration of machine learning. Their assistance, often in great detail, is essential to my development as a scientist.

Lastly, I would like to thank my family. It is obvious that I would not have been able to finish this thesis without them.

Contribution to Original Knowledge

The analysis presented in this thesis is the first search for charged Higgs bosons with masses between 110 GeV and 200 GeV using the Large Hadron Collider. Given the large mass of the Higgs boson, its productions are readily feasible in the Large Hadron Collider, which offers a collision energy that is an order of magnitude higher than that of other colliders. Traditionally, searches for charged Higgs bosons focus on the phase space where the Higgs mass is above 200 GeV, close to the WZ threshold. Given the fermiophobic nature of most charged Higgs models, the decay channel to WZ is the most prominent. This thesis searches for masses below 200 GeV, suppressing the WZ decay channel and favoring the $W\gamma$ channel. This analysis is the first to target the $W\gamma$ final state particles in order to give stringent constraints on the cross-section limits for charged Higgs boson masses below 200 GeV.

In the reconstruction of the invariant mass of the charged Higgs boson, this thesis also develops a novel method in the estimation of the neutrino particle's four-momentum, which uses a constraint fit to overcome the hurdle of unphysical initial parameters due to measuring uncertainties.

Contribution of Authors

This thesis presents the search for charged Higgs bosons in the ATLAS experiment, a large collaboration of more than 5000 members, including students, scientists, and technicians from 182 institutes. Since the field of high energy physics requires many experts from different backgrounds, a large portion of the ATLAS members spent a great majority of their contributions to hardware and software developments, alongside their scientific research conducted using data collected by the ATLAS detector. It is common practice that a paper published by the ATLAS collaboration must address all members in the author list as authors to give credit to their work as crucial pieces of the larger project.

The analysis part described in this thesis is carried out by the two teams at McGill University and Duke University. Two Ph.D. candidates are in charge of the main developments of the analysis: Zhelun Li, supervised by Professor Francois Corriveau and Professor Andreas Warburton at McGill, and Utsav Patel, supervised by Professor Mark Kruse at Duke. The contribution of authors in each chapter is listed in the following.

- Chapter 1: The author gives an introductory description of the analysis.
- Chapter 2: The author goes through the theoretical background of various physical theories used in this analysis.
- Chapter 3: The author summarizes different ATLAS detector components, built upon the collective effort of many scientific researchers and engineers in the AT-

LAS collaboration. The author has directly contributed to the upgrade of the muon detector described in Section 3.2.3 as a member of the McGill ATLAS group.

- Chapter 4: The author describes the reconstruction and selection of all physics objects used in this analysis. The reconstruction algorithm of each particle type is developed by its working group of many researchers in the ATLAS collaboration. The author had directly contributed to developing the b-tagging algorithm described in Section 4.5.2 and the missing transverse momentum reconstruction described in Section 4.7. The author also worked on the neutrino reconstruction algorithm discussed in Section 4.8.
- Chapter 5: The author gives the selection of events used to separate the Standard Model background events from signal process events. Utsav Patel has contributed to the design of analysis regions described in Section 5.2.
- Chapter 6: The author describes the estimation of background processes. In particular, the author contributed to the development of data-driven estimation of processes involving misidentified particles.
- Chapter 7: The author gives the uncertainties used in this analysis. The author is responsible for combining theoretical uncertainties and determining systematic uncertainties for the data-driven methods.
- Chapter 8: The author describes the results of the search. The author contributed to the generation of nominal and systematically varied histograms used in the fitting procedures. The author also collaborated with Utsav Patel to develop the fitting strategies using the TRexFitter framework.
- Chapter 9: The author concludes the analysis discussed in this thesis with a discussion on future research directions.

Table of Contents

Dedication	i
Abstract	ii
Abrégé	iii
Acknowledgements	iv
Contribution to Original Knowledge	vi
Contribution of Authors	vii
List of Figures	xviii
List of Tables	xxii
1 Introduction	1
2 Theoretical background	5
2.1 The Standard Model	5
2.1.1 Fermions	5
2.1.2 Gauge Bosons	8

2.1.3	Lagrangian Without The Higgs Boson	9
2.1.4	The Higgs Mechanism	13
2.2	The Georgi-Machacek Model	18
2.2.1	The Georgi-Machacek Lagrangian	18
2.2.2	H_5^+ Decay Channels	23
3	Experimental Setup	28
3.1	The Large Hadron Collider	28
3.1.1	Beam Preparation	30
3.1.2	Luminosity and Pileup	33
3.2	The ATLAS detector	36
3.2.1	Inner Detector	40
3.2.2	Calorimeter	43
3.2.3	Muon Spectrometer	49
3.2.4	Trigger	51
4	Object Reconstruction	54
4.1	Tracks and Clusters	55
4.1.1	Tracks	55
4.1.2	Vertices	57
4.1.3	Calorimeter Clusters	58

4.2	Electrons	59
4.3	Photons	63
4.4	Muon	64
4.5	Jets	68
4.5.1	Jet Reconstruction	68
4.5.2	B-tagging	71
4.6	Overlap Removal	72
4.7	Missing Transverse Momentum	72
4.8	Neutrino	75
4.8.1	W -mass Constraint	76
4.8.2	Complex Solutions	76
4.8.3	Higgs Mass	80
5	Event Selection	83
5.1	Datasets	83
5.1.1	Data	84
5.1.2	Monte Carlo Simulation	85
5.1.3	Triggers and Derivations	88
5.2	Event Selections and Analysis Regions	91
5.2.1	Inclusive Event Selection	91

5.2.2	Regions	93
5.2.3	Summary	99
6	Background Estimation	101
6.1	Monte Carlo sample overlap removal	101
6.2	Electrons Faking Photons	102
6.3	Jets Faking Photons	108
7	Systematic Uncertainty	115
7.1	Detector-related Systematics	115
7.2	Theoretical Uncertainties	118
7.3	Systematics of Data-driven Estimations	119
7.3.1	Electrons faking photons	120
7.3.2	Jets Faking Photons	120
8	Fitting methods and results	122
8.1	Fitting	122
8.1.1	Profile Likelihood Function	122
8.1.2	Fitting Strategy	125
8.2	Fitting Results with Asimov Datasets	125
9	Conclusion	139

Bibliography	154
Glossary	158
A Detector Resolution	159
B Good Run List	160
C Neutrino Reconstruction	161
D Event Selection	164
E Systematic uncertainties	166
E.1 Detector-related Systematics	166
E.2 Data-driven methods	170
E.3 Signal region	171
E.4 Validation region	172
E.5 W +jets control region	173
E.6 $W\gamma$ control region	174
E.7 Z +jets control region	175
E.8 $Z\gamma$ control region	176
F Fitting	177

List of Figures

2.1	A diagram of fundamental particles in the Standard Model of particle physics. Fermion particles that form matter are listed in the three columns on the left with the two families of quarks and leptons. On the right-hand side, two types of bosons are presented: the gauge boson, which carries fundamental forces, and the scalar Higgs boson [11].	6
2.2	A visual display of the Higgs potential. Note that the local maximum at $\phi = 0$ is not stable. States at the bottom of the potential energy well do not respect the symmetry embedded in the shape of the potential [16].	15
2.3	Loop induced Feynman diagram of $HW\gamma$ production [24].	24
2.4	Dependence of the branching ratio of $H_5^\pm \rightarrow W^\pm\gamma$ on m_5 and s_H , for $M_2 = 40$ GeV [20].	26
2.5	Feynman diagrams for the dominant Drell-Yan production processes involving H_5^+ [20].	27
3.1	The CERN accelerator complex [29]. The four main detectors, ATLAS, CMS, ALICE, and LHCb, are labeled by the yellow dots on the LHC ring. .	29
3.2	The layout of the LHC ring and various experiments [30].	31
3.3	Cross-section of cryodipole [31].	32

3.4	Cumulative luminosity versus day delivered to ATLAS during stable beams for high energy p-p collisions [36].	35
3.5	The Run-2 pileup distribution [38].	37
3.6	A diagram of the ATLAS detector with its subsystems [39].	38
3.7	A diagram of the ATLAS inner detector with its subsystems [43].	41
3.8	The inner detector of the ATLAS experiment shown in the R-z plane [45]. The vertical axis marks the radial distance orthogonal to the beam axis. . .	42
3.9	Interactions between incoming particles and different detector components of ATLAS [48].	44
3.10	The ATLAS calorimeter systems with its submodules: the electromagnetic calorimeter and the hadronic calorimeter [25].	45
3.11	Cutaway views of the LAr electromagnetic calorimeter and the hadronic tile calorimeter tower are shown on the left and right-hand side, respectively [49].	48
3.12	The ATLAS muon spectrometer [25].	50
3.13	The ATLAS trigger and data acquisition system [59].	52
4.1	Shapes of the LAr calorimeter pulse [67].	60
4.2	A schematic diagram of an electron's path through the ATLAS detector [68]. The solid red line represents the electron trajectory through the inner detector to the calorimeter. The dashed red line represents a photon produced from interactions with the detecting material.	61
4.3	Measured electron reconstruction efficiencies with respect to E_T in $Z \rightarrow ee$ events [68].	62

4.4	An example of the anti- k_t algorithm reconstructing jets at the parton-level [72].	70
4.5	Resolution of E_T^{miss} measured by RMS width as a function of truth E_T^{miss} in $W \rightarrow e\nu, W \rightarrow \mu\nu$, and $t\bar{t}$ events [80].	75
4.6	Comparison of neutrino z-momentum p_z^ν estimates in events with complex solutions in MC simulation of signal processes. The x-label is the difference between the truth momentum, p_z^ν , and the estimated momentum, \hat{p}_z^ν	79
4.7	Reconstructed Higgs mass from $pp \rightarrow H_5^\pm H_5^0$ and $pp \rightarrow H_5^\pm H_5^\mp$ processes with the fiveplet Higgs mass at 150 GeV.	81
5.1	The workflow of data and MC processing up to the analysis level.	84
5.2	Selections implemented in all regions in this analysis.	100
6.1	Fitting crystal ball (signal) and polynomial (background) functions to the invariant mass histogram of electron and photon. The data is obtained from the bin of $p_T > 27$ GeV and $ \eta < 0.6$ with the full run-2 data.	105
6.2	Fitting crystal ball (signal) and polynomial (background) functions to the invariant mass histogram of two electrons. The data is obtained from the bin of $p_T > 27$ GeV and $ \eta < 0.6$ with the full run-2 data.	105
6.3	Fake rates with respect to p_T and $ \eta $ in MC simulations. The last p_T bin is partially cut for better display.	106
6.4	Fake rates with respect to p_T and $ \eta $ in the full run-2 data. The last p_T bin is partially cut for better display.	107
6.5	Fake rate ratio between data and MC with respect to p_T and $ \eta $ in the full run-2 Data. The ratio is defined to be $\frac{FR_{data}}{FR_{MC}}$. The last p_T bin is partially cut for better display.	108

6.6	Isolation energy distribution of real photons in the bin of $160 \text{ GeV} < m^{\ell+\gamma+\text{MET}} < 180 \text{ GeV}$ of the $W\gamma$ control region.	110
6.7	Isolation energy distribution of real and fake photons in the bin of $160 \text{ GeV} < m^{\ell+\gamma+\text{MET}} < 180 \text{ GeV}$ of the $W\gamma$ control region.	111
6.8	The final fit in the bin of $160 \text{ GeV} < m^{\ell+\gamma+\text{MET}} < 180 \text{ GeV}$ of the $W\gamma$ control region. The cut at 2.45 GeV results from the tight isolation cut required by the tight photon definition as introduced in Section 4.3.	111
6.9	Comparison of the truth information and template fit estimate in the W +jets control region with the pseudo-dataset generated from $W\gamma$ and W +jets sample.	113
6.10	Comparison of the truth information and template fit estimate in the $W\gamma$ control region with the pseudo-dataset generated from $W\gamma$ and W +jets sample.	114
8.1	Pruning plot of systematics used in the fit for the $m_{H_5} = 150 \text{ GeV}$ mass point.	128
8.2	Nuisance parameters after the fit for the $m_{H_5} = 150 \text{ GeV}$ mass point.	129
8.3	Ranking plot of all nuisance parameters after the fit for the $m_{H_5} = 150 \text{ GeV}$ mass point.	130
8.4	Normalisation Factors of various processes' cross sections after the fit.	131
8.5	Correlation Matrix of all systematics after the pruning procedure.	132
8.6	Plots of the $W\gamma$ control region before and after the fit.	133
8.7	Plots of the $Z\gamma$ control region before and after the fit.	133
8.8	Plots of the W +jets control region before and after the fit.	134

8.9	Plots of the Z +jets control region before and after the fit.	134
8.10	Plots of the validation region before and after the fit.	135
8.11	Plots of the signal region before and after the fit.	135
8.12	Expected limit of the signal charged Higgs process within the mass range between 110 GeV and 200 GeV.	137
C.1	Comparison of charged Higgs boson mass estimation between different algorithms in the $H_5^\pm H_5^0$ samples with $m_{H_5} = 150$ GeV.	162
C.2	Comparison of charged Higgs boson mass estimation between different algorithms in the $H_5^\pm H_5^\mp$ samples with $m_{H_5} = 150$ GeV.	163

List of Tables

2.1	Mass eigenstates of Georgi-Machacek model's Higgs sector.	21
2.2	Couplings of the fiveplet states H_5 with one or two gauge bosons. Notice that all couplings with two gauge bosons are proportional to s_H	23
2.3	Decay channels for members of the scalar fiveplet at low mass, including possible off-shell decays [20].	24
3.1	Integrated luminosity in run-2 with uncertainties [35].	35
4.1	Cluster sizes for different particles in the EM calorimeter [65]. Converted photons are those that go through pair production before reaching the calorimeter.	58
4.2	Selection criteria of objects contributing to E_T^{miss} [80].	74
5.1	Summary of the $H_5^\pm H_5^0$ signal process event generation	87
5.2	Summary of the $H_5^\pm H_5^{\mp\mp}$ signal process event generation	87
5.3	Summary of the $H_5^+ H_5^-$ signal process event generation	87

5.4	Summary of the prompt background processes event generation. The $W\gamma$ and $Z\gamma$ background samples in this table contain events in which the photon is produced through QCD processes. The $EWKW\gamma$ and $EWKZ\gamma$ samples are those where the photon was produced through electroweak interactions.	89
5.5	Summary of the non-prompt background processes event generation.	89
5.6	List of triggers used in each year of the run-2 dataset.	90
5.7	Cutflow for various processes leading to the construction of the signal region.	94
5.8	Cutflow for various processes leading to the construction of the validation region.	95
5.9	Cutflow for various processes leading to the construction of the $W\gamma$ -dominant control region.	95
5.10	Cutflow for various processes leading to the construction of the W + jets-dominant control region.	96
5.11	Cutflow for various processes leading to the construction of the Z + jets-dominant control region.	97
5.12	Cutflow for various processes leading to the $Z\gamma$ -dominated Control Region.	98
5.13	Summary of cuts for each region	100
6.1	Overlap regions of various MC samples.	102
7.1	A table of the PDF sets used in the MC sample with alternative PDF sets.	118
A.1	Energy resolutions and η coverages of various components of the ATLAS detector.	159

D.1	Initial cutflows for various processes before constructing each signal, validation, and control region. This is for those regions with a final state of one lepton and at least one photon. The two tables are separated due to aesthetic reasons. The first cut in the lower table is the one placed after the last cut in the upper table.	165
E.1	Systematic uncertainties of electrons faking photons in each region. The uncertainties are expressed in percentages of the nominal values.	170
E.2	Systematic uncertainties of jets faking photons in each region. The uncertainties are expressed in percentages of the nominal values.	170
E.3	Relative effect of each nuisance parameter on the yields in the signal region before the fit. The signal sample used for this table has a mass of 150 GeV.	171
E.4	Relative effect of each systematic on the yields in the validation region before the fit. The signal sample used for this table has a mass of 150 GeV.	172
E.5	Relative effect of each systematic on the yields in the W +jets control region before the fit. The signal sample used for this table has a mass of 150 GeV.	173
E.6	Relative effect of each systematic on the yields in the $W\gamma$ control region. The signal sample used for this table has a mass of 150 GeV.	174
E.7	Relative effect of each systematic on the yields in the Z +jets control region before the fit. The signal sample used for this table has a mass of 150 GeV.	175
E.8	Relative effect of each systematic on the yields in the $Z\gamma$ control region. The signal sample used for this table has a mass of 150 GeV.	176
F.1	Yields of the analysis in all control regions before the fit.	177

F.2	Yields of the analysis in the validation region and the signal region before the fit.	178
F.3	Yields of the analysis in all control regions after the fit.	178
F.4	Yields of the analysis in the validation region and the signal region after the fit.	178

Chapter 1

Introduction

The Standard Model (SM) of particle physics is a theory that describes the fundamental building blocks of matter and their interactions. The theory was first proposed in the 1970s and has successfully described a wide range of physics phenomena at the subatomic level. After decades of experimental effort, the Standard Model is now considered one of the most well-tested theories in the world.

In the family of fundamental particles in the Standard Model, the Higgs boson plays the vital role of giving mass to all other particles with non-zero mass. Following the Standard Model's prediction, the Higgs boson was discovered in 2012 [1], which provides a key piece of the puzzle in completing the Standard Model.

The discovery of a scalar Higgs boson, as predicted by the Standard Model, has raised many discussions on the possibility of having a more complex Higgs sector. Beyond Standard Model (BSM) theories often propose a Higgs sector that contains many Higgs bosons, which are siblings to the existing one in the Standard Model. There have been searches for a charged Higgs boson in the ATLAS collaboration in the past few years. A search in the WZ channel using the Georgi-Machacek (GM) model [2] as the benchmark model reported no evidence of a charged Higgs boson with a mass between 200 GeV and

1000 GeV using run-1 data with a center of mass energy $\sqrt{s} = 8$ TeV and an integrated luminosity of $20.3 \pm 0.6 \text{ fb}^{-1}$ [3]. Another charged Higgs boson search in the ATLAS collaboration guided by the type-II seesaw mode [4] uses the same run-2 data as this search in the thesis with $\sqrt{s} = 13$ TeV and an integrated luminosity 140.1 fb^{-1} . It excludes doubly-charged Higgs bosons at 95% confidence level with mass limits at 350 GeV for the pair production mode and 230 GeV for the associated production mode in channels of $W^\pm W^\pm$ and $W^\pm Z$ [5]. These results in the ATLAS collaboration could be regarded as complementary to the search in this thesis as they used similar datasets to probe different phase spaces and decay channels.

The search for charged Higgs bosons is also performed in many other collaborations around the world since certain models could prefer decay channels that are also sensitive to other experiments. These results used quite different datasets and probed channels that are often distinct from those in the ATLAS searches. In 2013, collaborations at the Large Electron-Positron (LEP) Collider reported that the charged Higgs in doublet extensions were excluded below 80 GeV at 95% confidence level [6]. The LEP search was performed with a center of mass energy from 183 GeV to 209 GeV and an integrated luminosity of 2.6 fb^{-1} . The Higgs boson in the LEP searches were taken to be pair produced via coupling to the neutral Z boson and then subsequently decayed into a wide variety of final state particles such as $c\bar{s}c\bar{s}$, $\tau\nu\tau\nu$, etc. The BaBar experiment also performed charged Higgs searches by studying B -meson decays that are sensitive to type II two-Higgs-doublet model [7]. The full BaBar dataset with a center of mass energy of 10.58 GeV and an integrated luminosity of 426 fb^{-1} with 471 million $B\bar{B}$ events were used to determine the decay ratios of $\mathcal{R}(D) = \mathcal{B}(B \rightarrow D\tau^-\bar{\nu}_\tau)/\mathcal{B}(B \rightarrow D\ell^-\bar{\nu}_\ell)$ and $\mathcal{R}(D^*) = \mathcal{B}(B \rightarrow D^*\tau^-\bar{\nu}_\tau)/\mathcal{B}(B \rightarrow D^*\ell^-\bar{\nu}_\ell)$ [8]. This search considered the ratio of $\tan\beta/m_H$, where $\tan\beta$ is the ratio of vacuum expectation energies of the two Higgs doublets and m_H is the mass of the Higgs, in the range from 0.05 GeV^{-1} to 1 GeV^{-1} . The doublet model was eventually excluded at 99.8% confidence levels when both $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ channels were combined. The CDF collaboration had also conducted a search of the charged Higgs

boson through the channel of $t \rightarrow H^+b$ using its dataset of 193 fb^{-1} at $\sqrt{s} = 1.96 \text{ TeV}$. This search reported no evidence of charged Higgs but set a 95% confidence level upper limit of $\mathcal{BR}(t \rightarrow H^+b)$ around 0.8 for charged Higgs masses between 80 GeV and 160 GeV [9]. The $D\emptyset$ experiment searched for the charged Higgs in mass ranges between 180 GeV and 300 GeV using the dataset with 0.9 fb^{-1} at $\sqrt{s} = 1.96 \text{ TeV}$ [10]. It reported a 95% confidence level exclusion of Higgs mass above 180 GeV with $\tan\beta$ above 30. The H1 experiment also conducted a search of doubly charged Higgs bosons using 118 pb^{-1} of ep collision data at $\sqrt{s} = 300 \text{ GeV}$ between 1994 – 1997 and $\sqrt{s} = 318 \text{ GeV}$ between 1998 – 2000. The lower limit of the charged Higgs mass was determined to be 141 GeV at 95% confidence level [10].

This analysis uses the proton-proton collisions with an integrated luminosity of 140.1 fb^{-1} in the ATLAS detector within the Large Hadron Collider (LHC) at a center-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$. Monte Carlo (MC) samples are used to model the Standard Model backgrounds in the analysis. These simulated samples consist of proton collision events from various Standard Model processes that would yield final state particles consisting of a lepton, a photon, and a neutrino. Some processes, like those with misidentified final state photons, are hard to model using simulation and hence require data-driven methods to re-calibrate or entirely replace the MC samples. Systematic variations of various sources of uncertainties are also applied to both MC simulations and data-driven methods to estimate the total uncertainty of the backgrounds. The limit of the charged Higgs boson production cross-section is then computed by a likelihood fit. As required by the ATLAS collaboration, the analysis in this thesis is blinded at the time of writing. The blinding procedure prevents researchers from looking at the real data in the signal region where the final results are obtained. The results are therefore obtained with independently generated pseudo-data as discussed in Section 5.2.2.

The structure of this thesis is given below: Chapter 2 gives the theoretical details of the Standard Model and the GM model. The experiment setup of the Large Hadron Col-

lender and the ATLAS (A Toroidal LHC Apparatus) detector is then described in Chapter 3. Next, chapter 4 summarizes the algorithms that reconstruct physics objects from the signals collected by the ATLAS detector. Then, the event selection using reconstructed objects is discussed in Chapter 5. Next, Chapter 6 deals with the technical treatment of the background estimation using MC samples and data-driven methods. The systematic uncertainties associated with the background estimations are then discussed in Chapter 7. Finally, Chapter 8 presents the results of the charged Higgs boson's cross-section limits. The thesis ends with Chapter 9, which summarizes the analysis.

Chapter 2

Theoretical background

This chapter presents the theoretical models of particle physics. First, the Standard Model (SM) of particle physics is given in Section 2.1. The Standard Model is broken into three separate subsections: Section 2.1.1 introduces fermions which are building blocks of matter; Section 2.1.2 covers gauge bosons which are force carriers that are responsible for fundamental interactions between particles. Section 2.1.3 gives a theoretical description of the Standard Model particles without the Higgs boson. Then the Higgs boson and its interactions with other particles are introduced in Section 2.1.4. In Section 2.2, the Georgi-Machacek Model is presented. Section 2.2 discusses the motivation of searching in the $W\gamma$ decay mode from a theoretical perspective.

2.1 The Standard Model

2.1.1 Fermions

The Standard Model has 12 spin- $\frac{1}{2}$ fermions, each with its antiparticle. Some of these fermions could couple with each other to form composite particles like baryons and

Standard Model of Elementary Particles

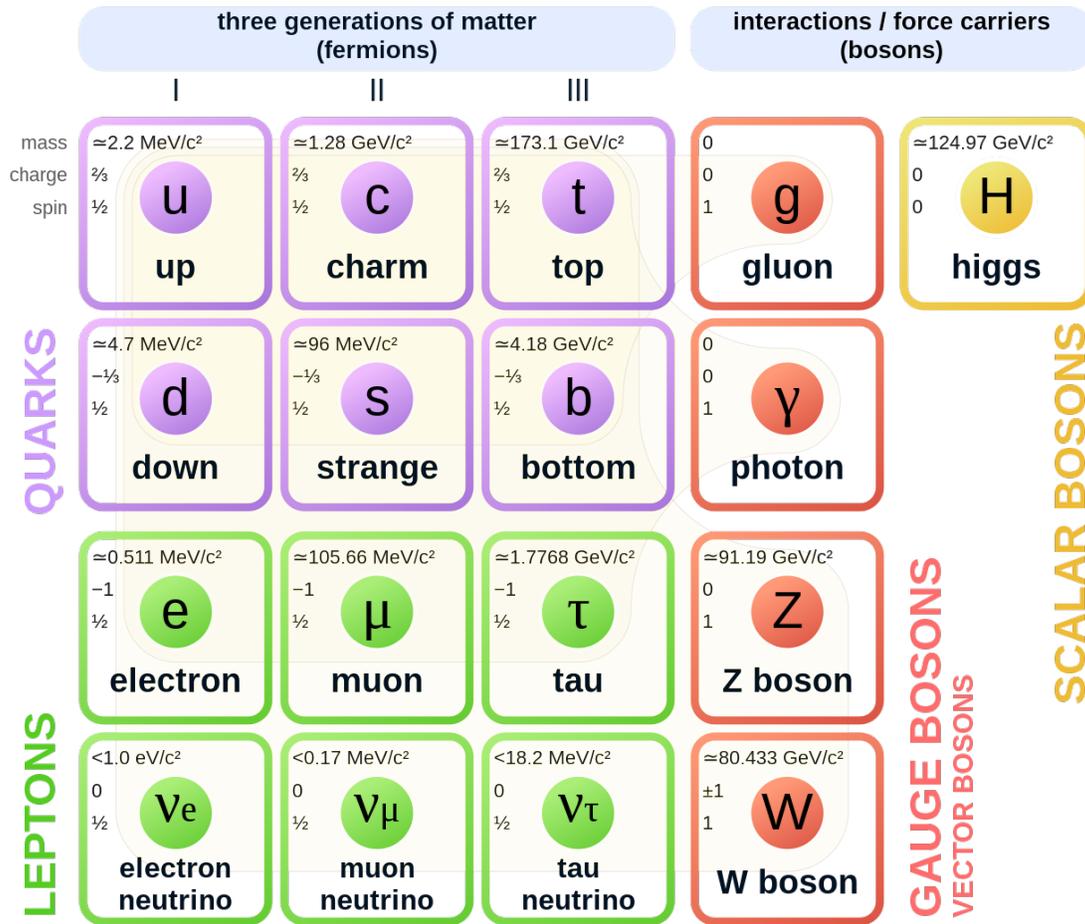


Figure 2.1: A diagram of fundamental particles in the Standard Model of particle physics. Fermion particles that form matter are listed in the three columns on the left with the two families of quarks and leptons. On the right-hand side, two types of bosons are presented: the gauge boson, which carries fundamental forces, and the scalar Higgs boson [11].

mesons, which are the building blocks of ordinary visible matter that is observed in the universe. The elementary fermions could be classified into two categories: quarks and leptons.

Quarks

The six flavors of quarks in the fermion family are *up* (u), *down* (d), *charm* (c), *strange* (s), *top* (t), and *bottom* (b), as shown in the top left of Figure 2.1. These quarks all have electric charges of $+\frac{2}{3}e$ or $-\frac{1}{3}e$, while their corresponding antiparticles have exactly the opposite of those charges. Other than the electric charge, quarks also possess color charges which allow them to interact through the strong force. Due to the phenomenon of color confinement, color-charged particles cannot be isolated. Therefore, the existence of quarks can only be inferred from composite particles, which are formed by having quarks combined through the attraction of the strong force. These composite particles consisting of quarks are called hadrons, which are further divided into baryons and mesons; these are made of odd and even numbers of quarks, respectively.

Two of the most stable hadrons are the protons and neutrons, forming the atomic nucleus when binding. Both protons and neutrons are baryons, each made of three quarks: uud for the proton and udd for the neutron. Unlike baryons, mesons, consisting of a quark and an antiquark, are short-lived particles only found in high-energy collisions of particles. In the field of High-energy physics, it is convenient to analyze hadrons using the parton model proposed by Richard Feynman. The parton model states that the hadrons could be viewed as a collection of point-like sub-particles named partons. The Parton Distribution Functions (PDF) [12] gives the probability of finding a given type of particle within the hadron that carries a certain momentum fraction. As shown in later chapters, the PDF is essential in the simulations of proton-proton collisions and is considered one of the significant sources of theoretical systematic uncertainties.

Leptons

Unlike quarks, leptons are elementary spin- $\frac{1}{2}$ particles that do not interact through the strong force. There are three generations of leptons, each with two particles: a charged

one with $-1e$ charge and its corresponding neutral particle, as shown in the left bottom corner in Figure 2.1. The three generations are of ascending order of the mass of their charged leptons.

The electron is the lightest particle among the three charged leptons, with a mass of only 0.511 MeV. Since there are no charged leptons with lower masses for electrons to decay into, the electron is the most stable lepton, which could be widely found in nature. It is known to be an essential constituent of atoms which consist of electrons trapped in an orbit around the nuclei. Heavier charged leptons are less stable as they quickly decay into electrons and neutrinos. The muon has a mean lifetime of $2.196 \cdot 10^{-6}$ s, sufficient for measurements in most modern-day collision experiments. However, due to the large mass, muons with moderate momentum are considered to be minimum ionizing particles (MIP) that barely interact with detecting material as they pass through. The elusive nature of MIP has imposed significant challenges in its measurement. The details of the ATLAS apparatus for muon measurement are introduced in Chapter 3.2.

2.1.2 Gauge Bosons

In the Standard Model, gauge bosons carry fundamental forces between interactions of all other particles. Photons are massless particles mediating electromagnetic interactions. Their lack of mass requires them to travel at exactly the speed of light, $2.99792458 \cdot 10^8$ m/s, in a vacuum. The W^+ , W^- , and the Z^0 are three massive spin-1 particles carrying the weak force while the massless spin-1 gluons mediate the strong force.

The electromagnetic interaction describes the attraction and repulsion of particles with electric charge. One example of the electromagnetic force at the atomic level would be the mutual attraction between the nucleus and the electron. The electromagnetic force could be viewed as the exchange of photons as mediators between the two attracting particles.

The strong interaction mediated by gluons is responsible for the force between color-charge-carrying particles. Unlike photons that mediate electromagnetic interactions without electric charges, gluons carry color charges, allowing them to also couple between themselves.

The strength of the weak interaction is much smaller than that of other interactions by orders of magnitude. All fermions in the Standard Model could couple to the W^\pm and Z^0 bosons to participate in the weak interaction, making it possible for fermions to change flavors in the process. The W^\pm boson has a mass of 80.379 GeV, while that of the Z^0 boson is 91.1876 GeV [13]. Their masses are acquired from the spontaneous breaking of the SU(2) gauge symmetry, which will be introduced in Section 2.1.4.

2.1.3 Lagrangian Without The Higgs Boson

The Lagrangian of all Standard Model particles mentioned above could be formulated by combining the Lagrangians of individual sectors. It is worth noting that the Higgs boson plays a unique role of giving mass to other particles in the Standard Model. Therefore, the Lagrangian of the Higgs sector is introduced on its own in Section 2.1.4 as part of the Higgs Mechanism.

Quantum Electrodynamics(QED)

In the Standard Model, the bispinor field for a spin- $\frac{1}{2}$ particle is denoted by ψ , while its Dirac adjoint is denoted by $\bar{\psi}$. In 1928, Paul Dirac proposed the Dirac equation, which gives rise to the following Lagrangian for a quantum theory of electrons described by the spinor ψ [14]:

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\psi\bar{\psi}. \quad (2.1)$$

The first term on the equation's right-hand side gives the spinor's kinetic term using the Dirac matrices γ^μ . The second term gives the mass term of the field with mass m .

To include the interactions in quantum electrodynamics, one must also introduce the Lagrangian for the gauge field of electromagnetism. The gauge field tensor of electromagnetism is given by $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, where A_μ is the electromagnetic four-potential. Adding the electromagnetic term to the Lagrangian in Equation 2.1 gives the full quantum electrodynamics Lagrangian:

$$\mathcal{L}_{QED} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + i\bar{\psi}\gamma^\mu D_\mu\psi - m\psi\bar{\psi}, \quad (2.2)$$

where the derivative in the second term in Equation 2.2 is now a covariant derivative given by the electromagnetic gauge field tensor A_μ and the external field B_μ :

$$D_\mu = \partial_\mu + ieA_\mu + ieB_\mu.$$

Quantum Chromodynamics

The quantum chromodynamics (QCD) Lagrangian consists of the quark field, a fundamental representation of the SU(3) gauge group. To include gluon interactions, the gluon field strength Tensor $G_{\mu\nu}^a$ is introduced in analogy to the electromagnetic field strength tensor $F_{\mu\nu}$:

$$G_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f_{bc}^a A_\mu^b A_\nu^c, \quad (2.3)$$

where f_{bc}^a are the structure constants of SU(3).

The complete quantum chromodynamics Lagrangian is then given by

$$\mathcal{L}_{QCD} = -\frac{1}{4}G_a^{\mu\nu}G_{\mu\nu}^a + \bar{\psi}_i(i\gamma^\mu D_\mu^{ij} - m\delta^{ij})\psi_j. \quad (2.4)$$

The gauge covariant derivative is defined as: $(D_\mu)_{ij} = \partial_\mu\delta_{ij} - ig(T_a)_{ij}A_\mu^a$ where T_a are infinitesimal SU(3) generators.

Electroweak interactions

At high energies, electromagnetism and the weak interaction are unified in a single description called the electroweak interaction which is of $SU(2)_L \times U(1)_Y$ symmetry. The subscript letter L indicates that only left-handed fermions transform under non-trivial representations of the SU(2) group. The subscript Y differentiates the U(1) group of weak hypercharge from that of the electromagnetism, which is often denoted by $U(1)_{em}$.

The generators of the $SU(2)_L$ group are called the weak isospin(T), while those of the $U(1)_Y$ group are called weak hypercharge(Y). There are three W gauge bosons associated with $SU(2)_L$ weak isospin and one B boson of $U(1)_Y$ weak hypercharge. The charges of gauge bosons associated with electroweak interactions are determined by the third component of the weak isospin and the hypercharge: $Q = T_3 + \frac{1}{2}Y$.

The electroweak interaction Lagrangian is given in Equation 2.5:

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_g + \mathcal{L}_f + \mathcal{L}_{Higgs} + \mathcal{L}_y, \\ \mathcal{L}_g &= -\frac{1}{4}W_a^{\mu\nu}W_{\mu\nu}^a - \frac{1}{4}B^{\mu\nu}B_{\mu\nu}, \\ \mathcal{L}_f &= \sum_j \bar{\psi}_{j,L}i\gamma^\mu D_\mu\psi_{j,L} + \sum_f \bar{\psi}_{f,R}i\gamma^\mu D_\mu\psi_{f,R}. \end{aligned} \quad (2.5)$$

The first term \mathcal{L}_g describes the interaction of the W vector bosons and the B vector bosons with the field strength tensors $W^{a\mu\nu}$ and $B^{\mu\nu}$. The second term \mathcal{L}_f describes the coupling between the gauge fields with fermions. Index j in Equation 2.5 runs through all three generations of fermions, and index f runs through the nine charged fermions. The left and right-handed fermion fields are defined by: $\psi_{L,R} = \frac{1 \mp \gamma_5}{2} \psi$, and the covariant derivatives are given by:

$$\begin{aligned} D_\mu \psi_L &= (\partial_\mu + ig \sum_{A=1}^3 T_{L,R}^A W_\mu^A + ig' \frac{1}{2} Y_{L,R} B_\mu) \psi_L, \\ D_\mu \psi_R &= (\partial_\mu + ig' \frac{1}{2} Y_{L,R} B_\mu) \psi_R, \end{aligned} \quad (2.6)$$

where T and Y are generators of the $SU(2)_L$ and $U(1)_Y$ groups respectively. g and g' are the corresponding coupling constants of the $SU(2)_L$ and $U(1)_Y$ interactions. The other two terms in Equation 2.5 are involved in the Higgs mechanism, which is described in detail in Section 2.1.4.

The electroweak interaction is often parameterized by W^\pm , Z , and γ , which are given by:

$$\begin{aligned} W^\pm &= \frac{1}{\sqrt{2}} (W_1 \mp iW_2), \\ \gamma &= \cos\theta_W B + \sin\theta_W W_3, \\ Z &= -\sin\theta_W B + \cos\theta_W W_3, \end{aligned} \quad (2.7)$$

where the angle θ_W is the weak-mixing angle given by the coupling constants:

$$\begin{aligned} \cos\theta_W &= \frac{g}{\sqrt{g^2 + g'^2}}, \\ \sin\theta_W &= \frac{g'}{\sqrt{g^2 + g'^2}}. \end{aligned} \quad (2.8)$$

2.1.4 The Higgs Mechanism

The Brout–Englert–Higgs mechanism [15], sometimes called the Higgs mechanism, was developed independently in the 1960s by a group of physicists to explain the origin of mass in the Standard Model. It is clear from Section 2.1.3 that the simplest form of mass will inevitably violate the gauge symmetry in the Lagrangian. For example, a mass term of $m\psi\bar{\psi}$ for the electroweak Lagrangian will have the following property:

$$\begin{aligned}
 \bar{\psi}\psi &= \frac{1}{4}\bar{\psi}[2 + 2\gamma_5^2]\psi \\
 &= \frac{1}{4}\bar{\psi}[1 + 2\gamma_5 + \gamma_5^2 + 1 - 2\gamma_5 + \gamma_5^2]\psi \\
 &= \bar{\psi}\frac{1 + \gamma_5}{2}\frac{1 + \gamma_5}{2}\psi + \bar{\psi}\frac{1 - \gamma_5}{2}\frac{1 - \gamma_5}{2}\psi \\
 &= \bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L.
 \end{aligned} \tag{2.9}$$

This simple form of the mass term treats the left-handed part ψ_L and the right-handed part ψ_R on equal footing. However, the non-trivial $SU(2)_L$ gauge symmetry only applies to left-handed fields as shown in Equation 2.5. Therefore in practice, the right-handed field is only a trivial singlet while the left-handed field is a $SU(2)$ doublet. Then, it is evident that one could not simply add the two fields, ψ_L and ψ_R , with different transformation properties.

Spontaneous Symmetry Breaking

Consider a Lagrangian for an isospin doublet ϕ of the following form:

$$\begin{aligned}
 \phi &= \begin{pmatrix} \phi^+(x) \\ \phi^0(x) \end{pmatrix}, \\
 \mathcal{L}_{Higgs} &= (D_\mu\phi)^\dagger(D_\mu\phi) - V(\phi),
 \end{aligned} \tag{2.10}$$

where D_μ is the covariant derivative with respect to the SU(2) gauge symmetry:

$$D_\mu\phi = \left(\partial_\mu + ig \sum_{A=1}^3 T_{L,R}^A W_\mu^A + ig' \frac{1}{2} Y_{L,R} B_\mu\right)\phi, \quad (2.11)$$

and $V(\phi)$ is a gauge invariant potential given by:

$$V(\phi) = \mu^2 \phi^\dagger \phi + \frac{\lambda}{4} (\phi^\dagger \phi)^2. \quad (2.12)$$

This Lagrangian respects the SU(2) symmetry, and the minimum energy state depends on the scalar self-interaction coupling constants μ and λ in the potential. There is only one straightforward global minimum for the potential at $\phi = 0$ when $\mu > 0$. When $\mu < 0$, the shape of the potential energy is often referred to as the Mexican hat shape, which consists of a local maximum at $\phi = 0$ and a set of global minima at the vacuum expectation value (VEV) v defined by:

$$\phi^\dagger \phi = v^2 = \frac{-2\mu^2}{\lambda}. \quad (2.13)$$

A diagram of the Mexican hat shape of the Higgs potential is shown in Figure 2.2. Even though the potential satisfies the SU(2) symmetry required in the Lagrangian, the minimum energy states of this potential break the symmetry. This is an example of the physical process known as spontaneous symmetry breaking, by which a physical system in a symmetric state spontaneously ends up in an asymmetric state [17].

WZ Masses

To see how the spontaneous symmetry-breaking process generates mass, it is useful to expand the solution near the minimum at $\frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}$ to obtain:

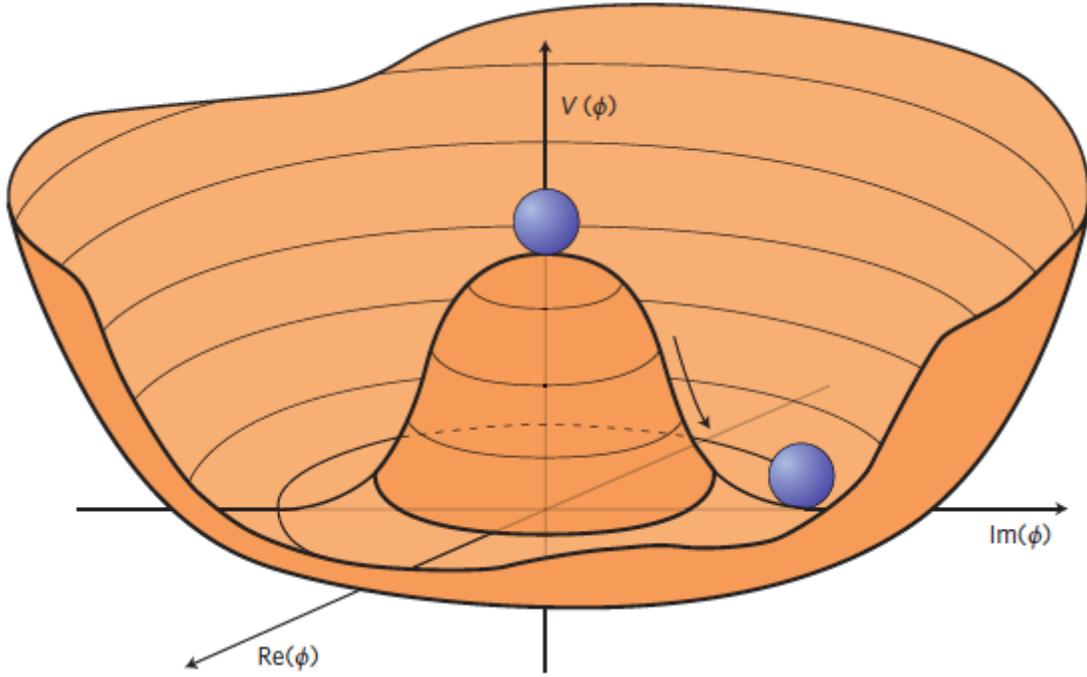


Figure 2.2: A visual display of the Higgs potential. Note that the local maximum at $\phi = 0$ is not stable. States at the bottom of the potential energy well do not respect the symmetry embedded in the shape of the potential [16].

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \quad (2.14)$$

The electroweak Lagrangian could be evaluated at this minimum energy state by acting the covariant derivative D_μ on ϕ as shown in Equation 2.11. After substituting the SU(2) generators into the covariant derivative, the mass term appears among many terms that come from expanding the derivative:

$$(D_\mu \phi)^\dagger (D_\mu \phi) = \frac{v^2}{8} [g^2 ((W_\mu^1)^2 + (W_\mu^2)^2) + (g' B_\mu - g W_\mu^3)^2] + \dots, \quad (2.15)$$

and it is useful to apply the parameterization introduced in Equation 2.7 to obtain the mass term as the following:

$$\frac{v^2}{8}[g^2((W_\mu^1)^2 + (W_\mu^2)^2) + (g'B_\mu - gW_\mu^3)^2] = \frac{g^2v^2}{8}W_\mu^+W_\mu^- + \frac{(g'^2 + g^2)v^2}{8}Z_\mu Z^\mu. \quad (2.16)$$

This mass term indicates that the mass of the two W^\pm bosons is $m_W = \frac{gv}{2}$, and that of the neutral Z boson is $m_Z = \frac{v}{2}\sqrt{g'^2 + g^2}$. Since the photon particle still preserves the $U(1)$ symmetry, it does not gain mass through the Higgs mechanism.

Fermion Masses

As shown in Equation 2.9, the simple construction of a mass term would not work in $SU(2)_L$. However, with the new Higgs field, ϕ , it is possible to construct a term in which the left-handed fields contract with the Higgs doublet while adding the right-handed singlet field. This contraction gives the so-called Yukawa term, the last term given in electroweak Lagrangian shown in Equation 2.5. The Yukawa term of the first generation of lepton is given as the following:

$$\mathcal{L}_{y,e} = -\lambda_e \sum_i \bar{\psi}_{i,L} \phi^i \psi_{f,R} + h.c., \quad (2.17)$$

where i is the index of the $SU(2)_L$ doublet and $h.c.$ denotes the hermitian conjugate of the term.

It is obvious that the term in Equation 2.17 is invariant under $SU(2)_L$ gauge transformations. Note that $\psi_{i,L}$ is a $SU(2)$ gauge doublet given by:

$$\psi_{e,L} = \begin{pmatrix} \nu_{e,L} \\ e_L \end{pmatrix} \quad (2.18)$$

where the first term and the second term represent the neutrino field and the massive lepton field, respectively. The right-handed field, $\psi_{e,R}$, is only a SU(2) singlet:

$$\psi_{e,R} = e_R. \quad (2.19)$$

Now the result of spontaneous symmetry breaking in the Yukawa term can be examined by substituting the solution from Equation 2.14 to obtain the overall Lagrangian in the following form:

$$\begin{aligned} \mathcal{L}_{y,e} &= -\lambda_e \sum_i \bar{\psi}_{i,L} \phi^i \psi_{f,R} + h.c \\ &= -\frac{1}{\sqrt{2}} \lambda_e \begin{pmatrix} \bar{\nu}_L \\ \bar{e}_L \end{pmatrix} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \psi_{f,R} + h.c \\ &= -\frac{1}{\sqrt{2}} \lambda_e \bar{e}_L v e_R + \dots \end{aligned} \quad (2.20)$$

Note that Equation 2.20 gives the electron mass with the vacuum expectation energy, v , of the Higgs field and its coupling constant, λ_e , to the electron.

The Yukawa coupling term above could be applied to all lepton generations, and it is observed to give no mass to the neutrino when the symmetry is broken because all neutrinos are left-handed. The lack of right-handed neutrino, ν_R , implies that a gauge invariant term for neutrinos could not be constructed in the form close to the one in Equation 2.17.

However, this is not a problem for quarks. Following the electroweak Lagrangian in Equation 2.5, it is clear that all quarks could, in principle, be right-handed. Once the same recipe is applied to quarks, the Lagrangian for the first generation of quarks is then given by:

$$\mathcal{L}_{y,q} = -\lambda_d \sum_i \bar{\psi}_{i,L} \phi^i \psi_{d,R} - \lambda_u \epsilon^{jk} \sum_k \bar{\psi}_{j,L} \phi^k \psi_{u,R} + h.c. \quad (2.21)$$

Then the spontaneous symmetry-breaking yields:

$$\begin{aligned}\mathcal{L}_{y,q} &= -\lambda_d \sum_i \bar{\psi}_{i,L} \phi^i \psi_{d,R} - \lambda_u \sum_i \bar{\psi}_{i,L} \phi^i \psi_{u,R} + h.c. \\ &= -\frac{1}{\sqrt{2}} \lambda_d \bar{d}_L v d_R - \frac{1}{\sqrt{2}} \lambda_u \bar{u}_L v u_R + \dots\end{aligned}\tag{2.22}$$

It is obvious that both the down and up quark obtained their mass through the Higgs mechanism:

$$m_d = \frac{1}{\sqrt{2}} \lambda_d v, m_u = \frac{1}{\sqrt{2}} \lambda_u.\tag{2.23}$$

2.2 The Georgi-Machacek Model

2.2.1 The Georgi-Machacek Lagrangian

This thesis discusses the search for charged Higgs bosons which are products of extensions of the Standard Model's scalar Higgs sector. Such extensions provide many interesting physical implications, offering possible explanations for the neutrino masses [18] and the recent discovery of the W mass anomaly where the W boson mass was measured to be significantly higher than the Standard Model prediction [19].

The Georgi-Machacek model [2] is an extension with a complex Higgs sector involving singly and doubly charged Higgs. Developed by Howard Georgi and Marie Machacek in 1985, the Georgi-Machacek model provides an extension compatible with the electroweak constraint of ρ , which is defined as the ratio of the two boson masses:

$$\rho = \frac{m_W^2}{m_Z^2 \cos^2 \theta} \approx 1\tag{2.24}$$

where θ is the experimentally measured weak mixing angle.

The Georgi-Machacek model considers the usual Higgs doublet and its bi-doublet form:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \text{ and } \Phi = \begin{pmatrix} \phi^{0*} & \phi^+ \\ -\phi^{+*} & \phi^0 \end{pmatrix} \quad (2.25)$$

and then a real triplet ξ as well as a complex triplet χ in addition to the Standard Model:

$$\xi = \begin{pmatrix} \xi_+ \\ \xi_0 \\ -\xi_+^* \end{pmatrix} \text{ and } \chi = \begin{pmatrix} \chi_{++} \\ \chi_+ \\ \chi_0 \end{pmatrix}. \quad (2.26)$$

.

The Higgs triplets field is defined in the bi-triplet form following the notation in Reference [20]:

$$X = \begin{pmatrix} \chi_0^* & \xi_+ & \chi_{++} \\ -\chi_+^* & \xi_0 & \chi_+ \\ \chi_{++}^* & -\xi_+^* & \chi_0 \end{pmatrix}. \quad (2.27)$$

The Lagrangian for fields is given by

$$\mathcal{L} = (D^\mu \phi)^\dagger (D_\mu \phi) + \frac{1}{2} \text{Tr}((D^\mu X)^\dagger (D_\mu X)). \quad (2.28)$$

After the spontaneous symmetry breaking, the model considers the following minimum solutions:

$$\langle \phi \rangle = \frac{1}{\sqrt{2}}v_\phi \quad \text{and} \quad \langle X \rangle = \begin{pmatrix} v_\chi & 0 & 0 \\ 0 & v_\chi & 0 \\ 0 & 0 & v_\chi \end{pmatrix}. \quad (2.29)$$

The W and Z mass constraint requires the vacuum expectation value to satisfy the following equation:

$$v_\phi^2 + 8v_\chi^2 \equiv v = \frac{1}{\sqrt{2}G_F} \approx (246 \text{ GeV})^2. \quad (2.30)$$

In order to construct a Lagrangian that respects the required SU(2) symmetry, the scalar potentials of the fields in the form of $X^\dagger X$ and $\Phi^\dagger \Phi$ are considered. It is interesting to note that contracting the aforementioned terms with SU(2) generators also produces terms invariant to the SU(2) group. Therefore, in the conventions of Reference [21], the most general scalar potential preserving the $SU(2)_L \times SU(2)_R$ symmetry is given by:

$$\begin{aligned} V(\Phi, X) = & \frac{\mu_2^2}{2} \text{Tr}(\Phi^\dagger \Phi) + \frac{\mu_3^2}{2} \text{Tr}(X^\dagger X) + \lambda_1 (\text{Tr}(\Phi^\dagger \Phi))^2 + \lambda_2 \text{Tr}(\Phi^\dagger \Phi) \text{Tr}(X^\dagger X) \\ & + \lambda_3 \text{Tr}(X^\dagger X X^\dagger X) + \lambda_4 \text{Tr}(X^\dagger X)^2 - \lambda_5 \text{Tr}(\Phi^\dagger \tau^a \Phi \tau^b) \text{Tr}(X^\dagger t^a X t^b) \\ & - M_1 \text{Tr}(\Phi^\dagger \tau^a \Phi \tau^b) (UXU^\dagger)_{ab} - M_2 \text{Tr}(X^\dagger t^a X t^b) (UXU^\dagger)_{ab} \end{aligned} \quad (2.31)$$

where the Roman letter indices loop through the three generator matrices when contracted. The three two-dimensional SU(2) generators, τ matrices, are defined using Pauli

matrices with $\tau^a = \frac{\sigma^a}{2}$ and the three three-dimensional SU(2) generators, t , are:

$$t^1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad t^2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad t^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \quad (2.32)$$

The matrix U in Equation 2.31 is given by:

$$U = \begin{pmatrix} -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ -\frac{i}{\sqrt{2}} & 0 & -\frac{i}{\sqrt{2}} \\ 0 & 1 & 0 \end{pmatrix}. \quad (2.33)$$

The fields can be re-organized into a fiveplet, a triplet, and two singlets:

	Doubly charged	Singly charged	Neutral
Fiveplet	$H_5^{++} = \chi^{++}$	$H_5^+ = \frac{\chi^+ - \xi^+}{\sqrt{2}}$	$H_5^0 = \sqrt{\frac{2}{3}}\xi^{0,r} - \sqrt{\frac{1}{3}}\chi^{0,r}$
Triplet	—	$H_3^+ = -s_H\phi^+ + c_H\frac{\chi^+ + \xi^+}{\sqrt{2}}$	$H_3^0 = -s_H\phi^{0,i} + c_H\chi^{0,i}$
Singlet	—	—	$H_1^0 = \phi^{0,r}, H_1^{0'} = \sqrt{\frac{1}{3}}\xi^{0,r} + \sqrt{\frac{2}{3}}\xi^{0,r}$

Table 2.1: Mass eigenstates of Georgi-Machacek model's Higgs sector.

Note that the mixing angle used in Table 2.1 is given by:

$$s_H \equiv \sin\theta_H = \frac{2\sqrt{2}v_\chi}{v}, \quad c_H \equiv \cos\theta_H = \frac{v_\phi}{v}. \quad (2.34)$$

The fiveplet and the triplet state in Table 2.1 are degenerate. Their masses are given by parameters of the scalar potential and vacuum expectation values:

$$\begin{aligned}
m_5^2 &= \frac{M_1}{4v_\chi} v_\phi^2 + 12M_2 v_\chi + \frac{3}{2} \lambda_5 v_\phi^2 + 8\lambda_3 v_\chi^2, \\
m_3^2 &= \frac{M_1}{4v_\chi} v^2 + \frac{\lambda_5}{2} v^2.
\end{aligned}
\tag{2.35}$$

The two singlets will then mix by an angle α to yield mass eigenstates h and H :

$$\begin{aligned}
h &= c_\alpha H_1^0 - s_\alpha H_1^{0'}, \\
H &= s_\alpha H_1^0 + c_\alpha H_1^{0'},
\end{aligned}
\tag{2.36}$$

where $c_\alpha = \cos\alpha$ and $s_\alpha = \sin\alpha$.

Using the same recipe for the spontaneous symmetry breaking in the standard model, the GM model Higgs yields the following mass mixing matrix in the W and Z basis:

$$M^2 \propto \begin{bmatrix} g^2 & 0 & 0 & 0 \\ 0 & g^2 & 0 & 0 \\ 0 & 0 & g^2 & -gg' \\ 0 & 0 & -gg' & g^2 \end{bmatrix}.
\tag{2.37}$$

Comparing Equation 2.37 with Equation 2.16, it is easy to see that the mixing is the same and hence the electroweak constraint in Equation 2.24 is preserved. This preservation of the electroweak constraint is one of the key motivations for the introduction of the GM model. Previous work had also linked the GM model to the W mass anomaly [19] and neutrino masses [22], providing stronger motivation for the search for a GM model charged Higgs boson.

2.2.2 H_5^+ Decay Channels

In the Yukawa coupling between the Higgs and fermions, the Higgs field is required to be a $SU(2)_L$ doublet to match the transformation property of the fermion doublet as shown in Equation 2.20. Since the fiveplet H_5 has no doublet representation, its couplings with leptons are strictly forbidden at the tree level. However, the fiveplet states couple to vector boson pairs with a coupling strength proportional to s_H , which is defined in Table 2.34. The coupling strength of the fiveplet states with one or two more gauge bosons is given in Table 2.2.

One gauge boson	Two gauge bosons
$g_{H_5^+ H_5^- \gamma} = e$	$g_{H_5^+ W^- Z} = -\frac{e^2 v}{2s_W^2 c_W} s_H$
$g_{H_5^+ H_5^- Z} = \frac{e}{2s_W c_W} (1 - 2s_W^2)$	$g_{H_5^0 W^+ W^-} = \frac{e^2 v}{2\sqrt{3}s_W^2} s_H$
$g_{H_5^+ H_5^- W^+} = \frac{e}{\sqrt{2}s_W}$	$g_{H_5^0 Z Z} = -\frac{e^2 v}{\sqrt{3}s_W^2 c_W^2} s_H$
$g_{H_5^+ H_3^- Z} = -\frac{e}{2s_W c_W} c_H$	$g_{H_5^{++} W^- W^-} = \frac{e^2 v}{\sqrt{2}s_W^2} s_H$
$g_{H_5^+ H_5^0 W^-} = \frac{\sqrt{3}e}{2s_W}$	—
$g_{H_5^+ H_3^0 W^-} = -\frac{ie}{2s_W} c_H$	—

Table 2.2: Couplings of the fiveplet states H_5 with one or two gauge bosons. Notice that all couplings with two gauge bosons are proportional to s_H .

Since the WZ decay channel, $H_5^\pm \rightarrow W^\pm Z$, is a tree-level decay, it is therefore expected to have a much higher decay width compared to that of the $W\gamma$ channel, $H_5^\pm \rightarrow W^\pm \gamma$, which consists of loops in the Feynman diagram [23]. Therefore, the $W\gamma$ decay channel could be favored when the WZ channel is suppressed. This suppression of the WZ channel could happen in two ways:

- $\sin\theta_H \ll 1$, which suppresses the vertex of $H_5^\pm W^\pm Z$.
- The fiveplet mass is below the WZ threshold, $m_5 < m_W + m_Z$, so that the production of at least one of the intermediate vector bosons is off-shell.

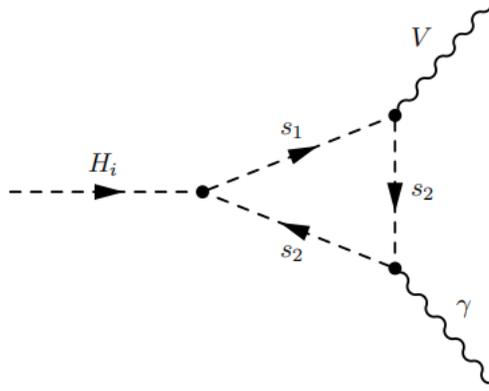


Figure 2.3: Loop induced Feynman diagram of $HW\gamma$ production [24].

This thesis examines the parameter space of small $\sin\theta_H$ and low fiveplet mass, which favors the $W\gamma$ decay channel. The mass of the fiveplet states is also assumed to be smaller than that of the triplet states, $m_5 < m_3$, so that H_5 will not decay into other scalar Higgs. Decays channels of the fiveplet members are listed in Table 2.3 [20]:

Particle	Decay Channels	Comment
H_5^\pm	$H_5^\pm \rightarrow W^\pm\gamma$ $H_5^\pm \rightarrow W^{\pm(*)}Z^{(*)}$	Loop-induced Suppressed by s_H^2 , off-shell
H_5^0	$H_5^0 \rightarrow \gamma\gamma$ $H_5^0 \rightarrow Z\gamma$ $H_5^0 \rightarrow Z^{(*)}Z^{(*)}$ $H_5^0 \rightarrow W^{\pm(*)}W^{\mp(*)}$	Loop-induced Loop-induced, phase space disfavoured Suppressed by s_H^2 , off-shell Suppressed by s_H^2 , off-shell
$H_5^{\pm\pm}$	$H_5^{\pm\pm} \rightarrow W^{\pm(*)}W^{\pm(*)}$	Suppressed by s_H^2 , off-shell

Table 2.3: Decay channels for members of the scalar fiveplet at low mass, including possible off-shell decays [20].

A plot of the $H_5^\pm \rightarrow W^\pm\gamma$ branching ratio dependence on m_5 and s_H from Reference [20] is shown in Figure 2.4, where the parameters of the model are fixed by:

$$\begin{aligned}
m_3^2 &= m_5^2 + \delta m^2, \\
m_H^2 &= m_5^2 = \frac{3}{2}\delta m^2 + \kappa_H v^2 s_H^2, \\
M_1 &= \frac{\sqrt{2}}{v}(m_5^2 + \frac{3}{2}\delta m^2 + \kappa_H v^2 s_H^2), \\
s_\alpha &= \kappa_\alpha s_H, \\
\delta m^2 &= (300 \text{ GeV}^2), \\
\kappa_\alpha &= -0.15 - \frac{m_5}{1000 \text{ GeV}}, \\
\kappa_H &= -\frac{m_5}{100 \text{ GeV}}, \\
\kappa_{\lambda_3} &= -\frac{\kappa_H^2}{10}.
\end{aligned} \tag{2.38}$$

The branching ratio in Figure 2.4 increases significantly as m_5 drops below the WZ threshold at around 171 GeV. This motivates the search in this thesis to be conducted in the parameter space of small values of s_H and low m_5 values between 110 GeV to 200 GeV. In this parameter space, it is obvious from the couplings in Table 2.2 that the leading production processes of signal events from proton collisions are those involving vertices between two Higgs and a vector boson:

- $pp \rightarrow H_5^\pm H_5^0 \rightarrow W^\pm \gamma + \gamma \gamma \rightarrow l\nu_l \gamma + X$,
- $pp \rightarrow H_5^\pm H_5^{\mp\mp} \rightarrow W^\pm \gamma + W^\mp W^\mp \rightarrow l\nu_l \gamma + X_2$,
- $pp \rightarrow H_5^+ H_5^- \rightarrow W^\pm \gamma + W^\mp \gamma \rightarrow l\nu_l \gamma + X_3$,

where X , X_2 and X_3 label other products produced in the decays. Since GM model Higgs bosons do not couple to fermions, they could only be produced from proton collisions via the Drell-Yan processes shown in Figure 2.5. The quarks within the proton collide to produce a vector boson, which then produces the Higgs pair via the HHV vertex.

In conclusion, according to the GM model, various signal processes involving charged Higgs bosons can be probed by measuring the lepton final state particles, $\ell + \gamma$, produced in proton-proton collisions. The Large Hadron Collider, capable of producing proton-proton collisions at extremely high energies, is thus an ideal experimental apparatus to conduct searches of the charged Higgs boson. The following chapter will introduce the Large Hadron Collider and the ATLAS detector, one of the main detectors that measure the final state particles produced in collisions.

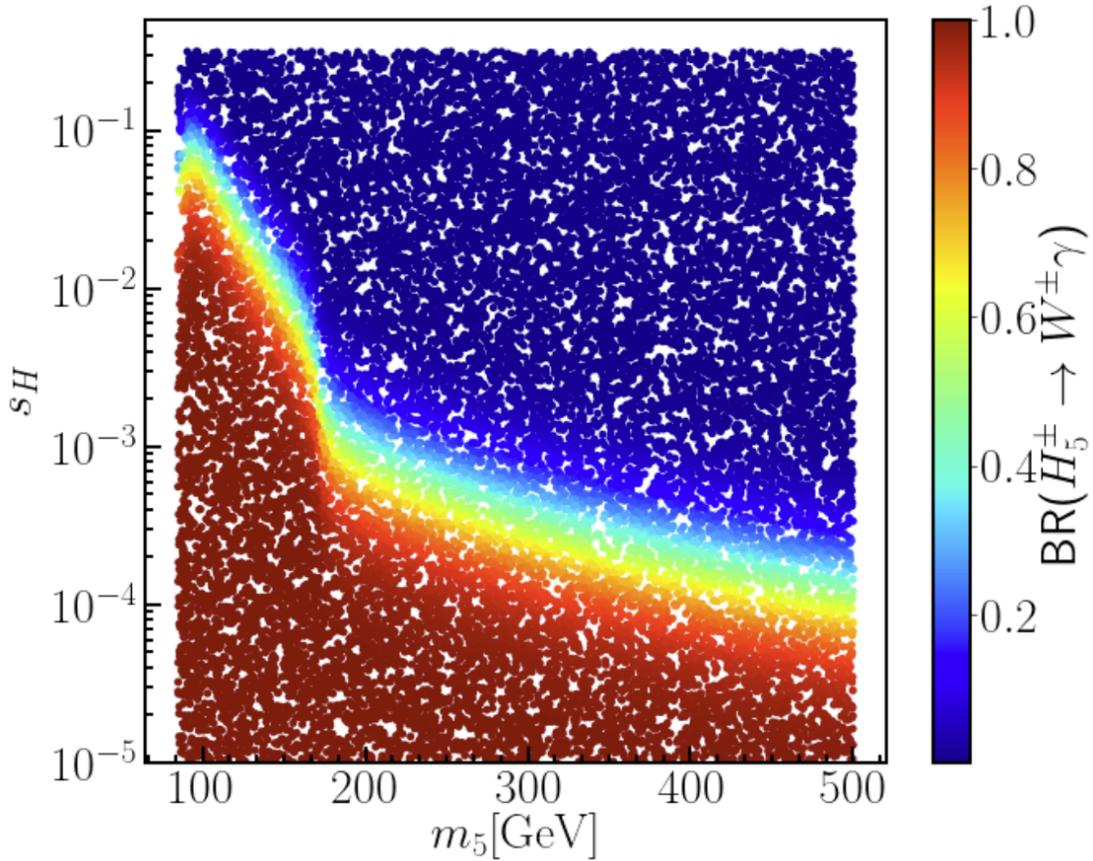


Figure 2.4: Dependence of the branching ratio of $H_5^\pm \rightarrow W^\pm \gamma$ on m_5 and s_H , for $M_2 = 40$ GeV [20].

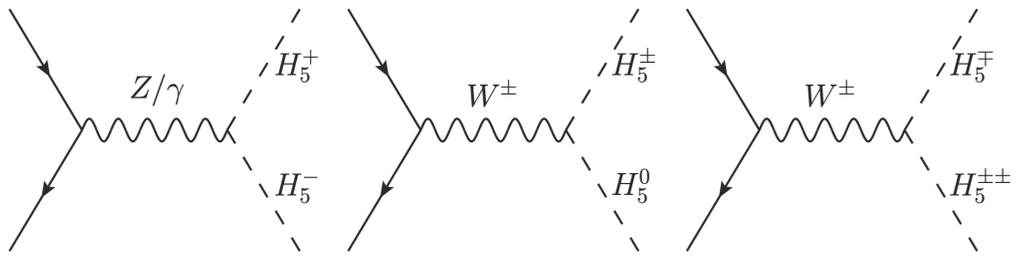


Figure 2.5: Feynman diagrams for the dominant Drell-Yan production processes involving H_5^+ [20].

Chapter 3

Experimental Setup

This thesis analyzes particle collision events collected by the ATLAS detector at the Large Hadron Collider. In this chapter, these two machines are introduced in order to provide an overview of the data collection process.

Section 3.1 describes the status of the Large Hadron Collider, with an overview of experiments in the CERN accelerator complex. In Section 3.2, various detector modules of the ATLAS detector are discussed.

3.1 The Large Hadron Collider

The LHC is the largest particle collider in the world, located on the French-Swiss border near Geneva. It is built in a circular tunnel approximately 100 meters underneath the ground, with a circumference of 27km. The LHC is mainly operating to study proton-proton ($p-p$) collisions, while other collisions involving heavy ions(e.g., p -Pb, Pb-Pb) are also produced. For proton-proton collisions, the LHC was designed to collide at a maximal center-of-mass energy of $\sqrt{s} = 14$ TeV, with each proton beam accelerated to a beam energy of 7 TeV.

There are four main detectors on the LHC ring: ATLAS [25], CMS [26], ALICE [27], and LHCb [28], as shown in Figure 3.1. The ATLAS and CMS detectors are general-purpose detectors capable of precise measurements of different final states produced by collisions in the LHC. The detector of the ALICE collaboration is optimized to study the quark-gluon plasma produced in nuclei collisions. The LHCb detector specializes in detecting particles in the forward direction to study the b quark, which helps us understand the matter-antimatter asymmetry in the universe.

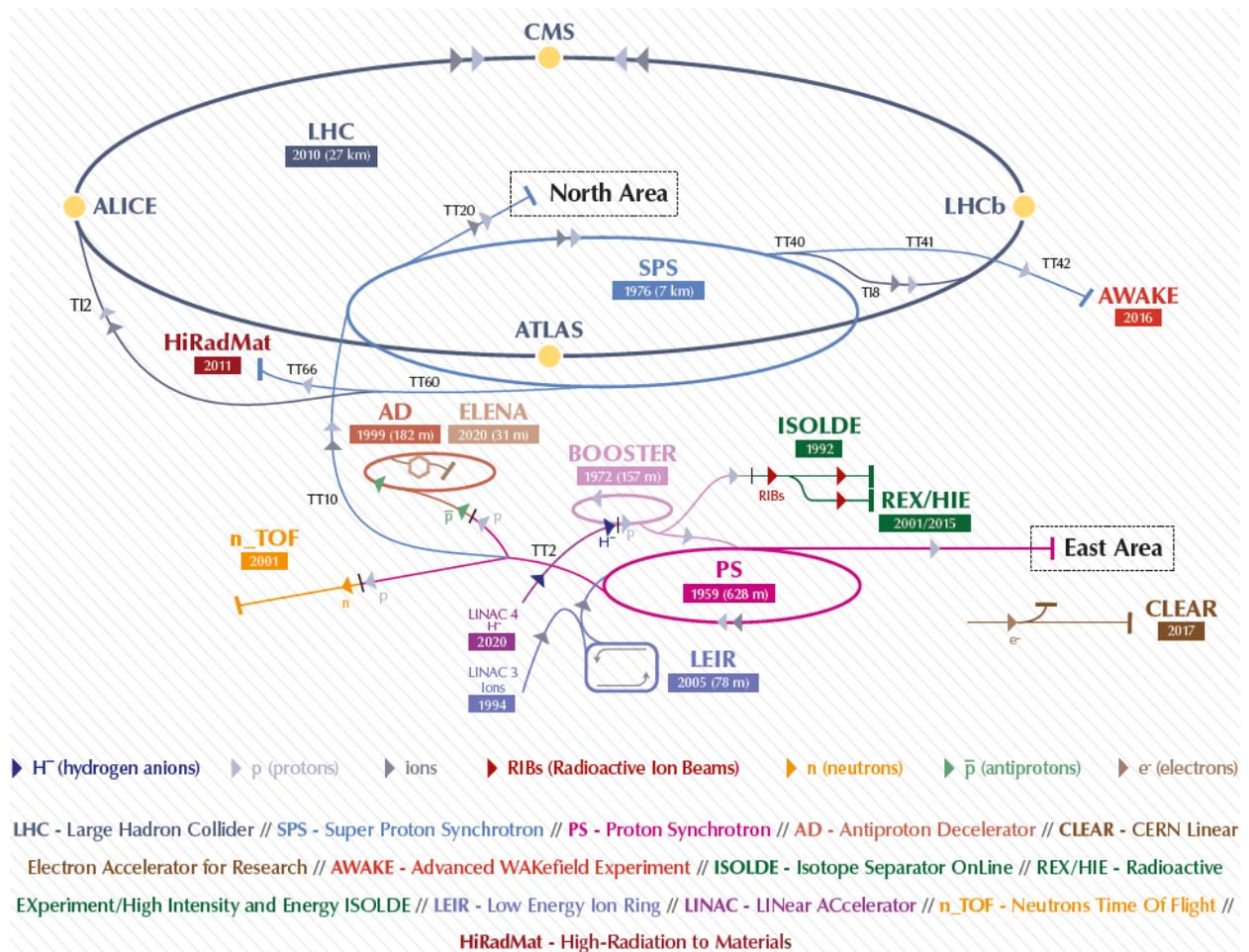


Figure 3.1: The CERN accelerator complex [29]. The four main detectors, ATLAS, CMS, ALICE, and LHCb, are labeled by the yellow dots on the LHC ring.

3.1.1 Beam Preparation

Injection chain and collisions

The proton beam in the LHC is first prepared and injected into the LHC ring by a series of devices in the accelerator complex shown in Figure 3.1. The first step in making the beam is ionizing hydrogen atoms to produce protons. After protons are extracted from the hydrogen gas, they are sent to the Linear Accelerator 2 in bunches to be accelerated to an energy of 50 MeV. The proton bunches are then directed toward the Proton Synchrotron Booster to increase their energy to 1.4 GeV. The energy of the bunches is then further increased from 1.4 GeV to 25 GeV as they enter the Proton Synchrotron. As the last step of the injection chain, the Super Proton Synchrotron can take the bunches from PS and accelerate them to 450 GeV, which is the desired energy of the proton bunches when they enter the LHC ring.

Inside the LHC ring, beam energy will then increase from 450 GeV to 6.5 TeV, which is the nominal collision energy of the run-2 period. The LHC ring consists of eight octants, each with its own "Point" where the experimental facilities are built. The two LHC rings cross at Point 1(ATLAS), 2(ALICE), 5(CMS), and 8(LHCb), as shown in Figure 3.2. The two beams inside the LHC rings are directed to rotate in opposite directions and are eventually brought to collide in the interaction points (IPs). After the beams have been given sufficient time to interact, the number of protons remaining in the beams will continue to decrease due to collisions. The whole beam will eventually be dumped into the dedicated facility at Point 6 to leave space for a refill.

Dipoles

The LHC has 1232 15 meters long dipole magnets with the same basic design, occupying about 2/3 of its circumference. Figure 3.3 shows a cross-sectional view of a dipole.

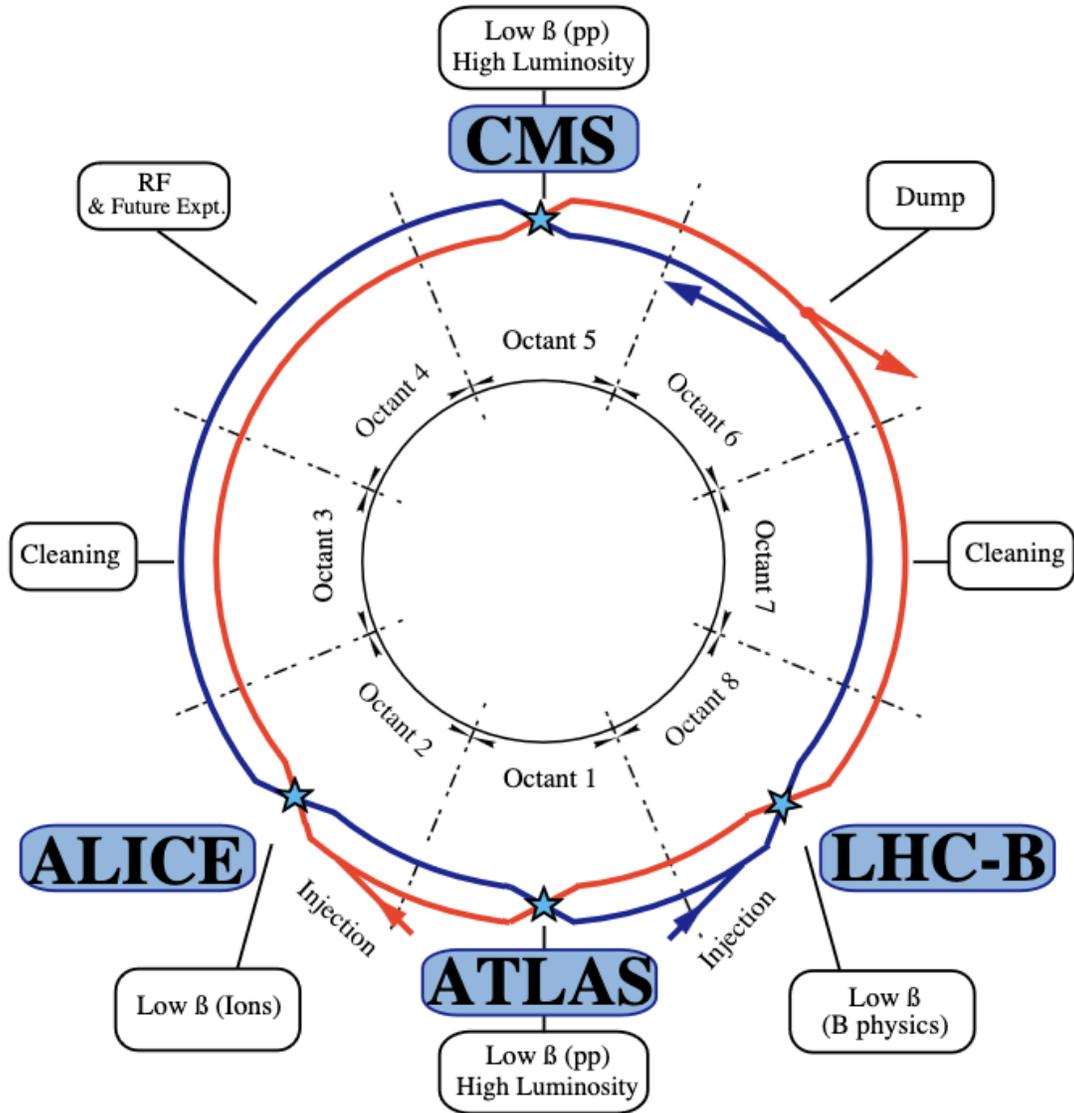


Figure 3.2: The layout of the LHC ring and various experiments [30].

The module is cooled to an extremely low temperature for the delivery of a strong magnetic field that bends the beam inside the LHC ring to fix its trajectory. In the intermediate thermal shield, the temperature is cooled to 55 K. Furthermore, the temperature is cooled down to only 1.9 K inside the iron yoke, yielding a magnetic field of 8.33 T with the superconducting Niobium-Titanium cable to constrain the motion of the beam precisely.

The dipoles are mainly used to bend the beam to have the same radius of motion as the radius of the LHC ring. Inside the cryodipole, as shown in Figure 3.3, quadrupoles are installed to assist in focusing the beam in the transverse plane. Higher-order multipoles are also used to gain more precise beam control as the energy increases.

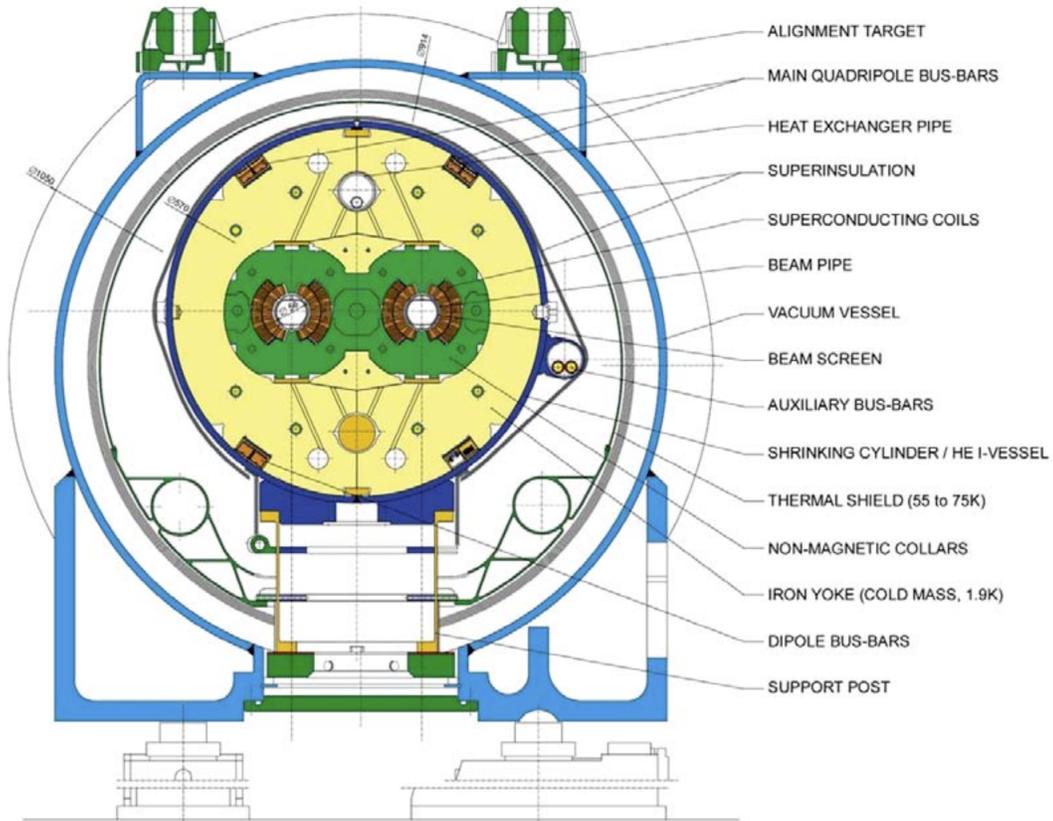


Figure 3.3: Cross-section of cryodipole [31].

RF cavities

The LHC has sixteen Radio Frequency (RF) cavities implemented to accelerate the beams. The RF cavities are superconducting metallic chambers that provide electromagnetic fields that oscillate at a frequency of 400 MHz with a maximal amplitude of up to 16 MV. Once the proton in the beam is accelerated to have the same frequency as that of the RF cavity, it is called a synchronous particle, which will no longer be subject to any

acceleration. Protons out of synchronization with the RF cavity will be accelerated or decelerated to match their frequency of passage with the frequency of RF cavities. The RF cavities help the LHC to focus the proton longitudinally and thus keep them in bunches separated by 25 ns in time, with each bunch typically containing up to 10^{11} protons.

3.1.2 Luminosity and Pileup

Luminosity

To describe the collider's capability of producing collisions, it is convenient to introduce a collider physics quantity called luminosity (L). The definition of luminosity is given in Equation 3.1:

$$L = \frac{1}{\sigma} \frac{dN}{dt} \quad (3.1)$$

where σ is the cross section, and N is the number of collisions that took place within time t . The luminosity above is often referred to as the instantaneous luminosity, which is in units of $m^{-2}s^{-1}$.

Since the cross-section σ embedded in N was canceled by the factor $\frac{1}{\sigma}$ in the front of Equation 3.1, the luminosity of a collider is independent of the rareness of the process under investigation. Therefore, this quantity is a collider parameter, which solely depends on the specifications of the collider itself. In particular, when two beams are directed to collide, it is convenient to express the luminosity in other collider-related quantities as shown in Equation 3.2:

$$L = f_{rev} \frac{N_1 N_2 N_b}{4\pi\sigma_x\sigma_y} S \quad (3.2)$$

where f_{rev} is the revolution frequency of the proton beam, N_1, N_2 are the number of particles in each bunch involved in the collision, and N_b is the number of bunches in each beam. σ_x and σ_y are parameters describing the width of the beam in the transverse plane, and S is the geometric scale factor that accounts for the effects of the crossing angle[13].

Accurate measurements of the instantaneous luminosity are imperative to all experiments at the LHC. For example, the ATLAS experiment uses the LUCID-2 detector [32] and the Beam Condition Monitor [33] to measure the instantaneous luminosity during the run-2 period between 2015 and 2018. The LUCID-2 detectors are Cherenkov detectors placed at $z = \pm 17$ m on both sides of the interaction point, while the BCM diamond detector measures complementary information from the bunch crossing. The two measurements are combined to estimate the visible rate per bunch crossing μ_{vis} . In each year of the LHC run, special calibration runs are performed to have van der Meer scans [34] yield the calibration constant σ_{vis} which relates μ_{vis} to the luminosity by Equation 3.3:

$$L = \frac{\mu_{vis} f_{rev}}{\sigma_{vis}}. \quad (3.3)$$

In the run-2 period, the instantaneous luminosity recorded in the LHC fluctuates around $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ [35]. Therefore, to quantitatively measure the number of collisions taking place in more extended periods, it is convenient to use the integrated luminosity, which is calculated by integrating the instantaneous luminosity over the period as shown in Equation 3.4:

$$\mathcal{L} = \int L dt. \quad (3.4)$$

The ATLAS measurement of integrated luminosity in each operating year between 2011 and 2018 is shown in Figure 3.4 [36]. These plots reveal that the LHC increased the

collision energy from 8 TeV to 13 TeV over the years and achieved an integrated luminosity in 2018 that is about ten times larger than that in 2011.

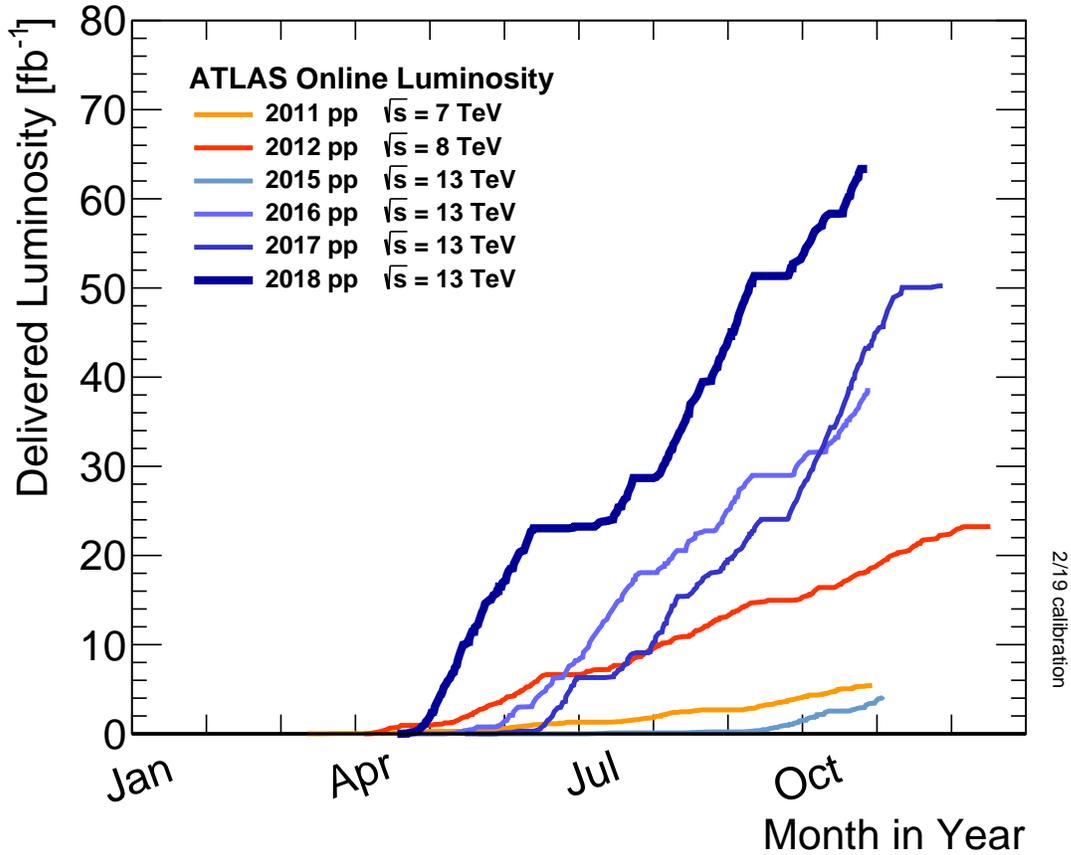


Figure 3.4: Cumulative luminosity versus day delivered to ATLAS during stable beams for high energy p-p collisions [36].

	2015	2016	2017	2018	Combined
Integrated Luminosity [fb^{-1}]	3.24	33.40	44.63	58.79	140.07
Total uncertainty [fb^{-1}]	0.04	0.30	0.50	0.64	1.17

Table 3.1: Integrated luminosity in run-2 with uncertainties [35].

Pileup interactions

Due to the large number of protons in each bunch, a typical bunch crossing yields dozens of proton-proton collisions near the interaction point. Particle tracks produced

in the collisions could be traced back to their vertices along the beam line, where the collisions took place. To focus on rare physics processes, the hard-scatter vertex is of great interest to LHC data analysis. Out of all the vertices in a collision, the hard-scatter vertex is the one with the highest transverse momentum (p_T) square, $\sum p_T^2$, over all associated tracks.

In contrast to the hard objects from the hard-scatter vertex and the intriguing physics phenomenon behind them, other vertices contain much softer outgoing particles, which are characterized by their low transverse momentum. These inelastic scattering vertices are called pileup vertices, and the event is referred to as the pileup event or simply the pileup. There are two kinds of pileup: the in-time pileup, produced by vertices in the current bunch crossing under investigation, and the out-of-time pileup from a previous or subsequent bunch crossing.

Since most BSM physics are produced in the high energy regime, they could only be produced by having head-on inelastic collisions that yield hard scattering. Pileup interactions provide soft particles that impede the ability to identify, measure, and associate hard-scatter objects for the study of rare physics processes. Therefore, various efforts have been used in the attempt to mitigate the effects of pileup [37]. Pileup interaction contamination is characterized by measuring the average number of interactions ($\langle\mu\rangle$) per bunch crossing. A plot of the ATLAS measurements of $\langle\mu\rangle$ in each of the run-2 operating years is shown in Figure 3.5.

3.2 The ATLAS detector

The physics analysis in this thesis is based on the data collected by the ATLAS (A Toroidal LHC ApparatuS) detector during the run-2 period of the LHC. In order to detect and identify all the secondary particles coming from the proton-proton collisions, the ATLAS detector provides a nearly 4π radians solid angle coverage with three main

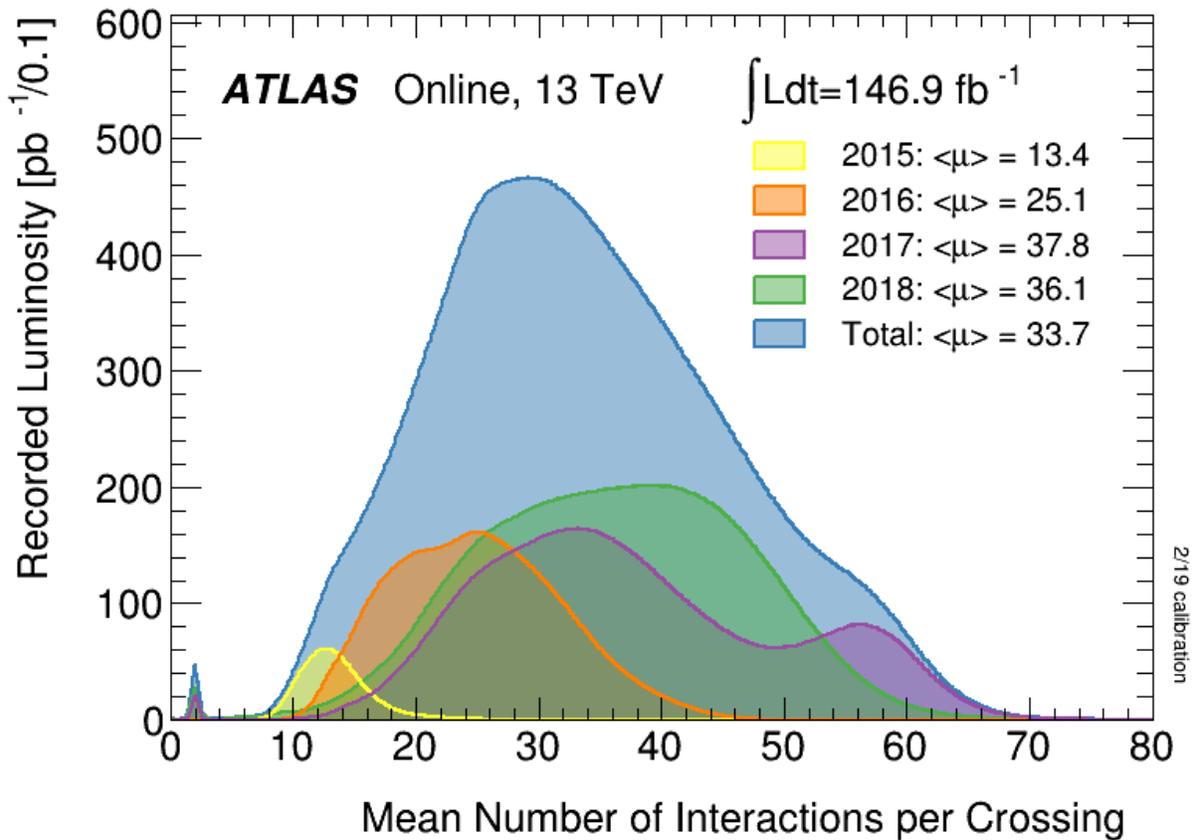


Figure 3.5: The Run-2 pileup distribution [38].

subdetector systems, each designed to specialize in measuring certain types of outgoing particles. Starting from the innermost system, the Inner Detector (ID) provides precise trajectory measurements of charged particles. In the middle layer, the calorimeter system determines energy deposits from charged and neutral particles. In the outermost layer, the muon spectrometer captures signals from muons, which leave negligible energy deposits in the calorimeters due to their nature of minimum ionizing particles. A diagram of the ATLAS detector with all of its subdetectors is shown in Figure 3.6.

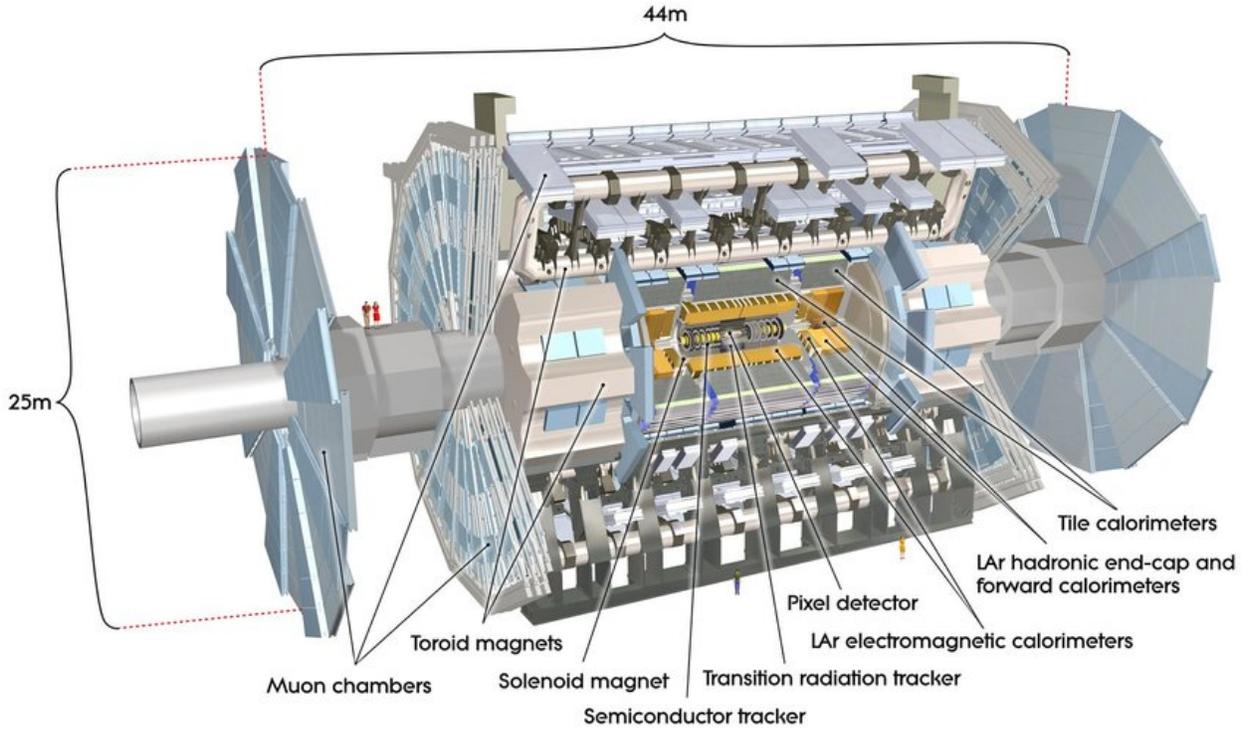


Figure 3.6: A diagram of the ATLAS detector with its subsystems [39].

Coordinate system

The convention of the coordinate system in the ATLAS experiment is to set the x -axis towards the center of the LHC ring, the y -axis pointing upward to the sky, and the z -axis along the beamline. The momentum vector of a measured particle is broken into three directions: p_x , p_y , and p_z . The transverse momentum p_T is defined as the magnitude of the momentum vector's projection onto the transverse plane:

$$p_T = \sqrt{p_x^2 + p_y^2}. \quad (3.5)$$

In practice, it is much easier to use the $\eta - \phi$ coordinates rather than the Cartesian coordinate system. In the $\eta - \phi$ coordinate system, the azimuthal angle ϕ is defined as the angle in the plane transverse to the beam axis \hat{z} and the value is set such that $\phi = 0$ cor-

responds to a direction pointed towards the center of the LHC ring. The pseudorapidity η is a quantity defined based on the polar angle θ , which is measured from the beam axis \hat{z} with $\theta = 0$ being towards \hat{z} and $\theta = \frac{\pi}{2}$ pointing upward in the transverse plane. The pseudorapidity is defined as the following:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \quad (3.6)$$

The pseudorapidity is 0 in the transverse plane, whereas $\eta = \infty$ points towards the beam axis. In the limit of large momentum compared to mass, the pseudorapidity is approximately the same as the rapidity y :

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}. \quad (3.7)$$

The difference in rapidity, and its pseudorapidity approximation, is a Lorentz invariant quantity. Lorentz invariant quantities are convenient in describing particles from proton-proton collisions since the proton constituents, partons, involved in the collision carry an unknown fraction of their momentum along the z-axis, which corresponds to a boost with an unknown magnitude.

The $\eta - \phi$ coordinate system (p_T, η, ϕ) covers the entire 4π radians solid angle to describe outgoing particles in all directions. In the following chapters, it is important to consider the distance in the $\eta - \phi$ plane ΔR :

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} \quad (3.8)$$

where $\Delta\phi$ and $\Delta\eta$ are differences in ϕ and η between two points in the $\eta - \phi$ plane.

3.2.1 Inner Detector

The inner detector (ID) [40, 41] of ATLAS measures trajectories of charged particles with a coverage of $|\eta| < 2.5$. The ID is encapsulated in the 2T magnetic field generated by the solenoid magnet. As charged particles travel through the inner detector, their paths get bent in the transverse plane to form circular trajectories, the radii of which depend on the charge and momentum of the particles under investigation. These measured trajectories are called tracks, which are used in many object reconstruction algorithms described in Chapter 4. For a charged particle with considerable momentum, its trajectory radius will be large enough such that the trajectory is close to a straight line. The large radius implies that the momentum resolution is better for low p_T particles, which the magnetic field could easily bend. In the performance study with cosmic muons, the ID momentum resolutions are shown to be around $\frac{\sigma_{p_T}}{p_T} = 1.6\% \pm 0.1\%$ at low momenta and around 50% at 1 TeV [42]. A schematic view of the ATLAS inner detector is shown in Figure 3.7 and Figure 3.8 with all three ID submodules: the pixel detectors, the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT).

Pixel detector

The pixel detector consists of three layers of silicon sensors in the barrel and the two end-cap regions, with a coverage of $|\eta| < 2.5$. Each silicon semiconductor sensor is $50 \times 400 \mu\text{m}$ in area and $250 \mu\text{m}$ in thickness. When charged particles traverse the pixel sensor, electrons inside the semiconductor are excited to the conduction band, leaving holes behind them. The electric field then accelerates electrons and holes to register in the electrodes, which results in a pulse read out by the dedicated electronics. The pixel sensor provides an excellent three-dimensional measurement of the track with a spatial resolution of $10 \mu\text{m}$ in the transverse plane and $115 \mu\text{m}$ along the beam axis. Over 67 million

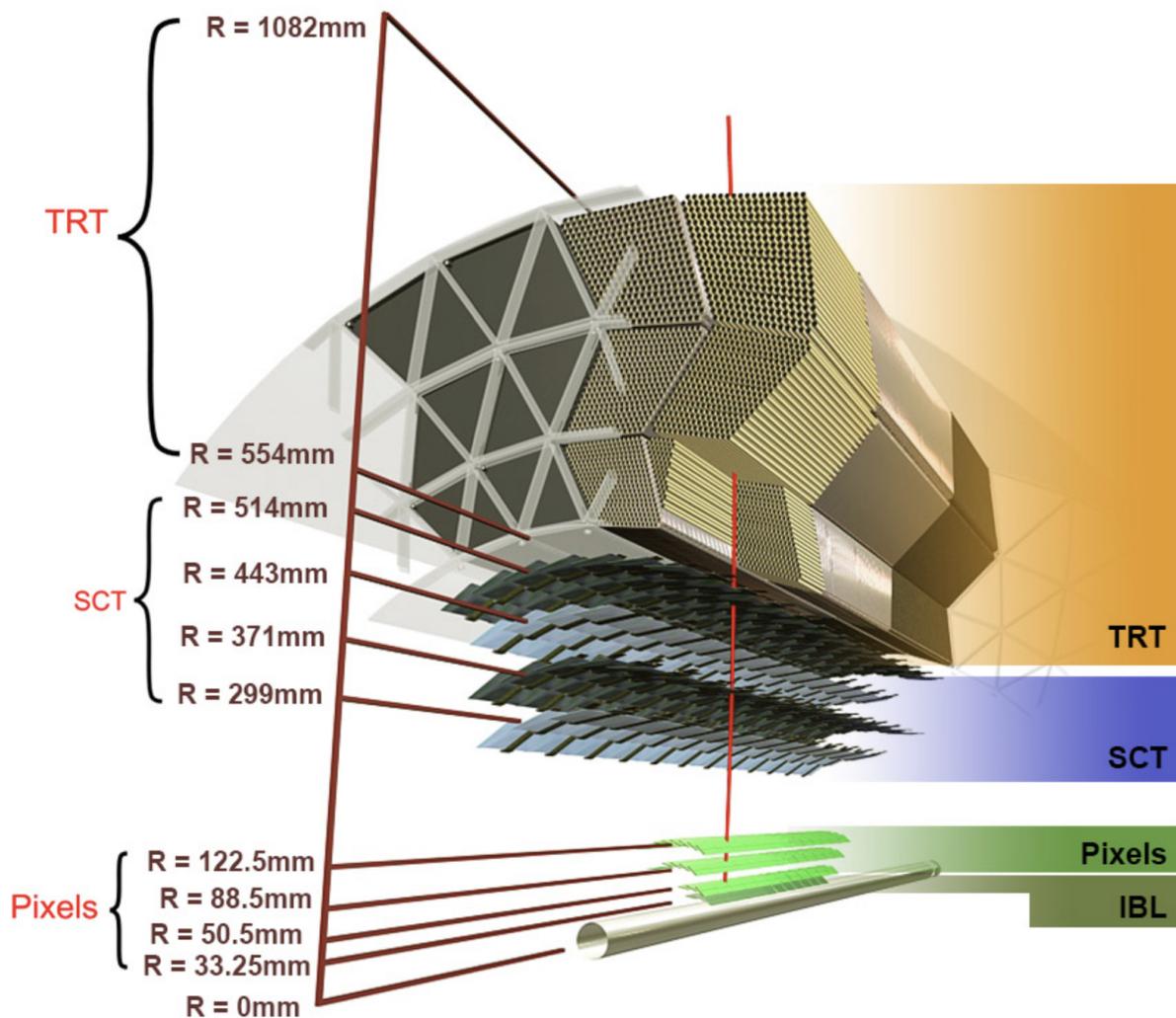


Figure 3.7: A diagram of the ATLAS inner detector with its subsystems [43].

pixel sensors are installed in the barrel region, while the end-cap regions have around 13 million pixel sensors.

An extra pixel layer called Ininsertable B-Layer (IBL) was installed before the run-2 data-taking period to improve vertexing and b-tagging resolutions [44]. The IBL consists of 12 million $50 \mu\text{m} \times 200 \mu\text{m}$ pixel sensors, which have a spatial resolution of $8 \mu\text{m}$ in the transverse plane and $40 \mu\text{m}$ in \hat{z} .

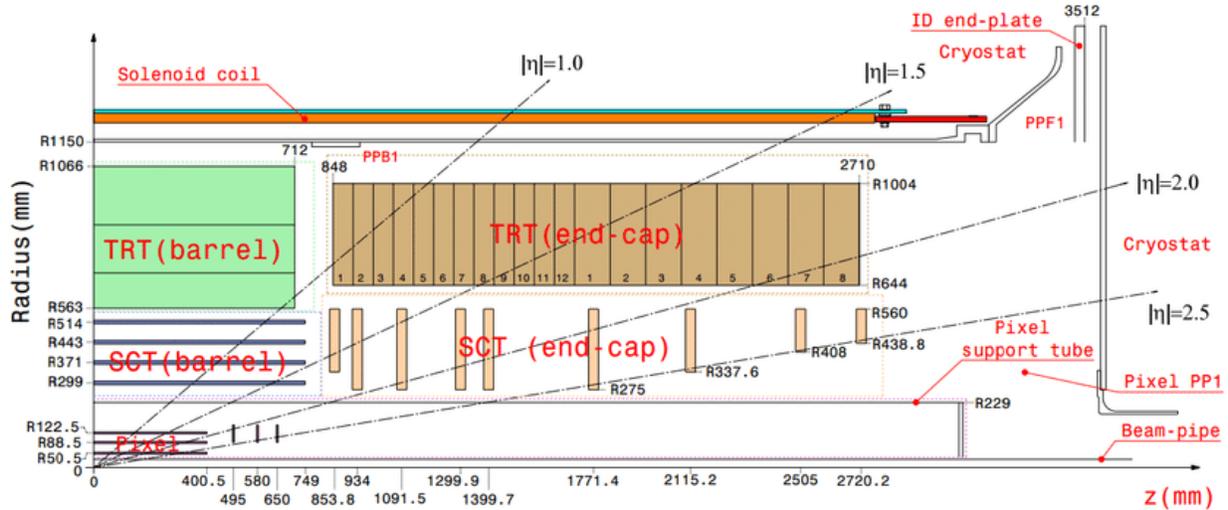


Figure 3.8: The inner detector of the ATLAS experiment shown in the R-z plane [45]. The vertical axis marks the radial distance orthogonal to the beam axis.

Semiconductor Tracker

Immediately outside of the pixel detector, the Semiconductor Tracker uses similar silicon sensors in 6 cm long microstrips. Each microstrip is $285\ \mu\text{m}$ wide with an $80\ \mu\text{m}$ pitch, providing a hit resolution of $17\ \mu\text{m}$ in the transverse plane and $580\ \mu\text{m}$ in \hat{z} . As shown in Figure 3.8, there are four layers of SCT in the barrel region and nine layers in each end-cap region.

Transition Radiation Tracker

The Transition Radiation Tracker [46] is the outermost component of the ID, covering the range of $|\eta| < 2$. The TRT exploits the phenomenon known as transition radiation, where radiations are emitted by particles moving at a constant velocity in an inhomogeneous medium [47].

The TRT is made of drift tubes with diameters of 4 mm. Each tube has a $31\ \mu\text{m}$ diameter gold-plated tungsten wire in the middle. The operating potential difference between

the wall and the wire is 1.5 kV, with the wire at ground potential. Each tube is filled with a gas mixture made of 70 % Xe, 27 % CO₂, and 3 % O₂. The central barrel region of the TRT contains 52,544 straw tubes parallel to the beam axis to cover $|\eta| < 1$, with each tube being 1.5 m in length. Each tube in the end-cap is 0.6 m long, arranged to be perpendicular to the beam axis. Each end-cap region consists of 122,880 straw tubes, covering the pseudorapidity of $1 < |\eta| < 2$.

Incoming charged particles will ionize the gas in the straw tube, leaving free electrons drifting toward the wire where the signals are amplified and read out. The radiation signal scales with the relativistic factor $\gamma = \frac{E}{m}$, implying that the detected radiation is more likely from electrons than other heavier particles, hence making it possible to identify electrons using TRT response [46].

3.2.2 Calorimeter

Calorimeters are detectors that measure the energy deposits of particle showers, which are induced by the interaction between the incoming particle and the calorimeter materials. A particle shower is a cascade of particles created by high-energy particles traversing dense matter. Through radiation and pair productions, new particles with less energy are created and eventually absorbed by the material. A schematic view of the particles' interactions with the calorimeter is shown in Figure 3.9. It is clear that photons, leptons, and hadrons deposit their energies in calorimeters through particle showers, while neutrinos and muons interact less with matter and escape the detector with minimal energy deposit.

The calorimeter systems have two components: the electromagnetic calorimeter (ECal) and the hadronic calorimeter (HCal), as shown in Figure 3.9 and Figure 3.10. The calorimeters are designed to fully absorb most incoming particles' kinetic energy to form an estimation of the original energy. The energy resolution is typically better for incident par-

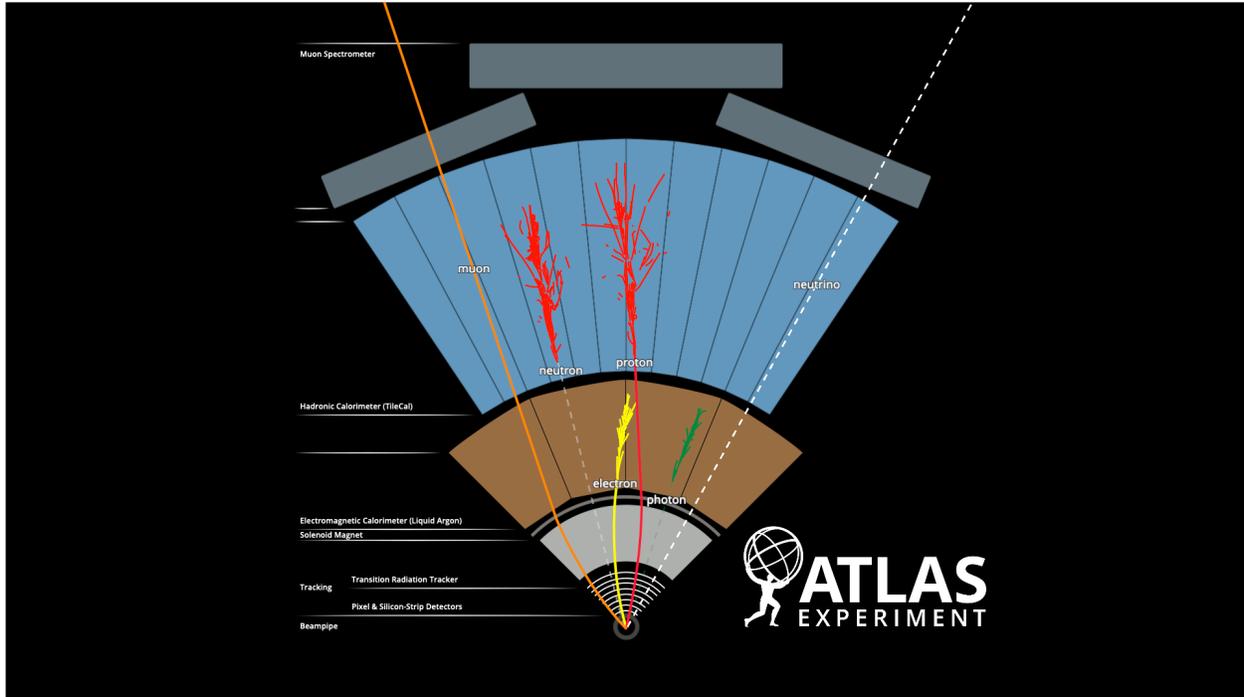


Figure 3.9: Interactions between incoming particles and different detector components of ATLAS [48].

ticles with higher energies. This improvement in resolution is due to the fact that the Poisson fluctuation in the shower is lower in percentage when the shower is larger in size and also that the readout noise is relatively smaller when the energy scale of the particle is higher. The parameterization is typically given by:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \quad (3.9)$$

where a is the constant for the term giving random fluctuations in the shower, b accounts for the readout noise, and c is an energy-dependent constant.

The ECal and the FCal are sampling calorimeters with alternating layers of dense absorbers and detecting materials. In the absorber layer, dense material is placed to interact with the incoming particle to initiate particle showers through either electromagnetic or

hadronic interactions. The shower is subsequently measured in the measuring layers, where energy deposits are converted to electronic signals. Since a large portion of the energy is deposited in the absorber, the calorimeter is effectively sampling to measure signals, hence the name sampling calorimeter. A typical parameter of calorimeters, especially for the electromagnetic calorimeter, is the radiation length (X_0), defined as the average distance over which an electron loses about 63.2% of its initial energy through Bremsstrahlung radiation. For example, the radiation length of liquid Argon is 14.2 cm [49] while that of lead is 0.5612 cm. For the hadronic calorimeter, the most relevant parameter is the interaction length, λ , which characterizes the mean distance a particle travels before going through an inelastic interaction. The interaction length of steel is typically around 17 cm.

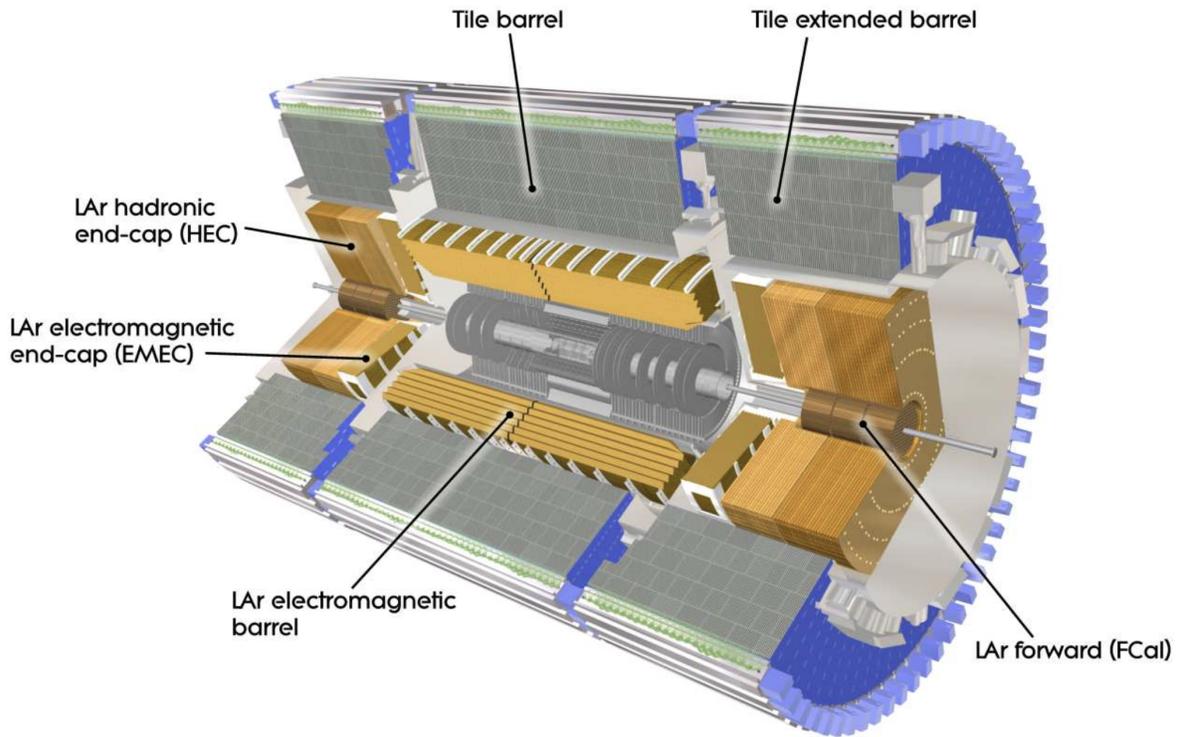


Figure 3.10: The ATLAS calorimeter systems with its submodules: the electromagnetic calorimeter and the hadronic calorimeter [25].

Electromagnetic calorimeter

The electromagnetic calorimeter is responsible for measuring most of the energy deposit of electrons and photons through electromagnetic showers. It consists of a barrel region with $|\eta| < 1.475$ and two end-cap regions with $1.375 < |\eta| < 3.2$ as shown in Figure 3.10. The barrel region contains two half-barrels with a gap of a few millimeters in between. Each half-barrel ECal has 16 equally sized modules covering half of the transverse plane. Each end-cap contains two coaxial wheels with eight modules. The larger wheel covers the pseudorapidity range of $1.375 < |\eta| < 2.5$, and the small one covers $2.5 < |\eta| < 3.2$.

The active material of the ECal is liquid Argon, while lead is used as the absorbing material, which initiates the electromagnetic shower. The ECal is designed to be symmetric in the transverse plane but with three layers in the radial direction. A cutaway view of the LAr ECal is shown on the left-hand side of Figure 3.11. The first layer of ECal consists of strip cells with a granularity of $\Delta\phi \times \Delta\eta = 0.0031 \times 0.1$ and a radial coverage of $4.3X_0$. The main goal of the first layer is to have a granularity fine enough to identify and measure the process of photonic π^0 decay, $\pi^0 \rightarrow \gamma\gamma$, with the pion's transverse momentum between 0.2 GeV and 2000 GeV [50]. The second layer of ECal comprises square cells with a granularity of $\Delta\phi \times \Delta\eta = 0.0245 \times 0.025$ and a radial length of $16X_0$. Since this layer has the largest depth, it is where most portions of the energy deposit take place. Finally, in the last layer, the segments have a much larger width in η with a granularity of $\Delta\phi \times \Delta\eta = 0.0245 \times 0.5$. The overall depth of the ECal is at least $22X_0$, which is enough to have the most high-energy electrons and photons deposit all their kinetic energy within the ECal. For the ATLAS LAr calorimeter, the energy resolution is given by [25]:

$$\frac{\sigma_E}{E_{\text{ECal}}} = \frac{10\%}{\sqrt{E}} \oplus 0.4\% \quad (3.10)$$

where the energy E is in unit of GeV.

Hadronic calorimeter

The hadronic calorimeter [51] is responsible for detecting hadrons that have escaped the electromagnetic calorimeter. The HCal consists of the Tile Calorimeter in the central barrel with $|\eta| < 1.7$ and the Hadronic End-cap Calorimeter (HEC) with $1.7 < |\eta| < 3.2$.

The active material in the tile calorimeter is plastic scintillators, while steel is used as the absorber material. Each tile calorimeter covers a radial distance of $2.28 \text{ m} < r < 4.25 \text{ m}$ with respect to the beam axis. A cutaway view of the tile calorimeter is shown on the right-hand side of Figure 3.11. As hadronic showers reach the scintillator layer, secondary particles in the shower will interact with the scintillating material to produce ultraviolet photons, which will be detected by the photo-multiplier tube (PMT). In the central barrel region, the first two layers have the granularity of $\Delta\phi \times \Delta\eta = 0.1 \times 0.1$ while the third layer has $\Delta\phi \times \Delta\eta = 0.2 \times 0.1$. The HEC uses liquid Argon as the active material and copper as the absorber to handle the high radiation environment in the high η region.

The overall resolution of the HCal is worse than that of the ECal due to the larger fluctuation of hadronic showers, which mainly consist of pions. The resolution of the ATLAS HCal is given by [52]:

$$\frac{\sigma_E}{E}_{\text{HCal}} = \frac{52\%}{\sqrt{E}} \oplus 5.7\%. \quad (3.11)$$

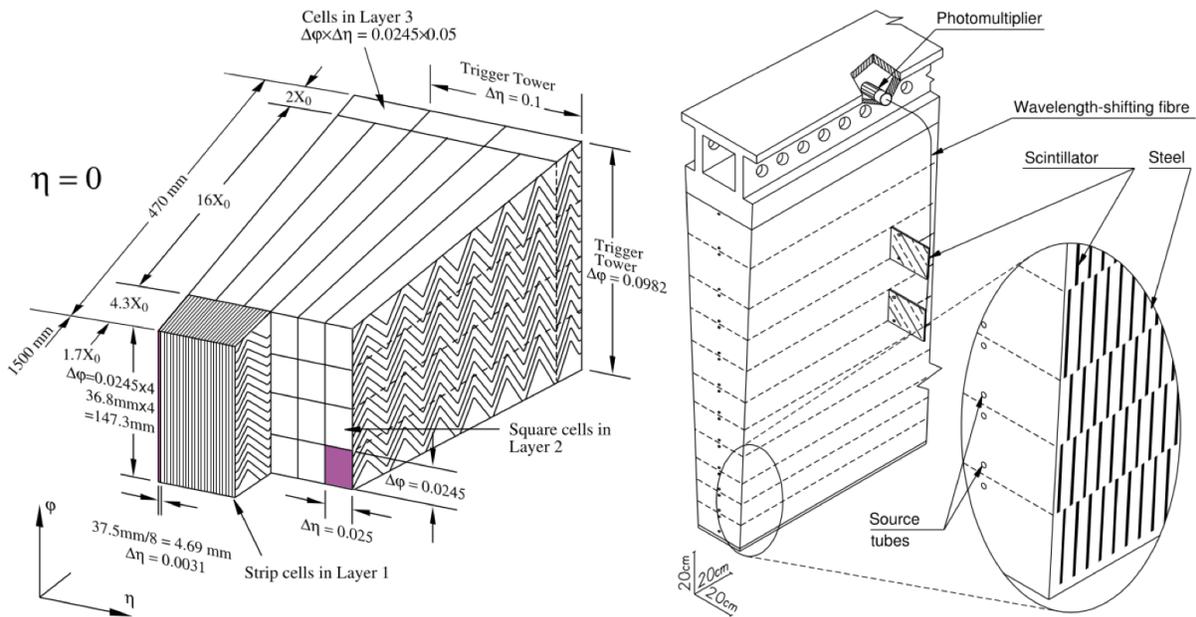


Figure 3.11: Cutaway views of the LAr electromagnetic calorimeter and the hadronic tile calorimeter tower are shown on the left and right-hand side, respectively [49].

Forward Calorimeter (FCAL)

The Forward Calorimeter (FCAL) covers the pseudorapidity region of $3.1 < |\eta| < 4.9$. To cope with the high amount of radiation in the forward region, the FCAL also uses liquid Argon as the active material. In the first layer, copper is used as the absorber to provide good resolution. In the next two outer layers, tungsten is used as the absorbing material to contain the hadronic shower because of its higher density. The forward calorimeter plays an important role in capturing particles in the forward region. As a result, the calorimeter helps determine the overall event energy. The implementation of the forward calorimeter leads to better calculations of the missing transverse momentum, which is an important physics analysis quantity defined in Section 4.7.

3.2.3 Muon Spectrometer

As introduced in Section 2.1.1, the muon particle is a minimum ionizing particle, which leaves negligible energy deposit in the calorimeter as it travels through. Since muons are charged particles, they could be measured by the inner detector of ATLAS. However, when the muon's transverse momentum is above the level of 100 GeV, the track of it in the inner detector is nearly a straight line, which yields poor resolutions in the momentum measurements. Therefore, to identify high-energy muons, the Muon Spectrometer (MS) [53] is placed outside the calorimeter, which absorbs all high-energy particles except for muons and neutrinos.

Like the inner detector, the muon spectrometer measures the trajectories of muons. The MS consists of a barrel region with $|\eta| < 1.4$ and two end-cap regions with $1.6 < |\eta| < 2.7$. The barrel region provides a strong magnetic field of 1 T using the barrel toroidal magnet. In contrast, the two end-cap regions have magnetic fields of lower magnitudes (0.5 T) provided by the two end-cap toroidal magnets. The barrel region has three concentric cylindrical layers at radial distances of 5 m, 7.5 m, and 10 m. The end-cap region has four disks centered at positions along the beam axis with measured $|z|$ coordinates of 7.4 m, 10.8 m, 14 m, and 21.5 m. The muon spectrometer consists of four subsystems, each designed with a specific role in either tracking measurements or muon triggers. The overall resolution of the MS is about 3% – 10% for momenta between 10 GeV and 1000 GeV [53].

A cutaway view of the muon spectrometer is shown in Figure 3.12 with all four subsystems: the Monitored Drift Tubes (MDTs), the Cathode Strip Chambers (CSCs), the Resistive Plate Chambers (RPCs) and the Thin Gap Chambers (TGCs).

The Monitored Drift Tubes

The MDTs [54] are drift tube chambers that cover the pseudorapidity range of $|\eta| < 2.7$. Each detector consists of a 30 mm cathode tube filled with 93%/7% mixture of Argon/CO₂

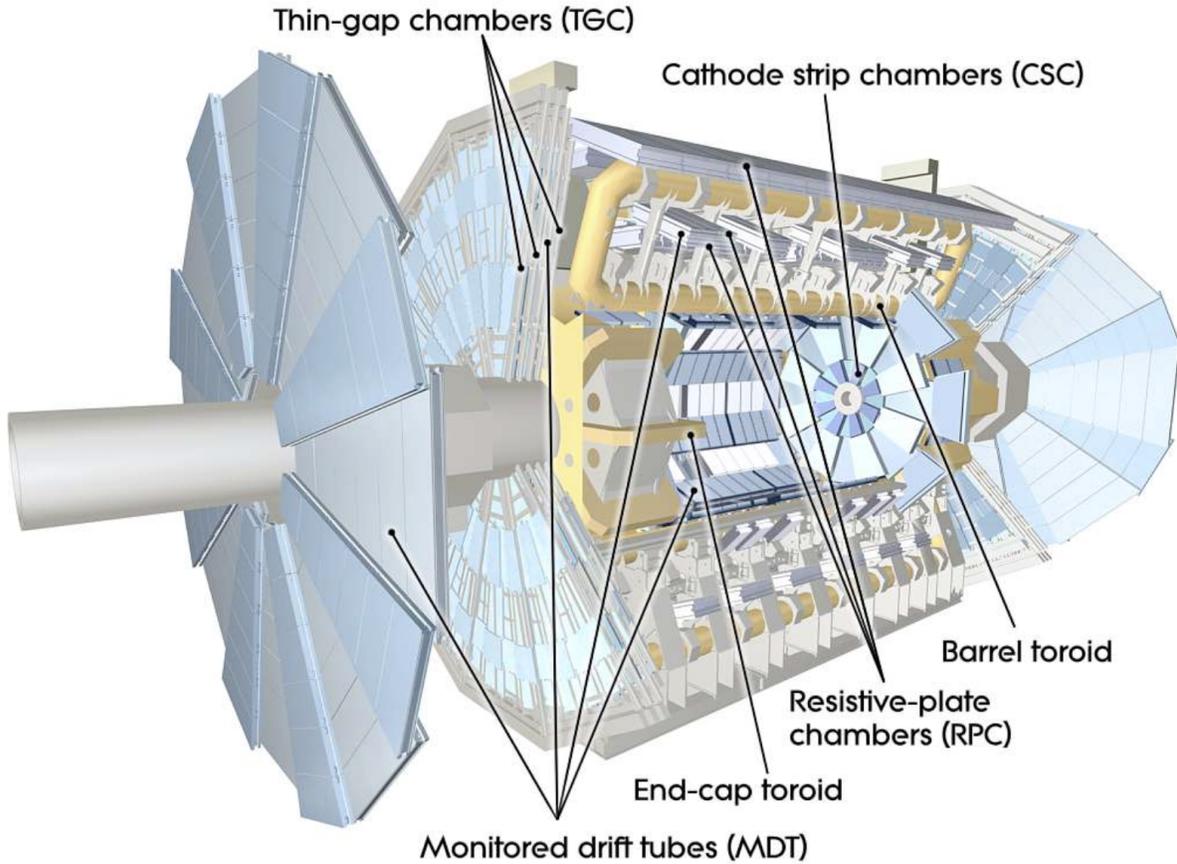


Figure 3.12: The ATLAS muon spectrometer [25].

gas and an anode wire placed at the center of the tube. The tungsten-rhenium anode wire inside the tube has an operating voltage of 3 kV. As muons traverse the tube, ionized electrons from the Argon gas will drift toward the anode wire with a maximum drift time of around 700 ns [25]. The position resolution of an MDT chamber ranges from 60 μm to about 80 μm [25].

The Cathode Strip Chambers

The CSCs [55] are multi-wire proportional chambers used in the forward region to handle the high particle flux, covering the pseudorapidity range of $2 < |\eta| < 2.7$. Each chamber operates at a voltage of 1.9 kV with Argon/ CO_2 gas inside the chamber. As

the charged particle passes through the gas, ionized electrons form an avalanche on the anode wire, which is read out to yield a precise position measurement with a resolution around $60\ \mu\text{m}$ [25].

The Resistive Plate Chambers

The Resistive Plate Chambers [56] are installed in the barrel region with $|\eta| < 2$ to provide muon triggers. Each RPC consists of two parallel plates with the 2 mm gap between them filled with a 94.7%/5%/0.3% mixture of $\text{C}_2\text{H}_2\text{F}_4$ - C_4H_{10} - SF_6 gas. The parallel plates operate at a voltage of 9.8 kV to provide a strong and uniform electric field in the gap. The spatial resolution of the RPCs is only about 1 cm, which is much worse than the MDTs. However, the signal from ionized electrons will be read out much faster, typically within 2 ns [57]. This motivates the usage of RPCs for muon trigger systems.

The Thin Gap Chambers(TGCs)

Similar to the RPCs, TGCs [58] are multi-wire chambers installed in the forward region. The primary purpose of the TGCs is to provide fast triggering and coordinate measurement of muons in the MS end-caps. Each TGC consists of $50\ \mu\text{m}$ gold-plated tungsten wires with a 1.8 mm spacing. The two-chamber planes have a gap of 2.8 mm, with 9 mm strips behind the cathode planes to read out signals. The gap is filled with a 55%/45% mixture of CO_2 /n-pentane gas at an operating voltage of 3.0 kV.

3.2.4 Trigger

The ATLAS trigger and data acquisition (TDAQ) system identifies and selects interesting physics events in real-time and saves them for offline processing. The ATLAS trigger system is a two-level trigger system that reduces the recorded event rate to about 1 kHz

from the collision rate at 40 MHz [59]. A schematic diagram of the ATLAS TDAQ system is shown in Figure 3.13.

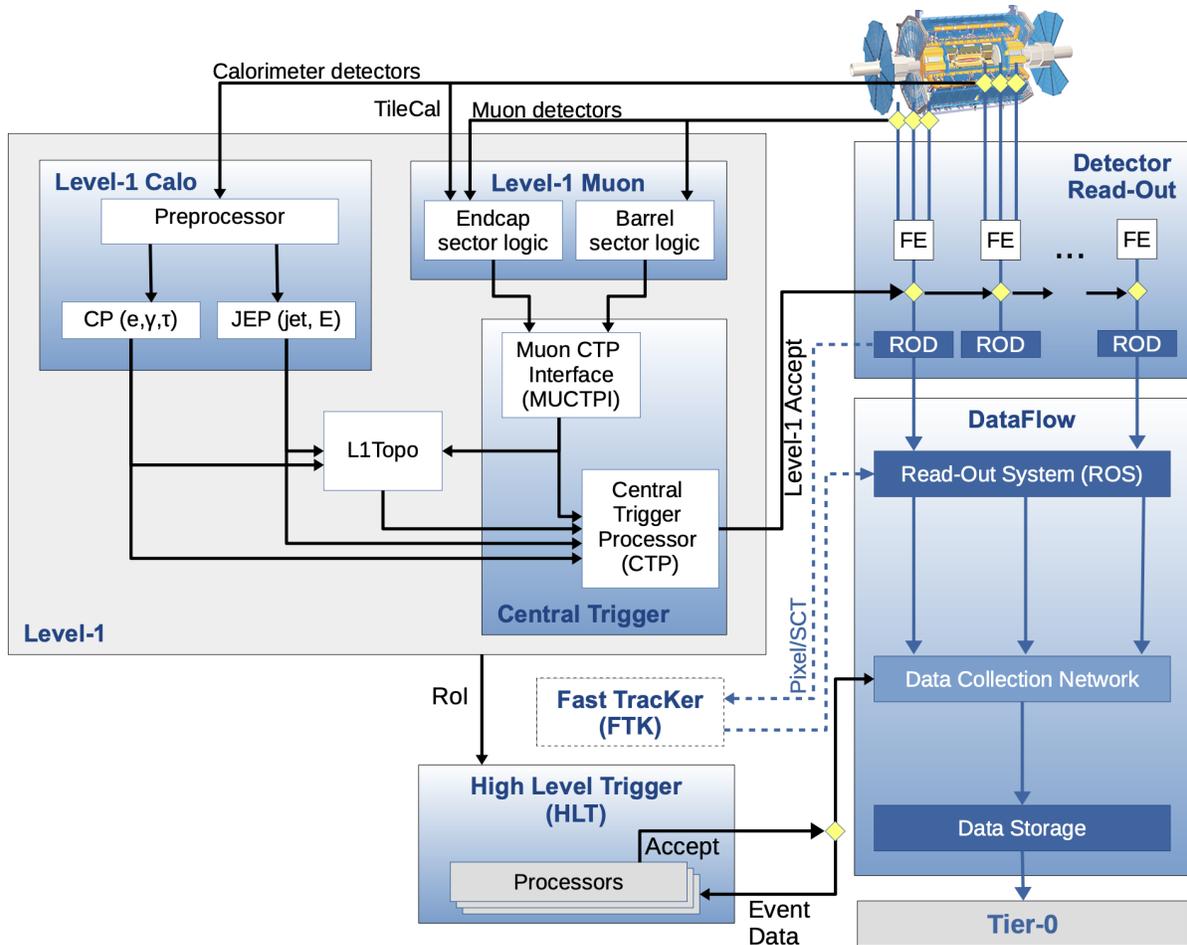


Figure 3.13: The ATLAS trigger and data acquisition system [59].

The level-1 (L1) triggers are hardware-based triggers in calorimeters and muon spectrometers. The L1 calorimeter trigger takes analog calorimeter information and feeds it into dedicated preprocessors to identify physics objects for event selections. The L1 muon trigger uses information collected by the RPCs in the barrel region and the TGCs in the end-cap region to trigger muon events as introduced in Section 3.2.3. As shown in Figure 3.13, all L1 information is directed toward the Central Trigger Processor to make the final decision for level-1. Once the decision is reached to accept the event, the front-end detec-

tor electronics will read out all detector data of the current event and send them to the readout driver for processing. The event information is then sent to the readout system to buffer the data. The L1 trigger accepts events at a rate of 100 kHz to feed them to further examinations under High-Level triggers (HLT).

The HLT is a software-based triggering system that further reduces the event rate down to about 1 kHz. The offline triggers are typically algorithms based on the Athena software [25], which is used in offline analysis. Dedicated Processing Units are used to execute these algorithms to provide rejections before data storage.

At this point, the LHC and the individual components of the ATLAS detector have been introduced. The physics data in this analysis consists of physics objects reconstructed from the raw data collected by the ATLAS detector. The following chapter will discuss various object reconstruction algorithms used in the ATLAS experiment.

Chapter 4

Object Reconstruction

After recording information from the detectors, physics objects must be reconstructed for the physics analysis. This chapter discusses various reconstruction algorithms for all physics objects used in this analysis. First, Section 4.1 will go through the reconstruction of intermediate objects, such as tracks, vertices, and clusters, using detecting signals. Then, the reconstruction of various particles is discussed in Section 4.2 (electrons), Section 4.3 (photons), and Section 4.4 (muons). In Section 4.5, particle jets and their reconstruction algorithm are introduced. The overlap removal algorithm for all aforementioned particles is discussed in Section 4.6. Section 4.7 introduces the missing transverse momentum, the reconstruction of which relies heavily on the reconstructions of all aforementioned particles. Finally, the reconstruction of neutrino using the missing transverse momentum is introduced in Section 4.8.

4.1 Tracks and Clusters

4.1.1 Tracks

Tracks are the trajectories of charged particles measured by the ATLAS inner detector. The track reconstruction algorithm has four stages: clustering, iterative combinatorial track finding, ambiguity solving, and track fitting [60].

Clustering

When energy deposits in sensor components exceed a pre-defined threshold, the pixels, and strips are grouped into clusters using connected component analysis [61]. The three-dimensional space-points are subsequently created using these clusters to represent the trajectory of a charged particle as it travels through the inner detector. A cluster could be a single-particle or a merged cluster, depending on whether the cluster corresponds to an energy deposit from one or multiple charged particles.

Iterative combinatorial track finding

From combinations of track clusters, track seeds are formed to give possible track candidates. Then, space-points are added to the track candidates using a Kalman filter [62]. This results in high primary particle reconstruction efficiencies, which could exceed 99% for muons.

Ambiguity solving

The ambiguity solver assigns a track score to each track candidate formed by track finding, with higher scores indicating a more plausible track. The track score is based

on subdetector resolutions, χ^2 fit of the track, and holes, which are intersections of trajectories with detectors without matching clusters. Finally, a track must pass all of the following criteria to enter the next stage of the reconstruction algorithm [60]:

- $p_T > 400$ MeV.
- $|\eta| < 2.5$.
- Minimum of 7 pixels and SCT clusters.
- Maximum of either one shared pixel cluster or two shared SCT clusters on the same layer.
- No more than two holes in the combined pixel and SCT detectors.
- No more than one hole in the pixel detector.
- $|d_0^{\text{BL}}| < 2.0$ mm.
- $|z_0^{\text{BL}} \sin(\theta)| < 3.0$ mm.

where d_0^{BL} is the transverse impact parameter with respect to the beam, θ is the polar angle of the track, and z_0^{BL} is the longitudinal distance between the point where d_0^{BL} is measured and the primary vertex, which is defined in Section 4.1.2.

Track fit

All track candidates that passed the ambiguity solver are fitted using a high-resolution fit at this stage. Then, additional neural networks determine each cluster's position and its uncertainty. Finally, the resulting fitted tracks are added to the final collection of tracks.

4.1.2 Vertices

Multiple collisions will occur along the beam axis in each bunch crossing of proton beams, which produces tracks of charged particles. A vertex is a physical location from which a group of tracks originates. Primary vertices are defined as the locations of inelastic interactions. In the run-2, there is typically no more than one such collision per bunch crossing. The primary vertex in each bunch crossing is the hard-scatter vertex, defined as the vertex with the highest sum of p_T^2 over all associated tracks. Other vertices in the same bunch crossing are called pile-up vertices, which represent locations of pile-up interactions. Each vertex is required to have at least two associated tracks that satisfy the following cuts [63]:

- $p_T > 400$ MeV.
- $|\eta| < 2.5$.
- Number of silicon hits ≥ 9 if $|\eta| \leq 1.65$.
- Number of silicon hits ≥ 11 if $|\eta| \leq 1.65$.
- IBL hits + B-layer hits ≥ 1 .
- A maximum of 1 shared module (1 shared pixel hit or 2 shared SCT hits).
- Pixel holes = 0.
- SCT holes ≤ 1 .

A vertex seed is first estimated to determine a vertex position. Then, tracks and the seed are iteratively fitted to estimate a better vertex position, with incompatible tracks being down-weighted to the end of the track list in each iteration [64]. After the vertex position is determined, all incompatible tracks are removed and used to fit new vertices. This process is repeated in order to find all vertices in the event.

4.1.3 Calorimeter Clusters

The energy of each particle shower is typically deposited across multiple cells in the calorimeter and thus requires clustering algorithms to group adjacent cells to reconstruct the incident particle. In the ATLAS experiment, two main clustering algorithms are used for the calorimeters [65]: the sliding-window and topological algorithms.

Sliding-window Algorithm

The sliding-window algorithm is primarily used to reconstruct electrons, photons, and tau leptons. The electromagnetic calorimeter is divided into grids in the $\eta - \phi$ plane for the window to slide on. Each grid has a fixed dimension of 0.025×0.025 in the $\eta - \phi$ plane, yielding 200×256 calorimeter towers.

A window with a fixed size in terms of calorimeter towers ($N_{\text{window}}^{\eta} \times N_{\text{window}}^{\phi} = 5 \times 5$) scans through all towers to find positions where the local energy deposit within the window exceeds the pre-defined threshold of 3 GeV. Clusters of particles are then built in that region with different sizes based on the hypothesized particle types [65]:

Particle type	Barrel ($N_{\text{cluster}}^{\eta} \times N_{\text{cluster}}^{\phi}$)	End-cap ($N_{\text{cluster}}^{\eta} \times N_{\text{cluster}}^{\phi}$)
Electron	3×7	5×5
Converted photon	3×7	5×5
Unconverted photon	3×5	5×5

Table 4.1: Cluster sizes for different particles in the EM calorimeter [65]. Converted photons are those that go through pair production before reaching the calorimeter.

Topological algorithm

The topological algorithm builds topoclusters of variable sizes using areas where energy deposits are significantly higher than the expected noise fluctuation as seeds. The

main variable used in the algorithm is the signal significance ζ_{cell}^{EM} which is the ratio of the signal energy deposit in the cell and the average noise:

$$\zeta_{cell}^{EM} = \frac{E_{signal}^{EM}}{\sigma_{noise,cell}^{EM}}. \quad (4.1)$$

The proto-cluster seeds are taken to be cells that satisfy $|\zeta_{cell}^{EM}| > 4$. Neighboring cells of each proto-cluster are subsequently merged into the cluster if the energy deposit in the cell satisfies $|\zeta_{cell}^{EM}| > 4$.

The signal development time is typically much larger than the 25 ns interval between consecutive bunch crossings. To shorten the signal development, the pulse is shaped to have a sharp peak and then decreases to negative values, which gives an overall amplitude integral of 0. A plot of the signal development is shown in Figure 4.1. Negative energy cells are typically from out-of-time pile-up interactions from consecutive bunch crossings. In the topological algorithm, absolute values of significance are used to include cells with negative energies to avoid bias from only including cells with positive energies [66].

4.2 Electrons

A schematic diagram of the electron's path to the calorimeter is shown in Figure 4.2. As electrons travel through the calorimeter, they initiate particle showers made of positrons, electrons, and photons. These shower particles are typically very collimated within the same cluster. Each electron candidate requires a charged particle's track measured by the inner detector and a localized energy deposit in calorimeter clusters. The track and cluster must also be close to each other in the $\eta - \phi$ plane.

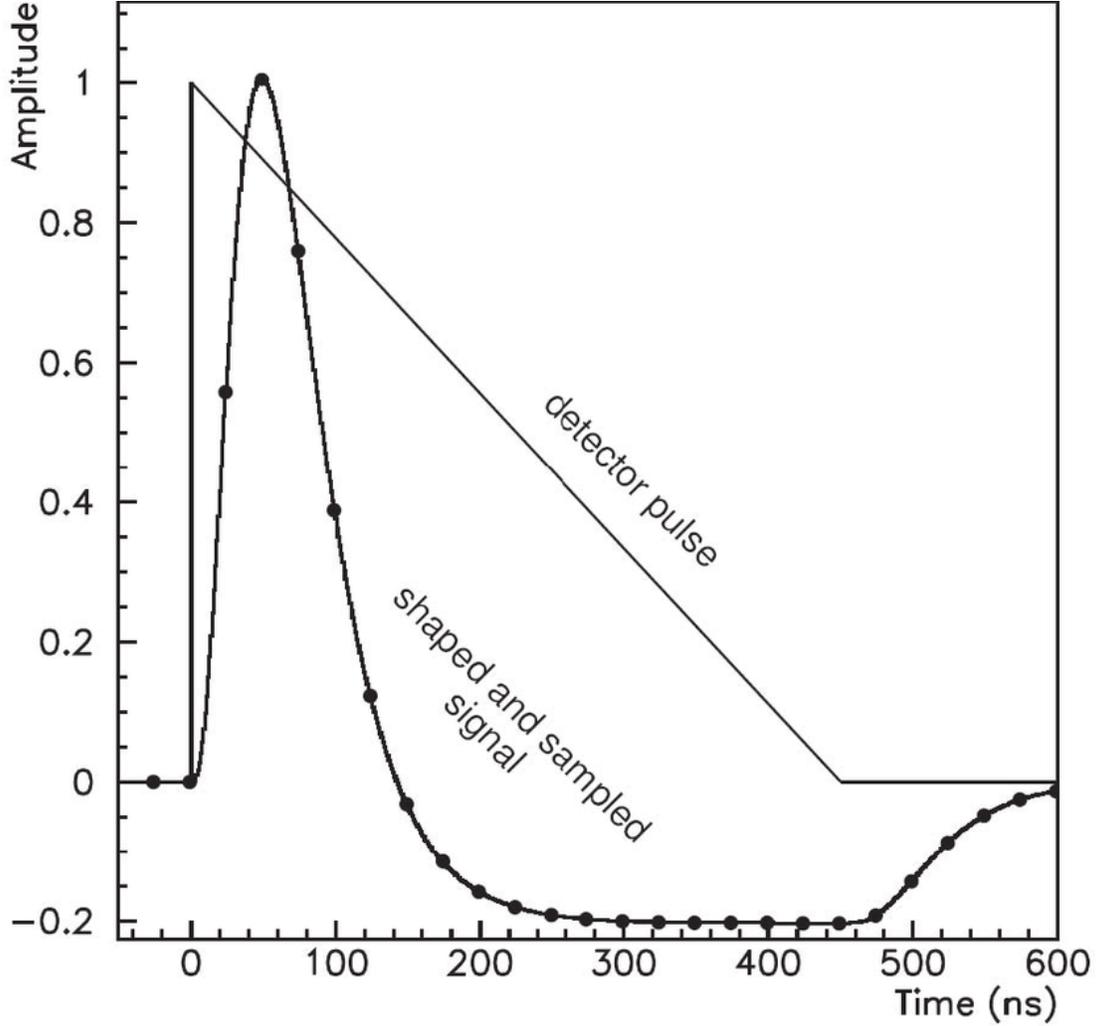


Figure 4.1: Shapes of the LAr calorimeter pulse [67].

The Gaussian Sum Filter [69] is applied to tracks and clusters to correct the energy loss of charged particles' interaction in the material. The electron track and its cluster are required to satisfy $|\eta_{\text{cluster}} - \eta_{\text{track}}| < 0.05$ and $-0.1 < \phi_{\text{cluster}} - \phi_{\text{track}} < 0.05$ [68]. After this step, a dedicated algorithm will be executed to assess the photon/electron ambiguity based on quantities measured by the inner detector. Finally, reconstructed clusters are built with a window size of 3×7 in the barrel region and 5×5 in the end-cap region, as shown in Table 4.1. The ATLAS collaboration has four likelihood-based working points for electron identification: VeryLoose, Loose, Medium, and Tight. A plot of reconstruction efficiencies of the working points, except for VeryLoose, is shown in Figure 4.3.

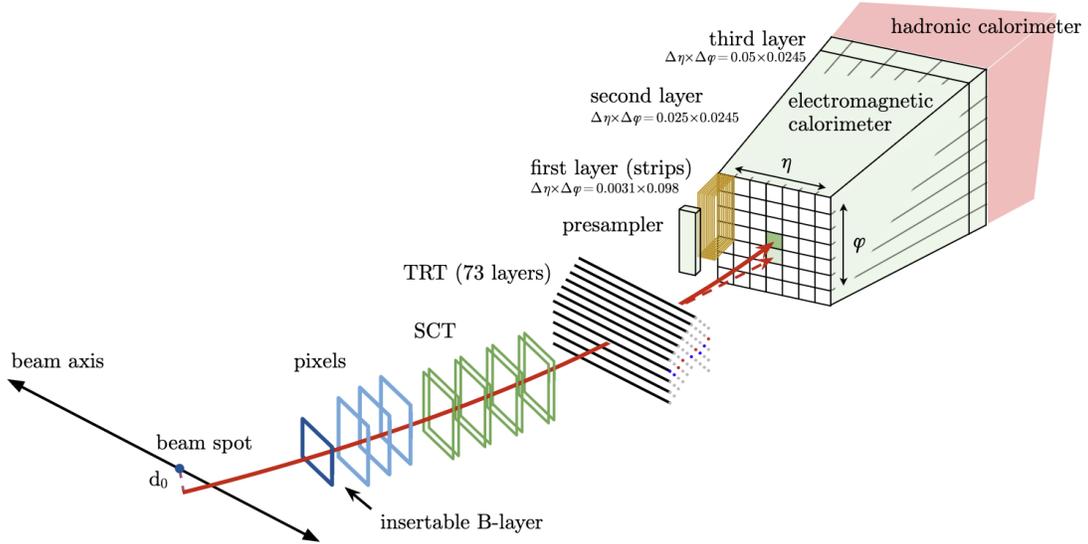


Figure 4.2: A schematic diagram of an electron's path through the ATLAS detector [68]. The solid red line represents the electron trajectory through the inner detector to the calorimeter. The dashed red line represents a photon produced from interactions with the detecting material.

Another working point for electrons is the isolation working point, which is based on energy deposit within a cone of $\Delta R = 0.2$ around the electron's energy deposit in the electromagnetic calorimeter. The sum of transverse energy deposits within the cone gives the variable *topoetcone20*. As for the tracks, the transverse momentum of all tracks within the cone will contribute to the variable *ptvarcone20*. For the isolation working point to be passed, the ratio of an electron's *topoetcone20* or *ptvarcone20* to its transverse momentum must be lower than certain cuts. Two electron isolation working points criteria are used in this analysis, FixedCutLoose and FixedCutTight, and their respective selection criteria are listed in the following.

For the calorimeter isolation, the following working points are used for each isolation:

- FixedCutLoose: $\frac{\text{topoetcone20}}{p_T} < 0.2,$

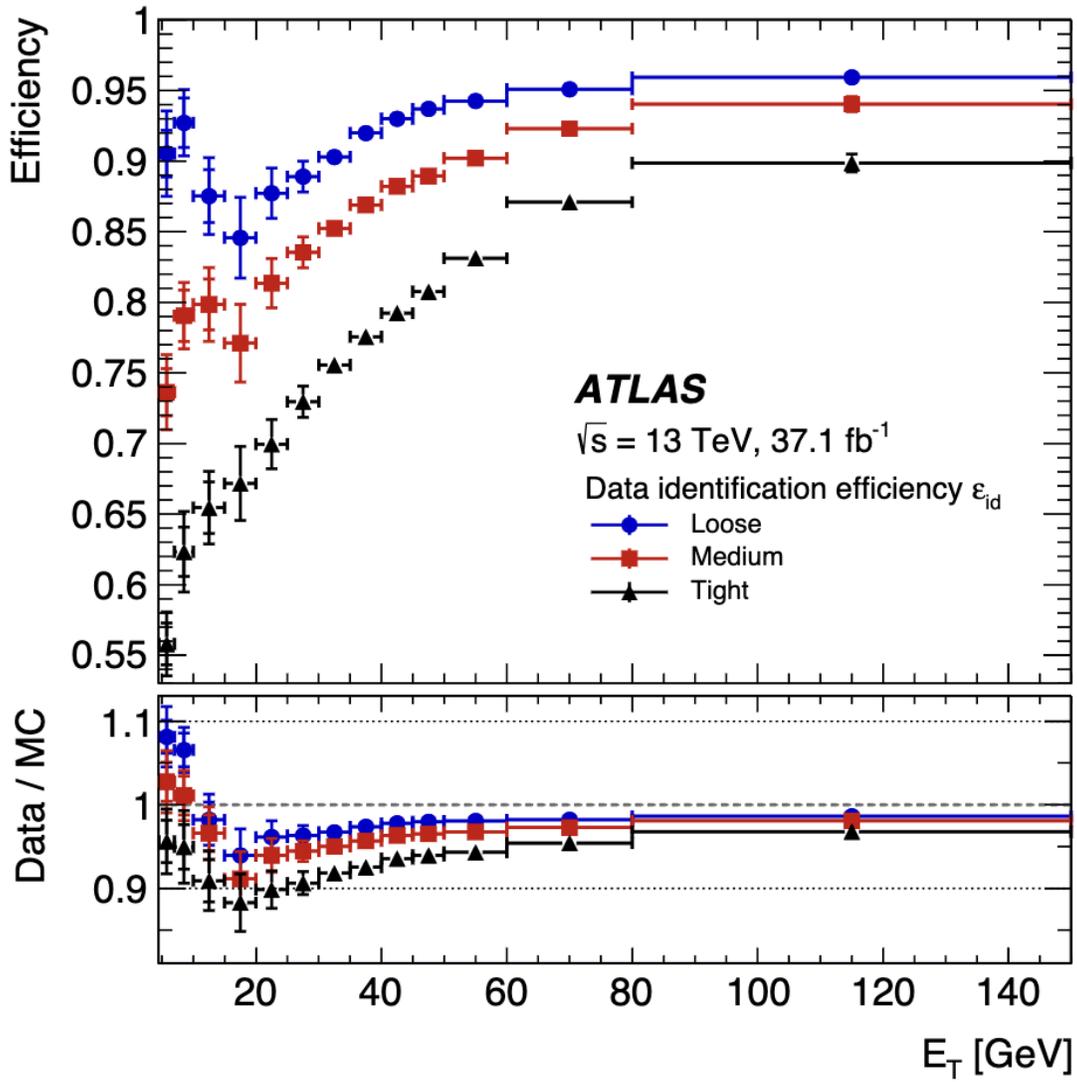


Figure 4.3: Measured electron reconstruction efficiencies with respect to E_T in $Z \rightarrow ee$ events [68].

- FixedCutTight: $\frac{topoetcone20}{p_T} < 0.06$.

For the track isolation, the following working points are used for each isolation:

- FixedCutLoose: $\frac{ptvarcone20}{p_T} < 0.15$,
- FixedCutTight: $\frac{ptvarcone20}{p_T} < 0.06$.

This analysis defines a "loose" electron as an electron candidate that passes the Loose-AndBLayerLH (Loose) identification working point and the FixedCutLoose isolation requirements. A "tight" electron is defined as a candidate that passes the TightLH (Tight) identification and the FixedCutTight isolation working points. All electron candidates in the crack region with a pseudorapidity range of $1.37 < |\eta| < 1.52$ are rejected. To ensure that the electron is associated with the primary vertex, the candidate is required to have a longitudinal impact parameter of $|z_0 \sin(\theta)| < 0.5$ mm and a transverse impact parameter d_0 lower than 5.0 mm.

4.3 Photons

The reconstruction of a photon depends primarily on whether the photon is converted or not. A converted photon is a photon that has converted into an electron-positron pair before reaching the electromagnetic calorimeter, while an unconverted photon is the opposite. In order to reconstruct converted photons, tracks with silicon hits and tracks reconstructed only in the TRT are used to reconstruct the secondary vertex. Candidate clusters with no matched vertices are considered to be unconverted photons. However, there still exists a photon/electron ambiguity. In the last step, arbitration is performed to finalize the particle type based on track hits, momentum, and E/p measurements. The reconstruction efficiency of photons passing tight identification working point is measured to be 45% - 60% at $E_T = 10$ GeV to 95% - 98% for $E_T > 100$ GeV [70].

The photon identification working points are similar to those defined for the electrons. However, due to the lack of tracks, photon identification is mainly based on calorimeter variables, such as shower shape, calorimeter energy deposit, and hadronic leakage. Combined with the photon candidate's $|\eta|$ and E_T , these variables give the following identification working points: Loose, Medium, and Tight.

The isolation working point of the photon is defined using the *topoetcone* variable, which characterizes the energy deposit around the photon candidate within a cone of a fixed radius. The main variables used are *topoetcone20*, which corresponds to the sum of the transverse energy of topo-clusters within the cone of $\Delta R < 0.2$, and *topoetcone40*, which corresponds to the sum of the transverse energy of topo-clusters within the cone of $\Delta R < 0.4$. In addition, all tracks within that same cone will have their transverse momenta contribute to the variable *ptvarcone*. In this analysis, two working points are considered: FixedCutLoose and FixedCutTight, which are defined by cutting the above-mentioned variables.

For the calorimeter energy deposit, the following is required:

- FixedCutLoose: $topoetcone20 < 0.065 E_T$,
- FixedCutTight: $topoetcone40 < 0.022 E_T + 2.45 \text{ GeV}$.

Moreover, for the track requirements, both the FixedCutLoose and FixedCutTight working points require $\frac{ptcone20}{E_T} < 0.05$, where *ptcone20* corresponds to the sum of transverse momenta of tracks within the cone of $\Delta R < 0.2$.

A "loose" photon is defined as a photon candidate that passes the Loose identification working point and the FixedCutLoose isolation requirement. Likewise, a "tight" electron is a candidate that passes Tight identification and FixedCutTight isolation requirements.

4.4 Muon

The muon reconstruction combines the reconstructions in the MS and ID, which are independently performed. In the ID, the muon is reconstructed in the same way as other charged particles.

MS reconstruction

In the MS, the first step of the reconstruction is to search for hit patterns in muon chambers to form segments. The hits from segments in different layers are subsequently fitted to create muon track candidates using a segment-seeded combinatorial search [71]. At least two matching segments are needed to build a track unless the track is in the barrel-endcap transition region, where one segment would suffice. Next, hits from each track candidate are fitted using a χ^2 fit. The candidate is accepted if the χ^2 value of the fit satisfies pre-determined criteria.

ID - MS combined reconstruction

After track candidates are built in the MS, the combined reconstruction uses ID, MS, and calorimeter inputs. There are four types of muons, the definition of which are based on the subdetectors used in the reconstruction [71]:

- Combined muon:

A combined muon uses a combined track fitted with hits from the ID and MS subdetectors. Muons are typically reconstructed using an Outside-in method, in which the candidates are reconstructed in the MS first and then extrapolated to the ID to match the ID track. Sometimes, the procedure is reverted to an Inside-out method, in which the muon is extrapolated outward to the MS from the ID. The Inside-out is used as a complementary method to the main combined muon's Outside-in method.

- Segment-tagged muons:

A Segment-tagged muon is a muon reconstructed from a track in the ID that is matched to at least one local track segment in the MDT or CSC subdetectors. The Segment-tagged muon is used for muons that traverse only one layer of MS chambers.

- Calorimeter-tagged muons:

The Calorimeter-tagged muons have their ID tracks matched to calorimeter energy deposits compatible with minimum ionizing particles. The identification scheme for this type of muons is optimized for the region of $|\eta| < 0.1$ and $15 \text{ GeV} < p_T < 100 \text{ GeV}$.

- Extrapolated muons:

Extrapolated muons are reconstructed based only on MS tracks and must be loosely compatible with the hypothesis that they originate from the interaction point. This is done in order to extend the muon reconstruction to the region of $2.5 < |\eta| < 2.7$, which is not covered by the ID.

The muon identification uses precision stations defined as MS stations, where muon candidates record at least three hits in the MDT or the CSC. In contrast to precision stations, a precision hole station is an MS station where the candidate records less than three hits in the MDT or CSC.

There are three muon working points: Loose, Medium, and Tight. Their selection criteria are listed as the following:

- Medium:
 - Combined Muon and Inside-out Muon reconstruction methods.
 - At least two precision stations for candidates with $|\eta| > 0.1$.
 - One precision station and no more than one precision hole station for candidates with $|\eta| < 0.1$.
- Loose:
 - Passes the Medium working point.

- Includes Calorimeter-Tagged and Segment-Tagged muons for candidates with $|\eta| < 0.1$.
- Low- p_T muons with $|\eta| < 1.3$ reconstructed by the Inside-out method are accepted if they are confirmed by the Segment-Tagged reconstruction method.
- Tight:
 - Combined Muons and Inside-out muons.
 - At least two precision stations.
 - Restrictions placed on the combined track fit χ^2 value.
 - Restrictions placed on transverse momenta measurements' similarity between measurements in ID and MS.

The isolation working point of the muon is similar to that of the electrons, for which the energy deposits within a cone are summed. The primary variable for calorimeter energy deposit is *topoetcone20*, which denotes the energy deposits within $\Delta R < 0.2$ of the candidate, and the transverse momenta of the track particles within the cone of $\Delta R < 0.3$ are summed into *ptvarcone30*.

The two isolation working points are FixedCutLoose and FixedCutTight. Their calorimeter isolation cuts are defined as the following:

- FixedCutLoose: $\frac{topoetcone20}{p_T} < 0.3$,
- FixedCutTight: $\frac{topoetcone20}{p_T} < 0.06$.

Furthermore, the track isolation cuts are:

- FixedCutLoose: $\frac{ptvarcone30}{p_T} < 0.15$,
- FixedCutTight: $\frac{ptvarcone30}{p_T} < 0.06$.

The "loose" muon is defined as a muon candidate that passes the Loose muon quality working point and the FixedCutLoose isolation criteria. A "tight" muon is a candidate that passes the Tight muon quality cuts and the FixedCutTight isolation criteria. In order to reject cosmic ray muons, the muon candidates must have a longitudinal impact parameter requirement of $|z_0 \sin(\theta)| < 0.5$ mm to ensure their association with the primary vertex.

4.5 Jets

Jets are collimated sprays of particles from the hadronization process of quarks or gluons produced in particle collisions. The reconstruction of a jet is more complicated than that of other particles due to its finite area. This section discusses the jet reconstruction algorithm used in the ATLAS experiment (Section 4.5.1) and the b-tagging algorithms for identifying b-jets (Section 4.5.2).

4.5.1 Jet Reconstruction

The constituents of jets are calorimeter topoclusters introduced in Section 4.1.3. Since a jet occupies a finite area in the calorimeter, its reconstruction algorithm requires grouping topoclusters. The jets reconstructed solely from calorimeter topoclusters are called EMtopojets. Information from the ID could also be used in the reconstruction to form Particle Flow (PFlow) jets.

EMtopojets

The most general k_t algorithm [72] uses topoclusters' four vectors as inputs and groups them based on the following distance measures:

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

$$d_{iB} = k_{ti}^{2p},$$
(4.2)

where $\Delta_{ij}^2 = (y_i - y_j)^2 - (\phi_i - \phi_j)^2$ is the distance between two clusters in the rapidity-azimuthal ($y - \phi$) plane, and k_{ti} is the transverse momentum of the cluster. For jets used in this analysis, the radius R is taken to be 0.4 in the $\eta - \phi$ plane.

In the ATLAS experiment, EMtopojets are reconstructed from topoclusters using the anti- k_t jets clustering algorithm, defined by setting $p = -1$ in Equation 4.2. For each cluster i , the algorithm computes d_{ij} by looping through all other clusters j and computes d_{iB} . If $d_{ij} < d_{iB}$, the algorithm combines the two clusters as constituents from the same jet and removes the j -th cluster from the list. If $d_{ij} > d_{iB}$, the i -th cluster is considered as a jet and hence removed from the list of clusters. This process is repeated until no more clusters are on the list. An example of the anti- k_t algorithm is shown in Figure 4.4.

PFlow jets

The PFlow algorithm [73] uses ID information to reconstruct charged hadrons while retaining calorimeter cluster information for neutral particles. The algorithm removes energy deposit overlaps between tracks and calorimeter clusters to combine both to reconstruct jets, producing the so-called PFlow jets. The PFlow workflow could be summarized as the following [73]:

1. The algorithm starts with well-measured tracks and calorimeter clusters.
2. Tracks and clusters are matched based on their relative distance in the $\eta - \phi$ plane defined by $\Delta R' = \sqrt{(\frac{\Delta\eta}{\sigma_\eta})^2 + (\frac{\Delta\phi}{\sigma_\phi})^2}$, where σ_η and σ_ϕ are cluster widths along the two angular directions, respectively.

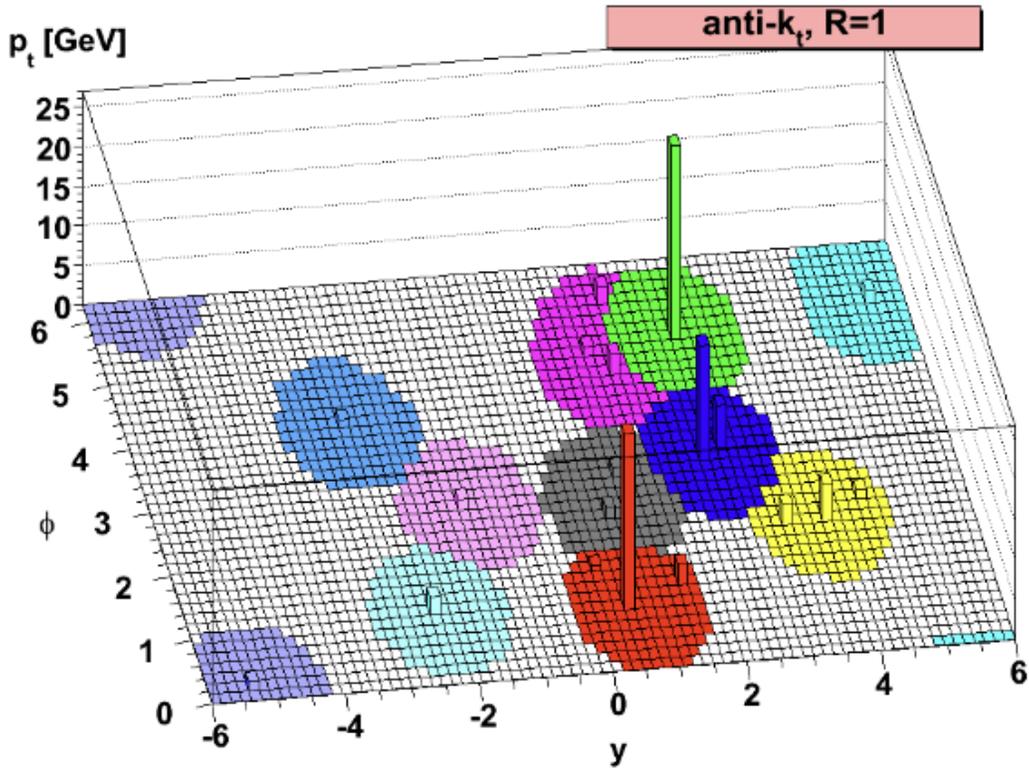


Figure 4.4: An example of the anti- k_t algorithm reconstructing jets at the parton-level [72].

3. Expected calorimeter energy deposit of the track particle is computed based on its position and momentum.
4. The algorithm assesses the situation and decides if more clusters need to be added to fully contain the energy deposit from the shower of the incident particle.
5. The expected energy deposit from the calorimeter is subtracted on a cell-by-cell basis from all matched clusters.
6. The remnant clusters are removed if consistent with the particle shower fluctuation.

The PFlow jets are reconstructed with the anti- k_t algorithm using remaining topoclusters and selected tracks matched with the primary vertex by requiring $|z_0 \sin \theta| < 2$ mm.

Jet working points

Jet candidates must pass the *BadLoose* [74] quality criteria to remove jets that are not associated with real calorimeter energy deposits. This effect could be due to detector issues, LHC beam issues, and cosmic ray particles.

Jets are also required to pass the Jet Vertex Tagger (JVT) [75] requirement in order to reject jets from pile-up interactions. The JVT algorithm grades the likelihood of each jet being a pile-up jet in a single score. There are three working points: Loose, Medium, or Tight. This analysis uses the Tight working point, which cuts the JVT score at $JVT > 0.7$ to operate at a 90(95)% efficiency for hard-scatter jets with transverse momentum in the range of 20-30 GeV (30-50 GeV).

4.5.2 B-tagging

The B-tagging algorithm [76] aims to identify b-jets, which are jets coming from decays of b-quarks. Due to the longer lifetime of the massive b-quark, the tracks from b-jets could typically be traced back to a displaced secondary vertex in the ID.

Various algorithms are implemented in ATLAS to tag b-jets, such as the IP2D/IP3D algorithm [77], SV1 algorithm [78], the boosted decision tree discriminant MV2 [77] and the DL1 [76] algorithm based on deep feed-forward neural network. In this analysis, the b-tagging task relies on the DL1 algorithm, which outputs a multidimensional array representing the probability of the candidate jet being a b-, c-, or light-flavor jet. This analysis uses the working point of DL1 at 77% efficiency.

4.6 Overlap Removal

It is common in collision experiments to observe signals that are close to one another in the $\eta - \phi$ plane. This requires the reconstruction algorithm to prioritize certain types of particles and remove the overlapping candidate. In the ATLAS experiment, the standard working point of overlap removal between electrons, muons, photons, and jets is achieved by the `OverlapRemovalTool` algorithm [79].

The algorithm first considers all electrons and muons in the final state and removes electrons that share common tracks with muons. For photons, if any lepton or jet falls within the cone of $\Delta R < 0.4$ in the $\eta - \phi$ plane around a photon, the photon is kept, and the other particle candidate is removed. The last step is the removal of lepton and jets. For electrons, the jet is removed, and the electron is kept if the relative distance between them is within $\Delta R < 0.2$. However, if the relative distance is within $0.2 < \Delta R < 0.4$, the electron is removed, and the jet is kept. For muons, the overlap removal algorithm targets jets within $\Delta R < 0.4$ of a muon. If the number of tracks associated with the jet is less than 3, then the jet is removed, and the muon is kept. Otherwise, the muon is removed, and the jet is kept.

4.7 Missing Transverse Momentum

There exist various undetectable particles in the standard model and its many extensions. An example of one such weakly interacting particle in the standard model is the neutrino, which traverses through all modules of the ATLAS detector without leaving any signal. Since direct measurement could not be performed, such particles produced in collisions are inferred from the imbalance of overall energy deposit in all directions.

Due to the conservation of momentum, the sum of all particles' momenta should be balanced along all directions, yielding a sum of vanishing magnitude. If all observed

particles momenta sum to a nonzero value, this implies the existence of weakly interacting particles with momentum magnitudes and directions capable of balancing the overall momentum. The sum of these undetected momenta is referred to as the missing momentum.

In practice, the protons' interacting constituents, known as partons, carry unknown percentages of the parent protons' momenta along the beam axis. The unknown fraction imposes great difficulty in estimating the missing momentum along the beam direction since the nonzero momentum sum of the initial state is unknown. However, since the beam is confined to the z-axis, the partons only possess negligible momentum in the transverse plane, causing the transverse momentum to sum to zero within the momentum resolution. This motivates the use of the missing transverse momentum, denoted by $\mathbf{E}_T^{\text{miss}}$, which is the projection of the three-dimensional missing momentum onto the transverse plane.

In the ATLAS experiment, the $\mathbf{E}_T^{\text{miss}}$ vector is constructed by the momentum contribution from several sources [80]:

$$\mathbf{E}_T^{\text{miss}} = - \sum_{\text{selected electrons}} \mathbf{p}_T^e - \sum_{\text{accepted photons}} \mathbf{p}_T^\gamma - \sum_{\text{accepted } \tau\text{-leptons}} \mathbf{p}_T^{\tau_{\text{had}}} - \sum_{\text{selected muons}} \mathbf{p}_T^\mu - \sum_{\text{accepted jets}} \mathbf{p}_T^{\text{jets}} - \sum_{\text{unused tracks}} \mathbf{p}_T^{\text{track}}. \quad (4.3)$$

The algorithm selects a list of physics objects likely from the hard scatter vertex as shown in the first five terms in Equation 4.3. The negative sign on the right-hand side of the equation ensures that the direction of $\mathbf{E}_T^{\text{miss}}$ is opposite to that of the sum of observed particles. This analysis used the soft term in the $\mathbf{E}_T^{\text{miss}}$ reconstruction, represented by the last term in Equation 4.3. The soft term uses all tracks associated with the hard scatter

vertex but is not used in reconstructing objects in other terms of the $\mathbf{E}_T^{\text{miss}}$. The selection criteria of objects used in Equation 4.3 are listed in Table 4.2 [80].

Particle type	Selections	Comments
Electrons (e)	$ \eta < 1.37$ or $1.52 < \eta < 2.47$. $p_T > 10$ GeV	Medium quality required.
Photons (γ)	$ \eta < 1.37$ or $1.52 < \eta < 2.47$. $p_T > 25$ GeV	Tight quality required. No overlap with e .
τ -leptons (τ_{had})	$ \eta < 1.37$ or $1.52 < \eta < 2.47$ $p_T > 20$ GeV	Narrow jets with few low associated tracks. Medium quality required. No overlap with e and γ .
Muons (μ)	$ \eta < 2.7$. $p_T > 10$ GeV	Medium quality required.
Jets	$ \eta < 4.5$. $p_T > 60$ GeV. — or — $2.4 < \eta < 4.5$ $20 < p_T < 60$ GeV — or — $ \eta < 2.4$ $p_T > 60$ GeV JVT > 0.59	Reconstruction quality passed. No overlap with e , γ , and τ_{had} . Dedicated removal with μ .

Table 4.2: Selection criteria of objects contributing to $\mathbf{E}_T^{\text{miss}}$ [80].

The performance of the $\mathbf{E}_T^{\text{miss}}$ reconstruction algorithm is tested in MC simulated by comparing the reconstructed $\mathbf{E}_T^{\text{miss}}$ values with the truth information. A plot of $\mathbf{E}_T^{\text{miss}}$ resolution versus the truth $\mathbf{E}_T^{\text{miss}}$ is shown in Figure 4.5 for $W \rightarrow e\nu$, $W \rightarrow \mu\nu$, and $t\bar{t}$ events [80]. The resolution of $\mathbf{E}_T^{\text{miss}}$ in the x and y axis is represented by the Root Mean Square width. The resolution of $\mathbf{E}_T^{\text{miss}}$ is significantly worse than that of other objects, primarily because the $\mathbf{E}_T^{\text{miss}}$ is prone to fluctuations caused by pile-up events and uncertainties from other objects' reconstructions. The resolution in each axis in the transverse plane is typically between 10 to 20 GeV, as shown in Figure 4.5.

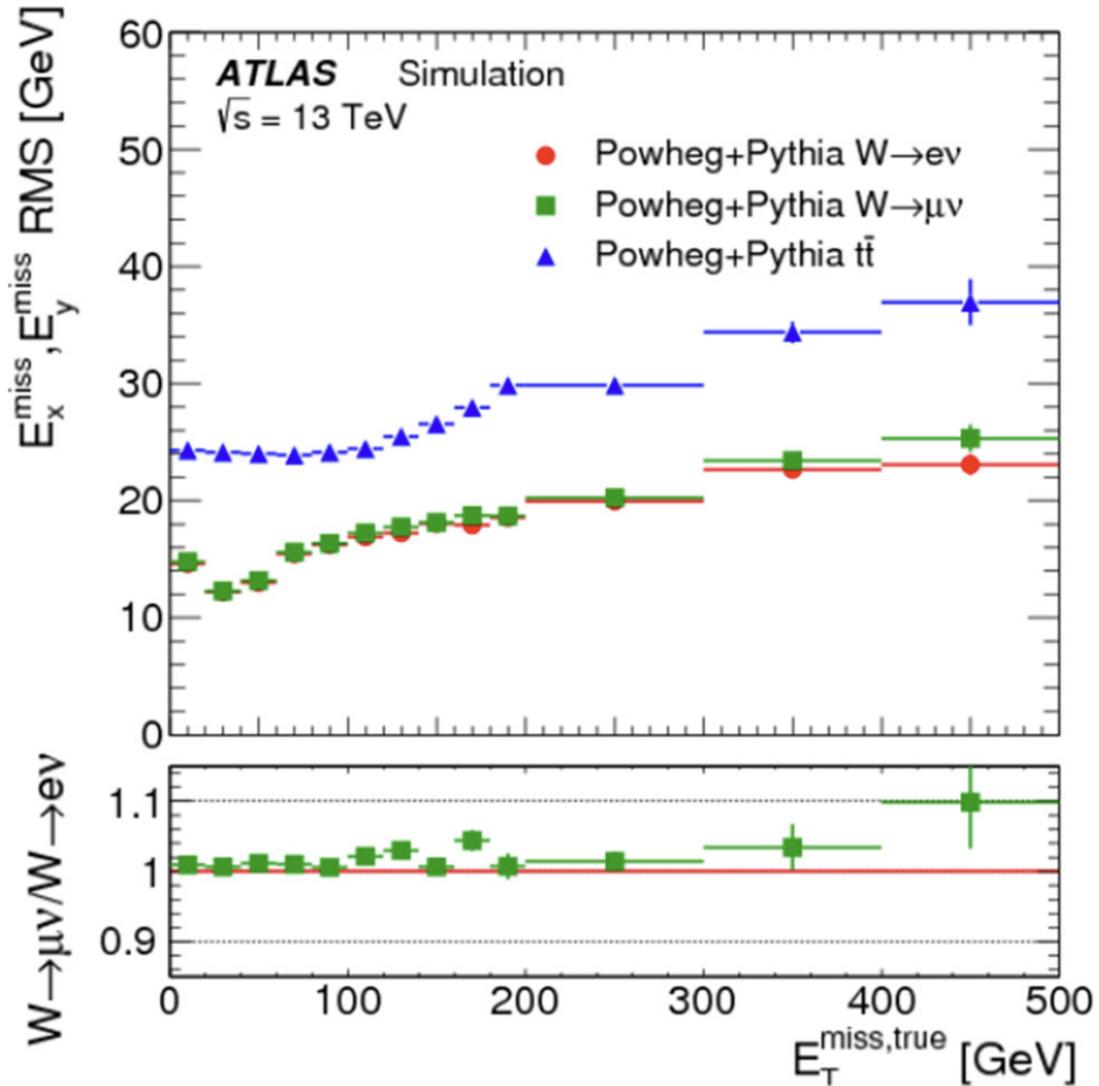


Figure 4.5: Resolution of E_T^{miss} measured by RMS width as a function of truth E_T^{miss} in $W \rightarrow e\nu$, $W \rightarrow \mu\nu$, and $t\bar{t}$ events [80].

4.8 Neutrino

The final state neutrino in the W decay imposes an obvious challenge in reconstructing the Higgs invariant mass. Since the detectors could not measure neutrinos, their four-momenta, which are needed for the Higgs mass reconstruction, must be inferred from extra constraints.

4.8.1 W -mass Constraint

In this analysis, the leading signal and background process, $H \rightarrow W(l\nu)\gamma$ and standard model $W(l\nu)\gamma$, respectively involve only one neutrino. This leads us to model the neutrino transverse momentum using the $\mathbf{E}_T^{\text{miss}}$ vector, assuming all other sources of $\mathbf{E}_T^{\text{miss}}$ are negligible compared to the neutrino contribution. However, the neutrino momentum along the z -axis (p_z^ν) is still left unknown. To tackle this problem, the W -mass constraint is used to solve for p_z^ν .

In the $W\gamma$ final state that is considered in this analysis, there is typically only one lepton, which originates from the decay of the W -boson, implying that the subsystem of lepton and neutrino should have an invariant mass close to the W -boson mass. This gives the following constraint expressed in energy (E^ν, E^ℓ) and momenta ($\mathbf{p}^\nu, \mathbf{p}^\ell$) of the neutrino and the lepton:

$$m_{\ell\nu} = \sqrt{(E^\nu + E^\ell)^2 - (\mathbf{p}^\nu + \mathbf{p}^\ell)^2} = m_W, \quad (4.4)$$

where $E_\nu = \sqrt{p_x^{\nu 2} + p_y^{\nu 2} + p_z^{\nu 2}} \approx \sqrt{(E_x^{\text{miss}})^2 + (E_y^{\text{miss}})^2 + p_z^{\nu 2}}$.

Once the neutrino transverse momentum is taken to be the same as the $\mathbf{E}_T^{\text{miss}}$ vector, it is obvious to observe that the only unknown in Equation 4.4 is p_z^ν . Solving this equation gives a p_z^ν estimation subject to a quadratic ambiguity. If the quadratic equation yields two real solutions, the solution with the smaller magnitude is picked. However, the quadratic equation could yield complex, non-physical solutions due to $\mathbf{E}_T^{\text{miss}}$ fluctuations.

4.8.2 Complex Solutions

If all parameters in Equation 4.4 are exact, the solutions will be real numbers since the momentum magnitude of a particle could not be a complex number. In practice, there are enough fluctuations in these parameters, especially those affiliated with the reconstruc-

tion of E_T^{miss} , so that the solutions could be complex. The parameters and the complex solutions are assumed to be close to the truth values since the fluctuation in E_T^{miss} is relatively small compared to the energy scale of neutrino in most cases. Therefore, the easiest method to deal with complex solutions is to take their real parts, which are assumed to be close to the truth p_z^ν :

$$p_z^\nu \approx \Re(p_z^{\nu, \mathbb{C}}) \quad (4.5)$$

where $p_z^{\nu, \mathbb{C}}$ is the solution to Equation 4.4 and \Re denotes the action of taking the real part. This method assumes that the solution's complex part is small and negligible, such that the real part has a magnitude close to the truth value.

E_T^{miss} variation

Alternatively, the algorithm could acknowledge that the E_T^{miss} is poorly measured and attempts to vary the E_T^{miss} near the nominal value to search for real solutions of Equation 4.4. In this method, the algorithm floats all three components of the neutrino momentum, which are constrained to be close to their respective nominal values. For the transverse momentum along the x and y-axis, the nominal values are those of the E_T^{miss} . For p_z^ν , it is constrained to be close to the real part of the complex solution ($\Re(p_z^{\nu, \mathbb{C}})$). Finally, it is useful to add the W -mass constraint that states the invariant mass of the lepton and neutrino system should be close to that of the W -boson. These constraints are combined to give the following cost function, in which each constraint is associated with its coefficient:

$$C(p_x^\nu, p_y^\nu, p_z^\nu) = \sum_{i=x,y} c_i (p_i^\nu - E_i^{\text{miss}})^2 + c_z (p_z^\nu - \bar{p}_z^\nu)^2 + c_m (m_{\ell\nu} - m_W)^2. \quad (4.6)$$

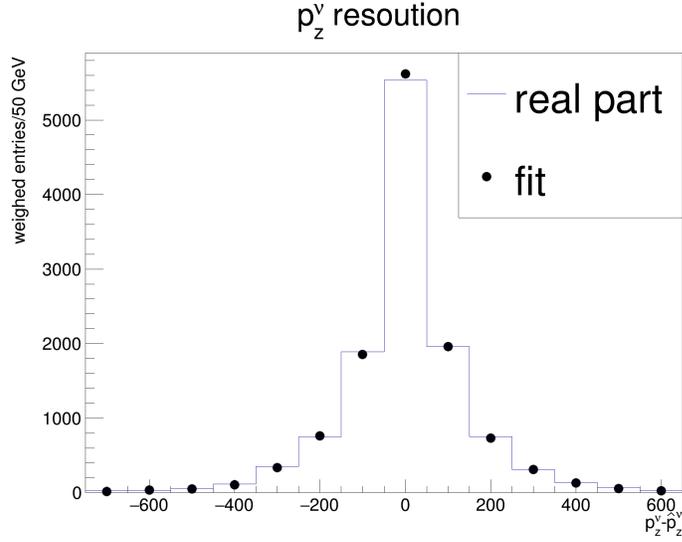
Minimizing Equation 4.6 would yield solutions of neutrino momenta that reconcile the nominal values with the W -mass constraints. The confidence of each constraint is characterized by its strength coefficients. In this analysis, these values are taken to be:

- $c_x = c_y = 100$
- $c_z = 10$
- $c_m = 1$

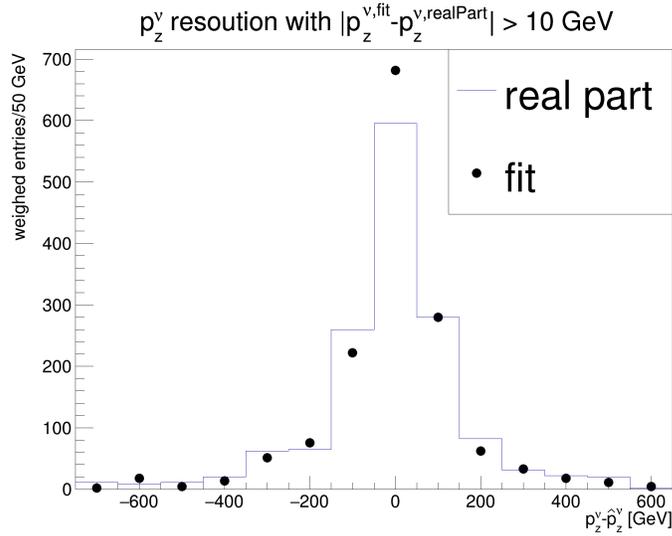
The estimates of the fit method are compared with those based on the real parts of the complex solutions in events where complex solutions were found. The results obtained from the signal process MC simulation are shown in Figure 4.6. For most events, the fluctuation is not large such that taking the real part of the complex solution gives similar performance compared to implementing the fit method. However, in events where the two methods do not agree, the fit method yields much better resolutions, as shown in Figure 4.6(b). The nominal values of coefficients were chosen to reflect the relative strength of the different constraint terms in the cost function. In practice, the coefficients in the fit were observed to be robust in large ranges of fluctuations around the nominal values. The coefficient c_x , c_y and c_z were tested with values from:

- $c_x, c_y \in \{0.1, 1, 10, 100, 1000\}$
- $c_z \in \{0.01, 0.1, 1, 10, 100\}$.

With combinations of choices listed above, the maximum variation in resolution was merely 0.36%. This motivates the use of nominal values for this analysis without further optimizations.



(a) Comparison in all events.



(b) Comparison in events where the fit estimates and the real part estimates differ by more than 10 GeV.

Figure 4.6: Comparison of neutrino z-momentum p_z^ν estimates in events with complex solutions in MC simulation of signal processes. The x-label is the difference between the truth momentum, p_z^ν , and the estimated momentum, \hat{p}_z^ν .

4.8.3 Higgs Mass

After the reconstruction of the neutrino momentum, the W boson momentum could be easily reconstructed by combining the four-momentum of the lepton and neutrino. At the same time, if the photon decayed from the charged Higgs could be selected by considering the photon closest to the lepton, the invariant mass of the charged Higgs candidate could be reconstructed by:

$$m_5 = m_{W\gamma} = m_{\ell\nu\gamma} = \sqrt{\left(\sum_{k \in \{\ell, \nu, \gamma\}} E^k\right)^2 - \left(\sum_{k \in \{\ell, \nu, \gamma\}} \mathbf{p}^k\right)^2}. \quad (4.7)$$

Figure 4.7 shows the distributions of the reconstructed invariant masses in signal Monte Carlo samples with $m_5 = 150$ GeV. It is clear that the invariant mass peaks around 150 GeV are reconstructed in both $H_5^\pm H_5^\mp$ and $H_5^\pm H_5^0$ samples. It is also clear that the reconstructed Higgs mass distribution has a much higher tail in Figure 4.7(a) than in Figure 4.7(b). This high tail is an inevitable effect coming from the incorrect selections of photons. Although the two processes are generated with similar mechanisms, there are more final state photons in the $H_5^\pm H_5^0$ samples that eventually resulted in a higher probability of incorrect association of photons with the charged Higgs decay vertices.

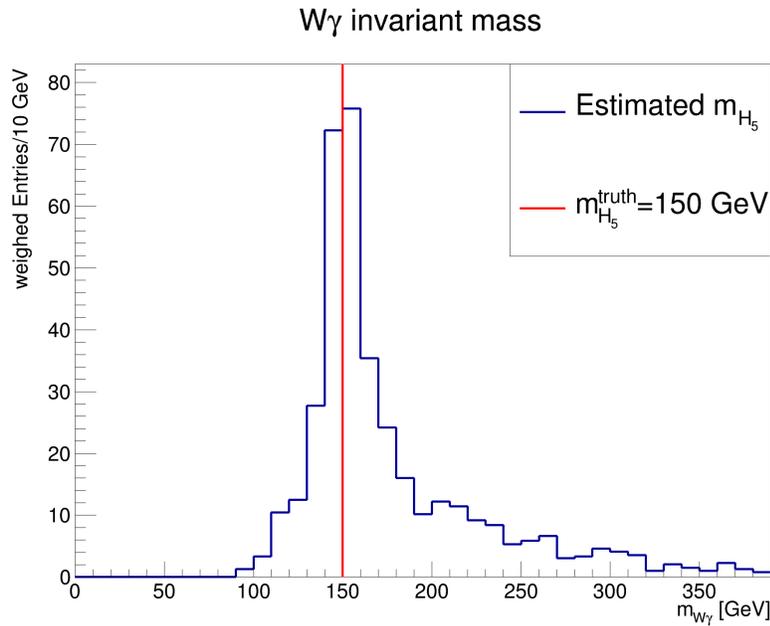
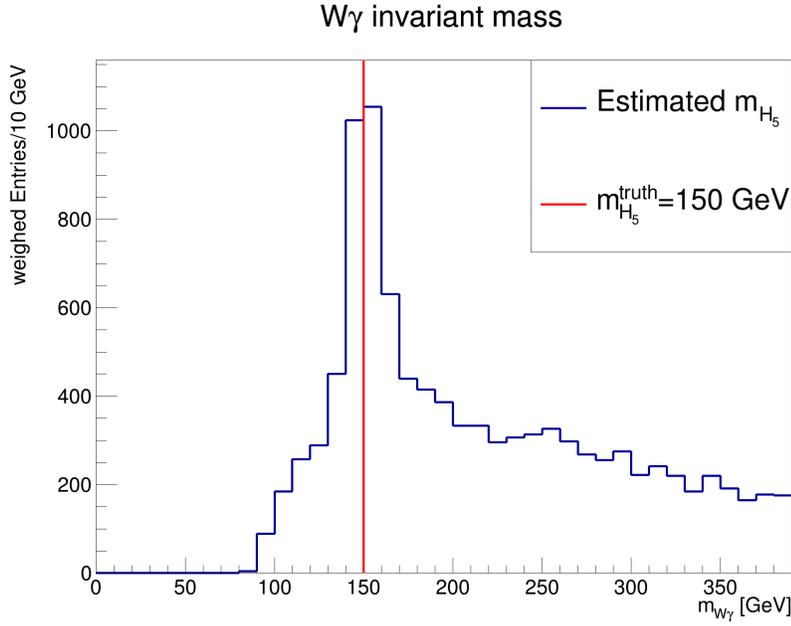


Figure 4.7: Reconstructed Higgs mass from $pp \rightarrow H_5^\pm H_5^0$ and $pp \rightarrow H_5^\pm H_5^\mp$ processes with the fiveplet Higgs mass at 150 GeV.

Using the reconstruction algorithms introduced in the chapter, most of the physics objects needed in this analysis can be reconstructed, including the invariant mass of the

charged Higgs boson in the signal samples. The next step will be to use these reconstructed objects to select interesting physics events where the analysis will be performed. The next chapter will discuss the event selections that build up various regions for the analysis.

Chapter 5

Event Selection

This chapter presents a short introduction to datasets and analysis regions. The analysis uses real data collected by the ATLAS detector and MC simulations of signal and background processes. These datasets are described in Section 5.1. To increase signal sensitivity in the analysis, these datasets are divided into different regions where the analysis is conducted. These regions are defined in Section 5.2.

5.1 Datasets

This section briefly introduces the data format used in the ATLAS experiment. Collision data collected by the ATLAS detector are reconstructed to be saved in the derivation format if the recorded event has passed the dedicated triggers. Similarly, the MC samples are generated at the event level and subsequently undergo detector simulation, which matches the detector performance of the ATLAS detectors. These simulated events are then reconstructed and saved in the same derivation format as data. Finally, the project in this thesis uses data and MC derivations for the analysis. A schematic diagram of the workflow is shown in Figure 5.1.

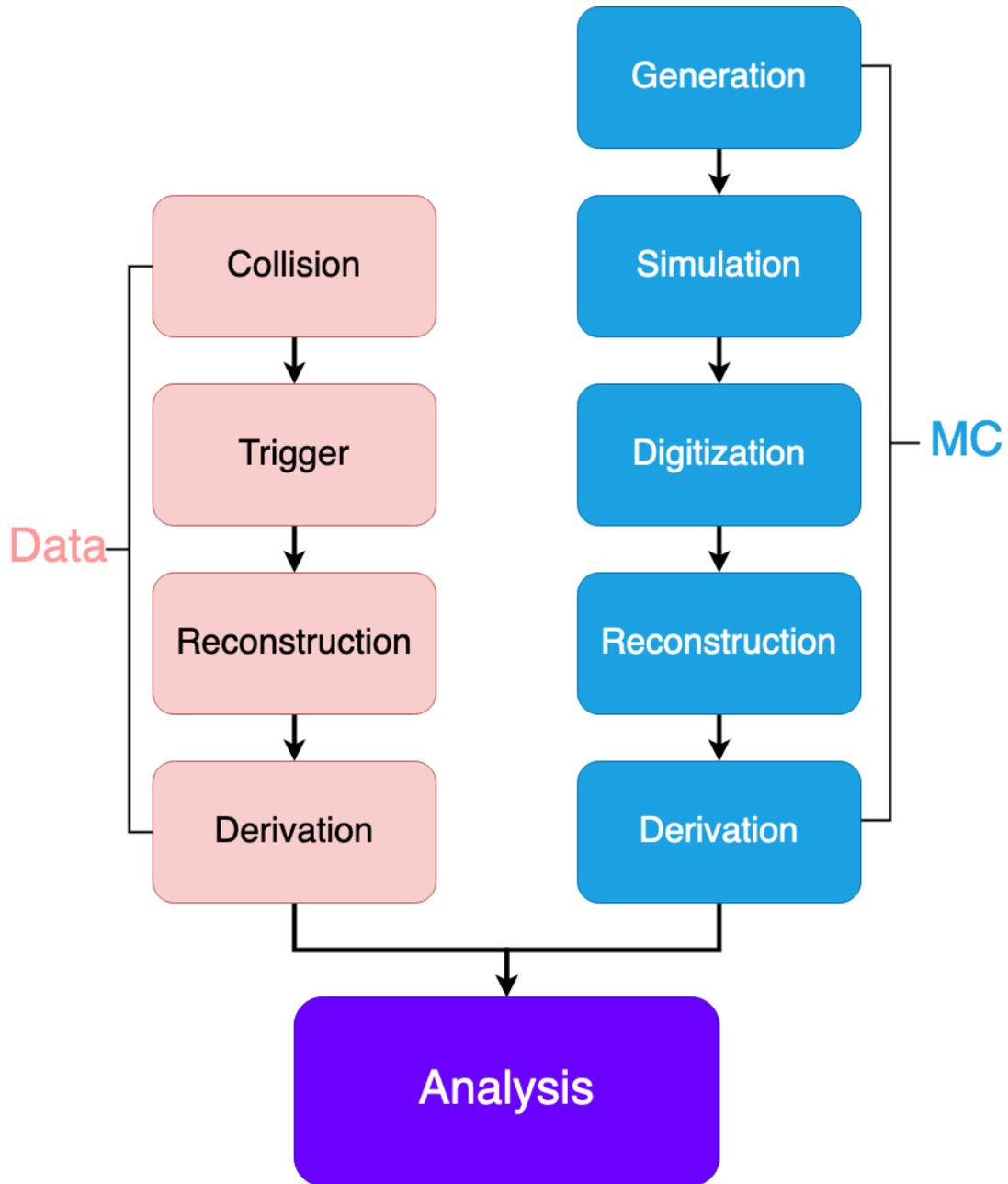


Figure 5.1: The workflow of data and MC processing up to the analysis level.

5.1.1 Data

This analysis uses the entire run-2 dataset recorded by the ATLAS detector at a center-of-mass energy of $\sqrt{s} = 13$ TeV from 2015 to 2018. All data used in this analysis must be in the Good Run List (GRL), which are lists containing good quality data recorded in

a given year. The integrated luminosity delivered by the LHC in each year of the run-2 is shown in Figure 3.4, and the total delivered luminosity of good quality data in the entire run-2 period is around 140.1 fb^{-1} . The overall measured luminosity uncertainty is estimated to be 0.83% [35]. The mean number of pile-up interactions per bunch crossing in the run-2 periods typically ranges from 15 to 60. The distribution of the mean number of interactions per crossing in the run-2 period was shown in Figure 3.5.

5.1.2 Monte Carlo Simulation

The MC generation consists of five steps, as shown in Figure 5.1. These steps can be summarized as the following [81]:

1. Generation: The first step is to sample from the hard scatter process matrix elements to produce events with particles out of the proton-proton collisions.
2. Simulation:
 - Parton shower: QCD hadronization processes are simulated for partons created by the collisions. This step produces QCD-related particles that the detectors will record.
 - Detector simulation: Particles' interactions with different detector components are simulated. In practice, this is often the most computationally expensive part of the MC generation, and fast simulation techniques are often used to avoid full simulation of energy deposits.
3. Digitization: This step converts the simulated energy deposit to digital signals similar to the raw data recorded by the ATLAS detector.
4. Reconstruction: The simulated samples are subject to the same reconstruction algorithms as the real data. These algorithms were described in detail in Chapter 4.

The ATLAS experiment splits MC generation into three campaigns, MC16a, MC16d, and MC16e, which correspond to data collected in year 2015+2016, 2017, and 2018, respectively.

Signal samples

This analysis considers the GM model with the $H_5^\pm \rightarrow W\gamma$ and $H_5^0 \rightarrow \gamma\gamma$ decays as the benchmark. These samples are simulated using the ATLAS fast simulation package, AtlFast-II [82], which is much less computationally expensive than the full ATLAS GEANT4 simulation [83]. Proton-proton collisions are generated at the parton level through Pythia 8 [84] using the NNPDF3.0NLO set of Parton Distribution Functions (PDF).

Signal samples of $H_5^\pm H_5^0$ and $H_5^\pm H_5^\mp$ were generated with masses between 110 GeV and 200 GeV in steps of 10 GeV while setting $s_H = \sin \theta_H = 1 \times 10^{-4}$ to further suppress the WZ decay channel. $H_5^\pm H_5^\mp$ processes were also generated at mass points of 150 GeV and 190 GeV, where the doubly charged Higgs decayed through $H_5^{\mp\mp} \rightarrow W^\mp \gamma W^\mp \gamma$. In the generation, matrix elements are calculated using MadGraph5_aMC@NLO v2.8.1 [85].

This analysis considers the leptonic decay channels of the W^\pm bosons with an overall final state of $\ell\nu_\ell\gamma$. A series of generator filter cuts is applied to select the desired events:

- Require at least one lepton (e, μ, τ) with at least $p_T > 7$ GeV. This filter allows for the presence of hadronic tau leptons.
- Require the leading lepton (e, μ) to have a $p_T > 20$ GeV.
- Require at least one photon with $E_T > 20$ GeV.

The event generation details of each signal process are given in Tables 5.1 - 5.3. These tables give the cross-section and number of events generated for each mass point. The

filter efficiency in the tables marks the fraction of events passing the filters. The k-factor gives corrections for the absence of high-order terms.

DSID	m_5 (GeV)	Events	Cross-Section [fb]	Filter Efficiency	k-factor
504671	110	50000	683.7	27.67%	1.0
504651	120	50000	959.6	27.52%	1.0
504652	130	50000	1305	27.40%	1.0
504653	140	50000	1724	27.42%	1.0
504672	150	50000	2212	27.30%	1.0
504655	160	50000	2773	27.33%	1.0
504656	170	50000	3398	27.33%	1.0
504657	180	50000	4089	27.38%	1.0
504673	190	50000	4845	27.49%	1.0
504659	200	50000	5660	27.54%	1.0

Table 5.1: Summary of the $H_5^\pm H_5^0$ signal process event generation

DSID	m_5 (GeV)	Events	Cross-Section [fb]	Filter Efficiency	k-factor
504676	150	50000	3.801×10^{-5}	49.51%	1.0
504677	190	50000	4.411×10^{-3}	49.06%	1.0

Table 5.2: Summary of the $H_5^\pm H_5^{\mp\mp}$ signal process event generation

DSID	m_5 (GeV)	Events	Cross-Section [fb]	Filter Efficiency	k-factor
504661	110	50000	9.191	36.74%	1.0
504662	120	50000	30.90	37.79%	1.0
504663	130	50000	74.79	38.35%	1.0
504664	140	50000	147.7	38.61%	1.0
504674	150	50000	254.5	39.01%	1.0
504666	160	50000	398.3	39.29%	1.0
504667	170	50000	580.9	39.60%	1.0
504668	180	50000	802.5	39.96%	1.0
504675	190	50000	1065	40.01%	1.0
504670	200	50000	1364	40.36%	1.0

Table 5.3: Summary of the $H_5^+ H_5^-$ signal process event generation

Background samples

The MC simulation also models the background processes and can be divided into two categories: prompt backgrounds and non-prompt backgrounds. Prompt backgrounds

refer to those processes in which the reconstruction of the detected final state particles match the truth type of the particle. On the contrary, non-prompt backgrounds contain misidentified final state particles, which are classified into the wrong particle type after the reconstruction algorithm.

The prompt backgrounds are summarised in Table 5.4, and the non-prompt backgrounds are summarised in Table 5.5. The leading processes are the $W\gamma$ and $Z\gamma$ samples produced by QCD interactions. High-order samples generated by the electroweak interaction are also used and marked with the label "EWK" in Table 5.4. In the non-prompt samples, the leading order contribution comes from W +jets and Z +jets samples. Although their cross sections are much larger than that of other prompt processes, their overall contributions are less prominent since only a small fraction of the events gives the non-prompt photon.

These samples were selected to give a reasonable estimate of the $\ell\nu\gamma$ final states. $W\gamma$ samples will have nearly identical final states with the signal samples, contributing the most in the signal region. Some of the other samples Table 5.4 involve more final state particles than those in the signal sample. For example, $Z\gamma$ samples will have two final state leptons from the Z decay instead of one. However, if the subleading lepton's transverse momentum is smaller than the p_T cut we implement, this could still result in a one-lepton final state. The non-prompt samples in Table 5.5 contributes to signal-like events via faking some of the final state particles with non-prompt ones,

5.1.3 Triggers and Derivations

Triggers

Various triggers are required to have been passed by this analysis to select signal events for the massive dataset recorded in the run-2 period. Due to the different running con-

DSID	Process	Generators	PDF	Cross-Section [pb]	Filter Efficiency	k-factor
700011	$Z\gamma \rightarrow e\bar{e}\gamma$ (QCD only)	Sherpa 2.2.8	NNPDF30NNLO	98.69	100%	1.0
700012	$Z\gamma \rightarrow \mu\bar{\mu}\gamma$ (QCD only)	Sherpa 2.2.8	NNPDF30NNLO	98.67	100%	1.0
700013	$Z\gamma \rightarrow \tau\bar{\tau}\gamma$ (QCD only)	Sherpa 2.2.8	NNPDF30NNLO	98.70	100%	1.0
700015	$W\gamma \rightarrow e\nu_e\gamma$ (QCD only)	Sherpa 2.2.8	NNPDF30NNLO	355.96	100%	1.0
700016	$W\gamma \rightarrow \mu\nu_\mu\gamma$ (QCD only)	Sherpa 2.2.8	NNPDF30NNLO	357.41	100%	1.0
700017	$W\gamma \rightarrow \tau\nu_\tau\gamma$ (QCD only)	Sherpa 2.2.8	NNPDF30NNLO	356.14	100%	1.0
410389	$t\bar{t}\gamma \rightarrow \ell\nu_\ell\gamma + \text{jets}$	MadGraph+Pythia8	A14NNPDF23LO	46.243	100%	1.16
363270	$\text{EWKW}\gamma \rightarrow e\nu_e\gamma$	MadGraph+Pythia8	NNPDF30	$8.3980 \cdot 10^{-1}$	100%	1.0
363271	$\text{EWKW}\gamma \rightarrow \mu\nu_\mu\gamma$	MadGraph+Pythia8	NNPDF30	$8.3980 \cdot 10^{-1}$	100%	1.0
363272	$\text{EWKW}\gamma \rightarrow \tau\nu_\tau\gamma$	MadGraph+Pythia8	NNPDF30	$8.3961 \cdot 10^{-1}$	100%	1.0
363267	$\text{EWKZ}\gamma \rightarrow e\bar{e}\gamma$	MadGraph+Pythia8	NNPDF30	$5.2126 \cdot 10^{-2}$	100%	1.0
363268	$\text{EWKZ}\gamma \rightarrow \mu\bar{\mu}\gamma$	MadGraph+Pythia8	NNPDF30	$5.1978 \cdot 10^{-2}$	100%	1.0
363269	$\text{EWKZ}\gamma \rightarrow \tau\bar{\tau}\gamma$	MadGraph+Pythia8	NNPDF30	$5.1992 \cdot 10^{-2}$	100%	1.0
366160	$WZ\gamma \rightarrow \ell\nu_\ell\ell'\ell'\gamma$	Sherpa 2.2.5	NNPDF30NNLO	$1.1788 \cdot 10^{-2}$	100%	1.0
366161	$WW\gamma \rightarrow \ell\nu_\ell\ell'\nu_{\ell'}\gamma$	Sherpa 2.2.5	NNPDF30NNLO	$8.6767 \cdot 10^{-2}$	100%	1.0
366162	$ZZ\gamma \rightarrow \ell\ell\ell'\ell'\gamma$	Sherpa 2.2.5	NNPDF30NNLO	$1.0375 \cdot 10^{-2}$	100%	1.0
700195	$Z\gamma\gamma \rightarrow e\bar{e}\gamma\gamma$	Sherpa 2.2.10	NNPDF30NNLO	1.3121	100%	1.0
700196	$Z\gamma\gamma \rightarrow \mu\bar{\mu}\gamma\gamma$	Sherpa 2.2.10	NNPDF30NNLO	1.3118	100%	1.0
700197	$Z\gamma\gamma \rightarrow \tau\bar{\tau}\gamma\gamma$	Sherpa 2.2.10	NNPDF30NNLO	1.3094	100%	1.0
700199	$W\gamma\gamma \rightarrow e\nu_e\gamma\gamma$	Sherpa 2.2.10	NNPDF30NNLO	2.0036	100%	1.0
700200	$W\gamma\gamma \rightarrow \mu\nu_\mu\gamma\gamma$	Sherpa 2.2.10	NNPDF30NNLO	1.9974	100%	1.0
700201	$W\gamma\gamma \rightarrow \tau\nu_\tau\gamma\gamma$	Sherpa 2.2.10	NNPDF30NNLO	2.0050	100%	1.0

Table 5.4: Summary of the prompt background processes event generation. The $W\gamma$ and $Z\gamma$ background samples in this table contain events in which the photon is produced through QCD processes. The $\text{EWKW}\gamma$ and $\text{EWKZ}\gamma$ samples are those where the photon was produced through electroweak interactions.

DSID	Process (all units in GeV)	Generators	PDF	Cross-Section [pb]	Filter Efficiency	k-factor
364100-364141	$Z + \text{jets}$	Sherpa 2.2.1	NNPDF30NNLO	$1981.7 - 1.4821 \cdot 10^{-1}$	97.51%	0.8215 - 1.0
364156-364197	$W + \text{jets}$	Sherpa 2.2.1	NNPDF30NNLO	19149 - 1.2343	97.02%	8.246 - 1.0
410470	fully-leptonic $t\bar{t}$	Powheg+Pythia8	A14NNPDF23LO	729.77	54.382%	1.1398
410471	fully-hadronic $t\bar{t}$	Powheg+Pythia8	A14NNPDF23LO	729.77	45.621%	1.1397
364350	$\gamma\gamma, 0 < m_{\gamma\gamma} < 50$	Sherpa 2.2.4	NNPDF30NNLO	93.499	100%	1.0
364351	$\gamma\gamma, 50 < m_{\gamma\gamma} < 90$	Sherpa 2.2.4	NNPDF30NNLO	139.04	100%	1.0
364352	$\gamma\gamma, 90 < m_{\gamma\gamma} < 175$	Sherpa 2.2.4	NNPDF30NNLO	51.822	100%	1.0
364353	$\gamma\gamma, 175 < m_{\gamma\gamma} < 2000$	Sherpa 2.2.4	NNPDF30NNLO	10.999	100%	1.0
364354	$\gamma\gamma, 2000 < m_{\gamma\gamma} < E_{CMS}$	Sherpa 2.2.4	NNPDF30NNLO	$7.0335 \cdot 10^{-4}$	100%	1.0
364541	$\gamma + \text{jets}, 17 < p_T^\gamma < 35$	Sherpa 2.2.2	NNPDF30NNLO	$4.1548 \cdot 10^5$	100%	1.0
364542	$\gamma + \text{jets}, 35 < p_T^\gamma < 70$	Sherpa 2.2.2	NNPDF30NNLO	$4.3976 \cdot 10^4$	100%	1.0
364543	$\gamma + \text{jets}, 70 < p_T^\gamma < 140$	Sherpa 2.2.2	NNPDF30NNLO	$4.5263 \cdot 10^3$	100%	1.0
364544	$\gamma + \text{jets}, 140 < p_T^\gamma < 280$	Sherpa 2.2.2	NNPDF30NNLO	376.03	100%	1.0
364545	$\gamma + \text{jets}, 280 < p_T^\gamma < 500$	Sherpa 2.2.2	NNPDF30NNLO	2.1864	100%	1.0
364546	$\gamma + \text{jets}, 500 < p_T^\gamma < 1000$	Sherpa 2.2.2	NNPDF30NNLO	1.4629	100%	1.0
364547	$\gamma + \text{jets}, 1000 < p_T^\gamma < E_{CMS}$	Sherpa 2.2.2	NNPDF30NNLO	$2.9864 \cdot 10^{-2}$	100%	1.0
364250	$ZZ \rightarrow \ell\ell\ell'\ell'$	Sherpa 2.2.2	NNPDF30NNLO	1.2516	100%	1.0
364253	$WZ \rightarrow \ell\ell\ell'\nu_{\ell'}$	Sherpa 2.2.2	NNPDF30NNLO	4.5724	100%	1.0
364254	$WW \rightarrow \ell\nu_{\ell'}\ell'\nu_{\ell'}$	Sherpa 2.2.2	NNPDF30NNLO	12.501	100%	1.0
364255	$WZ \rightarrow \ell\nu_{\ell'}\nu_{\ell'}\bar{\nu}_{\ell'}$	Sherpa 2.2.2	NNPDF30NNLO	3.2347	100%	1.0

Table 5.5: Summary of the non-prompt background processes event generation.

ditions in each year, the choice of trigger depends on the run period. For this analysis, the single lepton triggers, which select the single lepton in the final state particles of the signal samples, are implemented. The lepton triggers used in the ATLAS experiment are named after the p_T threshold and reconstruction quality such as "eXX" or "muXX", where "XX" labels the minimum p_T threshold in GeV, followed by likelihood-based identification working point labels such as *lhloose*, *lhmedium*, *lhtight*. The triggers also specify the isolation working point with the label *ivar*. These quality labels indicate that the lepton

Lepton	2015	2016-2018
Electron	<i>HLT_e24_lhmedium_L1EM20VH</i>	<i>HLT_e26_lhtight_nod0_ivarlose</i>
	<i>HLT_e60_lhmedium</i>	<i>HLT_e60_lhmedium</i>
	<i>HLT_e140_lhloose_nod0</i>	<i>HLT_e140_lhloose_nod0</i>
Muon	<i>HLT_mu26_ivarmedium</i>	<i>HLT_mu26_ivarmedium</i>
	<i>HLT_mu50</i>	<i>HLT_mu50</i>

Table 5.6: List of triggers used in each year of the run-2 dataset.

has passed certain identification and isolation requirements. If "nod0" is in the trigger, it indicates that the trigger does not use the transverse impact parameter (d_0) information. The label *L1EM20VH* specifies the p_T threshold in the electromagnetic calorimeter as discussed in Reference [86]. Table 5.1.3 lists the triggers used in this analysis. Each event from the run-2 dataset is accepted if it passes any of the triggers listed in this table.

Derivations

After the reconstruction of raw data has been performed, all events are saved in Analysis Object Data (AOD) format, which contains information on reconstructed objects in each event. These events must go through the derivation production to be saved in the Derived Analysis Object Data (DAOD) format. Compared to AODs, DAODs have some of their information skimmed while additional information is added. Information like jet reconstruction and flavor tagging is typically added at the DAOD level.

Our analysis uses the DAOD_PHYS derivation, a fairly new derivation that aims to be the general-use derivation for future analyses. DAOD_PHYS only occupies a small size of ~ 50 kB/event and thus will greatly ease the data storage difficulties, which are expected to be increasingly burdensome as the disk space required increases exponentially after the HL-LHC upgrade [87].

5.2 Event Selections and Analysis Regions

Only a subset of events in the overall dataset are used in the actual analysis. These events are selected using the basic inclusive event selection criteria. The selected events are then further divided into different analysis regions where the actual analysis takes place. There are three types of regions in this analysis, the signal region (SR), the validation region (VR), and the control region (CR). The signal region is designed to have the greatest signal sensitivity and is thus used to search for the signal processes. The validation region is a region that is close to the signal region but with fewer signal process contributions. The background estimation is checked in the validation region to ensure good modeling of background processes in the signal region. A control region is designed to have one of the background processes dominate while having minimal signal leakage. Nuisance parameters from various systematic uncertainties are determined in these regions and subsequently used in the signal region for the analysis. Kinematic cuts of different regions are introduced in Section 5.2.2. The MC simulation with the 110 GeV Higgs mass is used in the following sections to estimate the signal contribution.

5.2.1 Inclusive Event Selection

The inclusive selection in this analysis requires all selected objects to pass the tight selection criteria discussed in Chapter 4. All leptons are required to have a transverse momentum p_T larger than 27 GeV, while photons and jets are required to have $p_T > 25$ GeV.

A primary vertex with at least two tracks satisfying $p_T > 500$ MeV is also required. Events are also required to pass the GRL requirement and at least one of the triggers listed in Section 5.1. Additionally, events with calorimeter detection corruptions are removed based on the standard ATLAS procedure of *GoodCalo*. If the event contains at least one

unidentified jet passing the Loose JVT requirement, it is removed to reject events with jets incorrectly reconstructed from calorimeter noise or cosmic ray shower [88].

A series of kinematic variables are considered to categorize events into different regions. This analysis considers one lepton and one photon in the final state. Since the photon is typically highly boosted from the W decay, the leading photon and the photon closest to it in the final state are selected. For all photons in the final state, the closest photon is defined as the photon that has the shortest distance $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ in the $\eta - \phi$ plane with respect to the leading lepton.

Other kinematic variables are constructed using the selected lepton and photon to further discriminate against certain processes. The common ones used are listed below:

- $p_T^{\ell+\gamma} = \sqrt{(p_x^\ell + p_x^\gamma)^2 + (p_y^\ell + p_y^\gamma)^2}$
- $p_T^{\ell+\gamma+\text{MET}} = \sqrt{(p_x^\ell + p_x^\gamma + E_x^{\text{miss}})^2 + (p_y^\ell + p_y^\gamma + E_y^{\text{miss}})^2}$

These two variables give the transverse momentum of the subsystems of lepton, photon, and neutrino. In the case of $W(\ell\nu)\gamma$ final states in the signal samples, the whole system is produced from a heavy charged Higgs recoiling against another Higgs boson. This production mechanism thus favors final states with large transverse momentum in the $\ell + \gamma$ or $\ell + \nu + \gamma$ subsystem. The two variables are used extensively in the following subsections to fine-tune the signal process contributions in each region. A clear example is shown in Table 5.7, where the high cutting values in $p_T^{\ell+\gamma}$ and $p_T^{\ell+\gamma+\text{MET}}$ reduce much of the background contributions but keep most of the signal events.

5.2.2 Regions

Signal region

The signal region is constructed to have the greatest signal process sensitivity. First of all, the initial cuts on leptons, photons, and jets are implemented:

- Exactly one tight lepton with $p_T > 27$ GeV,
- One or more tight photon with $p_T > 25$ GeV,
- Less than two jets with $p_T > 25$ GeV ,
- No b-jet.

In the signal processes, the pair production of Higgs bosons typically gives highly boosted objects, motivating the placement of harder cuts on variables related to the Higgs boson. The event is required to have $E_T^{\text{miss}} > 50$ GeV and the transverse momentum of the lepton + photon system to be larger than 100 GeV. The event is subsequently required to have $p_T^{\ell+\gamma+\text{MET}} > 200$ GeV to suppress the standard model $W\gamma$ background process. A cutflow of various standard model background processes versus the signal process is shown in Table 5.7. The table uses V to denote the two vector bosons, W and Z .

In the construction of this region, MC simulations are used as a guide to model all backgrounds, including ones involving misidentified objects. The modeling of these events will be checked with the data-driven method in the analysis. In the signal region, it is clear that the $W\gamma$ background gives the leading contribution that is on par with the signal process. The signal process contribution in the signal region is about 42.5% of the total number of weighted events.

Background	Cuts (in GeV)			
	Initial Cuts	$E_T^{\text{miss}} > 50$	$p_T^{\ell+\gamma} > 100$	$p_T^{\ell+\gamma+\text{MET}} > 200$
$W\gamma$	466,181	118,744	51,495	4,712
$Z\gamma$	458,388	18,117	4,927	188
$Z + \text{jets}$	1,104,675	27,511	5,484	262
$W + \text{jets}$	37,645	14,852	2,316	407
$t\bar{t}\gamma$	2,483	1,514	933	48
$t\bar{t}$	3,459	2,148	1,543	25
Top	1,152	449	133	13
$VV\gamma$	625	397	296	21
EWKV γ	6,650	2,928	1,759	377
2γ	22,336	585	294	12
Signal	7,600	5,698	5,531	4,483

Table 5.7: Cutflow for various processes leading to the construction of the signal region.

Validation region

The validation region is designed to be close to the signal region and thus serves the purpose of validating background estimations. The validation region inherits all cuts in the signal region except the last one, which is modified to $100 \text{ GeV} < p_T^{\ell+\gamma+\text{MET}} < 200 \text{ GeV}$, to ensure its proximity with the signal region.

This region is orthogonal to the signal region by definition, with a signal-to-background ratio for the Higgs fiveplet mass point of 110 GeV at approximately 4.4%. The relatively lower signal leakage ensures that it is safe to perform validation. Since it is adjacent to the hyper-parameter space of the signal region, the validation region validates the background estimation for the signal region. Table 5.8 shows a cutflow of various processes in the validation region.

$W\gamma$ control region

Since $W\gamma$ is the leading background in both the signal and validation region, the $W\gamma$ control region is the most important control region in which the $W\gamma$ background

Background	Cuts (in GeV)			
	Initial Cuts	$E_T^{\text{miss}} > 50$	$p_T^{\ell+\gamma} > 100$	$100 < p_T^{\ell+\gamma+\text{MET}} < 200$
$W\gamma$	466,181	118,744	51,495	9,897
$Z\gamma$	458,388	18,117	4,927	716
$Z + \text{jets}$	1,104,675	27,511	5,484	860
$W + \text{jets}$	53,327	32,172	9,228	1,383
$t\bar{t}\gamma$	2,483	1,514	933	278
$t\bar{t}$	3,459	2,148	1,543	289
Top	1,152	449	133	40
$VV\gamma$	625	397	296	47
EWKV γ	6,650	2,928	1,759	602
2γ	22,336	585	294	46
Signal	7,600	5,698	5,531	624

Table 5.8: Cutflow for various processes leading to the construction of the validation region.

dominates. Once again, the last selection criterion in the signal region is modified to be $p_T^{\ell+\gamma+\text{MET}} < 100$ GeV to maintain the $W\gamma$ dominance while cutting down the signal contribution. In the end, the contribution of the signal process is estimated to be around 0.82% of the total number of weighted events in the $W\gamma$ control region. This criterion also ensures the region's orthogonality with all previously defined regions. The cutflow of various processes in the $W\gamma$ control region is shown in Table 5.9.

Background	Cuts (in GeV)			
	Initial Cuts	$E_T^{\text{miss}} > 50$	$p_T^{\ell+\gamma} > 100$	$p_T^{\ell+\gamma+\text{MET}} < 100$
$W\gamma$	466,181	118,744	51,495	36,887
$Z\gamma$	458,388	18,117	4,927	4,023
$Z + \text{jets}$	1,104,675	27,511	5,484	4,632
$W + \text{jets}$	53,327	32,172	9,228	7,380
$t\bar{t}\gamma$	2,483	1,514	933	637
$t\bar{t}$	3,459	2,148	1,543	1,299
Top	1,152	449	133	80
$VV\gamma$	625	397	296	228
EWKV γ	6,650	2,928	1,759	781
2γ	22,336	585	294	236
Signal	7,600	5,698	5,531	463

Table 5.9: Cutflow for various processes leading to the construction of the $W\gamma$ -dominant control region.

W +jets control region

In the W +jets control region, selections are taken to be similar to those of the signal region. The last two selections are modified to be $p_T^{\ell+\gamma} < 100$ GeV and $p_T^{\ell+\gamma+\text{MET}} < 200$ GeV, which yield a large contribution of events with misidentified photons. The W +jets process has no real photon but jets faking photons passing selections instead. The W +jets MC sample is used to model such events at this stage. These estimations will be checked using data-driven methods in Section 6.3. From Table 5.10, the W +jets background is observed to be the subleading background in the W +jets control region. However, the W +jets process still has a much higher contribution in this region compared to that in the $W\gamma$ control region shown in Table 5.9.

Background	Cuts (in GeV)			
	Initial Cuts	$E_{\text{T}}^{\text{miss}} > 50$	$p_T^{\ell+\gamma} < 100$	$p_T^{\ell+\gamma+\text{MET}} < 200$
$W\gamma$	466,181	118,744	67,248	66,557
$Z\gamma$	458,388	18,117	13,190	13,170
Z + jets	1,104,675	27,511	22,027	21,997
W + jets	53,327	32,172	28,871	28,855
$t\bar{t}\gamma$	2,483	1,514	581	572
$t\bar{t}$	3,459	2,148	604	602
Top	1,152	449	316	311
$VV\gamma$	625	397	101	99
EWKV γ	6,650	2,928	1,168	1,126
2γ	22,336	585	291	291
Signal	7,600	5,698	167	167

Table 5.10: Cutflow for various processes leading to the construction of the W + jets-dominant control region.

Z +jets control region

Similar to the W +jets control region, the Z +jets control region is designed to gain an understanding of background events involving electrons faking photons. These events are modeled using the Z +jets MC simulation, where the misidentified photon comes from

one of the electrons from the Z boson decay. In the actual analysis, electron faking photons estimation must be validated with data-driven methods introduced in Section 6.2. Since the Z +jets events do not have any neutrino in the leptonic channel, they contribute much more in the low $E_{\text{T}}^{\text{miss}}$ region. Cuts similar to those for the signal region are used except the $E_{\text{T}}^{\text{miss}}$ and $p_T^{\ell+\gamma+\text{MET}}$ cuts, which are reversed:

- $p_T^{\ell+\gamma} > 100$ GeV,
- $p_T^{\ell+\gamma+\text{MET}} < 200$ GeV.

As a result, the Z +jets contribution in the signal region is estimated to be on par with that of the $W\gamma$ contribution.

Background	Cuts (in GeV)			
	Initial Cuts	$E_{\text{T}}^{\text{miss}} < 50$	$p_T^{\ell+\gamma} > 100$	$p_T^{\ell+\gamma+\text{MET}} < 200$
$W\gamma$	466,181	347,437	136,343	134,012
$Z\gamma$	458,388	440,271	46,480	46,026
$Z + \text{jets}$	1,104,675	1,077,161	130,673	129,579
$W + \text{jets}$	53,327	108,092	13,441	13,277
$t\bar{t}\gamma$	2,483	890	564	555
$t\bar{t}$	3,459	1,312	900	898
Top	1,152	703	241	239
$VV\gamma$	625	228	126	122
EWKV γ	6,650	3,722	2,156	2,007
2γ	22,336	21,752	8,610	8,526
Signal	7,600	1,902	1,250	1,208

Table 5.11: Cutflow for various processes leading to the construction of the $Z + \text{jets}$ -dominant control region.

$Z\gamma$ control region

Although the analysis targets events with only one lepton, the standard model $Z\gamma$ process, which typically produces two leptons, is also considered one of the main backgrounds. The $Z\gamma$ events could pass the signal selections if one of the electrons produced is misidentified or is not reconstructed by the algorithm.

The $Z\gamma$ control region is designed to have the $Z\gamma$ contribution dominate. The $Z\gamma$ control region is required to have the following $Z\gamma$ initial cuts:

- Exactly one pair of same flavor opposite sign leptons above 27 GeV,
- Exactly one photon in the event,
- No b-jet.

These initial cuts are different from the ones defined for other regions. The two-lepton selection specifically targets the $Z\gamma$ process and filters out other processes effectively. Then, the event is required to have $|m_{\ell,\gamma} - m_Z| < 15$ GeV and $p_T^\gamma > 50$ GeV. The cutflow of these selections is shown in Table 5.12

Background	Cuts		
	$Z\gamma$ Initial cuts	$ m_{\ell,\gamma} - m_Z < 15$ GeV	$p_T^\gamma > 50$ GeV
$Z\gamma$	71,727	47,033	14,308
Z + jets	35,414	12,861	1,764
$t\bar{t}\gamma$	257	50	22
$t\bar{t}$	12	3	0
EWKV γ	366	282	151
VV γ	171	50	26
$W\gamma$	7	1	0
W + jets	0	0	0
2γ	0	0	0
γ + jets	0	0	0
Single Top	0	0	0
Signal	5	3	2

Table 5.12: Cutflow for various processes leading to the $Z\gamma$ -dominated Control Region.

Unblinding procedures

The official ATLAS experiment regulation requires the analysis to be blinded, preventing the researchers from looking at data in the signal region to reduce bias. However, the

MC sample statistics and data in regions other than the signal region are safe to be looked at. As shown in this section, all regions in this analysis were chosen to be completely orthogonal. As a result, the signal region phase space cannot be probed by looking at other regions.

The ATLAS collaboration requires all analysis groups to form editorial boards, which discuss methodologies used in the analysis. After thorough validations by the editorial board committee, the analysis group can unblind the signal region if approved. At the time of writing, the editorial board review is still ongoing. The blinding procedures led us to use Asimov datasets in the signal region to obtain expected results until the unblinding is officially approved.

An Asimov dataset is a pseudo-dataset constructed by sampling from the expected distribution. The number of pseudo-data events in a particular bin is sampled from a Poisson distribution with a mean value equal to the sum of contributions from all background processes. The Asimov dataset was used in the signal regions for all mass points considered in this analysis to ensure the blinding of data. For each masspoint, a separate Asimov dataset is generated based on its expected signal distribution.

5.2.3 Summary

With all the regions defined in the last section, Figure 5.2 and Table 5.2.3 summarize the regions in this analysis. It is important to note that all regions are orthogonal to each other.

The background processes will be estimated in each region and compared with data. In the control regions where signal leakage is limited to under 5% of the total number of events, good agreements with data are expected. As a result, important hyper-parameters of the backgrounds are measured there. After taking these measurements, the updated

background estimation is checked in the validation region. As the final step, the signal region is analyzed with the validated background estimation to search for new physics.

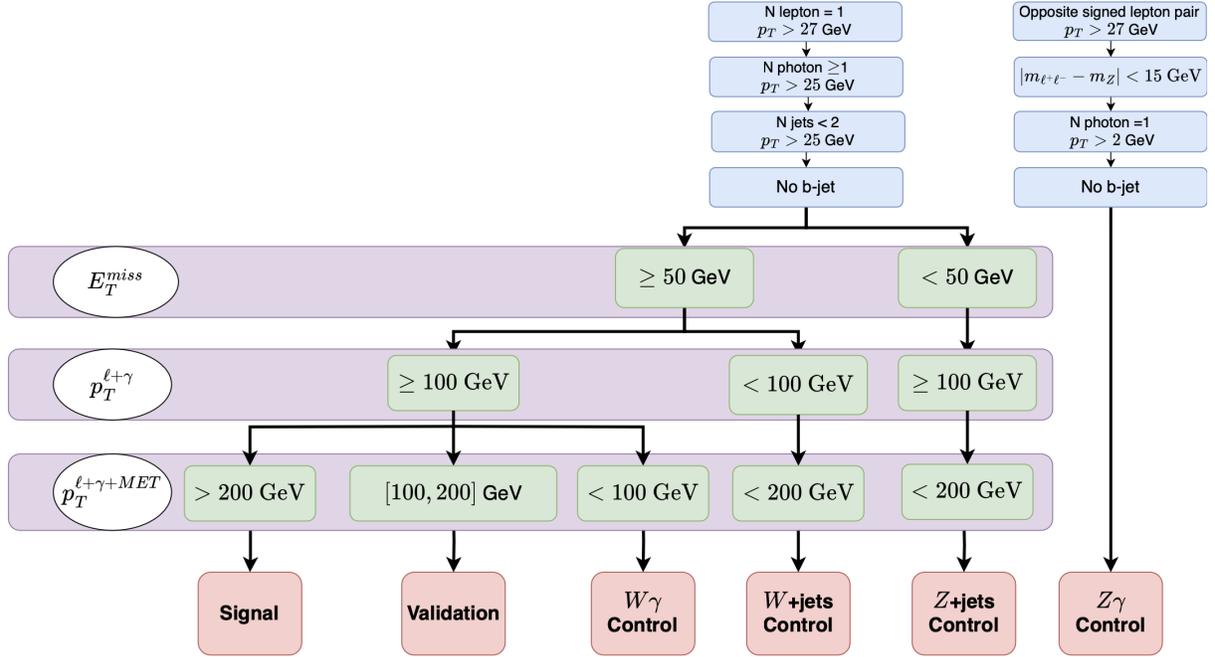


Figure 5.2: Selections implemented in all regions in this analysis.

Cut Type	Signal	Validation	$W\gamma$	$W + \text{jets}$	$Z + \text{jets}$	$Z\gamma$
Lepton	1 ℓ					2 same-flavour opposite-sign ℓ
Photon	$\geq 1\gamma$					1 γ
$ m_{\ell,\gamma} - m_Z $	$> 5 \text{ GeV}$					-
numJets	< 2					-
E_T^{miss}	$> 50 \text{ GeV}$				$< 50 \text{ GeV}$	-
$p_T^{\ell+\gamma}$	$> 100 \text{ GeV}$		$< 100 \text{ GeV}$		$> 100 \text{ GeV}$	-
$p_T^{\ell+\gamma+\text{MET}}$	$> 200 \text{ GeV}$	$> 100 \text{ GeV}$ $< 200 \text{ GeV}$	$< 200 \text{ GeV}$			-
$ m_{\ell,\ell} - m_Z $	-	-	-	-	-	$< 15 \text{ GeV}$
p_T^γ	-	-	-	-	-	$> 50 \text{ GeV}$
$\Delta R_{\ell,\gamma}$	< 1.0	< 1.0	-	-	-	-

Table 5.13: Summary of cuts for each region

Chapter 6

Background Estimation

This chapter presents the background estimation for the analysis. If not otherwise specified, background processes are modeled by MC simulations produced according to the workflow discussed in Section 5.1.2. However, backgrounds with misidentified objects are known to be susceptible to MC mis-modeling. Data-driven methods are used to estimate backgrounds with misidentified photons in the analysis.

In Section 6.1, the overlap removal method used in removing redundant MC events is discussed. There are two major sources of misidentified photons: electrons faking photons (Section 6.2) and jets faking photons (Section 6.3). The fake rates in MC and data are used to correct the estimation of electrons faking photons. For jets faking photons, the template fit method is used to develop a fully data-driven estimate.

6.1 Monte Carlo sample overlap removal

In this analysis, various MC samples are used to estimate background processes. These MC samples are typically generated independently with their generator-level selections that do not necessarily guarantee their orthogonality. For example, the leading back-

ground processes, $W\gamma$, and $Z\gamma$, include events with at least one final state photon. At the same time, W +jets and Z +jets events are inclusive samples that do not place any selection on the final state photon, which results in a potential overlap with the $V\gamma$ ($W\gamma$ and $Z\gamma$) samples. This parameter space overlap in MC samples is not to be confused with the object overlap introduced in Section 4.6, where the former is in the phase space of kinematic and dynamic variables while the latter is a spatial overlap.

A dedicated overlap removal scheme is used to treat overlapping events to ensure no double-counting. As shown in Table 6.1, two overlapping regions are defined for the removal algorithm.

Region	MC samples
No truth photon	Z +jets, W +jets, $t\bar{t}$
With truth photon	$W\gamma$, $Z\gamma$, EWK $V\gamma$, $VV\gamma$, $t\bar{t}\gamma$

Table 6.1: Overlap regions of various MC samples.

The multi-photon final states are not targeted in this analysis and are thus modeled by the high-order terms in the existing samples in Table 6.1. For those samples classified in the “no truth photon” region, an overlap removal is needed if a truth photon is found in any of their events. The event is discarded if a truth photon is found outside a cone of $\Delta R = 0.1$ from any lepton in the $\eta - \phi$ plane of an event from a “no truth photon” region sample.

6.2 Electrons Faking Photons

Due to the electron/photon ambiguity in the reconstruction algorithm discussed in Section 4.2 and Section 4.3, a major source of fake photons comes from electrons. This analysis considers a baseline selection of one electron + one photon. Such a selection is susceptible to di-electron contamination, in which one of the electrons fakes a photon to

yield an event with one electron + one photon. The easiest method to limit the contamination from the process is to include a cut on the invariant mass $m_{e\gamma}$ around to the Z mass window to limit Z +jets contributions. In the analysis regions defined in Table 5.2.3, this invariant mass cut with a 5 GeV window has already been included. However, this still leaves a small residual contribution of electrons faking photons outside the window.

Previously, the Z +jets MC simulation has been used to model events with electrons faking photons. In this section, a data-driven technique is implemented to validate the modeling of the MC simulation. To quantitatively measure the misidentification, it is convenient to define the Fake Rate ($F_{p_T, \eta}^{e \rightarrow \gamma}$), which represents the probability of a given electron faking a photon. To capture the p_T and η dependence, the fake rates are measured in slices of p_T and η with the following binning:

$$p_T \text{ bins} = [27, 35, 45, 55, 70, 90, 120, 2000] \text{ GeV},$$

$$|\eta| \text{ bins} = [0, 0.6], [0.6, 1.37], [1.52, 1.82], [1.82, 2.37].$$

The fake rate is measured using the tag and probe method. First, two control regions, which are orthogonal to all other analysis regions, are selected. The $e\gamma$ region is defined as:

- Exactly one tight electron with $p_T > 27$ GeV,
- Exactly one tight photon with $p_T > 27$ GeV,
- More than one jet with $p_T > 25$ GeV,
- No b-jet.

The di-electron (ee) region is defined as

- Exactly two tight electrons with $p_T > 27$ GeV,

- No tight photon with $p_T > 27$ GeV,
- More than one jet with $p_T > 25$ GeV,
- No b-jet.

The two regions above are required to have more than one jet to stay strictly orthogonal to the analysis regions. The rate is measured by taking the ratio of the number of electrons faking photons to the number of real electrons. In each slice of p_T and η , the invariant masses of $m_{e\gamma}$ and m_{ee} are built in the $e\gamma$ and ee regions, respectively. The invariant mass distribution consists of a polynomial-like background and a "signal" component that corresponds with the resonance of the Z boson. In the ee region, the number of events from the signal resonance represents the number of events with correctly tagged electrons. In the $e\gamma$ region, it is expected that there should be no resonance of $m_{e\gamma}$ near the Z mass, and therefore, the signal component represents the number of events with electrons faking photons.

In each slice of p_T and η , the number of events is measured in a window around the Z mass with invariant masses between 69.188 GeV and 110.188 GeV. The choice of binning is subject to an arbitration addressed by the systematic variation discussed in Section 7.3.

$$F_{p_T, \eta}^{e \rightarrow \gamma} = \frac{N_{e\gamma}^{sig}(p_T, \eta)}{N_{ee}^{sig}(p_T, \eta)}. \quad (6.1)$$

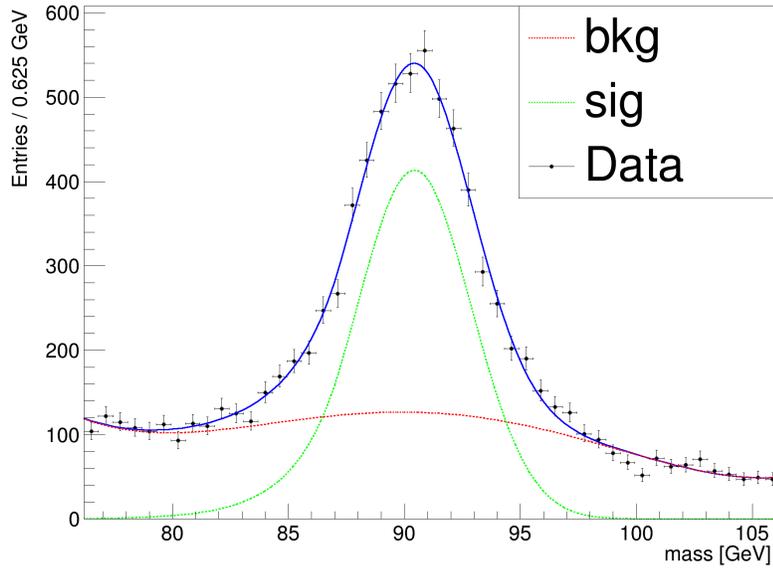


Figure 6.1: Fitting crystal ball (signal) and polynomial (background) functions to the invariant mass histogram of electron and photon. The data is obtained from the bin of $p_T > 27$ GeV and $|\eta| < 0.6$ with the full run-2 data.

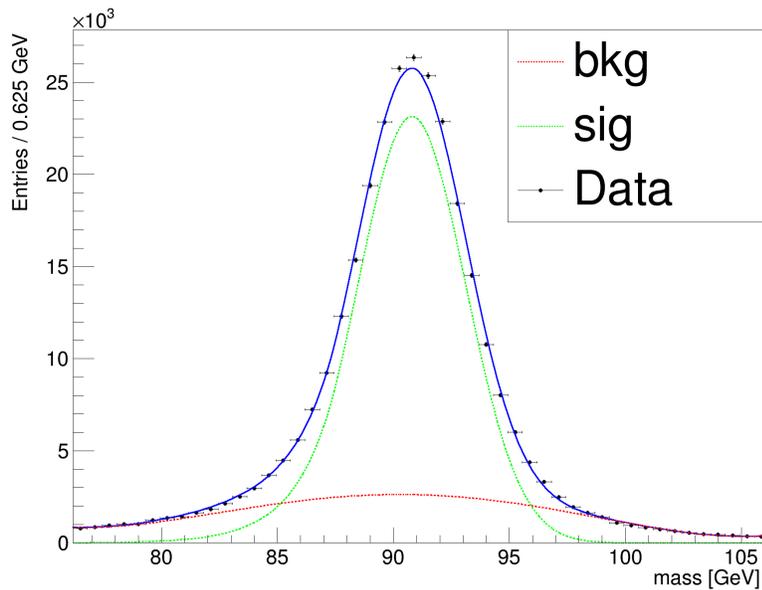


Figure 6.2: Fitting crystal ball (signal) and polynomial (background) functions to the invariant mass histogram of two electrons. The data is obtained from the bin of $p_T > 27$ GeV and $|\eta| < 0.6$ with the full run-2 data.

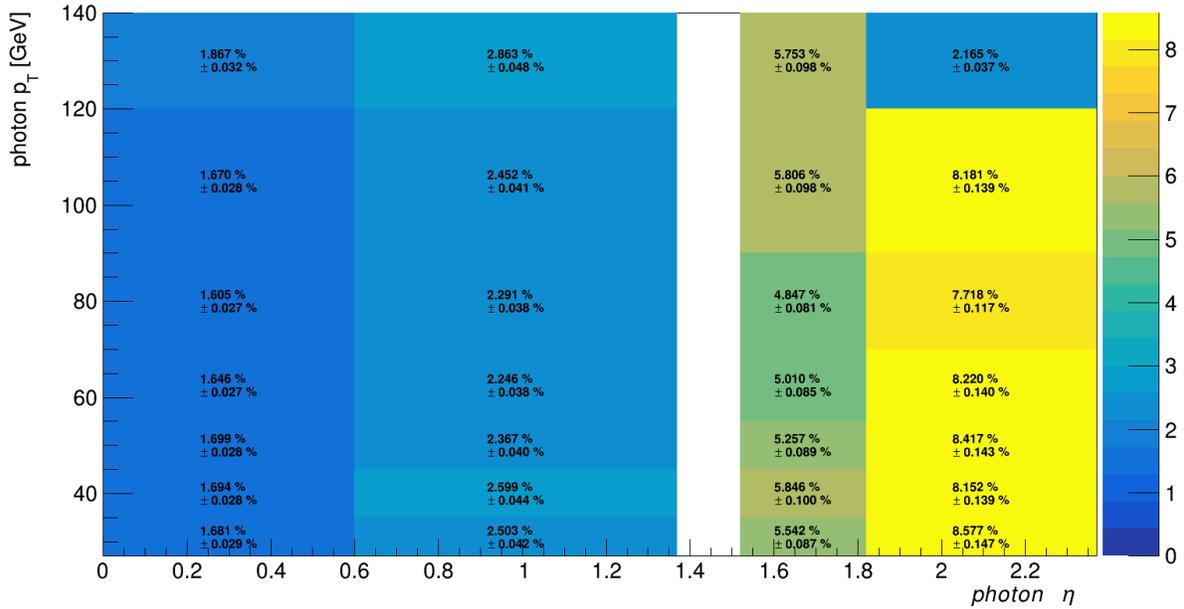


Figure 6.3: Fake rates with respect to p_T and $|\eta|$ in MC simulations. The last p_T bin is partially cut for better display.

N^{sig} in Equation 6.1 is the integrated number of events of the signal components that represent the Z resonance. To recover the signal component, the fits are performed with the signal component modeled by the double-sided crystal ball function and a fourth-order polynomial for the background. Fit results in the bin $p_T > 27$ GeV and $|\eta| < 0.6$ are shown in Figure 6.1 and Figure 6.2. Such fits are performed in all bins except for the $|\eta|$ bin between 1.37 and 1.52, where the photon object selection is excluded.

After the fits, the signal components are integrated to obtain N^{sig} in both regions and substituted the values to Equation 6.1 to get the fake rate. The fake rate is computed in both MC and data, with their measured fake rates shown in Figure 6.3 and Figure 6.4, respectively.

The ratio of the fake rates between data and MC is used to validate the modeling. The ratio is shown in Figure 6.5 with uncertainties that will be later introduced in Section 7.3. The ratio is close to one in most bins within the uncertainties, suggesting that the MC

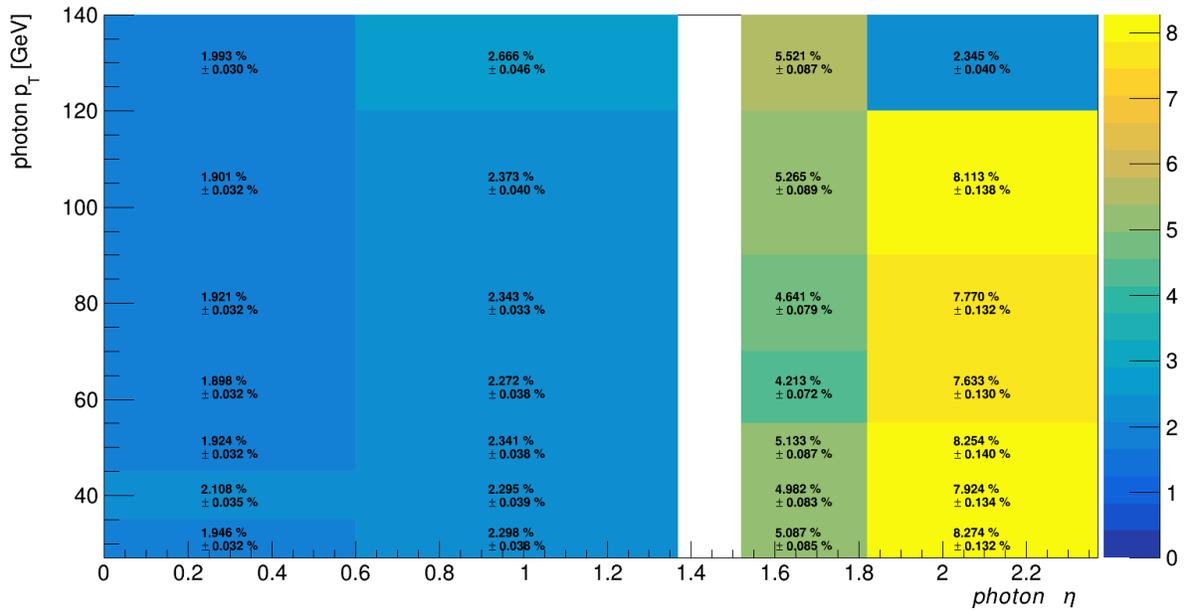


Figure 6.4: Fake rates with respect to p_T and $|\eta|$ in the full run-2 data. The last p_T bin is partially cut for better display.

simulation provided a reasonable estimation of misidentified photons from this particular source. Therefore, the MC simulations are used in the analysis to model the $e \rightarrow \gamma$ fake contributions in the analysis regions with the ratios shown in Figure 6.5 as corrections. The ratio is applied in each event to re-weight the event for a correct estimation.

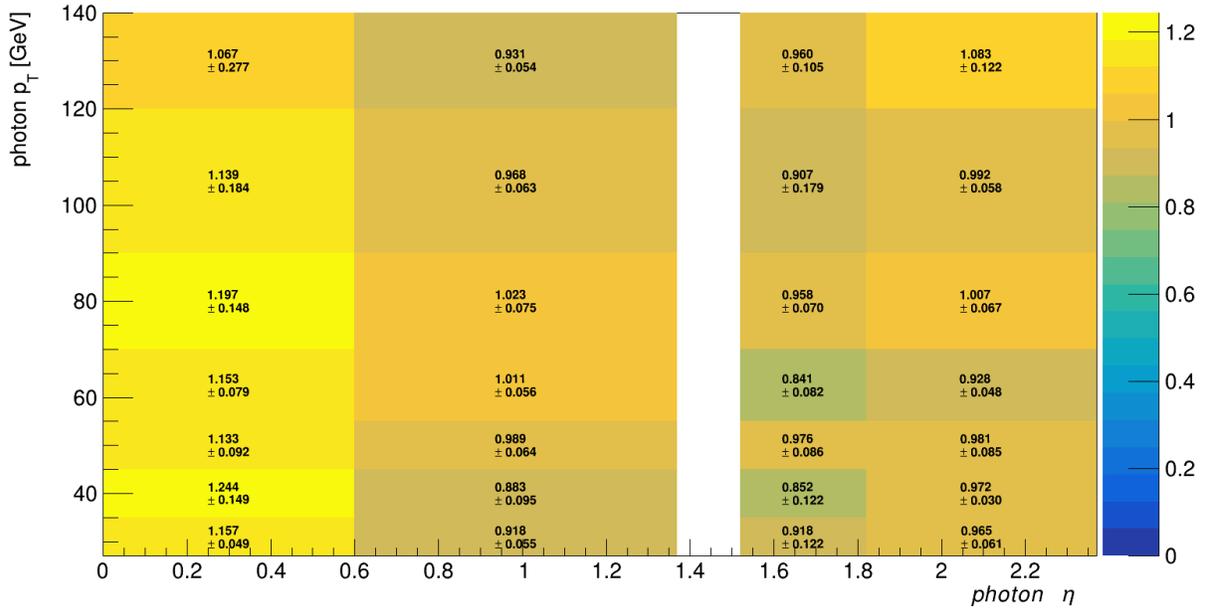


Figure 6.5: Fake rate ratio between data and MC with respect to p_T and $|\eta|$ in the full run-2 Data. The ratio is defined to be $\frac{FR_{data}}{FR_{MC}}$. The last p_T bin is partially cut for better display.

6.3 Jets Faking Photons

Another major source of misidentified objects in this analysis is jets faking photons, sometimes called the “QCD fakes”, given the QCD nature of hadronic jets. These non-prompt photons are known to be poorly modeled by MC simulation and thus require a data-driven approach.

For jets faking photons, it is straightforward to implement the template fit method, which estimates the fake photon contribution by deriving and fitting templates of real and fake photons. The variable in which the template is derived is taken to be the isolation energy (E^{iso}). The isolation energy is defined using E_T^{topo40} , which is the sum of transverse energy deposits of topological clusters within the cone of $\Delta R < 0.4$ centered at the photon. Equation 6.2 gives the exact definition of isolation energy:

$$E^{iso} = E_T^{topo40} - 0.022 \cdot p_T^\gamma. \quad (6.2)$$

The photon isolation energy is a good discriminating variable between real and QCD-fake photons. Since the real photons' energy deposits are more collimated, their isolation energies are much smaller. However, the misidentified jet occupies a much larger radius, typically a cone with $\Delta R = 0.4$. The large area will result in a much higher energy deposit from other jet constituents around the reconstructed photon and hence have much higher isolation energy. Therefore, by deriving isolation energy distributions for real and QCD-fake photons, they could be applied to the data to form a data-driven estimate of the QCD-fake photon contribution in the analysis regions. The workflow of the template fit method is given as the following:

1. Obtain the isolation energy templates of real photons (real template):

MC simulations are used to obtain the isolation energy templates of real photons. An example of the real photon template is shown in Figure 6.6, in which a Bukin function [89] was fitted to the simulated isolation energy of real photons from the $W\gamma$ MC samples.

2. Obtain the isolation energy templates of fake photons (fake templates):

The particle identification (PID) criteria are changed from Tight to Loose to construct a fake-enriched region to obtain the isolation energy templates of fake photons. The photons are specifically required to fail the tight selection but pass the LoosePrime4 selection, which selects photons passing the Loose selection but failing at least one of the four shower shape cuts in the Tight selection [90]. The LoosePrime4 selection ensures that the new region, where the fake template is to be fitted, is fake-dominant while being orthogonal to the analysis region that uses the Tight selection. The real and fake templates are fitted in this region, whereas another Bukin distribution models the fake template. The MC estimation constrains the contribution of real photons in this region. The shape of the simple template is fixed with the normalization factor floated. An example of such a fit is shown in Figure 6.7.

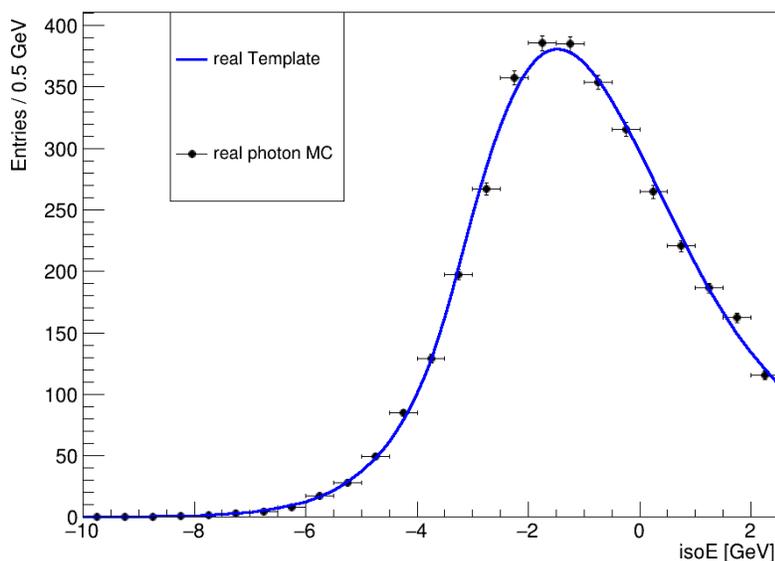


Figure 6.6: Isolation energy distribution of real photons in the bin of $160 \text{ GeV} < m^{\ell+\gamma+\text{MET}} < 180 \text{ GeV}$ of the $W\gamma$ control region.

3. The final fit:

The isolation energy templates of both the real and fake photons had been obtained at this point. These templates are then applied in the actual bins of the analysis regions where photons are required to pass the tight PID and isolation cuts. Figure 6.8 shows an example of this final fit. The fake photon contribution is obtained by integrating the fitted fake template up to 2.45 GeV, the cut on the isolation energy of tight photons.

The data-driven methods require sufficient data to work robustly. As a result, the procedure mentioned above was repeated for each bin of all analysis regions to obtain fake contributions unless there were less than 10 events with loose photons in that bin. This procedure ensures the template fit is not performed in statistically limited bins with negligible fake contributions. For those bins where the template fit was not performed, the fake photons originating from jets were estimated to have no contribution. The systematic uncertainties associated with this method are discussed in Section 7.3.

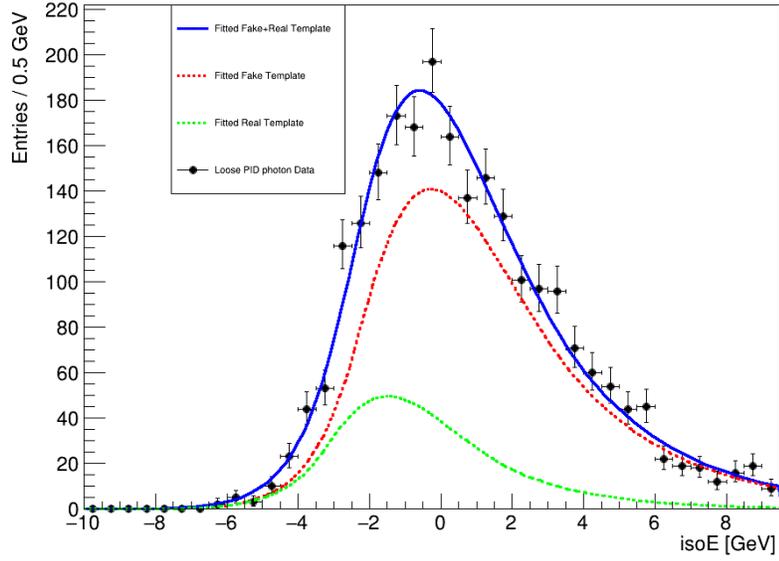


Figure 6.7: Isolation energy distribution of real and fake photons in the bin of $160 \text{ GeV} < m^{\ell+\gamma+\text{MET}} < 180 \text{ GeV}$ of the $W\gamma$ control region.

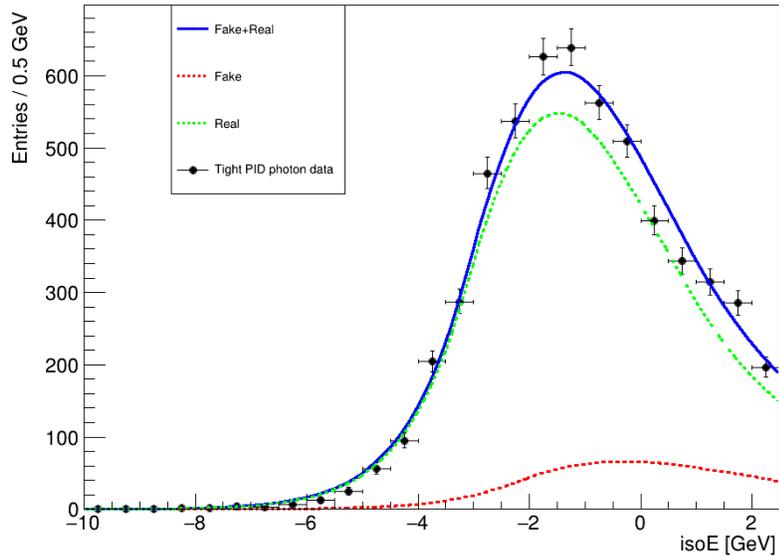


Figure 6.8: The final fit in the bin of $160 \text{ GeV} < m^{\ell+\gamma+\text{MET}} < 180 \text{ GeV}$ of the $W\gamma$ control region. The cut at 2.45 GeV results from the tight isolation cut required by the tight photon definition as introduced in Section 4.3.

To validate the method, pseudo-datasets are reconstructed using the MC simulation of $W\gamma$ and W +jets processes. These events are weighted using their corresponding MC weights when they are mixed in the pseudo-datasets. The $W\gamma$ samples are used to model real photons, and the W +jets samples are used to model jets faking photons. If the photon in the W +jets event was not from a jet, the event is discarded to ensure the purity of the jets faking photons process. Due to the lack of the statistics of MC simulation in the signal and validation regions, the validation tests are performed in the $W\gamma$ and W +jets control regions where significant jets faking photons contamination was expected. The comparisons of truth contribution versus estimations are shown in Figure 6.9 and Figure 6.10. Good agreement between the estimations and truth values is observed in both the W +jets and $W\gamma$ control regions. It is also clear that the fake contamination varied substantially from bin to bin and became nearly negligible at higher masses.

The background estimations introduced in this chapter will be used as nominal background modeling in the analysis to search for the charged Higgs boson. However, the systematic uncertainties must be estimated for all modeling approaches to allow the statistical treatment of the experimental data. The next chapter will discuss various systematic uncertainty sources associated with the background estimations.

W+jets control region closure test

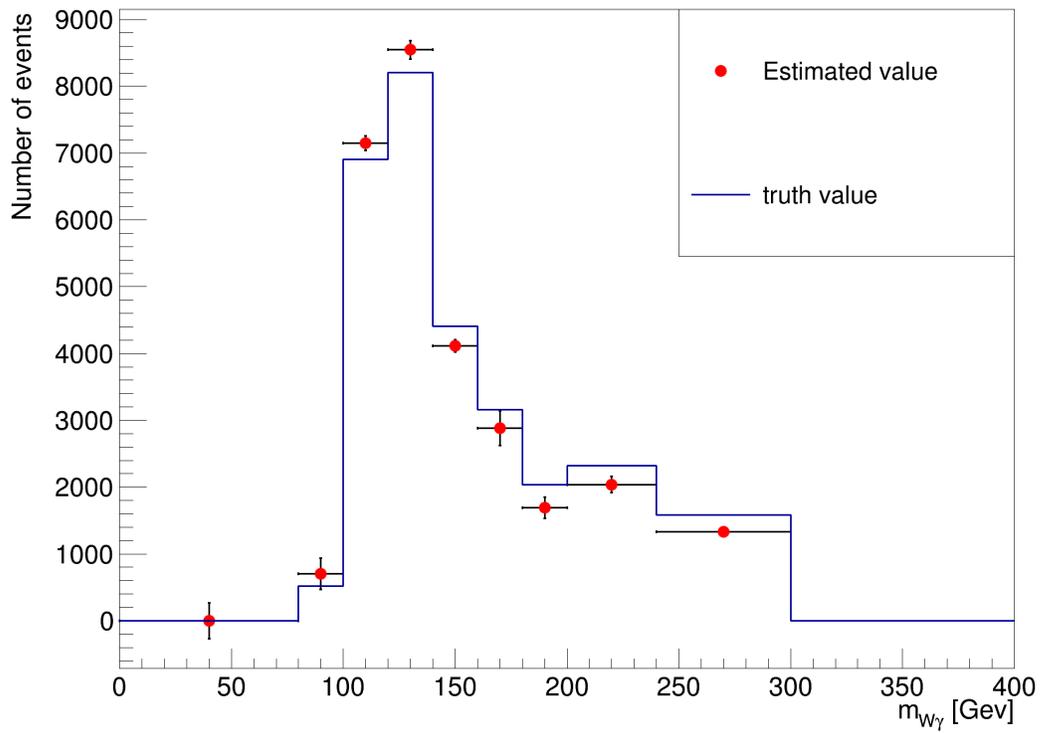


Figure 6.9: Comparison of the truth information and template fit estimate in the W +jets control region with the pseudo-dataset generated from $W\gamma$ and W +jets sample.

$W\gamma$ control region closure test

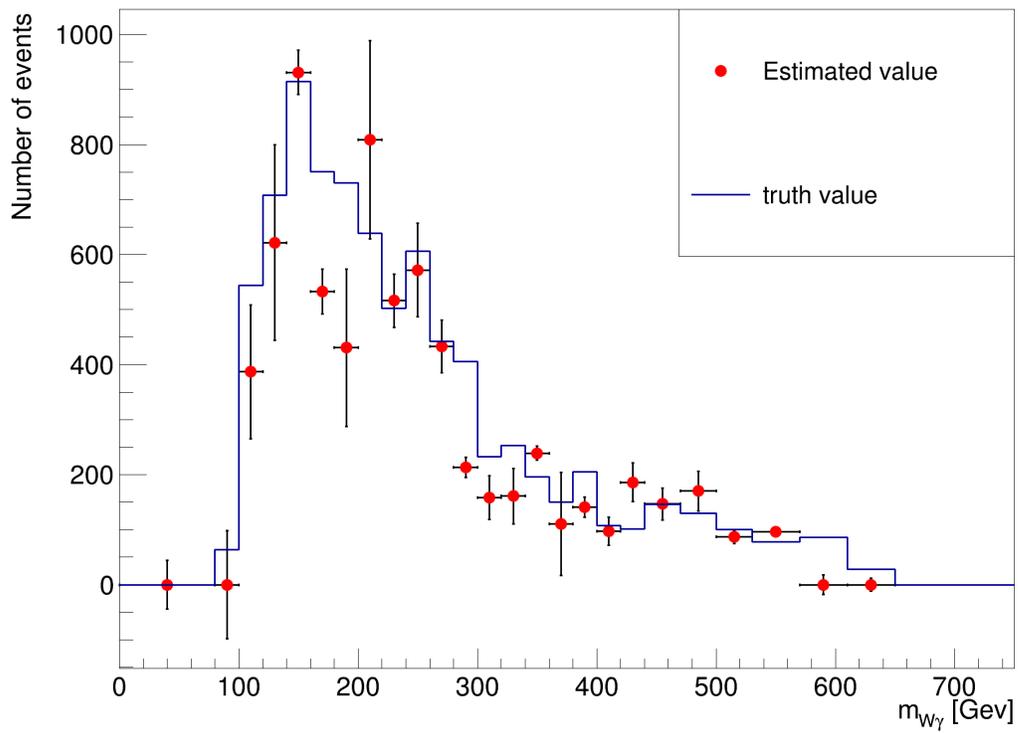


Figure 6.10: Comparison of the truth information and template fit estimate in the $W\gamma$ control region with the pseudo-dataset generated from $W\gamma$ and W +jets sample.

Chapter 7

Systematic Uncertainty

In this chapter, all systematic uncertainties (systematics) considered in this analysis are discussed. Systematic uncertainties are measurement errors that statistical treatments cannot mitigate. In other words, systematic uncertainties deal with variations of certain hyper-parameter choices. These systematic uncertainties are categorized into three main categories. In Section 7.1, the systematic uncertainties related to detectors are introduced. Section 7.2 discusses the theoretical systematic uncertainties. In Section 7.3, the systematic uncertainties associated with data-driven estimates introduced in Section 6.2 and Section 6.3 are discussed.

7.1 Detector-related Systematics

In the analysis, MC samples were generated with the detector responses that account for various detector limitations and reconstruction algorithms. The systematic variations of their associated nuisance parameters result in uncertainties that will eventually propagate through the statistical analysis. To account for these uncertainties, various working groups in the ATLAS collaboration worked to provide weight systematics, mainly for ob-

jects used in the analysis. For each systematic source, the uncertainty is recorded in the event weight corrections, typically with small deviations from unity. The full list of the systematics of this source is listed in Appendix E.1.

E_T^{miss} uncertainties

There are three E_T^{miss} variations considered for the uncertainty treatment [80]. The systematics involve the magnitude scaling of the soft term in the E_T^{miss} and also the variation in the Gaussian smearing of the soft term magnitudes in directions parallel or perpendicular to the vector sum of all hard objects' transverse momenta.

Leptons

In the reconstruction of leptons, scale factors of various forms are introduced to address the difference between data and simulated MC samples [68, 71]. These typically contain scale factors related to the identification, reconstruction, triggers, and isolation working points. For electrons, the variations of the energy resolution and scale are also used as systematic uncertainties. For muons, there are also extra uncertainties related to tracking devices, such as the ID smearing and the MS tracking smearing.

Photons

The photon systematics are associated with the identification and isolation working points. The variations in these working points are stored as scale factors in the datasets. Similar to the electron case, the analysis also considers the energy resolution and the energy scale of the photon as sources of systematic uncertainties.

Jets

Systematic uncertainties of jets are those developed by the JetEtMiss Combined Performance group. The two categories are Jet Energy Scale (JES) [91] and Fractional Jet p_T Resolution (JER) [92].

The JES uncertainties are derived from test-beam data, collision data, and MC simulations. These involve variations of nuisance parameters, η inter-calibration, flavor physics, pile-up corrections, and the JVT algorithm. Another important uncertainty for jets is associated with the punch-through effect, which models high-energy jets that escape the calorimeter. The JER uncertainties are obtained by taking the quadratic difference between the energy resolution in data and MC samples.

Flavour tagging

The flavor tagging uncertainties are those obtained from eigenvector variations [93]. The eigenvector decomposition was performed using the covariance matrix built from all relevant systematics. The leading eigenvector variations in each jet category are used and listed in Appendix E.1. In addition to eigenvector uncertainties, there are also sources of systematics from the extrapolation in the high p_T phase space.

Luminosity

The luminosity uncertainty of the full run-2 dataset is taken to be 0.83% [35].

7.2 Theoretical Uncertainties

First, the normalizations of the main backgrounds, W_γ and Z_γ , are floated to be fitted in the control regions. Then, there are also two sources of systematics associated with the theoretical modeling of processes: the Parton Distribution Functions uncertainties and the scale uncertainties. These uncertainties were evaluated separately and combined in quadrature using the PMGSystematicsTool [94].

Parton Distribution Function uncertainties

In this category, the PDF sets of MC samples are varied to estimate the impact of PDF-related systematics. The baseline PDF sets of main backgrounds like W_γ and Z_γ MC samples are modeled by the NNPDF30_nnlo_as_0118_hessian PDF sets, which model the physical processes up to the Next-to-Next-to-Leading Order (NNLO). The variations are then recorded by comparing them with PDF sets of PDF4LHC15_nnlo_30_pdfas. For minor background processes like Di-photon, γ -jets, and Single Top processes, the baseline PDF set is NNPDF30_nnlo_as_0118, while the alternative PDF set used for comparison is MMHT2014nnlo68cl. These PDF sets information is stored in the derived dataset and can be accessed easily during the analysis. Table 7.1 shows various processes and their PDF sets.

Process	Nominal PDF set	alternative PDF set
$W_\gamma ; Z_\gamma$:	NNPDF30_nnlo_as_0118_hessian	PDF4LHC15_nnlo_30_pdfas
Di-photon ; γ jets ; Single Top	NNPDF30_nnlo_as_0118	MMHT2014nnlo68cl
Signal	NNPDF30_nlo_as_0118	MMHT2014nlo68clas118

Table 7.1: A table of the PDF sets used in the MC sample with alternative PDF sets.

The PDF systematic uncertainty is estimated in the analysis by calculating the differences between histograms produced by nominal and alternative PDF sets. This calculation is performed in each bin of the histograms used in the analysis.

Renormalization and factorization scale uncertainties

Other than the PDF sets variation, the renormalization scale μ_R and the factorization scale μ_F were also considered major sources of theoretical systematic uncertainties. The scales were varied among three choices: 1 (nominal), 0.5, and 2, giving the following combinations of the two scales in a pairwise fashion:

$$\{(\mu_R, \mu_F)\} = \{(0.5, 0.5), (1, 0.5), (0.5, 1), (1, 1), (2, 1), (1, 2), (2, 2)\}. \quad (7.1)$$

In each bin of the analysis histograms, the difference between the nominal value and the varied value produced by each of these combinations was evaluated. The systematic uncertainty in a bin was taken to be the maximal difference out of all combinations.

7.3 Systematics of Data-driven Estimations

The data-driven methods were introduced in Section 6.2 and Section 6.3 to estimate misidentified objects. Various systematic variations were considered for these methods to account for the systematic uncertainties. In this section, the sources of systematics for the two data-driven methods are discussed separately.

The overall systematic uncertainty was calculated in each data-driven process by taking the maximum variation out of all sources. The uncertainty was also symmetrized such that the up and down variations were identical.

7.3.1 Electrons faking photons

The following variations were considered for the data-driven estimation of electrons faking photons.

- The fit errors. These are the uncertainties returned by the fit.
- Varying the invariant mass window used in the fit to be 15 GeV instead of the nominal 20 GeV.
- Switching the fitting function of the signal component from a Double-Sided Crystal Ball function to a Gaussian function.

7.3.2 Jets Faking Photons

The systematic variations for jets faking photons are:

- Varying the binning by combining adjacent bins. For example, the binning of the $W\gamma$ control region is changed as the following:
 - Nominal binning(GeV): 0,125,150,175,200,225,250,275,300,350,400,500,600,700
 - Varied binning(GeV): 0,150,200,250,300,400,500,700

The fake yield in the varied binning is computed and compared with the sum of fake yields from its corresponding bins in the nominal binning. A ratio of the variation is determined and assigned to the nominal value as the systematic variation of binning choice.

- Changing the LoosePrime4 PID working-point to LoosePrime2. This will impact the determination of the fake photon template in each bin and hence varies the estimation of fakes in the final fit.

At this point, all the ingredients for the data analysis have been introduced. The next chapter will proceed to start developing the statistical method that allows for the derivation of cross-section limits of the signal samples using the background estimations and their systematic uncertainties. Systematic uncertainties will go through the pruning procedure, which will be introduced in Section 8.2. This procedure will remove trivial systematics that do not result in changes more than 0.5% of the nominal values. The remaining systematics will be shown in Figure 8.1. The overall uncertainties, including all systematics, are shown in Appendix F. The relative effects of the remaining systematics are summarized in a series of tables from Appendix E.3 to Appendix E.8.

In the end, all of the systematics were fitted to obtain good data-MC agreements in the control regions. The values of these systematics are then extrapolated to the validation region and the signal region. For example, when a slight MC deficit in the $W\gamma$ control region was observed, the normalization factor of the MC sample was then pulled higher to make up the deficit. These effects will be shown in the next chapter.

Chapter 8

Fitting methods and results

With all the ingredients introduced in previous chapters, the analysis results will be discussed in this chapter. The analysis aims to search for BSM processes in the signal region, quantified by an excess of observed events compared to the SM background. The compatibility between data and combined SM background with signal processes is evaluated with statistical treatments.

The likelihood functions used in the statistical fit are introduced in Section 8.1. The fit, which uses background estimates, signal processes, and data, is implemented using the TRExFitter package [95] to extract signal significance and set limits on the free parameters of signal processes. Fitting results using Asimov datasets are shown in Section 8.2.

8.1 Fitting

8.1.1 Profile Likelihood Function

The statistical model starts with a marked Poisson model [96], which is subject to a varying signal strength adjustment. Following the usual convention, the signal and back-

ground events are labeled by S and B, respectively. For the marked Poisson model, a signal strength parameter, μ , is introduced to yield the following Poisson-based model for an arbitrary discriminating variable with a value of x_e in event e :

$$\mathcal{P}(\{x_1, \dots, x_n\}|\mu) = \text{Pois}(n|\mu S + B) \left[\prod_{e=1}^n \frac{\mu S f_S(x_e) + B f_B(x_e)}{\mu S + B} \right], \quad (8.1)$$

where $f_S(x_e)$ and $f_B(x_e)$ are the probability density functions of the signal and background, respectively. With the signal strength parameter, this model reflects the probability of observing n events with an expectation of $\mu S + B$ events. The Poisson distribution is given by $\text{Pois}(n|\nu) = \frac{\nu^n e^{-\nu}}{n!}$. The second term inside the square brackets in Equation 8.1 describes the probability density of obtaining x_e based on the weighted sum of $f_S(x_e)$ and $f_B(x_e)$.

In practice, the number of data events observed in each bin is fixed for the above equation, leaving only the strength μ to vary. The likelihood function, \mathcal{L} , can be constructed for a histogram by considering the product of Poisson probabilities of all bins, as shown in Equation 8.2 from Reference [96]:

$$\begin{aligned} \mathcal{L}(\mu, \vec{\theta}) &= \mathcal{P}(n_b|\mu) \\ &= \text{Pois}(n_{tot}|\mu S + B) \left[\prod_{b \in \text{bins}} \frac{\mu \nu_b^{\text{sig}} + \nu_b^{\text{bkg}}}{\mu S + B} \right] \\ &= \mathcal{N}_{\text{comb}} \prod_{b \in \text{bins}} \text{Pois}(n_b|\mu \nu_b^{\text{sig}} + \nu_b^{\text{bkg}}). \end{aligned} \quad (8.2)$$

where n_b is the data histogram, $\mu \nu_b^{\text{sig}}$ is the number of signal events, and ν_b^{bkg} is the number of background events in a particular bin, b . $\vec{\theta}$ is the set of known nuisance parameters. $\mathcal{N}_{\text{comb}}$ is a constant combinatorial factor that only has impacts on the determination of floated parameters and can, therefore, be neglected.

To obtain limits of a certain parameter μ , the profile likelihood ratio [97], λ , is defined based on the likelihood function:

$$\lambda(\mu) = \begin{cases} \frac{\mathcal{L}(\mu, \hat{\theta}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\theta}(\hat{\mu}))}, & \hat{\mu} \geq 0, \\ \frac{\mathcal{L}(\mu, \hat{\theta}(\mu))}{\mathcal{L}(0, \hat{\theta}(0))}, & \hat{\mu} < 0, \end{cases} \quad (8.3)$$

where $\hat{\mu}$ is the maximum likelihood estimator of the parameter μ and $\hat{\theta}(\mu)$ represents the conditional maximum likelihood estimator of θ with a fixed strength parameter μ .

One of the goals of this analysis is to obtain the upper limits of the cross-sections of the signal processes. This suggests that the test statistics, q_μ , could be defined following the procedures in Reference [97]:

$$q_\mu = \begin{cases} -2 \ln \lambda(\mu), & \hat{\mu} \leq \mu \\ 0, & \hat{\mu} > \mu \end{cases} = \begin{cases} -2 \ln \frac{\mathcal{L}(\mu, \hat{\theta}(\mu))}{\mathcal{L}(0, \hat{\theta}(0))}, & \hat{\mu} < 0, \\ -2 \ln \frac{\mathcal{L}(\mu, \hat{\theta}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\theta}(\hat{\mu}))}, & 0 \leq \hat{\mu} \leq \mu, \\ 0, & \hat{\mu} > \mu. \end{cases} \quad (8.4)$$

The statistics defined in Equation 8.4 were used in this analysis to test the compatibility of the measured data with the SM background plus the signal sample with a certain strength. The main statistical treatment considered is the computation of the p-value, which is defined as the probability of finding a value of the test statistics more extreme than the observed value. In the context of high-energy physics, the p-value is often interpreted as the probability of observing more events than the experimental outcome in the hypothesis of the signal strength being lower than a certain value. By computing the p-values with the statistics mentioned above, the exclusion limits [98] of the signal cross section times branch ratio, $\sigma \times \text{BR}$, could be obtained for each mass level considered.

8.1.2 Fitting Strategy

In all regions, this analysis considered histograms built on the reconstructed Higgs mass, $m_{W\gamma}$. Binned profile likelihood fits were performed using the TRExFitter program [95].

In the control regions, nuisance parameters are determined using the fits and extrapolated to the signal and validation regions. The normalizations of $W\gamma$ and $Z\gamma$ background processes are expressed as free parameters following a uniform distribution instead of the Gaussian distribution used to model other nuisance parameters. In practice, all systematics introduced in Chapter 7, including those related to detector effects, theoretical sources, and data-driven methods, will enter the fit and be determined in the control regions. In order to reduce the computational complexity, all systematics uncertainties in the shape and normalization that result in changes of less than 0.5% of the nominal values are pruned away.

8.2 Fitting Results with Asimov Datasets

The following section shows the results of fits using an Asimov dataset with the signal process at the mass point of 150 GeV. The generation of the signal samples follows the recipe introduced in Section 2.2.2. The free parameters are fixed as shown in Equation 2.38. The cross section and filter efficiency of the simulated signal sample can be found in Table 5.1 and Table 5.3. In the fit, uncertainties were pruned away if they resulted in changes less than 0.5% of the nominal values. In Figure 8.1, the full uncertainty prunings are shown. Since samples have their own systematics list, those systematics that are not found in a particular sample are colored grey in their corresponding boxes in Figure 8.1. The systematics with the green color are the ones kept after the pruning, while the red ones were pruned. The orange color labels the systematics which have their normaliza-

tion dropped, meaning that they were only considered for their impacts on the shape of the distribution. The yellow color labels the systematics that have only the normalization fitted but not the shape.

Another interesting variable is the pull of a parameter, which is a quantitative measurement of the incompatibility between the parameter's input value, θ_0 , and its fitted value that gives the maximum likelihood, $\hat{\theta}$. The pull is defined as

$$\text{pull}(\theta) = \frac{\hat{\theta} - \theta_0}{\Delta\theta} \quad (8.5)$$

where $\Delta\theta$ is the uncertainty of parameter η .

Another interesting parameter for this analysis is the impact, $\Delta\mu$, which measures the change in signal strength when a particular nuisance parameter fluctuates by exactly one standard deviation. The impact is given by

$$\Delta\mu^\pm(\theta) = \hat{\mu}_{\theta_0 \pm \Delta\theta} - \hat{\mu} \quad (8.6)$$

where $\hat{\mu}_{\theta_0 \pm \Delta\theta}$ is the change in fitted signal strength when the nuisance parameter is fixed to $\theta_0 \pm \Delta\theta$ and $\hat{\mu}$ is the fitted value of signal strength that gives the maximum likelihood. The pulls of nuisance parameters are shown in Figure 8.2, while the impacts are shown in the ranking plot in Figure 8.3.

The normalization of the $W\gamma$ and $Z\gamma$ background are floated in the fit. The best-fit values for these normalizations are shown in Figure 8.4. The correlations between the kept systematics are given in the correlation matrix shown in Figure 8.5. These quantities served the purpose of checking the variations before they were extrapolated to the signal region.

After these nuisance parameters are determined, the estimated background distribution varies in normalization and shape. Figure 8.6 to Figure 8.11 shows the pre-fit and post-fit distribution in all regions. The background estimation and Asimov data agreement are visibly improved after the fit is performed. The results of agreements are shown in tables in Appendix F. It could be clearly observed in Table F.1 and Table F.3 that the agreement in the control regions went from around 95% up to nearly 100% after the fitting. It is also clear from Figure 8.6 to Figure 8.9 that the χ^2/ndf labeled on the figure also dropped to near zero after the fitting. Since the fit varies nuisance parameters to reconcile the data-background agreement, it is expected that the χ^2/ndf values in control regions turn out to be small. It could be observed that in the validation and signal region where there is no pulling of the nuisance parameters, the χ^2/ndf values are much higher.

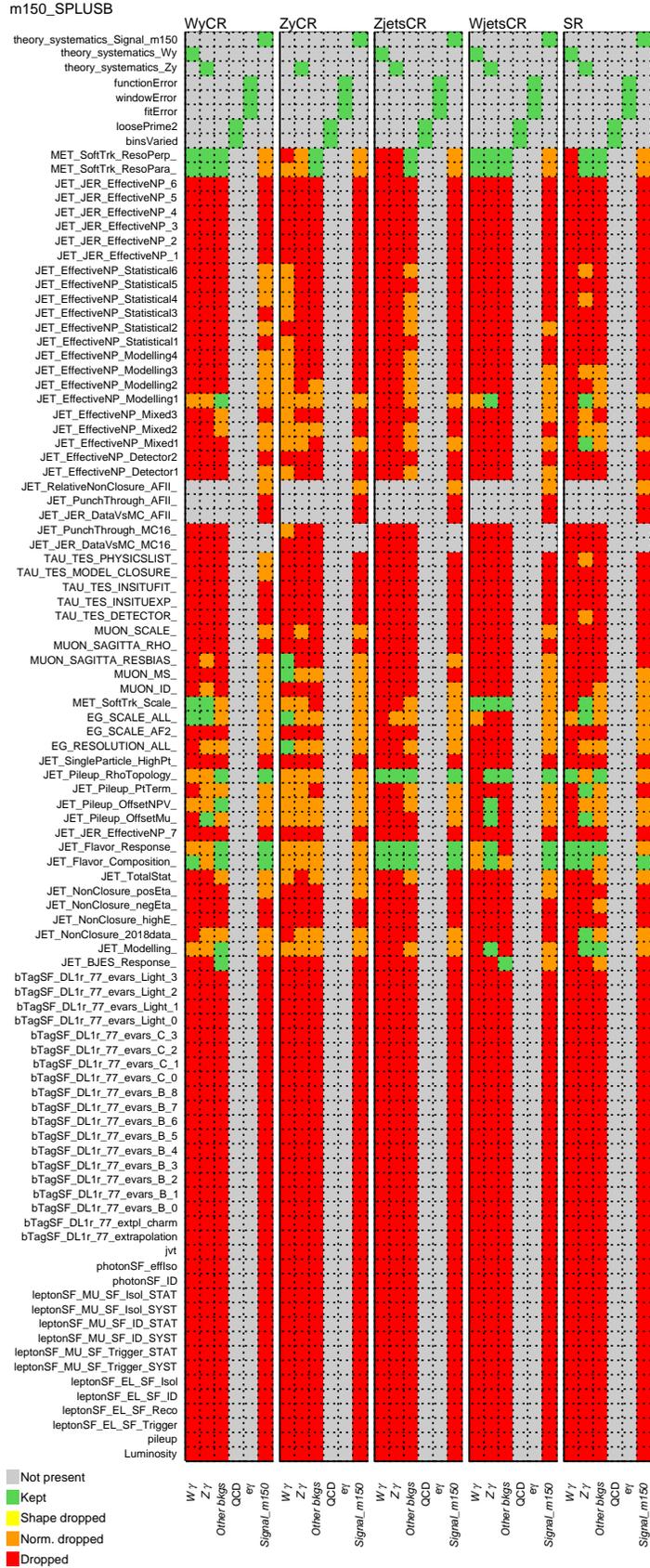


Figure 8.1: Pruning plot of systematics used in the fit for the $m_{H_5} = 150$ GeV mass point.

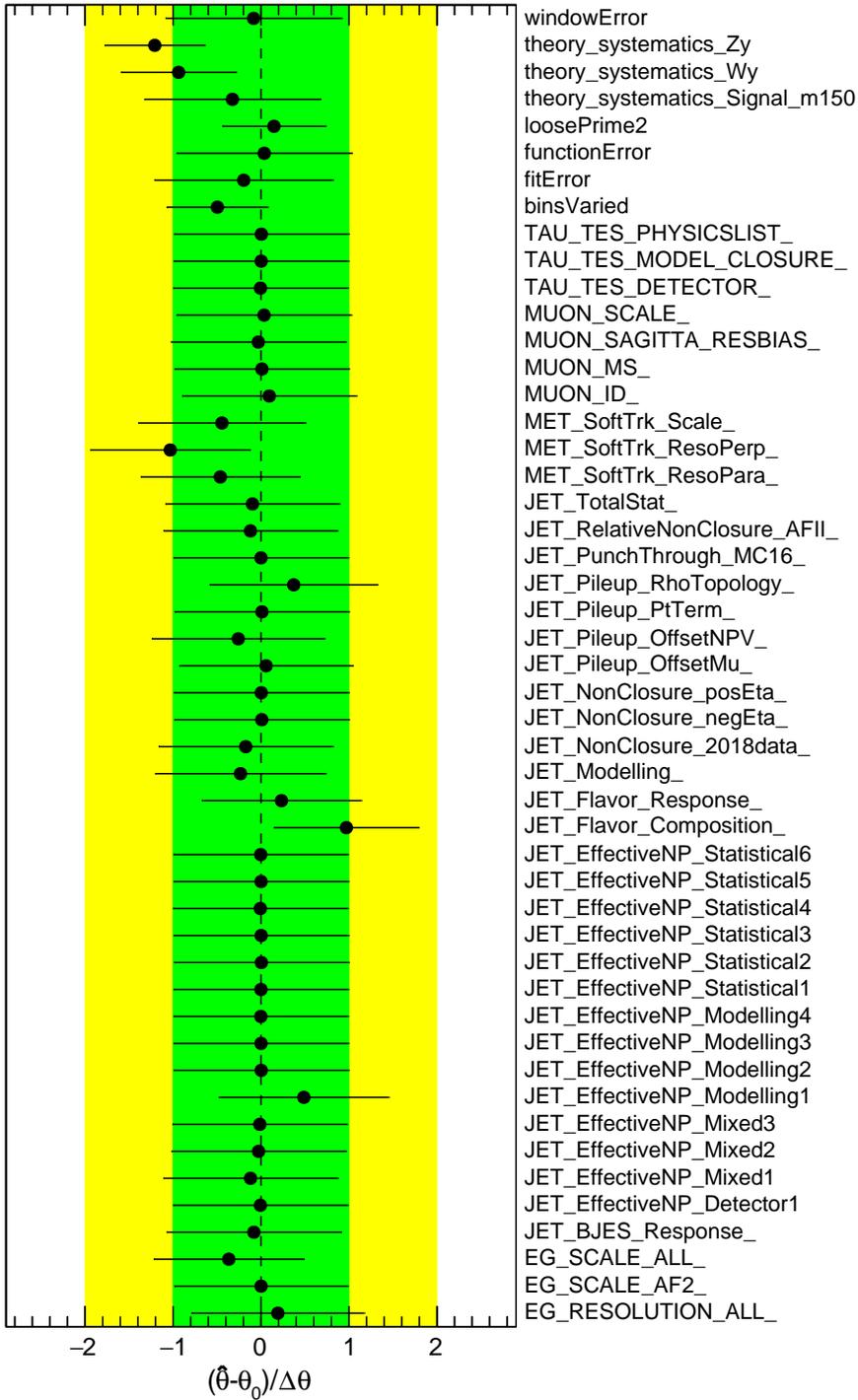


Figure 8.2: Nuisance parameters after the fit for the $m_{H_5} = 150$ GeV mass point.

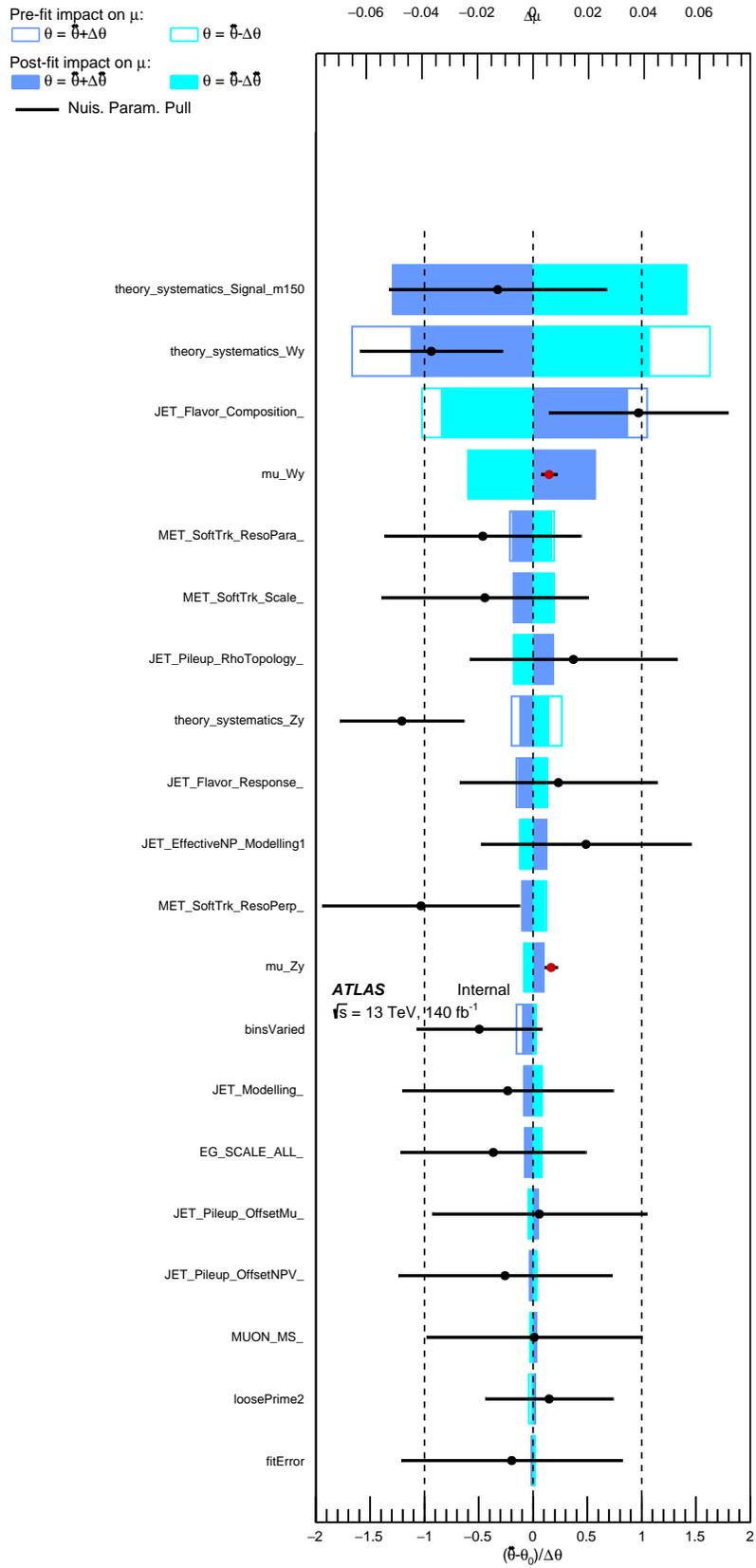


Figure 8.3: Ranking plot of all nuisance parameters after the fit for the $m_{H_5} = 150 \text{ GeV}$ mass point.

ATLAS Internal

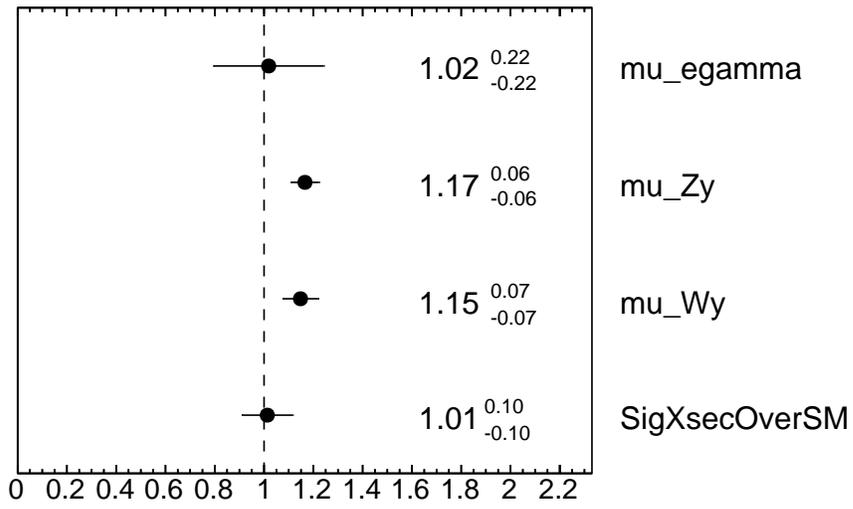


Figure 8.4: Normalisation Factors of various processes' cross sections after the fit.

ATLAS Internal

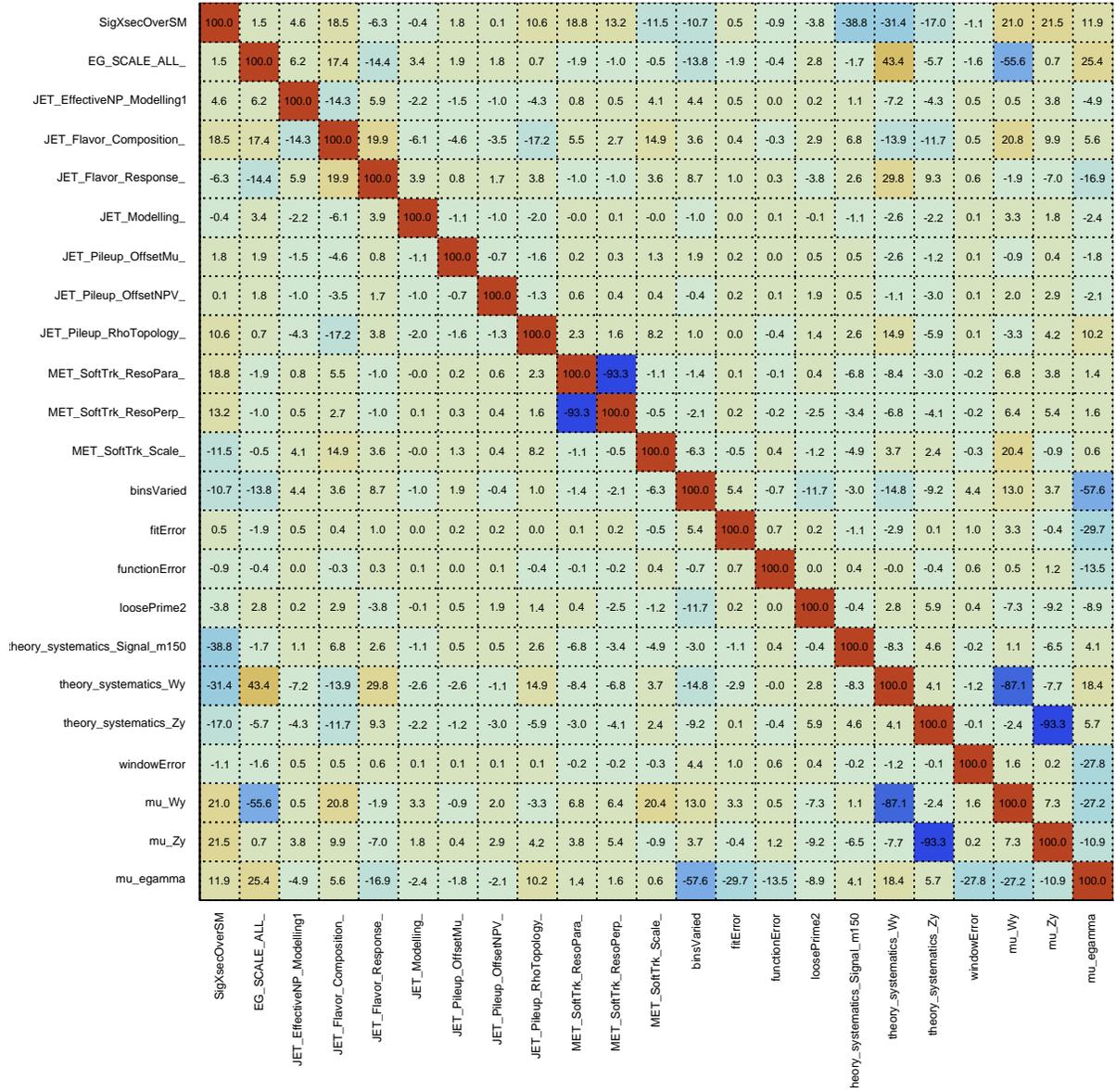
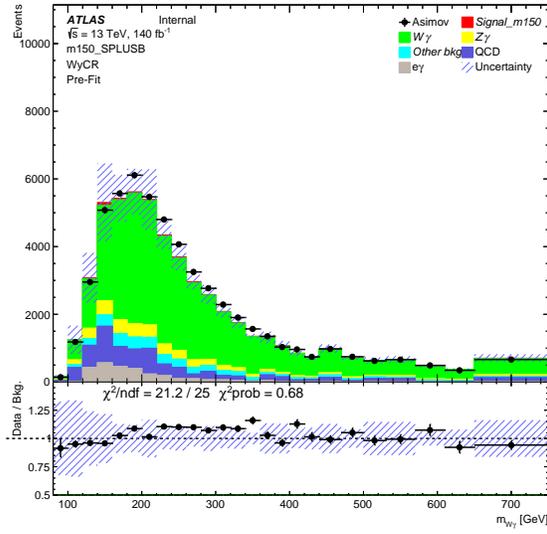
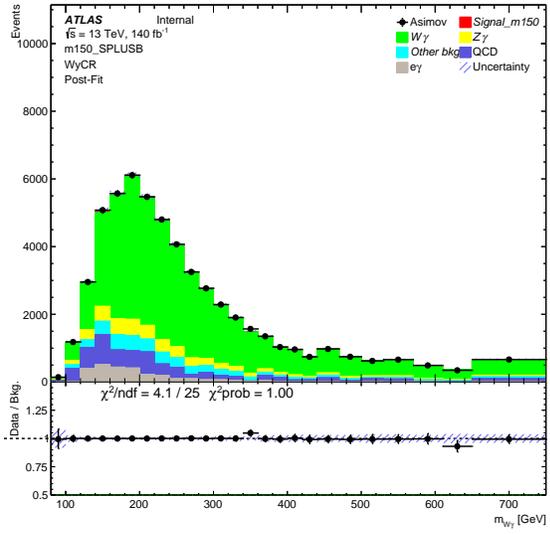


Figure 8.5: Correlation Matrix of all systematics after the pruning procedure.

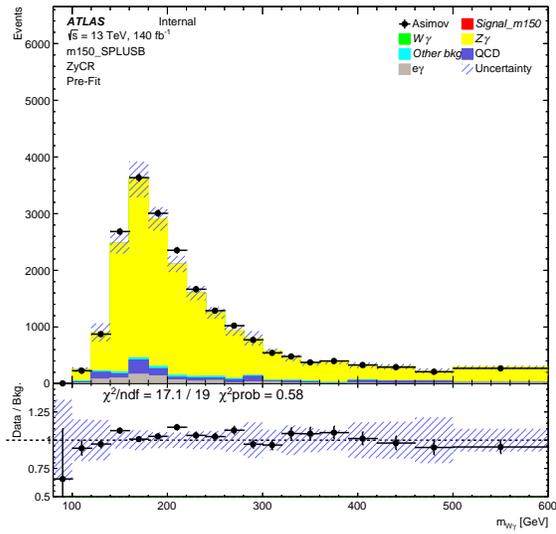


(a) $W\gamma$ control pre-Fit plot.

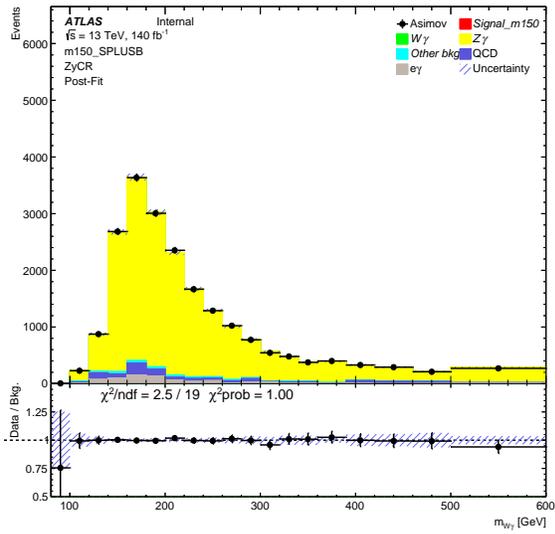


(b) $W\gamma$ control post-Fit plot.

Figure 8.6: Plots of the $W\gamma$ control region before and after the fit.

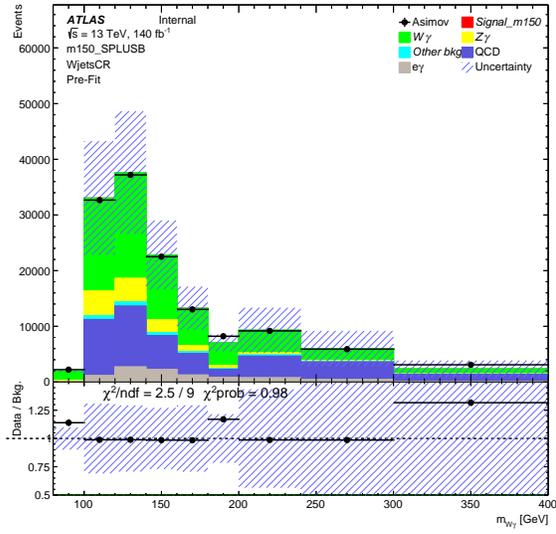


(a) $Z\gamma$ control pre-Fit plot.

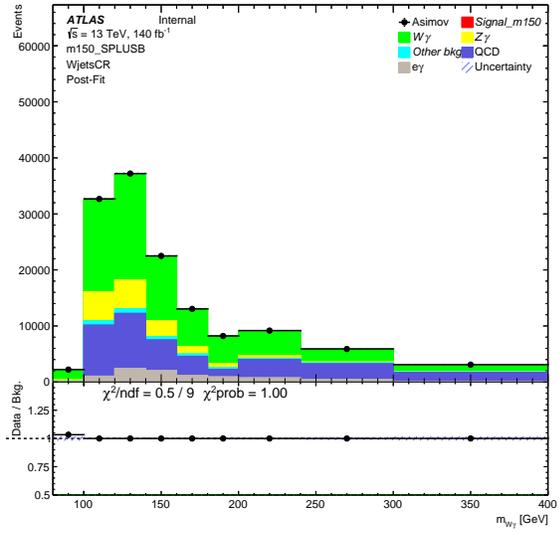


(b) $Z\gamma$ control post-Fit plot.

Figure 8.7: Plots of the $Z\gamma$ control region before and after the fit.

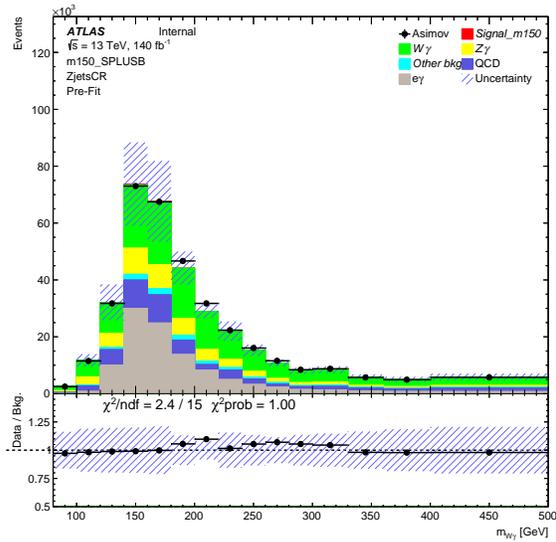


(a) W +jets control pre-Fit plot.

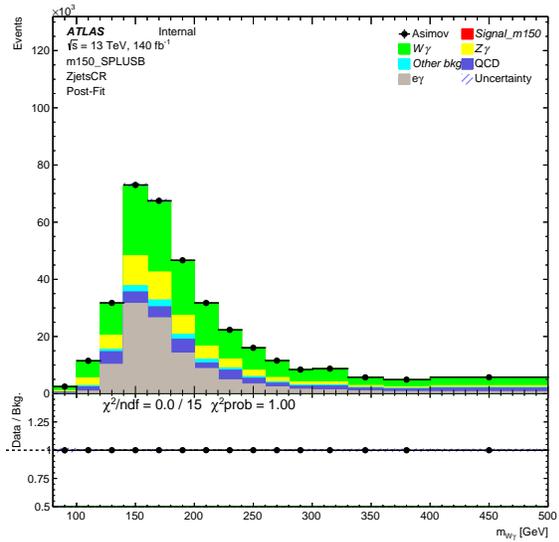


(b) W +jets control post-Fit plot.

Figure 8.8: Plots of the W +jets control region before and after the fit.

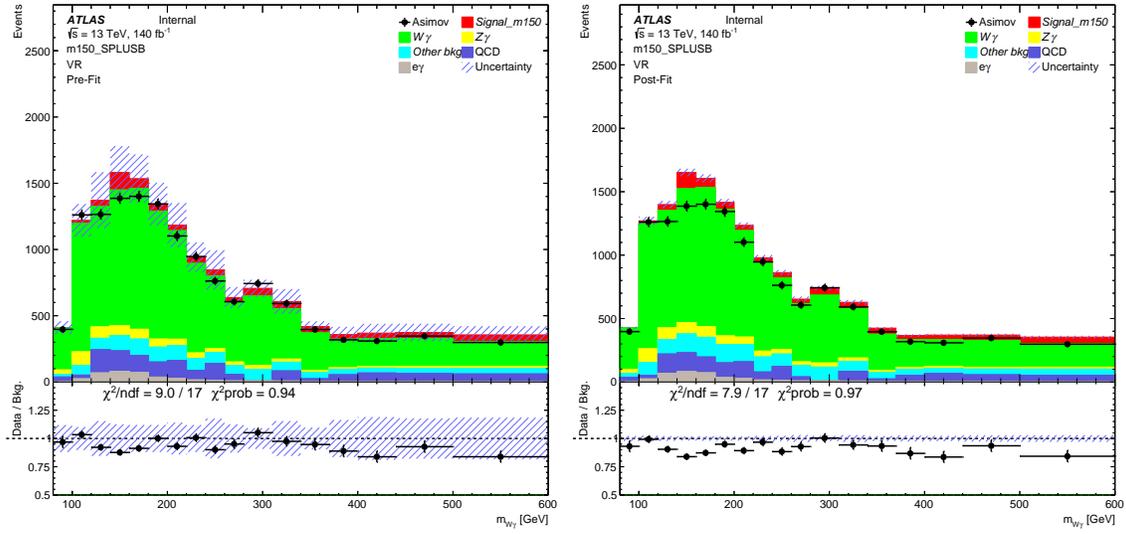


(a) Z +jets control pre-Fit plot.



(b) Z +jets control post-Fit plot.

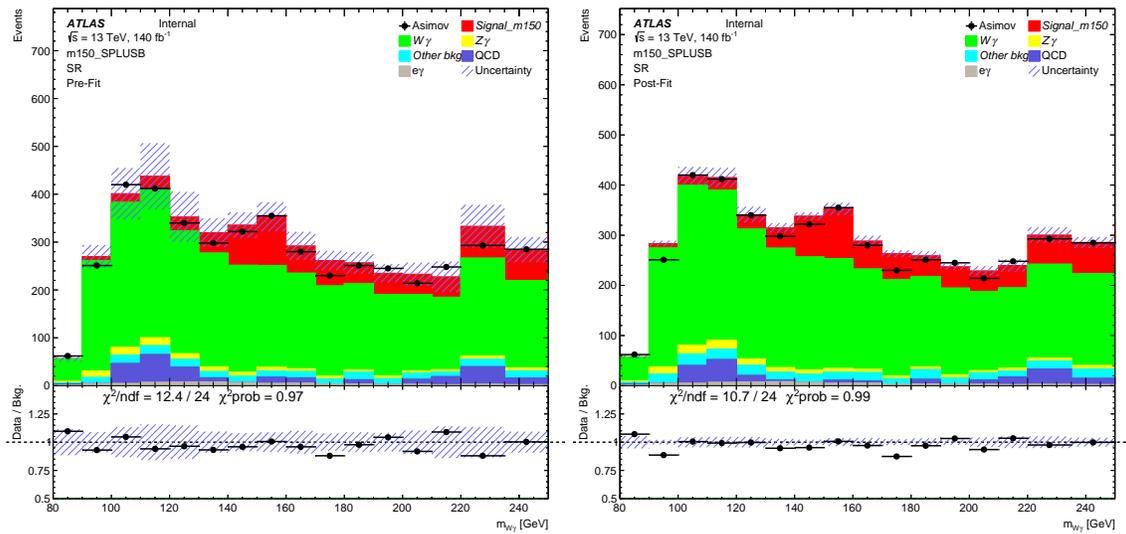
Figure 8.9: Plots of the Z +jets control region before and after the fit.



(a) Validation region pre-Fit plot.

(b) Validation region post-Fit plot.

Figure 8.10: Plots of the validation region before and after the fit.



(a) Signal region pre-Fit plot.

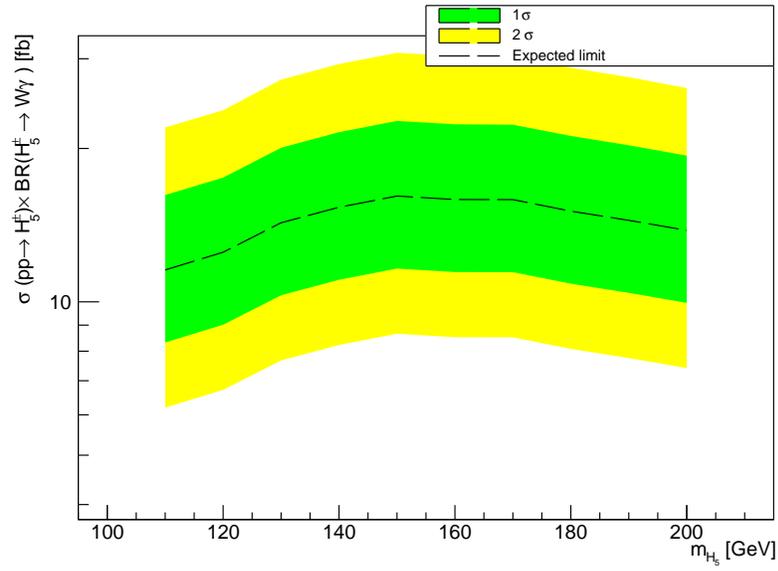
(b) Signal region post-Fit plot.

Figure 8.11: Plots of the signal region before and after the fit.

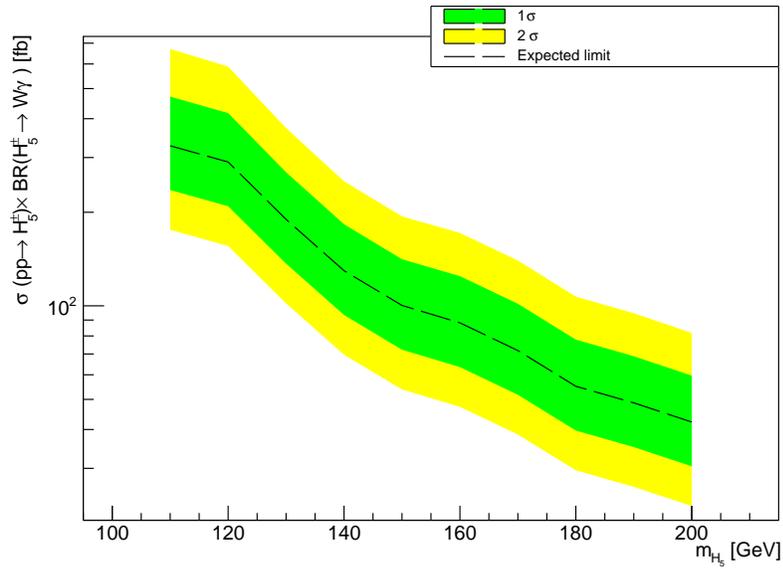
It is clear from Figure 8.1 that most nuisance parameters are dropped due to their negligible contributions. The pull of the remaining ones from pruning exhibits moderate pull, all within two standard deviations. The theory systematics of the $W\gamma$ and $Z\gamma$ process

have the largest pulls according to Figure 8.1. This is likely due to the observed deficit in the $W\gamma$ and $Z\gamma$ control region before the fit.

After the nuisance parameters are extrapolated from the control regions, the validation confirms good post-fit modeling of the background processes. At this point, the results shown in the post-fit distribution plot in Figure 8.11(b) could be interpreted. The expected limit is fitted using the profile likelihood method introduced in Section 8.1. 95% confidence level limits of both the $pp \rightarrow H_5^\pm H_5^0$ and $pp \rightarrow H_5^\pm H_5^\mp$ processes are shown in Figure 8.12. The cross section of the $pp \rightarrow H_5^\pm H_5^0$ process is constrained to have limits between 11.6 fb and 16.2 fb with a relatively flat curve shown in Figure 8.12(a). The flat curve is mostly due to the large tail of the reconstructed Higgs mass as shown in Figure 4.7(a), a phenomenon of both p_z' resolution and incorrect photon associations. However, the $pp \rightarrow H_5^\pm H_5^\mp$ process has a much more distinct peak with a lower tail since it only involves one final state photon, as shown in Figure 4.7(b). The lack of a tail in the reconstructed Higgs mass in the $pp \rightarrow H_5^\pm H_5^\mp$ sample essentially makes the fitting scheme more "peak-dependent". Since the cross-section of the signal sample is lower at higher masses, the 95% confidence level limit in Figure 8.12(b) decreases from 332.1 fb to 41.5 fb as mass increases. And since the $pp \rightarrow H_5^\pm H_5^\mp$ samples have lower cross sections compared to the $pp \rightarrow H_5^\pm H_5^0$ samples, the expected limit is much worse, especially at lower masses.



(a) Expected limit of the charged Higgs process with a singly charged Higgs and a neutral Higgs, $pp \rightarrow H_5^\pm H_5^0 \rightarrow W^\pm \gamma + \gamma\gamma$.



(b) Expected limit of the charged Higgs process with both singly and doubly charged Higgs, $pp \rightarrow H_5^+ H_5^- \rightarrow W^\pm \gamma + W^\mp \gamma$.

Figure 8.12: Expected limit of the signal charged Higgs process within the mass range between 110 GeV and 200 GeV.

The next step of the analysis will be to unblind the signal region after approval from the ATLAS collaboration to observe the real data. This should not change the extrapolation of the nuisance parameters since the real data is already used in the control regions. The values and pulls of these nuisance parameters will be identical to the results shown in this chapter.

However, the distribution in the signal region is expected to vary significantly. The signal process was injected into the Asimov data used in this analysis, meaning that the pseudo-data contains both the signal and backgrounds. This is clear from the invariant mass peak around 150 GeV in Figure 8.11. In the signal injection, the nominal cross section of the signal process is used to obtain the expected signal significance. Once the real data is used, it is likely that this peak around 150 GeV will be much less prominent. As a result, it is natural to speculate that the limit shown in Figure 8.12 will be much weaker with the unblinded real data.

Chapter 9

Conclusion

This thesis has presented the forecast results of the search for charged Higgs bosons using the GM model as the benchmark model. This analysis has yet to be approved to unblind at the time of writing and thus requires the use of Asimov datasets to derive the expected limits of the signal processes' cross-sections. Once unblinded, this analysis will be the first dedicated search for charged Higgs boson with masses below 200 GeV. The expected limits were calculated with Higgs boson mass points between 110 GeV and 200 GeV. The limits of the $H_5^\pm H_5^0$ signal process cross-section were calculated to be between 11.6 fb and 16.2 fb for all mass points, while those of the $H_5^\pm H_5^\mp$ process ranged from 332.1 fb to 31.5 fb.

The main discriminant used in the analysis is the invariant mass of the charged Higgs boson, which is reconstructed by estimating the four-momentum of the neutrino coming from the Higgs decay. MC simulations were used to model both the SM background and signal processes. For backgrounds involving misidentified photons, data-driven methods provided precise modeling. The electron faking photon process was modeled using corrections estimated by the tag and probe method. The jets faking photons contribution was estimated using the template fit method with templates derived from data and MC simulation.

There are still a few remaining parts of this analysis that are worth looking into in the future. First, the final state charged lepton could be misidentified, especially in the electron channel. Despite their small contribution in the signal region, it is still worthwhile to use data-driven techniques to model these processes, especially in fake-dominant control regions. These could be investigated using the tag and probe method similar to the one implemented in estimating electrons faking photons. The data-driven estimation of fake leptons is currently under investigation and will be adopted before the unblinding of the analysis. Another interesting extension of this analysis is to search for invariant mass peaks using `bumphunter` [99] over the entire reconstructed spectrum of charged Higgs mass. The `bumphunter` algorithm offers a model-independent search for mass peaks with corrections of statistical phenomena such as the look elsewhere effect. Finally, an alternative search strategy worth noting would be a completely model-dependent search. This analysis uses the GM model as a benchmark, while all cuts are designed to be model-agnostic. Alternatively, one could design cuts specifically optimized for a particular model. For example, a model-dependent search for a GM model charged Higgs will capitalize on the fact that three final state photon processes are rare in the standard model. Requiring these tighter cuts will likely yield much greater signal process sensitivity, providing more stringent constraints on the signal process cross-sections.

Bibliography

- [1] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, sep 2012. URL: <https://doi.org/10.1016%2Fj.physletb.2012.08.020>, doi:10.1016/j.physletb.2012.08.020.
- [2] Howard Georgi and Marie Machacek. Doubly charged Higgs bosons. *Nuclear Physics B*, 262(3):463–477, 1985. URL: <https://www.sciencedirect.com/science/article/pii/0550321385903256>, doi:[https://doi.org/10.1016/0550-3213\(85\)90325-6](https://doi.org/10.1016/0550-3213(85)90325-6).
- [3] ATLAS Collaboration. Search for a Charged Higgs Boson Produced in the Vector-Boson Fusion Mode with Decay $H^\pm \rightarrow W^\pm Z$ using pp Collisions at $\sqrt{s} = 8$ TeV with the ATLAS Experiment. *Phys. Rev. Lett.*, 114:231801, Jun 2015.
- [4] Pei-Hong Gu, He Zhang, and Shun Zhou. A Minimal Type II Seesaw Model. *Phys. Rev. D*, 74:076002, 2006. arXiv:hep-ph/0606302, doi:10.1103/PhysRevD.74.076002.
- [5] ATLAS Collaboration. Search for doubly and singly charged Higgs bosons decaying into vector bosons in multi-lepton final states with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 06:146, 2021. arXiv:2101.11961, doi:10.1007/JHEP06(2021)146.

- [6] Aleph, Delphi, L3, OPAL Collaborations, and the LEP working group for Higgs boson searches. Search for charged higgs bosons: Combined results using lep data, 2013. arXiv:1301.6065.
- [7] Svjetlana Fajfer, Jernej F. Kamenik, and Ivan Nisandzic. On the $B \rightarrow D^* \tau \bar{\nu}_\tau$ Sensitivity to New Physics. *Phys. Rev. D*, 85:094025, 2012. arXiv:1203.2654, doi:10.1103/PhysRevD.85.094025.
- [8] J. P. Lees et al. Measurement of an Excess of $\bar{B} \rightarrow D^{(*)} \tau^- \bar{\nu}_\tau$ Decays and Implications for Charged Higgs Bosons. *Phys. Rev. D*, 88(7):072012, 2013. arXiv:1303.0571, doi:10.1103/PhysRevD.88.072012.
- [9] A. Abulencia et al. Search for charged Higgs bosons from top quark decays in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ -TeV. *Phys. Rev. Lett.*, 96:042003, 2006. arXiv:hep-ex/0510065, doi:10.1103/PhysRevLett.96.042003.
- [10] V. M. Abazov et al. Search for charged Higgs bosons decaying to top and bottom quarks in $p\bar{p}$ collisions. *Phys. Rev. Lett.*, 102:191802, 2009. arXiv:0807.0859, doi:10.1103/PhysRevLett.102.191802.
- [11] Wikimedia Commons. Standard Model of Elementary Particles, 2019. URL: https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.
- [12] Georges Aad et al. Determination of the parton distribution functions of the proton using diverse ATLAS data from pp collisions at $\sqrt{s} = 7, 8$ and 13 TeV. *Eur. Phys. J. C*, 82(5):438, 2022. arXiv:2112.11266, doi:10.1140/epjc/s10052-022-10217-z.
- [13] Particle Data Group. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2020(8), 08 2020. 083C01. arXiv:https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673740/rpp2020-vol12-2015-2092_18.pdf, doi:10.1093/ptep/ptaa104.

- [14] Dirac Paul Adrien Maurice. The quantum theory of the electron. *Proc. R. Soc. Lond. A* 117:610–624, 1928. doi:<http://doi.org/10.1098/rspa.1928.0023>.
- [15] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>, doi:10.1103/PhysRevLett.13.508.
- [16] John Ellis, Mary K. Gaillard, and Dimitri V. Nanopoulos. A Historical Profile of the Higgs Boson. An Updated Historical Profile of the Higgs Boson. pages 255–274, 2016. 22 pages, 5 figures, update of arXiv:1201.6045, to be published in the volume. URL: <https://cds.cern.ch/record/2012465>, arXiv:1504.07217.
- [17] V. A. Miransky. *Dynamical symmetry breaking in quantum field theories*. 1994.
- [18] Cheng-Wei Chiang, Giovanna Cottin, and Otto Eberhardt. Global fits in the Georgi-Machacek model. *Phys. Rev. D*, 99:015001, Jan 2019.
- [19] T. Aaltonen et al. High-precision measurement of the W boson mass with the CDF II detector. *Science*, 376(6589):170–176, 2022. doi:10.1126/science.abk1781.
- [20] Heather E. Logan and Yongcheng Wu. Searching for the W decay of a charged Higgs boson. *Journal of High Energy Physics*, 2018(11), nov 2018. URL: [https://doi.org/10.1007/JHEP11\(2018\)121](https://doi.org/10.1007/JHEP11(2018)121), doi:10.1007/JHEP11(2018)121.
- [21] Katy Hartling, Kunal Kumar, and Heather E. Logan. The decoupling limit in the Georgi-Machacek model. *Physical Review D*, 90(1), jul 2014. URL: <https://doi.org/10.1103/PhysRevD.90.015007>, doi:10.1103/PhysRevD.90.015007.
- [22] Xiao Kang Du, Zhuang Li, Fei Wang, and Ying Kai Zhang. Explaining the CDF-II w -boson mass anomaly in the georgi-machacek extension models. *The European Physical Journal C*, 83(2), feb 2023. URL: <https://doi.org/10.1140/epjc/s10052-023-11297-1>, doi:10.1140/epjc/s10052-023-11297-1.

- [23] Rikard Enberg, Johan Rathsman, and Glenn Wouda. Higgs phenomenology in the stealth doublet model. *Physical Review D*, 91(9), may 2015. URL: <https://doi.org/10.1103/PhysRevD.91.095002>, doi:10.1103/PhysRevD.91.095002.
- [24] Céline Degrande, Katy Hartling, and Heather E. Logan. Scalar decays to $\gamma\gamma$, $Z\gamma$, and $W\gamma$ in the Georgi-Machacek model. *Phys. Rev. D*, 96:075013, Oct 2017. doi:10.1103/PhysRevD.96.075013.
- [25] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(08):S08003, aug 2008. URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08003>, doi:10.1088/1748-0221/3/08/S08003.
- [26] The CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004, aug 2008. URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08004>, doi:10.1088/1748-0221/3/08/S08004.
- [27] The ALICE Collaboration. The ALICE experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08002, aug 2008. URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08002>, doi:10.1088/1748-0221/3/08/S08002.
- [28] The LHCb Collaboration. The LHCb Detector at the LHC. *Journal of Instrumentation*, 3(08):S08005, aug 2008. URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08005>, doi:10.1088/1748-0221/3/08/S08005.
- [29] Renilde Vanden Broeck. THE CERN ACCELERATOR COMPLEX. Complexe des accélérateurs du CERN. 2019. URL: <https://cds.cern.ch/record/2693837>.
- [30] Eberhard Keil. The Large Hadron Collider LHC. 1996. revised version submitted on 2004-08-19 11:10:07. URL: <https://cds.cern.ch/record/316527>.

- [31] Lyndon Evans and Philip Bryant. LHC Machine. *Journal of Instrumentation*, 3(08):S08001, aug 2008. URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08001>, doi:10.1088/1748-0221/3/08/S08001.
- [32] G Avoni, et al. The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS. *JINST*, 13(07):P07017, 2018. URL: <http://cds.cern.ch/record/2633501>, doi:10.1088/1748-0221/13/07/P07017.
- [33] Marko Mikuz, V. Cindro, I. Dolenc, H. Kagan, G. Kramberger, Helmut Frais-Kölbl, Andrej Gorisek, E. Griesmayer, B. Macek, I. Mandic, M. Niegl, H. Pernegger, Dale Smith, Diane Tardif, W. Trischuk, Peter Weilhammer, and Marko Zavrtanik. The atlas beam conditions monitor. *Journal of Instrumentation*, 3:1914 – 1917, 02 2008. doi:10.1109/NSSMIC.2007.4436530.
- [34] S van der Meer. Calibration of the effective beam height in the ISR. Technical report, CERN, Geneva, 1968. URL: <https://cds.cern.ch/record/296752>.
- [35] ATLAS Collaboration. Luminosity determination in pp collisions at $\sqrt{s} = 13$ tev using the atlas detector at the lhc, 2022. arXiv:2212.09379.
- [36] ATLAS collaboration. Public ATLAS Luminosity Results for Run-2 of the LHC , 2023. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- [37] CMS collaboration. Pileup mitigation at CMS in 13 TeV data. *Journal of Instrumentation*, 15(09):P09018–P09018, sep 2020. URL: <https://doi.org/10.1088/1748-0221/15/09/p09018>, doi:10.1088/1748-0221/15/09/p09018.
- [38] J.T. Boyd. LHC Run-2 and Future Prospects. 2021. URL: <https://cds.cern.ch/record/2707815>, arXiv:2001.04370, doi:10.23730/CYRSP-2021-005.247.

- [39] ATLAS Collaboration. Studies of the performance of the atlas detector using cosmic-ray muons. *European Physical Journal C*, 71:1–36, 11 2010. doi:10.1140/epjc/s10052-011-1593-6.
- [40] *ATLAS inner detector: Technical Design Report, 1*. Technical design report. ATLAS. CERN, Geneva, 1997. URL: <http://cds.cern.ch/record/331063>.
- [41] S Haywood, L Rossi, R Nickerson, and A Romaniouk. *ATLAS inner detector: Technical Design Report, 2*. Technical design report. ATLAS. CERN, Geneva, 1997. URL: <http://cds.cern.ch/record/331064>.
- [42] ATLAS collaboration. Studies of the performance of the ATLAS detector using cosmic-ray muons. *The European Physical Journal C*, 71(3), mar 2011. URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-011-1593-6>, doi:10.1140/epjc/s10052-011-1593-6.
- [43] *ATLAS: a 25-year insider story of the LHC experiment*. Advanced series on directions in high energy physics. World Scientific, Singapore, 2019. URL: <https://cds.cern.ch/record/2636577>, doi:10.1142/11030.
- [44] Alessandro La Rosa. The atlas insertable b-layer: from construction to operation. 2016. URL: <https://arxiv.org/abs/1610.01994>, doi:10.48550/ARXIV.1610.01994.
- [45] ATLAS collaboration. Performance of the ATLAS transition radiation tracker in run 1 of the LHC: tracker properties. *Journal of Instrumentation*, 12(05):P05002–P05002, may 2017. URL: <https://doi.org/10.1088%2F1748-0221%2F12%2F05%2Fp05002>, doi:10.1088/1748-0221/12/05/p05002.
- [46] A Vogel. ATLAS Transition Radiation Tracker (TRT): Straw Tube Gaseous Detectors at High Rates. Technical report, CERN, Geneva, 2013. URL: <https://cds.cern.ch/record/1537991>.

- [47] V L Ginzburg. Transition radiation and transition scattering. *Physica Scripta*, 1982(T2A):182, jan 1982. URL: <https://dx.doi.org/10.1088/0031-8949/1982/T2A/024>, doi:10.1088/0031-8949/1982/T2A/024.
- [48] Sascha Mehlhase. ATLAS detector slice (and particle visualisations). 2021. URL: <https://cds.cern.ch/record/2770815>.
- [49] *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical design report. ATLAS. CERN, Geneva, 1996. URL: <https://cds.cern.ch/record/331061>, doi:10.17181/CERN.FWRW.FOOQ.
- [50] Deep Learning for Pion Identification and Energy Calibration with the ATLAS Detector. Technical report, CERN, Geneva, 2020. URL: <https://cds.cern.ch/record/2724632>.
- [51] *ATLAS tile calorimeter: Technical Design Report*. Technical design report. ATLAS. CERN, Geneva, 1996. URL: <http://cds.cern.ch/record/331062>, doi:10.17181/CERN.JRBJ.7028.
- [52] P. Adragna et al. Testbeam studies of production modules of the ATLAS tile calorimeter. *Nucl. Instrum. Meth. A*, 606:362–394, 2009. doi:10.1016/j.nima.2009.04.009.
- [53] *ATLAS muon spectrometer: Technical Design Report*. Technical design report. ATLAS. CERN, Geneva, 1997. URL: <https://cds.cern.ch/record/331068>.
- [54] S. Horvat, O. Kortner, H. Kroha, A. Manz, S. Mohr dieck, and V. Zhuravlov. Precision Drift Chambers for the Atlas Muon Spectrometer. *eConf*, C030626:FRAP13, 2003. URL: <https://cds.cern.ch/record/639745>.
- [55] T. Argyropoulos, K. A. Assamagan, B. H. Benedict, V. Chernyatin, E. Cheu, J. Deng, A. Gordeev, I. Gough Eschrich, V. Grachev, K. Johns, S. Junnarkar, A. Kandasamy, V. Kaushik, A. Khodinov, A. Lankford, X. Lei, R. Murillo Garcia, K. Nikolopoulos,

- P. O'Connor, C. L. Parnell-Lampen, W. Park, V. Polychronakos, R. Porter, M. Purohit, M. Schernau, B. K. Toggerson, and D. Tompkins. Cathode strip chambers in atlas : Installation, commissioning and in situ performance. In *2008 IEEE Nuclear Science Symposium Conference Record*, pages 2819–2824, 2008. doi:10.1109/NSSMIC.2008.4774958.
- [56] R. Santonico and R. Cardarelli. Development of resistive plate counters. *Nuclear Instruments and Methods in Physics Research*, 187(2):377–380, 1981. URL: <https://www.sciencedirect.com/science/article/pii/0029554X81903633>, doi:[https://doi.org/10.1016/0029-554X\(81\)90363-3](https://doi.org/10.1016/0029-554X(81)90363-3).
- [57] Giordano Cattani and (On behalf of the RPC group). The resistive plate chambers of the atlas experiment: performance studies. *Journal of Physics: Conference Series*, 280(1):012001, feb 2011. URL: <https://dx.doi.org/10.1088/1742-6596/280/1/012001>, doi:10.1088/1742-6596/280/1/012001.
- [58] Koichi Nagai. Thin gap chambers in atlas. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 384(1):219–221, 1996. BEAUTY '96. URL: <https://www.sciencedirect.com/science/article/pii/S0168900296010650>, doi:[https://doi.org/10.1016/S0168-9002\(96\)01065-0](https://doi.org/10.1016/S0168-9002(96)01065-0).
- [59] The ATLAS collaboration. Operation of the atlas trigger system in run 2. *Journal of Instrumentation*, 15(10):P10004, oct 2020. URL: <https://dx.doi.org/10.1088/1748-0221/15/10/P10004>, doi:10.1088/1748-0221/15/10/P10004.
- [60] ATLAS collaboration. Performance of the ATLAS track reconstruction algorithms in dense environments in LHC run 2. *The European Physical Journal C*, 77(10), oct 2017. URL: <https://doi.org/10.1140/epjc/s10052-017-5225-7>, doi:10.1140/epjc/s10052-017-5225-7.

- [61] Azriel Rosenfeld and John Pfaltz. Pfaltz, j.l.: Sequential operations in digital picture processing. *journal of the acm* 13(4), 471-494. *J. ACM*, 13:471–494, 10 1966. doi: 10.1145/321356.321357.
- [62] R. Frühwirth. Application of kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 262(2):444–450, 1987. URL: <https://www.sciencedirect.com/science/article/pii/0168900287908874>, doi:[https://doi.org/10.1016/0168-9002\(87\)90887-4](https://doi.org/10.1016/0168-9002(87)90887-4).
- [63] Vertex Reconstruction Performance of the ATLAS Detector at " $\sqrt{s} = 13$ TeV". Technical report, CERN, Geneva, 2015. URL: <https://cds.cern.ch/record/2037717>.
- [64] ATLAS collaboration. Reconstruction of primary vertices at the ATLAS experiment in run 1 proton–proton collisions at the LHC. *The European Physical Journal C*, 77(5), may 2017. URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-017-4887-5>, doi:10.1140/epjc/s10052-017-4887-5.
- [65] W Lampl, S Laplace, D Lelas, P Loch, H Ma, S Menke, S Rajagopalan, D Rousseau, S Snyder, and G Unal. Calorimeter Clustering Algorithms: Description and Performance. Technical report, CERN, Geneva, 2008. URL: <https://cds.cern.ch/record/1099735>.
- [66] ATLAS collaboration. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *Eur. Phys. J.*, C77:490, 2017. arXiv:1603.02934, doi:10.1140/epjc/s10052-017-5004-5.
- [67] Heather Abreu et al. Performance of the electronic readout of the atlas liquid argon calorimeters. *Journal of Instrumentation*, 5, 09 2010. doi:10.1088/1748-0221/5/09/P09003.

- [68] ATLAS collaboration. Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13$ TeV. *The European Physical Journal C*, 79(8), aug 2019. URL: <https://doi.org/10.1140/epjc/s10052-019-7140-6>, doi:10.1140/epjc/s10052-019-7140-6.
- [69] Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung. Technical report, CERN, Geneva, 2012. URL: <https://cds.cern.ch/record/1449796>.
- [70] ATLAS collaboration. Measurement of the photon identification efficiencies with the ATLAS detector using LHC run 2 data collected in 2015 and 2016. *The European Physical Journal C*, 79(3), mar 2019. URL: <https://doi.org/10.1140/epjc/s10052-019-6650-6>, doi:10.1140/epjc/s10052-019-6650-6.
- [71] ATLAS collaboration. Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV. *The European Physical Journal C*, 76(5), may 2016. URL: <https://doi.org/10.1140/epjc/s10052-016-4120-y>.
- [72] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti- k_t jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, apr 2008. URL: <https://doi.org/10.1088/1126-6708/2008/04/063>, doi:10.1088/1126-6708/2008/04/063.
- [73] ATLAS collaboration. Jet reconstruction and performance using particle flow with the ATLAS detector. *The European Physical Journal C*, 77(7), jul 2017. URL: <https://doi.org/10.1140/epjc/s10052-017-5031-2>.
- [74] Howtocleanjets2012. <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/HowToCleanJets2012>. Accessed: 2023-03-29.
- [75] Tagging and suppression of pileup jets with the ATLAS detector. Technical report, CERN, Geneva, 2014. All figures including auxiliary figures are available

- at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2014-018>. URL: <https://cds.cern.ch/record/1700870>.
- [76] ATLAS collaboration. ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV. *The European Physical Journal C*, 79(11), nov 2019. URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-019-7450-8>, doi:10.1140/epjc/s10052-019-7450-8.
- [77] Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run. Technical report, CERN, Geneva, 2017. URL: <https://cds.cern.ch/record/2273281>.
- [78] Secondary vertex finding for jet flavour identification with the ATLAS detector. Technical report, CERN, Geneva, 2017. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2017-011>. URL: <https://cds.cern.ch/record/2270366>.
- [79] D Adams, C Anastopoulos, A Andreazza, M Aoki, L Asquith, M Begel, F Bernlochner, U Blumenschein, A Bocci, S Cheatham, W Davey, P-A Delsart, P-O Deviveiros, A Dewhurst, D Duschinger, F Filthaut, P Francavilla, F Garberson, S Head, A Henrichs, A Hoecker, M Kagan, B Kersevan, TJ Khoo, B Lenzi, D Lopez Mateos, B Malaescu, Z Marshall, T Martin, C Meyer, A Morley, W Murray, M zur Nedden, R Nicolaidou, S Pagan Griso, G Pasztor, P Petroff, C Pizio, R Polifka, X Poveda, R Reece, F Ruehr, F Salvatore, R Sandstroem, T Scanlon, D Scheirich, S Schramm, A Schwartzman, K Suruliz, M Sutton, E Thompson, M Tripiana, A Tuna, S Viel, M Vincter, I Vivarelli, M Wielers, A Wildauer, and Z Zinonos. Recommendations of the Physics Objects and Analysis Harmonisation Study Groups 2014. Technical report, CERN, Geneva, 2014. URL: <https://cds.cern.ch/record/1743654>.
- [80] ATLAS collaboration. Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV. 78(11), nov

2018. URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-018-6288-9>, doi:10.1140/epjc/s10052-018-6288-9.
- [81] ATLAS collaboration. Introduction to MC generation, 2023. URL: https://atlassoftwaredocs.web.cern.ch/AnalysisSWTutorial/mc_intro/.
- [82] W Lukas. Fast Simulation for ATLAS: Atlfast-II and ISF. Technical report, CERN, Geneva, 2012. URL: <https://cds.cern.ch/record/1458503>, doi:10.1088/1742-6596/396/2/022031.
- [83] Marilena Bandieramonte, Joseph Boudreau, and Riccardo Maria Bianchi. Full-SimLight: ATLAS standalone Geant4 simulation. Technical report, CERN, Geneva, 2020. URL: <https://cds.cern.ch/record/2712863>, doi:10.1051/epjconf/202024502029.
- [84] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Computer Physics Communications*, 191:159–177, jun 2015. URL: <https://doi.org/10.1016%2Fj.cpc.2015.01.024>, doi:10.1016/j.cpc.2015.01.024.
- [85] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5: going beyond. *Journal of High Energy Physics*, 2011(6), jun 2011. URL: <https://doi.org/10.1007%2Fjhep06%282011%29128>, doi:10.1007/jhep06(2011)128.
- [86] W Buttinger. The ATLAS Level-1 Trigger System. Technical report, CERN, Geneva, 2012. URL: <http://cds.cern.ch/record/1456546>, doi:10.1088/1742-6596/396/1/012010.
- [87] Johannes Elmsheuser et al. Evolution of the ATLAS analysis model for Run-3 and prospects for HL-LHC. *EPJ Web Conf.*, 245:06014, 2020. doi:10.1051/epjconf/202024506014.

- [88] Data-Quality Requirements and Event Cleaning for Jets and Missing Transverse Energy Reconstruction with the ATLAS Detector in Proton-Proton Collisions at a Center-of-Mass Energy of $\sqrt{s} = 7$ TeV. Technical report, CERN, Geneva, 2010. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2010-038>. URL: <https://cds.cern.ch/record/1277678>.
- [89] A.Bukin. CERN ROOT: RooBukinPdf Class Reference, 2023. URL: <https://root.cern.ch/doc/master/classRooBukinPdf.html>.
- [90] The ATLAS collaboration. Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data. *Journal of Instrumentation*, 14(12):P12006–P12006, dec 2019. URL: <https://doi.org/10.1088/1748-0221/14/12/p12006>, doi:10.1088/1748-0221/14/12/p12006.
- [91] Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV. Technical report, CERN, Geneva, 2015. URL: <https://cds.cern.ch/record/2037613>.
- [92] ATLAS collaboration. Jet energy resolution in proton-proton collisions at $\sqrt{s} = 7$ TeV recorded in 2010 with the ATLAS detector. *The European Physical Journal C*, 73(3), mar 2013. URL: <https://doi.org/10.1140/epjc/s10052-013-2306-0>, doi:10.1140/epjc/s10052-013-2306-0.
- [93] Eigenvector recomposition: a new method to correlate flavor-tagging systematic uncertainties across analyses. Technical report, CERN, Geneva, 2022. URL: <https://cds.cern.ch/record/2809651>.
- [94] Andy Buckley, James Ferrando, Stephen Lloyd, Karl Nordström, Ben Page, Martin Rüfenacht, Marek Schönherr, and Graeme Watt. LHAPDF6: parton density access in the LHC precision era. *The European Physical Journal C*, 75(3), mar

2015. URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-015-3318-8>, doi:10.1140/epjc/s10052-015-3318-8.
- [95] Michele Pinamonti and Loic Valery. TRexFitter. <https://trexfitter-docs.web.cern.ch/trexfitter-docs/>. Accessed: 2023-04-28.
- [96] Kyle Cranmer, George Lewis, Lorenzo Moneta, Akira Shibata, and Wouter Verkerke. HistFactory: A tool for creating statistical models for use with RooFit and RooStats. Technical report, New York U., New York, 2012. URL: <https://cds.cern.ch/record/1456844>.
- [97] G Cowan, K Cranmer, E Gross, and et al. Asymptotic formulae for likelihood-based tests of new physics. Feb 2011. URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-011-1554-0>, doi:10.1140/epjc/s10052-011-1554-0.
- [98] A L Read. Presentation of search results: the cls technique. *Journal of Physics G: Nuclear and Particle Physics*, 28(10):2693, sep 2002. URL: <https://dx.doi.org/10.1088/0954-3899/28/10/313>, doi:10.1088/0954-3899/28/10/313.
- [99] Georgios Choudalakis. On hypothesis testing, trials factor, hypertests and the bump-hunter, 2011. arXiv:1101.0390.

Glossary

ATLAS (A Toroidal LHC Apparatus) One of the general purpose detector at the LHC.

BSM (Beyond Standard Model) Physics theories beyond the Standard Model.

CR (Control Region) A region in the phase space where nuisance parameters are determined.

CSC (Cathode Strip Chambers) Multi-wire proportional chambers implemented in the forward region of the muon spectrometer.

DAOD (Derived Analysis Object Data) A data storage format used for all data in this analysis.

Electromagnetic Calorimeter (ECal) The calorimeter system designed to measure energies of electrons and photons.

FCAL (Forward Calorimeter) The calorimeter system that covers the forward region of the detector.

GM (Georgi-Machacek) The Georgi-Machacek model of charged Higgs bosons.

GRL (Good Run List) A list of good quality data recorded in a given year.

HCal (Hadronic Calorimeter) The calorimeter system designed to measure energies of hadrons.

HEC (Hadronic End-cap Calorimeter) Hadronic calorimeter system in the end-cap region.

HLT (High-Level triggers) High-level triggers systems of the ATLAS detector.

ID (Inner Detector) The innermost component of the ATLAS detector that is capable of measuring trajectories of charged particles.

IP (Interaction Point) Designate points in the LHC where proton beams are brought to collide.

JER (Fractional Jet p_T Resolution) Systematics uncertainties related to the energy resolution of jets.

JES (Jet Energy Scale) Systematics uncertainties related to the energy scale of jets.

JVT (Jet Vertex Tagger) A algorithm that grades the likelihood of each jet being a pile-up jet in a single score.

L1 (Level-1) The first level of the ATLAS trigger system.

LAr (Liquid Argon) The active material of the ATLAS ECal system.

LHC (Large Hadron Collider) The collider facility from which datasets in this analysis were collected.

MC (Monte Carlo) Computer simulation based on random sampling.

MDTs (Monitored Drift Tubes) Drift tube chambers in the muon spectrometer.

MET (Missing transverse momentum) The projection of the three-dimensional missing momentum onto the transverse plane. In mathematical expression, it is denoted by E_T^{miss} .

MIP (Minimum Ionizing Particles) A particle in the particular momentum range such that it only deposits minimum energy as it traverses blocking material.

MS (Muon Spectrometer) A dedicated ATLAS detector component for the precise measurements of muons.

NNLO (Next-to-Next-to-Leading Order) The third order theoretical correction of a physical process .

PDF (Parton Distribution Functions) Probability distribution of individual's parton momentum as a fraction of the proton's momentum.

PFlow (Particle Flow) The algorithm that re-calibrates cluster energies by removing overlapping tracks.

PID (Particle Identification) Working points of particle reconstruction quality.

PMT (photo-multiplier tube) Phototubes that measure charged particles by multiplying the induced currents.

QCD (Quantum Chromodynamics) The part of the quantum field theory that describes the strong interaction.

RF (Radio Frequency) The oscillation frequency between 20 kHz and 300 GHz.

RPC (Resistive Plate Chambers) Muon trigger systems in the barrel region of the muon spectrometer.

SCT (Semiconductor Tracker) Silicon sensors that provide position measurements of the tracks.

SM (Standard Model) The standard model of particle physics.

SR (Signal Region) A region in the phase space where the signal processes are measured.

Systematics (Systematic uncertainties) Uncertainties related to systematic variations of nuisance parameters.

TDAQ (ATLAS trigger and data acquisition) The overall trigger system of the ATLAS experiment.

TGC (Thin Gap Chambers) Multi-wire chambers in the forward region of the muon spectrometer that provides triggering and position measurements.

TRT (Transition Radiation Tracker) Drift tubes implemented to identify particles based on the transition radiation effect.

VEV (Vacuum Expectation Value) The expectation value of a field, typically the Higgs field, in vacuum.

VR (Validation Region) A region in the phase space where the estimated background is compared with the recorded data to check the validity of modeling.

Appendix A

Detector Resolution

The following table from Reference [25] summarizes the energy resolutions and η coverage of various detector components of the ATLAS detector introduced in Section 3.2.

Detector Component	Resolution	Measurement η Coverage	Trigger η Coverage
Tracking	$\frac{\sigma_{p_T}}{p_T} = 0.05\%p_T \oplus 1\%$	$[-2.5, +2.5]$	
EM Calorimeter	$\frac{\sigma_E}{E} = 10\%/\sqrt{E} \oplus 0.7\%$	$[-3.2, +3.2]$	$[-2.5, +2.5]$
Hadronic Calorimeter -Barrel and endcap	$\frac{\sigma_E}{E} = 50\%/\sqrt{E} \oplus 3\%$	$[-3.2, +3.2]$	$[-3.2, +3.2]$
-Foward	$\frac{\sigma_E}{E} = 100\%/\sqrt{E} \oplus 10\%$	$3.1 < \eta < 4.9$	$3.1 < \eta < 4.9$
Muon Spectrometer	$\frac{\sigma_{p_T}}{p_T} = 10\%$ at $p_T = 1$ TeV	$[-2.7, +2.7]$	$[-2.4, +2.4]$

Table A.1: Energy resolutions and η coverages of various components of the ATLAS detector.

Appendix B

Good Run List

The GRLs for different run-2 years, which were introduced in Section 5.1.1, are the following:

- `data15.13TeV.periodAllYear.DetStatus-v89-pro21-02.Unknown.PHYS.StandardGRLAll.Good.25ns`
- `data16.13TeV.periodAllYear.DetStatus-v89-pro21-01.DQDefects-00-02-04.PHYS.StandardGRLAll.Good.25ns`
- `data17.13TeV.periodAllYear.DetStatus-v99-pro22-01.Unknown.PHYS.StandardGRLAll.Good.25ns.Triggerno17e33prim`
- `data18.13TeV.periodAllYear.DetStatus-v102-pro22-04.Unknown.PHYS.StandardGRLAll.Good.25ns.Triggerno17e33prim`

Appendix C

Neutrino Reconstruction

In Appendix C, the two neutrino reconstruction algorithms introduced in Section 4.8 are compared. In this analysis, two main algorithms are reconstructed to deal with complex numbers in the reconstruction of the neutrino momentum. In Figure C.1 and Figure C.2, the reconstructed $W\gamma$ invariant masses using the neutrino estimations from the two methods are compared.

The truth mass of the Higgs bosons in both plots is at exactly 150 GeV, as indicated by the red vertical lines. The estimates from the fitting algorithm are shown in black dots. The estimates obtained from taking the real parts of the complex solutions are plotted with solid bars. The plots were made by using events where the outputs of the two algorithms disagree by more than 10 GeV. As shown in these plots, the fit method obtains much better resolutions of the Higgs invariant mass.

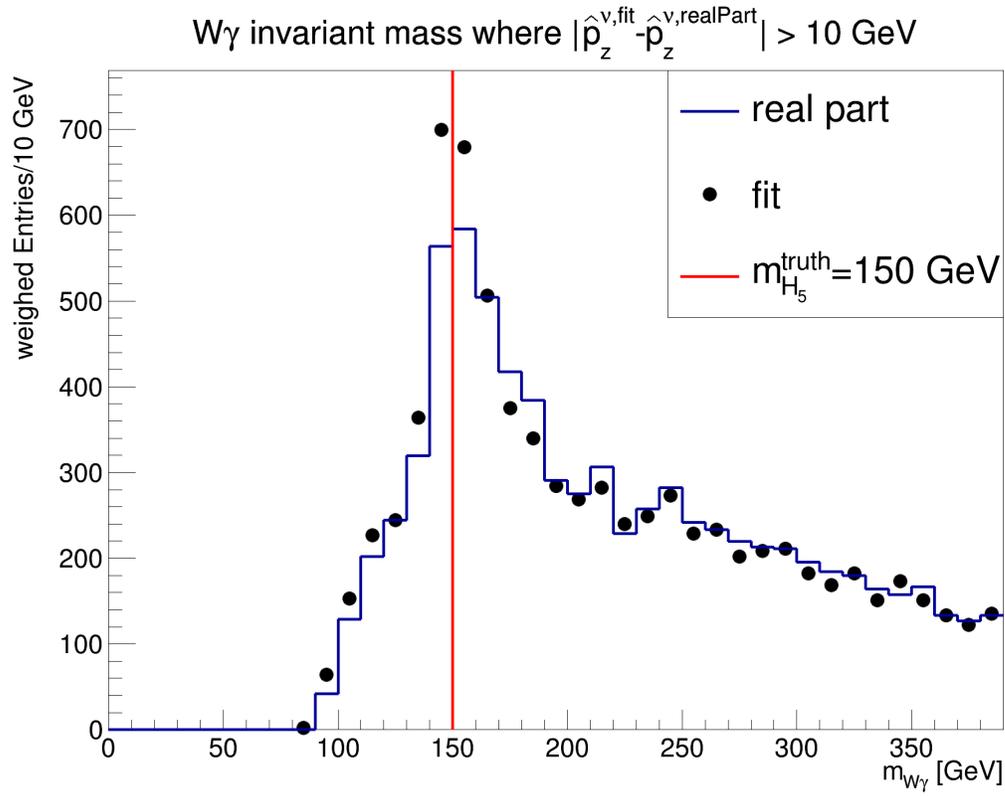


Figure C.1: Comparison of charged Higgs boson mass estimation between different algorithms in the $H_5^\pm H_5^0$ samples with $m_{H_5} = 150$ GeV.

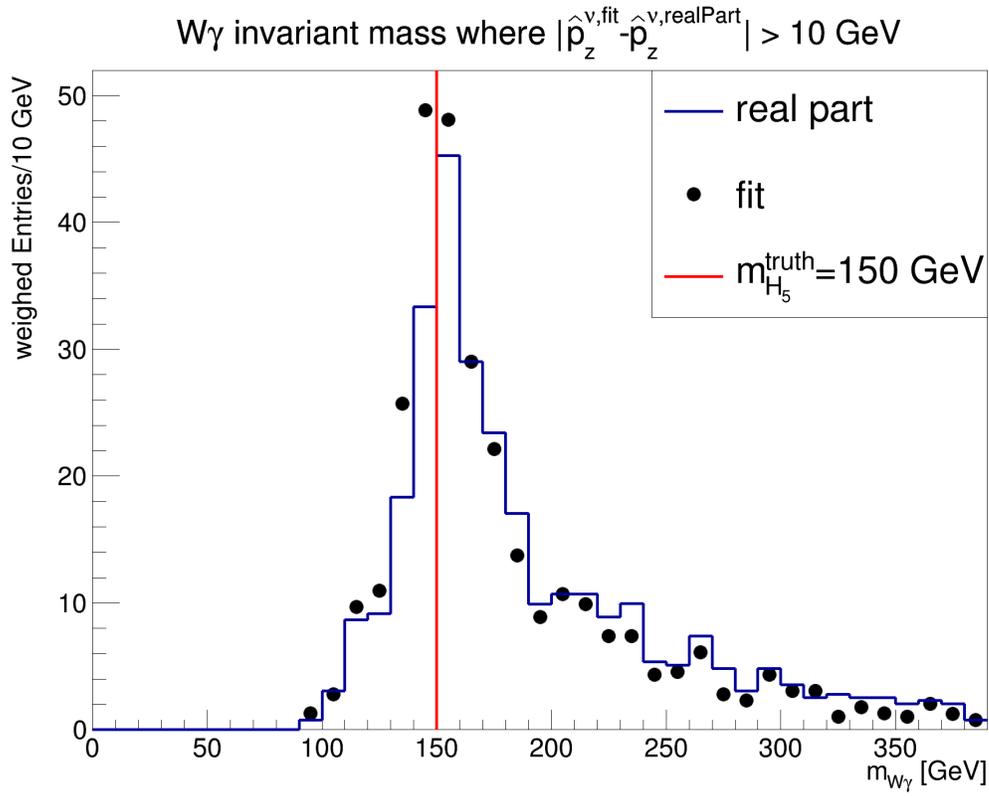


Figure C.2: Comparison of charged Higgs boson mass estimation between different algorithms in the $H_5^\pm H_5^\mp$ samples with $m_{H_5} = 150$ GeV.

Appendix D

Event Selection

In the following table, the initial cutflows of various processes are shown. These correspond to the event selections introduced in Section 5.2.1.

Sample	Cuts					
	GRL	GoodCalo	PriVtx	NoBadMuon	JetClean	Trigger
$W\gamma$	27,659,059	27,659,059	27,659,055	27,658,693	27,534,790	5,446,340
$Z\gamma$	19,151,785	19,151,785	19,151,740	19,151,314	19,041,076	7,107,553
$Z + \text{jets}$	424,449,117	424,449,117	424,448,744	424,437,542	422,106,167	227,162,136
$W + \text{jets}$	700,569,321	700,569,321	700,567,758	700,558,284	697,219,729	242,166,961
$t\bar{t}\gamma$	790,909	790,909	790,909	790,902	787,085	288,826
$t\bar{t}$	52,313,451	52,313,451	52,313,450	52,313,135	52,056,789	13,626,158
Top	12,622,833	12,622,833	12,622,831	12,622,750	12,622,740	3,484,327
$VV\gamma$	13,155	13,155	13,155	13,155	13,069	8,268
EWKV γ	641,282	641,282	641,279	641,270	637,570	191,169
2γ	25,197,108	25,197,108	25,196,960	25,196,960	25,101,643	79,228
$H_5^0 H_5^\pm$	30,137	30,137	30,37	30,133	30,034	16,364
$H_5^+ H_5^-$	505	505	505	505	503	282

Sample	Cuts				
	$n_\ell = 1$	$n_\gamma \geq 1$	$n_{\text{jet}} < 2$	No b-jets	$ m_{\ell,\gamma} - m_Z < 5 \text{ GeV}$
$W\gamma$	5,442,638	891,208	502,909	490,416	466,181
$Z\gamma$	5,337,012	814,648	562,516	550,615	458,388
$Z + \text{jets}$	144,689,777	2,733,245	1,043,671	944,107	337,516
$W + \text{jets}$	242,149,767	91,742	41,080	39,158	37,645
$t\bar{t}\gamma$	277,815	72,355	5,549	2,696	2,483
$t\bar{t}$	13,122,650	74,830	8,615	3,730	3,459
Top	5,586	1,271	701	687	625
$VV\gamma$	3,483,930	5,839	2,228	1,283	1,152
EWKV γ	155,166	19,872	7,434	7,208	6,650
2γ	79,091	29,466	25,269	24,946	22,336
$H_5^0 H_5^\pm$	16,107	15,491	9,085	8,990	8,763
$H_5^+ H_5^-$	270	173	62	60	58

Table D.1: Initial cutflows for various processes before constructing each signal, validation, and control region. This is for those regions with a final state of one lepton and at least one photon. The two tables are separated due to aesthetic reasons. The first cut in the lower table is the one placed after the last cut in the upper table.

Appendix E

Systematic uncertainties

Appendix E gives the details of the systematic uncertainties introduced in Chapter 7. Appendix E.1 covers the detector-related systematics discussed in Section 7.1. In Appendix E.2, the contribution of systematics uncertainties associated with the data-driven methods, which were introduced in Section 7.3, are summarized. The remaining parts of this Appendix give the relative effects of all systematics uncertainties.

E.1 Detector-related Systematics

E_T^{miss} uncertainties

There are three E_T^{miss} variations considered for the uncertainty treatment [80]:

- MET_SoftTrk_Scale: Scaling the magnitude of the soft term in the E_T^{miss} .
- MET_SoftTrk_ResoPara: Gaussian smearing of the soft term magnitude in the direction parallel to the vector sum of all hard objects' \vec{p}_T .

- MET_SoftTrk_ResoPerp: Gaussian smearing of the soft term magnitude in the direction perpendicular to the vector sum of all hard objects' \vec{p}_T .

Leptons

- Electron:
 - Scale factor of electron identification: leptonSF_EL_SF_ID
 - Scale factors of electron reconstruction: leptonSF_EL_SF_Reco
 - Scale factors of electron triggers: leptonSF_EL_SF_Trigger
 - Scale factors of electron isolation: leptonSF_EL_SF_Isol
 - Electron energy resolution: EG_RESOLUTION_ALL
 - Electron energy scale: EG_SCALE_[ALL,AF2]
- Muon:
 - Inner detector smearing: ATLAS_MUON_ID
 - Muon spectrometer track smearing: ATLAS_MUON_MS
 - Charge-independent scale momentum: ATLAS_MUON_SCALE
 - Charge-dependent scale momentum: MUON_SAGITTA_[RHO,RESBIAS]
 - Scale factors of muon identification: leptonSF_MU_SF_ID_[SYST,STAT].
 - Scale factors of track-to-vertex association (TTVA): leptonSF_MU_SF_TTVA_[SYST,STAT].
 - Scale factors of muon triggers: leptonSF_MU_SF_Trigger_[SYST,STAT]
 - Scale factors of muon isolation: leptonSF_MU_SF_Isol_[SYST,STAT]
- Tau:
 - Tau Energy Scale:
 - * TAUS_TRUEHADTAU_SME_TES_PHYSICSLIST

- * TAUS_TRUEHADTAU_SME_TES_INSITUFIT
- * TAUS_TRUEHADTAU_SME_TES_INSITUEXP
- * TAUS_TRUEHADTAU_SME_TES_DETECTOR
- * TAUS_TRUEHADTAU_SME_TES_MODEL_CLOSURE

Photons

- Photon identification: photonSF_ID
- Photon isolation: photonSF_effIso
- Photon energy resolution: EG_RESOLUTION_ALL
- Photon energy scale: EG_SCALE_[ALL,AF2]

Jets

- JES:
 - Effective nuisance parameters:
 - * Detector-related: EffectiveNP_Detector[1,2]
 - * Modelling-related: EffectiveNP_Modelling[1,2,3,4]
 - * Mixing both aspects: EffectiveNP_Mixed[1,2,3]
 - * Statistics-related: EffectiveNP_Modelling[1,2,3,4,5,6]
 - η inter-calibration:
 - * Modelling-related: EtaIntercalibration_Modelling
 - * Statistics-related: EtaIntercalibration_Total_Stat
 - * Non-closure: EtaIntercalibration_Total_[2018data,highE,negEta,posEta]
 - Flavour-related :

- * Flavor_[Composition,Response]
- * BJES_Response
- Punch-through effect: PunchThrough_[AFII,MC16]
- Pileup-correction:
 - * Pileup_offsetMu
 - * Pileup_offsetNPV
 - * Pileup_PtTerm
 - * Pileup_RhoTopology
- Single High p_T particle response: SingleParticle_HighPt
- JVT scale factor: jvt
- JER:
 - In situ measurement of the JER for MC versus data: JER_DataVsMC_[AFII,MC16]
 - Effective nuisance parameters:EffectiveNP_[1,2,3,4,5,6,7_restTerm]

Flavour tagging

- Eigenvector variations:
 - B-jets: bTagSF_DL1r_77_eigenvars_B_[0,1,2,3,4,5,6,7,8]
 - C-jets: bTagSF_DL1r_77_eigenvars_C_[0,1,2,3]
 - Light-flavour jets: bTagSF_DL1r_77_eigenvars_Light_[0,1,2,3]
- High- p_T extrapolation:
 - bTagSF_DL1r_77_extrapolation
 - bTagSF_DL1r_77_extrapolation_from_charm

E.2 Data-driven methods

	$W\gamma\text{CR}$	$Z\gamma\text{CR}$	VR	ZjetsCR	WjetsCR	SR
Window size:	4.58%	4.44%	5.04%	4.52%	4.41%	5.88%
Fitting function:	2.60%	1.94%	2.66%	2.78%	3.58%	2.37%
Fitting error:	4.52%	5.13%	5.03%	4.58%	4.04%	6.70%

Table E.1: Systematic uncertainties of electrons faking photons in each region. The uncertainties are expressed in percentages of the nominal values.

	$W\gamma\text{CR}$	$Z\gamma\text{CR}$	VR	ZjetsCR	WjetsCR	SR
Binning variation:	9.50%	29.4%	17.1%	4.37%	1.49%	45.7%
Loose PID Selection:	4.41%	16.7%	47.5%	4.64%	1.94%	27.1%

Table E.2: Systematic uncertainties of jets faking photons in each region. The uncertainties are expressed in percentages of the nominal values.

E.3 Signal region

	$W\gamma$	$Z\gamma$	Other bkg	QCD	$e\gamma$	Signal
JET_Flavor_Composition	-0.0111 / 0.0111	0.0208 / -0.0208	6.89e-10 / -1.4e-09	0 / 0	0 / 0	-0.013 / 0.013
JET_Flavor_Response	-0.0158 / 0.0158	-0.0439 / 0.0439	-0.0149 / 0.0149	0 / 0	0 / 0	1.45e-08 / -1.79e-09
JET_Pileup_RhoTopology	-0.0136 / 0.0136	2.43e-11 / -2.78e-09	-0.0129 / 0.0129	0 / 0	0 / 0	5.1e-09 / 5.96e-09
theory_systematics_Wy	0.107 / -0.107	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
EG_SCALE_ALL	1.71e-09 / -8.54e-10	-0.0162 / 0.0162	3.53e-09 / 6.77e-12	0 / 0	0 / 0	3.01e-09 / 8.11e-09
JET_Modelling	0 / 0	0.0212 / -0.0212	0.0114 / -0.0114	0 / 0	0 / 0	1.01e-08 / -6.39e-09
JET_NonClosure_2018data	0 / 0	0.0102 / -0.0102	8.47e-09 / 1.25e-11	0 / 0	0 / 0	9.76e-09 / 2.8e-09
JET_Pileup_OffsetMu	0 / 0	0.0113 / -0.0113	5.65e-09 / -9.18e-09	0 / 0	0 / 0	-3.8e-09 / -1.51e-09
JET_Pileup_PtTerm	0 / 0	0.0102 / -0.0102	-1.15e-11 / 7.08e-09	0 / 0	0 / 0	0 / 0
MET_SoftTrk_Scale	0 / 0	-0.0375 / 0.0375	5.68e-09 / 3.52e-09	0 / 0	0 / 0	-1.29e-09 / 4.09e-09
JET_EffectiveNP_Mixed1	0 / 0	0.0137 / -0.0137	-1.69e-11 / -4.23e-09	0 / 0	0 / 0	1.44e-09 / -2.73e-09
JET_EffectiveNP_Modelling1	0 / 0	0.0182 / -0.0182	6.35e-09 / 1.2e-08	0 / 0	0 / 0	-6.1e-09 / -8.4e-09
MET_SoftTrk_ResoPara	0 / 0	-0.0335 / 0.0335	-0.131 / 0.131	0 / 0	0 / 0	-6.46e-10 / -3.23e-09
MET_SoftTrk_ResoPerp	0 / 0	-0.0361 / 0.0361	-0.127 / 0.127	0 / 0	0 / 0	5.53e-09 / 3.8e-09
theory_systematics_Zy	0 / 0	0.157 / -0.157	0 / 0	0 / 0	0 / 0	0 / 0
JET_Pileup_OffsetNPV	0 / 0	2.76e-09 / -6.9e-09	4.25e-09 / -7.08e-09	0 / 0	0 / 0	5.31e-09 / 8.68e-09
EG_RESOLUTION_ALL	0 / 0	4.13e-09 / 1.39e-09	7.07e-09 / 2.83e-09	0 / 0	0 / 0	7.61e-09 / -1.44e-09
TAU_TES_DETECTOR	0 / 0	1.38e-08 / 2.12e-11	0 / 0	0 / 0	0 / 0	0 / 0
TAU_TES_PHYSICSLIST	0 / 0	-1.39e-09 / 9.67e-09	0 / 0	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Mixed2	0 / 0	8.33e-09 / 1.33e-09	4.96e-09 / -3.54e-09	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Mixed3	0 / 0	9.66e-09 / 1.38e-09	0 / 0	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Modelling3	0 / 0	-5.49e-09 / 4.11e-09	4.89e-12 / 3.53e-09	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Statistical4	0 / 0	5.5e-09 / 1.4e-09	0 / 0	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Statistical6	0 / 0	8.28e-09 / 6.66e-12	0 / 0	0 / 0	0 / 0	0 / 0
JET_BJES_Response	0 / 0	0 / 0	2.05e-11 / -2.85e-09	0 / 0	0 / 0	0 / 0
JET_NonClosure_negEta	0 / 0	0 / 0	-1.13e-08 / -4.24e-09	0 / 0	0 / 0	0 / 0
JET_TotalStat	0 / 0	0 / 0	7.05e-10 / -6.36e-09	0 / 0	0 / 0	3.66e-09 / -3.95e-09
MUON_ID	0 / 0	0 / 0	7.78e-09 / 2.83e-09	0 / 0	0 / 0	7.25e-09 / 7.53e-09
JET_EffectiveNP_Modelling2	0 / 0	0 / 0	-2.13e-09 / -1.41e-09	0 / 0	0 / 0	0 / 0
binsVaried	0 / 0	0 / 0	0 / 0	0.375 / -0.355	0 / 0	0 / 0
loosePrime2	0 / 0	0 / 0	0 / 0	0.125 / -0.117	0 / 0	0 / 0
fitError	0 / 0	0 / 0	0 / 0	0 / 0	0.067 / -0.067	0 / 0
windowError	0 / 0	0 / 0	0 / 0	0 / 0	0.0588 / -0.0588	0 / 0
functionError	0 / 0	0 / 0	0 / 0	0 / 0	0.0237 / -0.0237	0 / 0
theory_systematics_Signal_m150	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0.0591 / -0.0591
EG_SCALE_AF2	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-7.25e-09 / -2.15e-09
MUON_MS	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-4.09e-09 / 1.16e-08
JET_RelativeNonClosure_AFII	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	7.18e-10 / -5.24e-09

Table E.3: Relative effect of each nuisance parameter on the yields in the signal region before the fit. The signal sample used for this table has a mass of 150 GeV.

E.4 Validation region

	$W\gamma$	$Z\gamma$	Other bkg	QCD	$e\gamma$	Signal
JET_Flavor_Response	-0.0146 / 0.0146	-0.0362 / 0.0362	2.46e-13 / -1.39e-08	0 / 0	0 / 0	-2.83e-10 / -1.22e-08
theory_systematics_Wy	0.113 / -0.113	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
JET_Flavor_Composition	3.13e-09 / 4.69e-09	0.0506 / -0.0506	4.64e-09 / 4.63e-09	0 / 0	0 / 0	-0.0139 / 0.0139
JET_Pileup_RhoTopology	-4.69e-09 / 4.69e-09	2.71e-09 / -9.51e-09	-0.0207 / 0.0207	0 / 0	0 / 0	-0.0112 / 0.0112
EG_SCALE_ALL	4.69e-09 / -6.26e-09	-1.36e-09 / 2.72e-09	3.09e-09 / 3.09e-09	0 / 0	0 / 0	1.7e-09 / 1.64e-08
JET_Modelling	0 / 0	0.0275 / -0.0275	-7.72e-09 / 6.18e-09	0 / 0	0 / 0	5.67e-10 / 9.35e-09
JET_Pileup_OffsetMu	0 / 0	0.0343 / -0.0343	-9.27e-09 / -9.26e-09	0 / 0	0 / 0	6.23e-09 / -8.5e-10
JET_Pileup_OffsetNPV	0 / 0	0.0134 / -0.0134	-1.08e-08 / 6.17e-09	0 / 0	0 / 0	7.37e-09 / 9.35e-09
MET_SoftTrk_Scale	0 / 0	-0.0528 / 0.0528	7.73e-09 / 6.17e-09	0 / 0	0 / 0	-1.98e-09 / -3.97e-09
JET_EffectiveNP_Modelling1	0 / 0	0.0165 / -0.0165	-7.72e-09 / -9.27e-09	0 / 0	0 / 0	8.22e-09 / -1.64e-08
MET_SoftTrk_ResoPara	0 / 0	-0.0452 / 0.0452	-0.163 / 0.163	0 / 0	0 / 0	-6.52e-09 / -2.55e-09
MET_SoftTrk_ResoPerp	0 / 0	-0.0481 / 0.0481	-0.164 / 0.164	0 / 0	0 / 0	-4.82e-09 / -9.35e-09
theory_systematics_Zy	0 / 0	0.139 / -0.139	0 / 0	0 / 0	0 / 0	0 / 0
JET_Pileup_PtTerm	0 / 0	6.79e-09 / 1.5e-08	0 / 0	0 / 0	0 / 0	-8.5e-09 / 1.1e-08
JET_BJES_Response	0 / 0	0 / 0	-3.08e-09 / 1.54e-09	0 / 0	0 / 0	0 / 0
EG_RESOLUTION_ALL	0 / 0	0 / 0	1.54e-09 / -4.63e-09	0 / 0	0 / 0	2.27e-09 / 1.05e-08
JET_EffectiveNP_Mixed1	0 / 0	0 / 0	-1.55e-09 / 3.09e-09	0 / 0	0 / 0	0 / 7.93e-09
binsVaried	0 / 0	0 / 0	0 / 0	0.16 / -0.16	0 / 0	0 / 0
loosePrime2	0 / 0	0 / 0	0 / 0	0.0604 / -0.0604	0 / 0	0 / 0
fitError	0 / 0	0 / 0	0 / 0	0 / 0	0.0503 / -0.0503	0 / 0
windowError	0 / 0	0 / 0	0 / 0	0 / 0	0.0504 / -0.0504	0 / 0
functionError	0 / 0	0 / 0	0 / 0	0 / 0	0.0266 / -0.0266	0 / 0
theory_systematics_Signal_m150	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0.0739 / -0.0739
JET_NonClosure_2018data	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	1.42e-08 / -6.23e-09
EG_SCALE_AF2	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-2.83e-10 / 4.25e-09
MUON_ID	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	2.27e-09 / 1.64e-08
MUON_MS	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-8.22e-09 / 1.05e-08
JET_RelativeNonClosure_AFII	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	2.83e-10 / -1.5e-08
JET_EffectiveNP_Mixed2	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	9.63e-09 / -1.36e-08
JET_EffectiveNP_Statistical3	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-6.23e-09 / 8.5e-09

Table E.4: Relative effect of each systematic on the yields in the validation region before the fit. The signal sample used for this table has a mass of 150 GeV.

E.5 W +jets control region

	$W\gamma$	$Z\gamma$	Other bkg	QCD	$e\gamma$	Signal
MET_SoftTrk_Scale	-0.0133 / 0.0133	-0.0674 / 0.0674	-0.0114 / 0.0114	0 / 0	0 / 0	5.46e-09 / 1.13e-08
MET_SoftTrk_ResoPara	-0.0119 / 0.0119	-0.0617 / 0.0617	-0.19 / 0.19	0 / 0	0 / 0	-1.06e-08 / -4e-09
MET_SoftTrk_ResoPerp	-0.0103 / 0.0103	-0.0609 / 0.0609	-0.191 / 0.191	0 / 0	0 / 0	-1.42e-08 / 4.37e-09
theory_systematics_Wy	0.0568 / -0.0568	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
JET_Flavor_Composition	8.21e-09 / -7.3e-09	0.0743 / -0.0743	-7.64e-09 / -7.64e-09	0 / 0	0 / 0	-0.0207 / 0.0207
JET_Flavor_Response	-1.83e-09 / 9.13e-09	-0.0289 / 0.0289	0 / 0	0 / 0	0 / 0	0.012 / -0.012
EG_SCALE_ALL	-2.1e-08 / 3.65e-09	0 / 0	0 / 0	0 / 0	0 / 0	-4.73e-09 / 1.46e-09
JET_EffectiveNP_Modelling1	-9.13e-09 / 2.74e-09	0.0166 / -0.0166	0 / 0	0 / 0	0 / 0	-4.37e-09 / 4e-09
JET_Modelling	0 / 0	0.0395 / -0.0395	0 / 0	0 / 0	0 / 0	1.06e-08 / 5.46e-09
JET_Pileup_OffsetMu	0 / 0	0.0488 / -0.0488	0 / 0	0 / 0	0 / 0	-1.6e-08 / 1.09e-09
JET_Pileup_OffsetNPV	0 / 0	0.0216 / -0.0216	0 / 0	0 / 0	0 / 0	1.6e-08 / -3.64e-10
JET_Pileup_RhoTopology	0 / 0	0.0322 / -0.0322	-0.027 / 0.027	0 / 0	0 / 0	-0.0164 / 0.0164
theory_systematics_Zy	0 / 0	0.0492 / -0.0492	0 / 0	0 / 0	0 / 0	0 / 0
JET_BJES_Response	0 / 0	0 / 0	-0.0131 / 0.0131	0 / 0	0 / 0	-1.75e-08 / -1.49e-08
binsVaried	0 / 0	0 / 0	0 / 0	0.0595 / -0.0595	0 / 0	0 / 0
loosePrime2	0 / 0	0 / 0	0 / 0	0.0218 / -0.0218	0 / 0	0 / 0
fitError	0 / 0	0 / 0	0 / 0	0 / 0	0.0404 / -0.0404	0 / 0
windowError	0 / 0	0 / 0	0 / 0	0 / 0	0.0441 / -0.0441	0 / 0
functionError	0 / 0	0 / 0	0 / 0	0 / 0	0.0358 / -0.0358	0 / 0
theory_systematics_Signal_m150	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0.11 / -0.11
JET_NonClosure_posEta	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-1.46e-08 / -2e-08
JET_TotalStat	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-3.64e-10 / -6.55e-09
JET_Pileup_PtTerm	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	9.46e-09 / 1.6e-08
EG_RESOLUTION_ALL	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	2.91e-09 / -5.46e-09
EG_SCALE_AF2	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	1.35e-08 / -1.42e-08
MUON_ID	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	9.46e-09 / -8.73e-09
MUON_MS	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	5.09e-09 / -1.16e-08
MUON_SAGITTA_RESBIAS	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	7.64e-09 / 2.18e-09
MUON_SCALE	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-1.42e-08 / 1.42e-08
JET_RelativeNonClosure_AFII	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-8.73e-09 / -1.27e-08
JET_EffectiveNP_Detector1	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	4.73e-09 / 5.46e-09
JET_EffectiveNP_Mixed1	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-1.78e-08 / 4.73e-09
JET_EffectiveNP_Mixed3	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-1.71e-08 / -1.2e-08
JET_EffectiveNP_Modelling2	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-2.55e-09 / 4.37e-09
JET_EffectiveNP_Modelling3	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-1.56e-08 / 2.18e-09
JET_EffectiveNP_Statistical2	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	6.55e-09 / -6.19e-09

Table E.5: Relative effect of each systematic on the yields in the W +jets control region before the fit. The signal sample used for this table has a mass of 150 GeV.

E.6 $W\gamma$ control region

	$W\gamma$	$Z\gamma$	Other bkg	QCD	$e\gamma$	Signal
JET_Flavor_Composition	-0.0108 / 0.0108	-9.6e-10 / 6.25e-09	-0.0243 / 0.0243	0 / 0	0 / 0	-0.0255 / 0.0255
EG_SCALE_ALL	0.015 / -0.015	0.0143 / -0.0143	-3.25e-09 / 3.79e-09	0 / 0	0 / 0	2.74e-09 / 7.98e-09
MET_SoftTrk_Scale	-0.0129 / 0.0129	-0.0434 / 0.0434	1.09e-09 / -1.63e-09	0 / 0	0 / 0	3.51e-09 / 8.27e-09
MET_SoftTrk_ResoPara	-0.0113 / 0.0113	-0.0414 / 0.0414	-0.168 / 0.168	0 / 0	0 / 0	-1.9e-09 / 5.89e-09
MET_SoftTrk_ResoPerp	-0.0108 / 0.0108	-0.0407 / 0.0407	-0.168 / 0.168	0 / 0	0 / 0	-5.67e-09 / -3.29e-10
theory_systematics_Wy	0.0608 / -0.0608	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
JET_Modelling	1.89e-09 / -1.26e-09	7.69e-09 / 3.36e-09	-0.0161 / 0.0161	0 / 0	0 / 0	3.36e-09 / 1.27e-09
JET_Flavor_Response	-2.31e-09 / 9.88e-09	2.4e-09 / 1.75e-12	0.0127 / -0.0127	0 / 0	0 / 0	0.0158 / -0.0158
JET_Pileup_OffsetNPV	-2.31e-09 / 3.15e-09	-1.92e-09 / 3.36e-09	-0.0118 / 0.0118	0 / 0	0 / 0	1.59e-09 / -5.75e-10
JET_Pileup_RhoTopology	1.47e-09 / 5.47e-09	-3.84e-09 / 9.13e-09	-0.0383 / 0.0383	0 / 0	0 / 0	-0.0182 / 0.0182
JET_EffectiveNP_Modelling1	-4.42e-09 / 4.84e-09	3.85e-09 / 4.8e-10	-0.0164 / 0.0164	0 / 0	0 / 0	-1.69e-08 / -1.03e-08
JET_Pileup_OffsetMu	0 / 0	0.0106 / -0.0106	-1.38e-08 / -2.72e-10	0 / 0	0 / 0	4.66e-10 / -1.2e-08
theory_systematics_Zy	0 / 0	0.0717 / -0.0717	0 / 0	0 / 0	0 / 0	0 / 0
JET_NonClosure_2018data	0 / 0	-9.13e-09 / 5.29e-09	-1.89e-09 / -1.62e-09	0 / 0	0 / 0	-2.79e-09 / 1.08e-08
JET_Pileup_PtTerm	0 / 0	-1.92e-09 / 4.22e-15	7.31e-09 / 1.43e-08	0 / 0	0 / 0	-2.47e-10 / 3.49e-09
EG_RESOLUTION_ALL	0 / 0	1.06e-08 / 7.21e-09	-8.12e-10 / -1.35e-09	0 / 0	0 / 0	6.49e-09 / -1.54e-11
MUON_ID	0 / 0	-8.17e-09 / -4.32e-09	0 / 0	0 / 0	0 / 0	-5.07e-09 / 1.03e-08
MUON_SAGITTA_RESBIAS	0 / 0	-2.4e-09 / 4.8e-10	0 / 0	0 / 0	0 / 0	-2.95e-09 / -4.07e-09
JET_BJES_Response	0 / 0	0 / 0	-0.0159 / 0.0159	0 / 0	0 / 0	0 / 0
JET_TotalStat	0 / 0	0 / 0	1.06e-08 / 1.08e-09	0 / 0	0 / 0	7.62e-09 / -6.4e-09
JET_EffectiveNP_Mixed2	0 / 0	0 / 0	7.58e-09 / 5.96e-09	0 / 0	0 / 0	-3.21e-09 / -5.9e-09
JET_EffectiveNP_Mixed3	0 / 0	0 / 0	9.2e-09 / -8.12e-10	0 / 0	0 / 0	0 / 0
binsVaried	0 / 0	0 / 0	0 / 0	0.135 / -0.135	0 / 0	0 / 0
loosePrime2	0 / 0	0 / 0	0 / 0	0.152 / -0.152	0 / 0	0 / 0
fitError	0 / 0	0 / 0	0 / 0	0 / 0	0.0452 / -0.0452	0 / 0
windowError	0 / 0	0 / 0	0 / 0	0 / 0	0.0458 / -0.0458	0 / 0
functionError	0 / 0	0 / 0	0 / 0	0 / 0	0.0261 / -0.0261	0 / 0
theory_systematics_Signal_m150	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0.104 / -0.104
JET_NonClosure_posEta	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	7.44e-09 / -5.69e-09
EG_SCALE_AF2	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-3.95e-09 / 7.71e-09
MUON_MS	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-3.69e-09 / -9.37e-09
MUON_SCALE	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	1.31e-08 / -5.22e-09
TAU_TES_MODEL_CLOSURE	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	3.79e-09 / -2.04e-09
TAU_TES_PHYSICSLIST	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	2.22e-09 / -6.6e-09
JET_RelativeNonClosure_AFII	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	8.38e-09 / -5.48e-10
JET_EffectiveNP_Detector1	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	7.66e-09 / 8.34e-09
JET_EffectiveNP_Mixed1	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-5.3e-09 / 7.26e-09
JET_EffectiveNP_Modelling2	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	7.37e-09 / 1.02e-08
JET_EffectiveNP_Modelling3	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	3.99e-09 / 4.82e-09
JET_EffectiveNP_Modelling4	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	8.85e-09 / 2.56e-09
JET_EffectiveNP_Statistical2	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	3.75e-09 / -4.34e-09
JET_EffectiveNP_Statistical4	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	3.84e-10 / 8.4e-09
JET_EffectiveNP_Statistical5	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	1.88e-09 / 1.13e-08
JET_EffectiveNP_Statistical6	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-8.45e-09 / 3.25e-09

Table E.6: Relative effect of each systematic on the yields in the $W\gamma$ control region. The signal sample used for this table has a mass of 150 GeV.

E.7 Z +jets control region

	$W\gamma$	$Z\gamma$	Other bkg	QCD	$e\gamma$	Signal
JET_Flavor_Composition	-0.0225 / 0.0225	-0.0271 / 0.0271	-0.0234 / 0.0234	0 / 0	0 / 0	-0.0222 / 0.0222
JET_Flavor_Response	0.0123 / -0.0123	0.0139 / -0.0139	0.0124 / -0.0124	0 / 0	0 / 0	0.0135 / -0.0135
JET_Pileup_RhoTopology	-0.0163 / 0.0163	-0.0204 / 0.0204	-0.022 / 0.022	0 / 0	0 / 0	-0.0156 / 0.0156
theory_systematics_Wy	0.0468 / -0.0468	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
theory_systematics_Zy	0 / 0	0.0364 / -0.0364	0 / 0	0 / 0	0 / 0	0 / 0
EG_SCALE_ALL	0 / 0	-1.2e-08 / 3.99e-09	8.27e-09 / -3.18e-09	0 / 0	0 / 0	3.65e-09 / 2.84e-09
MET_SoftTrk_ResoPara	0 / 0	0 / 0	-0.0403 / 0.0403	0 / 0	0 / 0	-1.62e-09 / 1.99e-09
MET_SoftTrk_ResoPerp	0 / 0	0 / 0	-0.0404 / 0.0404	0 / 0	0 / 0	7.7e-09 / -1.78e-08
JET_Modelling	0 / 0	0 / 0	4.45e-09 / -3.82e-09	0 / 0	0 / 0	-5.27e-09 / -8.51e-09
JET_TotalStat	0 / 0	0 / 0	1.59e-08 / 7.64e-09	0 / 0	0 / 0	0 / 0
JET_Pileup_OffsetNPV	0 / 0	0 / 0	-5.09e-09 / -6.36e-09	0 / 0	0 / 0	-1.18e-08 / 2.03e-09
JET_Pileup_PTerm	0 / 0	0 / 0	8.91e-09 / 1.59e-08	0 / 0	0 / 0	1.62e-08 / -5.27e-09
EG_RESOLUTION_ALL	0 / 0	0 / 0	9.54e-09 / 3.18e-09	0 / 0	0 / 0	4.87e-09 / -6.49e-09
MET_SoftTrk_Scale	0 / 0	0 / 0	6.36e-10 / 1.15e-08	0 / 0	0 / 0	-3.24e-09 / 5.68e-09
JET_EffectiveNP_Detector1	0 / 0	0 / 0	-1.27e-09 / 7.64e-09	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Mixed1	0 / 0	0 / 0	6.36e-10 / 5.73e-09	0 / 0	0 / 0	-8.11e-10 / -2.43e-09
JET_EffectiveNP_Mixed2	0 / 0	0 / 0	-1.27e-09 / 2.55e-09	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Modelling1	0 / 0	0 / 0	-1.08e-08 / -5.09e-09	0 / 0	0 / 0	-1.38e-08 / 2.07e-08
JET_EffectiveNP_Modelling2	0 / 0	0 / 0	-1.91e-09 / 4.24e-14	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Modelling3	0 / 0	0 / 0	-5.73e-09 / 7e-09	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Modelling4	0 / 0	0 / 0	-3.18e-09 / 5.09e-09	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Statistical2	0 / 0	0 / 0	1.27e-09 / 1.91e-09	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Statistical3	0 / 0	0 / 0	1.27e-08 / 8.27e-09	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Statistical4	0 / 0	0 / 0	-9.54e-09 / 8.27e-09	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Statistical6	0 / 0	0 / 0	7.64e-09 / 1.27e-08	0 / 0	0 / 0	0 / 0
binsVaried	0 / 0	0 / 0	0 / 0	0.548 / -0.548	0 / 0	0 / 0
loosePrime2	0 / 0	0 / 0	0 / 0	0.233 / -0.233	0 / 0	0 / 0
fitError	0 / 0	0 / 0	0 / 0	0 / 0	0.0458 / -0.0458	0 / 0
windowError	0 / 0	0 / 0	0 / 0	0 / 0	0.0452 / -0.0452	0 / 0
functionError	0 / 0	0 / 0	0 / 0	0 / 0	0.0278 / -0.0278	0 / 0
theory_systematics_Signal_m150	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0.0652 / -0.0652
JET_NonClosure_2018data	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-2.03e-09 / 1.14e-08
JET_Pileup_OffsetMu	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	1.42e-08 / 5.27e-09
EG_SCALE_AF2	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	5.68e-09 / 2.03e-09
MUON_ID	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-8.11e-10 / -1.22e-09
MUON_SAGITTA_RESBIAS	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-4.05e-10 / 5.27e-09
JET_RelativeNonClosure_AFII	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-1.22e-08 / -7.3e-09

Table E.7: Relative effect of each systematic on the yields in the Z +jets control region before the fit. The signal sample used for this table has a mass of 150 GeV.

E.8 $Z\gamma$ control region

	$W\gamma$	$Z\gamma$	Other bkg	QCD	$e\gamma$	Signal _{m 150}
EG_RESOLUTION_ALL	0.0517 / -0.0517	8.45e-10 / 8.62e-09	-6.21e-09 / -5.88e-09	0 / 0	0 / 0	-6.43e-09 / 6.62e-09
EG_SCALE_ALL	-0.029 / 0.0301	-4.09e-09 / 4.99e-09	9.31e-09 / -1.26e-08	0 / 0	0 / 0	-1.89e-09 / -7.37e-09
MUON_MS	0.0372 / -0.0372	-2.84e-09 / -6.54e-09	6.77e-10 / -7.4e-09	0 / 0	0 / 0	-7.56e-09 / -5.29e-09
MUON_SAGITTA_RESBIAS	0.036 / -0.036	0 / 0	0 / 0	0 / 0	0 / 0	6.43e-09 / -1.13e-09
JET_Modelling	-2.74e-09 / 1.01e-09	-4.85e-09 / 9.38e-09	-1.51e-09 / -9.24e-09	0 / 0	0 / 0	-3.78e-09 / -9.83e-09
JET_TotalStat	-9.27e-09 / 6.95e-09	0 / 0	-7.89e-09 / 2.69e-09	0 / 0	0 / 0	-1.55e-08 / 3.4e-09
JET_Flavor_Composition	-1.71e-08 / 2.59e-09	-3.3e-09 / 1.04e-09	1.68e-09 / -3.37e-10	0 / 0	0 / 0	6.43e-09 / -3.4e-09
JET_Flavor_Response	-8.15e-09 / 1.83e-08	-7.8e-09 / 5.62e-09	3.19e-09 / 3.86e-09	0 / 0	0 / 0	-4.54e-09 / 0
JET_Pileup_OffsetMu	-1.06e-08 / 2.59e-08	-8.1e-09 / 1.91e-10	-7.89e-09 / -6.13e-09	0 / 0	0 / 0	9.45e-09 / -1.89e-09
JET_Pileup_OffsetNPV	1.32e-08 / -7.03e-09	4.91e-10 / 1.48e-08	2.55e-10 / -5.21e-09	0 / 0	0 / 0	-7.94e-09 / 1.13e-09
JET_Pileup_PtTerm	-1.64e-09 / 1.64e-09	-1.36e-09 / 1.36e-09	0 / 0	0 / 0	0 / 0	4.54e-09 / -1.17e-08
JET_Pileup_RhoTopology	-3.93e-09 / 6.54e-09	4.91e-09 / 6.84e-09	4.29e-09 / -9.32e-10	0 / 0	0 / 0	-3.02e-09 / 9.07e-09
MET_SoftTrk_Scale	6.95e-09 / -6.95e-09	-1.88e-09 / 1.5e-09	0 / 0	0 / 0	0 / 0	3.02e-09 / -3.78e-09
JET_PunchThrough_MC16	-4.63e-09 / 4.63e-09	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Detector1	-1.04e-08 / 8.11e-09	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Mixed1	-5.22e-09 / 7.54e-09	1.47e-08 / -1.83e-08	0 / 0	0 / 0	0 / 0	7.56e-10 / -2.27e-09
JET_EffectiveNP_Mixed2	-1.53e-08 / 1.53e-08	-9.79e-09 / 1.08e-08	-6.3e-09 / -4.03e-09	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Mixed3	-1.77e-08 / 1.53e-08	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Modelling1	5.57e-09 / -7e-09	-4.74e-09 / -5.56e-09	3.11e-09 / 5e-10	0 / 0	0 / 0	4.16e-09 / 1.7e-08
JET_EffectiveNP_Modelling2	-1.68e-08 / 1.68e-08	0 / 0	-1.68e-09 / -5.96e-09	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Modelling3	1.25e-08 / -1.25e-08	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Modelling4	-3.19e-09 / 3.19e-09	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Statistical1	-2e-08 / 2e-08	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Statistical3	6.66e-09 / -6.66e-09	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Statistical4	9.54e-09 / -9.54e-09	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Statistical5	4.06e-09 / -4.06e-09	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
JET_EffectiveNP_Statistical6	1.59e-08 / -1.59e-08	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0
MET_SoftTrk_ResoPara	2.03e-09 / 1.54e-08	4.31e-09 / 1.11e-08	-0.0873 / 0.0873	0 / 0	0 / 0	-1.29e-08 / 1.13e-08
theory_systematics_Zy	0 / 0	0.0745 / -0.0745	0 / 0	0 / 0	0 / 0	0 / 0
JET_NonClosure_2018data	0 / 0	-3e-10 / 2.4e-09	7.57e-10 / 8.39e-10	0 / 0	0 / 0	-9.45e-09 / 9.07e-09
MUON_SCALE	0 / 0	9.84e-09 / 2.29e-09	0 / 0	0 / 0	0 / 0	5.67e-09 / 2.65e-09
MET_SoftTrk_ResoPerp	0 / 0	2.75e-09 / 3e-09	-0.0872 / 0.0872	0 / 0	0 / 0	-1.74e-08 / 1.74e-08
binsVaried	0 / 0	0 / 0	0 / 0	0.262 / -0.262	0 / 0	0 / 0
loosePrime2	0 / 0	0 / 0	0 / 0	0.149 / -0.149	0 / 0	0 / 0
fitError	0 / 0	0 / 0	0 / 0	0 / 0	0.0513 / -0.0513	0 / 0
windowError	0 / 0	0 / 0	0 / 0	0 / 0	0.0444 / -0.0444	0 / 0
functionError	0 / 0	0 / 0	0 / 0	0 / 0	0.0194 / -0.0194	0 / 0
theory_systematics_Signal_m150	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0.376 / -0.376
JET_NonClosure_posEta	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-3.74e-08 / 3.82e-08
EG_SCALE_AF2	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-1.29e-08 / 1.4e-08
MUON_ID	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-4.92e-09 / -8.32e-09
JET_RelativeNonClosure_AFII	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	-3.78e-09 / -3.78e-09

Table E.8: Relative effect of each systematic on the yields in the $Z\gamma$ control region. The signal sample used for this table has a mass of 150 GeV.

Appendix F

Fitting

This Appendix summarizes the yields in all regions after the fitting process discussed in Section 8.2.

	$W\gamma$ CR	$Z\gamma$ CR	Z+jets CR	W+jets CR
$W\gamma$	$36,282 \pm 2,672$	0.4 ± 0.2	$133,104 \pm 9,30$	$66,871 \pm 9,558$
$Z\gamma$	$3,969 \pm 437$	$17,488 \pm 1,340$	$45,940 \pm 3,235$	$13,252 \pm 2,920$
Other bkg	$3,523 \pm 870$	355 ± 44	$11,990 \pm 965$	$2,996 \pm 890$
QCD	$6,323 \pm 1,324$	$1,112 \pm 337$	$46,330 \pm 27,698$	$40,677 \pm 5,999$
$e\gamma$	$2,830 \pm 257$	765 ± 57	$103,320 \pm 9,410$	$9,374 \pm 1,347$
Signal(m_{150} GeV)	544 ± 63	10 ± 4	$1,176 \pm 101$	164 ± 28
Total background	$53,472 \pm 3,717$	$19,730 \pm 1395$	$341,861 \pm 34,284$	$133,335 \pm 1,838$
Data	55,729	20,435	348,080	133,983
ratio (bkg/data)	$95.9\% \pm 6.7\%$	$96.5\% \pm 6.8\%$	$98.2\% \pm 9.9\%$	$99.5\% \pm 13.7\%$

Table F.1: Yields of the analysis in all control regions before the fit.

	VR	SR
$W\gamma$	$9,754 \pm 1,138$	$4,470 \pm 495$
$Z\gamma$	701 ± 128	173 ± 31
Other bkg	$1,235 \pm 288$	337 ± 62
QCD	$1,350 \pm 235$	375 ± 145
$e\gamma$	375 ± 32	77 ± 7
Signal(m150 GeV)	842 ± 69	$1,661 \pm 104$
Total background	$14,256 \pm 1,252$	$7,093 \pm 541$
Asimov data	13,473	6,981
ratio (bkg/Asimov data)	$105.8\% \pm 9.2\%$	$101.6\% \pm 6.8\%$

Table F.2: Yields of the analysis in the validation region and the signal region before the fit.

	$W\gamma$ CR	$Z\gamma$ CR	Z+jets CR	W+jets CR
$W\gamma$	$38,274 \pm 1025$	0.5 ± 0.05	$144,138 \pm 8,623$	$68,662 \pm 2,127$
$Z\gamma$	$4,389 \pm 228$	$18,322 \pm 216$	$50,453 \pm 3,087$	$16,000 \pm 1,587$
Other bkg	$4,146 \pm 818$	398 ± 41	$12,532 \pm 894$	$3,612 \pm 782$
QCD	$5,713 \pm 681$	961 ± 188	$33,376 \pm 15,815$	$36,736 \pm 1,414$
$e\gamma$	$2,705 \pm 545$	754 ± 154	$106,428 \pm 20,124$	$8,761 \pm 1,771$
Signal m150	496 ± 49	9 ± 3	1149 ± 106	144 ± 14
Total background	$55,724 \pm 238$	$20,444 \pm 143$	$348,076 \pm 2,245$	$133,915 \pm 372$
Data	55,729	20,435	348,080	133,983
ratio (bkg/data)	$99.9\% \pm 0.4\%$	$100.1\% \pm 0.7\%$	$100.0\% \pm 0.7\%$	$99.9\% \pm 0.3\%$

Table F.3: Yields of the analysis in all control regions after the fit.

	VR	SR
$W\gamma$	$9,975 \pm 417$	$4,483 \pm 171$
$Z\gamma$	779 ± 69	177 ± 14
Other bkg	$1,538 \pm 29$	396 ± 62
QCD	$1,251 \pm 134$	304 ± 77
$e\gamma$	377 ± 76	76 ± 165
Signal m150	817 ± 70	$1,623 \pm 130$
Total background	$14,737 \pm 274$	$7,059 \pm 815$
Asimov data	13,473	6,981
ratio (bkg/Asimov data)	$109.4\% \pm 2.0\%$	$116.1\% \pm 1.3\%$

Table F.4: Yields of the analysis in the validation region and the signal region after the fit.