

**Estimation of the fake/non-prompt lepton background  
in the search for a fermiophobic low-mass charged  
Higgs boson through  $W^\pm\gamma$  resonances with ATLAS**

Celine Beier

Department of Physics  
McGill University, Montreal  
September 2023

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

©Celine Beier, 2023

# Abstract

Charged Higgs bosons are a proposed theoretical extension to the Standard Model of particle physics. Currently, a search for a fermiophobic charged Higgs boson  $H_5^\pm$  with a mass between 110 GeV and 200 GeV through the  $H_5^\pm \rightarrow W^\pm\gamma$  decay is being developed. The analysis is based on data collected in the ATLAS experiment using proton-proton collisions with a center of mass energy of  $\sqrt{s} = 13$  TeV and an integrated luminosity of  $140 \text{ fb}^{-1}$ .

The final state of the  $H_5^\pm \rightarrow W^\pm\gamma \rightarrow \ell^\pm\nu_\ell\gamma$  decay consists of one electron or muon and at least one photon. This analysis is therefore subject to a fake/non-prompt lepton background which includes events where charged leptons produced in semi-leptonic decays of hadrons or through photon conversions pass the lepton selection criteria. Additionally, other objects in the detector can be misreconstructed as electrons or muons.

This thesis presents a data-driven estimate of the fake/non-prompt lepton background for the  $H_5^\pm \rightarrow W^\pm\gamma$  analysis consistent with ATLAS-internal guidelines and practices. The fake factor method relies on defining a control region where the fake efficiency, the probability that a fake/non-prompt lepton passes the selection criteria used in the analysis, is measured. The fake efficiency is then extrapolated into the analysis regions, where the expected background can be calculated. Statistical and systematic uncertainties on the fake/non-prompt lepton background estimates are evaluated.

# Résumé

Des bosons de Higgs chargés ont été proposés comme extension théorique du Modèle Standard de la physique des particules. Une recherche pour un boson de Higgs chargé fermiophobique  $H_5^\pm$ , avec une masse entre 110 GeV et 200 GeV et se désintégrant suivant  $H_5^\pm \rightarrow W^\pm\gamma$ , est présentement en cours. L'analyse utilise les données accumulées par l'expérience ATLAS avec des collisions proton-proton à une énergie du centre de masse de  $\sqrt{s} = 13$  TeV et une luminosité intégrée de  $140 \text{ fb}^{-1}$ .

L'état finale de la désintégration  $H_5^\pm \rightarrow W^\pm\gamma \rightarrow \ell^\pm\nu_\ell\gamma$  comporte un électron ou un muon et au moins un photon. L'analyse est donc très sensible au bruit de fond leptoniques "feint/non prompt" qui inclut des événements où des électrons produits dans des désintégrations semi-leptoniques de hadrons ou par conversion photonique satisfont le critère de sélection des leptons. De plus, d'autres objets observés par le détecteur peuvent être faussement reconstruits comme électrons ou muons.

Cette thèse présente un estimé basé sur les données elles-mêmes de bruit de fond leptoniques pour l'analyse  $H_5^\pm \rightarrow W^\pm\gamma$  tout en étant consistant avec les instructions et méthodes propres à ATLAS. La méthode du facteur feint consiste à définir une région de contrôle où le taux feint, c'est-à-dire la probabilité qu'un lepton "feint/non prompt" satisfasse les critères de sélection utilisés dans l'analyse, est mesuré. Le taux feint est alors extrapolé dans les régions d'analyse, où les bruits de fonds attendus peuvent alors être calculés. Les incertitudes statistiques et systématiques de ces estimés sont aussi évaluées.

# Acknowledgments

Particle physics is a collaborative effort and this thesis would not have been possible without the work of many scientists, engineers and technicians in the ATLAS collaboration.

I'm extremely grateful to my supervisors, Professor Dr. Francois Corriveau and Professor Dr. Andreas Warburton, for allowing me to join the McGill ATLAS group, providing me an interesting research project, and patiently supporting me throughout my Master's degree.

Many thanks to the  $H_5^\pm \rightarrow W^\pm \gamma$  working group for allowing me to contribute to their analysis. I'm especially grateful to Zhelun Li and Utsav Patel for helping me navigate the analysis framework, build my analysis on their work, and solve many problems along the way. Thank you to Alessandro Ambler for his patient explanations on the fake/non-prompt lepton background and the amazing CERN physicists who responded to my questions on various email lists.

I would also like to thank Professor Dr. Jean Barrette for his mentoring and the amazing job at the Rutherford and McPherson Collections.

The biggest thanks and all the hugs to my friends in Montreal and around the world! Vielen lieben Dank an meine Familie für all ihre Unterstützung!

## Contribution of the Author

The fake/non-prompt lepton background estimate is meant to be incorporated into the  $H_5^\pm \rightarrow W^\pm\gamma$  analysis. Accordingly, many elements of the analysis have been used in this thesis, including the definitions of the object selections and analysis regions as well as the estimations of the prompt backgrounds and the electron to photon and jet to photon fake backgrounds. They were developed by the  $H_5^\pm \rightarrow W^\pm\gamma$  working group, consisting of Utsav Patel, Zhelun Li, Professor Dr. Mark Kruse, Professor Dr. Andreas Warburton, Professor Dr. Francois Corriveau and Dr. Xuan Yang.

I developed the analysis of the fake/non-prompt lepton background in Chapter 8. The techniques are based on guidelines from the ATLAS Isolation and Fake Forum. I have used other ATLAS analyses as references on the practical implementation of these methods.

Parts of my analysis are done with the Analysis Top framework, which was developed from contributions from ATLAS scientists all over the world.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>High-Energy Particle Physics</b>	<b>3</b>
2.1	The Standard Model . . . . .	3
2.1.1	Fundamental Forces and Gauge Bosons . . . . .	3
2.1.2	Fundamental Fermions . . . . .	3
2.1.3	The Higgs Boson . . . . .	4
2.2	Physics beyond the Standard Model . . . . .	4
2.3	Search for $H_5^\pm \rightarrow W^\pm\gamma$ . . . . .	5
<b>3</b>	<b>The ATLAS Experiment at the LHC</b>	<b>7</b>
3.1	The Large Hadron Collider . . . . .	7
3.2	The ATLAS Detector . . . . .	7
3.2.1	The Inner Detector . . . . .	9
3.2.2	The ATLAS Calorimeters . . . . .	10
3.2.3	The Muon Spectrometer . . . . .	12
3.3	The ATLAS Trigger System . . . . .	14
<b>4</b>	<b>Datasets</b>	<b>16</b>
4.1	Trigger Selection . . . . .	16
4.2	Data Samples . . . . .	16
4.3	Monte Carlo Samples . . . . .	17
<b>5</b>	<b>Object Reconstruction and Selection</b>	<b>18</b>
5.1	Electrons . . . . .	18
5.1.1	Electron Reconstruction . . . . .	18
5.1.2	Electron Identification . . . . .	18
5.1.3	Electron Isolation . . . . .	19
5.1.4	Electron Selection . . . . .	20
5.2	Muons . . . . .	20
5.2.1	Muon Reconstruction . . . . .	20
5.2.2	Muon Identification . . . . .	21
5.2.3	Muon Isolation . . . . .	22
5.2.4	Muon Selection . . . . .	22
5.3	Photons . . . . .	23
5.4	Jets . . . . .	23

5.5	Missing Transverse Energy . . . . .	24
5.6	Overlap Removal . . . . .	24
<b>6</b>	<b>Event Selection</b>	<b>25</b>
6.1	$H_5^\pm \rightarrow W^\pm\gamma$ Analysis Regions . . . . .	25
6.2	Fake Efficiency Control Region . . . . .	27
<b>7</b>	<b><math>H_5^\pm \rightarrow W^\pm\gamma</math> Backgrounds</b>	<b>29</b>
7.1	Electron to Photon Fakes . . . . .	29
7.2	Jet to Photon Fakes . . . . .	30
<b>8</b>	<b>Fake/Non-Prompt Lepton Background for <math>H_5^\pm \rightarrow W^\pm\gamma</math></b>	<b>32</b>
8.1	Fake Factor Method . . . . .	32
8.2	Determination of the Fake Efficiency and Fake Factor . . . . .	34
8.3	Fake/Non-Prompt Lepton Background Estimates in the Analysis Regions . .	43
8.4	Statistical and Systematic Uncertainties . . . . .	52
8.4.1	Statistical Uncertainties . . . . .	53
8.4.2	MET Bias . . . . .	53
8.4.3	$ d_0 /\sigma(d_0)$ Bias . . . . .	53
8.4.4	Theory Uncertainties from MC Simulations . . . . .	54
8.4.5	Fake/Non-Prompt Lepton Composition . . . . .	55
<b>9</b>	<b>Conclusion and Outlook</b>	<b>61</b>
	<b>Bibliography</b>	<b>64</b>

# List of Figures

1	Feynman diagrams for the main production mechanisms for $H_5^\pm$ . . . . .	6
2	Schematic view of the CERN accelerator complex . . . . .	8
3	Schematic view of the ATLAS detector . . . . .	9
4	Schematic view of the ATLAS Inner Detector . . . . .	11
5	Schematic view of the ATLAS calorimeter system . . . . .	12
6	Schematic view of the ATLAS muon spectrometer . . . . .	13
7	The ATLAS trigger system in Run 2 . . . . .	15
8	Fake efficiency estimates in the electron channel for data . . . . .	37
9	Fake efficiency estimates in the electron channel for $Z + jets$ MC . . . . .	37
10	Fake efficiency estimates in the muon channel for data . . . . .	38
11	Fake efficiency estimates in the muon channel for $Z + jets$ MC . . . . .	38
12	Fake efficiency estimates in the electron channel for data and $Z + jets$ MC projected into the $p_T$ and $\eta$ dimensions . . . . .	39
13	Fake efficiency estimates in the muon channel for data and $Z + jets$ MC projected into the $p_T$ and $\eta$ dimensions . . . . .	39
14	Fake factor estimates in the electron channel for data . . . . .	40
15	Fake factor estimates in the electron channel for $Z + jets$ MC . . . . .	40
16	Fake factor estimates in the muon channel for data . . . . .	41
17	Fake factor estimates in the muon channel for $Z + jets$ MC . . . . .	41
18	Fake factor estimates in the electron channel for data and $Z + jets$ MC pro- jected into the $p_T$ and $\eta$ dimensions . . . . .	42
19	Fake factor estimates in the muon channel for data and $Z + jets$ MC projected into the $p_T$ and $\eta$ dimensions . . . . .	42
20	Fake/non-prompt electron yield estimates in the validation region . . . . .	46
21	Fake/non-prompt muon yield estimates in the validation region . . . . .	46
22	Fake/non-prompt electron yield estimates in the signal region . . . . .	47
23	Fake/non-prompt muon yield estimates in the signal region . . . . .	47
24	Fake/non-prompt electron yield estimates in the $W^\pm\gamma$ control region . . . . .	48
25	Fake/non-prompt muon yield estimates in the $W^\pm\gamma$ control region . . . . .	48
26	Fake/non-prompt electron yield estimates in the $Z^0\gamma$ control region . . . . .	49
27	Fake/non-prompt muon yield estimates in the $Z^0\gamma$ control region . . . . .	49
28	Fake/non-prompt electron yield estimates in the $W^\pm + jets$ control region . . . . .	50
29	Fake/non-prompt muon yield estimates in the $W^\pm + jets$ control region . . . . .	50
30	Fake/non-prompt electron yield estimates in the $Z^0 + jets$ control region . . . . .	51

31	Fake/non-prompt muon yield estimates in the $Z^0 + jets$ control region . . .	51
32	Truth composition of electrons in the fake efficiency control region . . . . .	58
33	Truth composition of muons in the fake efficiency control region . . . . .	58
34	Truth composition of electrons in the validation region . . . . .	59
35	Truth composition of muons in the validation region . . . . .	59
36	Truth composition of electrons in the signal region . . . . .	60
37	Truth composition of muons in the signal region . . . . .	60

# List of Tables

1	Triggers used in the analysis . . . . .	16
2	Good Run Lists used in the analysis . . . . .	17
3	Luminosity of the data . . . . .	17
4	Isolation working point definitions for electrons . . . . .	19
5	Selection cuts for electrons . . . . .	20
6	Isolation working point definitions for muons . . . . .	22
7	Selection cuts for muons . . . . .	22
8	Isolation working point definitions for photons . . . . .	23
9	Definition of the signal region . . . . .	26
10	Definition of the validation region . . . . .	27
11	Definition of the background control regions . . . . .	28
12	Backgrounds for the $H_5^\pm \rightarrow W^\pm\gamma$ analysis . . . . .	31
13	Bin edges for the fake efficiency calculation . . . . .	35
14	Fake/non-prompt lepton estimates in the analysis regions . . . . .	43
15	Bin edges in $m_{W^\pm\gamma}$ in the analysis regions . . . . .	44
16	Statistical and systematic uncertainties on the fake/non-prompt electron back- ground . . . . .	52
17	Statistical and systematic uncertainties on the fake/non-prompt muon back- ground . . . . .	52
18	Comparison of fake/non-prompt leptons from light flavored jets to fake/non- prompt leptons from heavy flavored jets . . . . .	56

## List of Abbreviations

<b>BSM</b>	Beyond the Standard Model
<b>CB</b>	Combined Muon
<b>CSC</b>	Cathode Strip Chambers
<b>CT</b>	Calorimeter-Tagged Muon
<b>ECAL</b>	Electromagnetic Calorimeter
<b>FCAL</b>	Forward Calorimeter
<b>GRL</b>	Good Run Lists
<b>HCAL</b>	Hadronic Calorimeter
<b>HEC</b>	Hadronic End-cap Calorimeter
<b>HL-LHC</b>	High-Luminosity LHC
<b>HLT</b>	High Level Trigger
<b>ID</b>	Inner Detector
<b>IFF</b>	ATLAS Isolation and Fake Forum
<b>IO</b>	Inside-Out Combined Muon
<b>JVT</b>	Jet-Vertex Tagger
<b>LAr</b>	Liquid Argon
<b>LHC</b>	Large Hadron Collider
<b>MC</b>	Monte Carlo
<b>MDT</b>	Monitored Drift Tubes
<b>ME</b>	Muon-Spectrometer Extrapolated Muon
<b>MET</b>	Missing Transverse Energy
<b>MS</b>	Muon Spectrometer
<b>PMG</b>	ATLAS Particle Modeling Group

<b>ROI</b>	Region of Interest
<b>RPC</b>	Restrictive Plate Chambers
<b>SCT</b>	Semiconductor Tracker
<b>SM</b>	Standard Model
<b>ST</b>	Segment-Tagged Muon
<b>TGC</b>	Thin Gap Chambers
<b>TRT</b>	Transition Radiation Tracker
<b>WP</b>	Working Point

# 1 Introduction

The Standard Model of particle physics summarizes our current understanding of the particles in the universe and their interactions. It has been very successful in providing descriptions of many particle physics processes and predicting particles before their experimental observation. However, several observed phenomena cannot be explained in the Standard Model framework, for example neutrino masses, dark matter, the matter-antimatter symmetry or gravity. Many theories have been developed to extend the Standard Model to include these observations.

For instance, several Standard Model extensions describe a larger Higgs sector with additional Higgs bosons, for example electrically charged Higgs bosons. Currently, an analysis is searching for a fermiophobic charged Higgs boson through the  $H_5^\pm \rightarrow W^\pm\gamma$  decay channel [1, 2]. In contrast to previous analyses, Higgs boson masses between 110 GeV and 200 GeV are considered, a mass region that has not been probed with the Large Hadron Collider (LHC). The LHC is the world's most energetic particle accelerator and provides proton-proton collisions with a center of mass energy of  $\sqrt{s} = 13$  TeV. The  $H_5^\pm \rightarrow W^\pm\gamma$  analysis is using data collected with the ATLAS experiment at the LHC with an integrated luminosity of  $140 \text{ fb}^{-1}$ .

For an accurate descriptions of the measured signal, backgrounds from events that have similar signatures in the detector need to be accounted for. While most backgrounds can be simulated, backgrounds related to the misidentification and misattribution of particles have to be estimated with data-driven methods as their simulation cannot be expected to be accurate. For the  $H_5^\pm \rightarrow W^\pm\gamma$  analysis three data-driven background calculations need to be considered: electrons misidentified as photons, jets misidentified as photons and fake/non-prompt leptons. The fake/non-prompt lepton background includes objects in the detector that are incorrectly reconstructed as leptons as well as real leptons from different production processes that are misattributed.

This thesis aims to estimate the fake/non-prompt lepton background for the  $H_5^\pm \rightarrow W^\pm\gamma$  analysis. Based on ATLAS-internal guidelines, the fake factor method is used [3]. It relies on measuring the fake efficiency, the probability that a fake/non-prompt lepton passes the lepton selection criteria, in a well-understood control region. The results are then extrapolated into the analysis regions, where the expected background can be calculated.

In this thesis, Chapter 2 gives a brief introduction into the Standard Model, searches for physics beyond the Standard Model and the  $H_5^\pm \rightarrow W^\pm\gamma$  analysis. Chapter 3 describes the ATLAS detector used to collect the data for this analysis which is detailed in Chapter 4. The reconstruction of particles from detector information and applied selection criteria are

summarized in Chapter 5, followed by the description of the event selection in Chapter 6. An overview of the backgrounds for the  $H_5^\pm \rightarrow W^\pm \gamma$  analysis is given in Chapter 7 and Chapter 8 details the estimation of the fake/non-prompt lepton background. Chapter 9 contains a summary and outlook.

## 2 High-Energy Particle Physics

Particle physics investigates the fundamental constituents of the universe, elementary particles, and the four fundamental forces: electromagnetism, the weak and strong interactions and gravity.

### 2.1 The Standard Model

The Standard Model (SM) is the theoretical framework summarizing our current understanding of particle physics. It describes all known elementary particles and their interactions. The following short introduction is based on [4, 5, 6].

#### 2.1.1 Fundamental Forces and Gauge Bosons

In the SM, three of the four fundamental forces, the electromagnetic, weak and strong interactions, are described by a quantum field theory. The quanta of the corresponding gauge fields are spin 1 gauge bosons which mediate the interaction. Electromagnetism, described by quantum electrodynamics, and the weak interaction can be combined within the frame of the electroweak theory. The gauge boson for electromagnetism is the photon, while the force carriers for the weak interaction are the  $W^+$ ,  $W^-$  and  $Z^0$  bosons. The strong interaction, described by quantum chromodynamics, is exchanged by eight gluons.

In order to interact via one of the three forces a particle is required to carry the corresponding charge: an electric charge for electromagnetic interactions, a weak isospin for the weak force or a colour charge (red, blue, green and the respective anti-colours) for the strong force.

Gravity is the fourth fundamental interaction which cannot yet be explained within the framework of the SM.

#### 2.1.2 Fundamental Fermions

The SM includes 12 spin  $\frac{1}{2}$  matter particles called fermions. They are classified into three generations with the corresponding particles differing from each other in their masses and lifetimes. With the exception of neutrinos, particles in a higher generation have a higher mass and will decay into lighter, more stable particles. Common matter is made up from particles of the first generation while particles in the second and third generation can be produced in high-energy environments such as astrophysical sources or particle accelerators.

Each generation includes two quarks and two leptons, a charged lepton (electron, muon, tau) and a neutrino.

All fermions carry a weak isospin and can therefore interact via the weak force. Electrons, muons and taus as well as quarks also hold an electric charge while only quarks carry a colour charge. Quarks are never observed as single particles but only in colour-neutral composites called hadrons.

For every fundamental fermion there exists an anti-particle. Particle and anti-particle have the same mass and lifetime, but hold opposite charges.

### 2.1.3 The Higgs Boson

The Higgs boson is the spin 0 scalar boson of the Higgs field. Quarks, charged leptons and  $W^+$ ,  $W^-$  and  $Z^0$  bosons interact with this field through which they acquire mass. The Higgs boson was predicted by the SM and discovered in 2012 by the ATLAS [7] and CMS [8] experiments.

## 2.2 Physics beyond the Standard Model

The SM has a large number of free parameters that are tuned to match experimental observations instead of being derived from theoretical principles [4]. With this approach, the SM had remarkable predictive success with for example the tau [9], the tau neutrino [10] and the Higgs boson [11] being experimentally observed after their theoretical prediction. However, the SM is not a complete description of all natural phenomena, and there are a number of open questions:

In the SM neutrinos cannot acquire mass through interaction with the Higgs field like other SM particles and are therefore predicted to be massless. However, neutrino oscillation was experimentally observed [12], a process that can only occur if neutrinos have mass and explanations of neutrino mass require physics beyond the SM [13]. The SM is also inconsistent with cosmological observations which conclude that baryonic matter, the matter described in the SM, only accounts for up to 5% of the critical density in the universe. The nature of dark matter and dark energy, which account for 25% and 70% of the critical density respectively, is so far unknown [14]. Similarly, there is no current explanation for the matter-antimatter asymmetry that produced our matter-dominated universe [4]. Finally, the SM only includes three out of four fundamental forces and does not provide a description of gravity.

Many theoretical extensions have been developed to include these processes and phenomena into the SM framework. The predictions of these beyond Standard Model (BSM) theories can be tested with particle accelerator experiments.

### 2.3 Search for $H_5^\pm \rightarrow W^\pm \gamma$

In the SM, the Higgs mechanism assumes a minimal Higgs sector with a single scalar boson. This is the simplest model choice, however, other BSM theories with an extended Higgs sector can be constructed. One possibility is the inclusion of electrically charged Higgs bosons [4]. These would interact electromagnetically in addition to their interaction with the Higgs field and therefore require new search strategies. Multiple searches for charged Higgs bosons coupling to fermions, especially top quarks, have been performed, for example in [15, 16, 17]. In contrast, fermiophobic charged Higgs particles preferably interact with bosons. They have been proposed by different BSM models, like the Georgi-Machacek model [18], the Stealth Doublet model [19], and the two-Higgs-doublet model [20].

In the Georgi-Machacek model the SM Higgs sector, which consists of a complex scalar doublet, is extended by a real triplet and a complex triplet. The resulting physical Higgs fields can be grouped by their transformation properties under  $SU(2)$  symmetry into one quintuplet, one triplet and two singlets. The Higgs bosons of the quintuplet group,  $H_5^{++}$ ,  $H_5^+$  and  $H_5^0$ , all have the same mass  $m_5$ . In contrast to the triplet bosons they do not have a complex scalar doublet component in their construction and are therefore fermiophobic, but do couple to vector bosons. For  $m_5 < m_3$  the quintuplet Higgs bosons will only decay into SM bosons [21].

$H_5^\pm$  can be produced in Drell-Yan reactions in proton-proton collisions through three main production channels [21]:

$$pp \rightarrow H_5^\pm + H_5^0 \quad (2.1)$$

$$pp \rightarrow H_5^\pm + H_5^{\mp\mp} \quad (2.2)$$

$$pp \rightarrow H_5^+ + H_5^- \quad (2.3)$$

The Feynman diagrams are shown in Figure 1. The cross sections only depend on the mass

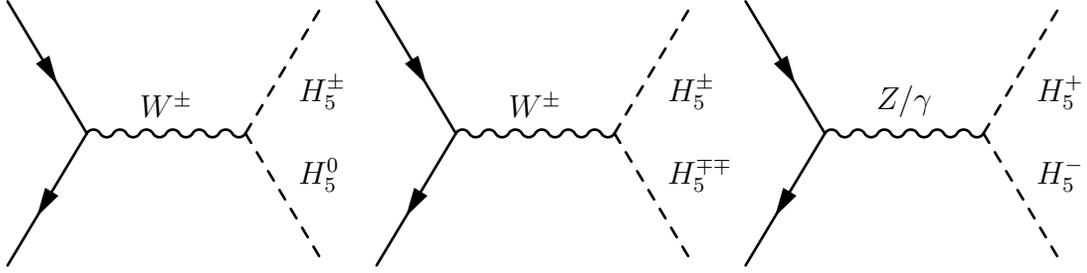


Figure 1: Feynman diagrams for the main production mechanisms for  $H_5^\pm$  [21].

of the charged Higgs bosons  $m_5$ , and the mechanisms are ordered above from highest to lowest cross section [21].

There have been multiple searches for  $H_5^\pm$  Higgs bosons decaying at tree-level into a  $W^\pm$  and a  $Z^0$  boson:

$$H_5^\pm \rightarrow W^\pm Z^0 \quad (2.4)$$

These analyses, for example in [22, 23, 24], have focused on masses above 200 GeV. Currently, an analysis using data collected by the ATLAS experiment from 2015-2018 in  $\sqrt{s} = 13$  TeV proton-proton collisions is searching for  $H_5^\pm$  with a mass between 110 GeV and 200 GeV decaying into a  $W^\pm$  boson and a photon [1, 2]:

$$H_5^\pm \rightarrow W^\pm \gamma \quad (2.5)$$

This decay occurs at one loop and is expected to have a lower branching ratio than the competing tree-level decay  $H_5^\pm \rightarrow W^\pm Z^0$ . However, if the charged Higgs boson has a mass below the  $W^\pm Z^0$  threshold at 171 GeV, the tree-level decay is suppressed and the decay  $H_5^\pm \rightarrow W^\pm \gamma$  can be expected to be the main process [21].

The produced  $W^\pm$  boson will further decay:

$$W^\pm \rightarrow \ell^\pm \nu_\ell \quad (2.6)$$

so that the final state of the Higgs boson decay is  $\ell^\pm \nu_\ell \gamma$ . Only electrons and muons in the final state are considered as those particle have lifetimes long enough to be directly detected.

## 3 The ATLAS Experiment at the LHC

### 3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the world's most powerful particle accelerator. It has a circumference of 27 km and is located about 100 m underground at the Swiss-French boarder at the European Organization for Nuclear Research CERN. The LHC was designed to facilitate proton-proton collisions as well as collisions involving heavy ions for four main experiments, the general-purpose detectors ATLAS [25] and CMS [26], the LHCb experiment [27] specialized on physics involving a  $b$  quark and the ALICE detector [28] focused on studying the quark-gluon plasma created in heavy ion collisions [29].

Before injection into the LHC, protons and heavy ions need to be pre-accelerated in the CERN accelerator system shown in Figure 2. For protons, the gradual increase of particle energy and the beam bunching happens in linear (LINAC 2) and circular (BOOSTER, PS and SPS) accelerators. At an energy of 450 GeV proton bunches from the SPS are injected into two counter-rotating vacuum beam pipes which make up the LHC. Acceleration then happens through electric fields in radio-frequency cavities while bending and focusing is achieved with superconducting magnets. The collision of proton bunches happens in four interaction points with a frequency of 40 MHz [29].

The first proton-proton collisions at the LHC were recorded in 2009 and since then there have been three data collection periods called runs. Between runs, the LHC is periodically shutdown to allow for repairs, maintenance and upgrades on the accelerator and experiments [31]. Run 1 (2011-2012) operated with beam energies of 3.5 TeV and 4 TeV corresponding to center of mass energies of  $\sqrt{s} = 7$  TeV and  $\sqrt{s} = 8$  TeV [31]. During Run 2 (2015-2018) the beam energy was increased to 6.5 TeV ( $\sqrt{s} = 13$  TeV) and currently Run 3 (2022-2026) is collecting data with a targeted maximum beam energy of 7 TeV ( $\sqrt{s} = 14$  TeV) [32, 33].

In the next years, the LHC will undergo a number of upgrades aimed to upkeep and improve its performance. The new configuration, known as the High-Luminosity LHC (HL-LHC), will provide improved statistics and accordingly higher chances of observing new physics phenomena [34].

### 3.2 The ATLAS Detector

ATLAS (**A** Toroidal LHC Apparatu**S**) is a multi-purpose detector centered around a beam interaction point of the LHC. A schematic diagram is shown in Figure 3. The ATLAS

## The CERN accelerator complex *Complexe des accélérateurs du CERN*

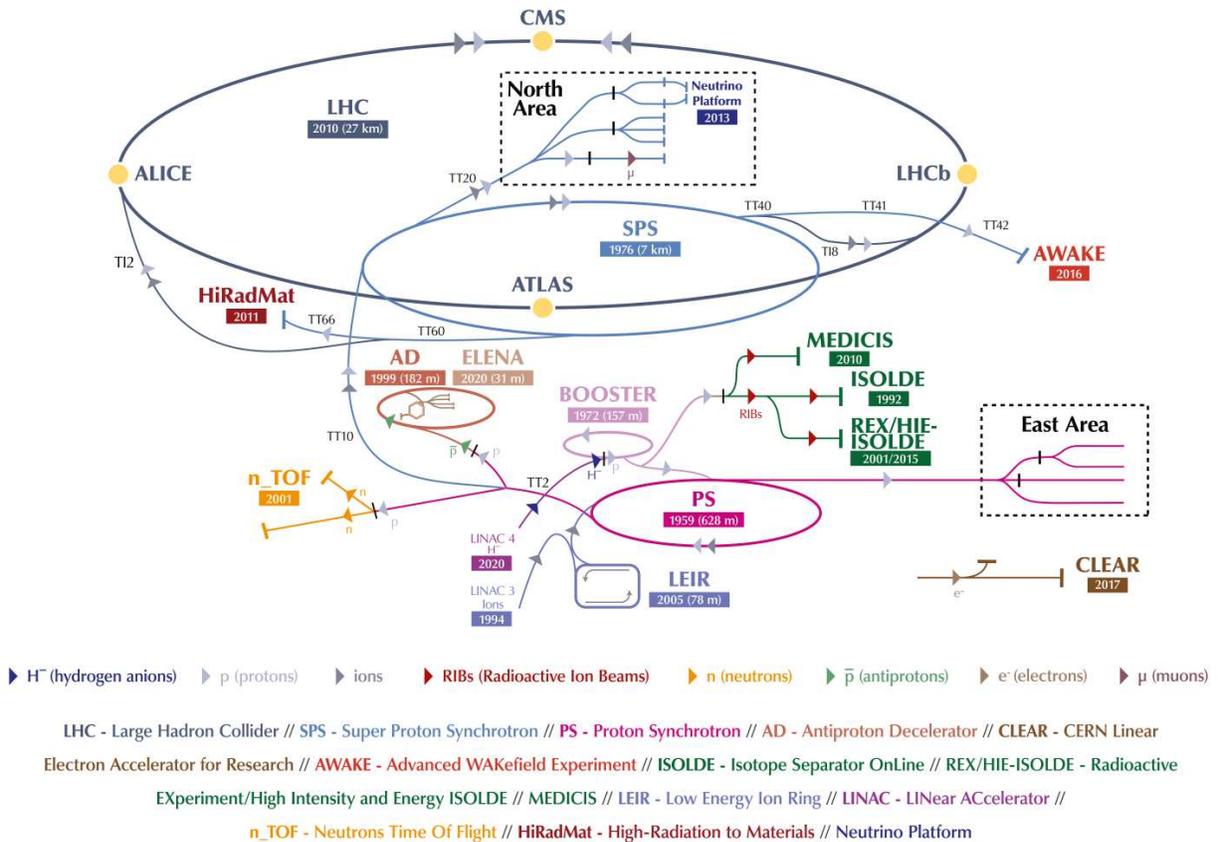


Figure 2: Schematic view of the CERN accelerator complex in 2022 [30].

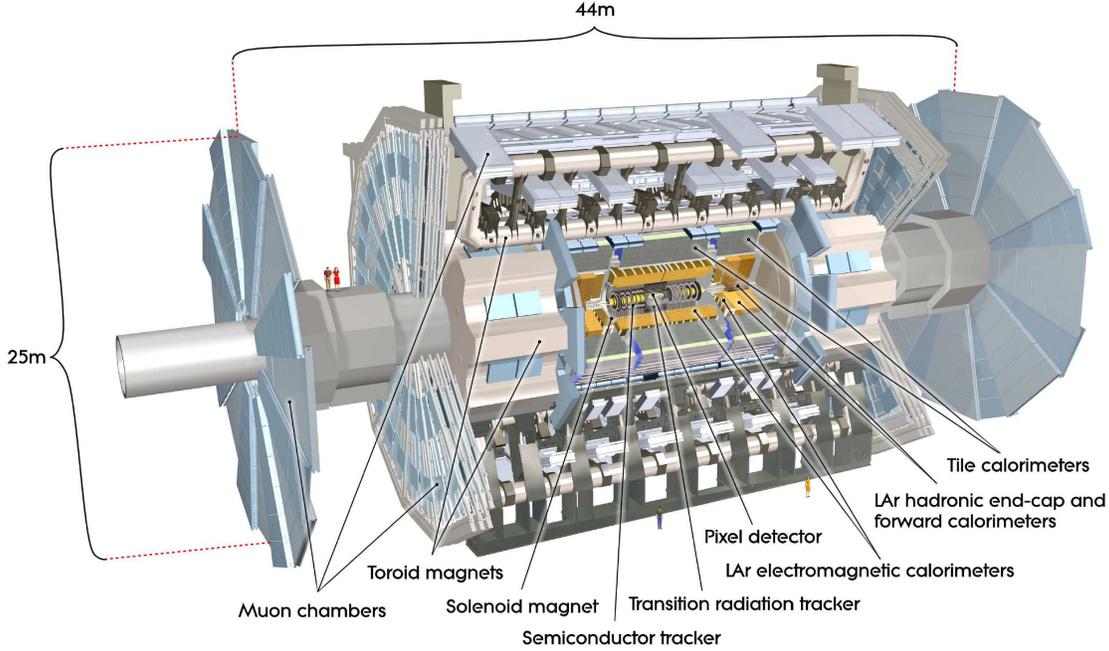


Figure 3: Schematic view of the ATLAS detector [25].

detector has a cylindrical geometry that is forward-backward symmetric. Its subsystems are often different for the barrel, the central cylindrical region surrounding the beam pipe, and two end-cap regions attached to both sides of the barrel [25]. In the following, the main subsystems of the ATLAS detector are briefly described in their configuration for Run 2 when the data used for this thesis was collected. The detector was changed during the shutdowns before and after Run 2 to upkeep and improve operation and will undergo major changes in the scope of the high-luminosity upgrade [25, 34].

### 3.2.1 The Inner Detector

The Inner Detector (ID), shown in Figure 4, is meant to record the tracks of charged particles close to the proton-proton interaction point, which is critical for reconstructing primary and secondary vertices. It also measures the momentum and charge of particles [25].

The detectors closest to the beam pipe are three layers of pixel detectors which provide three track measurement points. The pixel detectors are segmented silicon semiconductor detectors, p-n-junctions with an applied reverse bias which increases the depletion region between the two sides. In the depletion region charged particles create electron-hole pairs which can be detected as an electric pulse on the electrodes [35]. There are 1744 pixel sensor modules, each with 46000 pixels, which gives a total of about 80.4 million readout channels

[25]. A fourth layer of pixel detectors, the B-Layer, was installed in 2014 between the beam pipe and the innermost layer of pixel detectors [36].

The middle layer of the ID is the Semiconductor Tracker (SCT), a silicon semiconductor detector split into micro-strips. In the barrel region there are 4 double layers which provide 8 tracking points, while in the end-cap region 9 double layers are installed. In total the SCT provides 6.3 million readout channels [25].

The outer ID layer is made up from the Transition Radiation Tracker (TRT). It is built from 298,000 individual straw chambers, gas filled tubes with an electric wire in the middle. Charged particles passing through the detector ionize the gas and the resulting charged particles create a current in an electric field applied between the chamber walls and the wire. On average, the TRT provides 36 hits per passing particle [25, 35]. The space between the individual tubes is filled with material for the creation of transition radiation. Transition radiation is electromagnetic radiation emitted by relativistic charged particles when crossing the boundary between two materials with different dielectric properties. The photons can be detected in the straw chambers where accordingly the measured signal for particles that emit transition radiation will be much stronger. Since the transition radiation effect is stronger for light, highly energetic particles, its detection is used for the identification of electrons [35].

The ID is located within a solenoid magnet which supplies a magnetic field of 2 T. Moving charged particle trajectories are bent in the magnetic field due to the Lorentz force and the direction and magnitude of the bending can be used to determine the charge and momentum of the passing particle [25, 37].

### 3.2.2 The ATLAS Calorimeters

Calorimeters stop a moving particle and measure the deposited energy. High energy photons, electrons and hadrons create secondary particles when interacting with a material, which in turn interact and produce more particles. These particle cascades, called electromagnetic or hadronic showers, continue until all the primary particle's energy has been absorbed by the medium [35]. ATLAS has two calorimeter systems, shown in Figure 5, one for electromagnetic and one for hadronic showers. Both are sampling calorimeters, which means that they possess different, alternating materials used to generate the showers (absorbing layer) and to measure the deposited energy (active layer) [25, 35].

The inner calorimeter is the Liquid Argon (LAr) Electromagnetic Calorimeter (ECAL), which includes a barrel part and two end-cap components. The absorber material is lead-

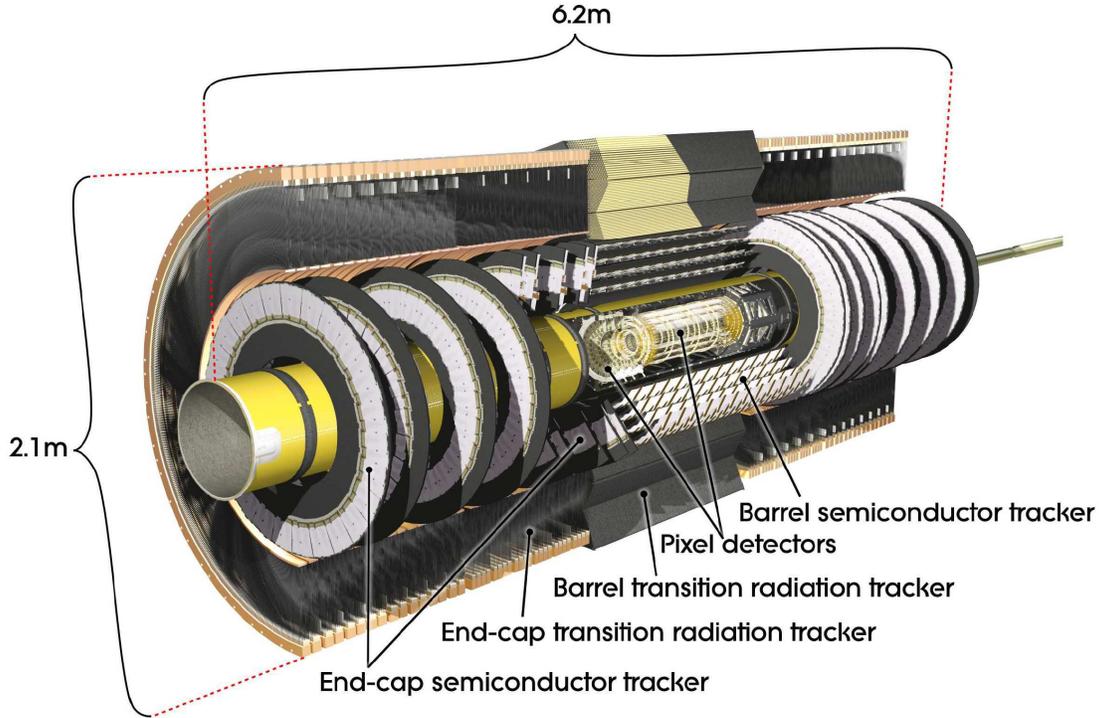


Figure 4: Schematic view of the ATLAS Inner Detector [25].

stainless steel in an accordion geometry, while the active layer is liquid argon. Passing charged particles ionize the argon and an applied voltage generates a current which can be recorded [25, 37].

The ECAL is surrounded by the Hadronic Calorimeter (HCAL) which stops most hadronic particles that pass through the ECAL. The barrel region is further divided into three individual barrel detectors, which consist of layers of steel as the absorber and polystyrene scintillators as the active material. Ionizing particles create scintillation photons which are collected by wavelength-shifting fibers and guided to the back of the structure where they are recorded by photo-multiplier tubes [25, 35]. On the end-cap sides the Hadronic End-cap Calorimeter (HEC) is very similar in design. It also relies on liquid argon as the active material, but uses copper plates as the absorber [25].

The Forward Calorimeter (FCal) is a subdetector in the end-cap region of the calorimeter system. It consists of liquid argon as the active material and three layers of absorbers. The layer closest to the interaction point is made out of copper and absorbs most electromagnetically interacting particles while the outer two layers are made out of tungsten which is better suited to trigger showers of hadrons [25, 37].

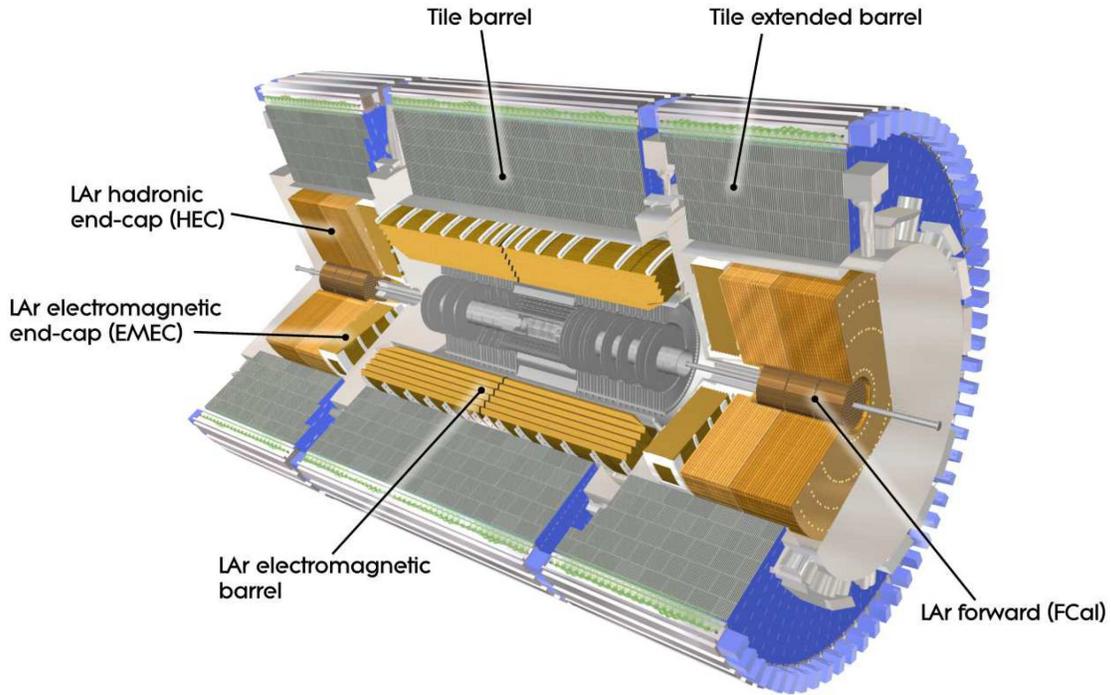


Figure 5: Schematic view of the ATLAS calorimeter system [25].

### 3.2.3 The Muon Spectrometer

Muon tracking and momentum measurement take place in the outermost regions of the ATLAS detector, the Muon Spectrometer (MS) (Figure 6). It is surrounded by two toroidal magnets generating a magnetic field in which muons will be bent and therefore the information from the muon tracking chambers, the Monitored Drift Tubes (MDT) and the Cathode Strip Chambers (CSC), can be used reconstruct the muon momentum [25].

The MDT provide tracking over most of the area around the collision point. They consist of 354,000 individual drift tubes arranged in 1172 chambers which increase in size with their distance from the interaction point [38]. A drift tube is cylindrical cathode filled with a gas mixture of argon and carbon-dioxide and a wire anode along its center. When a muon passes the chamber it produces ionization charges which can be read out [35].

The other part of the muon tracking system, the CSC cover the forward region close to the beam pipe where most muons are expected, so the detector needs to operate in a region with larger particle fluxes, higher track densities and more intense radiation compared to the barrel area covered by the MDT. The CSC consist of four layered rings round the beam pipe, which corresponds to four track measurement points. They are made from multi-wire

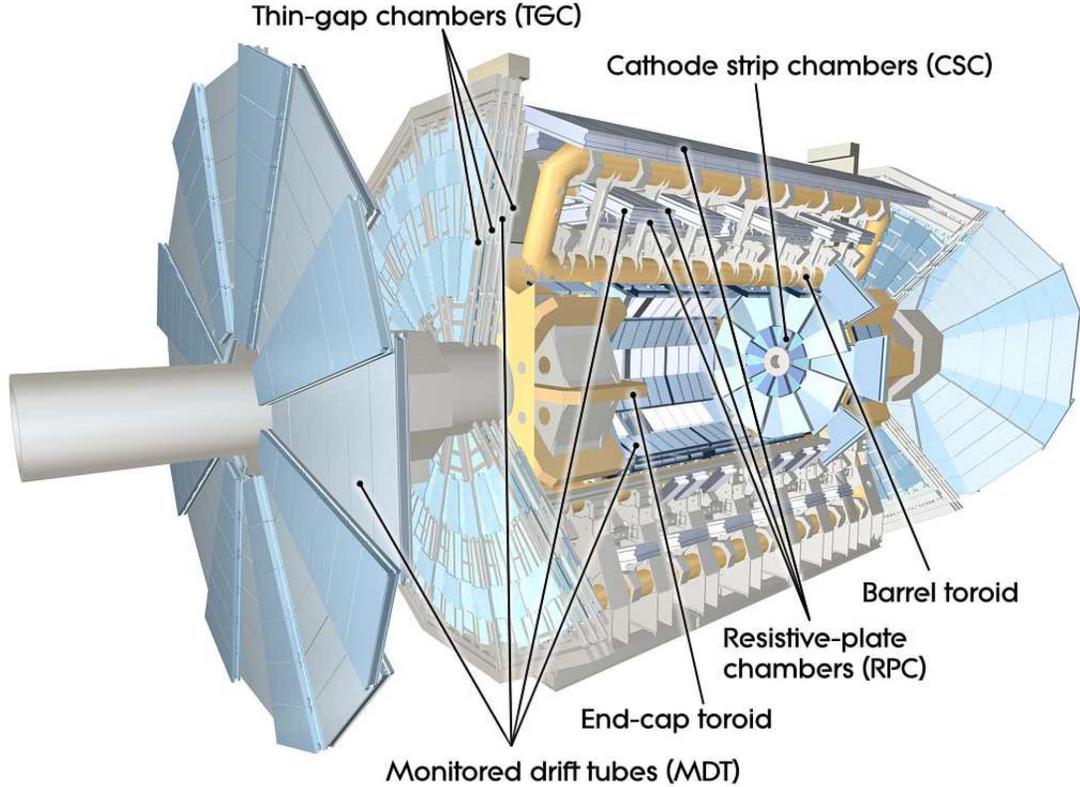


Figure 6: Schematic view of the ATLAS muon spectrometer [25].

proportional chambers filled with an argon and carbon-dioxide mixture. Muons will ionize the argon atoms and an applied voltage between the anode wires and the cathode chamber walls will cause an ionization avalanche that can be detected [25, 35].

The Muon Spectrometer also includes the Resistive Plate Chambers (RPC) in the barrel region and the Thin Gap Chambers (TGC) in the end-cap regions. These two subsystems, optimized for fast recognition of a passing muon, allow for a triggering on muon tracks [25]. The RPC consist of 3 trigger stations, each with 2 detector layers, allowing for 6 possible hits of a passing muon. In total, the RPC are made out of 3,700 individual gaseous parallel-electrode plate detectors. This detection principle relies on a muon ionizing the gas and the electric field applied between two parallel electrodes causing avalanche ionization [25, 35, 39]. In the end-cap region the TGC are made out of multi-wire proportional chambers operated in quasi-saturated mode which are arranged into seven detector layers [25].

### 3.3 The ATLAS Trigger System

The LHC bunch crossing rate at the interaction point is around 40 MHz, which means collisions in the ATLAS detector can be recorded every 25 ns. Not all of these events can be saved, partially due to limitations on the available storage space, but also because of technical limits for reading out the full detector for each event. Therefore the trigger system aims to select specific events for permanent storage [40]. The ATLAS trigger system was redesigned after Run 1 to keep up with the higher luminosity of the LHC during Run 2 and will be adapted to the new requirements in the course of the HL-LHC upgrade as well [41, 42].

The Run 2 trigger system, schematically shown in Figure 7, consists of two levels, which make trigger decisions in parallel to the detector readout. Only if an event is accepted in the trigger does it get sent to permanent storage, otherwise it is discarded [40]. The first trigger stage (Level-1) is a hardware trigger, which reduces the event rate from about 40 MHz to 100 kHz. It relies on electronic detector components that can be read out quickly so that the Level-1 decision time for an “accept” is only  $2.5 \mu\text{s}$ . The Level-1 input information comes from the ECAL and HCAL, where hits can indicate the presence of electrons, photons, taus and jets as well as an overall high total energy of the event or a high missing transverse energy, as well as the muon trigger system, which evaluates data from the RPC and TGC for muons in the event. Together the information from the calorimeters and muon system are used to determine Regions of Interest (ROI), which are then passed to the second stage of the trigger, the high level trigger (HLT) [40]. The HLT is a software trigger which uses more detailed information from all detector components to select events based on selection algorithms that define characteristics which make events interesting for further physics analysis. The HLT has a decision time of about 200 ms and reduces the event rate to about 1 kHz [40].

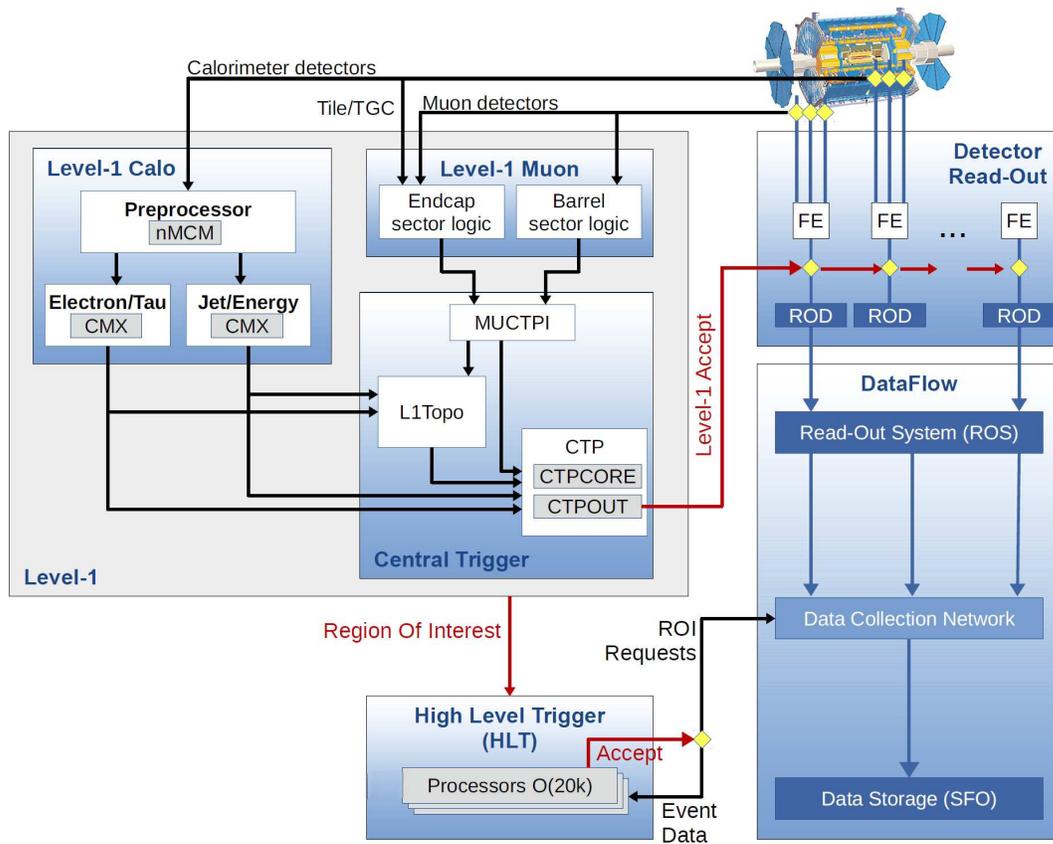


Figure 7: The ATLAS trigger system in Run 2. Adapted from [40].

## 4 Datasets

The  $H_5^\pm \rightarrow W^\pm\gamma$  analysis uses data collected with the ATLAS detector in Run 2 from 2015-2018. In the following, the data samples and Monte Carlo simulations are described.

### 4.1 Trigger Selection

The analysis uses unprescaled single-electron and single-muon triggers, which are tuned to select all events with at least one electron or muon. The trigger applies some identification and isolation requirements to the leptons as well as a selection for a specific minimum transverse momentum.

Since the LHC run conditions changed repeatedly during Run 2 and improvements in the trigger algorithms were made, different triggers are used for each year of Run 2. The details of the trigger definitions can be found in the ATLAS trigger menus for 2015 [43], 2016 [44], 2017 [45] and 2018 [46]. The triggers for this analysis are summarized in Table 1.

Year	Electron Level-1	Muon Level-1
2015	EM20VH	L1_MU20 (L1_MU15)
2016-2018	EM22VHI	L1_MU20
Year	Electron HLT	Muon HLT
2015	e24_lhmedium_L1EM20VH e60_lhmedium e120_lhloose	mu20_iloose_L1MU15 mu50
2016-2018	e26_lhtight_nod0_ivarloose e60_lhmedium_nod0 e140_lhloose_nod0	mu26_ivarmedium mu50

Table 1: List of Level-1 triggers and HLT used in the analysis by year. For 2015, the L1\_MU15 Level-1 trigger is only used in combination with mu20\_iloose\_L1MU15 [47].

### 4.2 Data Samples

The data for this analysis was collected during Run 2 when the LHC was operating at  $\sqrt{s} = 13$  TeV. During data-taking the various conditions of the detector are recorded in Good Run Lists (GRL) which are used to select the periods when all subsystems of the ATLAS detector were working properly and the recorded events can be used for analysis.

Year	Good Run List
2015	data15_13TeV_periodAllYear_DetStatus-v89-pro21-02.Unknown.PHYS.StandardGRL.All.Good.25ns
2016	data16_13TeV_periodAllYear_DetStatus-v89-pro21-01.DQDefects-00-02-04.PHYS.StandardGRL.All.Good.25ns
2017	data17_13TeV_periodAllYear_DetStatus-v99-pro22-01.Unknown.PHYS.StandardGRL.All.Good.25ns.Triggerno17e33prim
2018	data18_13TeV_periodAllYear_DetStatus-v102-pro22-04.Unknown.PHYS.StandardGRL.All.Good.25ns.Triggerno17e33prim

Table 2: Good Run Lists used in the analysis by year.

Year	Integrated Luminosity [ $\text{pb}^{-1}$ ]
2015	3,244.54
2016	33,402.2
2017	44,630.6
2018	58,791.6

Table 3: Integrated luminosity for the data defined in the GRLs for each year of Run 2 [48].

The GRL used are listed in Table 2 [48]. The luminosity for each year is given in Table 3. The total luminosity is  $140 \text{ fb}^{-1}$  with an uncertainty of 1.7% [48].

The output files after reconstruction are too large to be analyzed directly. They are reduced to smaller file versions, called derivations, according to specific analysis needs. This analysis uses the data files produced in the DAOD.PHYS derivation framework with the p4356 and p5314 p-tags. Details on the derivations can be found in [49].

### 4.3 Monte Carlo Samples

Monte Carlo (MC) simulations describing SM and BSM processes are used to model the expected distributions of signal and background processes for the analysis. The events from proton-proton collisions are simulated with MC generators for the given physics theory input. They are then processed with a detector simulator [50, 51]. After this step, the MC datasets are in the same format as the measured data files, but also contain truth information about the simulated process. The object reconstruction is then identical to data. The MC samples are reduced in size for physics analysis with the DAOD.PHYS derivation framework, the p-tags used are p4355 and p5313 [49].

## 5 Object Reconstruction and Selection

In the ATLAS detector different particles will leave characteristic signatures in the detector subsystems which can be used for particle identification and the reconstruction of particle properties. In the following the reconstruction and object selection of electrons, muons, photons and jets will be described in more detail as these physics objects are used in the  $H_5^\pm \rightarrow W^\pm \gamma$  analysis.

### 5.1 Electrons

#### 5.1.1 Electron Reconstruction

The reconstruction of electrons is based on their characteristic signatures in the ATLAS detector, a charged-particle track in the ID and an energy deposit in the ECAL. High-energy electrons will lose a significant amount of their energy to bremsstrahlung when interacting with the detector. The radiated photons will convert into electron-positron pairs, however usually these secondary particles will be emitted close to the original electron. Accordingly, the calorimeter energy deposits are measured in topologically connected clusters to account for all energy from the primary electron and nearby secondary particles. If a calorimeter cluster can be spatially matched to an ID track the object becomes an electron candidate. Since bremsstrahlung emission can occur on the inside of the detector it is possible to match multiple ID tracks to the same cluster, all connected to the same primary electron [52, 53].

#### 5.1.2 Electron Identification

The electron candidates reconstructed from the detector signatures will include fake and non-prompt electrons from background processes such as jets misidentified as electrons, semileptonic decays of heavy quarks and photons converting into electrons. Placing requirements on variables related to the electron candidate's identification and isolation aims to preferentially select the wanted real electrons [52, 53].

Real electrons can be selected using a likelihood-based identification variable. Characteristic features of real electrons, such as the longitudinal and lateral shower shapes in the ECAL, the track quality, track-cluster matching and particle identification variables from the TRT are used as discrimination variables which are combined into a single score. Increasing threshold values on the resulting discriminant define three identification working points (WP) for physics analysis: loose, medium and tight. A higher ID WP corresponds to a higher

Calorimeter-based isolation	
FCTight	$E_T^{iso}(\Delta R < 0.2)/E_T < 0.06$
FCLoose	$E_T^{iso}(\Delta R < 0.2)/E_T < 0.2$
Tracking-based isolation	
FCTight	$p_T^{iso}(\Delta R < 0.2)/p_T < 0.06$
FCLoose	$p_T^{iso}(\Delta R < 0.2)/p_T < 0.15$

Table 4: Isolation WP definitions for electrons used in the analysis with the isolation variables  $E_T^{iso}(\Delta R)$  and  $p_T^{iso}(\Delta R)$  [53].  $E_T$  is the electron’s transverse energy and  $p_T$  its transverse momentum.

background rejection but also a lower identification efficiency. The LooseAndBLayer WP uses the same discriminant threshold as the Loose WP but adds an additional requirement of a hit in the B-Layer [52, 53].

### 5.1.3 Electron Isolation

A typical feature of real electrons is that there is very little detector activity in the area surrounding the electron. In ATLAS, the activity around an electron candidate is quantified in two isolation variables,  $E_T^{iso}(\Delta R)$  and  $p_T^{iso}(\Delta R)$ . Both variables refer to a cone with radius  $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$  around the electron’s direction where where  $\eta$  is the electron’s pseudorapidity and  $\phi$  the azimuthal angle with respect to the beam pipe orientation [52].

The isolation variable  $E_T^{iso}(\Delta R)$  is based on information from the ECAL. For an event, the calorimeter cells are grouped into topological clusters. A topological cluster starts from a calorimeter cell that measures an energy deposit of at least four times more than the expected noise and extends in all three spatial directions to include cells that have energy deposits of at least twice the noise level.  $E_T^{iso}(\Delta R)$  is calculated by summing the transverse energies of all topological clusters with centers within a cone of radius  $\Delta R$  around the electron candidate. The transverse energy contribution from the electron is removed [52].

The other isolation variable  $p_T^{iso}(\Delta R)$  is obtained from tracking information. Analogous to the calorimeter-based isolation variable, the transverse momenta of tracks within a cone of radius  $\Delta R$  around the electron track are summed up, excluding the electron’s contribution [52].

The calorimeter- and track-based isolation variables are used to define isolation working points. The WPs used in this analysis are FCTight and FCLoose, the corresponding requirements on the isolation variables are listed in Table 4.

	<b>Baseline</b>	<b>Tight</b>
$p_T$	$> 27$ GeV	$> 27$ GeV
ID WP	LooseAndBLayer	Tight
Isolation WP	FCLoose	FCTight
$ z_0 \sin(\theta) $	$< 0.5$ mm	$< 0.5$ mm
$ d_0 /\sigma(d_0)$	$< 5$	$< 5$

Table 5: Summary of the cuts included in the two electron selections for this analysis.

#### 5.1.4 Electron Selection

The pseudorapidity range of  $1.37 < |\eta| < 1.52$  corresponds to a crack between the electromagnetic calorimeters in the barrel and end-cap regions. Accordingly, electrons reconstructed in this region are disregarded [25].

For the  $H_5^\pm \rightarrow W^\pm \gamma$  analysis two electron selections are defined as shown in Table 5. The baseline selection is inclusive of the tight selection, which places stronger cuts on electron candidates to ensure the selection of real electrons with little contribution from fake/non-prompt particles. Electrons are required to have a transverse momentum  $p_T > 27$  GeV to be compatible with the triggers used for the analysis, described in Section 4.1. The cuts on the longitudinal impact parameter of  $|z_0 \sin(\theta)|$  and on the significance of the transverse impact parameter  $|d_0|/\sigma(d_0)$  with the transverse impact parameter  $d_0$  its uncertainty  $\sigma(d_0)$  are included to select electrons that are originating from the primary vertex.

## 5.2 Muons

### 5.2.1 Muon Reconstruction

Muon reconstruction in ATLAS can rely on reconstructed tracks in the ID and MS as well as information from the ECAL and HCAL. Depending on which information is available and used for the reconstruction, five different muon types can be defined [54]:

- Combined Muon (CB): Tracks in the ID are matched to tracks in the MS in a combined fit while taking into account energy deposits in the ECAL and HCAL. Silicon-Associated Forward Muon (SiF) is a sub-type for the  $|\eta| > 2.5$  region where MS tracks are matched to tracks from the pixel and SCT detectors.
- Inside-Out Combined Muon (IO): A reconstructed track from the ID is extrapolated to the MS and becomes a muon candidate if at least three hits in the MS are found.

In contrast to the Combined Muon, this technique does not need a reconstructed track in the MS.

- Muon-Spectrometer Extrapolated Muon (ME): A reconstructed track from the MS that does not have a matching track in the ID is extrapolated to the beam line.
- Segment-Tagged Muon (ST): A track in the ID is matched to at least one reconstructed MS segment and the parameters for the muon candidate are taken from the ID track.
- Calorimeter-Tagged Muon (CT): A reconstructed track from the ID is extrapolated through the ECAL and HCAL. If energy deposits characteristic for a minimum-ionising particle are found along the projected line, the object becomes a muon candidate with its parameters taken from the ID track.

### 5.2.2 Muon Identification

For muon candidates, a precision station is defined as a MS segment in which the muon has three or more hits in the MDT or CSC. A precision hole station is a MS segment where at least three hits are recorded, but also at least three hits are missing given the muon trajectory and detector structure. The  $q/p$  compatibility is defined with the charge  $q$  to momentum  $p$  ratios measured in the ID and MS and the respective uncertainties as [54]:

$$q/p \text{ compatibility} = \frac{|(\frac{q}{p})_{ID} - (\frac{q}{p})_{MS}|}{\sqrt{\sigma^2(\frac{q}{p})_{ID} + \sigma^2(\frac{q}{p})_{MS}}} \quad (5.1)$$

With these parameters three identification working points are defined, which are called the muon quality [54]:

- Medium: For  $|\eta| < 2.5$  muons are required to be of the CB or IO type and need to have at least two precision stations. The exception is the region of  $|\eta| < 0.1$  where the muons can also have one precision station and at most one precision hole station. All muon candidates are required to have  $q/p < 7$ . For the  $2.5 < |\eta| < 2.7$  region ME muons and SiF Muons with at least three precision stations are included.
- Loose: All medium muons are included, and in addition CT muons and ST Muons are accepted in the  $|\eta| < 0.1$  range. The  $|\eta| < 1.3$  region is extended to include IO muons with at least one precision station if they are also reconstructed as ST muons.

Calorimeter-based isolation	
FCTight	$E_T^{iso}(\Delta R < 0.2)/E_T < 0.06$
FCLoose	$E_T^{iso}(\Delta R < 0.2)/E_T < 0.3$
Tracking-based isolation	
FCTight	$p_T^{iso}(\Delta R < 0.3)/p_T < 0.06$
FCLoose	$p_T^{iso}(\Delta R < 0.3)/p_T < 0.15$

Table 6: Isolation WP definitions for muons used in the analysis with the isolation variables  $E_T^{iso}(\Delta R)$  and  $p_T^{iso}(\Delta R)$  [53].  $E_T$  is the muon’s transverse energy and  $p_T$  its transverse momentum.

	<b>Baseline</b>	<b>Tight</b>
$p_T$	$> 27$ GeV	$> 27$ GeV
Muon Quality	Loose	Tight
Isolation WP	FCLoose	FCTight
$ z_0 \sin(\theta) $	$< 0.5$ mm	$< 0.5$ mm
$ d_0 /\sigma(d_0)$	$< 3$	$< 3$

Table 7: Summary of the cuts included in the two muon selections for this analysis.

- **Tight:** The muons are required to have the CB or IO type and at least two precision stations. For the track fit the normalized  $\chi^2$  value is restricted to  $\chi^2 < 8$  and there are additional requirements on the  $q/p$  compatibility depending on the transverse momentum  $p_T$  and pseudorapidity  $|\eta|$  of the muon.

### 5.2.3 Muon Isolation

The isolation variables  $E_T^{iso}(\Delta R)$  and  $p_T^{iso}(\Delta R)$  for muon candidates are defined analogous to the isolation variables for electron candidates (Section 5.1.3). The WPs used in this analysis are FCTight and FCLoose; their definitions are summarized in Table 6 [53].

### 5.2.4 Muon Selection

Analogous to the electron selection, two muon selections are defined for this analysis which are summarized in Table 7. Muons are required to have a transverse momentum  $p_T > 27$  GeV to comply with restrictions from the trigger selections. Cuts on the longitudinal impact parameter of  $|z_0 \sin(\theta)|$  and the significance of the transverse impact parameter  $|d_0|/\sigma(d_0)$  ensure that the muon is originating from the primary vertex.

Calorimeter-based isolation	
FCTight	$E_T^{iso}(\Delta R < 0.4) < 0.022E_T + 2.45 \text{ GeV}$
FCLoose	$E_T^{iso}(\Delta R < 0.2)/E_T < 0.65$
Tracking-based isolation	
FCTight	$p_T^{iso}(\Delta R < 0.2)/E_T < 0.05$
FCLoose	$p_T^{iso}(\Delta R < 0.2)/E_T < 0.05$

Table 8: Isolation WP definitions for photons used in the analysis with the isolation variables  $E_T^{iso}(\Delta R)$  and  $p_T^{iso}(\Delta R)$  [55].

### 5.3 Photons

Photons are reconstructed from clustered energy deposits in the ECAL that cannot be matched to a track in the ID. Unconverted photons are clusters that cannot be matched to an electron track or a photon-electron conversion vertex while converted photons are reconstructed from clusters that are matched to a conversion vertex where the photon was converted into electrons via pair production [55].

Similar to electrons, a likelihood-based identification is used for photons: Characteristic variables like the shower shape and the energy amount deposited in the calorimeters are used to define a discriminant on which different cuts can be placed to define the analysis WPs. For this analysis, photons are required to pass the Tight WP [55]. The photon isolation variables  $E_T^{iso}(\Delta R)$  and  $p_T^{iso}(\Delta R)$  are the same as for electron candidates. The isolation WP definitions are listed in Table 8, the WP used in the analysis is FCTight [55].

In addition to the identification and isolation requirements, photons in this analysis need to have a transverse momentum of  $p_T > 27 \text{ GeV}$ .

### 5.4 Jets

For the reconstruction of jets, the products of hadronization of quarks and gluons in the detector, both tracking information from the ID as well as the ECAL and HCAL are used. The reconstruction is done with the `AntiKt4EMPFLOW` algorithm, which relies on an iterative process to cluster energy deposits in the calorimeters into jets and combine those with tracking information [56].

To reduce the background from jets produced in non-hard scatter events the likelihood-based discriminant jet-vertex tagger (JVT) is defined. It combines information on the measured vertexes and the momentum of tracks into a single variable, on which three working points, Loose, Medium and Tight, are defined. In this analysis, jets are required to pass the Tight

working point [57]. Additionally, all jets need to have a transverse momentum of  $p_T > 25$  GeV.

B-jets, jets produced by a  $b$  quark, can be used in analysis to restrict or isolate processes involving a  $t$  quark. To distinguish b-jets from jets induced by other hadrons the deep learning algorithm DL1r is used. It relies on characteristic features of b-jets, like the existence of secondary vertices and higher impact parameters of the tracks to assign a probability that an observed jet is from a b-quark. The ATLAS standard is to place a cut on the DL1r output discriminant that corresponds to a 77% efficiency in b-tagging, which is adopted in this analysis [58].

## 5.5 Missing Transverse Energy

The missing transverse energy (MET) is used to indicate the presence of particles that cannot be detected in the ATLAS detector, especially neutrinos. It is calculated by adding the transverse momenta of all detected leptons, photons and jets and including a soft term to account for low-energy tracks recorded in the ID. Since the momentum in the transverse plane is preserved, an MET different from zero indicates the presence of neutrinos in an event [59].

## 5.6 Overlap Removal

The reconstruction processes described above can lead to the reconstruction of two different objects from the same tracks or energy deposits. In this case, only one of the reconstructed objects is kept, according to standard ATLAS overlap removal rules: If an electron and a muon share a track, the electron is removed. For an electron and a jet which overlap with a distance  $\Delta R < 0.2$  the electron is kept, while for  $0.2 < \Delta R < 0.4$  the electron is disregarded. If a muon overlaps with a reconstructed jet with  $\Delta R < 0.4$ , the muon is kept if there are two or less tracks reconstructed; for more reconstructed tracks the jet is kept. For the overlap of a lepton and a photon with  $\Delta R < 0.4$  the photon is removed, while for the overlap between a jet and a photon the photon is kept [60].

## 6 Event Selection

All events used for physics analysis must be collected in a data collection period that is a member of the GRL described in Section 4.2. In addition, all events that are considered to have unreliable calorimeter data are disregarded. The events must pass one of the unpre-scaled single lepton triggers listed in Section 4.1. To ensure the selection of hard-scatter events, at least one vertex reconstructed from two or more tracks with  $p_T > 500$  MeV is required for each event. Additionally, any event with unidentified jets, which for example could stem from cosmic rays or calorimeter malfunctions, is removed.

### 6.1 $H_5^\pm \rightarrow W^\pm \gamma$ Analysis Regions

The signal region aims to select events which are the result of the decay process of a charged Higgs boson:

$$H_5^\pm \rightarrow W^\pm \gamma \rightarrow \ell^\pm \nu_\ell \gamma \quad (6.1)$$

Cuts on the number of reconstructed objects are used to select these signal events. In the signal region, events are required to have exactly one charged lepton that passes the tight selection defined in Sections 5.1.4 and 5.2.4. This analysis only includes final states with electrons or muons; decays of the  $W^\pm$  boson into taus are not considered. The events need at least one photon, less than two jets and no  $b$ -jets.

For the definition of the signal region it is also important to minimize the background from other processes that produce a similar signature in the detector. This is done by applying cuts on kinematic variables: Events with a neutrino in the final state, like the signal process, have a higher MET than events without neutrinos as neutrinos cannot be detected in the ATLAS detector. Requiring events to have a minimum MET cuts events from background processes without a neutrino, for example events involving  $Z^0$  bosons which can leptonically decay into two leptons:

$$Z^0 \rightarrow \ell^+ \ell^- \quad (6.2)$$

Events involving a  $W^\pm$  boson and a jet,  $W^\pm + jets$  events, have a final state of  $\ell^\pm \nu_\ell + jet$  which, when the jet is misidentified as a photon, can appear as the signal signature. Leptons and photons from fake processes tend to have a lower  $p_T$  than real leptons and photons, so

Signal Region	
Object Cuts	$N_{muon} = 1$ or $N_{electron} = 1$ $N_{photon} \geq 1$ $N_{jets} < 2$ $N_{b-jets} = 0$
Kinematic Cuts	$MET > 50$ GeV $p_T^{\ell+\gamma} > 100$ GeV $p_T^{\ell+\gamma+MET} > 200$ GeV $ m_{\ell+\gamma} - m_Z  > 5$ GeV

Table 9: Summary of the object and kinematic cuts used to define the signal region in the  $H_5^\pm \rightarrow W^\pm\gamma$  analysis [1].

placing a lower bound on  $p_T^{\ell+\gamma}$ , the sum of the transverse momenta of the charged lepton and photon, reduces the number of fake backgrounds in the signal sample. Similarly, a cut on the combined transverse momenta of the lepton, the photon, and the MET,  $p_T^{\ell+\gamma+MET}$ , favors signal over background events as signal events are more likely to be reconstructed with higher transverse momenta for the final state particles. Finally, there is a background from electron-to-photon fakes, events where an electron is wrongly identified as a photon. This background stems mostly from  $Z^0 + jets$  events where the  $Z^0$  boson decays into two electrons. It can be cut out by requiring the invariant mass of the lepton and photon system  $m_{\ell+\gamma}$  to be different than the mass of the  $Z^0$  boson. The signal region selection including the object and kinematic cuts is summarized in Table 9.

The  $H_5^\pm \rightarrow W^\pm\gamma$  analysis is blinded, which means the signal region cannot be analyzed before the whole analysis framework is developed. This procedure is required for analysis in the ATLAS collaboration and ensures that the analysis is built without biases from expected results.

To test the analysis framework before application to the signal region, a validation region is defined. The validation region is kinematically close to the signal region, but dominated by the background processes and contains only up to 10% of signal events. Its definition is therefore very similar to the selection of the signal region, except for the cut on  $p_T^{\ell+\gamma+MET}$ , which is changed to be orthogonal to the signal region. The validation region selection is shown in Table 10.

The  $H_5^\pm \rightarrow W^\pm\gamma$  analysis also requires the definition of background-dominated control regions. They are defined by inverting cuts from the signal region that were placed in the signal selection to cut out specific background processes. The control region are orthogonal to the signal and validation regions and each other. The respective selections are listed in

Validation Region	
Object Cuts	$N_{muon} = 1$ or $N_{electron} = 1$ $N_{photon} \geq 1$ $N_{jets} < 1$ $N_{b-jets} = 0$
Kinematic Cuts	$MET > 50$ GeV $p_T^{\ell+\gamma} > 100$ GeV $100\text{GeV} < p_T^{\ell+\gamma+MET} < 200$ GeV $ m_{\ell+\gamma} - m_Z  > 5$ GeV

Table 10: Summary of the object and kinematic cuts used to define the validation region in the  $H_5^\pm \rightarrow W^\pm\gamma$  analysis [1].

Table 11.

## 6.2 Fake Efficiency Control Region

For the estimation of the fake lepton background for the  $H_5^\pm \rightarrow W^\pm\gamma$  analysis a control region enriched in fake/non-prompt leptons is required. The fake/non-prompt leptons are selected with a “tag-and-probe” method on  $Z^0 \rightarrow \ell^+\ell^-$  events: The control region selects events with three electrons or muons that pass the baseline selection criteria defined in Sections 5.1.4 and 5.2.4. Two of these leptons are being “tagged” to the leptonic decay of a  $Z^0$  boson by requiring them to have the same flavor and opposite signs. Their invariant mass has to be within  $\pm 10\%$  of the  $Z^0$  boson mass and they have to pass the tight lepton selections. The unpre-scaled single lepton trigger must be triggered and matched to one of the  $Z^0$  tag leptons. The third baseline lepton, the probe lepton, is required to have opposite flavor to the tag leptons to avoid any ambiguity of which leptons are the result of the  $Z^0$  boson decay. The probe lepton is likely to be a fake/ non-prompt lepton as there are very few processes that will result in this detector signature. The main background processes are events with a  $W^\pm$  and a  $Z^0$  boson which will also often have a neutrino from the  $W^\pm$  decay in the final state. Accordingly, the amount of  $W^\pm Z^0$  background in the fake lepton control region can be reduced by requiring the events to have a MET of  $< 40$  GeV.

$W^\pm\gamma$ Control Region	
Object Cuts	$N_{muon} = 1$ or $N_{electron} = 1$ $N_{photon} \geq 1$ $N_{jets} < 2$ $N_{b-jets} = 0$
Kinematic Cuts	$MET > 50$ GeV $p_T^{\ell+\gamma} > 100$ GeV $p_T^{\ell+\gamma+MET} < 200$ GeV $ m_{\ell+\gamma} - m_Z  > 5$ GeV
$Z^0\gamma$ Control Region	
Object Cuts	$N_{muon} = 2$ or $N_{electron} = 2$ opposite sign of the leptons $N_{photon} = 1$ $N_{b-jets} = 0$
Kinematic Cuts	$ m_{\ell+\ell} - m_Z  < 15$ GeV $p_T^\gamma > 50$ GeV
$W^\pm + jets$ Control Region	
Object Cuts	$N_{muon} = 1$ or $N_{electron} = 1$ $N_{photon} \geq 1$ $N_{jets} < 2$ $N_{b-jets} = 0$
Kinematic Cuts	$MET > 50$ GeV $p_T^{\ell+\gamma} < 100$ GeV $p_T^{\ell+\gamma+MET} < 200$ GeV $ m_{\ell+\gamma} - m_Z  > 5$ GeV
$Z^0 + jets$ Control Region	
Object Cuts	$N_{muon} = 1$ or $N_{electron} = 1$ $N_{photon} \geq 1$ $N_{jets} < 2$ $N_{b-jets} = 0$
Kinematic Cuts	$MET < 50$ GeV $p_T^{\ell+\gamma} > 100$ GeV $p_T^{\ell+\gamma+MET} < 200$ GeV $ m_{\ell+\gamma} - m_Z  > 5$ GeV

Table 11: Summary of the object and kinematic cuts used to define the different background-dominated control regions in the  $H_5^\pm \rightarrow W^\pm\gamma$  analysis [1].

## 7 $H_5^\pm \rightarrow W^\pm\gamma$ Backgrounds

Some SM processes can have similar detector signatures as the  $H_5^\pm \rightarrow W^\pm\gamma$  decay and pass the signal selection. They must be accounted for to obtain an accurate number of signal events in the signal region. All background processes considered are listed in Table 12. They can be divided into two categories; prompt backgrounds, where the lepton and photon in the final state are real particles, and fake/non-prompt backgrounds which are created by particle misreconstruction and misattribution [1].

Prompt backgrounds are described well in MC simulations which are used to account for their contribution to the analysis regions. In contrast, fake/non-prompt backgrounds generally do not need to be well modeled in simulations and must be estimated with data-driven methods. In the  $H_5^\pm \rightarrow W^\pm\gamma$  analysis there are three data-driven background estimations that describe electrons faking photons, jets faking photons and fake/non-prompt leptons [1].

### 7.1 Electron to Photon Fakes

To estimate the background from electrons misreconstructed as photons, events from the  $Z^0 + jets$  control region with either two electrons or one electron and one photon in the final state are selected. The invariant mass of the two particles is required to be close to the  $Z^0$  boson mass. Additionally, it is fitted with a Crystal Ball function to describe the leptonic decay of a  $Z^0$  boson and a polynomial function to include background. Only events that are part of the signal component are considered. These requirements ensure the selection of  $Z^0$  boson decays into two electrons and photons reconstructed in the final state are very likely to be misidentified electrons. With these events, a fake rate  $F^{e\rightarrow\gamma}$  is calculated by dividing the number of events with fake photons by the number of events with two electrons [1]:

$$F^{e\rightarrow\gamma} = \frac{N_{e\gamma}^{signal}}{N_{ee}^{signal}} \quad (7.1)$$

The fake rates measured in data and MC simulations are in good agreement indicating that electron to photon fakes are well simulated. Accordingly, this background contribution is estimated with MC simulations. To account for any small differences in the fake rates, the ratio of the fake rates measured in data and MC is used to re-weight all MC events [1].

## 7.2 Jet to Photon Fakes

Real photons and jets misreconstructed as photons can be differentiated with the isolation energy  $E^{iso}$  [1]:

$$E^{iso} = E_T(\Delta R < 0.4) - 0.022p_T \quad (7.2)$$

where  $E_T(\Delta R < 0.4)$  is the sum of transverse energies of topological clusters located within a cone with  $\Delta R < 0.4$  around the photon and  $p_T$  is the photon's transverse momentum. Real photons typically have low isolation energies due to high  $p_T$  and small energy deposits in close proximity, while fake photons from misidentified jets are characterized by large energy deposits resulting in large isolation energies [1]. Since real photons are well modeled in MC simulations, a function describing the isolation energy of real photons is obtained by fitting MC events in the  $W^\pm\gamma$  control region with a Bukin distribution [61]. A fake photon sample is selected from data in the  $W^\pm\gamma$  control region where photons are required to pass the Loose but fail the Tight identification WP which makes them very likely to be jets misidentified as photons. Analogous to real photons, a template for the fake photon isolation energy distribution is derived by fitting a Bukin distribution. The two templates for real and fake photons are combined into a single fit function, which is then applied to the isolation energy distribution in the analysis regions. The fake photon yield estimate is the result of this fit [1].

Prompt Backgrounds	
$W^\pm\gamma \rightarrow \ell^\pm\nu_\ell\gamma$	Same final state as the signal process
$Z^0\gamma \rightarrow \ell^\pm\bar{\ell}^\mp\gamma$	Leptons are not reconstructed or fall outside of the detector acceptance which appears as MET
$t\bar{t}\gamma \rightarrow \ell^\pm\nu_\ell\gamma + jets$	Jets are not reconstructed or fall outside of the detector acceptance which appears as MET
$Z^0Z^0\gamma \rightarrow \ell^\pm\bar{\ell}^\mp\ell'^\pm\bar{\ell}'^\mp\gamma$ $W^\pm W^\pm\gamma \rightarrow \ell^\pm\nu_\ell\ell'^\pm\nu_{\ell'}\gamma$ $W^\pm Z^0\gamma \rightarrow \ell^\pm\nu_\ell\ell'^\pm\bar{\ell}'^\mp\gamma$	Leptons are not reconstructed or fall outside of the detector acceptance which appears as MET
$Z^0\gamma\gamma \rightarrow \ell^\pm\bar{\ell}^\mp\gamma\gamma$ $W^\pm\gamma\gamma \rightarrow \ell^\pm\nu_\ell\gamma\gamma$	Photons and leptons are not reconstructed or fall outside of the detector acceptance which appears as MET
Fake/Non-prompt Backgrounds	
$Z^0 + jets$ $W^\pm + jets$	Jets are misidentified as leptons and photons; particles that are not reconstructed or fall outside the detector acceptance can appear as MET
$t\bar{t} \rightarrow \ell^\pm\nu_\ell\ell'^\pm\nu_{\ell'} + jets$ $t\bar{t} \rightarrow jets$	Jets are misidentified as leptons and photons
$\gamma\gamma$	Photons are misidentified as leptons
$\gamma + jets$	Jets are misidentified as leptons
$Z^0Z^0 \rightarrow \ell^\pm\bar{\ell}^\mp\ell'^\pm\bar{\ell}'^\mp$ $W^\pm W^\pm \rightarrow \ell^\pm\nu_\ell\ell'^\pm\nu_{\ell'}$ $W^\pm Z^0 \rightarrow \ell^\pm\nu_\ell\ell'^\pm\bar{\ell}'^\mp$ $W^\pm Z^0 \rightarrow \ell^\pm\nu_\ell\nu_{\ell'}\bar{\nu}_{\ell'}$	Electrons are misidentified as photons; photons are radiated from charged particles

Table 12: Summary of the prompt and non-prompt backgrounds for the  $H_5^\pm \rightarrow W^\pm\gamma$  analysis with explanations on how the processes can be misidentified as a signal event [1].

## 8 Fake/Non-Prompt Lepton Background for $H_5^\pm \rightarrow W^\pm\gamma$

The object reconstruction in ATLAS aims to select real leptons, particles created either directly in the proton-proton scattering or through the decay of a short-lived non-hadronic resonance, over fake and non-prompt leptons. Fake leptons are objects in the detector that are incorrectly reconstructed as leptons. In contrast, non-prompt leptons are correctly reconstructed, but originate from other SM processes like semi-leptonic decays of hadrons or photon conversions and are misattributed [3].

Fake/non-prompt leptons are difficult to simulate as they are created in various misidentification processes, some of which have non-perturbative theoretical descriptions so that simulations can not be expected to be accurate. Additionally, simulations would need to account for all possible particle interactions, details in the detector design and material responses, resulting in very complex calculations. Since the identification and isolation requirements during particle reconstruction are optimized to select real leptons, a simulation would also need to include a very large number of events to ensure that after object reconstruction the sample size still allows for statistical precision. The required large sample size makes fake/non-prompt lepton background simulations too resource-heavy to be practical. Therefore, the fake/non-prompt lepton background is estimated using a data-driven method which relies on measuring the rates at which fake/non-prompt leptons are selected in well-understood control regions and extrapolating the results into the analysis regions [3].

For ATLAS analyses standard methods for the estimation of the fake/non-prompt lepton background have been developed by the ATLAS Isolation and Fake Forum (IFF) and are detailed in [3, 62]. In the following, the fake factor method is described and applied to estimate the fake/non-prompt lepton background for  $H_5^\pm \rightarrow W^\pm\gamma$ .

### 8.1 Fake Factor Method

The fake factor method relies on two lepton selections, baseline and tight, which are described in Sections 5.1.4 and 5.2.4. Leptons that pass the baseline selection but not the tight selection are referred to as loose. The fraction of real leptons in the baseline sample that pass the tight selection is called the real efficiency  $\epsilon_r$ :

$$\epsilon_r = \frac{N_r^t}{N_r^b} \quad (8.1)$$

where  $N_r^b$  and  $N_r^t$  are the numbers of real leptons in the baseline and tight samples. Anal-

ogous, the fake efficiency  $\epsilon_f$  relates the numbers of fake/non-prompt leptons in the baseline sample  $N_f^b$  and tight sample  $N_f^t$ :

$$\epsilon_f = \frac{N_f^t}{N_f^b} \quad (8.2)$$

If the baseline and tight selections are chosen well, the real efficiency will be much larger than the fake efficiency. Real leptons and real efficiencies are well modeled in MC simulations and therefore the numbers of real lepton events in the baseline, tight and loose samples can be directly obtained from MC simulations for any given region. The fake efficiency and the number of fake/non-prompt lepton events in the different samples need to be measured in data [3].

The number of fake/non-prompt leptons in the baseline sample  $N_f^b$  is composed of fake/non-prompt leptons in the tight sample  $N_f^t$  and in the loose sample  $N_f^l$ :

$$N_f^b = N_f^t + N_f^l \quad (8.3)$$

With the fake efficiency the number of fake/non-prompt loose leptons can be expressed as:

$$N_f^l = (1 - \epsilon_f)N_f^b \quad (8.4)$$

The loose sample  $N^l$  consists of real and fake/non-prompt leptons:

$$N^l = N_r^l + N_f^l \quad (8.5)$$

Combining Equations 8.4 and 8.5 yields an expression for the number of fake/non-prompt leptons in the baseline sample:

$$N_f^b = \frac{1}{(1 - \epsilon_f)}(N^l - N_r^l) \quad (8.6)$$

With the fake efficiency this can be transformed into the number of fake/non-prompt leptons in the tight sample:

$$N_f^t = \frac{\epsilon_f}{(1 - \epsilon_f)}(N^l - N_r^l) = F(N^l - N_r^l) \quad (8.7)$$

Here, the fake factor  $F$  is defined as

$$F = \frac{\epsilon_f}{(1 - \epsilon_f)} \quad (8.8)$$

Therefore, the fake factor method allows to estimate the number of fake/non-prompt leptons in an analysis region from the fake factor, the measured number of events that pass the loose selection and the number of real lepton events in the loose sample which can be obtained from MC simulations [3]. The advantage of the fake factor method over other fake/non-prompt lepton estimation techniques is that it does not depend on the yield of events in the tight selection and is therefore blind to the contents of the signal region [3].

## 8.2 Determination of the Fake Efficiency and Fake Factor

The fake efficiency is measured in the fake-enriched control region described in Section 6.2. This region was chosen for the  $H_5^\pm \rightarrow W^\pm\gamma$  analysis as it is kinematically close to the signal region and has been used by other ATLAS analysis for the fake/non-prompt lepton background estimation [63, 64]. The probe lepton in the fake efficiency control region is very likely to be fake and is used to calculate the fake efficiency according to:

$$\epsilon_f = \frac{N^t}{N^b} \quad (8.9)$$

where  $N^t$  and  $N^b$  are the numbers of events where the probe lepton passes the tight and baseline selections respectively. The fake efficiency control region has a background from real  $W^\pm Z^0$  events which is minimized in the construction of the region by introducing a MET cut of  $< 40$  GeV. The remaining  $W^\pm Z^0$  background is modeled with MC simulations and the truth-matched MC events are subtracted from the number of measured events. This transforms Equation 8.9 into:

$$\epsilon_f = \frac{N^t - N_{W^\pm Z^0, MC}^t}{N^b - N_{W^\pm Z^0, MC}^b} \quad (8.10)$$

The fake efficiency depends on the lepton kinematics and the IFF recommends binning in the lepton's transverse momentum  $p_T$  and pseudorapidity  $\eta$  [3]. The bin edges are chosen to create similar statistics for  $N^t - N_{W^\pm Z^0, MC}^t$  in each bin and are listed in Table 13. Although all leptons in the analysis are required to have  $p_T > 27$  GeV, probe leptons with lower  $p_T$  are included in the fake efficiency calculation in separate bins. While these are not needed for the

	Electron Channel	Muon Channel
$p_T [GeV]$	[10,15,20,27,35,50,1000]	[10,15,27,35,1000]
$ \eta $	[0,1.37],[1.52,2.5]	[0,2.5]

Table 13: Bins edges in  $p_T$  and  $\eta$  used for the fake efficiency calculation.

final fake/non-prompt lepton background estimate, they allow for a better comprehensive understanding of the behavior of the fake efficiency with varying  $p_T$ . For the electron channel, statistics allow for 3 bins in the  $p_T$  dimension above 27 GeV, while for the muon channel only two bins are created. For the same reason the  $\eta$  dimension in the electron channel is divided into two bins according to the different calorimeters used for particle reconstruction while in the muon channel a single  $\eta$  bin is used.

Even with the smaller number of bins in the muon channel the number of events in each bin is very low and  $N^t - N_{W^\pm Z^0, MC}^t$  is only  $\sim 60$  events per bin. This results in high relative statistical uncertainties. Particularly in the high  $p_T$  bin the background from real  $W^\pm Z^0$  events increases significantly compared to the number of measured events. This is a result of fake/non-prompt muons being predominantly created in the decay of heavy flavored hadrons and therefore they are very unlikely to have a high  $p_T$ . In comparison, fake/non-prompt electrons are also created through the misreconstruction of light flavored jets and from photon conversions, both of which can result in particles with high  $p_T$  [3].

The fake efficiency calculated with data and MC simulation is shown in Figures 8 and 9 for the electron channel and in Figures 10 and 11 for the muon channel. To allow for a better comparison between the results obtained with data and MC simulation, Figures 12 and 13 show projections of the calculated fake efficiencies into the two binning dimensions. Only  $p_T$  bins with  $p_T > 27$  GeV which are used for the fake/non-prompt background calculation are included in the projections. The diagrams show that for fake/non-prompt electrons MC simulations do not model the fake efficiency well and underestimate the measured value. This disagreement is the reason a data-driven method is necessary for the fake/non-prompt lepton background. In contrast, the behavior of fake muons is generally well described in the MC simulation, with the exception of the low  $p_T$  bin.

The uncertainties on the fake efficiency are statistical uncertainties where lower event numbers result in higher relative uncertainties. Accordingly in the muon channel the statistical uncertainties are very large and the small uncertainties on MC stem from higher statistics in the MC simulation than in data.

The fake factors (Equation 8.8) resulting from data and MC simulation are shown in Figures 14 and 15 for the electron channel and Figures 16 and 17 for the muon channel. Analogous

to the fake efficiency, Figures 18 and 19 show projections of the fake factor into the binning dimensions. The statistical uncertainties on the fake efficiency are analytically propagated to the fake factor.

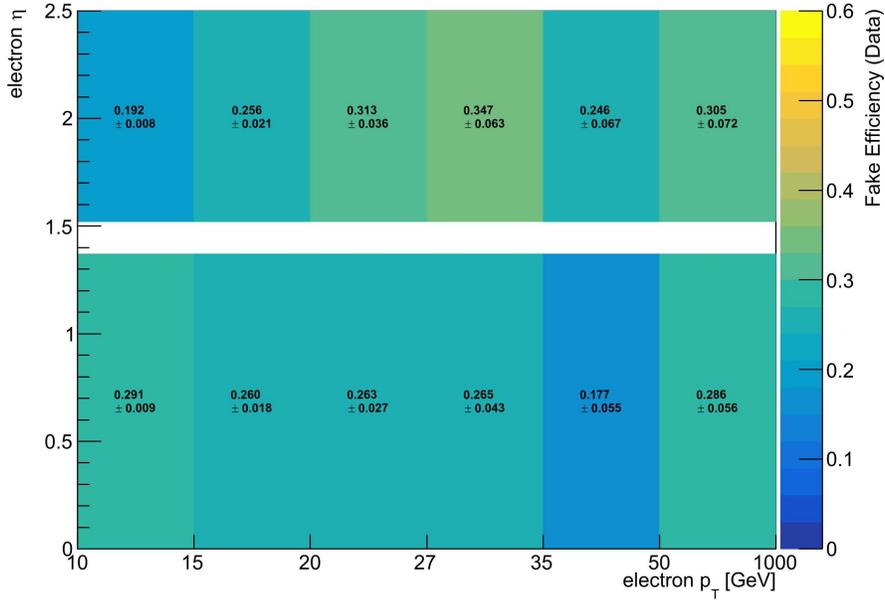


Figure 8: Fake efficiency estimates in the electron channel calculated with data. Only statistical uncertainties are shown.

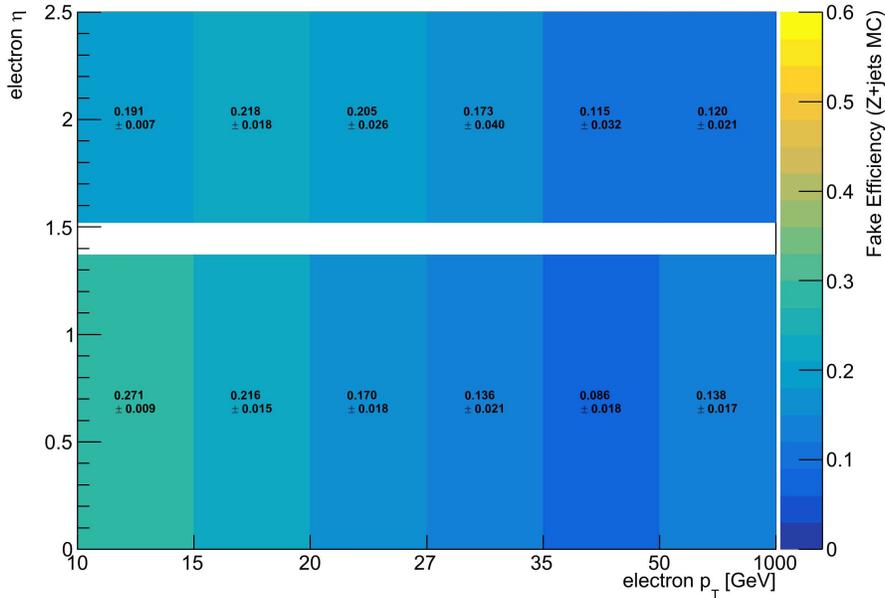


Figure 9: Fake efficiency estimates in the electron channel calculated with  $Z + jets$  MC. Only statistical uncertainties are shown.

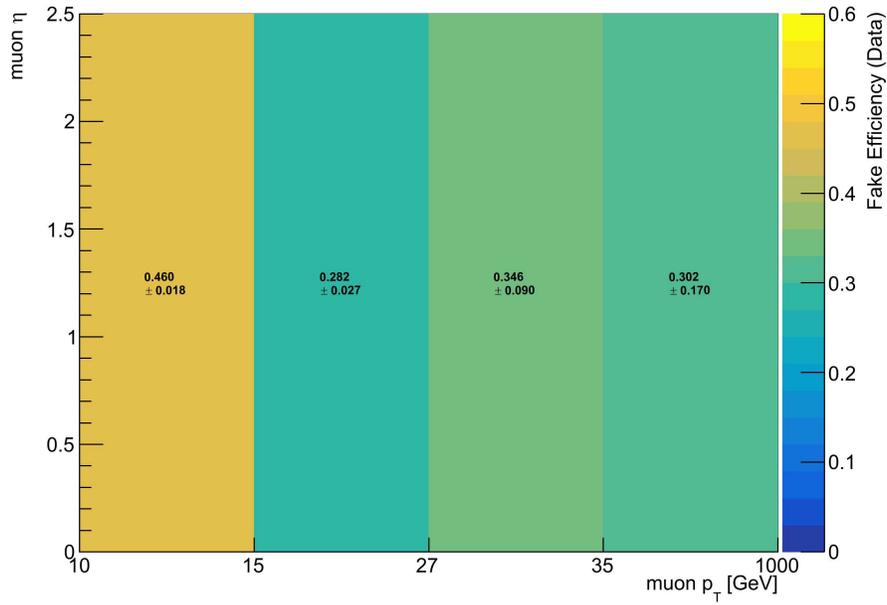


Figure 10: Fake efficiency estimates in the muon channel calculated with data. Only statistical uncertainties are shown.

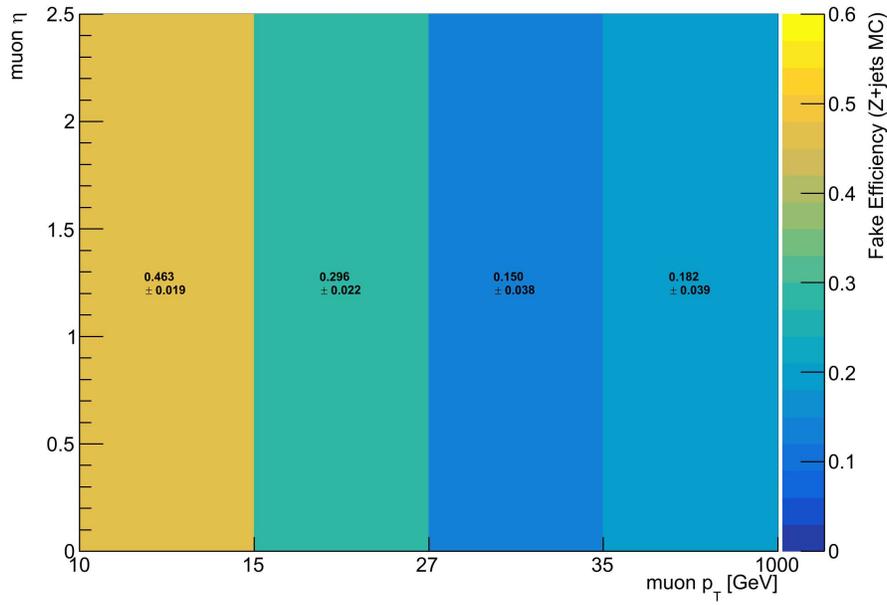


Figure 11: Fake efficiency estimates in the muon channel calculated with  $Z + jets$  MC. Only statistical uncertainties are shown.

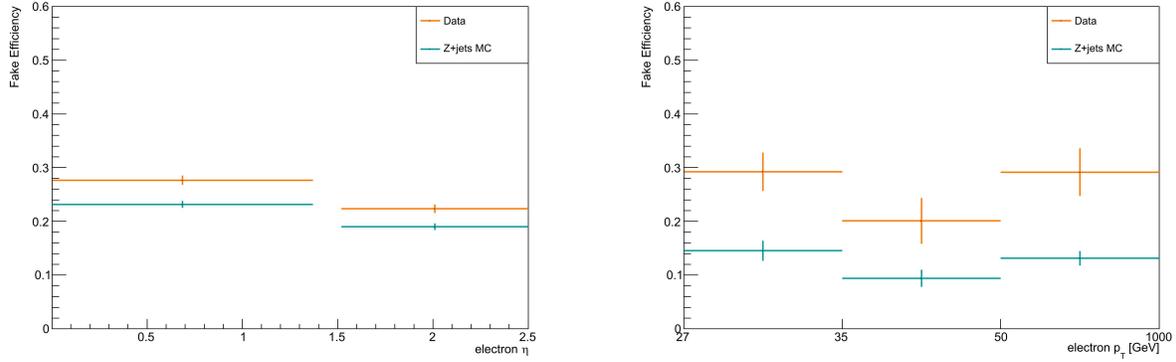


Figure 12: Fake efficiency estimates in the electron channel for data and  $Z + jets$  MC projected into the  $p_T$  and  $\eta$  dimensions. Only bins with  $p_T > 27$  GeV are included. Only statistical uncertainties are shown.

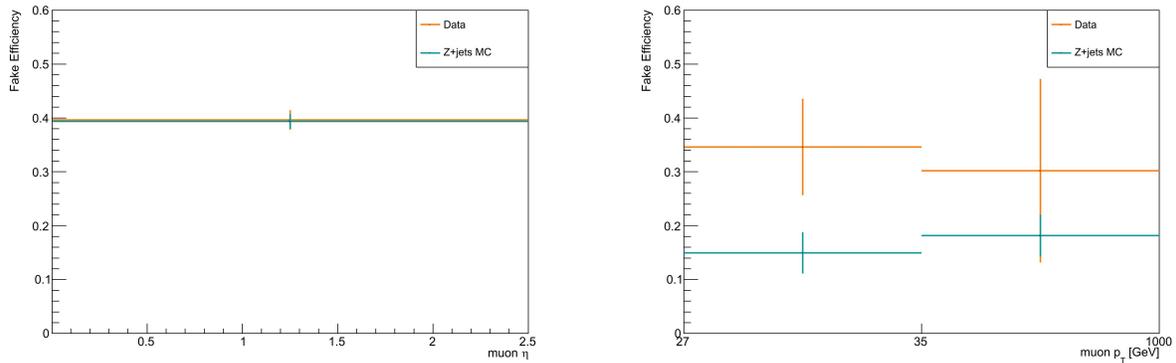


Figure 13: Fake efficiency estimates in the muon channel for data and  $Z + jets$  MC projected into the  $p_T$  and  $\eta$  dimensions. Only bins with  $p_T > 27$  GeV are included. Only statistical uncertainties are shown.

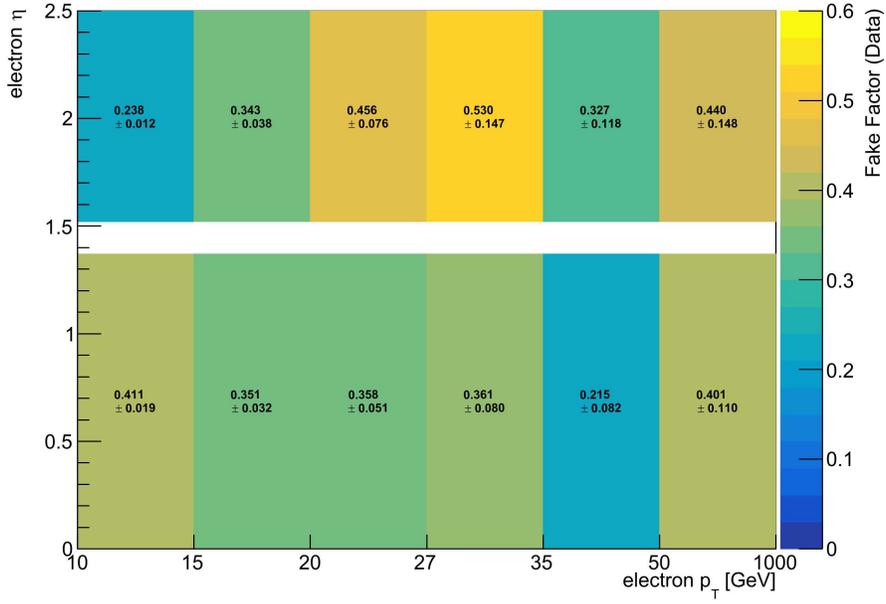


Figure 14: Fake factor estimates in the electron channel calculated with data. Only statistical uncertainties are shown.

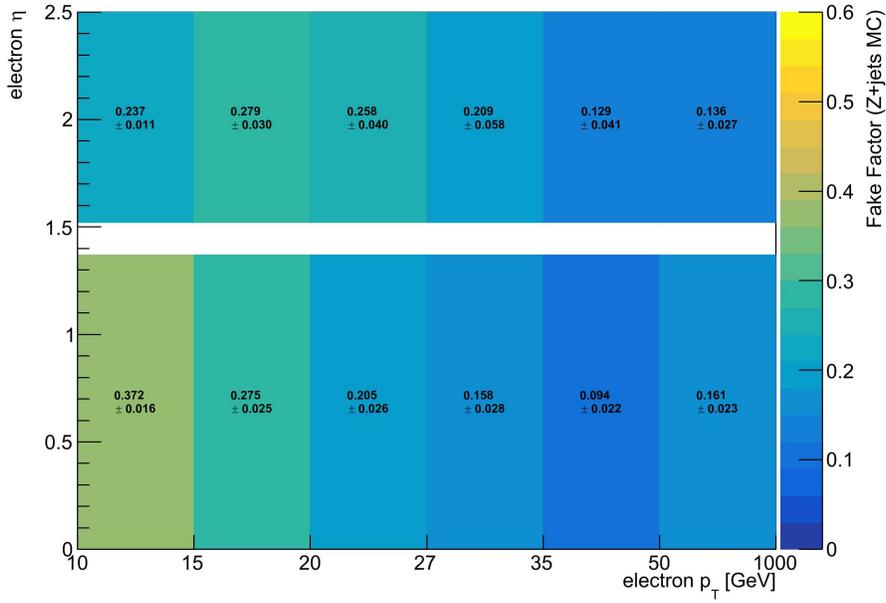


Figure 15: Fake factor estimates in the electron channel calculated with  $Z + jets$  MC. Only statistical uncertainties are shown.

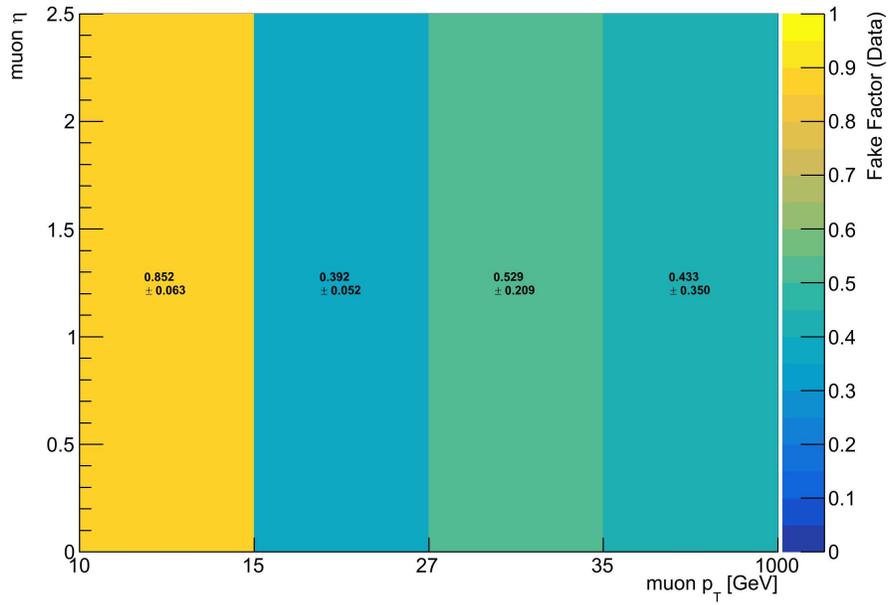


Figure 16: Fake factor estimates in the muon channel calculated with data. Only statistical uncertainties are shown.

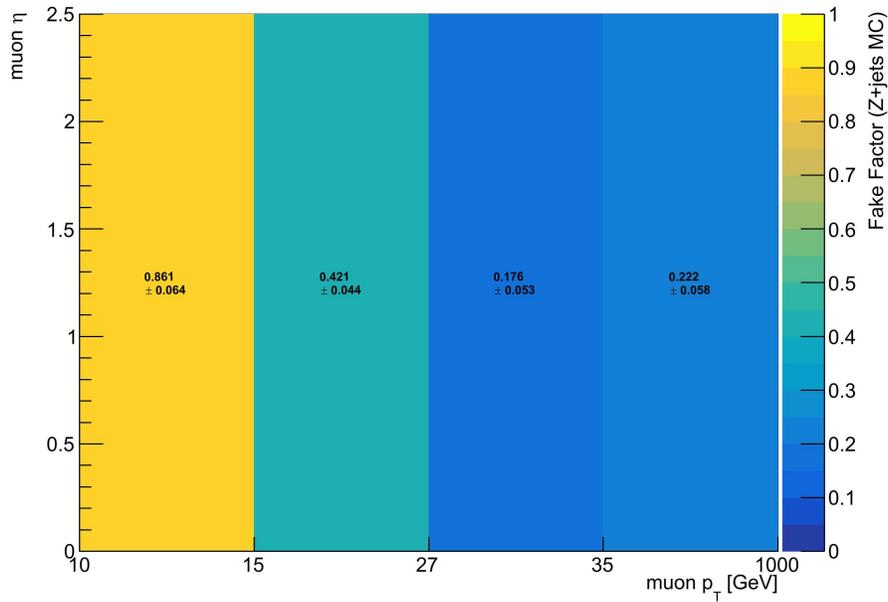


Figure 17: Fake factor estimates in the muon channel calculated with  $Z + jets$  MC. Only statistical uncertainties are shown.

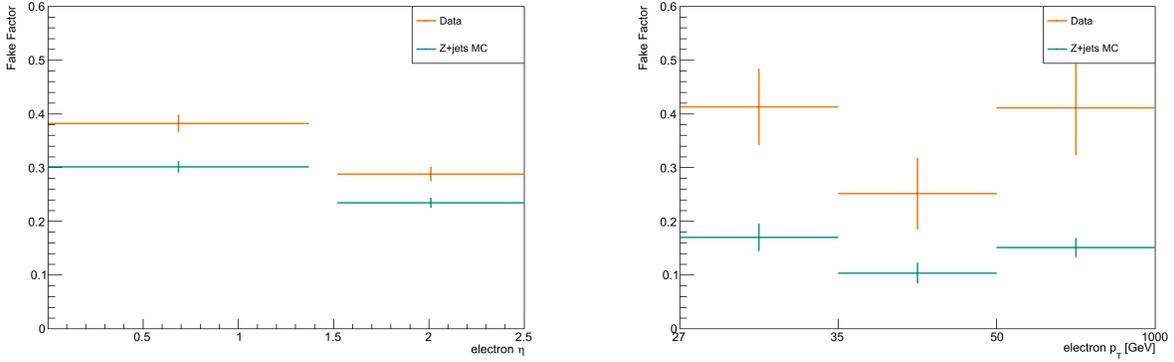


Figure 18: Fake factor estimates in the electron channel for data and  $Z + jets$  MC projected into the  $p_T$  and  $\eta$  dimensions. Only bins with  $p_T > 27$  GeV are included. Only statistical uncertainties are shown.

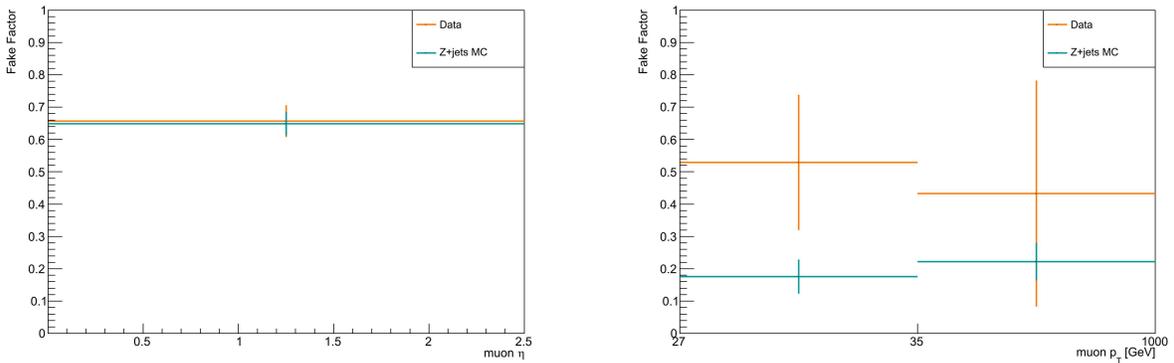


Figure 19: Fake factor estimates in the muon channel for data and  $Z + jets$  MC projected into the  $p_T$  and  $\eta$  dimensions. Only bins with  $p_T > 27$  GeV are included. Only statistical uncertainties are shown.

	Electron Fakes	Muon Fakes
Validation Region	$38 \pm 5$	$10 \pm 5$
Signal Region	$27 \pm 5$	$3 \pm 2$
$W\gamma$ Control Region	$118 \pm 11$	$37 \pm 11$
$Z\gamma$ Control Region	$-1 \pm 3$	$4 \pm 3$
$W + jets$ Control Region	$172 \pm 8$	$133 \pm 13$
$Z + jets$ Control Region	$1699 \pm 31$	$318 \pm 30$

Table 14: Estimated numbers of fake/non-prompt lepton events in the analysis regions. Only statistical uncertainties are shown.

### 8.3 Fake/Non-Prompt Lepton Background Estimates in the Analysis Regions

The number of fake/non-prompt lepton events in the analysis regions  $N_f^l$  can be calculated with Equation 8.7. The number of loose events  $N^l$  is measured in data while the expected number of real lepton events in the loose sample  $N_r^l$  is estimated from MC simulations. The fake factor is applied as a function of  $p_T$  and  $\eta$  on an event-by-event basis so that Equation 8.7 becomes

$$N_f^l = \sum_{Data, i=1}^{N_{Data}^l} F_i - \sum_{MC, j=1}^{N_{MC}^l} \omega_j F_j \quad (8.11)$$

where  $F_i$  and  $F_j$  are the fake factors appropriate for the respective leptons and  $\omega_j$  is the MC event weight. MC samples are typically created with at least one order of magnitude more events than data. This allows to minimize statistical fluctuations and event weights are then used to scale the simulation to the measured luminosity.

Table 14 shows the fake/non-prompt lepton yield estimates for the analysis regions defined in Section 6.1. With the exception of the  $Z^0\gamma$  control region, in all regions more fake/non-prompt electrons are expected than fake/non-prompt muons. In the analysis muon candidates are required to have recorded hits in the MS to be identified as muons. Since the ECAL and HCAL sufficiently stop other particles before reaching the MS, this provides a unique identification criterion and misidentifications are unlikely to occur. Simulations of particles in the ATLAS detector confirm that the dominant contribution to the fake/non-prompt lepton background in the muon channel are non-prompt muons from heavy flavored hadron decays [3]. Electrons on the other hand are reconstructed from information from the ID and ECAL, detector subsystems where other particles will also be detected which allows for possible misidentifications. In the electron channel, the fake/non-prompt lepton

	$m_{W\gamma}[GeV]$ Bin Edges
Validation Region	[0, 80, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 310, 340, 370, 400, 440, 500, 600]
Signal Region	[0, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 235, 250, 270, 290, 310, 340, 370, 400, 450, 500]
$W\gamma$ Control Region	[0, 80, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 380, 400, 420, 440, 470, 500, 530, 570, 610, 650, 750]
$Z\gamma$ Control Region	[0, 80, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 320, 340, 360, 390, 420, 460, 500, 600]
$W + jets$ Control Region	[0, 80, 100, 120, 140, 160, 180, 200, 240, 300, 400]
$Z + jets$ Control Region	[0, 80, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300, 330, 360, 400, 500]

Table 15: Bin edges in  $m_{W^\pm\gamma}$  in the analysis regions used for  $H_5^\pm \rightarrow W^\pm\gamma$ .

background has multiple significant contributions, fakes from misidentified light flavored jets, non-prompt electrons from heavy flavored jets and non-prompt electrons produced in photon conversions. The decay of light flavored hadrons can also produce non-prompt electrons, but the relative contribution is negligible compared to the other sources [3]. The multiple processes to produce fake/non-prompt electrons compared to muons can explain the higher number of fake/non-prompt electrons.

Generally, the control regions including jets have higher numbers of fake/non-prompt leptons than the corresponding control regions with photons. This is expected, as the sources for fake/non-prompt leptons are mainly jets which are either misidentified as electrons or produce non-prompt muons in hadronic decays.

The  $Z^0\gamma$  control region has a noticeably small number of fake/non-prompt leptons. The control region selection is optimized for the leptonic  $Z^0$  boson decay and two final state leptons are tagged very precisely to this decay by requiring them to have opposite sign and an invariant mass close to the  $Z^0$  boson mass. Fake/non-prompt leptons are almost completely cut out. The negative estimate in the electron channel is a statistical artifact produced by negative weights of MC events which are subtracted from a small number of measured events in the loose sample.

The  $H_5^\pm \rightarrow W^\pm\gamma$  analysis investigates the distribution of the invariant mass of the  $W^\pm\gamma$  system where a difference between data and the expected background would indicate a BSM charged Higgs boson. To include the fake/non-prompt lepton background in the analysis, the number of fake/non-prompt lepton events in each  $m_{W^\pm\gamma}$  bin used in the analysis needs

to be determined. The  $m_{W\pm\gamma}$  bins for the different analysis regions are listed in Table 15. Figures 20 - 31 show the fake/non-prompt lepton contribution for each bin with statistical uncertainties. Some bins have negative estimates arising from statistical effects where the real background is larger than the number of measured events in the loose sample. For incorporation into the analysis, the fake/non-prompt lepton yields for these bins will be set to zero as negative event numbers are unphysical. In cases where the statistical and systematic uncertainties make the results consistent with zero, this is not required.

During the development of the analysis the signal region is blinded and the validation region is used to check the modeling of background processes. In the validation region (Figures 20 and 21) the fake/non-prompt lepton background makes up less than 1% of the region's composition and is very small compared to other background processes [1]. Without the fake/non-prompt lepton background each bin in the validation region shows agreement between the background estimations and measured events due to relatively large uncertainties on the background components. Adding the fake/non-prompt lepton background in each analysis bin does not significantly change this validation [1]. The fake factor method used to estimate the fake/non-prompt lepton background is therefore validated.

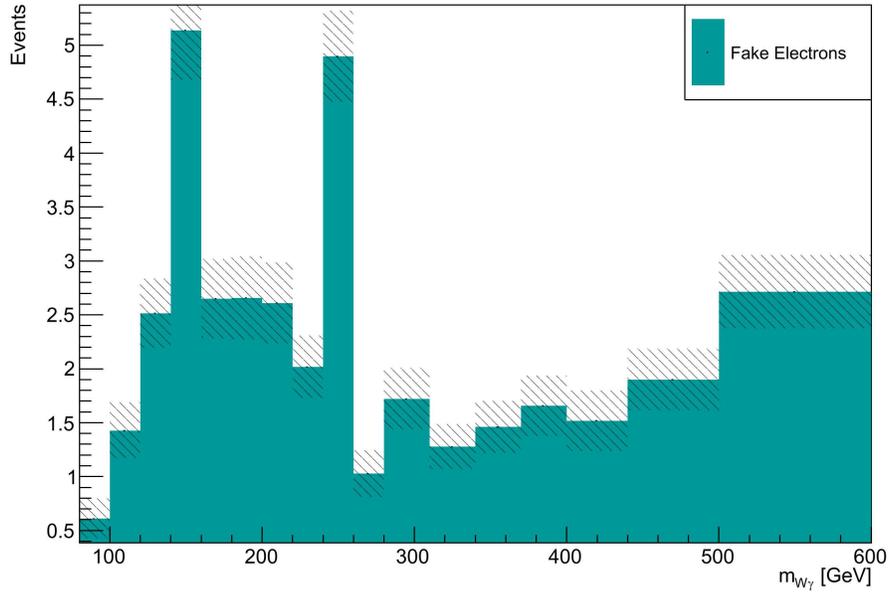


Figure 20: Fake/non-prompt electron yield estimates in the validation region. Only statistical uncertainties are shown.

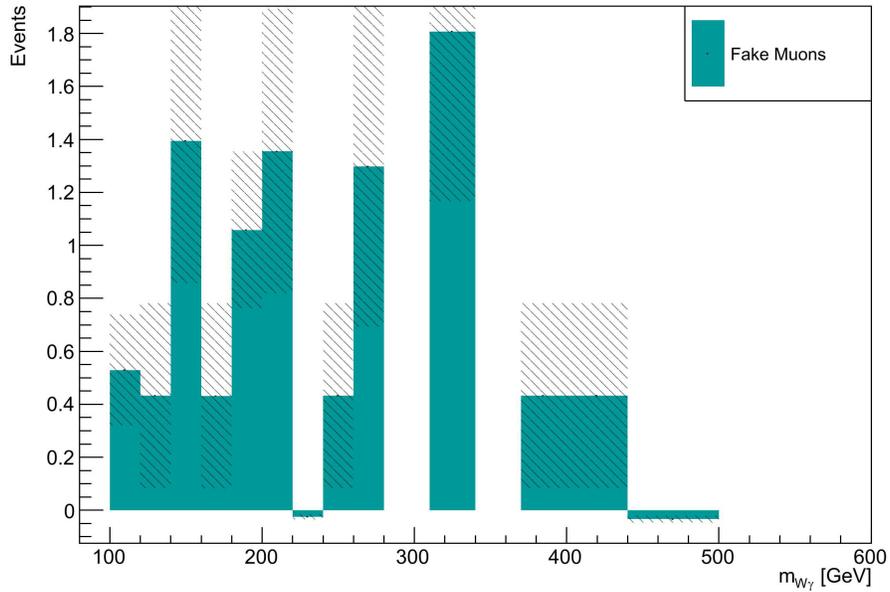


Figure 21: Fake/non-prompt muon yield estimates in the validation region. Only statistical uncertainties are shown.

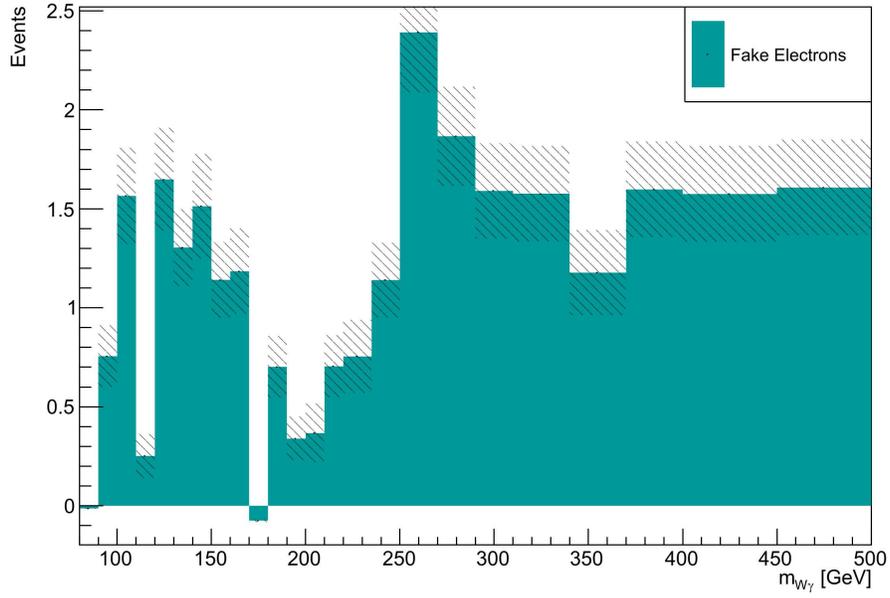


Figure 22: Fake/non-prompt electron yield estimates in the signal region. Only statistical uncertainties are shown.

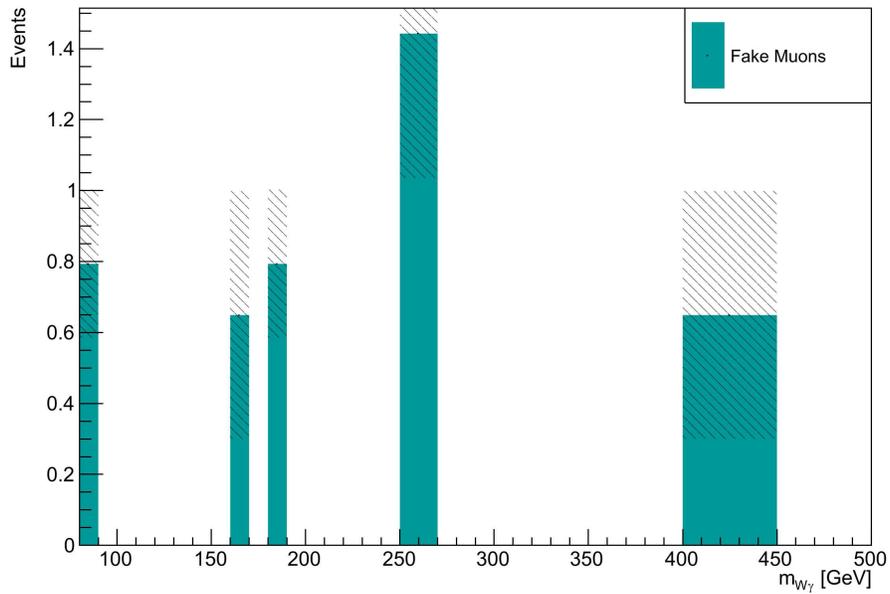


Figure 23: Fake/non-prompt muon yield estimates in the signal region. Only statistical uncertainties are shown.

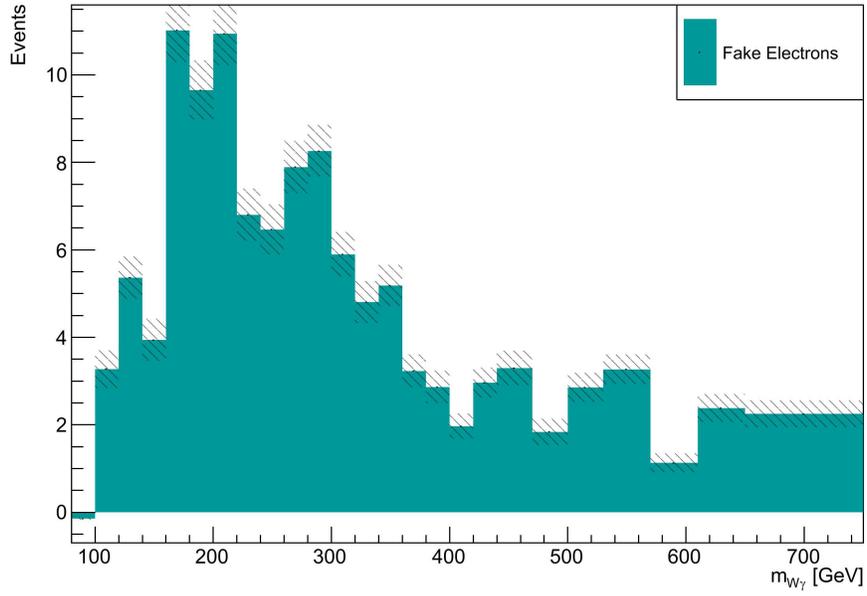


Figure 24: Fake/non-prompt electron yield estimates in the  $W^\pm\gamma$  control region. Only statistical uncertainties are shown.

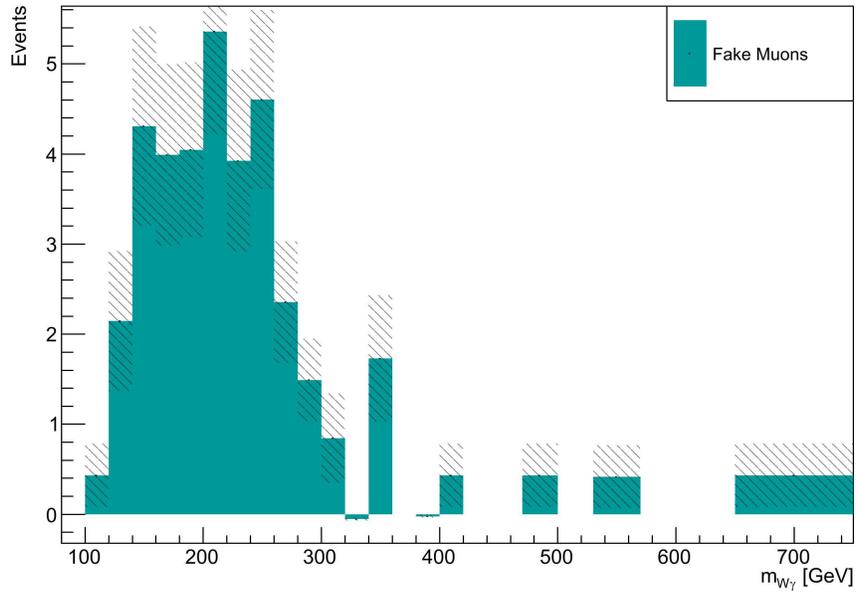


Figure 25: Fake/non-prompt muon yield estimates in the  $W^\pm\gamma$  control region. Only statistical uncertainties are shown.

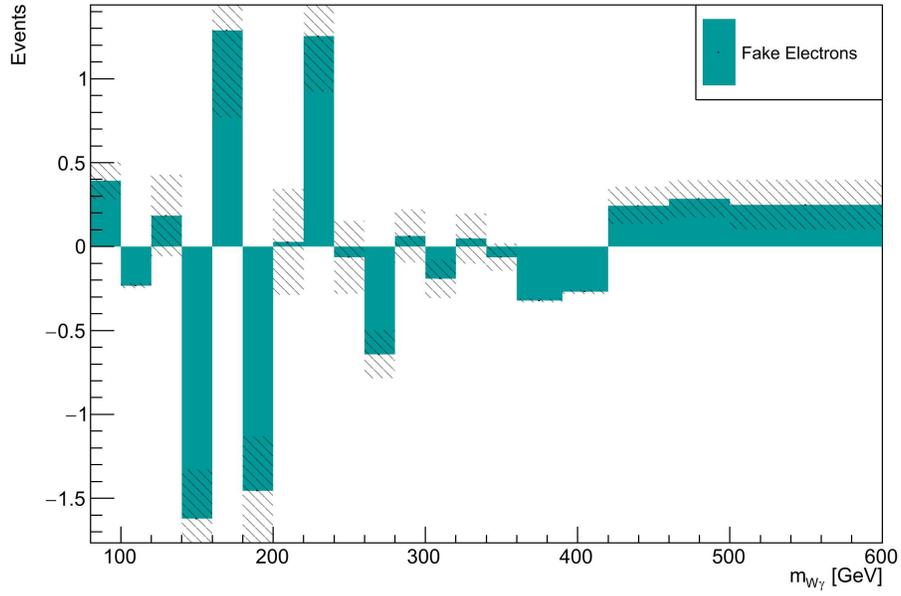


Figure 26: Fake/non-prompt yield electron estimates in the  $Z^0\gamma$  control region. Only statistical uncertainties are shown.

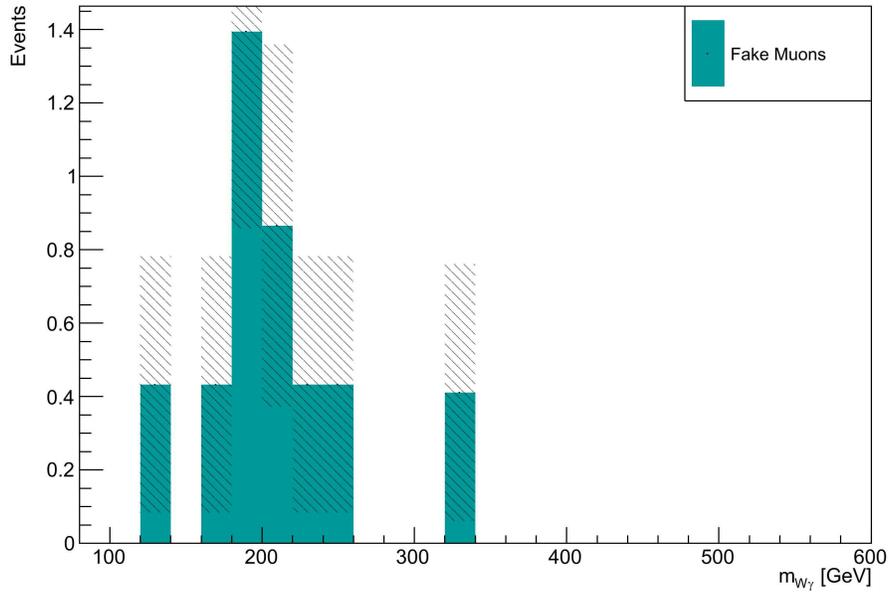


Figure 27: Fake/non-prompt muon yield estimates in the  $Z^0\gamma$  control region. Only statistical uncertainties are shown.

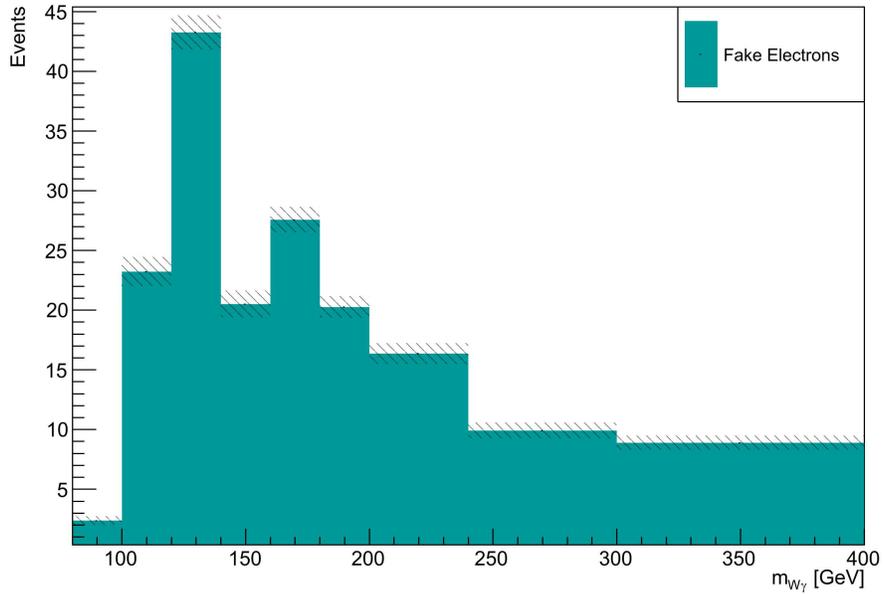


Figure 28: Fake/non-prompt electron yield estimates in the  $W^\pm + jets$  control region. Only statistical uncertainties are shown.

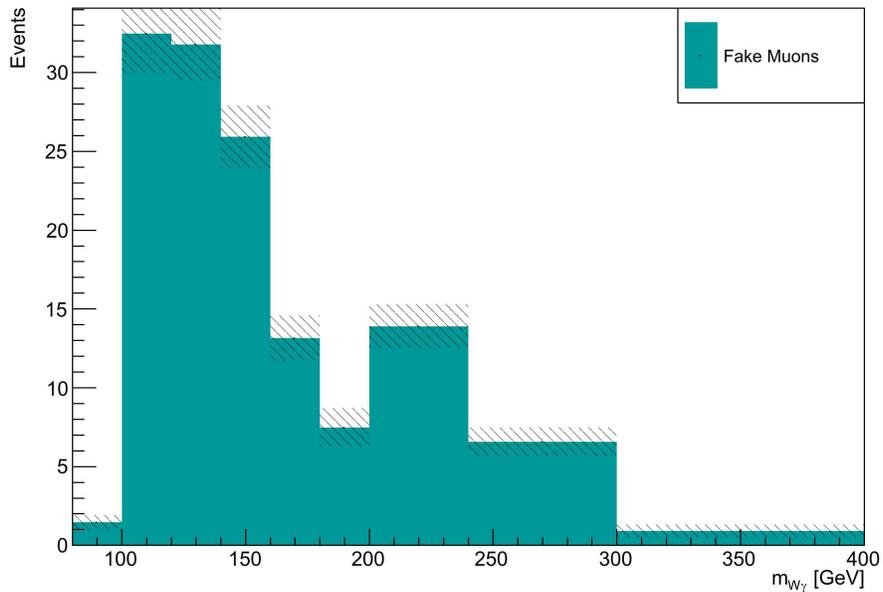


Figure 29: Fake/non-prompt muon yield estimates in the  $W^\pm + jets$  control region. Only statistical uncertainties are shown.

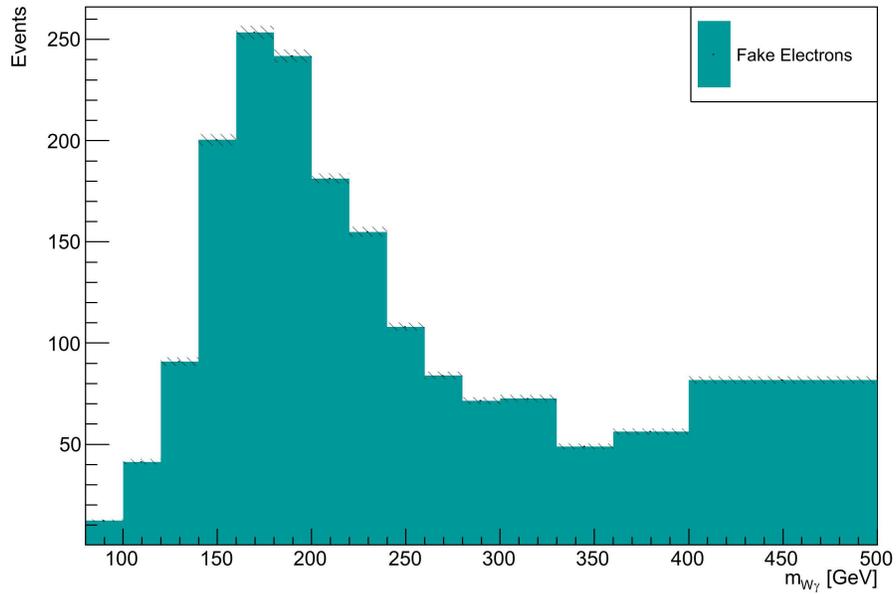


Figure 30: Fake/non-prompt electron yield estimates in the  $Z^0 + jets$  control region. Only statistical uncertainties are shown.

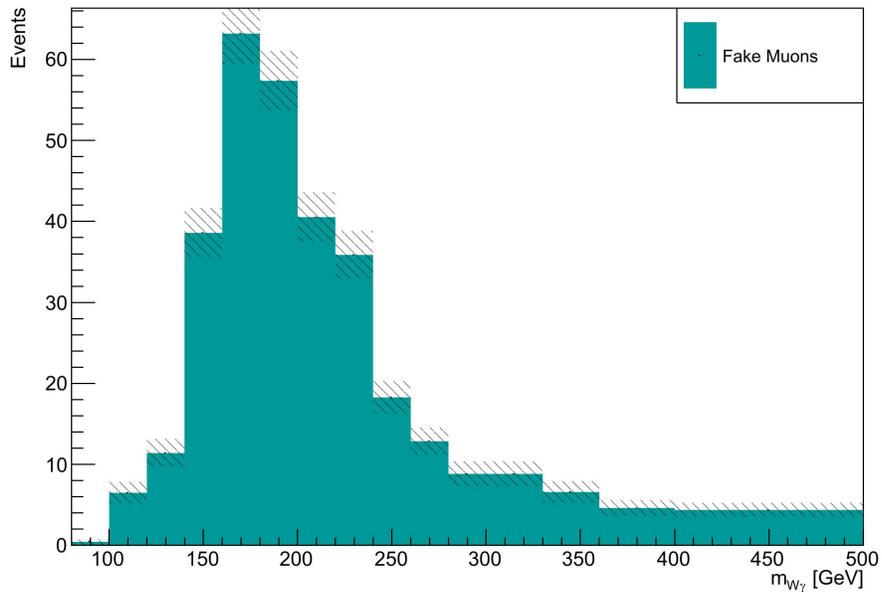


Figure 31: Fake/non-prompt muon yield estimates in the  $Z^0 + jets$  control region. Only statistical uncertainties are shown.

	Validation Region	Signal Region
<b>Nominal</b>	38	27
MET Up	-4	-2
MET Down	$\pm 0$	$\pm 0$
$ d_0 /\sigma(d_0)$ Down	$\pm 0$	$\pm 0$
Theory Up	+19	+13
Theory Down	-19	-14
Lepton Composition	$\pm 0$	$\pm 0$
<b>Systematic</b>	+19 -19	+13 -14
<b>Statistical</b>	$\pm 5$	$\pm 5$

Table 16: Statistical and systematic uncertainties on the estimated numbers of fake/non-prompt electron events.

	Validation Region	Signal Region
<b>Nominal</b>	10	3
MET Up	-1	$\pm 0$
MET Down	+1	$\pm 0$
$ d_0 /\sigma(d_0)$ Down	$\pm 0$	$\pm 0$
Theory Up	+4	+1
Theory Down	-5	-2
Lepton Composition	$\pm 0$	$\pm 0$
<b>Systematic</b>	+4 -5	+1 -2
<b>Statistical</b>	$\pm 5$	$\pm 2$

Table 17: Statistical and systematic uncertainties on the estimated numbers of fake/non-prompt muon events.

## 8.4 Statistical and Systematic Uncertainties

The fake/non-prompt lepton yield estimates have associated statistical and systematic uncertainties. Based on the recommendations by the IFF four different sources of systematic uncertainties are evaluated in the validation region and signal region [3, 62]. The quadratic sum of these individual systematics yields the overall systematic uncertainty. All considered uncertainties are summarized in Tables 16 and 17 and are described in detail in the following.

### 8.4.1 Statistical Uncertainties

The fake efficiencies have uncertainties accounting for statistical fluctuations between the bins in which they are calculated. As recommended by the IFF, these uncertainties are analytically propagated to the fake factors and the fake/non-prompt lepton yields [3]. The statistical uncertainties for all analysis regions are shown in Table 14. In the electron channel, high statistics in the fake efficiency control region result in relative statistical uncertainties of  $\pm 13\%$  in the validation region and  $\pm 19\%$  in the signal region. As can be seen in Table 16, these statistical uncertainties are small compared to the systematic uncertainty. In contrast, in the muon channel low number of events in the control region lead to high statistical uncertainties of  $\pm 50\%$  in the validation region and  $\pm 66\%$  in the signal region. Accordingly, in the muon channel statistical uncertainties are dominant over systematic uncertainties (Table 17).

### 8.4.2 MET Bias

In the definition of the fake efficiency control region (Section 6.2) a MET cut was introduced to reduce the background from real  $W^\pm Z^0$  events. The cut value of  $< 40$  GeV is an arbitrary choice and therefore its effect on the fake/non-prompt lepton yields needs to be taken into account. This is done by changing the MET cut by  $\pm 10$  GeV and calculating the fake efficiency, fake factor and fake/non-prompt lepton yields as described above with this new control region. All other definitions for the fake efficiency control region, the analysis regions and bins remain unchanged. The results in the validation and signal regions are shown in Tables 16 and 17. The fake/non-prompt lepton yields are not affected by reducing the MET cut to 30 GeV while a higher MET cut of 50 GeV results in a lower number of fake/non-prompt lepton yields. The changes in the electron channel correspond to  $-11\%$  of the nominal value in the validation region and  $-7\%$  in the signal region. In the muon channel only the estimate in the validation region is reduced by  $-10\%$  while in the signal region due to the low statistics no changes are observed.

### 8.4.3 $|d_0|/\sigma(d_0)$ Bias

The definition of the baseline selection (Sections 5.1.4 and 5.2.4) has an influence on the measured fake efficiency. Particularly the significance of the transverse impact parameter  $|d_0|/\sigma(d_0)$  can change the fake/non-prompt lepton composition, as a higher cut value discriminates against non-prompt leptons from heavy flavored jets, whereas misreconstructed

leptons from light flavored jets are less affected. For the baseline selection  $|d_0|/\sigma(d_0) < 5$  is required for electrons and  $|d_0|/\sigma(d_0) < 3$  for muons, which are the same  $|d_0|/\sigma(d_0)$  cuts as for the tight lepton selection. To evaluate the effect of this choice on the fake/non-prompt lepton yield estimates the  $|d_0|/\sigma(d_0)$  cut is removed. The fake efficiency, fake factor and fake/non-prompt lepton yields are then calculated with all other definitions unchanged. This results in the same number of fake/non-prompt leptons and therefore no systematic uncertainty needs to be assigned.

#### 8.4.4 Theory Uncertainties from MC Simulations

In the fake efficiency control region the background from real  $W^\pm Z^0$  events is modeled by MC simulations and uncertainties on the  $W^\pm Z^0$  background will influence the fake efficiency, especially in the muon channel, where the control region includes more background  $W^\pm Z^0$  events than  $Z^0 + fake$  events. The MC simulation relies on several physics variables, for example cross sections, parton distribution functions, factorization scales and renormalization values [3]. These parameters have associated uncertainties which will propagate to the  $W^\pm Z^0$  background estimation.

There are ATLAS-internal methods to extract systematic uncertainties on the fake efficiencies, fake factors and fake/non-prompt lepton background estimates from the uncertainties on the physics parameters used in simulated backgrounds. However, for analyses where the fake/non-prompt lepton background is only a very small contribution in the signal region and a full calculation of its systematics from real lepton backgrounds is out of proportion, the IFF recommends to simplify the process by assigning a symmetric uncertainty that safely includes all associated systematics while still being relatively small compared to other backgrounds and their uncertainties in the signal region [62].

In particular, the IFF refers to the ATLAS analysis searching for strongly produced supersymmetric particles using final states with multiple jets and two or three leptons [65]. This analysis assigns a systematic uncertainty of  $\pm 50\%$  for both fake/non-prompt electrons and muons in regions where the fake efficiency calculated in data varies significantly from the one measured in MC simulations as is the case in fake efficiency control region. This is in agreement with the results of the ATLAS analysis investigating the invariant mass spectrum of final states with four leptons [64], which uses a very similar method and fake efficiency control region to this analysis.

Tables 16 and 17 include a systematic uncertainty of  $\pm 50\%$  due to the theory uncertainties on the MC simulated background. It is the largest systematic uncertainty which predominantly

determines the overall systematic uncertainty. However, since the fake/non-prompt lepton background is a very small contribution in the signal region and since its uncertainty is negligible compared to the uncertainties from all other prompt and non-prompt backgrounds, this is sufficient for this analysis.

#### 8.4.5 Fake/Non-Prompt Lepton Composition

Fake/non-prompt leptons are created in several different processes which can be expected to have different fake efficiencies. In this analysis, no separation between fake/non-prompt leptons from different sources is made and a single average fake efficiency is used. Accordingly, to be able to extrapolate the fake efficiency from the fake efficiency control region to the application regions the contributions from different processes creating fake/non-prompt leptons should be approximately equal in all regions.

The origins of fake/non-prompt leptons in data cannot be identified and therefore MC simulations have to be used to measure the fake/non-prompt lepton compositions. As shown in the calculations of the fake efficiencies and fake factors, MC simulations do not model fake/non-prompt processes accurately and the fake/non-prompt lepton compositions from simulations are therefore only approximations. The truth composition of fake/non-prompt leptons is obtained with the IFFTruthClassifierTool [66]. It allows to distinguish between different categories of fake/non-prompt leptons. Fake leptons from light flavored jets are categorized as “Light Flavored Decays” which also includes non-prompt electrons from semi-leptonic decays of light flavored hadrons. However simulation studies on the origin of fake/non-prompt leptons show that this contribution is negligibly small [3]. Non-prompt leptons from heavy flavored jets are classified in the “b-Hadron Decays” or “c-Hadron Decays” categories and non-prompt electrons from photon conversions will be labeled as “Prompt Photon Conversion”. Leptons sorted into “Prompt Electrons” and “Prompt Muons” as well as all other categories are considered real leptons for this analysis, with the exception of “Unknown” and “KnownUnknown” which indicate problems with the MC dataset so that the origin of a lepton cannot be identified.

The truth composition of the fake efficiency control region for the electron channel is shown in Figure 32. As intended, fake/non-prompt lepton events are dominant in this region and the majority of events include fake electrons from light flavored jets, followed by non-prompt electrons from heavy flavored jets. Non-prompt electrons from photon conversions have a negligibly small contribution. Figures 34 and 36 show the truth electron composition in the validation region and signal region where the main contribution to the fake/non-prompt

	Electron Channel	Muon Channel
Fake Efficiency Control Region	$3.0 \pm 0.3$	$0.14 \pm 0.02$
Validation Region	$11 \pm 5$	$0.1 \pm 0.1$
Signal Region	$90 \pm 270$	$0.5 \pm 0.3$

Table 18: Ratios of the number of events with fake/non-prompt leptons from light flavored jets to the number of events with fake/non-prompt leptons from heavy flavored jets in the fake efficiency control region, validation region and signal region. Only statistical uncertainties are shown.

lepton background stems from non-prompt electrons from photon conversions. This is not ideal as this process is not significantly represented in the fake efficiency control region and therefore not included in the fake efficiency. Further studies are required to investigate if this contribution is an artifact from estimating the truth composition using MC simulations, which are only approximations, and the detailed origins of these non-prompt electrons [62]. To account for non-prompt electrons from photon conversions in the fake efficiency, a different fake efficiency control region region would be required.

For fake/non-prompt electrons originating from jets the main contribution are light flavored jets misreconstructed as electrons. To compare the truth compositions in the fake efficiency control region and the two application regions, the ratio of fake electrons from light flavored jets to non-prompt electrons from heavy flavored jets is calculated and shown in Table 18. The large statistical uncertainties on the ratios originate from the low number of fake/non-prompt lepton events in the validation and signal region. The ratios in the validation and signal regions are in agreement with the ratio in the fake efficiency control region within  $2\sigma$ , and therefore the extrapolation of the fake efficiency from the fake efficiency control region into the application regions is justified for fake/non-prompt leptons originating from jets.

The truth compositions for the muon channel are shown in Figures 33, 35 and 37 for the fake efficiency control region, validation region and signal region respectively. In all three regions the main contribution to the fake/non-prompt muon background comes from misattributed non-prompt muons from heavy flavored jets. Statistics are very limited in the validation and signal regions, and accordingly the ratios of the number of events with fake/non-prompt muons from light to heavy flavored jets (Table 18) show large uncertainties. The ratios in the validation and signal regions are in  $2\sigma$  agreement with the ratio in the fake efficiency control region, so the extrapolation of the fake efficiency into the analysis regions in the muon channel is validated.

Due to the limited statistics in the validation region and signal region, the lepton composition is not examined in dependence on the lepton's  $\eta$  or  $p_T$ . Analyses and simulations show that

while non-prompt muons from heavy favored decays are the only significant contribution in the muon channel, for electrons non-prompt electrons dominate in the low  $p_T$  range while for high  $p_T$  misreconstructed light flavored jets and photon conversion become dominant [3]. The IFF recommends to include a systematic uncertainty if the fake/non-prompt lepton compositions in the fake efficiency control region and the application regions are significantly different. This is the case in the electron channel, where the validation and signal regions include significant contributions from non-prompt electrons from photon conversions that are not present in the fake efficiency control region. However, the relative contributions from fake/non-prompt electrons created from jets and fake/non-prompt muons from all origins are in agreement between the fake efficiency control region and the application regions and therefore, a predominant fraction of the fake/non-prompt lepton background is well represented in all regions. Additionally, the ATLAS analysis investigating the invariant mass spectrum of four leptons found that even large differences in fake/non-prompt lepton compositions do not have to correspond to different fake efficiencies [64]. Given that the fake/non-prompt lepton background estimation used in this analysis is verified in the validation region, possible effects from different fake/non-prompt lepton compositions on the fake efficiency should be safely included in the  $\pm 50\%$  assigned for the theory uncertainties from MC simulations [64, 65].

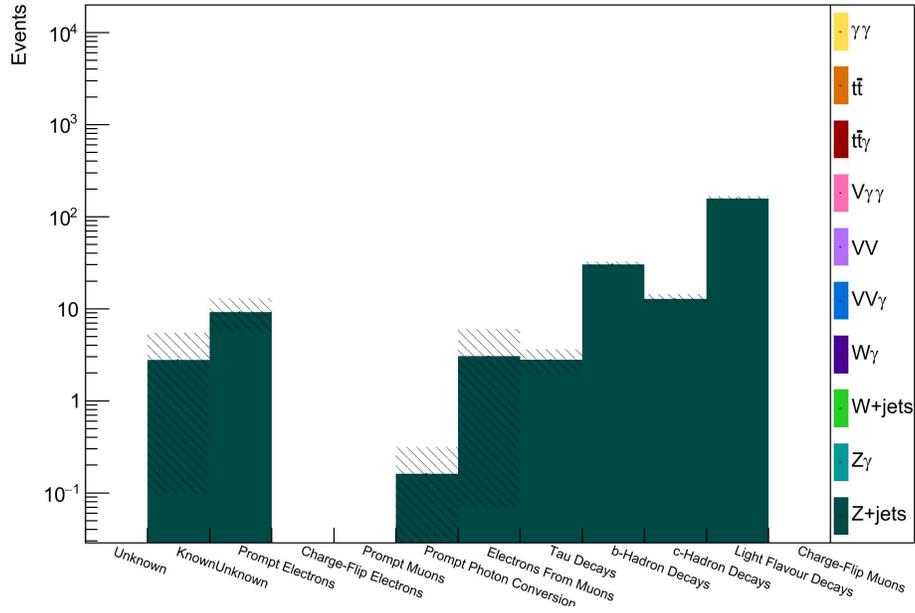


Figure 32: Truth composition of electrons in the fake efficiency control region. Only statistical uncertainties are shown.  $V$  is signifying a vector boson, either  $W^\pm$  or  $Z^0$ .

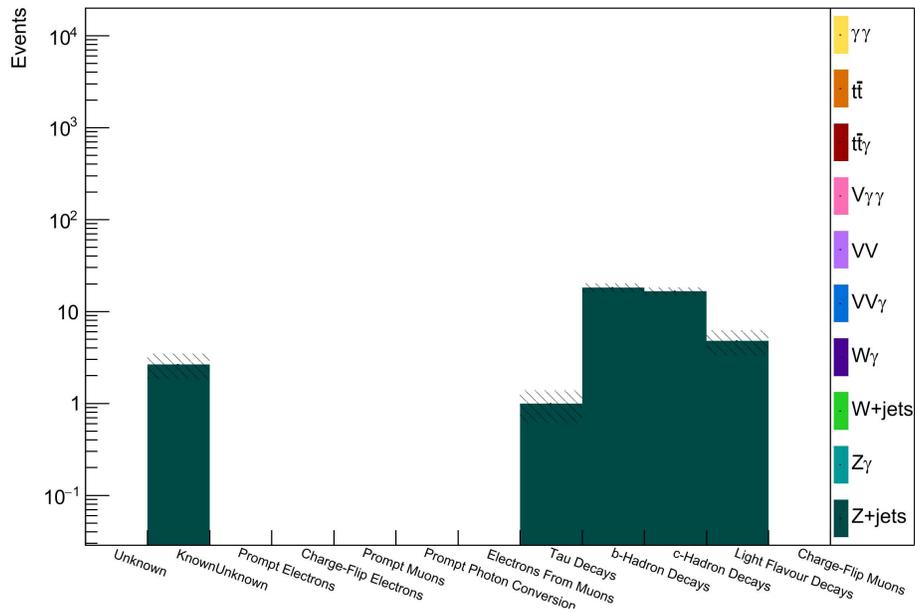


Figure 33: Truth composition of muons in the fake efficiency control region. Only statistical uncertainties are shown.  $V$  is signifying a vector boson, either  $W^\pm$  or  $Z^0$ .

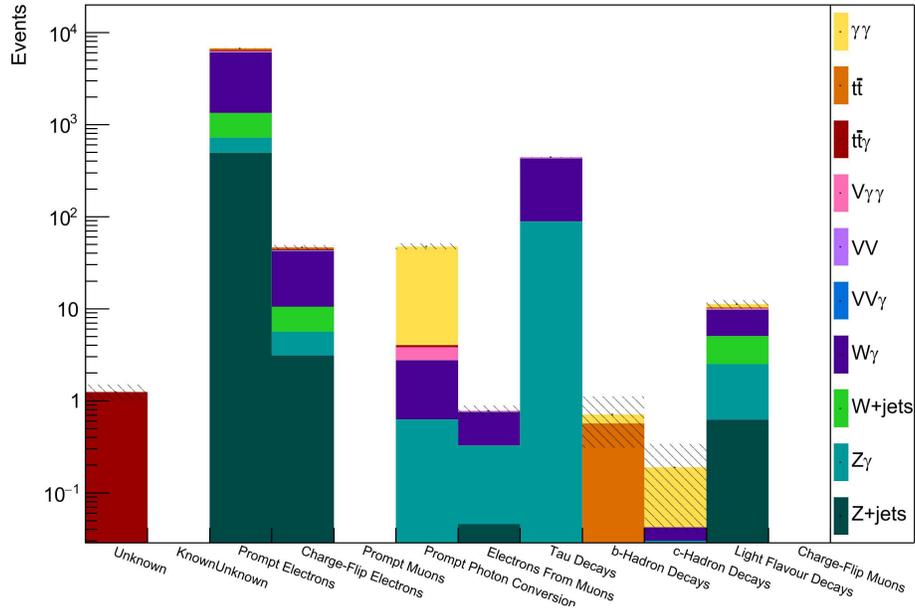


Figure 34: Truth composition of electrons in the validation region. Only statistical uncertainties are shown.  $V$  is signifying a vector boson, either  $W^\pm$  or  $Z^0$ .

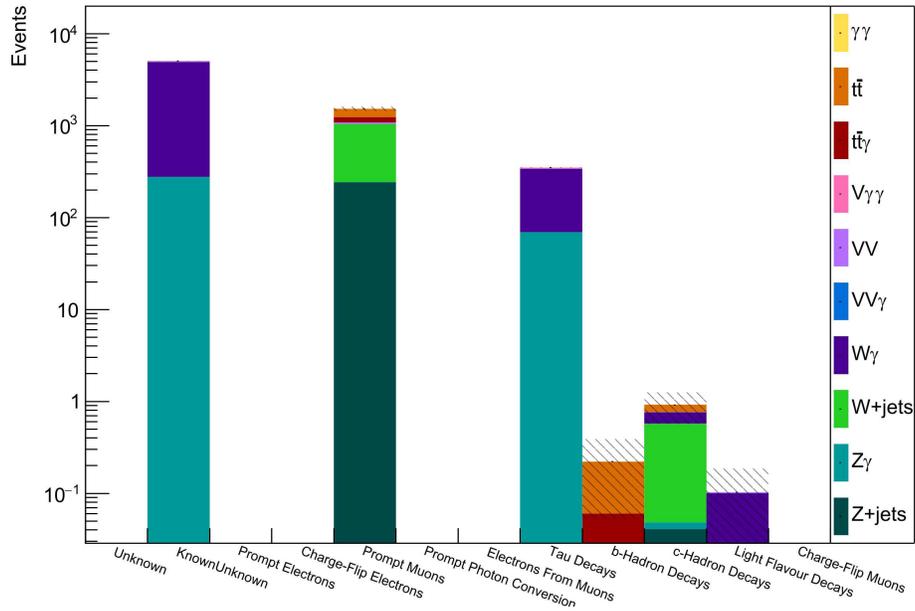


Figure 35: Truth composition of muons in the validation region. Only statistical uncertainties are shown.  $V$  is signifying a vector boson, either  $W^\pm$  or  $Z^0$ .

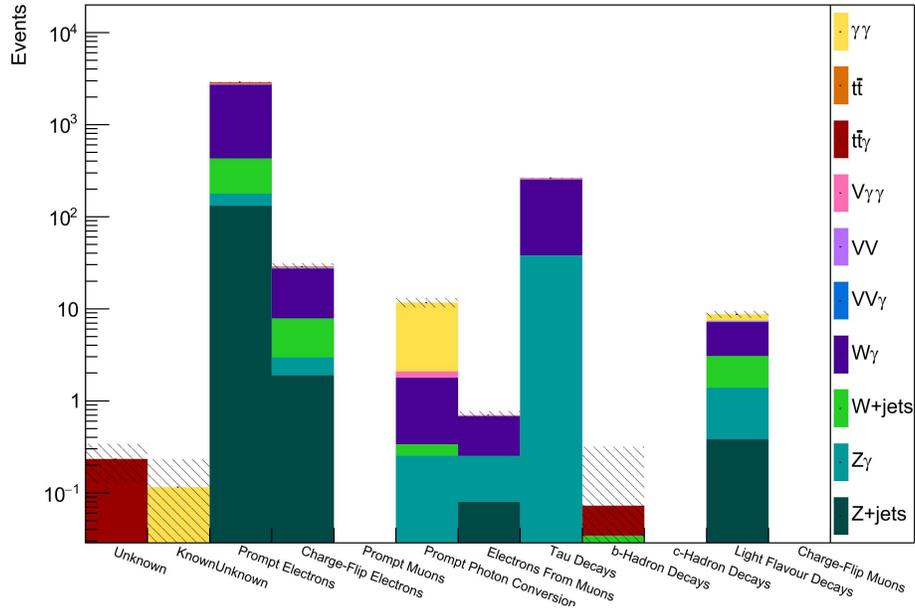


Figure 36: Truth composition of electrons in the signal region. Only statistical uncertainties are shown.  $V$  is signifying a vector boson, either  $W^\pm$  or  $Z^0$ .

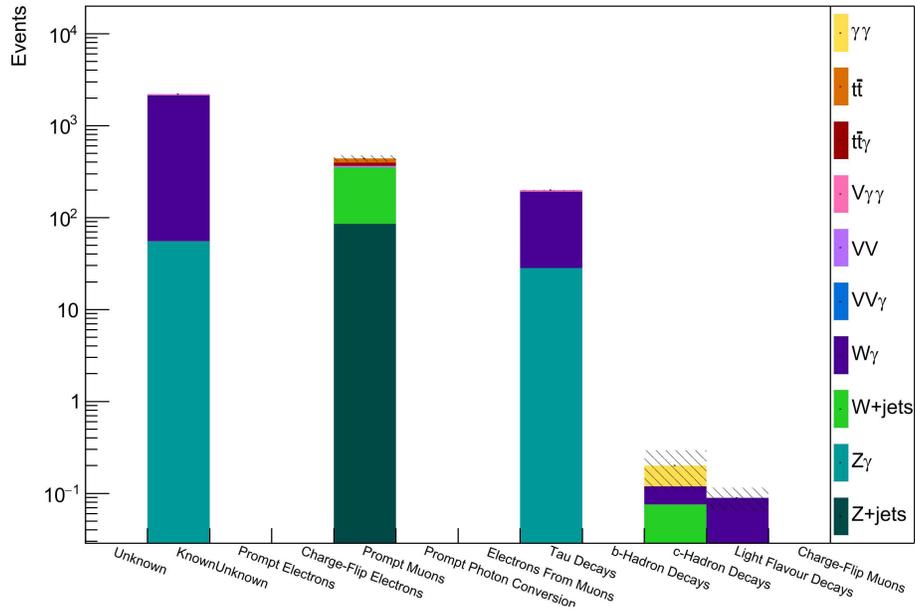


Figure 37: Truth composition of muons in the signal region. Only statistical uncertainties are shown.  $V$  is signifying a vector boson, either  $W^\pm$  or  $Z^0$ .

## 9 Conclusion and Outlook

The ongoing  $H_5^\pm \rightarrow W^\pm\gamma$  analysis is searching for a low-mass fermiophobic charged Higgs boson in  $\sqrt{s} = 13$  TeV proton-proton collisions at the LHC with the ATLAS detector. To achieve an accurate measurement of signal events, all possible sources of background must be accounted for. The majority of background events are prompt backgrounds which can be estimated with Monte Carlo simulations. In contrast, the three contributing fake/non-prompt backgrounds, electrons faking photons, jets faking photons and fake/non-prompt leptons, must be estimated with data-driven methods.

This thesis presents an estimation of the fake/non-prompt lepton background for the  $H_5^\pm \rightarrow W^\pm\gamma$  analysis with the fake factor method. This technique relies on defining a control region enriched in fake/non-prompt leptons where the fake efficiency is measured. The control region and the analysis regions are kinematically close but orthogonal, and the fake efficiency is assumed to be the same. From the fake efficiency, the expected number of fake/non-prompt lepton events in the analysis regions can be calculated. For the  $H_5^\pm \rightarrow W^\pm\gamma$  analysis the fake/non-prompt lepton background has a very small contribution below 1%.

The extrapolation of the fake efficiency into the analysis regions is only possible if the fake/non-prompt leptons are created in similar processes. To verify this, the fake/non-prompt lepton compositions in the validation and signal regions are compared to the lepton origins in the fake efficiency control region. Overall, good agreement is found and therefore applying the fake efficiency in the analysis regions is justified. However, in the validation and signal regions there is a contribution from electrons from photon conversions that is not present in the fake efficiency control region. Further investigations are required to determine whether these leptons are introduced by approximating the lepton composition from MC simulations and their detailed origins. Electrons from photon conversions are not typically a large contributor to the fake/non-prompt lepton background and the IFF has very few recommendations on how to include them in background analyses [62].

The statistical and systematic uncertainties of the fake/non-prompt lepton background estimates are evaluated. For the electron channel, systematic uncertainties are dominant, whereas in the muon channel due to low statistics in the fake efficiency control region the statistical uncertainties are larger. The systematic uncertainties mainly stem from theory uncertainties on the MC simulations, where a flat and large uncertainty was assigned that includes all associated systematics. To improve the systematic errors, these uncertainties could be fully evaluated following the recommendations from the ATLAS Particle Modeling Group (PMG) [62]. However, since the fake/non-prompt lepton background has a small

contribution to the signal region and the assigned large relative uncertainties are very small compared to the uncertainties from other backgrounds, this would not significantly improve the background estimation.

The implemented techniques to estimate the fake/non-prompt lepton background are validated in the validation region. Due to the small contribution from the fake/non-prompt lepton background and the large relative uncertainties on the other background components, the validation would also hold without the fake/non-prompt lepton background added. Accordingly, for this analysis including the fake/non-prompt lepton background does not significantly improve the background description and the agreement between the background estimates and measured data in the validation region. However, if in possible future analysis versions the uncertainties on the other backgrounds could be reduced, accounting for the fake/non-prompt lepton background would become more important. In that case, the validation of the fake/non-prompt lepton background estimates shows that the fake factor method provides a correct description of this background.

In performance studies, the fake factor method shows accurate and robust estimates in all possible analysis scenarios and it was therefore chosen to calculate the fake/non-prompt lepton background. The main disadvantage of the fake factor method is that the relative uncertainty on the background estimates can be significantly larger than the ones obtained with other fake/non-prompt background estimation techniques [3]. According to the IFF, the choice of applied techniques should depend on the complexity of the fake lepton background and its contribution to the overall background of the analysis [3]. Since the fake/non-prompt lepton background is a small contribution to the signal region, the fake factor method provides a suitable estimate. However, there would be several benefits from including other fake/non-prompt lepton estimation methods.

The fake factor method assumes that real efficiencies are well modeled in MC simulations. In contrast, for a fake/non-prompt lepton background estimate with the matrix method the real efficiency also needs to be measured in a separate control region. Ideally, the fake factor method and the matrix method should yield the same result, and any observed difference would indicate either mismodeling of real leptons or mismeasurement of the real efficiency. Therefore, including the matrix method into the analysis would provide an opportunity to check the assumption of well modeled real leptons [3].

For analyses with low statistics the Poisson likelihood matrix method can in some scenarios yield lower statistical uncertainties than the fake factor and matrix methods. It treats the number of expected fake/non-prompt leptons as free parameters which are maximized in a likelihood fit to find the most probable result. The fake/non-prompt lepton yields are

constrained to be non-negative, which is a large advantage of this method as it avoids unphysical negative results from statistical fluctuations [3].

In summary, the methods presented in this thesis provide a validated estimation of the fake/non-prompt lepton background which can be included in the  $H_5^\pm \rightarrow W^\pm \gamma$  analysis.

## Bibliography

- [1] U. Patel, Z. Li, M. Kruse, et al. *Search for a Fermiophobic Beyond the Standard Model Charged Higgs Boson through  $W\gamma$  Resonances for Masses Below 200 GeV*. Tech. rep. ATL-COM-PHYS-2022-1156. CERN, Dec. 2022. URL: [cds.cern.ch/record/2843316/](https://cds.cern.ch/record/2843316/).
- [2] Zhelun Li. “Search for a Fermiophobic Charged Higgs Boson in Proton-Proton Collisions with the ATLAS Detector”. PhD thesis. McGill University, 2023.
- [3] The ATLAS Collaboration. “Tools for Estimating Fake/Non-Prompt Lepton Backgrounds with the ATLAS Detector at the LHC”. In: (Nov. 2022). DOI: [10.48550/arXiv.2211.16178](https://doi.org/10.48550/arXiv.2211.16178).
- [4] M. Thomson. *Modern Particle Physics*. Cambridge University Press, Sept. 2013. DOI: [10.1017/cbo9781139525367](https://doi.org/10.1017/cbo9781139525367).
- [5] D. Griffiths. *Introduction to Elementary Particles*. John Wiley and Sons, Dec. 1987. DOI: [10.1002/9783527618460](https://doi.org/10.1002/9783527618460).
- [6] M. Peskin and D. Schroeder. *An Introduction to Quantum Field Theory*. CRC Press, 1995. DOI: [10.1201/9780429503559](https://doi.org/10.1201/9780429503559).
- [7] The ATLAS Collaboration. “Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC”. In: *Physics Letters B* 716.1 (Sept. 2012). DOI: [10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020).
- [8] The CMS Collaboration. “Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC”. In: *Physics Letters B* 716.1 (Sept. 2012). DOI: [10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021).
- [9] M. Kobayashi and T. Maskawa. “CP-Violation in the Renormalizable Theory of Weak Interaction”. In: *Progress of Theoretical Physics* 49.2 (Feb. 1973). DOI: [10.1143/PTP.49.652](https://doi.org/10.1143/PTP.49.652).
- [10] M. Perl, G. Abrams, A. Boyarski, et al. “Evidence for Anomalous Lepton Production in  $e^+ - e^-$  Annihilation”. In: *Physical Review Letters* 35.22 (Dec. 1975). DOI: [10.1103/PhysRevLett.35.1489](https://doi.org/10.1103/PhysRevLett.35.1489).
- [11] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Physical Review Letters* 13.9 (Aug. 1964). DOI: [10.1103/PhysRevLett.13.321](https://doi.org/10.1103/PhysRevLett.13.321).
- [12] The SNO Collaboration. “Combined Analysis of All Three Phases of Solar Neutrino Data from the Sudbury Neutrino Observatory”. In: *Physical Review C* 88.2 (Aug. 2013). DOI: [10.1103/PhysRevC.88.025501](https://doi.org/10.1103/PhysRevC.88.025501).

- [13] S. Bilenky and S. Petcov. “Massive Neutrinos and Neutrino Oscillations”. In: *Reviews of Modern Physics* 59.3 (July 1987). DOI: 10.1103/RevModPhys.59.671.
- [14] The Particle Data Group. “Review of Particle Physics”. In: *Progress of Theoretical and Experimental Physics* 2022.8 (Aug. 2022). DOI: 10.1093/ptep/ptac097.
- [15] The CMS Collaboration. “Search for a Charged Higgs Boson in  $pp$  collisions at  $\sqrt{s} = 8$  TeV”. In: *Journal of High Energy Physics* 2015.11 (Nov. 2015). DOI: 10.1007/JHEP11(2015)018.
- [16] The ATLAS Collaboration. “Search for Charged Higgs Bosons Produced in Association with a Top Quark and Decaying via  $H^{pm} \rightarrow \tau\nu$  using  $pp$  Collision Data Recorded at  $\sqrt{s} = 13$  TeV by the ATLAS Detector”. In: *Physics Letters B* 759 (Aug. 2016). DOI: 10.1016/j.physletb.2016.06.017.
- [17] The ATLAS Collaboration. “Search for Charged Higgs Bosons Decaying via  $H^{\pm} \rightarrow \tau^{\pm}\nu$  in Fully Hadronic Final States using  $pp$  Collision Data at  $\sqrt{s} = 8$  TeV with the ATLAS Detector”. In: *Journal of High Energy Physics* 2015.3 (Mar. 2015). DOI: 10.1007/JHEP03(2015)088.
- [18] H. Georgi and M. Machacek. “Doubly Charged Higgs Bosons”. In: *Nuclear Physics B* 262.3 (Dec. 1985). DOI: 10.1016/0550-3213(85)90325-6.
- [19] R. Enberg, J. Rathsman, and G. Wouda. “Higgs Phenomenology in the Stealth Doublet Model”. In: *Physical Review D* 91.9 (May 2015). DOI: 10.1103/PhysRevD.91.095002.
- [20] A. Pich and P. Tuzon. “Yukawa Alignment in the Two-Higgs-Doublet Model”. In: *Physical Review D* 80.9 (Nov. 2009). DOI: 10.1103/PhysRevD.80.091702.
- [21] H. Logan and Y. Wu. “Searching for the  $W\gamma$  Decay of a Charged Higgs Boson”. In: *Journal of High Energy Physics* 2018.11 (Nov. 2018). DOI: 10.1007/JHEP11(2018)121.
- [22] The CMS Collaboration. “Search for Charged Higgs Bosons Produced via Vector Boson Fusion and Decaying Into a Pair of  $W$  and  $Z$  bosons using proton-proton collisions at  $\sqrt{s} = 13$  TeV”. In: *Physical Review Letters* 119.14 (Oct. 2017). DOI: 10.1103/PhysRevLett.119.141802.
- [23] The ATLAS Collaboration. “Search for Resonant  $WZ$  Production in the Fully Leptonic Final State in Proton-Proton Collisions at  $\sqrt{s} = 13$  TeV with the ATLAS Detector”. In: *Physics Letters B* 787 (Dec. 2018). DOI: 10.1016/j.physletb.2018.10.021.
- [24] The CMS Collaboration. “Measurement of Electroweak  $WZ$  Boson Production and Search for New Physics in  $WZ +$  Two Jets Events in  $pp$  Collisions at  $\sqrt{s} = 13$  TeV”. In: *Physics Letters B* 795 (Aug. 2019). DOI: 10.1016/j.physletb.2019.05.042.

- [25] The ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *Journal of Instrumentation* 3.08 (Aug. 2008). DOI: 10.1088/1748-0221/3/08/S08003.
- [26] The CMS Collaboration. “The CMS Experiment at the CERN LHC”. In: *Journal of Instrumentation* 3.08 (Aug. 2008). DOI: 10.1088/1748-0221/3/08/S08004.
- [27] The LHCb Collaboration. “The LHCb Detector at the LHC”. In: *Journal of Instrumentation* 3.08 (Aug. 2008). DOI: 10.1088/1748-0221/3/08/S08005.
- [28] The ALICE Collaboration. “The ALICE Experiment at the CERN LHC”. In: *Journal of Instrumentation* 3.08 (Aug. 2008). DOI: 10.1088/1748-0221/3/08/S08002.
- [29] L. Evans and P. Bryant. “LHC Machine”. In: *Journal of Instrumentation* 3.08 (Aug. 2008). DOI: 10.1088/1748-0221/3/08/S08001.
- [30] E. Lopienska. *The CERN Accelerator Complex, Layout in 2022*. Feb. 2022. URL: [cds.cern.ch/images/CERN-GRAPHICS-2022-001-1](https://cds.cern.ch/images/CERN-GRAPHICS-2022-001-1).
- [31] R. Alemany-Fernandez, E. Bravin, L. Drosdal, et al. “Operation and Configuration of the LHC in Run 1”. Internal Note CERN-ACC-NOTE-2013-0041. May 2014. URL: [cds.cern.ch/record/1631030](https://cds.cern.ch/record/1631030).
- [32] R. Steerenberg, M. Albert, R. Alemany-Fernández, et al. “Operation and Performance of the CERN Large Hadron Collider During Proton Run 2”. In: *Proceedings of the 10th International Particle Accelerator Conference IPAC2019* (May 2019). DOI: 10.18429/JACOW-IPAC2019-MOPMP031.
- [33] S. Fartoukh, S. Kostoglou, M. Solfaroli Camillocci, et al. *LHC Configuration and Operational Scenario for Run 3*. Tech. rep. CERN-ACC-2021-0007. CERN, Nov. 2021. URL: [cds.cern.ch/record/2790409](https://cds.cern.ch/record/2790409).
- [34] G. Apollinari, I. Bejar Alonso, O. Brüning, et al. *High-Luminosity Large Hadron Collider (HL-LHC) Technical Design Report V. 0.1*. Tech. rep. CERN-2020-010. CERN, Sept. 2017. DOI: 10.23731/CYRM-2017-004.
- [35] C. Grupen and B. Shwartz. *Particle Detectors*. Cambridge University Press, Mar. 2008. DOI: 10.1017/CB09780511534966.
- [36] H. Pernegger. “The Pixel Detector of the ATLAS Experiment for LHC Run-2”. In: *Journal of Instrumentation* 10.06 (June 2015). DOI: 10.1088/1748-0221/10/06/C06012.

- [37] The ATLAS Collaboration. *ATLAS Detector and Physics Performance. Technical Design Report. Volume 1*. Tech. rep. CERN-LHCC-99-014. CERN, May 1999. URL: [cds.cern.ch/record/391176](https://cds.cern.ch/record/391176).
- [38] The ATLAS collaboration. “Resolution of the ATLAS Muon Spectrometer Monitored Drift Tubes in LHC Run 2”. In: *Journal of Instrumentation* 14.09 (Sept. 2019). DOI: 10.1088/1748-0221/14/09/P09011.
- [39] The ATLAS Collaboration. “Performance of the ATLAS RPC Detector and Level-1 Muon Barrel Trigger at  $\sqrt{s} = 13$  TeV”. In: *Journal of Instrumentation* 16.06 (June 2021). DOI: 10.1088/1748-0221/16/07/P07029.
- [40] A. Ruiz Martinez on behalf of the ATLAS collaboration. “The Run-2 ATLAS Trigger System”. In: *Journal of Physics: Conference Series* 762 (Oct. 2016). DOI: 10.1088/1742-6596/762/1/012003.
- [41] F. Pastore on behalf of the ATLAS collaboration. “The ATLAS Trigger System: Past, Present and Future”. In: *Nuclear and Particle Physics Proceedings* 273-275 (Apr. 2016). DOI: 10.1016/j.nuclphysbps.2015.09.167.
- [42] A. Barton on behalf of the ATLAS collaboration. “ATLAS: HL-LHC Upgrades and B-Physics Measurements”. In: *19th International Conference on B-Physics at Frontier Machines* (Oct. 2021). DOI: 10.22323/1.391.0061.
- [43] The ATLAS Collaboration. *2015 Start-Up Trigger Menu and Initial Performance Assessment of the ATLAS Trigger Using Run-2 Data*. Tech. rep. ATL-DAQ-PUB-2016-001. CERN, Mar. 2016. URL: [cds.cern.ch/record/2136007](https://cds.cern.ch/record/2136007).
- [44] The ATLAS Collaboration. *Trigger Menu in 2016*. Tech. rep. ATL-DAQ-PUB-2017-001. CERN, Jan. 2017. URL: [cds.cern.ch/record/2242069](https://cds.cern.ch/record/2242069).
- [45] The ATLAS Collaboration. *Trigger Menu in 2017*. Tech. rep. ATL-DAQ-PUB-2018-002. CERN, June 2018. URL: [cds.cern.ch/record/2625986](https://cds.cern.ch/record/2625986).
- [46] The ATLAS Collaboration. *Trigger Menu in 2018*. Tech. rep. ATL-DAQ-PUB-2019-001. CERN, Oct. 2019. URL: [cds.cern.ch/record/2693402](https://cds.cern.ch/record/2693402).
- [47] The ATLAS Collaboration. “Performance of the ATLAS Muon Triggers in Run 2”. In: *Journal of Instrumentation* 15.09 (Sept. 2020). DOI: 10.1088/1748-0221/15/09/p09015.
- [48] The ATLAS Collaboration. *Good Run Lists For Analysis Run 2*. Dec. 2022. URL: [twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/GoodRunListsForAnalysisRun2](https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/GoodRunListsForAnalysisRun2).

- [49] The ATLAS Collaboration. *Full List Of P-Tags And Patterns*. Aug. 2023. URL: [twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/FullListOfPtagsAndPatterns](https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/FullListOfPtagsAndPatterns).
- [50] The ATLAS Collaboration. “The ATLAS Simulation Infrastructure”. In: *The European Physical Journal C* 70.3 (Sept. 2010). DOI: 10.1140/epjc/s10052-010-1429-9.
- [51] S. Agostinelli, J. Allison, K. Amako, et al. “Geant4 - A Simulation Toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (July 2003). DOI: 10.1016/S0168-9002(03)01368-8.
- [52] The 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”. In: *The European Physical Journal C* 79.8 (Aug. 2019). DOI: 10.1140/epjc/s10052-019-7140-6.
- [53] The ATLAS Collaboration. “Performance of Electron and Photon Triggers in ATLAS During LHC Run 2”. In: *The European Physical Journal C* 80.1 (Jan. 2020). DOI: 10.1140/epjc/s10052-019-7500-2.
- [54] The ATLAS Collaboration. “Muon Reconstruction and Identification Efficiency in ATLAS Using the Full Run 2  $pp$  Collision Dataset at  $\sqrt{s} = 13$  TeV”. In: *The European Physical Journal C* 81.578 (June 2021). DOI: 10.1140/epjc/s10052-021-09233-2.
- [55] The ATLAS Collaboration. “Electron and Photon Performance Measurements with the ATLAS Detector Using the 2015-2017 LHC Proton-Proton Collision Data”. In: *Journal of Instrumentation* 14.12 (Dec. 2019). DOI: 10.1088/1748-0221/14/12/P12006.
- [56] The ATLAS Collaboration. “Jet Reconstruction and Performance Using Particleflow with the ATLAS Detector”. In: *The European Physical Journal C* 77.7 (July 2017). DOI: 10.1140/epjc/s10052-017-5031-2.
- [57] The ATLAS Collaboration. *Tagging and Suppression of Pileup Jets*. Tech. rep. ATLAS-PHYS-PUB-2014-001. CERN, May 2014. URL: [cds.cern.ch/record/1700870](https://cds.cern.ch/record/1700870).
- [58] The ATLAS Collaboration. “ATLAS  $b$ -Jet Identification Performance and Efficiency Measurement with  $t\bar{t}$  Events in  $pp$  Collisions at  $\sqrt{s} = 13$  TeV”. In: *The European Physical Journal C* 79.11 (Nov. 2019). DOI: 10.1140/epjc/s10052-019-7450-8.
- [59] The ATLAS Collaboration. “Performance of Missing Transverse Momentum Reconstruction with the ATLAS Detector Using Proton–Proton Collisions at  $\sqrt{s} = 13$  TeV”. In: *The European Physical Journal C* 78.11 (Nov. 2018). DOI: 10.1140/epjc/s10052-018-6288-9.

- [60] D. Adams, C. Anastopoulos, A. Andreazza, et al. *Recommendations of the Physics Objects and Analysis Harmonisation Study Groups 2014*. Tech. rep. ATL-PHYS-INT-2014-018. CERN, July 2014. URL: [cds.cern.ch/record/1743654](https://cds.cern.ch/record/1743654).
- [61] R. William. *RooBukinPDF*. CERN. July 2007. URL: [root.cern.ch/doc/master/RooBukinPdf\\_8cxx\\_source.html](https://root.cern.ch/doc/master/RooBukinPdf_8cxx_source.html).
- [62] The ATLAS Isolation and Fake Forum. *Fake Lepton Backgrounds*. Tech. rep. CERN, May 2022. URL: [twiki.cern.ch/twiki/bin/view/AtlasProtected/~FakeLeptonBackgrounds](https://twiki.cern.ch/twiki/bin/view/AtlasProtected/~FakeLeptonBackgrounds).
- [63] The ATLAS Collaboration. *Observation of  $W\gamma\gamma$  Triboson Production in Proton-Proton Collisions at  $\sqrt{s} = 13$  TeV with the ATLAS Detector*. Tech. rep. ATLAS-CONF-2023-005. CERN, Mar. 2023. URL: [cds.cern.ch/record/2853334/](https://cds.cern.ch/record/2853334/).
- [64] Z. Grout, M. Goblirsch-Kolb, P. Bhattarai, et al. *Unfolded Four Lepton Mass and Angular Distributions*. Tech. rep. CERN, Dec. 2019. URL: [cds.cern.ch/record/2704274/](https://cds.cern.ch/record/2704274/).
- [65] B. Abbott, J. Arguin, S. Berlendis, et al. *Search for Strongly-Produced Superpartners in Final States with Same-Sign or Three Leptons and Jets in 2015+2016 pp Collisions Data at  $\sqrt{s} = 13$  TeV*. Tech. rep. ATL-COM-PHYS-2016-1616. CERN, Nov. 2017. URL: [cds.cern.ch/record/2231789/](https://cds.cern.ch/record/2231789/).
- [66] The ATLAS Isolation and Fake Forum. *Truth Classification*. Oct. 2022. URL: [gitlab.cern.ch/atlas/athena/-/tree/21.2/PhysicsAnalysis/~AnalysisCommon/TruthClassification](https://gitlab.cern.ch/atlas/athena/-/tree/21.2/PhysicsAnalysis/~AnalysisCommon/TruthClassification).