Observation of Electroweak VBS ZZ Production and Search for New Gauge boson Z' in four-lepton Final State with the ATLAS Experiment

by
Shuzhou Zhang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Physics)
in The University of Michigan
2023

Doctoral Committee:

Professor Bing Zhou, Chair
Professor Christine Aidala
Professor James Liu
Professor Junjie Zhu
Professor Qiong Yang

Shuzhou Zhang

shuzhouz@umich.edu

ORCID iD: 0000-0001-9039-9809

Acknowledgements

I stand here today, filled with immense gratitude and humility, as I acknowledge the remarkable individuals who have provided unwavering support throughout my Ph.D. journey. This thesis would not have been possible without the love, guidance, and encouragement of these extraordinary people.

First and foremost, I would like to express my deepest gratitude to my loving parents, who have been my pillars of strength throughout this journey. To my father, Mr. Shaoxian Zhang, who instilled in me the value of hard work and perseverance, and to my mother, Ms. Xiaodong Fan, who taught me the importance of kindness and empathy, I dedicate this thesis. Your unwavering belief in my abilities and your constant support, both emotional and financial, have been invaluable in bringing me to this point in my life. Thank you for your patience and understanding during the long days and late nights, for your words of encouragement during moments of doubt, and for your unconditional love and support throughout my life.

I am immensely grateful to my advisor, Prof. Bing Zhou, for her invaluable guidance, mentorship, and support. Her vast knowledge, expertise, and enthusiasm have been instrumental in shaping my research and academic growth. Prof. Bing Zhou has been not only an exceptional mentor but also a source of inspiration, pushing me to excel and reach new heights. I feel incredibly fortunate to have had the opportunity to learn from and work with such a dedicated and compassionate advisor.

Additionally, I would like to express my gratitude to the UM ATLAS group professors, including Prof. Bing Zhou, Jianming Qian, Tom Schwarz, Dan Amidei, Junjie Zhu, and Jay Chapman, as well as Tiesheng Dai, for fostering a supportive university environment and securing the funding that has made this work possible. I would like to specifically thank Dr. Tiesheng Dai for providing me with valuable instructions and guidance related to the hardware work.

I would also like to thank the members of my thesis committee, Prof. James Liu, Junjie Zhu, Christine Aidala, and Qiong Yang, for their insightful feedback, constructive criticism, and patience throughout this process. Your expertise and thought-provoking questions have helped to refine my research and contributed to my personal and professional growth.

To my fellow graduate students and collaborators, particularly Zhe Yang, Jing Li, Dr. Aaron White, Bing Li, Cong Geng, and Zirui Wang, I extend my gratitude for the ca-

maraderie, stimulating intellectual discussions, and shared experiences. Your friendship, support, and humor have made the long hours spent in the group more enjoyable and significantly contributed to my overall well-being during this journey. Specifically, Dr. Bing Li, Cong Geng, and Zirui Wang provided patient and invaluable guidance on data analysis while also offering substantial support during my time in France. Dr. Aaron White shared numerous valuable insights on the thesis and supplied well-crafted plots that have been integrated into this thesis. I sincerely appreciate their assistance in both my professional and daily life, which has been truly impactful and deeply valued.

I would also like to extend my gratitude to the members of the ATLAS Collaboration who have diligently constructed and operated the detector and computing infrastructure that has served as the foundation for my work. Both their efforts and mine have been made possible by the generous support of various national funding agencies, including the Department of Energy. Additionally, I would like to express my appreciation to the dedicated engineers and staff at CERN who ensure the smooth operation of the LHC and maintain the site, as well as the diligent ATLAS secretaries who have consistently provided assistance in resolving the numerous challenges that arise from residing in France.

I am grateful for the funding support from the department of energy and the physics department of UM, which has enabled me to pursue my research and complete this thesis.

Lastly, I would like to express my heartfelt appreciation to my friends and extended family members for their love, support, and understanding throughout this journey. Your unwavering faith in me and your constant encouragement has been invaluable, and I am eternally grateful for each and every one of you.

Thank you all for being a part of this incredible journey. I am forever indebted to you for helping me achieve this milestone.

Contents

List of Figures

List of Tables

List of Abbreviations

ATLAS A Toroidal LHC Apparatus.

AUC Area Under ROC Curve.

BDT Boost Decision Tree.

BSM Beyond Standard Model.

CERN The European Organization for Nuclear Research.

CMS Compact Muon Solenoid.

CR Control Region.

DNN Deep Neural Network.

GBDT Gradient Boost Decision Tree.

Jet A narrow cone of hadrons and other particles produced by the hadronization of a quark or gluon.

JVF Jet Vertex Fraction.

JVT Jet Vertex Tagger.

LHC The Larger Hadron Collider.

LO Leading Order.

MC Mento Carlo simulation.

MET Missing Transverse Energy.

NLO Next Leading Order.

PDF Parton Distribution Function.

pDNN Parameterized Deep Neural Network.

QGC Quartic Gauge Coupling.

ROC Receiver Operating Characteristic.

Sherpa Monte Carlo generator for the Simulation of High-Energy Reactions of Particles.

SM The Standard Model.

SR Signal Region.

TGC Triple Gauge Coupling.

VBS Vector Boson Scattering.

VBSZZ Vector Boson Scattering with a Z-boson pair production.

Abstract

This thesis presents physics results from two analyses using data corresponding to an integrated luminosity of 139 fb^{-1} recorded with the ATLAS detector at a center-of-mass energy ($\sqrt{s}$) of 13 TeV at the Large Hadron Collider.

The first analysis result reported in this thesis is the observation of electroweak vector boson scattering (VBS) in the production of ZZ associating with two forward/backward hadronic jets, referred to as $ZZjj$ production. The electroweak VBS is a very important probe to study the dynamics of the spontaneous electroweak symmetry breaking, which explains the origin of the masses of elementary particles through their interactions with the Higgs field.

Among all processes related to vector-boson scattering, the electroweak production of a Z-boson pair associating with two jets is a rare and important one. The analysis used two different final states originating from the decays of the Z-boson pair: one contains four charged leptons and another contains two charged leptons and two neutrinos. Using the multivariate analysis based on boosted decision trees to reduce overwhelm QCD $ZZjj$ background, the overall signal significance of 5.7σ over the background-only hypothesis model is obtained in the analysis. The electroweak VBS $ZZjj$ production cross-section, 0.70 ± 0.18 fb, is measured, which is one of the smallest measured by ATLAS and is consistent with the Standard Model prediction.

The second analysis is a search for a new vector boson Z' with the four-muon (4μ) final state. The new gauge boson Z' is predicted by $L_\mu - L_\tau$ models to address observed phenomena that the Standard Model can not explain. The search examines the 4μ final state, using a deep learning neural network classifier to separate the Z' signal from the Standard Model background events. The di-muon invariant masses in the 4μ events are used to extract the Z' resonance signature. No significant excess of events is observed over the predicted background. Upper limits at a 95% confidence level on the Z' production cross-section times the decay branching fraction of $pp \to Z'\mu\mu \to 4\mu$ are set from 0.31 to 4.3 fb for the Z' mass ranging from 5 to 81 GeV. The corresponding common coupling strengths, $g_{Z'}$, of the Z' boson to the second and third generation leptons above 0.003 – 0.2 have been excluded.

Chapter 1

Introduction

This dissertation work is conducted with the ATLAS experiment [1] at the Large Hadron Collider (LHC) [2] at CERN [3]. An overview of fundamental particles, their interactions, and other basic particle physics analysis units and concepts are briefly introduced in this chapter.

1.1 Elementary Particles and Interactions

Elementary particles are building blocks of the physical world we are living in. Interactions between particles from the matter. Particle physics is all about space-time and energy to study the fundamental building blocks of our universe, the interactions among them, and the law that governs them. Our current understanding is the Standard Model [4] of particle physics. The research in experimental high energy physics is to test the SM with high precision and to search for its breakdown to advance our understanding of new physics.

1.1.1 Particles

Particles and interactions between them form the world. According to composition, particles can be classified into composite particles and elementary particles. All elementary particles in the SM are shown in Figure 1.1. They are quarks and leptons, and force carriers (gluon, photon, W, and Z). The Higgs boson is unique, which generates masses for elementary particles. There are many composite particles that are bound states of quarks, referred to as hadrons. The simplest composite particles are protons and neutrons, which are the components of nuclei formed by gluons (strong interaction force carriers) and quarks.

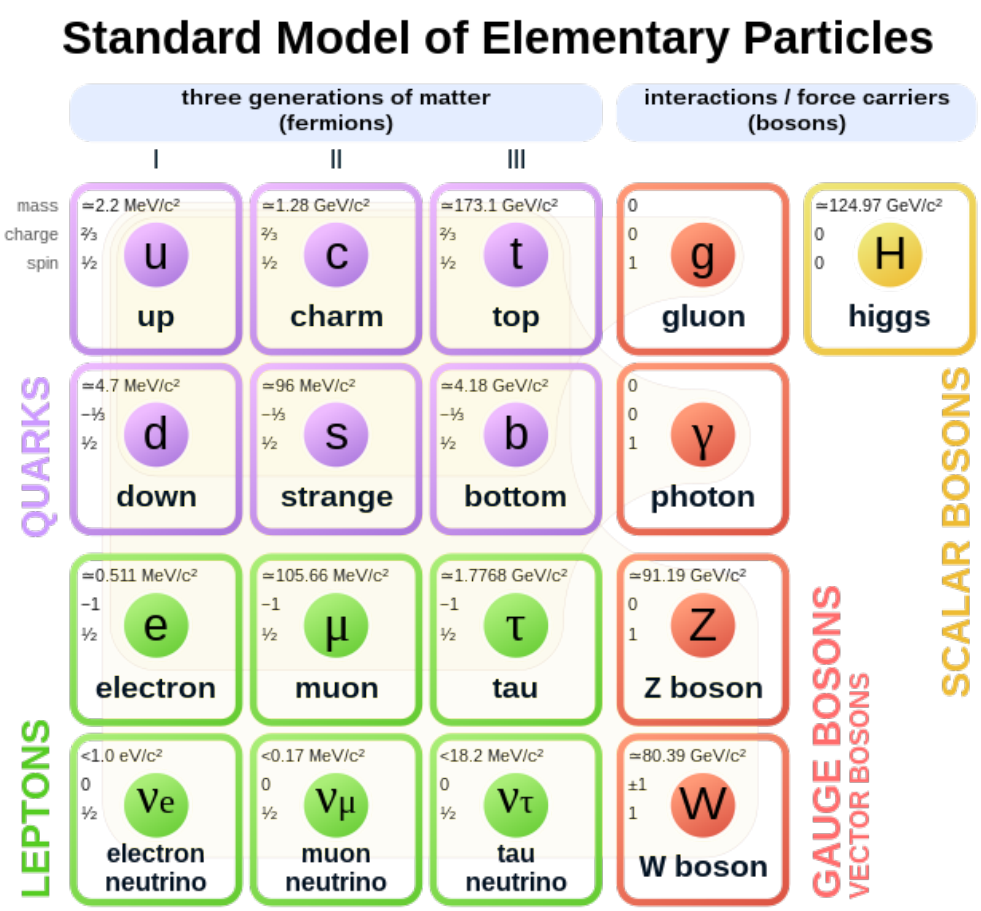


Figure 1.1: Elementary particles of the Standard Model.

Based on the spin (an intrinsic property of a particle), particles can also be grouped into fermions and bosons. Boson refers to a particle with an integer spin and follows Bose-Einstein statistics, while fermion refers to a particle with a half-odd integer spin (spin 1/2, spin 3/2, etc) and follow Fermi-Dirac statistics. Fermions obey the Pauli exclusion principle, which states no more than one identical fermion can occupy the same quantum state simultaneously.

There are four types of elementary spin-1 bosons in the SM. These are all interaction mediator particles, the force carriers. The photon mediates the electromagnetic force between charged particles. The photon itself doesn't carry any charge and has zero rest mass. $W^{\pm}$ and Z bosons are mediate particles of weak interactions and carry weak charges. There are eight gluons which are the force carriers of the strong interaction, The eight different gluons carry different color charges referred to as red(r), blue(b), and green(g).

The SM predicted many new particles, including top quark, τ-lepton, gluon, $W^{\pm}$, and Z bosons. They were all discovered in colliding beam experiments. The properties of these particles were measured with high precision and agree well with SM predictions. One scalar

boson (spin-0), the Higgs boson (H^0), was introduced in the SM to break the electroweak symmetry to generate particle masses. It was the last SM particle discovered in 2012 by the experiments, of ATLAS and CMS, at the LHC [5,6]. Great efforts have been made to measure the Higgs boson properties in the past 10 years. So far the measured properties are consistent with the SM prediction. This thesis topic, "observation of vector boson scattering" was one of the major physics goals at the LHC to study electroweak symmetry breaking dynamics and measure the high energy behavior of the Higgs boson interactions with massive bosons.

The major properties of these particles (mass, charge, spin) are summarized in Table 1.1.

Table 1.1: Particles of the SM listed along with their symbol and major properties [7]. One should note that neutrinos are massless in the SM. (*The graviton G is not considered as part of the SM.)

		Name	Symbol	Generation	Charge	Spin	Mass [MeV/c^2]
Fermions	Leptons	Electron	e	1st	-1	1/2	0.511
		Muon	μ	2nd	-1	1/2	105.7
		Tau	τ	3rd	-1	1/2	1776.8
		Electron Neutrino	ν_e	1st	0	1/2	$< 2 \times 10^{-6}$
		Muon Neutrino	ν_μ	2nd	0	1/2	$< 2 \times 10^{-6}$
		Tau Neutrino	ν_τ	3rd	0	1/2	$< 2 \times 10^{-6}$
	Quarks	Up	u	1st	2/3	1/2	2.2 ± 0.5
		Charm	c	2nd	2/3	1/2	$1.275 \pm 0.035 \times 10^3$
		Top	t	3rd	2/3	1/2	$173.0 \pm 0.4 \times 10^3$
		Down	d	1st	-1/3	1/2	4.7 ± 0.5
		Strange	s	2nd	-1/3	1/2	95 ± 9
		Bottom	b	3rd	-1/3	1/2	$4.18 \pm 0.04 \times 10^3$
Bosons		Photon	γ		0	1	$< 1 \times 10^{-24}$
		Gluon	g		0	1	0
		Z boson	Z		0	1	91.1876×10^3
		W boson	$W^\pm$		± 1	1	80.39×10^3
		Higgs boson	H		0	0	125.18×10^3
		Graviton*	G		0	2	$< 1 \times 10^{-38}$

Elementary fermions compose matters in the universe. In the SM, there are 24 different elementary fermions (if their anti-particles are counted separately). They are six different flavor quarks in three generations: up and down, strange and charm, bottom, and top, six leptons (electron, muon, tauon, and corresponding neutrinos), along with the corresponding antiparticles. Leptons carry integer electric charges, while quarks carry a fractional integer electric charge. Quarks also carry color charges (r, g, b) and they are bounded together by gluons, which means the isolated quark doesn't exist (color confinement).

There is a hypothetical elementary boson with spin-2, the graviton (G), which might be

the force carrier of gravity. But gravity is not included in the SM. Therefore, the graviton is not an SM particle and it hasn't been observed yet.

1.1.2 Interactions

Interactions between particles are also very important in our understanding of how the universe is formed and evolved. There are four fundamental interactions, also called fundamental forces. They are strong, weak, and electromagnetic (EM) interactions, as well as gravity. The EM and gravitational interactions have an infinite range and can be observed at the macro-scale in our daily lifes. while the strong and weak forces are short-range interactions. Their interaction ranges are at a subatomic distance level, and can only be observed in particle physics experiments. Fig. 1.2 shows examples of electromagnetic, strong, and weak interactions described by the Standard Model.

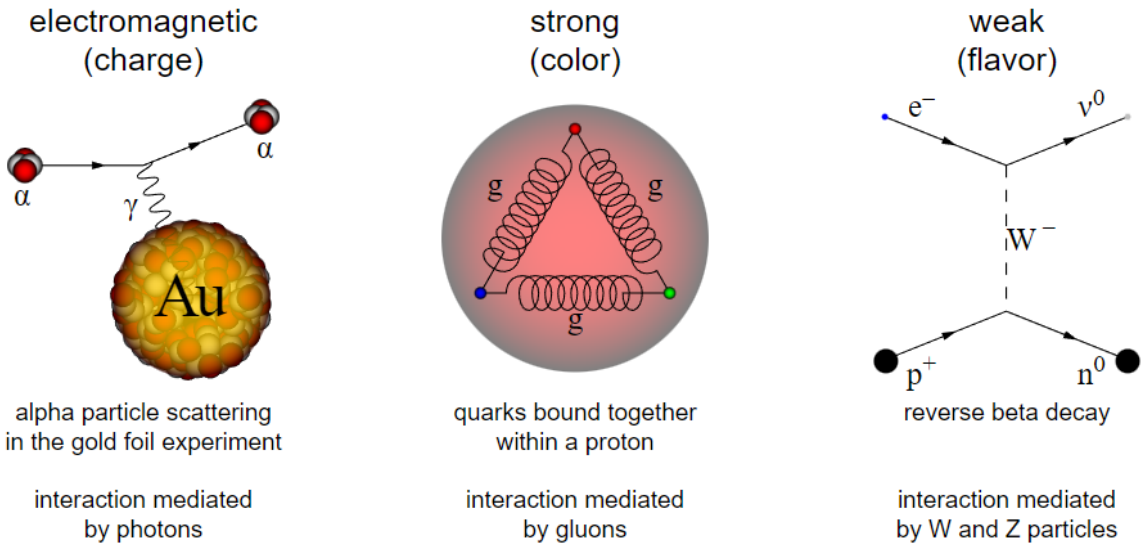


Figure 1.2: Examples of electromagnetic, strong, and weak interactions are described by the Standard Model.

The electromagnetic, strong, and weak force carriers are discrete quantum fields in particle physics theory, which means the interactions between particles are through exchanging the mediate particles (force carriers). The strong interaction is mediated by gluons and can be found between quarks, as well as hardons, which are composed of quarks. The weak interaction is carried by $W^{\pm}$ and Z bosons and is responsible for nuclear beta decay. The electromagnetic interactions are mediated by photon (γ), which is responsible to form atoms between electrons and nucleus, as well as molecules of matter. Unlike the three interactions

described above, the mediate particle of gravity called the graviton, hasn't been observed. According to Einstein's general theory of relativity, gravitation comes from the curvature of space-time. Table 1.2 summarizes the force carriers and properties of the fundamental interactions.

Table 1.2: Fundamental force carriers, relative strength, and the force range

interaction	theory	force carrier	relative strength	range (m)
Weak	Electroweak theory (EWT)	$W^{\pm}$, Z	10^{25}	10^{-18}
Strong	Quantum Chromodynamics (QCD)	gluons (g)	10^{38}	10^{-15}
Electromagnetic	Quantum Electrodynamics (QED)	γ	10^{36}	∞
Gravity	General relativity (GR)	graviton (hypothetical)	1	∞

1.2 Units used in Particle Physics

According to Einstein's mass-energy equivalence, $E = mc^2$, massive particles (such as the Higgs boson, top quark, etc) can be produced by high-energy particle collisions. Energy can be converted into mass and create new heavy unseen particles in colliding beams. In particle physics, a unit called electron-volt eV is used to describe the energy that a particle acquires. By definition, an electron-volt is the amount of kinetic energy gained by a single electron accelerating from rest through an electric potential difference of one volt in the vacuum. One electron-volt equals $1.602176634 \times 10^{-19}$ J under the 2019 redefinition of the standard international system (SI) base units. The particle mass and momentum can also use the energy unit of eV/c^2 and eV/c, respectively. In special relativity, the energy(E), momentum(p), and mass(m) of a particle are related by the equation

$$E^2 = (pc)^2 + (mc^2)^2. \tag{1.1}$$

At the particle rest frame (p=0), the above equation becomes $E = mc^2$. Therefore, particle mass is considered as the rest energy of the particle. It is an invariant quantity in 4-dimensional space-time Lorentz transformations, therefore we sometimes also call it as invariant mass.

One kilo-electronvolt (KeV) is one thousand electron-volts. Our room temperature is approximately KeV scale. One mega-electronvolt (MeV) is one million electron-volt. The rest energy of an electron is roughly 0.5 MeV and the rest energy of an up-quark is 1.9 MeV. Giga-electronvolt (GeV) and tera-electronvolt (TeV) are also widely used in particle physics. The rest mass of the heaviest observed elementary particle, the top quark, is about 172 GeV. Collision energy in the most powerful collider, the Large Hadron Collider, is designed to collide proton-protons at a center of mass energy of 14 TeV.

In particle physics studies we often use "natural unit" by taking $c = \hbar = \kappa_b = 1$, where $\hbar$ is Planck's constant $\hbar = \frac{h}{2\pi} = 1.05457 \times 10^{-34}$ J s $= 6.58212 \times 10^{-16}$ eV s, and κ_b is the Boltzmann constant, which equals 1.38×10^{-23} J/K. It is clear that in nature unit, energy, momentum, and mass are all in units of eV.

Using nature unit, there is only one fundamental dimension, which we can take to be energy and express it in eV, or KeV, or MeV, or GeV to various powers. For example,

$$\hbar c = (6.582x10^{-16}\ eV\ s)(3 \times 10^{8}\ m/s) = 19.73 \times 10^{-8}\ eV\ m = 197.3\ MeV\ fm, \quad (1.2)$$

where $1\ fm = 10^{-15}\ m$, which is the size of a proton. Using $\hbar c = 1 = 0.1973\ GeV\ fm$, we can easily express the distance and time in units of GeV^{-1}:

$$1\ fm = 5.058\ GeV^{-1}, \quad and \quad 1\ sec = 1.517 \times 10^{24}\ GeV^{-1}. \quad (1.3)$$

Temperature can also be expressed in units of eV/κ_B. Table 1.3 summarizes the relationship between the energy unit eV and the conversation to the SI unit of basic physics quantities.

Table 1.3: Expression of physics measurements in units related to electron-volt (eV). In the natural unit, $c = \hbar = \kappa_b = 1$.

Measurement	Energy unit	Conversion in SI unit of the energy unit
Energy	eV	1.60218×10^{-19} J
Mass	eV/c^2	1.78266×10^{-36} kg
Momentum	eV/c	5.34429×10^{-28} kg·m/s
Temperature	eV/k_B	1.16045×10^{4} K
Time	$\hbar$/eV	6.58212×10^{-16} s
Distance	$\hbar c$/eV	1.97327×10^{-7}

In colliding beam experiments, another commonly used variable is the center-of-mass energy, represented as $\sqrt{s}$, of the collider. For example, the LHC was operated at $\sqrt{s} = 13$ TeV in Run II. In fact, s is one of the Mandelstam variables [8] used in theoretical calculations. There are three Mandelstam variables s, u, and t. For one collision that has two incoming

particles and two outgoing particles after the collision, $s = (p_1 + p_2)^2 = (p_3 + p_4)^2$, $t = (p_1 - p_3)^2 = (p_2 - p_4)^2$ and $u = (p_1 - p_4)^2 = (p_3 - p_2)^2$, where p_1 and p_2 are four-momentum of two incoming particles, while p_3 and p_4 are four-momentum of two outgoing particles. It is clear that $\sqrt{s} = \sqrt{(p_1 + p_2) \cdot (p_1 + p_2)}$, which is the total energy of the system in the central mass frame of a collider.

The transverse momentum p_T is also widely used in physics analysis. Transverse momentum is defined as the momentum projection in the direction perpendicular to the beam line. Since the initial system before collision has a total momentum of zero, the negative sum of all transverse momenta of the produced particles after the collision is defined as the missing transverse energy, which is denoted as E_T^{miss}. Missing transverse energy is often caused by particles, such as neutrinos, escaping from the detection.

1.3 Cross-section and Luminosity

The most important quantity to describe an interaction process is the "cross-section" (σ), which represents the probability of the reaction between particles. The cross-section is determined by physics (the interaction that causes the particle scattering) and is independent of specific experimental detectors. It is a physical quantity with the dimension of area, and the commonly used unit is the barn (b) defined as

$$1\ b = 10^{-24}\ cm^2. \tag{1.4}$$

Roughly speaking, one barn equals the area of a uranium atom. In high-energy physics, commonly used units are nanobarn (nb), picobarn (pb), and femtobarn (fb).

$$1\ mb = 10^{-3}b, \quad 1\ \mu b = 10^{-6}b, \quad 1\ nb = 10^{-9}b, \quad 1\ pb = 10^{-12}b, \quad 1\ fb = 10^{-15}b. \tag{1.5}$$

Using the reduced Planck's constant, barn can also be expressed in units related to electron-volt using conversion $\hbar^2 c^2/\mathrm{GeV}^2 = 0.3894$ mb.

In colliding beam experiment the event rate from a certain physics interaction process is proportional to the underlying physics cross-section. The number of events produced also depends on the luminosity (L) of the collider, which is the beam particle flux rate in unit of $cm^{-2}s^{-1}$. Luminosity describes the number of particles passing through one unit area per unit time. The most commonly used unit for time-integrated luminosity is inverse femtobarn (fb^{-1}). The conventional unit for time-integrated luminosity is often used to describe the data size collected by a detector. The ATLAS detector collected data with time-integrated luminosity 139 fb^{-1} during Run II, which means we expect to get 139 events from a physics

process whose cross-section is 1 fb. Therefore, the number of events for a certain interaction process produced in a time period equals cross-section times the integrated luminosity,

$$N_{event} = \sigma \times \int Ldt. \tag{1.6}$$

We often use units of fb and fb^{-1} for cross-section and integrated luminosity, respectively, in the calculations. The cross-section unit in barns and the conversion to the energy unit are given in Table1.4.

Table 1.4: Re1ation between electron-volt and barn

unit	barn	GeV^{-2}
	1 mb	2.56819
	1 pb	2.56819×10^{-9}
	0.389379 mb	1
	0.389379 pb	10^{-9}

1.4 Thesis Topics of physics analyses

In this thesis, two analyses are presented. Both analyses are based on data collected by the ATLAS detector at the LHC at CERN with an integrated luminosity of 139 fb^{-1}. The data were recorded with the ATLAS experiment from proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV during the LHC Run II program (2015 - 2018). This thesis focuses on four-lepton final states for both analyses.

The first analysis is the observation of a rare ZZ production associated with two jets via electroweak interaction. The goal of this analysis is to study the electroweak symmetry-breaking dynamics through the vector boson scattering (VBS) in the $ZZjj$ channel. The experimental signature is in the four-lepton (4ℓ and $2\ell2\nu$) final states associating with two forward and backward hadronic jets. The second analysis is to search for a new gauge boson Z' predicted by theoretic models beyond SM (BSM) at a relatively low mass range in the four-muon final state.

Two analyses were conducted with similar event selections. However, the physics goals of the two analyses are different. The first study is a measurement of a physics process predicted by the SM, but never observed before, while the second study is looking for new physics signals. Multi-variate methods are used in the two analyses, but the techniques are different. The first analysis used "Gradient boosted decision trees". The second analysis

used a relatively new model, "parameterized neural network", which is widely used in high energy physics recently.

Results from these two analyses have been published as the ATLAS physics papers:

1 "Observation of electroweak production of two jets and a Z-boson pair", published on Nature Physics 19, 237–253 [9].

2 "Search for a new Z' gauge boson in 4μ events with the ATLAS experiment", accepted for publication on JHEP (arXiv:2301.09342) [10].

The organization of the thesis is the following. Chapter 2 briefly describes the particle physics theory; Chapter 3 describes the LHC and the ATLAS detector; Chapter 4 presents the physics object reconstruction with the ATLAS detector; Chapter 5 introduces the Statistical models used in analyses; Chapter 6 reports the analysis and results on vector-boson-scattering with the $ZZjj$ production; Chapter 7 reports the Z' search method and results. The summary and prospects are given in the last chapter.

Chapter 2

Theory

In mathematical physics, Yang–Mills theory [11] is a gauge theory based on a special unitary group $SU(N)$, or more generally any compact, reductive Lie algebra. Yang–Mills theory seeks to describe the behavior of elementary particles using these non-abelian Lie groups and is at the core of the unification of the electromagnetic force and weak forces (i.e. $SU(2) \times U(1)$) as well as Quantum Chromodynamics, the theory of the strong force (based on $SU(3)$). Thus it forms the basis of our understanding of the Standard Model of particle physics.

The name of the Standard Model is given in the 1960s - 1970s to a theory of fundamental particles and how they interact. It is based on quantum field theory and incorporated all that was known about subatomic particles at the time and predicted the existence of additional new particles as well. The SM development has been over the course of more than half of the twentieth century, sometimes driven forward by new experimental discoveries and sometimes by theoretical advances. The theory is established with a close interplay between experimental discoveries and new theoretical ideas. The cornerstones of the theory are based on two principles, gauge invariance, and spontaneous symmetry breaking, which will be described in this chapter. The SM has successfully explained and predicted a wide variety of experimental results. It summarizes our current best understanding of particle physics with an elegant mathematical presentation.

There are seventeen named elementary particles in the SM, including 12 matter particles: 6 quarks and 6 leptons, 4 gauge bosons, and 1 scalar, organized into the chart shown in Fig. 1.1. The SM predicted particles, top quark, τ-lepton and its associated neutrino (ν_τ), gluon, $W^\pm$ and Z, and Higgs boson were all observed in particle physics colliding beam experiments. The last particles discovered were the W and Z bosons in 1983 at CERN, the top quark in 1995 at Fermilab, the tau neutrino in 2000 at Fermilab, and the Higgs boson in 2012 at CERN. A brief description of the SM particles' properties and fundamental

forces acting on particles have been introduced in Section 1.1. The dynamics of the particle interactions and the SM theoretical formulation are described in this chapter.

The SM combines quantum mechanics, relativistic kinematics, and conservation laws observed in nature. The theory is constructed with a Lagrangian density function which contains the kinematic terms of the gauge fields and particles, and the interaction vertices. The particle's equation of motion (physics law) can be derived from the Lagrangian. The theory (Lagrangian) must be invariant under symmetry transformations. The gauge fields (the quanta are force carriers) and interactions are naturally introduced to the frame of the theory, Lagrangian, by the requirement of gauge invariance when the particle wave function $\psi(x)$ transforms with a local phase angle $\alpha(x)$, such as $\psi(x) \rightarrow \psi'(x) = e^{i\alpha(x)}\psi(x)$ (where x is 4-dimensional space-time coordinates).

The SM contains theories of QED (Quantum Electrodynamics), QCD (Quantum Chromodynamics), and Electroweak (EW) theory. They are gauge theories to describe electromagnetic, strong, and weak interactions, based on gauge symmetry groups, $SU(3)_C \times SU(2)_L \times U(1)_Y$, where C, L, and Y denote "Color", "Left-handed", and "Hyper-charge", respectively. QED is the first successful gauge theory based on U(1) symmetry developed in the 1950s by Shinichiro Tomonaga [12], Julian Schwinger [13, 14], Richard Feynman [15–17], and Freeman Dyson [18]. It combines Maxwell's theory of electromagnetism (EM), special relativity, and Quantum Field theory and describes all interactions of light with matter, and those of charged particles with one another. The force carrier of EM interactions is a massless photon γ. Because the behavior of atoms and molecules is primarily electromagnetic in nature, all of atomic physics can be considered a test laboratory for the theory. QED has been tested with extremely high precision in electron $g-2$ experiment [19]. The development of QCD followed very successful QED, based on $SU(3)_C$ symmetry transformation to describe the strong interactions [20]. Glashow, Weinberg, and Salam unified the EM and weak interactions and developed EW theory based on $SU(2)_L \times U(1)_Y$ symmetry transformations [21] [22] [23]. Gauge invariant requires massless gauge bosons and fermions, which contradicts the observations, weak gauge bosons W and Z and quarks and charged leptons are massive. To resolve the particle mass origin problem in theory the Higgs mechanism, a spontaneous electroweak symmetry breaking is introduced in the SM to generate masses of the elementary particles by Peter W. Higgs, F. Englert, and R. Brout in 1960s [24] [25].

The SM has been tested in particle physics experiments with extremely high precision. However, gravity is not included in the SM and it cannot explain dark matter and dark energy observed in the universe with astrophysics experiments. Also, there are some anomalies observed in experiments deviating at the 3-4σ level from the SM predictions. There are many theoretical models extending the SM aiming to address physics questions that cannot

be answered by SM. Searching for new physics predicted by BSM becomes one of the major physics goals of particle physics research.

The mathematical presentation of the SM QCD and EW theory, and BSM theory related to this thesis work are described in this chapter.

2.1 Quantum Chromodynamics

The modern theory of strong interactions is Quantum Chromodynamics (QCD), which is a gauge theory as part of the SM. The basic ingredient of QCD is that each of the six flavors or types of quark, $u,\ d,\ s,\ c,\ b$, and t has an additional quantum number, color, which takes the values of red(r), green(g), and blue(b), and they are called color charges of strong interactions. Anti-quarks carry anti-colors, $\bar{r},\ \bar{g}$, and $\bar{b}$. Each quark's wave function is presented in 3 components of colors,

$$q \ = \ \begin{pmatrix} r \\ g \\ b \end{pmatrix}. \tag{2.1}$$

There is an unbroken non-chiral SU(3) gauge symmetry acting on three color charges. Thus, there are 8 massless gauge bosons (gluons), $G^i(i = 1,\ 8)$, and each gluon carries two different color charges. They can be represented by the combination of r, g, b colors in table 2.1.

Table 2.1: Eight gluons in the QCD theory of the SM

$(r\bar{b} + b\bar{r})/\sqrt{2}$	$-i(r\bar{b} - b\bar{r})/\sqrt{2}$
$(r\bar{g} + g\bar{r})/\sqrt{2}$	$-i(r\bar{g} - g\bar{r})/\sqrt{2}$
$(b\bar{g} + g\bar{b})/\sqrt{2}$	$-i(b\bar{g} - g\bar{b})/\sqrt{2}$
$(r\bar{r} - b\bar{b})/\sqrt{2}$	$(r\bar{r} + b\bar{b} - 2g\bar{g})/\sqrt{2}$

Quarks interact through gluons by exchanging the color charges, as shown in Fig. 2.1. A strong gauge coupling g_s constant is introduced in the QCD Lagrangian which gives the strong fine structure constant $\alpha_s = g_s^2/4\pi$ as a measure of the strong interaction strength.

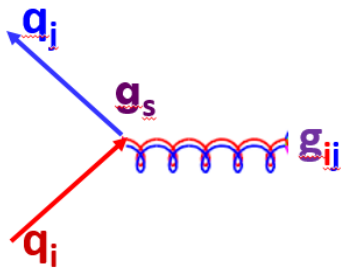


Figure 2.1: Quark-quark interaction through a gluon exchange. Each quark carries a different color charge, and the gluon carries two different colors. In the interaction vertex, the color charge is conserved. g_s shown in the diagram is the strong coupling constant.

The matrices representation of generators of SU(3) group are shown in equation 2.2.

$$G^i = \frac{\lambda^i}{2} \tag{2.2}$$

In equation 2.2, λ^i are the Gell-Mann matrices, which are a set of eight linearly independent 3×3 traceless Hermitian matrices. The eight Gell-Mann matrices are shown in equation 2.3.

$$\lambda^1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda^2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\lambda^4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \lambda^5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} \tag{2.3}$$

$$\lambda^6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \lambda^7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \quad \lambda^8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

The QCD Lagrangian is constructed by the gluon field kinematics, the quark field kinematics, and interactions with the gluon field, as well as the quark mass terms:

$$L_{QCD} = -\frac{1}{4} G^i_{\mu\nu} G^{i\mu\nu} + \sum_r \bar{q}^\alpha_r i \not{D}^\beta_\alpha q_{r\beta} - \sum_r m_r \bar{q}^\alpha_r q_{r\alpha} \tag{2.4}$$

where $G_{\mu\nu}$ denotes the field tensor, which can be expanded by using the SU(3) symmetry

group structure constant f_{ijk} as:

$$G^i_{\mu\nu} = \partial_\mu G^i_\nu - \partial_\nu G^i_\mu - g_s f_{ijk} G^j_\mu G^k_\nu, \tag{2.5}$$

and q denotes the quark field. The structure constants f_{ijk} are completely antisymmetric in the three indices. The values of structure constants can be calculated using the Gell-Mann matrices shown in 2.3. Non-zero values of the structure constants are shown in table 2.2.

$$[\lambda^i, \lambda^j] = 2i f_{ijk} \lambda^k \tag{2.6}$$

Table 2.2: Non-zero values of SU(3) structure constants (completely anti-symmetric)

$f_{123} = 1$	$f_{345} = \frac{1}{2}$	$f_{147} = \frac{1}{2}$
$f_{367} = -\frac{1}{2}$	$f_{156} = -\frac{1}{2}$	$f_{458} = \frac{\sqrt{3}}{2}$
$f_{246} = \frac{1}{2}$	$f_{678} = \frac{\sqrt{3}}{2}$	$f_{257} = \frac{1}{2}$

The gauge covariant derivative of the quark field is given by:

$$D^{\mu\beta}_\alpha = (D^\mu)_{\alpha\beta} = \partial^\mu \delta^\beta_\alpha + \frac{i g_s}{\sqrt{2}} G^{\mu\beta}_\alpha. \tag{2.7}$$

Inserting the expression of the covariant derivative 2.7 into the Lagrangian function we see that the first term represents the energy-momentum of the quarks and the second term represents the quark-gluon interactions with a coupling constant g_s which is related to the bare QCD coupling constant α_s:

$$\alpha_s = \frac{g_s^2}{4\pi} \tag{2.8}$$

The constant discussed above is called the bare coupling constant because the higher-order corrections are not taken into account. After absorbing the higher-order corrections, the coupling constant then becomes:

$$\alpha_s(\mu_R^2) = \frac{g_s^2(\mu_R^2)}{4\pi} \tag{2.9}$$

where μ_R is called the re-normalization scale. Normally μ_R can be treated as the energy scale, such as $\mu_R^2 = |q|^2$ where q is the four-momentum carried by intermediate gluons. Using Q^2

referring to the QCD interaction energy, the renormalization group equation is expressed as:

$$\frac{dg_s^2}{dlnQ^2} \equiv 4\pi\beta(g_s^2) = bg_s^4 + O(g_s^6) + \cdots \tag{2.10}$$

which contains the one-loop (the first term), two-loop (the 2nd term), and so on. For QCD, the coupling constant as a function of energy Q^2 is calculated as:

$$\alpha_s(Q^2) = \frac{1}{b_2^{n_q} ln\frac{Q^2}{\Lambda^2}}, \tag{2.11}$$

where n_q is the number of quarks that are lighter than Q and b_2 is the two loop term in the β function. Λ is the energy scale. The observed running corresponds to $\Lambda \sim 100 - 400$ MeV.

The coupling constant α_s becomes very small when Q^2 is large, which means α_s becomes small at a short distance. Therefore we call $\alpha_s(Q^2)$ as the running coupling in QCD. When the distance is very small (or when quarks carry very large energy) gluons and quarks can be treated as weakly coupled at a short distance, which referred to as "asymptotic freedom". At lower energy ($<< 1\ GeV$), the α_s becomes very large. In this case, the perturbation theory no longer works for QCD. This fact means the strong interaction will become very strong at low energies (or large distances) which leads to the confinement of particles carrying color charges. Experimentally, we have never observed free quarks and gluons, they are confined onsite hadrons, the bound-states of quarks.

QCD plays an essential role in cross-section calculations at the LHC. The protons in the initial state are a 'bag' of 'free' quarks and gluons, called partons, their energy and momentum are characterized by the parton density distribution functions measured in experiments,

$$f_i^{(a)}(x_a, Q^2), \tag{2.12}$$

which represents number of density of species i with momentum fraction x_a of hadron (or proton) a seen by probe with resoling power Q^2 given by QCD perturbation theory. Fig. 2.2 shows the measured PDF functions.

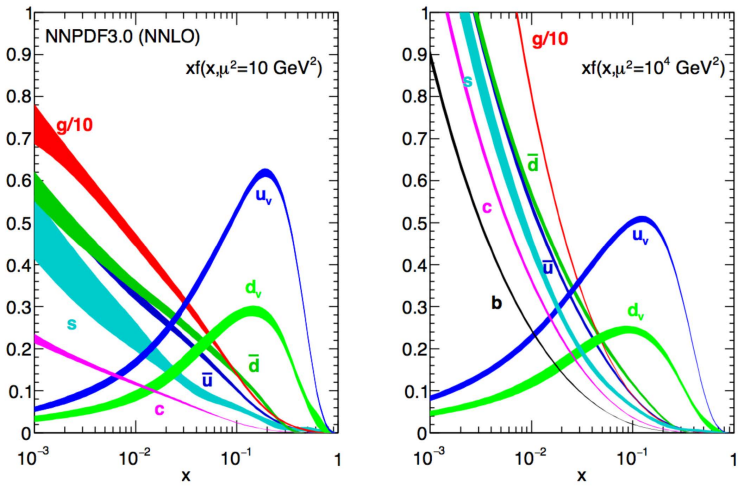


Figure 2.2: The NNPDF3.0 NNLO PDF set for $Q^2 = 10$ GeV^2 and $Q^2 = 104$ GeV^2 [26].

The quarks and gluons are fragmented into hadronic jets (formed by highly boosted hadrons) after they are produced through QCD hard scattering process. The basic hadronization process is shown in Fig 2.3. Hadronic jets are measured in experiments as important physics objects in analysis.

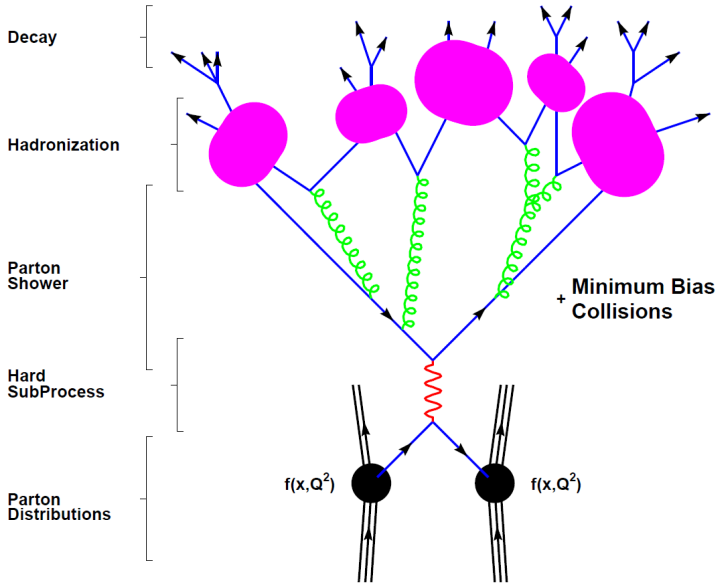


Figure 2.3: Basic structure of a QCD hard scattering at the LHC and the quark hadronization process. The pink-colored objects represent the final state color-free hadrons.

2.2 The electroweak theory

The Electroweak theory unified two fundamental interactions: electromagnetic interaction and weak interaction. The theory development started after the parity violation of weak interaction discovery in 1956 [27], physicists began to search for theories that can relate electromagnetic and weak interaction together. In 1964, Abdus Salam and John Ward [22] came up with a theory that predicts a massless photon and three massive gauge bosons with manually broken symmetry later Weinberg [21] had a similar idea that a set of symmetries produce the electroweak force. Weinberg also predicted rough masses of the W and Z boson and suggested the new electroweak theory is renormalizable. Sheldon Glashow [23] worked out the correct symmetry group to unify the two interactions. Steven Weinberg, Abdus Salam, and Sheldon Glashow were awarded the 1979 Nobel Prize in Physics because of their contribution to the unification of electromagnetic and weak interaction.

The EW theory is constructed based on symmetry group of $SU(2)_L \times U(1)_Y$. The symmetry group of transformation is generated by quantum numbers, I (isospin) and Y (hypercharge). The corresponding gauge bosons are $W^1_\mu, W^2_\mu, W^3_\mu$ for $SU(2)_L$ and $\mathcal{B}_\mu$ for $U(1)_Y$. The corresponding gauge coupling constants are g and g'. The hypercharge is defined as

$$Y = 2(Q - I_3), \tag{2.13}$$

that relates isospin 3rd component I_3 with the electric charge Q of the particle [28, 29]. By construction, left-handed leptons have $Y = -1$, and right-handed leptons have $Y = -2$. In the EW theory, the description of fermions is separated into the chiral left-handed and right-handed fields in the following form:

$$\begin{aligned} \psi_L(x) &= \frac{1}{2}(1 - \gamma^5)\psi(x) \\ \psi_R(x) &= \frac{1}{2}(1 + \gamma^5)\psi(x). \end{aligned} \tag{2.14}$$

The left-handed fields form the weak-isospin doublet L_f which include

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \quad \begin{pmatrix} u \\ d \end{pmatrix}_L, \quad \begin{pmatrix} c \\ s \end{pmatrix}_L, \quad \begin{pmatrix} t \\ b \end{pmatrix}_L, \tag{2.15}$$

where the projection of the third component of isospin , I_3, is $\frac{1}{2}$ and $-\frac{1}{2}$, respectively for two fields of the doublets. While the right-handed fields R_f (singlet) include

$$e_R, \quad \mu_R, \quad \tau_R, \quad u_R, \quad d_R, \quad c_R, \quad s_R, \quad t_R, \quad b_R. \tag{2.16}$$

Massless particles always have the same helicity and chirality and experiments observed only left-handed neutrino and right-handed antineutrino [27] [30]. Therefore massless neutrinos do not have a right-handed field in the SM.

The Lagrangian can then be written down in gauge fields and fermions terms:

$$\mathcal{L} = \mathcal{L}_{gauge} + \mathcal{L}_{fermions}. \tag{2.17}$$

The gauge term is given by

$$\mathcal{L}_{gauge} = -\frac{1}{4}\sum_{l=1}^{3} F^{l}_{\mu\nu}F^{l\mu\nu} - \frac{1}{4}f_{\mu\nu}f^{\mu\nu}, \tag{2.18}$$

where the field tensors are defined as

$$\begin{aligned} F^{l}_{\mu\nu} &= \partial_\mu W^{l}_{\nu} - \partial_\nu W^{l}_{\mu} - g\sum_{j,k=1}^{3}\epsilon_{jkl}W^{j}_{\mu}W^{k}_{\nu} \\ f_{\mu\nu} &= \partial_\mu \mathcal{B}_\nu - \partial_\nu \mathcal{B}_\mu, \end{aligned} \tag{2.19}$$

here the ϵ_{jkl} is structure constants of SU(2) symmetry group. And the gauge covariance derivatives acting on left-handed and right-handed fermions are

$$D_{\mu,L} = \partial_\mu + ig'\frac{Y}{2}\mathcal{B}_\mu + ig\frac{\tau}{2}\cdot W_\mu \tag{2.20}$$

$$D_{\mu,R} = \partial_\mu + ig'\frac{Y}{2}\mathcal{B}_\mu, \tag{2.21}$$

where g and g' are gauge coupling constants related to SU(2) isospin and U(1) hypercharge, respectively, and $\tau/2$ are 2×2 matrices as the $SU(2)$ generators, which are often denoted as L^i where $i = 1, 2, 3$. τ^2 are the Pauli matrices, which are shown in equation 2.22.

$$\tau^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tau^2 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \quad \tau^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.22}$$

This construction is consistent with the fact that charged-current weak interaction only affects left-handed fermions. Also, it would be difficult mathematically to incorporate 2-dimensional matrices with right-handed states where the neutrinos are missing. With those definitions above, and denoting the left-handed fermions as L_f and right-handed-fermions

as R_f, the fermion Lagrangian in the gauge covariance derivative form is

$$\begin{aligned}\mathcal{L}_{fermions} &= \bar{L_f} i\gamma^\mu D_{\mu,L} L_f + \bar{R_f} i\gamma^\mu D_{\mu,R} R_f \\ &= \bar{L_f} i\gamma^\mu (\partial_\mu + ig'\frac{Y}{2}\mathcal{B}_\mu + ig\frac{\tau}{2}\cdot W_\mu) L_f + \bar{R_f} i\gamma^\mu (\partial_\mu + ig'\frac{Y}{2}\mathcal{B}_\mu) R_f,\end{aligned} \tag{2.23}$$

where the L_f and R_f are summed over all fermion fields concerned. One should notice that there is no fermion mass term, $\mathcal{L}_m = -m(\bar{L}_f R_f + \bar{R}_f L_f)$, included in the Lagrangian since the fermion mass term violates gauge invariance.

As described previously the four gauge bosons in EW theory are massless. However, weak force mediators $W^\pm$ and Z bosons are massive. To solve the mass generation problems, the Higgs mechanism was introduced in the EW theory, and will be described in the Section followed.

2.3 Higgs mechanism

To generate the masses of fermions and gauge bosons in the EW theory, the Englert–Brout–Higgs mechanism (also referred to as Higgs mechanism) [24] [25]was proposed in the middle of 1960 by introducing a complex doublet of spin-0 scalar Higgs field in theory:

$$\phi = \begin{pmatrix}\phi^+ \\ \phi^0\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}\phi^1 - i\phi^2 \\ \phi^3 - i\phi^4\end{pmatrix}. \tag{2.24}$$

The Lagrangian of the Higgs scalar field can be written down generally in the EW theory:

$$\mathcal{L}_\phi = (D_L^\mu \phi)^\dagger (D_{\mu,L}\phi) - V(\phi), \tag{2.25}$$

where $D_{\mu,L}$ is the same left-handed gauge covariant derivative,

$$D_{\mu,L} = \partial_\mu + ig'\frac{Y}{2}\mathcal{B}_\mu + ig\frac{\tau}{2}\cdot W_\mu, \tag{2.26}$$

and the potential energy term is

$$\begin{aligned}V(\phi) &= \mu^2(\phi^\dagger\phi) + |\lambda|(\phi^\dagger\phi)^2 \\ &= \frac{1}{2}\mu^2(\sum_{i=1}^{4}\phi_i^2) + \frac{1}{4}\lambda(\sum_{i=1}^{4}\phi_i^2)^2,\end{aligned} \tag{2.27}$$

the two parameters introduced in the scalar Higgs field potential are λ and μ, which represent the strength of the self-coupling term and mass of the scalar field, respectively.

Under certain choice of the vacuum state ϕ_0 the complex scalar and Higgs potential under the vacuum state can be converted to:

$$\begin{aligned} \phi \rightarrow \phi_0 = \langle 0|\phi|0\rangle &= \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ \nu\end{pmatrix} \\ V(\phi) \rightarrow V(\nu) &= \frac{1}{2}\mu^2\nu^2 + \frac{1}{4}\lambda\nu^4 \end{aligned} \tag{2.28}$$

From equation 2.28, when $\mu^2 > 0$, the minimum point of $V(\phi)$ is $\nu = 0$ and the gauge group $SU(2) \times U(1)$ is unbroken, while in case of $\mu^2 < 0$ the minimum point of $V(\phi)$ is no longer at $\nu = 0$. The new minimum point becomes $\nu = \sqrt{\frac{-\mu^2}{\lambda}}$ and the $SU(2) \times U(1)$ symmetry is broken spontaneously. Fig. 2.4 shows the potential energy shapes for $\mu^2 > 0$, and $\mu^2 < 0$ (every ground state breaks the symmetry).

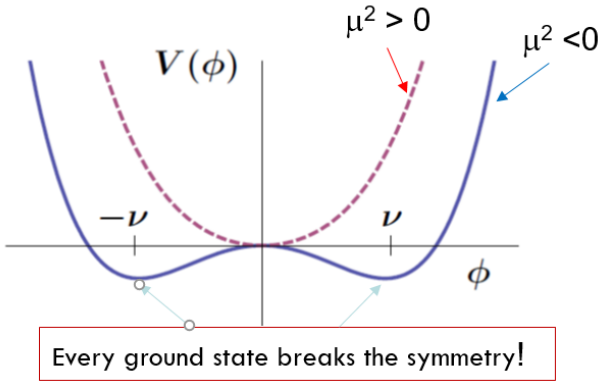


Figure 2.4: The potential energy shape for $\mu^2 > 0$ and $\mu^2 < 0$ (symmetry breaking in ground state).

At the minimum energy point (vacuum state), the generator related to L^1, L^2, and $L^3 - Y$ are also spontaneously broken, according to equation 2.29 and equation 2.22, which means the symmetry of weak interaction is broken and the W and Z boson will be massive.

$$L^i\phi_0 = \frac{\tau^i}{2}\frac{1}{\sqrt{2}}\begin{pmatrix}0\\ \nu\end{pmatrix} \neq 0,\ i = 1, 2, 3, \quad Y\phi_0 = \frac{I}{2}\frac{1}{\sqrt{2}}\begin{pmatrix}0\\ \nu\end{pmatrix} \neq 0 \tag{2.29}$$

As for the electromagnetic interaction, from equation 2.30, the electric charge Q remains 0 under the vacuum state. The electromagnetic interaction remains unbroken and the photon remains massless.

$$Q\phi_0 = (L^3 + Y)\phi_0 = 0 \tag{2.30}$$

The scalar field expansion around the ground state can be expressed as

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ \nu + H(x) \end{pmatrix} \tag{2.31}$$

where the Higgs filed $H(x)$ is the exited state of the scalar file $\phi(x)$.

Now we go back to examine the first term of the scalar Lagrangian 2.25 by inserting the expression of the scalar field in equation 2.31,

$$T(\phi) = (D_\mu \phi)^\dagger D^\mu \phi = \text{Vacuum } (\nu) \text{ terms} + \text{Higgs terms.} \tag{2.32}$$

The vacuum terms will generate masses for $W^\pm$ and Z bosons, and the Higgs terms include the Higgs field kinematics and interactions of the Higgs field with gauge fields.

The vacuum term in equation 2.32 can be expressed as,

$$(D_\mu \phi)^\dagger D^\mu \phi = \frac{1}{2}(0 \ \nu)[\frac{g}{2}\tau^i W^i_\mu + \frac{g'}{2}B_\mu]^2 \begin{pmatrix} 0 \\ \nu \end{pmatrix}. \tag{2.33}$$

This equation can be rewritten by replacing W^i using $W^\pm$ and $\tau^\pm$, as defined in equation 2.34 below,

$$\begin{aligned} W^\pm &= \frac{W^1 \mp iW^2}{\sqrt{2}}, \quad \tau^\pm = \frac{\tau^1 \pm i\tau^2}{2} \\ \tau^i W^i &= \tau^3 W^3 + \sqrt{2}\tau^+ W^+ + \sqrt{2}\tau^- W^-, \end{aligned} \tag{2.34}$$

therefore, the vacuum terms become:

$$\begin{aligned} &\frac{g^2\nu^2}{4} W^{+\mu} W^-_\mu + \frac{1}{2}(g^2 + g'^2)\frac{\nu^2}{4}[\frac{-g'\mathcal{B}_\mu + gW^3_\mu}{\sqrt{g^2 + g'^2}}]^2 \\ &= M_W^2 W^{+\mu} W^-_\mu + \frac{M_Z^2}{2} Z^\mu Z_\mu, \\ &\text{where Z boson is defined as: } Z_\mu = \frac{-g'\mathcal{B}_\mu + gW^3_\mu}{\sqrt{g^2 + g'^2}}. \end{aligned} \tag{2.35}$$

In equation 2.35, W^+ and W^- are physical W bosons that mediate the weak charged current in the SM and Z_μ represents the physical Z boson. $W^\pm$ and Z bosons acquired masses via the interactions with the scalar Higgs field in vacuum where the symmetry is broken,

$$M_W = \frac{g\nu}{2}, \quad M_Z = \frac{g_Z \nu}{2} = \frac{M_W}{\cos\theta_W}, \tag{2.36}$$

where θ_W is the weak angle, which is defined as:

$$\tan\theta_W = \frac{g'}{g}, \quad \cos\theta_W = \frac{g}{g_Z}, \quad \sin^2\theta_W = 1 - \frac{M_W^2}{M_Z^2}. \tag{2.37}$$

The electromagnetic interaction gauge field, photon (A_μ), can also be rewritten using gauge fields $\mathcal{B}_\mu$ and W_μ^3, as shown in equation below:

$$A_\mu = \frac{g\mathcal{B}_\mu + g'W_\mu^3}{\sqrt{g^2 + g'^2}}. \tag{2.38}$$

The A_μ field (γ) is orthogonal to the Z_μ field and there is no mass term of A_μ in the Higgs Lagrangian, which indicates that the photon remains massless. In the equations discussed above, the constant vacuum expectation value, ν, is called the weak scale, which can be measured using the mass of W boson. ν can also be calculated using the Fermi constant G_F. The measured vacuum expectation value is around 246 GeV. As for the weak angle, θ_W can be calculated using the mass of the W and Z boson, according to equation 2.39 below,

$$M_W = M_Z \times \cos\theta_W \sim \frac{(\pi\alpha\sqrt{2}G_F)^{\frac{1}{2}}}{\sin\theta_W}. \tag{2.39}$$

We now re-write the complete scalar Lagrangian 2.25 using expressions 2.31, 2.34, and 2.36 as the following,

$$\begin{aligned} \mathcal{L}_\phi = &\ M_W^2 W^{\mu+}W_\mu^-(1+\frac{H}{\nu})^2 + \frac{1}{2}M_Z^2 Z^\mu Z_\mu(1+\frac{H}{\nu})^2 + \frac{1}{2}(\partial_\mu H)^2 - V(\phi) \\ V(\phi) = &\ -\frac{\mu^4}{4\lambda} - \mu^2 H^2 + \lambda\nu H^3 + \frac{\lambda}{4}H^4. \end{aligned} \tag{2.40}$$

It is clear that the interaction coupling between the Higgs field and the gauge fields are proportional to the mass-squared of the gauge bosons. The potential $V(\phi)$ terms include the Higgs boson mass and self-interactions. The second term of the Higgs potential $V(\phi)$ gives the Higgs mass (which predicts existence of a physical scalar particle):

$$M_H = \sqrt{-2\mu^2} = \sqrt{2\lambda}\nu \tag{2.41}$$

The third and fourth terms in the Higgs potential indicate the tri-Higgs and four-Higgs self-coupling vertices.

Fermions also acquire masses via the Higgs mechanism and spontaneous symmetry breaking. The SM introduced a Yukawa sector in the Lagrangian, which describes the interactions between fermions and Higgs boson, without breaking the $SU(2)_L \times U(1)_Y$ symmetry. The

interaction terms take the following form:

$$\mathcal{L}_{Yukawa} = -g_f \bar{\psi}_L \phi \psi_R + h.c., \tag{2.42}$$

where g_f is the Yukawa coupling constant between the Higgs field and a specific fermion, and the $h.c.$ means Hermitian conjugate of the form terms. Substituting ϕ with the Higgs ground state term (2.31) after the symmetry breaking, the Lagrangian becomes

$$\begin{aligned} \mathcal{L}_{Yukawa} &= -\frac{g_f \nu}{\sqrt{2}}(\bar{\psi}_L \psi_R + \bar{\psi}_R \psi_L) - \frac{g_f}{\sqrt{2}}(\bar{\psi}_L \psi_R + \bar{\psi}_R \psi_L) H \\ &= -m_f \bar{\psi}\psi - \frac{m_f}{\nu}\bar{\psi}\psi H, \end{aligned} \tag{2.43}$$

where $m_f \equiv g_f \nu/\sqrt{2}$ is the mass of the fermion. There are 9 massive fermions in the SM theory, so there are 9 Yukawa Lagrangian terms in the form of (2.42), each with different Yukawa constants that have to be determined by experiments. The second term of equation (2.43) describes the interaction between fermion and the Higgs boson. The interaction strength is proportional to the fermion mass-squared.

2.4 The CKM matrix

Physicists found that the quark mass eigenstates (described by QCD) are not exactly the same as the quark weak interaction eigenstate (described by EW theory). The EW Lagrangian (2.23) determines that the EW interaction only happens for particles within the same generation. However, the strangeness-up changing weak charged current is observed in experiments, indicating that the weak eigenstates are not exactly the same as the QCD eigenstates (or mass eigenstates) and there are mixing between the two. Without loss of generality in theory, by assuming there is no mixing of $I_3 = \frac{1}{2}$ quark states, the mixing of $I_3 = -\frac{1}{2}$ can be written down as the CKM matrix [31, 32], which shows the relation between quarks weak eigenstates and mass eigenstates and it is a unitary matrix. The weak eigenstate can be transformed from the mass eigenstate following the matrix equation:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.44}$$

The elements of the CKM matrix are measured in experiments and their values are given

below [33]:

$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 0.97370 \pm 0.00014 & 0.2245 \pm 0.0008 & 0.00382 \pm 0.00024 \\ 0.221 \pm 0.004 & 0.987 \pm 0.011 & 0.0410 \pm 0.0014 \\ 0.0080 \pm 0.0003 & 0.0388 \pm 0.0011 & 1.013 \pm 0.030 \end{pmatrix} \quad (2.45)$$

The mixing angles between quark mass and weak eigenstates make the flavor changing between different generations possible. As for leptons, the weak and mass eigenstates are the same, which means there is no mixing between different lepton mass eigenstates and weak eigenstates. Flavor changing between different types of leptons by exchanging weak gauge bosons are not in the SM. Violation of lepton flavors hasn't been observed yet.

The relative branching ratio of quark decays can be estimated from the corresponding row or column of the CKM matrix. For example, $|V_{tb}|$ is much larger than the other two elements $|V_{td}|$ and $|V_{ts}|$ in the third row of the CKM matrix, which means top quarks are mostly decay into bottom quarks through weak decays. For example, $t \to bW$.

2.5 The vector boson scattering

In addition to generating masses, the Higgs mechanism also addressed the unitarity problem in cross-section calculation of longitudinally polarized W and Z bosons scattering [34]. This thesis is to observe the vector boson scattering (VBS) process in the $ZZjj$ channel. The related physics background is described in this section.

Since the Higgs boson was observed by the CMS and ATLAS experiment in 2012, lots of studies have been performed on the Higgs boson and its interaction with other particles. The Non-zero mass of weak gauge bosons and fermions is only one of the reasons why we need the Higgs boson. The second reason is a problem concerning vector boson polarization and the issue of the unitarity of VBS. The unitarity condition is equivalent to the requirement of the sum of probabilities of all possible final states evolving from a particular initial state should equal one. The sum of probabilities must be in principle calculated to infinite order in perturbative expansion, which is of course impossible to achieve practically.

From the gauge boson sector of the EW theory Lagrangian described in section 2.2, the weak gauge bosons have triple gauge coupling (TGC) and quartic gauge couplings (QGC) vertices. As an example of VBS process, the Feynman diagrams of the WW scattering are shown in Fig. 2.5. Other VV ($V = W, Z$) scattering diagrams are similar.

The transition probability from the initial VV state to the final VV state can be calcu-

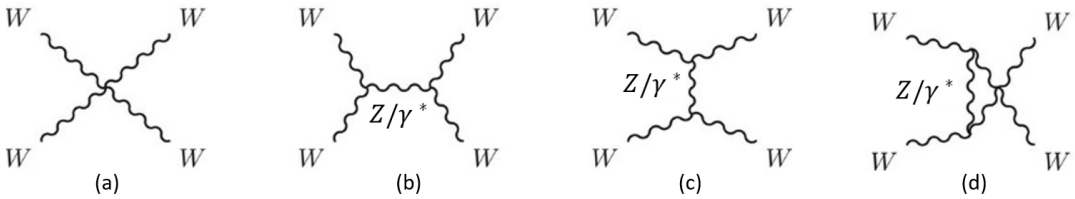


Figure 2.5: Feynman diagrams of WW scattering, including QGC (a) and TGC (b) - (c) vertices (in the $s,\ t$ and u channels.

lated based on quantum field theory:

$$|U_{fi}|^2 = |(2\pi)^4(\sum_k p_{f_k} - p_1 - p_2)\mathcal{M}_{fi}|^2, \tag{2.46}$$

where i and f denote "initial" and "final" states and k counts for the final state produced particles. The matrix $\mathcal{M}_{fi}$ is the transition amplitude from the initial state to the final state. In order to calculate the transition amplitude, we calculated the covariant form for the longitudinal polarization of four vectors $\epsilon_L(P)$ of the vector boson. We simplify the interaction as a $VV \to VV$ process and present each vector boson in its 4-dimensional energy-momentum vector: $(p_1,\ p_2) \to (k_1,\ k_2)$. In the center of mass (CM) frame, and considering the initial $VV = WW$, the polarization four-vector can be expressed as [35]:

$$\begin{aligned} \epsilon(p_1) &= \frac{p_1^\mu}{m_W} - \frac{2m_W}{s}p_2^\mu \\ \epsilon(p_2) &= \frac{p_2^\mu}{m_W} - \frac{2m_W}{s}p_1^\mu \end{aligned} \tag{2.47}$$

If the final state of VV is also WW, produced from Feynman diagrams with QGC and TGC vertices (including s, t, and u channels), the transition amplitude should be proportional to $\mathcal{M} \sim \epsilon_L\epsilon_L\epsilon_L\epsilon_L$ [36], which can be calculated as in reference [35]:

$$i\mathcal{M}_t^{\gamma+Z} = -i\frac{g^2}{4m_W^4}[(s-u)t - 3m_W^2(s-u) + \frac{8m_W^2}{s}u^2] \tag{2.48}$$

$$i\mathcal{M}_s^{\gamma+Z} = -i\frac{g^2}{4m_W^4}[s(t-u) - 3m_W^2(t-u)] \tag{2.49}$$

The transition amplitude (proportional to $(\mathcal{M}_s + \mathcal{M}_t)^2$) in $s+t$ channels will diverge very fast at a high energy scale as s^2:

$$i\mathcal{M}_{s+t}^{\gamma+Z} = -i\frac{g^2}{4m_W^4}[u^2 + O(s,\ t,\ u)]. \tag{2.50}$$

And the transition amplitude of QGC vertices are [35]:

$$i\mathcal{M}_4 = i\frac{g^2}{4m_W^4}[s^2 + 4st + t^2 - 4m_W^2(s+t) - \frac{8m_W^2}{s}ut]. \tag{2.51}$$

The sum of the transition amplitude of gauge vertices (s+t channel and QGC) is:

$$i\mathcal{M}_{gauge} = i\mathcal{M}_t^{\gamma+Z} + i\mathcal{M}_s^{\gamma+Z} + i\mathcal{M}_4 = -i\frac{g^2 u}{4m_W^2} + O((\frac{E}{m_W})^0) \tag{2.52}$$

From equation 2.50 to equation 2.52, the QGC vertices cancel part of the divergence at high energy, but the total transition amplitude is still proportional to the square of energy s. It will violate unitary at high energy $\sqrt{s} \sim 1$ TeV.

In the SM scalar Lagrangian (2.40), the Higgs boson can also interact with the W and Z boson. Additional Feynman diagrams of the WW scattering involving the Higgs boson are shown in Fig. 2.6. After including the Higgs boson, the transition amplitude of the Higgs

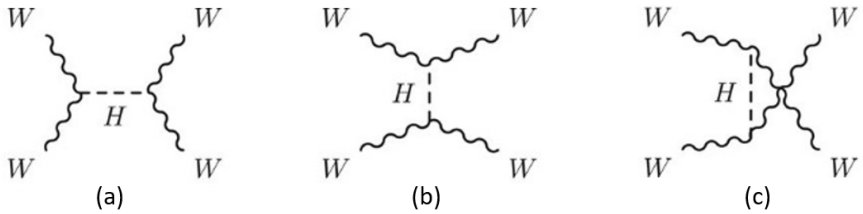


Figure 2.6: Feynman diagrams of WW scattering involving with Higgs boson interactions through the s, t and u channels.

diagram is (in the limit of $s, t \gg m_h$) given by [35]:

$$i\mathcal{M}_H = i\frac{g_h^2}{4m_W^2}u \tag{2.53}$$

Therefore, the total VBS amplitude (gauge boson and Higgs) is no longer diverting:

$$i\mathcal{M}_{\text{gauge}} + iM_H = \frac{g^2 m_h^2}{4m_W^2}. \tag{2.54}$$

The total transition amplitude becomes a constant after taking the contribution of Feynman diagrams involving the Higgs boson exchanging into account. The Higgs boson can cancel the divergence at high energy. The same arguments also can be applied to the opposite sign W boson scattering process [35], [36]. The corresponding transition amplitude can be

calculated as:

$$
\begin{aligned}
i\mathcal{M}_{gauge} &= -i\frac{g^2}{4m_W^2}u + O(s^0) \\
i\mathcal{M}_{H} &= i\frac{g_h^2}{4m_W^2}u + O(s^0)
\end{aligned}
\tag{2.55}
$$

Similarly, the lowest order Feynman diagrams of WZ scattering process are the WWZZ contact interaction [36]. Likewise, the divergence resulting from the sum of gauge-boson-only diagrams is exactly canceled by the Higgs diagrams.

For the ZZ scattering process, the statement is different [36], since it can only occur via the Higgs exchange (both s and t channel). However, in any real hardon-hardon experiment, this process cannot be separated from the dominant WWZZ process, where three additional graphs contribute in the lowest order, including the WWZZ contact interaction, t channel W exchange, and s channel Higgs exchange. Once again, the Higgs exchange provides cancellation of unwanted divergence in the ZZ scattering process. Fig.2.7 shows the theoretically calculated VBS cross-sections as a function of a center-of-mass energy E_{CM} for VV scattering processes with/without involving the Higgs boson.

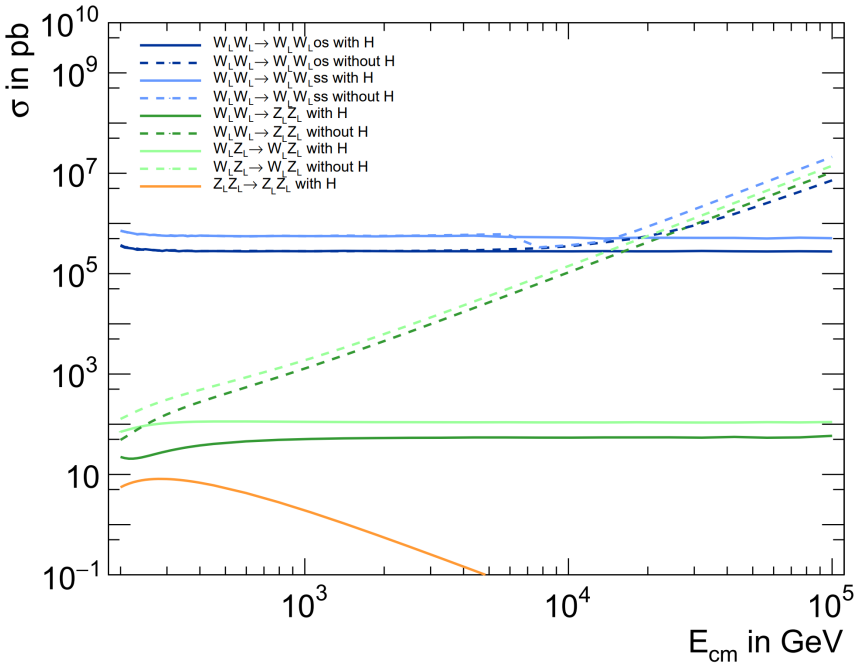


Figure 2.7: Cross sections of different vector boson scattering processes in the CM frame. The Higgs boson cancels divergence at high energy [37].

2.6 New gauge boson Z' predicted by $L_\mu - L_\tau$ models

The new gauge boson Z' is predicted by the highly motivated gauged theory, $L_\mu - L_\tau$ models [38], which is the simplest extension of the SM to address the observed g-2 of the muon anomalous magnetic dipole moment [39–42] and the B physics anomalies [43–46]. These models also aim to probe the particle physics and cosmology outstanding questions related to the dark matter and neutrino mass [47–49]. For a relatively low mass of Z', the most promising experimental signature would be a moderate excess of 4μ events with one $\mu^+\mu^-$ pair peaking around the Z' mass.

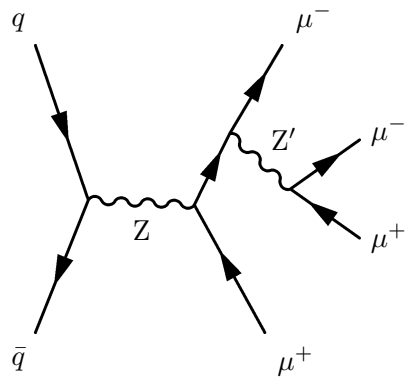


Figure 2.8: Tree-level Feynman diagram of the Z' production in radiation Drell-Yan process: $q\bar{q} \to Z \to \mu^+\mu^- Z' \to \mu^+\mu^-\mu^+\mu^-$. The Z' shown in the diagram is a new leptophilic vector boson predicted by the $L_\mu - L_\tau$ model.

The difference between the muon number and tau number, $L_\mu - L_\tau$, is one of the global symmetries present in the SM. The $U(1)_{L_\mu - L_\tau}$ symmetry is anomaly free and therefore can be gauged. The $U(1)_{L_\mu - L_\tau}$ symmetry is broken, resulting in a neutral, color singlet, massive Z' that couples only to μ and τ leptons and left-handed neutrinos at tree level. The Z' field kinematics, mass, and the muon and tau leptons are described by the Lagrangian below:

$$\mathcal{L}_{Z'} = -\frac{1}{4}F'_{\alpha\beta}F'^{\alpha\beta} + \frac{1}{2}M^2_{Z'}Z'^{\alpha}Z'_{\alpha} - g_{Z'}Z'_{\alpha}(\bar{\ell}_2\gamma^{\alpha}\ell_2 + \bar{\mu}\gamma^{\alpha}\mu - \bar{\ell}_3\gamma^{\alpha}\ell_3 - \bar{\tau}\gamma^{\alpha}\tau) \tag{2.56}$$

where $F'_{\alpha\beta} = \partial_\alpha Z'_\beta - \partial_\beta Z'_\alpha$ is the field strength tensor, and $\ell_i = (\nu_i, e_i)^T$ (i=2,3, denotes for the 2nd and the 3rd generation left-handed lepton doublet). μ and τ in the Lagrangian represent the right-handed charged muon and tau singlets. The parameters of the model is the mass $M_{Z'}$ and the coupling $g_{Z'} = \epsilon g'$. Here g' is the SM U(1) coupling constant ($g' = e/cos\theta_w = 0.357$), and ϵ is the scaled fraction of g'. The parameter space of $(M_{Z'}, g_{Z'})$ has not been strongly constrained in experiments since such Z' can not be directly produced

from the e^+e^- or the hadron colliders. Searching for the Z' has become a hot topic in both theoretical and experimental communities [50–52]. Fig. 2.8 shows the most promising Z' production through the radioactive Z boson decay process at the LHC.

Chapter 3

ATLAS at LHC

3.1 The Large Hardon Collider

3.1.1 Overview of the LHC

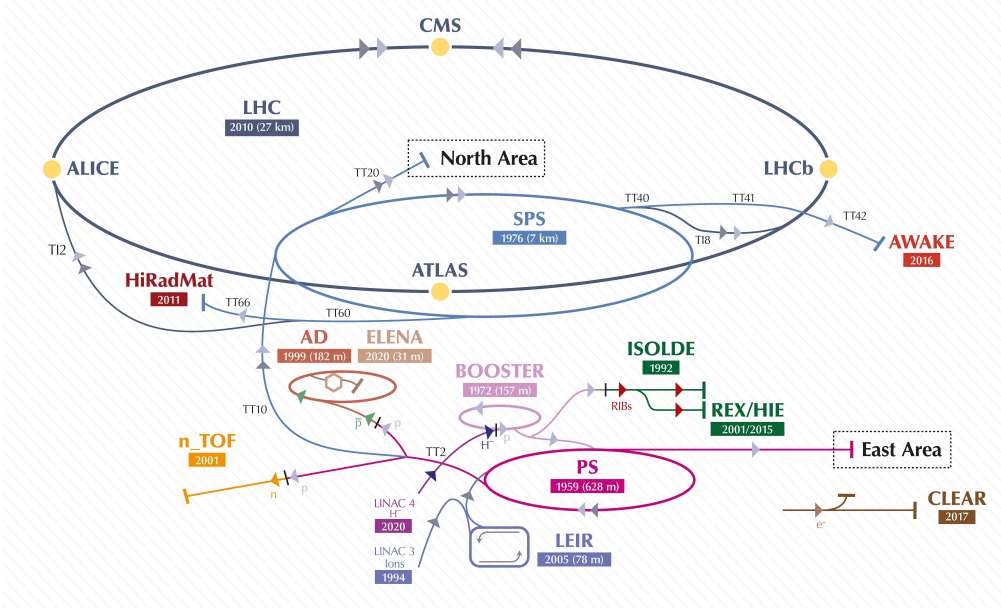
The Large Hadron Collider (LHC) is the biggest machine and the most powerful collider that humans ever built in history, which was built by the European Organization for Nuclear Research (CERN) from 1998 to 2008. It consists of a 27-kilometer ring of superconducting magnets with a number of accelerating structures to boost the energy of particles. The collider lies in a tunnel with a depth of 175m under France–Switzerland border near Geneva.

In total there are four crossing points that allow particles to collide on LHC and seven detectors are built around those crossing points. Each detector has a different design to serve different experimental purposes. The two largest detectors on the LHC are the ATLAS (A Toroidal LHC Apparatus) and the CMS (Compact Muon Solenoid) detector. Another two large detectors on the LHC are ALICE (A Large Ion Collider Experiment) and LHCb (LHC-beauty). There are also four smaller detectors: TOTEM (Total Cross Section, Elastic Scattering, and Diffraction Dissociation), LHCf (LHC-forward), MoEDAL (Monopole and Exotics Detector At the LHC), and FASER (Forward Search Experiment). The overall structure of LHC and the CERN collider complex is shown in figure 3.1.

The most important goal of the LHC is to discover the Higgs boson, which has been accomplished in 2012, and high precision measurement of the Higgs boson's properties, its interactions with other particles, and other Standard Model predictions. Besides the Standard Model measurement and discovery, the LHC is also a very important place to conduct beyond Standard Model searches, such as searching for the large family of new particles predicted by super-symmetric theories and other unresolved questions in physics.

During most run time, the LHC is operated under proton-proton beam collision, which

serves most of the experiments on the LHC. The LHC can also serve for heavy ion experiments, such as lead-lead ion beam collision and proton-lead collision. Normally the LHC will perform heavy ion collisions one month per year. The first proton-proton collision on the LHC was achieved in 2010. Each proton in the beam carried an energy of 3.5 TeV. The $\sqrt{s}$ (total collision energy in the center-of-mass frame) of the first collision is 7 TeV. The operation period from 2010 to 2014 is called Run I. During Run I $\sqrt{s}$ of the proton-proton collision on the LHC is 7-8 TeV. After finishing the first upgrade in 2015, the collision energy on LHC is increased to 13 TeV and this operation period is called Run II. After 2018, the LHC was shut down for three years for the phase II upgrade. In 2022, the LHC upgrade work is finished the LHC will enter the Run III operation period. The LHC draws about 120 MW of electrical power from the French electrical grid and generates about 140 TB of data during operation.



(a)

Figure 3.1: Structure of the Large Hardon Collider (LHC) and CERN accelerator complex in 2019 [53].

3.1.2 Accelerator design

Accelerators are the heart of the LHC. Accelerators boost particle beams to the target energy and maintain their energy when the particle is traveling in the circle since the particle will lose energy due to radiation. As shown in figure 3.1, the LHC accelerator complex is formed by a series of accelerators. Before being injected into the 27 km long main circle, the photon beams will be accelerated by a series of smaller accelerators and the energy

will increase gradually. The first accelerator is the Linear particle accelerator (LINAC4). Hydrogen ions (H^-) will be injected into LINAC4 and be accelerated to 160 MeV. LINAC4 is connected with Proton Synchrotron Booster (PSB), where the two electrons in the Hydrogen ions are separated with a hydrogen nucleus leaving only one proton. The proton will be accelerated to 2 GeV in PSB. Then, the proton beam enters the Proton Synchrotron (PS) and its energy will be increased to 26 GeV. The last accelerator before proton beams enter the LHC main accelerator is the Super Proton Synchrotron (SPS), where energy reaches 450 GeV then proton beams will be injected into the main ring of LHC. In the main ring of LHC, proton beams will be accelerated to their peak energy (13 TeV in Run II period) and will be circulated for 5 to 24 hours while collisions happen at the four crossing points.

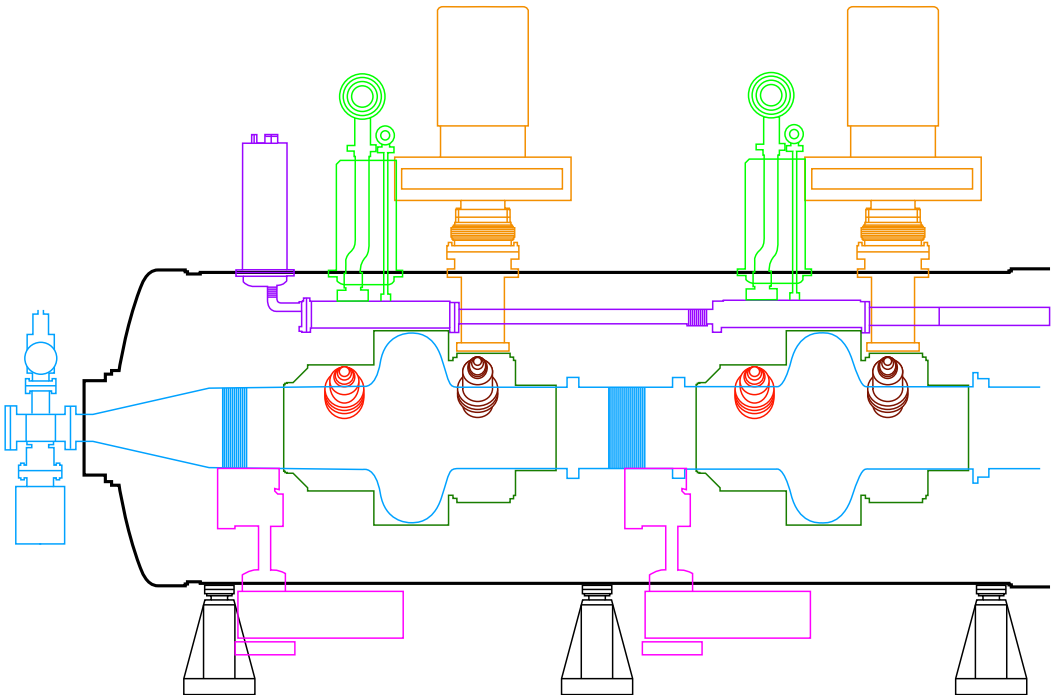


Figure 3.2: Schematic side-view of an LHC accelerating module showing two of the four cavities (blue). The cylindrical vacuum tank is shown in black. Inside, the beamline is shown in blue. The cavities are housed inside helium-filled cryomodules (dark green) fed by liquid helium baths (purple). Quench valves attached to the helium baths are shown in light green. Each superconducting cavity is driven by the variable power couplers shown in orange. Two other couplers control the higher-order resonance modes in the cavity: a broadband HOM coupler (red) and a narrow-band HOM coupler (brown). The resonance of each cavity is tuned by the elastic deformation of the chamber by a motor system (pink).

The accelerator of LHC is based on radio-frequency (RF) cavities, as shown in figure 3.2. LHC uses superconducting radio-frequency and the RFs are operated under the temperature

of 4.5K. RFs are made of copper and sputtered with niobium to achieve good thermal conductivity. To maintain the low temperature, a liquid helium cooling system is used by each RF.

A high voltage of 2 MV is applied to each radio frequency to generate a strong electric field of 5.3 MV/m. Each RF is powered by a 500 KW klystron, which is coupled with the RF by a waveguide with adjustable length. Changing the length of the waveguide can control the *quality factor*[1] of RF.

In LHC, there are eight RFs per beam, which means the voltage gradient per beam is 16 MV. Eight RFs are divided into two groups and RFs in the same group share one cryostat, which maintains their superconducting temperature.

3.1.3 Magnet design

The magnet is another crucial part of the LHC, which generate a magnetic field to hold the beam and keep particle circling in the LHC. There are different types of magnets installed on LHC for different purposes: dipole, quadrupole, and sextupole magnets, etc. Dipole magnets can bend the beam so that particles can travel in the circle of LHC. In total 1232 dipole magnets are installed on the LHC. Quadrupole magnets are used to change the shape of the beam, in other words, focus and defocus the beam when it is traveling. LHC uses 392 Quadrupole magnets. Sextupole magnets correct beam characteristics including chromaticity introduced by the quadrupoles and LHC uses 2464 sextupole magnets.

There are some other magnets such as octupole, decapole correctors, and kicker magnets. All those magnets installed on LHC keep the beam traveling in LHC and keep the beam shape correct.

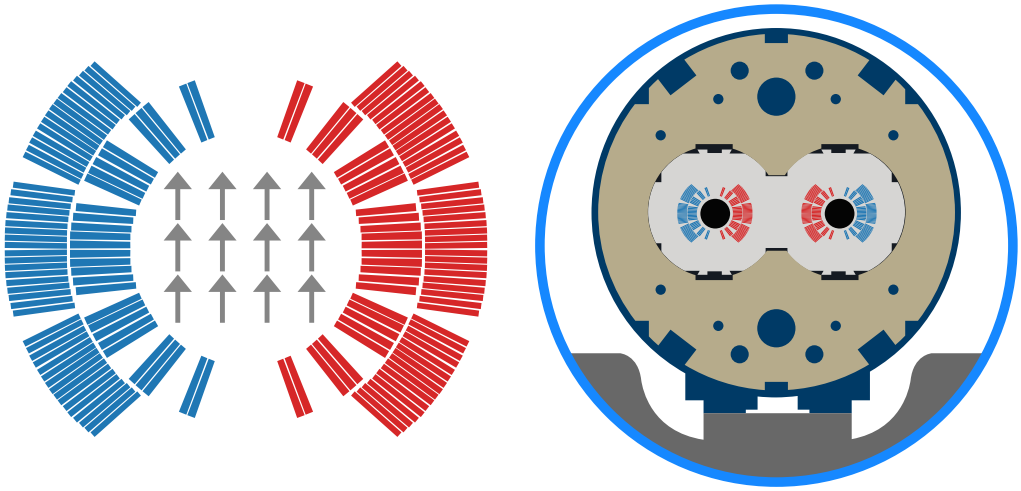
Figure 3.3 shows the structure of dipole magnets installed on the LHC and coil winding.

3.1.4 Operation and performance of the LHC

Operation period of the LHC

The operation of LHC and the ATLAS detector can be divided into the Run I, Run II, and Run III periods. Run I, also called the first operational run, started in 2009 and ended in 2013. At the early stage in Run I, the beam energy is 3.5 TeV ($\sqrt{s} = 7$ TeV) and then it was increased to 4 TeV ($\sqrt{s} = 8$ TeV). During Run I the ATLAS detector recorded data with an integrated luminosity of 4.9 fb^{-1} at 7 TeV $\sqrt{s}$ and data with an integrated luminosity of 20 fb^{-1} at 8 TeV $\sqrt{s}$. The luminosity of a collider is defined in equation 3.1, where N is the

[1]Peak energy lost per cycle, which is used to adjust the peak voltage in the cavity.



(a) Coil winding (b) Dipole magnet

Figure 3.3: (a) Cross section of the coils made from superconductive cables wound around the beamline to produce a homogeneous dipole magnetic field. Each rectangle represents a flat Nb-Ti cable. These are grouped in a configuration that produces a smooth internal magnetic field, indicated by the arrows. The cables are separated by layers of copper. The cables are color-coded: red(blue) cables carry current into(out of) the page. (b) Cross section of the main bending dipole magnets. Two beamlines (black) and coils (blue/red) are encased by austenitic steel collars (light grey). The collar is embedded in a large iron yoke (dark grey) and submerged in a liquid helium vessel (dark blue). The arrangement is held in a cryostat. For scale, the centers of the beamlines are separated by 19 cm. (figures by Aaron White).

number of particles per bunch, k_b is the number of bunches, σ_x^* and σ_y^* is the horizontal and vertical beam sizes at the interaction point and F is the geometrical reduction factor arising from the crossing angle [54].

$$L = \frac{N^2 k_b f}{4\pi\sigma_x^*\sigma_y^*} F \tag{3.1}$$

Equation 3.1 can be rewritten as equation 3.2 if the beams are round and the beta functions of both beams have the same values in both planes, where γ is the usual relativistic factor, ϵ_n is the normalized emittance, and β^* is the value of the beta function at the interaction point [54].

$$L = \frac{N^2 k_b f \gamma}{4\pi\epsilon_n\beta^*} F \tag{3.2}$$

After a long shutdown and upgrade of LHC and the ATLAS detector from 2013 to 2015,

the energy of the beam was increased to 6.5 TeV ($\sqrt{s}$ = 13 TeV) and ATLAS started the second operational run (Run II). The Run II period starts in 2015 and ended in 2018. The recorded luminosity by ATLAS and recorded during good operation of the detector during Run II are shown in figure 3.4.

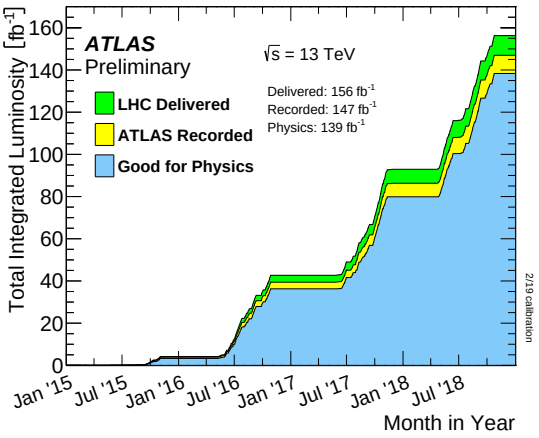


Figure 3.4: Total luminosity of *pp* collisions delivered by the LHC in Run 2 (green), recorded data by the ATLAS detector (yellow) and the portion of this recorded during good operation of ATLAS (blue) [55].

Two analyses in this thesis are both based on the ATLAS full Run II data set with an integrated luminosity of 139 fb^{-1} at $\sqrt{s}$ = 13 TeV.

From 2018 to 2022, the LHC shut down again (the Long Shutdown 2) and upgraded to the High Luminosity Large Hadron Collider (HL-LHC) project which will increase the luminosity by a factor of 10. The LHC became operational again in April 2022 and enters its third operational run (Run III).

Performance of the LHC

Table 3.1 summarized key performance parameters of the LHC pp collision during the Run 2 operation period from 2015 to 2018 [56] and performance parameters of Run 1 (2012) are summarized in table 3.2.

In table 3.1, the number of physics operation days in the table means the number of days the LHC performs pp collision. Pile-up (μ) refers to the average number of particle interactions per bunch crossing. Pile-up will affect object and event reconstruction since pp collisions in addition to the collisions of interest are introduced. The pile-up distributions of different years in the LHC Run 2 are shown in figure 3.5 [2].

[2]Public ATLAS Luminosity Results for Run-2 of the LHC

Parameter	2015	2016	20117	2018
IR1/IR5 Peak L (10^{34} cm^{-2}s^{-1})	0.5	1.4	2.11	2.1
Average pile-up (μ)	13	25	38	37
Max. number of bunches	2244	2220	2556 / 1868	2556
Max. train length (bunches)	144	96	144 / 128	144
Emittance injection (μm)	≈ 3.0	≈ 1.6	≈ 1.6	≈ 1.4
Bunch pop. start of stable beams (10^{11})	1.0-1.25	1.0-1.25	1.0-1.25	1.0-1.25
Emittance start of stable beams (μm)	≈ 3.5	≈ 2.2	≈ 2.2	≈ 1.9
RF Voltage injection (MV)	6	6	6	6 / 4
RF Voltage collisions (MV)	12	12	12	10
β^* IR1/5 (cm)	80	40	40-30	30-25
Half crossing angle IR1/5 (cm)	145	185 / 140	150-120	160-130
IR1 crossing sign	-1	-1	+1	+1
Max. stored energy (MJ)	280	270	330	320
IR1/IR5 integrated L (fb^{-1})	4.2	39.7	50.6	66
IR8 integrated L (fb^{-1})	0.36	1.87	1.98	2.46
IR2 integrated L (fb^{-1})	9	13	19	27
Commissioning duration (days)	58	28	24	17
Average length of stable beams (hours)	6.8	11.2	8.2	8.3
No. days of physics operation	88	146	140	145
Machine availability (%)	69	76	83	79
Stable beams efficiency (%)	35	49	49	49

Table 3.1: Evolution of the LHC performance in regular pp collision during Run 2 (2015-2018) [56].

Parameter	2010	2011	2012
Beam Energy	3.5	3.5	4.0
β^* in IP 1 and 5 (m)	2.0 / 3.5	1.5 / 1.0	0.6
Bunch spacing (ns)	150	75 / 50	50
Max. number of bunches	368	1380	1380
Max. bunch intensity (protons per bunch)	1.2×10^{11}	1.45×10^{11}	1.7×10^{11}
Normalized emittance at injection	≈ 2.0	≈ 2.4	≈ 2.5
Peak L (10^{34}cm^{-2}s^{-1})	0.021	0.37	0.77
Max. number of events per bunch crossing	4	17	37
Stored beam energy (MJ)	≈ 28	≈ 110	≈ 140

Table 3.2: Evolution of the LHC performance in regular pp collision during Run 1 (2010-2013) [57].

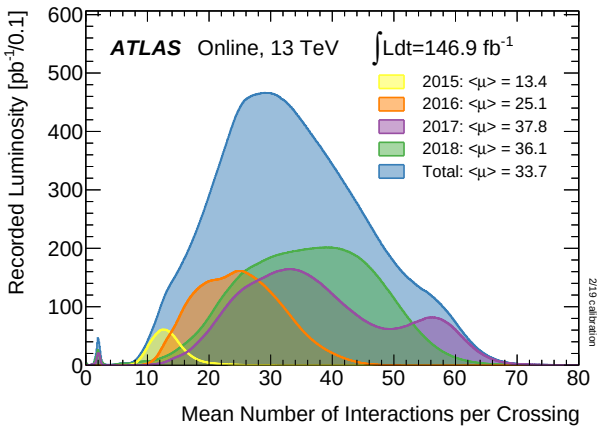


Figure 3.5: Luminosity-weighted distribution of the mean number of interactions per crossing (pile-up μ) for 2015 – 2018 pp collision data at $\sqrt{s} = 13$ TeV recorded by ATLAS during Run 2.

3.2 The ATLAS detector

3.2.1 Overview of the ATLAS detector

The ATLAS detector (A Toroidal LHC Apparatus) is the largest general-purpose particle detector built around the LHC, which is 25 meters in diameter, 46 meters long, and weighs about 7 thousand tons. The ATLAS detector is designed to serve different experimental purposes so that it can receive a wide range of signals. It can measure properties, such as energy, and momentum, of different kinds of particles. Different from other general-purpose particle detectors installed on Tevatron or LEP, the ATLAS detector can handle the extremely high beam energy and high collision rate produced by the LHC. To achieve this goal, the ATLAS detector is a very complex system, with more than 3000 km of cables, and contains many sub-systems.

As shown in figure 3.6, there are four major sub-systems on LHC: the inner detector, calorimeters, magnet system, and the muon spectrometer. The magnet system of the ATLAS detector produces a strong magnetic field that can bend charged particle while it travels through the detector. Particles produced at the collision point will pass through the inner detector first, where the inner detector can measure the particle's track precisely. After passing through the inner detector, particles will enter the calorimeter, which can measure the particle's energy. Finally, the muon spectrometer can identify muons and measure their tracks.

In the ATLAS experiment, there are three commonly used coordinates that describe the ATLAS detector. A Cartesian coordinate system, a cylindrical coordinate system and a Spherical coordinate.

The origin of the Cartesian coordinate system is set at the center of the inner tracker. The Z-axis is parallel to the beamline. The X-axis points toward the center of the LHC ring and Y axis point to the top of the ATLAS detector. The detector side with positive z coordinates is defined as the A-side, while the detector side with negative z coordinates is defined as the C-side.

As for the cylindrical coordinate system, the origin point and z-axis are the same as in the Cartesian coordinate system. The azimuthal angle ϕ is defined with respect to the Cartesian x axis, and the radius ρ is defined from the z axis. During the operation of the ATLAS detector.

[3]Chen *Computer generated cut-away view of the ATLAS detector showing its various components.*, http://commons.wikimedia.org/wiki/File:ATLAS_Drawing.jpg,CCBY-SA2.0,https://commons.wikimedia.org/w/index.php?curid=18438360

[3]Aaron White

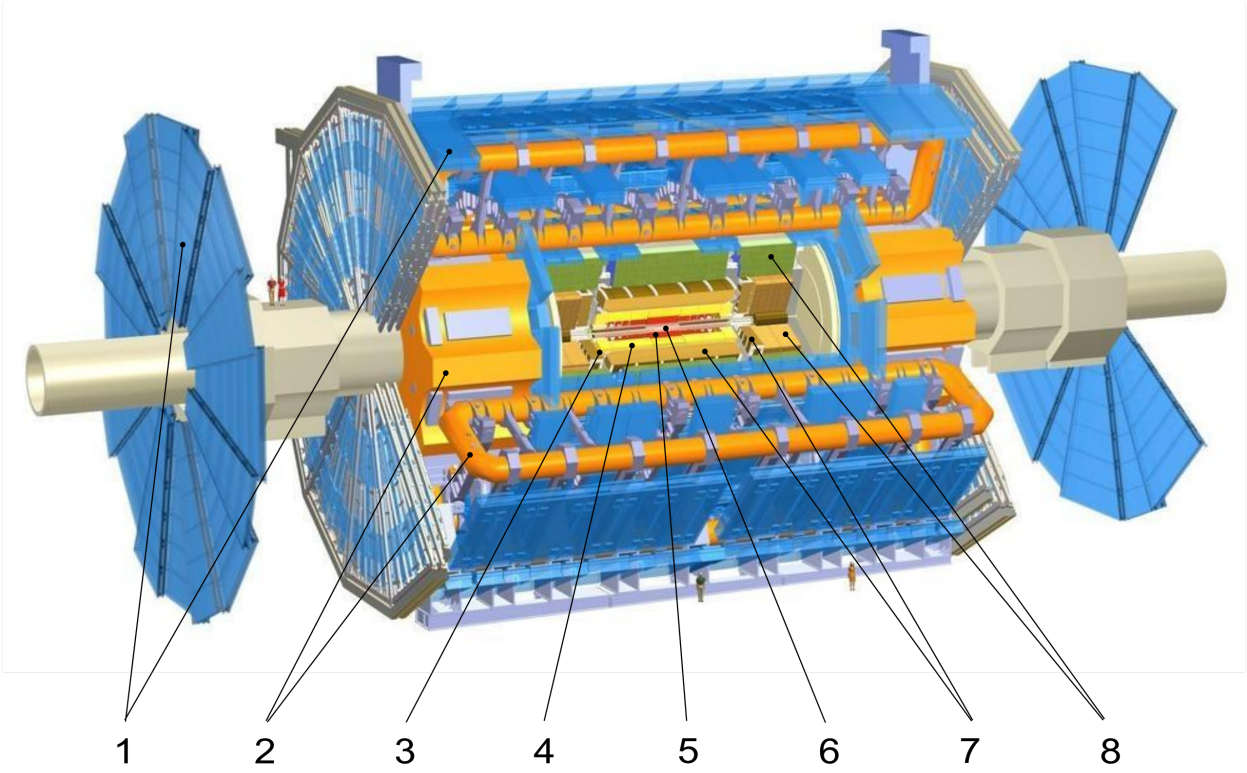


Figure 3.6: Computer generated cut-away view of the ATLAS detector showing its various components.
Moun Spectrometer: (1) Forward regions (End-caps). (1) Barrel region (Barrel).
Magnet System: (2) Toroid Magnets. (3) Solenoid Magnet.
Inner Detector: (4) Transition Radiation Tracker. (5) Semi-Conductor Tracker. (6) Pixel Detector.
Calorimeters: (7) Liquid Argon Calorimeter. (8) Tile Calorimeter.
Illustration by Chen[3].

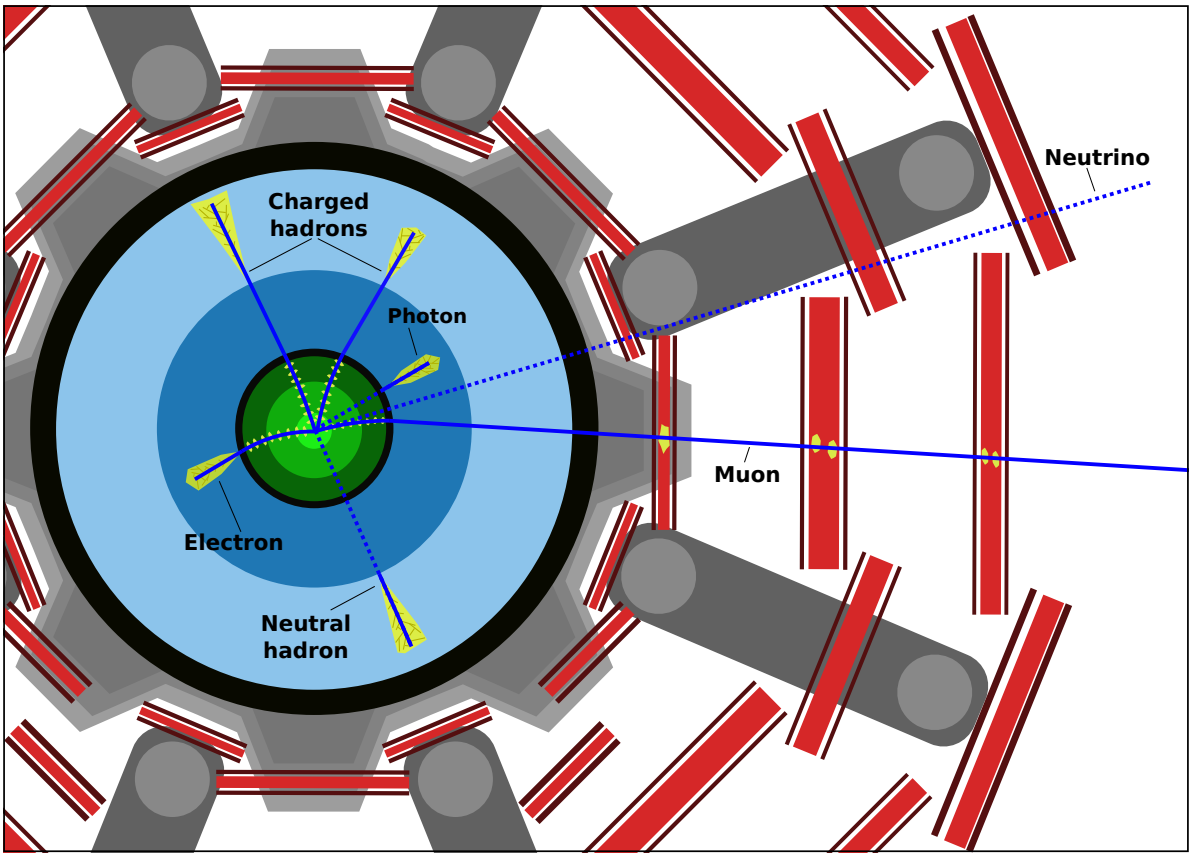


Figure 3.7: A cross-section of ATLAS looking along the beamline. The detectors are color-coded: green for the inner detector, blue for the calorimeter, and red for the muon spectrometer. Several particle paths are shown in blue. Dashed lines represent a particle passing through the detectors without interacting. Yellow areas indicate electromagnetic or hadronic showers in the calorimeters and other interactions with the detector. Outside the inner solenoid, the muon is bent parallel to the beamline by a toroidal magnetic field.

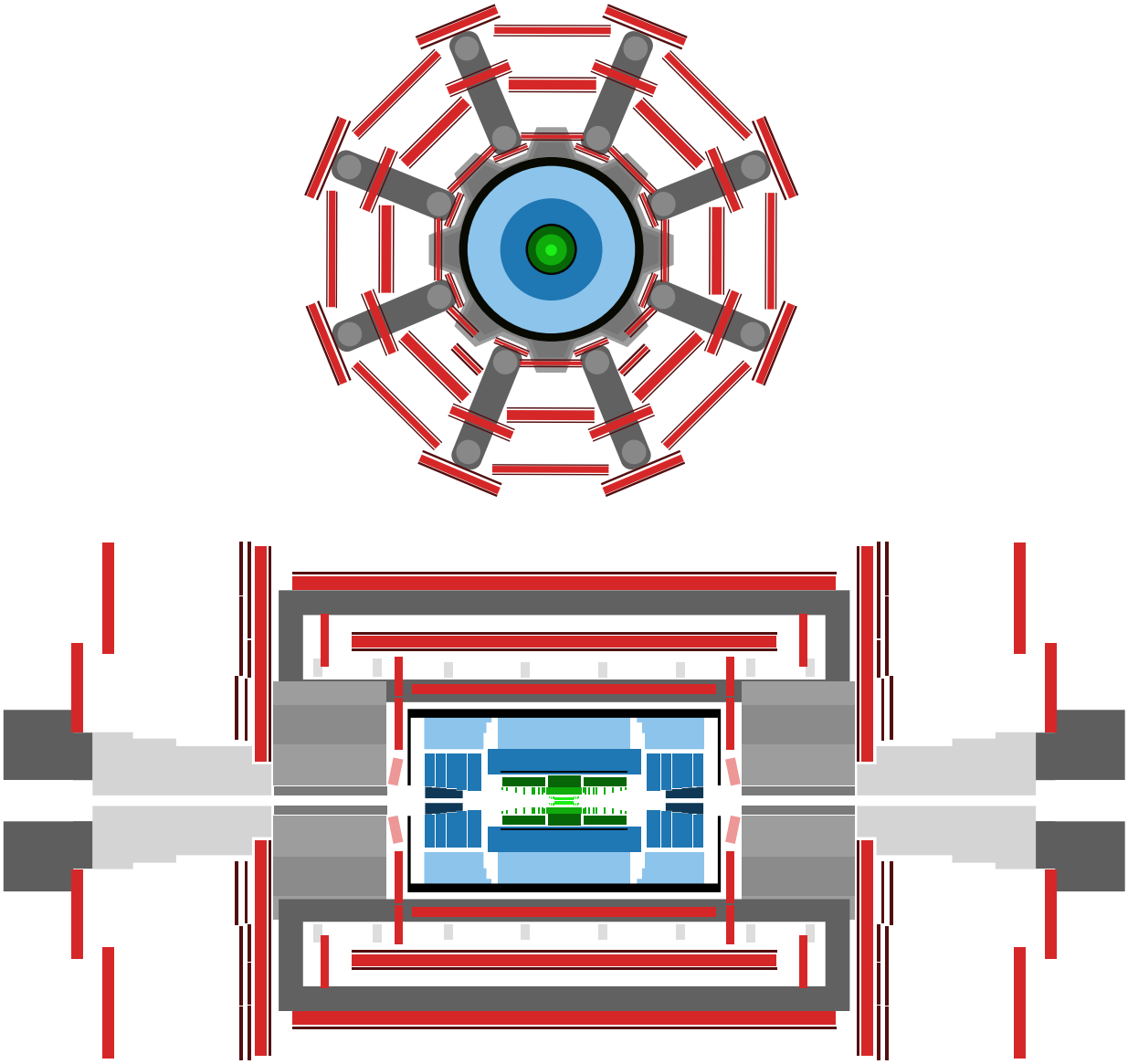


Figure 3.8: Layout of the ATLAS detector with subsystems coded by color. **Top:** a view of the barrel looking along the beamline in the z direction. **Bottom:** a view of the detector from the side looking in the x direction. **Red:** muon system with CSC (light), MDT (medium), and trigger (dark) chambers. **Blue:** calorimeter system with tile (light), LAr (medium), and FCal (dark) calorimeters. **Green:** inner detector with pixel (light), strip (medium), and TRT (dark) detectors. **Grey:** magnet system with ECT (light), BT (medium), and CS (black) magnets.

The Spherical coordinate is commonly used to describe physics interactions and tracks of particles. The origin point is the same as the Cartesian coordinate system and cylindrical coordinate system, which locates at the center of the inner tracker. The radius r is defined from the center of the detector and polar angle θ is defined with respect to the z axis. In particle physics, the polar angle can be used as a particle's *pseudorapidity* (η), which is one of the important proprieties to describe a particle's track in the ATLAS experiment. The definition of pseudorapidity is:

$$\eta \equiv -\ln\left(\tan\frac{\theta}{2}\right) \tag{3.3}$$

The ATLAS detector can identify different particles by their different interaction pattern in different sub-systems. As shown in figure 3.6, from the collision point, particles will travel through different layers of the detector. Different particles, such as electrons, muons, and photons, have different patterns in different layers. Figure 3.7 shows interactions between different particles and different layers of the ATLAS detector. For example, charged particles such as muons and electrons can be detected by the inner detector and leave tracks and others will pass through the inner detector directly. Electrons will deposit all of their energy in the calorimeter leaving an electromagnetic radiation shower while muons can pass through the calorimeter and leave tracks in the muon spectrometer. Particles with different energy and momentum will have tracks with different radius. For particles will different signs of charge, their track will have different directions in the magnetic field.

3.2.2 Inner detector

The innermost system of the ATLAS detector is the inner detector, which is designed to precise measures the position of the collision point and tracks of charged particles. As shown in figure 3.9, the inner detector contains three major components: the pixel detector, the silicon-strip tracker, and the transition radiation tracker. The inner detector is built around the beampipe and covers the region where collisions happen. The inner detector is about seven meters in length and 2.3 meters in diameter. The inner detector has an end-cap at both sides and can cover the region with $|\eta| < 2.5$.

The pixel detector and silicon-strip detector are solid-state silicon detectors. Both of them use semiconductors to detect charged particles. The semiconductor detector is made of two different types of semi-conductor materials: p-type (a trivalent impurity, which can be viewed as positive charges, (like Boron, Aluminum, etc.) is added to an intrinsic or pure semiconductor) and n-type (a donor impurity, which can provide extra electrons, (Arsenic, Phosphorus, etc.) is added to an intrinsic or pure semiconductor) semiconductor materials. The semiconductor detector contains many p-type and n-type layers.

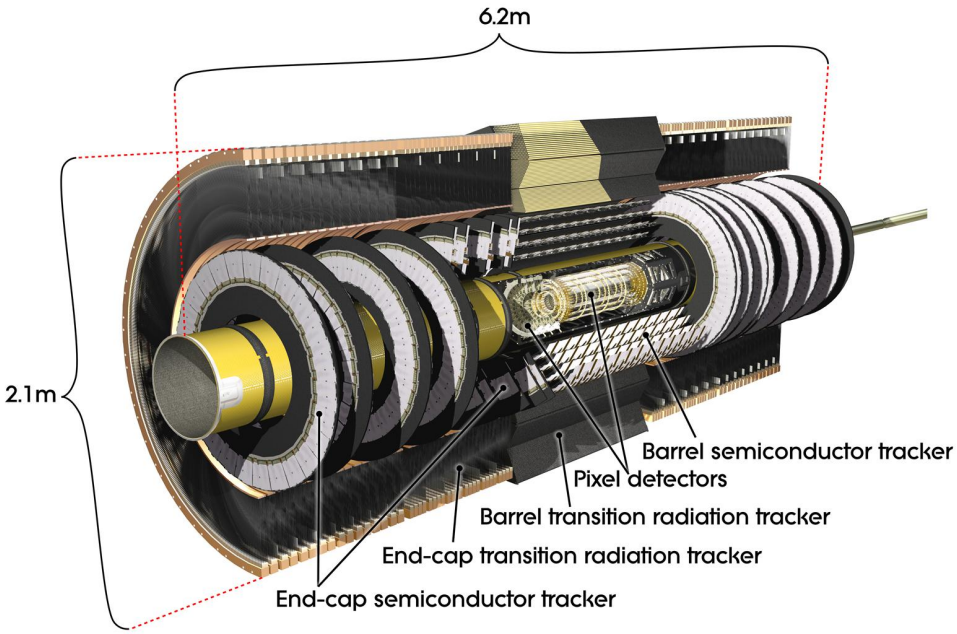


Figure 3.9: Computer-generated image of the inner detector with a section removed for visibility. Illustration by Joao Pequenao[4].

In the semiconductor detector, p-type and n-type materials are connected with each other, which is called the At the interface between p-type and n-type layer, extra electrons in the n-type side will move to the p-type side, which can produce a depletion region and an electric field at the interface region. Figure 3.10 shows the structure of a typical p-n junction.

When charged particles produced by the proton-proton collision pass the p-n junction, atoms will be ionized by the charged particles. Ionized electrons will drift under the electric field in the p-n junction so that the detector can detect the current and receive signals.

The pixel detector

As shown in figure 3.9, the pixel detector is the innermost detector of the inner detector. The pixel detector contains many small pixel-like silicon detectors, shown in figure 3.11, groups in the grid, and each pixel has its individual readout channel, which looks like the CMOS (complementary metal-oxide semiconductor) used in the digital camera. Using this unique structure, the pixel detector can measure particle position precisely since the system knows which pixels pick up signals. The size of each pixel is about $50 \times 400 \mu m^2$. In the

[4]Joao Pequenao. *Computer generated image of the ATLAS inner detector*, 2008. `https://cds.cern.ch/record/1095926`

[5]Wikipedia *Semiconductor*. `https://en.wikipedia.org/wiki/Semiconductor`

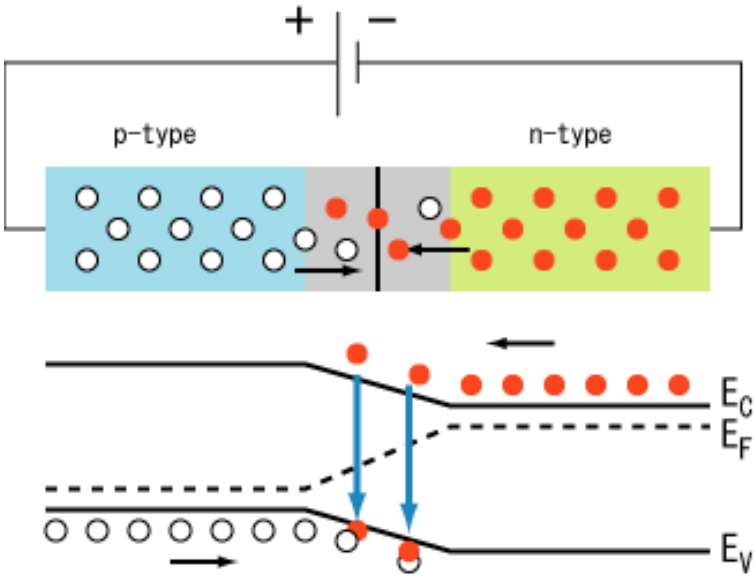


Figure 3.10: P-N junction Diode. Illustration by Wikipedia.[5]

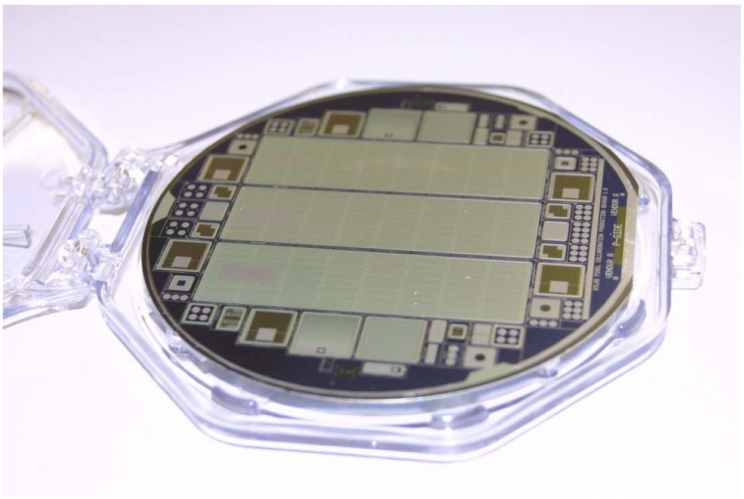


Figure 3.11: A photograph of a 4-inch diameter ATLAS pixel sensor wafer (p-side view) [58]

barrel, these are grouped into arrays of 24×160 pixels for readout. Sixteen arrays are located on a "module", which serves as the repeating basis of the various pixel layers. Four layers of detectors are arranged in the barrel, with the innermost layer located inside the beampipe, with radii of 2.9 cm, 5.1 cm, 8.9 cm, and 12.3 cm. In each end-caps, four disks are arranged between z=11 cm and z=20 cm [59, 60].

The silicon strip detector

After passing through the pixel detector, particles will enter the silicon-strip detector. Similar to the pixel detector, charged particles will ionize atoms around the p-n junction of the silicon-strip detector and produce a measurable electric current. Semi-conductors are in strip shape instead of pixel. The barrel silicon strip detector contains four layers and the end-cap detector contains nine disks on each side. Each barrel strip contains modules with two silicon strip layers. In each layer, the silicon wafers are divided into strips 126 mm in length. The strips are oriented differently in the barrel depending on their layer, with some running parallel to the beam axes and others offset at an angle of 40 mrad [61].

The transition radiation tracker

The outermost part of the inner detector is the transition radiation tracker. The transition radiation tracker (TRT) belongs to the gas detector category. This detector contains many gaseous proportional-mode drift tubes which are filled with a mixture of gas. The mixture of gas is made of xenon (70%), carbon dioxide (27%), and oxygen (3%). The tube wall is made of wound Kapton and in the center, there is a gold-plated wire, which is $31\mu m$ in diameter. During operation, a negative voltage of -1.5 kV is applied to the tube wall to create a strong electric field inside the drift tube. When a charged particle enters the drift tube, gas atoms will be ionized under the strong electric field. Ions with positive charges will drift to the tube wall and electrons will drift to the wire. During this process, those electrons and ions will further ionize gas atoms to create more charged particles so that they can produce high enough electric current that can be picked up by the detector. The current will be amplified and read out by the front-end electronics. The transition radiation tracker is a complement to the pixel and silicon-strip detector, which can provide a large tracking area at a much lower cost. Compared with the gas detector, the semi-conductor detector can provide more precise tracks and location information but the cost is too high. The choice of TRT is a balance between precision and budget.

3.2.3 Calorimeters

After passing through the inner detector, particles produced by collision will enter calorimeters. The calorimeter system of the ATLAS detector contains two sub-systems, the Liquid Argon Calorimeter (LAr) and the Tile Calorimeter. The Liquid Argon Calorimeter is designed to measure energy deposits through electromagnetic interactions, while the Tile Calorimeter is designed to measure energy deposits of hadronic jets.

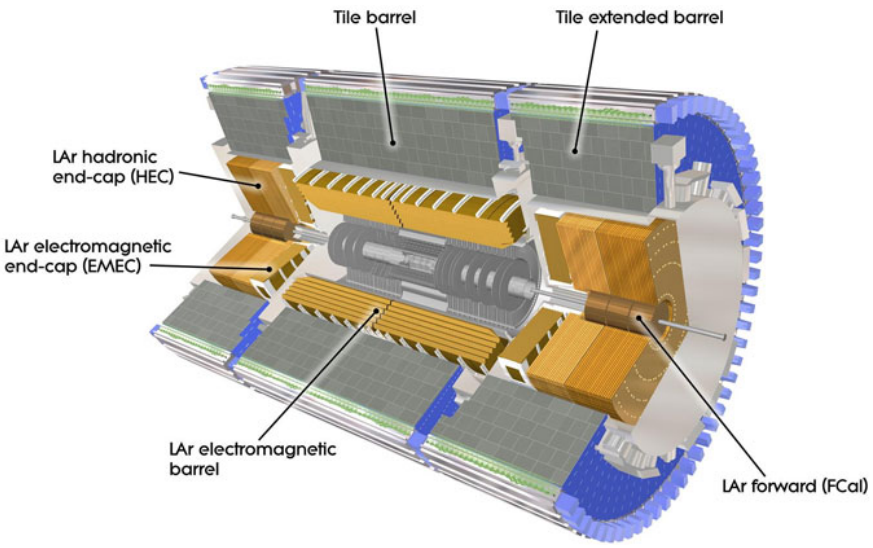


Figure 3.12: Computer-generated image of the calorimeters with a section of the removed for visibility. Illustration by Joao Pequenao[6].

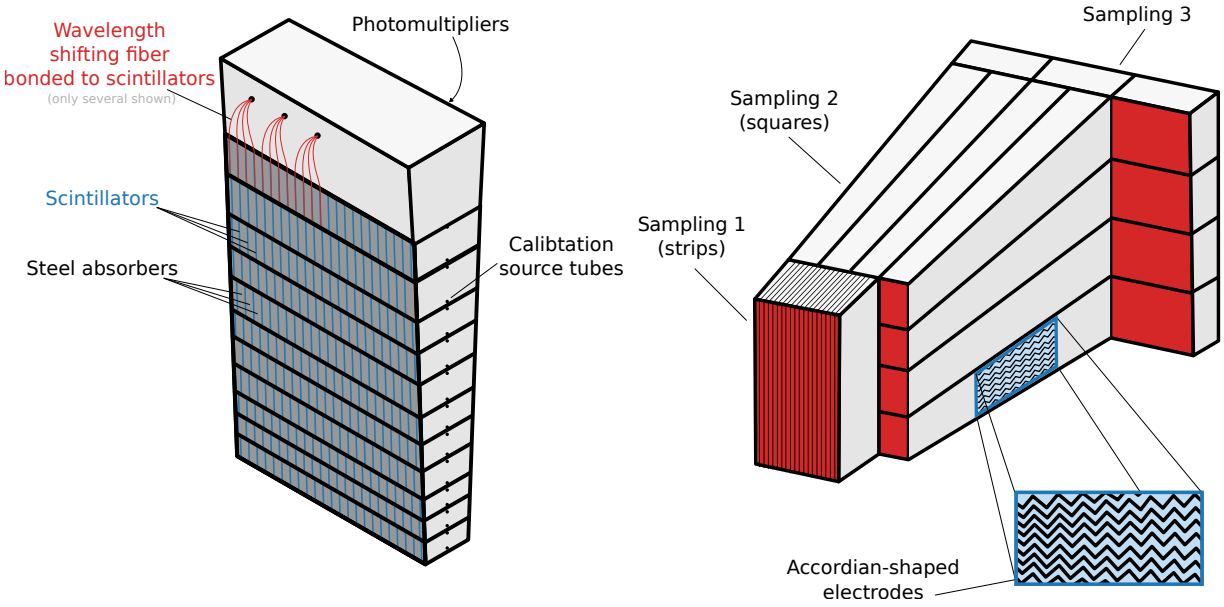
The calorimeter system can be divided into two major types: the electromagnetic calorimeter and the hadronic calorimeter. The electromagnetic calorimeter is designed to measure electrons' energy deposit through the bremsstrahlung radiation. It can also measure the energy loss of photons through electron-positron pair production. Most parts of the LAr calorimeter are used as electromagnetic calorimeters in the ATLAS detector. As shown in figure 3.13, the LAr calorimeter is located inside of the Tile calorimeter. Particles like electrons and photons will be totally absorbed by the LAr calorimeter and deposit all of their energy in the LAr calorimeter. Hadronic jets and pass through the LAr calorimeter and enter the Tile calorimeter, which belongs to the hadronic calorimeter. The hadronic calorimeter measures a particle's energy loss through hadronic interaction. The hadronic calorimeter is mainly used to measure the energy of hadronic jets produced by the proton-proton collision, which is very important to the ATLAS experiment.

The LAr calorimeter

The LAr calorimeters are used as both the electromagnetic calorimeter and the hadronic calorimeter in the ATLAS detector, which is between the inner detector and the Tile calorimeter. According to the structure of the calorimeter shown in figure 3.13, the LAr calorimeter has four major components: the LAr electromagnetic barrel, the LAr electromagnetic end-cap (EMEC), the LAr LAr forward calorimeter (FCal). The LAr electromag-

[6]Joao Pequenao. *Computer Generated image of the ATLAS calorimeter*, 2008. `https://cds.cern.ch/record/1095927`

netic barrel contains two half-cylinder electromagnetic calorimeters and covers $|\eta| < 3.2$. The end-cap parts and LAr forward cover $|\eta| < 4.8$.



(a) (b)

Figure 3.13: (a) A module in the TileCal consisting of the staggered scintillator (blue) and steel (grey) layers read out on both sides by fiber optics connected to photomultipliers. Source tubes in the TileCal allow radioactive sources to be inserted for calibration. (b) A section from the LAr, consists of towers (red) composed of accordion-shaped electrodes submerged in liquid argon.

The LAr barrel and end-cap electromagnetic c are designed to measure the energy of electrons and photons through electromagnetic interactions. As discussed in the previous section, electrons interact with the LAr calorimeters through bremsstrahlung radiation and photons interact with the LAr calorimeters through electron-positron pair production. Interaction rates of both interactions are related to the thickness of the material that particles travel through, which is called the radiation length, X_0. Energy loss in terms of travel distance is:

$$-\frac{dE}{dx} = \frac{E}{X_0} \tag{3.4}$$

Since the energy of photons is related to their frequency, frequency change in terms of the travel distance is:

$$-\frac{dw}{dx} = \frac{1}{\lambda_{\text{pair}}} e^{-x/\lambda_{\text{pair}}}; \quad \lambda_{\text{pair}} = \frac{9}{7} X_0. \tag{3.5}$$

Secondary particles produced by both interactions will continue to interact with the material and produce showers of electrons and photons. If the interaction material is thick enough all energy of the initial electron or photon will be absorbed. As shown in figure 3.13, the LAr electromagnetic calorimeters contain many accordion-shaped electrodes made of copper. Those electrodes are submerged in liquid argon. During operation, a high voltage of 2 kV will be applied to the copper electrodes. Between copper electrodes are steel-clad lead absorbers. Those absorbers will interact will particles and produce showers of electrons and photons. The absorbers are also submerged in liquid argon, which works as the active material. Incident particles will ionize argon atoms and the produced electrons and ions will drift to electrodes. Electrodes are grouped into towers of different shapes, including narrow strip towers, square sampling towers, and wide trigger towers. When a particle enters the LAr electromagnetic calorimeter it will pass the strip towers first, which can provide information on the particle's energy loss before reaching the calorimeter. After that, the particle will enter the sampling tower. The radiation length X_0 of sampling towers is about 20. Most of the energy of electrons and photons shower will be absorbed by the sampling towers. The last part the particle may pass through is the trigger tower and the total radiation length at the barrel region is about 22. The design of the barrel and EMEC calorimeters provide a relative energy resolution $\sigma_a = 10\%$ and $\sigma_c = 0.7\%$ [62].

As for the LAr hadronic calorimeter, similar to the LAr electromagnetic calorimeter, it also contains electrodes made of copper. The difference is that hadronic calorimeters use flat-shaped electrodes. What's more, instead of using steel-clad lead, its absorbers also use copper.

The Tile calorimeter

The previous section discusses the design of LAr calorimeter of the ATLAS detector. The LAr calorimeter is surrounded by the Tile calorimeter, which is mainly designed to measure the energy of hadronic jets through hadronic interactions. Different from the LAr hadronic calorimeter, the Tile calorimeter consists of absorbers and scintillators. The absorber will interact with the incoming hadronic particles to produce showers and scintillators will detect the produced charged particles.

Hadronic particles that enter the calorimeter will lose energy through hadronic interactions with atoms of the calorimeter. Different from the electromagnetic interactions mentioned in the previous sections, the physics quantity to describe the ability that the material absorbs the particle's energy through hadronic interaction is the average nuclear interaction length. Effective hadronic depth of the TileCal is $\sim 7\lambda_I$ [63], in which. The Tile calorimeter uses sheets of steel as absorb to interact with the incoming particles and absorb their energy.

Similarly to the electromagnetic shower of electrons and photons, the interaction between the incoming hardons and absorber atoms will produce secondary light hardons and neutrinos. This process will continue to produce more hadrons that form the hadronic shower in the hadronic calorimeter. When charged particles in the hadronic shower enter scintillators, scintillate light will be produced which can be detected by the scintillators.

The Tile calorimeter only contains three barrel parts. The central barrel Tile calorimeter is 5.6 meters long and covers the region with $|\eta| < 1.0$. Along with the barrel part is two extended Tile calorimeter which are 2.9 meters in length and extends the coverage of pseudorapidity from $|\eta| < 1.0$ to $|\eta| < 1.7$. The barrels have an inner radius of 2.3 m and an outer radius of 4.2 m. The central barrel covers a pseudorapidity of $|\eta| < 1$ and the extended barrels provide coverage up to $|\eta| < 1.7$. The energy resolution of the Tile design is $\sigma_a = 50\%$ and $\sigma_c = 7\%$ [64].

3.2.4 Muon system

Outside of the Calorimeter system is the muon system, which is the largest system of the ATLAS detector. The main purpose of the muon system is to measure the energy and transverse momentum of muons, as well as reconstruct muon tracks. The basic structure of the muon system is shown in figure 3.14. The Muon system of the ATLAS detector can be divided into the barrel and end-cap parts.

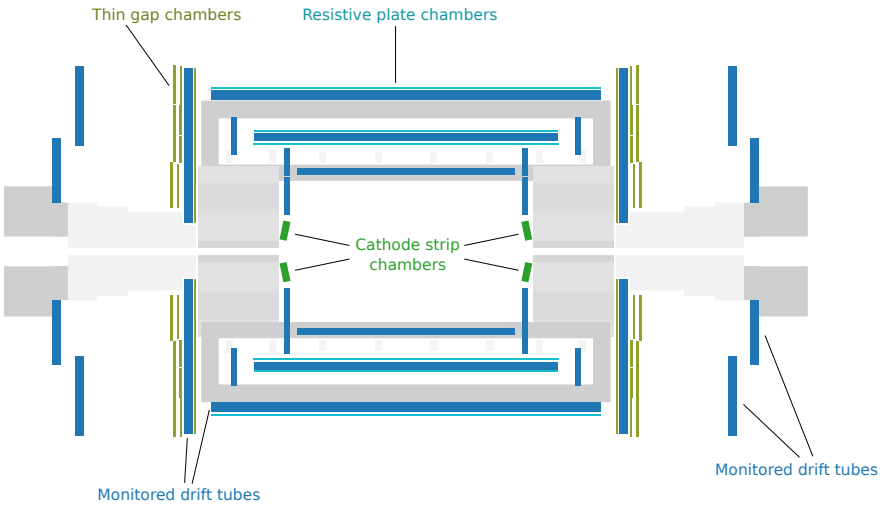


Figure 3.14: ATLAS muon system, with chambers of various types, labeled.

Monitored drift tubes (MDT)

Chambers formed by the Monitored drift tubes (MDT) are important parts of the ATLAS muon system. Those chambers are precision tracking chambers in the ATLAS muon system. These MDT chambers contain about 35000 drift tubes Monitored drift tubes in ATLAS are 3cm in diameter and their length is in the range of 0.6m to 6.1m. These tubes are in about 12000 chambers. Each MDT chamber has two multi-layers and each multi-layer has 3 or four layers of drift tubes. Typically, an MDT chamber is 2x4 m^2 in area and 0.5 meters high.

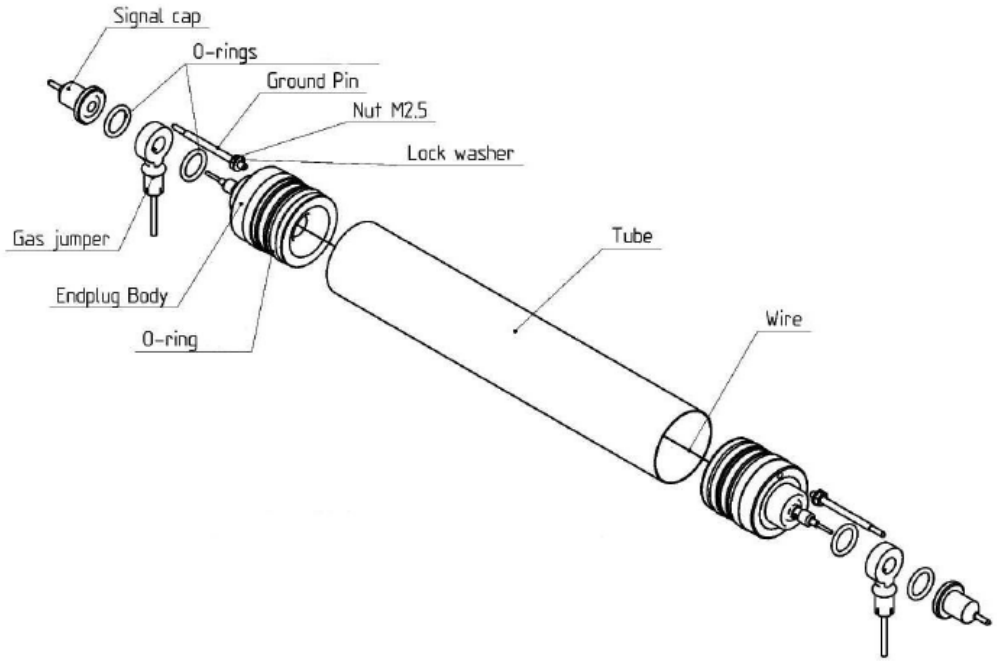


Figure 3.15: Exploded view of monitored drift tube (MDT) [65].

The tubes are made of aluminum. In the center of each tube, gold-plated tungsten wire, which is 50 μm in diameter, is used as the electrode. End-plugs fix the tungsten wire at each end of this tube. Those end-plugs enable tungsten wire to remain in the center of the tube within 10 μm of uncertainty. A high voltage of 3080 V is applied to the wire during operation.

The MDT tubes are filled with a mixture of argon (91%), nitrogen (4%), and methane (5%) at an absolute pressure of 3 bar. When muons enter the drift tubes, they will ionize the gas atoms. Under the strong electric field the electrons will drift to the wire and ions will drift to the tube wall. Electrons arriving at the central wire can produce current and current will be amplified by front-end electronics and then processed by the back-end electronics.

The MDT chambers contain a trigger system that can provide information about the

time muon arrives. Ionized electrons nearest to the anode wire reach the wire first and the back-end electronics are sensitive to the leading edge of current signals from the MDT chamber. The time difference between muon passing the trigger chamber and back-end electronics receiving signals from the tube is the drift time. Using the closest approach as radius and drawing a cycle we can get the drift circle. The Muon track is a tangent line of this drift circle. Distance between the central wire and muon track can be retrieved using the drift time and the function between drift time and drift distance in the drift tube, which is called the R-T function.

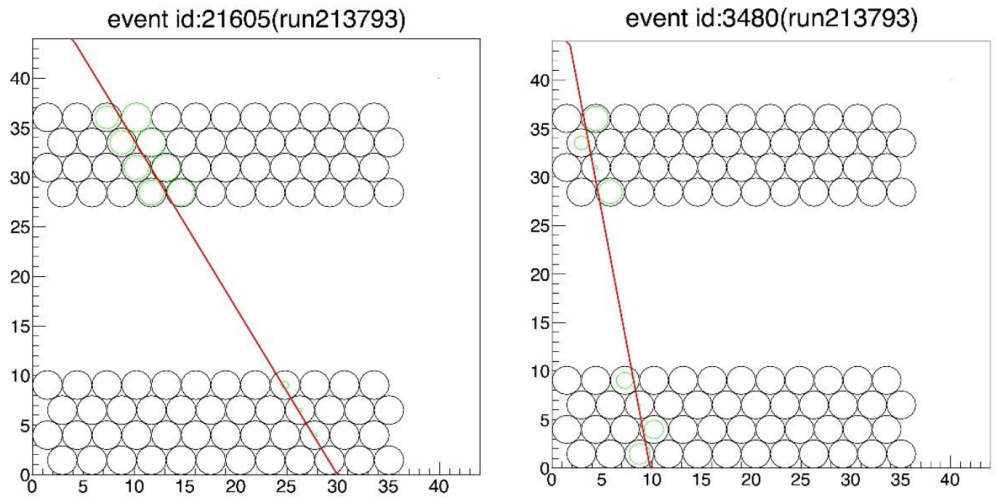
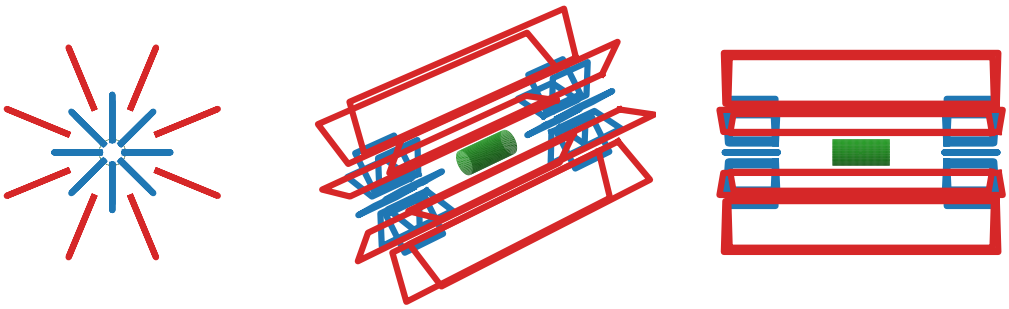


Figure 3.16: Muon tracks reconstructed by MDT chambers from cosmic ray test. Redline is the reconstructed muon tracks and green cycles are drift cycles calculated using the R-T function and drift time. The radius of the drift cycle is the drift distance.

Since only drift distance is known for the MDT chambers, the actual location where muon passes through a drift tube will locate on a cycle inside that tube, of which the radius equals the drift distance. In one event, all tubes that the moun pass through will have one drift cycle and the muon track is the common tangent line of those cycles. Figure 3.16 shows some reconstructed muon tracks by the MDT chamber during a cosmic ray test.

3.2.5 Magnet system

The momentum of charged particles can be calculated using their track radius in the magnetic field. To measure high-energy particles' transverse momentum, a strong magnetic field is needed for the ATLAS detector. The magnetic field in the ATLAS detector is about 4 T and the strong magnetic field is provided by a well-designed magnet system and superconducting magnets. The magnet system can be divided into three parts: the Barrel Toroid,



(a) (b) (c)

Figure 3.17: Illustration of the geometry of the magnet system field coils showing (a) the endcap, (b) the perspective, and (c) the side view. The Central Solenoid (green) is inside the eight windings of the Barrel Toroid (red). End-Cap Toroids (blue) appear on either side.

the end-cap Toroid, and the central Toroid. The layout of the three parts is shown with a different color in figure 3.17.

The innermost part is the Central Toroid magnetic system, which is located in the region of the inner detector. The size of the Central Toroid is 5.3 meters in length and 2.6 meters in diameter. It can produce a magnetic field of 2.6 T for the inner detector.

The largest part of the magnetic system is the Barrel Toroid, which produces a magnetic field in the barrel region, mainly for the muon systems. As shown in figure 3.17, it contains eight large coils surrounding the muon system. Those coils are air-cooled and can produce a strong magnetic field about 4 T.

At the end-cap region, there is one End-Cap Toroid on both sides. They are designed to produce a magnetic field for end-cap region detectors. Similar to the Barrel Toroid, each End-Cap Toroid also contains eight coils. Each End-Cap Toroid is 5 meters long. Its inner diameter is 1.65 meters and its outer diameter is 10.7 meters. The LHC beam pipe penetrates through the End-Cap Toroid. The End-Cap Toroids can provide a magnetic of 4.1 T for the end-cap region detectors. To maintain superconducting status, the End-Cap Toroids are cooled by the liquid helium.

3.2.6 Trigger system

The ATLAS detector has a trigger system to filter events taken from collisions. The trigger system is designed to have the capability to select interesting events from huge amounts of data. It selects events based on events' hit information taken from different systems of the ATLAS detector (such as inner tracker and calorimeters).

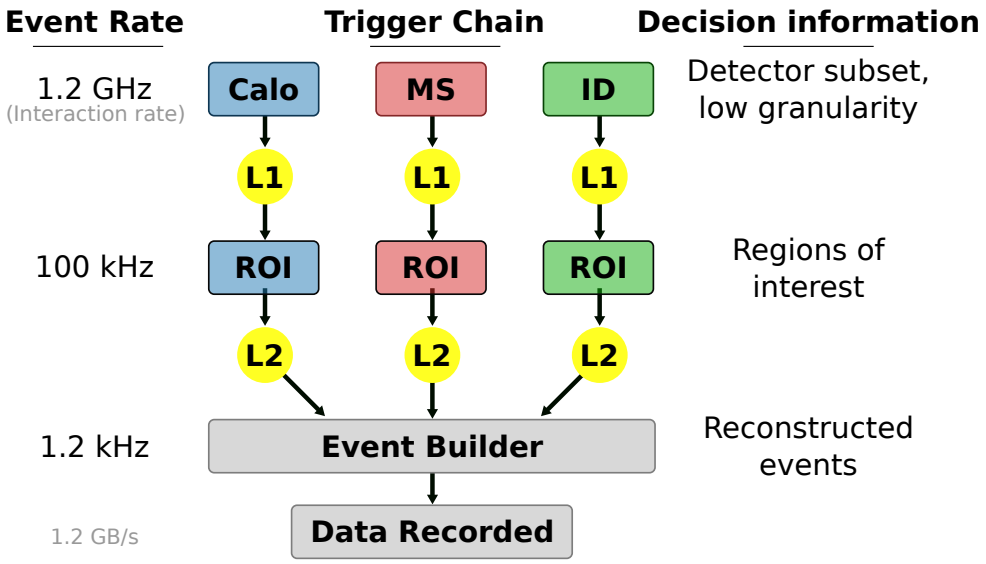


Figure 3.18: Overview of the architecture of the ATLAS' trigger system.

Figure 3.18 shows the architecture of the ATLAS trigger system. The trigger system can be divided into two parts: the Level-1 (L1) trigger and the Level-2 (L2) trigger. The interaction rate of the ATLAS detector is about 1.2 GHz. The L1 trigger reduces the event rate from 1.2 GHz to about 100 kHz and the L2 trigger further reduces the event rate to about 1.2 kHz.

The L1 trigger controls each individual detector (such as ID, the calorimeters, muon system) to accept or reject signals from one event based on the signatures of particles recorded by the detector. Events accepted by the L1 trigger will be stored in the readout buffers (ROB) for further processing.

Events passed through the L1 trigger will be further filtered by the L2 trigger. The L2 trigger selects events from the regions of interest (ROI) based on $\eta - \phi$ information. Selected events will be stored in the event buffer and then will be recorded as reconstructed events.

Chapter 4

Physics Objects Reconstruction

This chapter provides a summary of the reconstruction methods used to identify physics objects in this thesis, which include electrons, muons, jets (including b-tagged jets), and missing transverse energy. They can be reconstructed from the hits, tracks, and energy deposits inside the ATLAS detector.

Electrons and muons can be directly identified and reconstructed when they pass through different parts of the ATLAS detector. However, due to the color confinement of quarks, the detector can only observe cluster-like energy deposits formed by the hadronization of quarks and gluons, which are called jets. Since the detectors cannot observe neutrinos directly, the missing transverse energy is reconstructed as a physics object to represent the transverse energy carried by undetectable particles.

The object reconstruction process involves three major steps. The first step is the classification of signals detected by the detectors into different objects, such as electrons and jets. The second step is the object identification process, which involves algorithms that accept and reject candidate objects at different levels. The final step is the isolation cut, which involves algorithms that determine whether the candidate objects are isolated from other objects and whether they originated directly from the primary or promote interactions.

4.1 Electrons

Electrons produced from proton-proton collisions inside the ATLAS detector will pass through the Pixel Detector, the Silicon Tracker (SCT), the Transition Radiation Tracker (TRT), and finally be absorbed in the Electromagnetic Calorimeter (EM). A schematic illustration of the path of an electron through the detector is shown in figure 4.1.

4.1.1 Seed-cluster and track reconstruction

The first step of electron reconstruction is the reconstruction of electron candidates by matching the reconstructed tracks and seed-clusters of electrons in the $\eta \times \phi$ space.

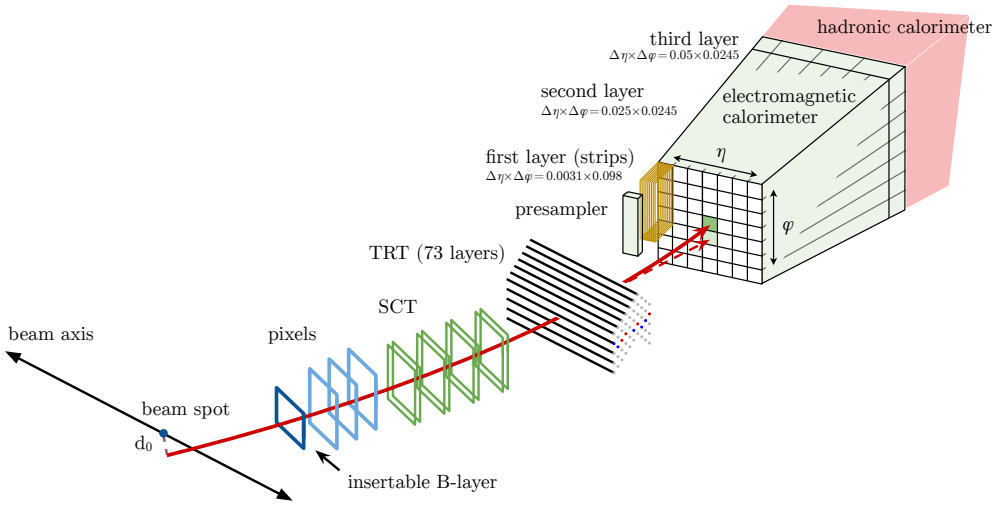


Figure 4.1: A schematic illustration of the path of an electron through the detector. The red trajectory shows the hypothetical path of an electron, which first traverses the tracking system (pixel detectors, then silicon-strip detectors, and lastly the TRT) and then enters the electromagnetic calorimeter. The dashed red trajectory indicates the path of a photon produced by the interaction of the electron with the material in the tracking system [66].

The seed-clusters are reconstructed inside the electromagnetic calorimeter. As shown in figure 4.1, the electromagnetic calorimeter contains the presampler, the first layer, the second layer, and the third layer. The electromagnetic calorimeter is divided into 200×256 elements (towers) in the $\eta \times \phi$ space and the size of each element is $\Delta\eta \times \Delta\phi = 0.25 \times 0.25$, corresponding to the granularity of the second layer of the EM calorimeter. Energy deposited in each tower by the electron is summed using energy collected in the presampler, the first layer, the second layer as well as the third layer at the same position in the $\eta \times \phi$ space. Then, a sliding-window algorithm with a window size equal to 3×5 towers in the $\eta \times \phi$ space is used to seed electromagnetic-energy cluster candidates. This process is repeated until this has been performed for every element in the calorimeter and the center of the sliding window is moved with a step size 0.025 in either the η or ϕ direction. The summed transverse energy in each cluster is required to be higher than 2.5 GeV. If two reconstructed candidates are close to each other in the $\eta \times \phi$ space, the cluster with higher E_{T} will be kept [66].

Tracks are reconstructed by the pixel detector and SCT using 'hit' in the inner tracker layers. The pixel detector and SCT assemble clusters from these hits. After that, the tracks

will be reconstructed in three steps: pattern recognition, ambiguity resolution, and TRT extension. The pattern recognition algorithm uses the pion hypothesis for the model of energy loss from interactions of the particle with the detector material. According to the hypothesis used in the pattern recognition, track candidates with p_{T} larger than 0.4 GeV are fitted using the ATLAS Global χ^2 Track Fitter. To increase the reconstruction efficiency, a second fit based on the electron hypothesis will be used to track candidates who failed in the fit and has small $\eta \times \phi$ separation to the EM clusters [66].

After track and seed-cluster are reconstructed, they are matched together to form the electron candidates if their separation in the $\eta \times \phi$ space satisfy $-0.10 < q \times (\phi_{\mathrm{track}} - \phi_{\mathrm{cluster}}) < 0.05$ and $|\eta_{\mathrm{cluster}} - \eta_{\mathrm{track}}| < 0.05$, where q is the sign of the electric charge of the particle [66].

4.1.2 Identification

After electron candidates are reconstructed, a method called likelihood (LH) identification is used to assess the probability of a candidate electron being produced by a physical electron passing through the detector. The main objective is to distinguish prompt electrons (signal) from other types, such as jets, photon conversion electrons, and hadronic decay electrons (background). For each candidate electron, fourteen quantities denoted as $\vec{x}$ are measured. These quantities characterize various properties, including the energy distribution across calorimeter layers, the impact parameter from the ID, the momentum lost by the track over time, the TRT response, and the $\eta - \phi$ match between the track and calorimeter cluster. The probability density functions (PDFs) for these quantities, denoted as $\vec{P}_{S(B)}$, are measured from simulation data for signal and background, respectively. The likelihood of a candidate being signal (background) is given in Equation 4.1.

$$L_{S(B)}(\vec{x}) = \prod_{i=1}^{14} P_{S(B),i}(x_i) \tag{4.1}$$

The discriminant d_L is defined in Equation 4.2 and peaks near one for signal, and zero for background.

$$d_L = \frac{L_S}{L_S + L_B} \tag{4.2}$$

The likelihood (LH) identification of electrons is based on increasingly restrictive thresholds of d_L, which define the `VeryLoose`, `Loose(AndBLayer)`, `Medium`, and `Tight` working points. For this thesis, the `LooseAndBLayer` and `Medium` LH identification working points

are used. Both require at least two-pixel hits and seven total hits in the silicon ID, as well as at least one-pixel hit in the innermost working pixel layer. For electrons with E_{T} =40 GeV, these working points have efficiencies of 88% and 80%, respectively.

Electrons reconstructed with a path that travels directly through a broken calorimeter cell are labeled as *BADCLUSELECTRON*. These electrons have a poor E_{T} measurement, and it is beneficial to exclude them from consideration.

4.1.3 Isolation

In this thesis, electrons produced by the signal model are produced promptly from the primary interaction point (IP), which means the electrons should be isolated from other objects. After the identification of electrons, an isolation algorithm will be used to determine whether the electrons are isolated. The activity within a $\eta - \phi$ cone of size $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ is measured using charged tracks and calorimeter energy deposits. A variable cone size of $\Delta R = \min(0.2, 10\mathrm{GeV}/p_{\mathrm{T}})$ is used to count tracks with $p_{\mathrm{T}} > 1$ GeV around the electron. The p_{T} of the tracks within this cone, excluding the electron's own tracks, is summed to define $p_{\mathrm{T}}^{\mathrm{varcone20}}$. The electron's tracks are plural to account for bremsstrahlung radiation converting to secondary electrons, which are counted as part of the electron candidate if their extrapolated track falls within $\Delta\eta + \Delta\phi = 0.05 \times 0.1$ of the primary calorimeter cluster. Meanwhile, a fixed cone size of $\Delta R = 0.2$ is used to sum up the activity in the calorimeters. First, the energy from the electron is subtracted within an area of $\Delta\eta + \Delta\phi = 0.125 \times 0.175$. Energy from pileup effects is also subtracted, and the remaining E_{T} is summed to define $E_{\mathrm{T}}^{\mathrm{topocone20}}$.

This thesis uses the isolation working point `FixedCutLoose`. The efficiency of this requirement for prompt electrons is approximately 99%.

4.2 Muons

Muons used in this thesis are reconstructed mainly based on information from the inner detectors (ID) and the muon spectrometer (MS). The calorimeters also play important roles in the determination of track parameters, muon energy correction due to energy loss in the calorimeters, and reconstruction of MS-independent tagging of ID tracks as muon candidates [67].

4.2.1 Muon track reconstruction

Muon track reconstruction inside the ID is similar to the case of the electron, where a global χ^2 fit on hit clusters in the ID is used. Muon track reconstruction in the MS

starts from the identification of short straight-line local track segments in the individual MS station. Then, track segments from different MS stations are combined together to form the preliminary track candidates and a parabolic trajectory based on a first-order approximation of the muon bending in the magnetic field. After that, the 3-D muon track candidates are reconstructed by combining the precision measurement in the bending plane and measurements of the second coordinate from the trigger detectors. A global χ^2 fit of the muon trajectory is used to remove outlier hits and add hits that are around the trajectory but are not included in the preliminary track candidates. Finally, the MS track of muon is reconstructed by a track fit using the updated hit information inside the MS.

Muons can be categorized into five types according to different reconstruction strategies: combined (**CB**), inside-out combined (**IO**), muon-spectrometer extrapolated (**ME**), segment-tagged (**ST**), and calorimeter-tagged (**CT**) muon.

The combined muons are reconstructed by the combined fit using tracks reconstructed in the MS and ID. Energy loss of muon in the calorimeters is also taken into account.

IO muons are reconstructed using a complementary inside-out algorithm, which extrapolates ID tracks to the MS and searches for at least three loosely-aligned MS hits. The ID track, the energy loss in the calorimeters, and the MS hits are then used in a combined track fit [67].

The ST muons are identified by requiring that an ID track extrapolated to the MS satisfies tight angular matching requirements to at least one reconstructed MS segment.

The ME muons are reconstructed using the MS track and the location of the interaction points when the MS track cannot be matched to the ID track. The ME muon is used to extend muon acceptance outside the ID coverage.

The CT muons are identified by extrapolating ID tracks through the calorimeters to search for energy deposits consistent with a minimum-ionizing particle [67].

4.2.2 Identification

After muons are reconstructed, additional requirements are also applied to muons in different analyses. To satisfy the requirements of various physics analyses, different sets of requirements on muons of different types are developed. The sets of requirements are called the Working Points (**WP**) of muons. Currently, there are three commonly used WPs: `Loose`, `Medium`, `Tight`, in order of increasing muon purity and decreasing selection efficiency. There are also two additional WPs designed for analyses focusing on the high p_{T} and low p_{T} muons: `Low-`p_{T} and `High-`p_{T} WP. These working points admit or reject muons based on several variables: the number of precision stations of a muon, the number of precision hole

stations, the q/p compatibility, and the ϕ'.

The precision station is defined as an MS station that the muon left at least three hits in the MDT or CSC detector, while the precision hole station refers to the MS station that does not satisfy the requirements of the precision station and misses at least three hits expected given the muon's track.

For CB and IO muons, the q/p compatibility is defined as equation 4.3, where q/p_{ID} and q/p_{MS} denotes the ratio of charge and momentum measured in the ID and MS. σ in the denominator is the corresponding uncertainty [67].

$$q/p \text{ compatibility} = \frac{|q/p_{\mathrm{ID}} - q/p_{\mathrm{MS}}|}{\sqrt{\sigma^2(q/p_{\mathrm{ID}}) + \sigma^2(q/p_{\mathrm{MS}})}} \tag{4.3}$$

ϕ' is defined as the ratio of the absolute difference between muon p_{T} measured by the ID and MS over the muon p_{T} measured by the combined fit using tracks from ID and MS, as shown in equation 4.4 [67].

$$\phi' = \frac{|p_{T,\mathrm{ID}} - p_{T,\mathrm{MS}}|}{p_{T,\mathrm{CB}}} \tag{4.4}$$

Muons used in this thesis are based on the Medium and Loose WPs. The Medium WP is the baseline muon identification WP for ATLAS physics analyses. In the region covered by ID ($|\eta| < 2.5$) except the region $|\eta| < 0.1$, the Medium WP only accepts CB and IO muons with at least two precision stations. In the region $|\eta| < 0.1$, the requirements are loose to at least one precision station and the number of precision hole stations should be less than two. To extend muon acceptance outside of the ID coverage, the Medium WP also accepts ME and SiF muon with at least three precision stations in the region $2.5 < |\eta| < 2.7$. All muons accepted by the Medium WP are required to satisfy q/p compatibility < 7.

In addition to all muons accepted by the Medium WP, the Loose WP also accepts CT and ST muons in the region $|\eta| < 0.1$. To increase the selection efficiency of low p_{T} muons, in the region $|\eta| < 1.3$ IO muons that $p_{\mathrm{T}} < 7$ GeVand have at least one precision station can also pass the Loose WP.

4.2.3 Isolation

The muons of interest in this thesis, like electrons, originate promptly from the interaction point, either through the decay of a vector boson, and are expected to be isolated from other particles in the event. On the other hand, muons from semi-leptonic decays and hadronic decays are produced in close proximity to other particles.

To identify the muons of interest, the concept of isolation is quantified by the sum of

tracks in a variable-sized cone around the muon, and four related variables are defined. The first variable, $p_{\mathrm{T}}^{\mathrm{varcone30}}$, is defined as the scalar sum of p_{T} for tracks within a cone size of $\Delta R = \min(0.3, 10\mathrm{GeV}/p_{\mathrm{T}})$. Only tracks with $p_{\mathrm{T}} > 1$ GeV are counted, and the muon's p_{T} is excluded. The second and third variables are $E_{\mathrm{T}}^{\mathrm{topocone20}}$ and $p_{\mathrm{T}}^{\mathrm{cone20}}$, which are defined as the scalar sum of E_{T} or p_{T}, respectively, within a cone size of $\Delta R = 0.2$. The fourth variable, $p_{\mathrm{T}}^{\mathrm{neu20}}$, is similar to $E_{\mathrm{T}}^{\mathrm{topocone20}}$, but it sums up neutral E_{T} within a cone size of $\Delta R = 0.2$.

These variables are used to define the three isolation working points in this thesis. The first, FixedCutTightTrackOnly, simply requires $p_{\mathrm{T}}^{\mathrm{varcone30}}/p_{\mathrm{T}} < 0.06$. The second, FixedCutPflowLoose, requires both $(p_{\mathrm{T}}^{\mathrm{varcone30}} + 0.4p_{\mathrm{T}}^{\mathrm{neu20}})/p_{\mathrm{T}} < 0.16$ and $(p_{\mathrm{T}}^{\mathrm{cone20}} + 0.4p_{\mathrm{T}}^{\mathrm{neu20}})/p_{\mathrm{T}} < 0.16$. The third, FixedCutLoose, requires both $p_{\mathrm{T}}^{\mathrm{varcone30}}/p_{\mathrm{T}} < 0.15$ and $E_{\mathrm{T}}^{\mathrm{topocone20}}/p_{\mathrm{T}} < 0.30$. While the efficiency of these isolation requirements varies with p_{T}, in general, fewer than 1% of prompt muons are lost [67].

4.2.4 Bad Muon Veto

In the high-p_{T} regime, it becomes difficult to accurately reconstruct muons due to the small bending radius in the magnetic field. A criteria named *Bad Muon Veto* (BMV) is used to address this by ignoring poorly reconstructed muons in the tails of the relative p_{T} resolution distributions, $\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}}$, given in Equation 4.5.

$$\frac{\sigma(p)}{p} = \left(\frac{p_0}{p_{\mathrm{T}}} \oplus p_1 \oplus p_2 \times p_{\mathrm{T}}\right) \tag{4.5}$$

The parameters p_0, p_1, and p_2 are measured for the MS and ID in different η regions. The first term describes uncertainty in energy loss as a muon travels through detector material and becomes less impactful at higher p_{T}. The second term covers multiple scattering and irregularities in the magnetic field. The third term dominates at high-p_{T} and describes the intrinsic spatial resolution of the muon detectors, including the accuracy of their alignment [67]. A cut is made on the relative uncertainty:

$$\frac{\sigma(q/p)}{(q/p)} < C(p_{\mathrm{T}}) \cdot \sigma_{rel}^{exp}. \tag{4.6}$$

Here $C(p_{\mathrm{T}})$ is a p_{T}-dependent coefficient which is equal to 2.5 when $p_{\mathrm{T}} < 1$ TeV and decreases linearly above this. The application of the BMV reduces efficiency by 7% for high-p_{T} muons, while removing poorly reconstructed muons that should not be considered for analysis.

4.3 Jets

In the observation of electroweak ZZ production in the four-lepton final states, The initial state vector bosons are radiated from quarks and then scatter into another pair of vector bosons in the final state. The VBS final states will have two back-to-back quarks with large angular separation. Quarks traveling through the ATLAS detector will produce narrow cones of hadrons, gluons, and other particles by hadronic interactions, which are called jets. The two back-to-back jets are key features of VBS events and are important to distinguish the VBS process and other processes.

4.3.1 Anti-k_t jet clustering algorithm

Jets used in this thesis is based on the anti-k_t jet clustering algorithm [68]. k_t in its names donates the p_{T} of jets. The algorithm starts with a list of topo-clusters. Topo-clusters are the short name for topologically-grouped noise-suppressed clusters of calorimeter cells, which can be called protojets. The topo-clusters are constructed from seed cells with energy more than four times the sum of electronic and pile-up noise. In this thesis, topo-clusters are calibrated by the electromagnetic (EM) scale so the reconstructed jets are called the EMTopo jets.

The anti-k_t jet clustering algorithm reconstructs jets by recursively merging closet protojets together until there is no protojet left. After all protojets are merged, the list of jets will be created. The detailed reconstruction steps are shown in algorithm 1 [69].

Algorithm 1 Jet clustering algorithm

Start with the initial list of protojets.
For protojet i in the list
repeat
  Define the self distance as $d_{iB} = k_{ti}^{2p}$ of the protojet i.
  For another protojet j in the list, define the distance to protojet i as $d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p})\frac{\Delta_{ij}^2}{R^2}$.
  Find the smallest of all d_{ij} including d_{iB} as $d_{\min}$.
  if d_{ij} is the $d_{\min}$ **then**
    Merge protojet i and j together as a new protojet k. p_{T} of the new protojet equals $k_{tk} = k_{ti} + k_{tj}$.
  else if d_{iB} is the $d_{\min}$ **then**
    Protojet i is not mergable. Remove it from the list of protojets and add it to the list of jets.
  end if
until No more protojet exist and finish constructing the list of jets.

In the algorithm 1, Δ_{ij} denotes the angular separation between protojet i and j, which is

shown in equation 4.7. y_i and ϕ_i are respectively the rapidity and azimuth of protojet i [69].

$$\Delta_{ij} = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \tag{4.7}$$

When two protojets i and j are merged as a new protojet k, the rapidity y_k and azimuth ϕ_k of the merged protojet k equals the weighted average of values of protojets i and j, which is shown in equation 4.8.

$$\begin{aligned} y_k &= \frac{k_{ti} \times y_i + k_{tj} \times y_j}{k_{tk}} \\ \phi_k &= \frac{k_{ti} \times \phi_i + k_{tj} \times \phi_j}{k_{tk}} \end{aligned} \tag{4.8}$$

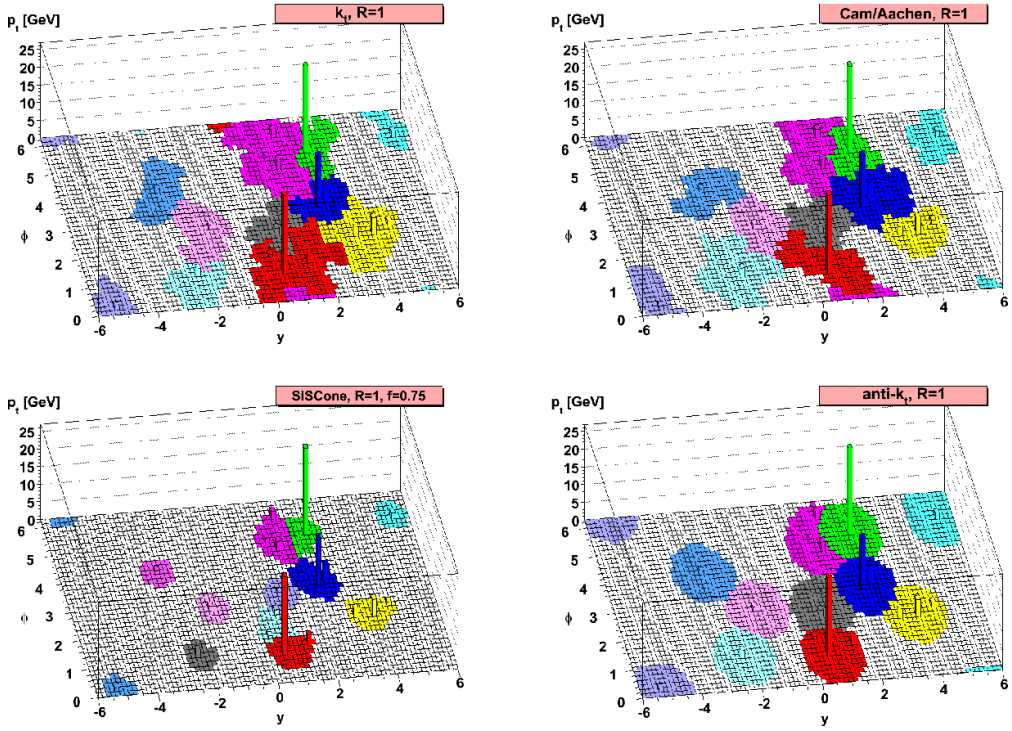


Figure 4.2: A sample parton-level event, together with many random soft "ghosts", clustered with four different jets algorithms, illustrating the "active" catchment areas of the resulting hard jets. For k_t and Cam/Aachen the detailed shapes are in part determined by the specific set of ghosts used and change when the ghosts are modified [68].

The parameter R used in the jet clustering algorithm is analogous to the cone size parameter. In practice, the value of R should be of order 1. In this thesis, jets are reconstructed using the jet clustering algorithm with $R = 0.4$. The jet clustering algorithm also depends

on the parameter p. For $p = 1$, the jet clustering algorithm is the inclusive k_t algorithm. The case of $p = 0$ corresponds to the inclusive Cambridge/Aachen algorithm. The jet clustering algorithm used in this thesis is based on $p = -1$, which denotes the name of anti-k_t. The shape of the jet reconstructed by the anti-k_t algorithm is not influenced by soft radiation [68]. Examples of reconstructed jet shapes using the jet clustering algorithm with different p values are shown in figure 4.2.

4.3.2 B-tagging jet

Jets originating from bottom quarks and the decay of B-hadrons, known as b-jets. The b-jets are important to physics analysis related to the top quark since $t \to b + W$ has the highest branching ratio in the top quark's decay. In this thesis, b-jets are used to define control regions to estimate backgrounds from $t\bar{t}$ productions. To distinguish b-jets from other light flavor jets, a multivariate discriminant, MV2c10, is used. The MV2 algorithm combines a gradient-boosted decision tree (GBDT) [70] and other low-level algorithms based on track information of jets. The GBDT uses the kinematic of jets as input variables and is trained on the hybrid $t\bar{t} + Z'$ sample [71]. An identification that tags 85% of b-jets defines the "b-tag working point" that is useful for rejecting events containing b-jets and is used in this thesis.

4.3.3 Jet vertex tagger

Proton-proton collisions at the LHC result in not only the hard-scatter interaction of interest but also additional interactions, which are referred to as pile-up interactions. The pile-up events will introduce additional transverse energy flow, which will be subtracted from the signal interaction of interest. In the jet reconstruction process, the Local fluctuations in the pileup activity, however, may result in spurious pileup jets. To reduce the effect of pile-up jets, a multivariate discriminant called the jet-vertex-tagger (JVT) is used. The JVT is constructed using the k-nearest neighbor (KNN) algorithm in the 2-D R_{p_T} and corrected-jet-vertex-fraction (corrJVF) space [72]. R_{p_T} is defined as the ratio of scalar p_{T} sum of tracks associated with the jet and the primary interaction vertex over the calibrated jet p_{T}, which is shown in equation 4.9.

$$R_{p_T} = \frac{\sum_k p_T^{\mathrm{trk}_k}(\mathrm{PV}_0)}{p_T^{\mathrm{jet}}} \tag{4.9}$$

The definition of corrected jet-vertex-fraction is similar to R_{p_T} shown in equation 4.9, while the denominator of corrJVT replaced by the scalar sum p_{T} of all associated jets with

correction for average scalar sum p_{T} to pileup tracks associated with a jet, which is shown in equation 4.10. The term $p_T^{\mathrm{PU}} = \sum_{n\geq 1}\sum_l p_T^{\mathrm{trk}_l}(\mathrm{PV}_n)$ represents the scalar sum p_{T} of tracks associated with the jet but originate from any of the pileup interaction vertices [72].

$$\mathrm{corrJVF} = \frac{\sum_k p_T^{\mathrm{trk}_k}(\mathrm{PV}_0)}{\sum_l p_T^{\mathrm{trk}_l}(\mathrm{PV}_0) + \frac{\sum_{n\geq 1}\sum_l p_T^{\mathrm{trk}_l}(\mathrm{PV}_n)}{k\cdot n_{\mathrm{trk}}^{\mathrm{PU}}}} \tag{4.10}$$

For each point in the 2-D R_{p_T}-corrJVF space, the ratio of the number of hard-scatter (signal) jets over the number of hard-scatter and pileup jets found in the local neighborhood around the point using the training sample equals the relative probability for a jet at that point to be the hard-scatter jet. For each point, the local neighborhood is defined as the 100 closet neighbors around the point based on the 2-D Euclidean distance [72].

In this thesis, a cut on the JVT is used to select jets from the hard-scatter vertex.

4.4 Missing transverse momentum

Due to the properties of neutrinos, they rarely interact with the materials of the ATLAS detector and cannot be directly measured and reconstructed. However, in the center-of-mass (COM) frame, the total p_{T} of the initial state is almost zero, since two proton beams collide with each other head-on. Therefore, the sum of p_{T} of all particles produced by the collision should also be close to zero. As a result, the negative sum of p_{T} of all produced objects equals the sum of p_{T} carried by neutrinos and other undetectable particles, if any. This quantity is known as the missing transverse momentum E_T^{miss}. To obtain the total measured p_{T} of the event, the transverse momenta of muons, electrons, and the remaining tracks in the ID are summed.

Since this thesis focuses on the four-lepton final state produced by the Z (Z* and Z') decay, there should be no neutrinos produced. The amount of E_T^{miss} should be close to zero for the $ZZ \to 4l$ events. Although the E_T^{miss} is not used in the signal region selections, the E_T^{miss} is useful to define control regions to estimate the non-prompt lepton backgrounds.

Chapter 5

Commonly used Statistical Methods in Particle Physics

This chapter introduces statistics methods used in this thesis, including statistical fitting methods used to calculate limit and significance and multivariate analysis methods used to do event classifications.

5.1 Statistical hypothesis testing

In particle physics, theories need to be examined by real data using a statistical method called statistical hypothesis testing, which is a method of statistical inference to decide whether the real data support a certain hypothesis. In particle physics, we often have a null hypothesis, normally the background-only model, in which the interaction or theory doesn't exist. The null hypothesis is often denoted as H_0. In contrast, the hypothesis in which the interaction or model exists is called the alternative hypothesis, the signal-plus-background model. The alternative hypothesis can be denoted as H_1 or H_a. Hypothesis testing is designed to calculate the string of evidence against the null hypothesis.

5.1.1 The p-value

The p-value is an important concept in statistical hypothesis testing. The P-value describes the probability of getting a test result at least as extreme as the observed result under the null hypothesis. If the p-value is very small using observed data, the null hypothesis is very unlikely to be correct. Normally the p-value can be gotten by calculating the portion of the area of results as extreme as observed data under the probability density distribution following the null hypothesis. Assume the number of events in one experiment should

observe an unknown distribution T under the null hypothesis and the observed number of events is t, the p-value is:

$$\begin{aligned} p &= Pr(T \geq t|H_0) \quad \text{for the one-sided right-tail test} \\ p &= Pr(T \leq t|H_0) \quad \text{for the one-sided left-tail test} \\ p &= 2\text{min}Pr(T \leq t|H_0), Pr(T \geq t|H_0) \quad \text{for the two-sided test} \end{aligned} \tag{5.1}$$

5.1.2 Statistical significance

Followed by the definition of p-value, the statistical significance level, often denoted as *alpha*, describes the level of probability that the null hypothesis is true. In particle physics, the statistical significance level is described using multiples of the standard deviation sigma σ of the normal distribution. Two sigmas, 2σ means the p-value of the null hypothesis equals 5%. Normally we use a much stricter level in particle physics. Under a normal distribution, 3σ, p-value equals 2.7×10^{-3}, means the evidence that the null hypothesis is not true is found and 5σ, p-value equals 5.7×10^{-7} means that the null hypothesis is proved incorrect, which physicists will call a discovery.

5.1.3 Likelihood function

In statistics, the likelihood function represents the probability a hypothesis is correct conditional on observed data. In particle physics experiments, the number of events in one experiment normally follows the Poisson distribution. To measure the strength of one interaction, normally a factor denoted as μ, which is called the signal strength, will be applied to the number of signal events. The parameter μ is also called the parameter of interest since the goal is to find the best μ value fitting observed data. $\mu = 0$ corresponds to the background-only model and a positive μ value corresponds to signal in addition to the background. There are also other parameters that will affect the likelihood function such as systematic uncertainties. Those parameters are called the nuisance parameters and are often denoted as θ. The likelihood function as a function of μ and θ can be written as equation 5.2.

$$L(\mu, \theta|\text{data}) = \sum_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \tag{5.2}$$

In equation 5.2, s_j represents a number of signal events and b_j represents the number of background events at each bin. n_j is the number of events observed in the experiment at bin

j. The μ value can be calculated by maximizing the value of the likelihood function based on the observed data. Normally, in addition to the region where the μ value will be directly calculated, several regions will also be set to constrain the scale factor on backgrounds. Those regions are called control regions. Including the control region, the likelihood function will become equation 5.3.

$$L(\mu, \theta|\text{data}) = \sum_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \sum_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-m_k} \tag{5.3}$$

m_k is the number of background events predicted by theory and u_k is the observed number of events at each bin.

5.1.4 The profile likelihood ratio and CLs method

In experimental particle physics, the profile likelihood ratio [73] and CLs method [74] are commonly used to set upper limits or significance of discovery of new signals. The upper limit means the largest intrinsic intensity that the source can have and yet have a given probability of remaining undetected. The statistical upper limit can be calculated using the CLs method based on the likelihood ratio of the alternative and null hypotheses. For example, if the observed number of events is x, the likelihood under the null hypothesis is $L(H_0|x)$ and the likelihood under the alternative hypothesis is $L(H_1|x)$. The likelihood ratio of the alternative hypothesis over the null hypothesis $\frac{L(H_1|x)}{L(H_0|x)}$, which can be represented as Λ, represents the level that observed data x prefers the alternative hypothesis compared to the null hypothesis. The negative log transformation of the likelihood ratio is the q-value in statistics, which is shown in equation 5.4.

$$\begin{aligned} \Lambda &= \frac{L(H_1|x)}{L(H_0|x)} \\ q &= -2ln\Lambda \end{aligned} \tag{5.4}$$

As described in the previous section, the likelihood function is a function of the signal strength μ and nuisance parameters θ, and the likelihood ratio Λ can be written as a function of μ since only μ is the parameter of interest, as shown in equation 5.5.

$$\Lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}}_\mu|n)}{L(\hat{\mu}, \hat{\theta}|n)} \tag{5.5}$$

In the denominator of equation 5.5, $\hat{\mu}$ and $\hat{\theta}$ are values of μ and θ that maximize the value

of likelihood function with all parameters floating, which can be called the unconditional maximum-likelihood estimator. In the numerator, $\hat{\hat{\theta}}$ means the value of θ that maximizes the value of L for a fixed value of μ. $\hat{\hat{\theta}}_\mu$ is also called the conditional maximum-likelihood estimator of θ. All of those values can be set by statistical fitting given the number of predicted or observed the number of events n. The q value as a function of the signal strength μ then can be written as equation 5.6.

$$q_\mu = \begin{cases} -2ln\Lambda(\mu) = -2ln\frac{L(\mu,\hat{\hat{\theta}}_\mu|n)}{L(\hat{\mu},\hat{\theta}|n)} & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} \tag{5.6}$$

In statistics Wilks' theorem [75], the distribution of $2ln\frac{L(H_1|x)}{L(H_0|x)}$, or $-2ln\frac{L(H_0|x)}{L(H_1|x)}$, asymptotically approaches the chi-squared (χ^2) distribution when the sample size approaches ∞ if x follows hypothesis H_0. Degree freedom of the χ^2 distribution equals the difference between parameter spaces of the likelihood function on the numerator and denominator ($df_1 - df_0$). Since in equation 5.5, the difference of parameter space is only the signal strength μ, the distribution of Λ follows a chi-square distribution with degree of freedom (k) equals 1. The relationship between the value of q_u and the significance level can be found in table 5.1.

Table 5.1: $\Pr(q \geq q_\mu)$ and q_μ value under chi-square distribution with k=1

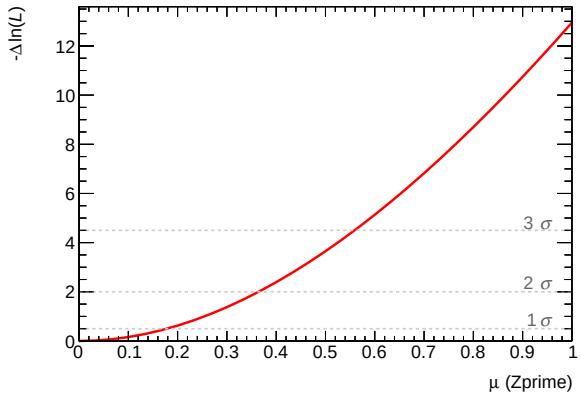
$\Pr(q \geq q_\mu)$	0.1	0.05	0.025	0.01
value of q_μ	2.71	3.84	5.02	6.63

The probability that the Q value is larger than the observed Q value under the s+b model is called the confidence level of the s+b model (CL_{s+b}). In contrast, the probability that the Q value is larger than the observed Q value under the b-only model is called the confidence level of the b-only model (CL_b). Then the CL_s is defined in equation 5.7

$$\begin{aligned} CL_{s+b} &= Pr(q_\mu \geq q_\mu^{obs}|s+b) \\ CL_b &= Pr(q_0 \geq q_0^{obs}|b) \\ CL_s &= \frac{CL_{s+b}}{CL_b} \end{aligned} \tag{5.7}$$

If the $CL_s < 5\%$, the s+b hypothesis can be rejected at the 95% confident level. For new physics, the signal strength is normally unknown. In this case, the 95% upper limit on the interaction means that the value of signal strength will let the s+b hypothesis be rejected at the 95% confident level. The 95% upper limit can be calculated using a likelihood scan.

The likelihood scan means value of μ will be continuously scanned until the $CL_s < 5\%$ and at that time the corresponding value of μ is the 95% upper limit on the signal strength.



(a)

Figure 5.1: Likelihood scan plot under the background-only model. The value of Y-axis equals $\frac{\Lambda}{2}$. The corresponding value of μ when $-\Delta\ln(L)$ equals 1.92 (the 2σ dash line) is the 95% upper limit on signal strength μ.

The significance of the discovery of new signals means the confidence level that the background-only model can be rejected. Since $\mu = 0$ corresponds to the background-only model, we can calculate the Q value q_0 when $\mu = 0$. When $CL_b < 5.7 \times 10^{-7}$, the derivation between data and background-only model equals 5σ and we can claim a 5σ discovery of the signal.

5.2 Multivariate analysis methods

Nowadays, more and more multivariate analysis methods are used in experimental high-energy physics. In this thesis, two different MVA methods are used in the two studies: boosted decision tree and neural network. This sector provides a brief introduction to the MVA methods used in this thesis.

5.2.1 Boosted decision tree

Decision tree

The decision tree [76] is a tree-based method that is used to classify different objects or regress different numerical values and it is a non-parametric supervised learning method. Non-parametric means the decision tree model doesn't contain any learnable parameters and

the model does not make explicit assumptions about the functional form of $p(x|Y = c_k)$ or $P(Y = c_k|X = x)$. Supervised learning means the model is trained on labeled data.

The decision tree methods recursively split the input space into rectangular boxes. At each step, the decision tree asks a question about one specific variable and split all events from the previous selection into its left or right leaf node. When the splitting stops, the node is called the leaf node of the decision tree, and each left node will be labeled with its majority class (classification tree) or numerical value (regression tree). The optimal goal is that each box should contain nodes mostly from the same class. When the construction of the decision tree is finished, each event will be labeled with the label of the box it falls into. Since each split will split the input into two rectangular, each split of the decision is linear. The decision tree model is a combination of many different splits and can have very complex decision boundaries so it is a non-linear model.

At each node, the decision tree will greedily choose the best split. Quality is defined as the decrease of impurity or loss. The most commonly used impurity measurements are the Gini index and entropy. The definition of the Gini index and entropy is:

$$\begin{aligned} \text{Gini(m)} &= \sum_{k=1}^{n} p_k(m)(1 - p_k(m) \\ \text{Entropy(m)} &= -\sum_{k=1}^{n} p_k(m) log(p_k(m)) \end{aligned} \tag{5.8}$$

where $p_k(m)$ refers to the purity of class k in node m:

$$p_k(m) = \frac{1}{n_m} \sum_{x_i \in R_m} I(y_i = c_k) \tag{5.9}$$

Then the quality of the split is defined as (the impurity is represented by l):

$$Q_{\text{split}} = l_{\text{parent}} - (\frac{N_{\text{left}}}{N_{\text{parent}}} l_{\text{left}} + \frac{N_{\text{right}}}{N_{\text{parent}}} l_{\text{right}}) \tag{5.10}$$

The decision tree will maximize the decrease of impurity at each split, which means it is a greedy method without global optimization.

The biggest advantage of the decision tree method is that it is easy to construct and easy to interpret. The tree model can be visualized to show the variables used in each split. But, as mentioned before, the decision tree is a greedy method and lacks global optimization ability. What's more, the ancestor node's selection has a big impact on its descendant nodes, which means if the splits of the first few nodes are biased due to noisy events the tree structure will be completely different. In another word, the simple decision tree model

is not stable.

Boosting methods

To solve the disadvantages of the simple decision tree model discussed in the previous section, three different methods are proposed: bagging, random forest, and boosting.

Bagging means fitting many different trees and each tree is trained on bootstrapped training data. The classification result is chosen by using a majority vote from all trees. By using different bootstrap training data for each tree, the variance of the model is reduced.

The random forest method is based on bagging. The random forest method not only uses bootstrapped training data for each tree but also uses a sampled subset of training variables for each tree. By using a subset of training variables for each tree, the random forest method further reduces model variance compared with the bagging method.

The main idea of boosting methods is fitting many different trees and each tree is based on the fitting result from its previous rounds. The most commonly used boosting methods are Ada-boost and gradient-based boosting methods. In this thesis, gradient boost decision tree [70] and XGBoost [77] are used so that this sub-section will focus on gradient-based boosting methods.

In a nutshell, the main idea behind gradient-based boosting methods is to find the global minimum points of the loss function. The most commonly used optimization method for finding minimum points of a function is gradient descent. The idea is to take repeated steps in the opposite direction of the gradient (or approximate gradient) of the function at the current point because this is the direction of the steepest descent. If θ is a parameter of one function that needs to be updated and θ_t means the parameter at step t. At step t, the parameter θ will be updated based on:

$$\begin{aligned} \theta_t &= \theta_{t-1} - \alpha \cdot g \\ g &= \frac{\partial f(x, \theta_{t-1})}{\partial \theta_{t-1}} \end{aligned} \tag{5.11}$$

where α is learning rate that controls the speed of parameter θ moving towards minimum point and g is the gradient of objective function $f(x, \theta)$. Continuously updating parameters following the direction of the negative gradient the objective function will reach the local or global minimum point.

The tree construction policy of the gradient boost decision tree method is based on gradient descent. If x_i refers to input variable of sample i and y_i refers to the label of sample i, the tree construction method is:

Algorithm 2 Gradient Boost Decision Tree

Initialize the first tree f_0, the growth policy is the same as decision tree
From $t = 1$ to T,
repeat
At step t, calculate the gradient of each sample from previous step:$\tilde{y_i} = -\frac{\partial L(y_i, F_{t-1}(x))}{\partial F_{t-1}(x)}$
Fit the t th tree: $\omega^* = \text{argmin}_\omega \sum_{i=1}^{N} (\tilde{y_i} - f_t(x_i, \omega))^2$
Find the best learning rate: $\rho^* = \text{argmin}_\rho \sum_{i=1}^{N} L(y_i, F_{t-1}(x_i) + \rho \cdot f_t(x_i)$
Add the t th tree: $F_t = F_{t-1} + \rho \cdot f_t$
until Loss doesn't decrease or reach the maximum step T

The gradient boost model is based on the regression tree, which means each node's score is not set by a majority vote. Since for each tree, the objective function is the negative loss of model from the last step, each node's score equals the weighted average of gradients of all samples in the current node, which is shown in equation 5.12:

$$\omega_k = \frac{\sum_{i \in n_k} g_i}{n_k} \tag{5.12}$$

where k refers to node k and n_k is all samples that fall into node k.

Similar to the single classification tree, the node split of each tree in the boost decision tree is also a greedy method and its goal is to minimize the loss function. Assume g_i is the negative loss of sample x_i from model $F_{t-1}(x_i, \omega)$ at step $t-1$, the objective function will be (i refers to all events in the parent node that need to be split):

$$\begin{aligned} &\min_{s,d} \sum_{i \in L} (g_i - \bar{g}_L)^2 + \sum_{i \in R} (g_i - \bar{g}_R)^2 - \sum_{\forall g_i} (g_i - \bar{g})^2 \\ &= \sum_{i \in L} g_i^2 - n_L \cdot \bar{g}_L^2 + \sum_{i \in R} g_i^2 - n_R \cdot \bar{g}_R^2 - (\sum_{\forall g_i} g_i^2 - n \cdot \bar{g}^2) \end{aligned} \tag{5.13}$$

Remove constant values from equation 5.13, the equation can be written as:

$$\begin{aligned} \max_{s,d} \quad & n_L \cdot \bar{g}_L^2 + n_R \cdot \bar{g}_R^2 \\ &= n_L \cdot (\frac{1}{n_L} \sum_{i \in L} g_i)^2 + n_R \cdot (\frac{1}{n_R} \sum_{i \in R} g_i)^2 \\ &= \frac{(\sum_{i \in L} g_i)^2}{n_L} + \frac{(\sum_{i \in R} g_i)^2}{n_R} \end{aligned} \tag{5.14}$$

In short, when splitting a node, GBDT will choose a variable split that can maximize the average sum square of sample loss in the new left and right child node.

The final predicted value of the gradient boost decision tree model is a weighted sum of values from all trees.

The gradient boost decision has good performance on many problems in different areas. This method is widely used in both high-energy physics analysis and the industry world to solve real-world problems. The Facebook company built a very successful ads recommendation system based on the gradient-boosted decision tree and logistic regression and the model provided a huge amount of revenue to the company.

The gradient boost decision tree uses the gradient descent method and only the first-order gradient is used in GBDT. XGBoost is a popular gradient-based decision tree model that uses both the first and second-order gradient. The parameter updating method that XGBoost used is the Newton method. In the Newton method, each parameter is not updated simply towards the direction of the negative gradient but uses the first-order and second-order gradient to find the updating direction. Another difference between the GBDT and XGBoost methods is that the XGBoost method includes the L2 regulation term into its loss function that can control tree size and node value in order to reduce the model's variance.

Algorithm 3 Newton method

At iteration t-1, the model parameter is θ_{t-1}

repeat

The second Taylor expansion of loss function L: $L(\theta_t) \approx L(\theta_{t-1}) + L'(\theta_{t-1})\Delta\theta + L''(\theta_{t-1})\frac{\Delta\theta^2}{2}$.

Use g and h to represent first-order and second-order gradient: $L(\theta_t) \approx L(\theta_{t-1}) + g\Delta\theta + h\frac{\Delta\theta^2}{2}$.

Minimize $g\Delta\theta + h\frac{\Delta\theta^2}{2}$.

Let: $\frac{\partial(g\Delta\theta + h\frac{\Delta\theta^2}{2})}{\partial\Delta\theta} = 0$, $\Delta\theta = -\frac{g}{h}$.

Update model parameter: $\theta_t = \theta_{t-1} - \frac{g}{h}$.

until Reach the maximum iteration T

In contrast to the gradient-boosted decision tree method, node values of the XGBoost tree are calculated directly using the Newton method. The objective function of the XGBoost is the sum of the loss function and regulation terms:

$$\text{obj}(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{5.15}$$

where Ω is the regulation term. According to the Newton method, the second-order Tyler expansion of loss function will be used:

$$\text{obj}(\theta) = \sum_{i=1}^{n} [l(y_i, \hat{y}_i^{t-1}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \sum_{k=1}^{K} \Omega(f_k) \tag{5.16}$$

In equation 5.16, the sum of the loss function is across all samples. This equation can be changed by summing up all samples in the same node first. The regularization term includes the sum of the square of all leaves' score, which is similar to the L2 regularization, and the number of nodes in each tree (T in equation 5.17), which controls the complexity of each tree. The equation will become:

$$\begin{aligned}\text{obj}^t &\approx \sum_{i=1}^{n}[g_i w_{q(x_i)} + \frac{1}{2}h_i w_{q(x_i)}^2] + \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2 \\ &= \sum_{j=1}^{T}[(\sum_{i\in I_j} g_i)w_j + \frac{1}{2}(\sum_{i\in I_j} h_i + \lambda)w_j^2] + \gamma T\end{aligned}$$

where I_j means the set of samples in jth node of the tree. The equation can be further simplified by using $I_j = \{i|q(x_i) = j\}$, $G_j = \sum_{i\in I_j} g_i$ and $H_j = \sum_{i\in I_j} h_i$:

$$\text{obj}^t = \sum_{j=1}^{T}[G_j w_j + \frac{1}{2}(H_j + \lambda)w_j^2] + \gamma T \tag{5.17}$$

Based on the Newton method, the best score of node j will be $w_j^* = -\frac{G_j}{H_j+\lambda}$, while in the GBDT method, the score is: $w_j^* = -\frac{G_j}{n_j}$. The optimized objective function is:

$$\text{obj}^* = -\frac{1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j + \lambda} + \gamma T \tag{5.18}$$

When choosing a new split, gain from the split is the change of objective function and it is defined as:

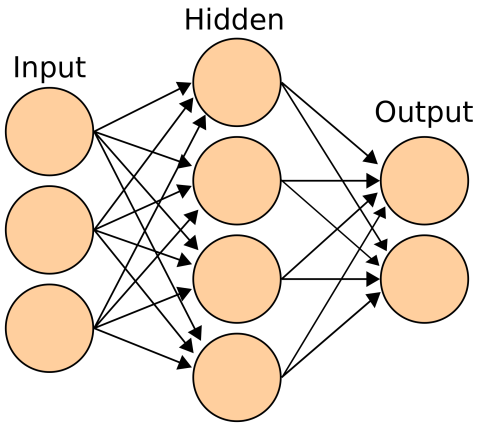
$$\text{Gain} = \frac{1}{2}[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda}] - \gamma \tag{5.19}$$

At each node, the split that can achieve the maximum gain will be chosen. Similar to the GBDT method, the final predicted value of the gradient boost decision tree model is a weighted sum of values from all trees.

5.2.2 Artificial neural network

The Artificial Neural Network (ANN), also called a Neural Network (NN), is a multivariate analysis model inspired by the human brain [78]. The Neural Network is based on a collection of connected artificial neurons. A typical neural network contains an input layer, hidden layers, and an output layer, as shown in figure 5.2

Figure 5.2 shows a fully connected neural network, which means all nodes between two neighbor layers are connected together. The first layer is called the input layer, which



(a)

Figure 5.2: Structure of fully connected Artificial Neural Network (ANN) with one hidden layer[1].

normally transforms low-dimensional input features into high-dimensional features. The last layer is called the output layer, which is used to output scores for different classes (normally on output node represents the score of one class). Since a fully connected neural network is formed by stacked layers, fully connected neural work can also be called a Multi-Layered Perceptron (MLP).

The transformation inside each layer is shown in equation 5.20, which is also known as the forward propagation.

$$\begin{aligned} z_l &= w_l^T \cdot a_{l-1} + b \\ a_l &= \sigma(z_l) \end{aligned} \tag{5.20}$$

in which w_l refers to the weight matrix of the lth layer and b is the bias term of the l the layer. In the fully connected neural network, each layer contains the weight matrix, bias term, and an activation function. The activation function is very important because it will do a non-linear transformation on the outputs. Without activation functions, the neural network is only a linear function no matter how many hidden layers it has. In equation 5.20, $\sigma(z_l)$ refers to the activation function and a_l refers to the output of lth layer after the activation function.

Since the purpose of the activation function is to do non-linear transformation, there are many different activation functions that can be used in the neural network. The equation shows the four most commonly used activation functions and their plots are shown in figure

5.3.

$$
\begin{aligned}
\text{Sigmoid: } \sigma(x) &= \frac{1}{1+e^{-x}} \\
\tanh\text{: } \sigma(x) &= \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \\
\text{ReLU: } \sigma(x) &= \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases} \\
\text{Leaky ReLU: } \sigma(x) &= \begin{cases} 0.01x & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases}
\end{aligned}
\tag{5.21}
$$

The ReLU activation function is the most commonly used activation function in hidden layers of neural works and the biggest advantage is the high speed. Leaky ReLU is a revised version of ReLU, which is used to solve the gradient vanishing problem of ReLU. The sigmoid function is often used in the output layer of a neural network and it can transform the input's range into [0, 1], which can be directly used as the probability of binary class classification. The output. The tanh function is often used in the recurrent neural network (RNN), which is used on time-series data.

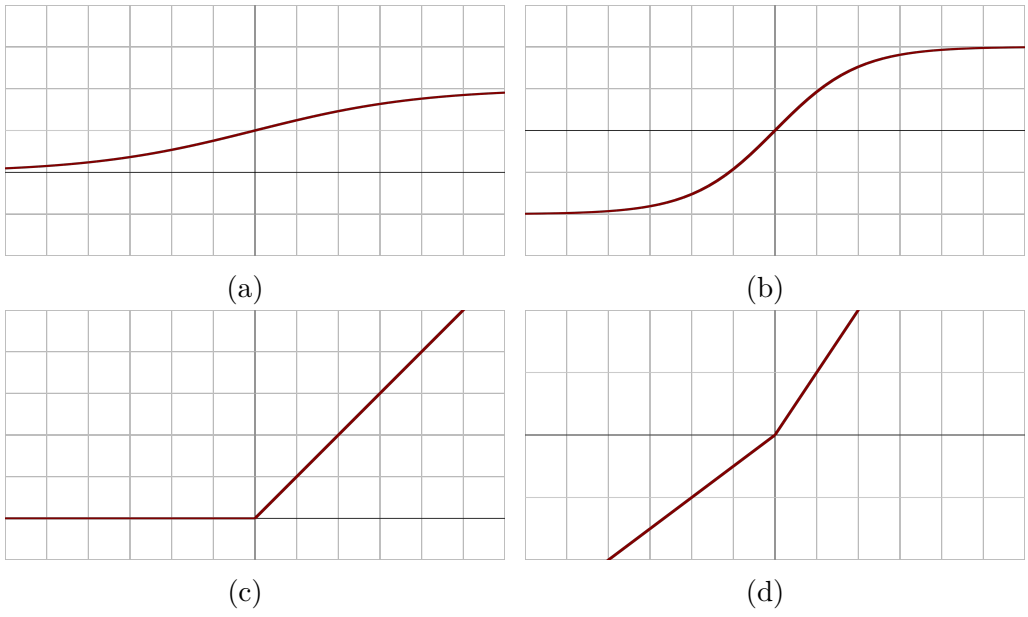


Figure 5.3: Commonly used activation functions in the neural network. (a): Sigmoid function. (b): Hyperbolic tangent (tanh) function. (c): Rectified linear unit (ReLU) function [79]. (d): Leaky rectified linear unit (Leaky ReLU) function [80].

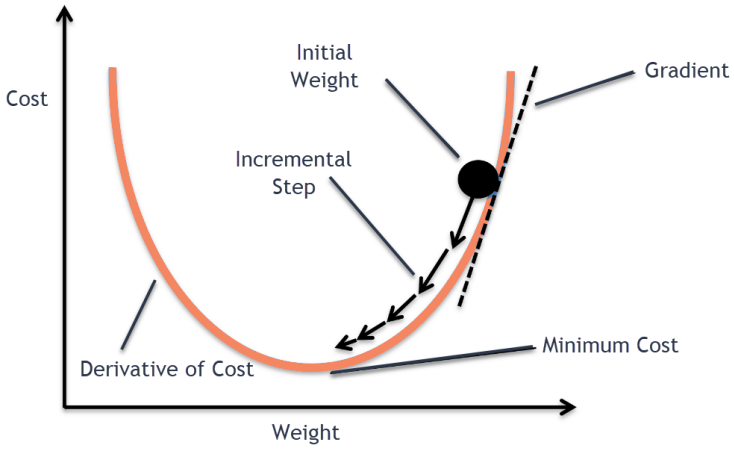
As shown in equation 5.20, each layer contains a weight matrix w and bias term b, which

are learnable parameters of a neural network. The training process of the neural network is updating those parameters. The basic idea behind updating parameters of the neural network is gradient descent, which is shown in equation 5.11. The first step is to calculate the gradient of each parameter, which is done by back-propagation (BP) [81]. The back-propagation calculates the gradient of each layer by using the chain rule, which computes the gradient of one layer each time and iterates backward to the last layer. Gradient propagation process between two neighbor layers (layer l and layer $l+1$) is shown in equation:

$$\begin{aligned} \frac{\partial L(\hat{y}, y)}{\partial w^l} &= \frac{1}{n}\delta^l \cdot a^{(l-1)T} \\ \delta^l &= w^{(l+1)T} \cdot \delta^{l+1} * \sigma'^l(z^l) \end{aligned} \tag{5.22}$$

in which $L(\hat{y}, y)$ is the loss of the neural network corresponding to prediction and ground truth. n refers to the number of samples in each batch. During the training process of the neural network, each layer's gradient is calculated by back-propagation and those parameters are updated by the optimizer based on gradients.

As shown in equation 5.11, gradient descent is one of the commonly used optimizers in the neural network. Gradient descent will update each parameter towards the direction of its negative gradient and the speed is controlled by a constant called learning rate. In this way, the loss of the neural network will move toward the local or global minimum point. This process is shown in figure 5.4.



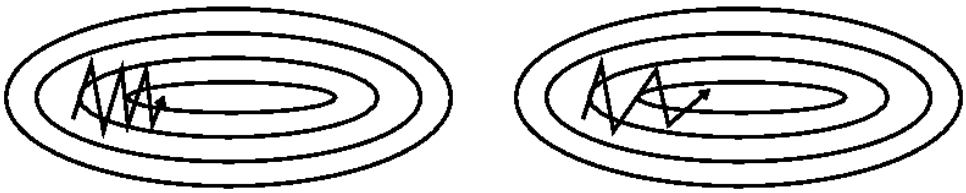
(a)

Figure 5.4: Illustration of gradient descent (GD) process used in the neural network. [2]

In the normal gradient descent, the gradient refers to the average of all samples' gradients.

In practice, calculating gradients of all samples requires loss of computing source, especially for large-size neural networks, and loss of input samples. There are two variants of gradient descent: stochastic gradient descent (SGD) and mini-batch gradient descent. The stochastic gradient descent (SGD) method randomly chooses one sample from the input and calculates gradients based on it. SGD method can reduce the time cost of gradient descent significantly since each time only one sample is used. The disadvantage of SGD is that updates taken towards the minimum point are noisy, which can often lead the gradient descent into other directions and it may take more steps to reach the minimum point. As for the mini-batch gradient descent, the gradient is calculated using a subset of all input samples, which is also called a batch. The batch size is often chosen as 32, 64, 128, etc, which is due to the architecture of modern GPUs. Compare with SGD, mini-batch gradient descent avoids noisy updates in SGD and it is faster than gradient descent.

Simple gradient descent optimizers have some disadvantages. One disadvantage is it is hard for gradient descent to navigate ravines. Ravines mean areas or point the absolute value of gradient in one direction is much larger or smaller than gradients in other directions. If the gradients are visualized as a counterplot, ravines mean an area the surface curves much deeper in one direction than in other directions. In these areas, the gradient descent method oscillates around the ravine and may not be able to reach the optimal point.



(a)

Figure 5.5: Gradient descent without momentum (left) and with momentum (right). Momentum reduces oscillations in the vertical direction. [3]

Momentum is a method that can help the gradient descent method to quickly pass through ravines and reduce oscillations [81]. This method borrows the concept of "momentum" from physics. As we know, in physics momentum is defined as:

$$\vec{p} = m\vec{v} \tag{5.23}$$

in which m refers to object's mass and $\vec{v}$ refers to the velocity. If a ball is pushed down a hill, the ball will accelerate while rolling down and the ball will also accumulate momentum. If the momentum is large enough the ball can easily pass through ravines or bumps. The

math format of gradient descent of momentum is:

$$\begin{aligned} m_t &= \gamma m_{t-1} + \eta \nabla_\theta L(\theta_{t-1}) \\ \theta_t &= \theta_{t-1} - m_t \end{aligned} \tag{5.24}$$

in which m is called velocity, which is the accumulation of gradients. Parameter θ is updated with negative velocity.

As mentioned before, the gradient descent method has one hyper-parameter: the learning rate, which controls the scale of each update to the parameter θ. The choice of learning rate will affect the performance of the neural network. Generally speaking, a larger learning rate will let loss decreases faster but loss may not be able to reach the minimum point. A smaller learning rate can guarantee loss reaches the minimum point but it will take a longer time. With a constant learning rate, the loss will be noisy when close to the minimum point. In this case, learning rate decay is a commonly used technique to choose different learning rates at different training stages. Equation 5.25 shows one commonly used learning rate decay method. n refers to epoch number and α_0 is the initial learning rate.

$$\alpha_n = \frac{1}{1 + \eta \cdot n} \cdot \alpha_0 \tag{5.25}$$

The learning rate decay method provides different learning rates for different epochs, which makes loss decrease fast in the first few epochs and makes loss decrease slowly when close to the minimum point. Learning rate decay uses the same learning rate in different directions. From the counterplots shown in figure 5.5, it is clear that larger learning rates should be used in directions with smaller gradients and smaller learning rates should be applied to directions with larger gradients. In this case, oscillation can be reduced and loss can decrease fast along directions with smaller gradients. One optimizer that solves this problem is the Root Mean Square Propagation (RMSProp).

The goal of RMSProp is to accelerate the optimization process. In other words, the RMSProp algorithm is designed to speed up the loss decrease and reduce oscillation in different directions. The math format of RMSProp is:

$$\begin{aligned} G_t &= (1 - \beta) G_{t-1} + \beta g_t \cdot g_t \\ \theta_t &= \theta_{t-1} - \frac{\alpha}{\sqrt{G_t + \epsilon}} g_t \end{aligned} \tag{5.26}$$

In equation 5.26, G_t is the running average of the square of gradients at step t, which is also called the second order gradient of the neural network, and g_t is the gradient at step

t. By introducing G_t, RMSProp will update parameters slower in directions with larger gradients and update parameters faster in directions with smaller gradients.

The RMSProp algorithm introduces the concept of the second-order gradient of the neural network. The algorithm that combines momentum and the second-order gradient together is called Adam [82]. Nowadays Adam along with the gradient descent algorithm are the most commonly used optimization algorithms. The math format of Adam is shown in:

$$\begin{aligned} m_t &= (1-\beta_1)m_{t-1} + \beta_1 g_t \\ G_t &= (1-\beta_2)G_{t-1} + \beta_2 g_t \cdot g_t \\ \hat{m_t} &= \frac{m_t}{1-\beta_1^t}, \quad \hat{G}_t = \frac{G_t}{1-\beta_2^t} \\ \theta_t &= \theta_{t-1} - \frac{\alpha}{\sqrt{\hat{G}_t + \epsilon}}\hat{m}_t \end{aligned} \tag{5.27}$$

In equation 5.27, $\hat{m_t}$ and $\hat{G}_t$ are corrected first-order momentum and second-order momentum. Their denominator is used to cancel bias in the first few updates.

The choice between using gradient descent with momentum or Adam as optimizer for the neural network is tricky in practice. Gradient descent with momentum with fine-tuned learning rate can achieve a better result than Adam, while Adam can let loss decrease faster. Based on this fact, an algorithm, Switching from Adam to SGD (SWTAS), combines these two algorithms together. The SWATS algorithm uses Adam in the first few epochs and then switches to the gradient descent with momentum. The switching point and learning rate of GD are determined by the SWTAS algorithm to get the best performance.

There are other optimization algorithms that can be used in neural network training, such as Adagrad, Nesterov Accelerated Gradient, AdaDelta, etc. Since in this thesis only gradient descent and Adam are used, other optimizers won't be introduced in this thesis.

Algorithm 4 Neural Network

Initialize weights of each layer in the neural network, normalize input features
repeat
  Forward propagation and calculate the loss on the training dataset.
  Calculate gradients of each layer's weights based on backward propagation.
  Update each layer's weights using optimization algorithms.
  Calculate loss on the validation dataset.
until Reach the maximum iteration T or loss stops decreasing.

As the number of layers of the neural network increases, gradient vanishing or gradient explosion may happen. From the forward propagation equation 5.20 and backward propaga-

tion 5.22, the output score or gradient will time other matrices and numbers when passing through one layer. Suppose the weight is a very large or very smaller number. In that case, the score and gradient will become infinite or zero after passing through several layers, which is called the gradient explosion or gradient vanishing. The float number used in modern computers ranges from 1.2×10^{-38} to 3.4×10^{38}. If a number is very large it will become 'Nan' and the neural network can not continue the training process. If the gradient becomes zero it will get stuck at a local point or the loss decrease will become very slow. To avoid this problem, the output of each layer needs to be controlled in a certain range, which introduces an algorithm called batch normalization [83]. Batch normalization will control inputs to each layer's activation function in a certain range, such as [0, 1], to make sure the gradient of activation function and output is in a reasonable range. Using the Sigmoid activation function as an example. From figure 5.3, the Sigmoid function is close to a linear function when x is in a range of [0, 1]. When $|x|$ becomes larger, the gradient of $\sigma(x)$ is close to zero. To make sure the gradient of $\sigma(x)$ does not vanish during the backward propagation, x needed to be controlled in the linear region of $\sigma(x)$, which means $|x| \in [0, 1]$. The batch normalization algorithm will normalize each layer's inputs to the activation function based on means and variances of inputs in the current batch, as shown in equation 5.28.

$$\hat{x} = \alpha \frac{x - \bar{x}}{\sigma + \epsilon} + \beta \tag{5.28}$$

In equation 5.28, $\hat{x}$ is input to the activation function after batch normalization. $\bar{x}$ and σ are mean and standard deviation of input feature x in current batch. α and β are two learnable parameters that will be updated together with the weight matrix w during the neural network training. They control the scale and central value of normalized $\hat{x}$.

Besides batch normalization, weight initialization of the neural network is also needed to be carefully chosen. The weights of a neural network can not be initialized as zero. The reason is that if all weights are set to zero all weights in the same layer will keep equaling each other due to the symmetric structure of the neural network. Usually, each node's weight is initialized randomly according to a normal or uniform distribution. However, the mean and variance of the normal and uniform distribution may affect the performance of the neural network. Based on the forward propagation function 5.20, each node's score is the sum of n_l number (n_l refers to the number of connections coming in at layer l, which can also be called as 'fan-in'), as shown in equation:

$$z_l^i = \sum_{k=1}^{n_l} w_{l,k}^i a_{l-1,k} \tag{5.29}$$

Assuming the standard derivation (std) of incoming features is one, each $w_{l,k}^i$ should have

the std of $\frac{1}{n_l}$ in order to keep z_l^i in the range [0, 1]. Similarly, from the backward propagation equation 5.22, each gradient is each node's score is the sum of n_{l+1} numbers. Each number in the weight matrix should have the std of $\frac{1}{n_{l+1}}$ in order to keep the gradient of layer l in the range [0, 1].

Based on the discussion above, the standard derivation of distribution that weight initialization used should be as what is shown in equation 5.30, and the mean value should be zero.

$$\sigma = \frac{2}{\sqrt{n_l + n_{l+1}}} \tag{5.30}$$

In other worlds, range of the uniform distribution is $[-\frac{\sqrt{6}}{\sqrt{n_l+n_{l+1}}}, \frac{\sqrt{6}}{\sqrt{n_l+n_{l+1}}}]$. This weight initialization is called the Xavier initialization [84].

5.2.3 Parameterized neural network

The neural network is an important multi-variate method to classify signals and backgrounds in particle physics. Normally in the application of neural networks in particle physics, the number of different kinds of signals and backgrounds are fixed so that the neural networks will have a fixed number of output classes and will be trained on data set with a fixed number of labels. In searches for new physics, searches for new particles, for example, the mass of the new particle is usually unknown. The search for the new particle will normally be performed in a continuous range of the particle's mass. Variables of signals at different mass points will have different distributions. In addition, limits at different mass points need to be set separately so that signals at different mass points can not be set using the same label during the neural network training.

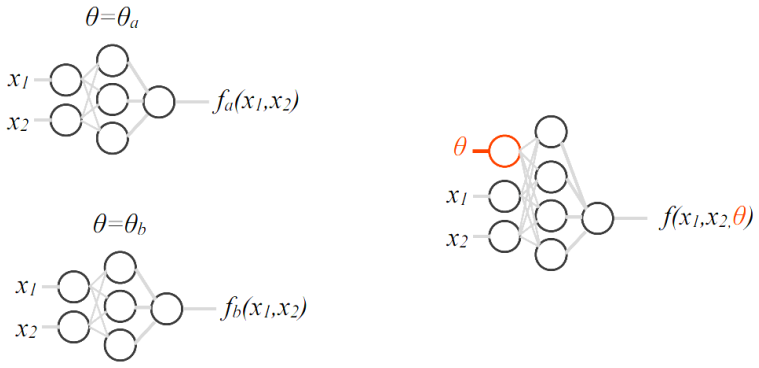


Figure 5.6: Parameterized neural network with a tagging variable θ.

One method to resolve this problem is training an individual neural network at each mass

point with all backgrounds and the signal at that mass point. However, the cost of training lots of neural networks will be too large if the training data sets are very large and the number of mass points is large. What's more, when the neural networks are applied to real data, we can only get limits at mass points in the neural network training. In another word, the neural networks can't learn how to classify signals from backgrounds at mass points not included in the training data sets. To resolve the problems discussed above, we could add a tagging variable θ as one of the training variables of the neural network, which is shown in figure 5.6.

The tagging variable is used to tag signals at different mass points so that the neural network could learn to classify signals from backgrounds at different mass points. For example, we could use the mass of the new particle directly as the tagging variable θ. Signals at different mass points will have different values of θ. However, the new particle doesn't exist in all backgrounds so we need to assign the tagging variable θ to backgrounds. Since θ is only used to identify different signals, we need to make sure that the neural network should separate signals and backgrounds based on the tagging variable only. In this case, the values of θ of backgrounds are set by random sampling based on the distribution of the signals. After that, distributions of θ of signals and backgrounds are identical. Since in backgrounds values of θ are randomly set, the correlations between θ and other variables are zero. The parameterized neural network is able to learn correlations between θ and other variables to separate signals at different mass points. For signals and backgrounds with the same value of θ, the parameterized neural network is able to separate them based on the distributions and correlations between other input variables. When applying the parameterized neural network to observed data, the value of θ can be set as the mass points that we would like to calculate limits at.

The idea behind the parameterized neural network is similar to the multi-task neural network, which has a shared bottom larger and multiple task towers following the shared bottom layer. Since the tagging variables θ of backgrounds are randomly sampled according to the distributions of signals, it is similar to splitting the background into different mass points and training individual neural networks at different mass points. The biggest advantage of using one network instead of individual networks is the parameterized neural network can achieve good performance at mass points not included in the training. In addition, the risk of over-fitting at each mass point is reduced since the parameterized neural network is optimized for all mass points. What's more, the performance of the parameterized neural network at one mass point can also benefit from other mass points.

Chapter 6

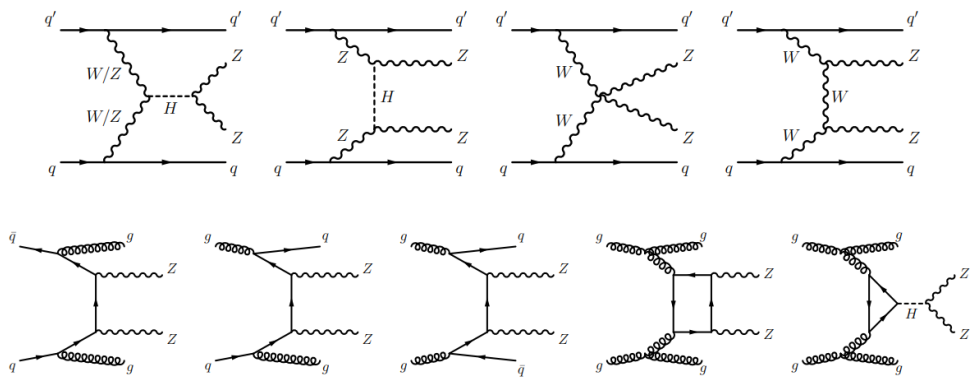
Observation of VBS ZZ Production

After the discovery of the Higgs boson, both the ATLAS and CMS have performed several searches and measurements for the vector boson scattering in different massive VV (W and Z boson) channels. The first observed vector boson scattering channel is the electroweak production of the same-sign WW with two jets. The CMS experiment also observed the same process with a significance of 5.5σ using partial run 2 data with the integrated luminosity of 35.9 fb^{-1} in 2017 [85], while the ATLAS experiment also observed the same-sign WW scattering at 6.5σ significance using partial run 2 data with the luminosity of 36.1 fb^{-1} in 2019 [86]. Later, the vector boson scattering of the WZ boson pair production was also observed. In 2018, the ATLAS experiment published the observation of the electroweak production of WZ boson pair in association with two jets with a significance of 5.3σ based on the partial Run 2 dataset with the integrated luminosity of 36.1 fb^{-1} [87]. CMS experiment observed the same process at 6.8σ based on the full Run 2 dataset with the integrated luminosity of 137 fb^{-1} [88].

Before this analysis, the electroweak production of $ZZjj$ hasn't been observed due to the small production cross-section and leptonic decay branching fraction. The CMS collaboration conducted a search for the electroweak $ZZjj$ production using partial Run 2 data with an integrated luminosity of 35.9 fb^{-1} and observed (expected) signal with a significance of 2.7 (1.6)σ in 2017, which was not enough to claim "evidence" [89]. This analysis takes advantage of full Run 2 data with an integrated luminosity of 139 fb^{-1} and multi-variate analysis method (gradient boosted decision tree) and observed this rare process with the signal significance of 5.7σ. The signal strength of the electroweak $ZZjj$ process is also measured in this analysis and the result is consistent with the Standard Model prediction. In this analysis, the inclusive $ZZjj$ production (EW and QCD) is measured in dedicated phase space. Another improvement in this analysis compared with the CMS previous study is that both final states with either four charged leptons and two jets ($lllljj$) or two charged

leptons, two neutrinos and two jets ($llννjj$) are included in the analysis, while only the four charged leptons plus two jets ($lllljj$) final state was included in the CMS study. In 2021, the CMS collaboration published another search for electroweak $ZZjj$ production in the $lllljj$ final state using full Run 2 data with an integrated luminosity of 137 fb^{-1} and observed (expected) the signal significance of 4.0 (3.5) σ, therefore, claimed "evidence" of electroweak $ZZjj$ production [90].

The typical VBS process at the LHC is initialed by quarks or gluons in the incoming proton beams. The initial state vector bosons are radiated from quarks and then scatter into another pair of vector bosons in the final state. The signature of VBS events will have two back-to-back jets with large angular separation. The two jets come from the two quarks that radiate vector bosons in the initial state. This feature has good discriminate power to separate VBS signals from other backgrounds. Thus the pure electroweak VV production associated with two jets is one of the most promising channels to study the VBS process. Some typical Feynman diagrams of the VBS (ZZ) process at LHC are shown in figure 6.1.



(a)

Figure 6.1: Typical diagrams for the production of ZZjj, including the relevant EW VBS diagrams (first row) and QCD diagrams (second row) [91].

In this chapter, details of studies of the observation of electroweak VBS ZZ production are presented. Since this thesis mainly focuses on the four lepton final states, only the studies of VBSZZ in the $llll$ channel will be introduced in detail. This chapter covers the object and event selection, background estimation, multivariate analysis, and final results of the four lepton channels.

6.1 Signal and background simulation

Signal processes of EWK production are modeled using the Powheg-Boxv2 event generator [92] with matrix elements (ME) calculated at next to leading order (NLO) in perturbative QCD (pQCD) and with the NNPDF3.0LO parton distribution functions (PDF). The contributions from triboson and VH processes in $llljj$ and $ll\nu\nu jj$ channels were estimated using the MadGraph5 _aMC@NLO 2.6.1 event generator [93] with ME calculated at leading order (LO) in pQCD with the NNPDF3.0LO PDF. Reweighting factors were calculated as a function of m_{jj} from the MadGraph5 _aMC@NLO events and applied to the POWHEG-V2 events. The effect is found to be below a few percent levels.

The QCD (non-gg) production of $4l$ and $2l2\nu$ + 2jets are modeled using Sherpa 2.2.2 [94] with the NNPDF3.0NNLO [95] PDF set, wherein the $ll\nu\nu jj$ final state both the decay from Z and W bosons are included in the simulation.

The gg induced ZZ + 2jets in the $4l$ channel is modeled using Sherpa 2.2.2 with the NNPDF3.0NNLO PDF (an additional 1.7 k-factor [96] is applied), and by gg2VV in $2l2\nu$ channel. Contribution from ggH Higgs is included as well.

The diboson productions from $WW \to lvqq$ and $WZ \to llqq$ are modeled using Powheg-Box with the CT10 PDF set [97] used in the ME calculations. The diboson productions from $WZ \to lllv$ are modeled with Sherpa 2.2.2 with the NNPDF3.0NNLO PDF. The other diboson channels are not included due to negligible contributions.

The triboson productions (VVV, V=W, Z) are modeled using Sherpa 2.2.2 with the NNPDF3.0NNLO PDF.

For the generation of $t\bar{t}$ events, Powheg-Box v2 [98–100] was used with the NNPDF3.0 PDF set in the ME calculation. Single-top-quark events produced in Wt final states were generated with Powheg-Box v2 with the CT10 PDF set [97] used in the ME calculations. Single top-quark production via s- or t-channels was generated by Powheg-Box v1 [98–100]. This generator uses the four-flavor scheme for the NLO QCD matrix element (ME) calculations together with the fixed four-flavour PDF set CT10f4 [97]. The $t\bar{t}V$ ($V = W, Z$) processes were generated at LO with MG5_aMC@NLO v2.2.2 [101] with the NNPDF2.3 PDF set.

The Z + jets process (with $Z/\gamma* \to ee/\mu\mu/\tau\tau$) was modelled using Sherpa 2.2.1 [94] with the NNPDF3.0NNLO PDF. The ME was calculated for up to two partons with next-to-leading-order (NLO) accuracy in QCD and up to four partons with leading-order (LO) accuracy using Comix [102] and OpenLoops [103].

The parton shower was modeled with Pythia8.186 [104] with the NNPDF2.3 [105] PDF set and the A14 set of tuned parameters [106] for all the samples except those from Sherpa,

where parton shower was done within SHERPA.

The detector response for simulated events was simulated within a framework [107] based on GEANT4 [108]. Furthermore, simulated events were processed with the same reconstruction software used for data. In order to account for the different particle reconstruction efficiencies measured in data and simulation, correction factors are derived in dedicated measurements and applied to simulated events.

In addition, several privately produced samples have been used to check the interference effect between EWK and QCD processes as summarized in section 6.4.5.

6.2 Object and event selections

This section describes the object and event selections in this analysis. The object definitions and reconstruction algorithms are introduced in chapter 4. Since the background rate is much higher in the $ll\nu\nu jj$ channel, we used tighter cuts on object selections compared to the $llljj$ channel.

6.2.1 Object selection

In this analysis, physics objects such as muons, electrons, jets, and missing transverse energy are used to reconstruct events and select control samples for background estimation in this analysis. Reconstruction algorithms of different objects have been introduced in chapter 4. Table 6.1 and table 7.5 summarize the selection cuts of muons and electrons in both the $llljj$ and $ll\nu\nu jj$ channels.

	$ZZ \to 4\ell$	$ZZ \to 2\ell 2\nu$
Identification	Loose	Combined with Medium ID
Kinematic cuts	$p_{\mathrm{T}} > 7$ GeV, $\lvert\eta\rvert < 2.7$	$p_{\mathrm{T}} > 20$ GeV, $\lvert\eta\rvert < 2.5$
Impact cuts	$\lvert d_0^{\mathrm{BL}}/d_{error}^{\mathrm{BL}}\rvert < 3.0$ $\lvert z_0^{\mathrm{BL}} \cdot \sin(\theta)\rvert < 0.5$ mm	$\lvert d_0^{\mathrm{BL}}/d_{error}^{\mathrm{BL}}\rvert < 3.0$ $\lvert z_0^{\mathrm{BL}} \cdot \sin(\theta)\rvert < 0.5$ mm
Isolation	FixedCutLoose	Loose

Table 6.1: Muon object definition.

Jets are reconstructed with the anti-k_T algorithm introduced in section 4.3.1 with a radius parameter of R = 0.4. To further reduce the effect of pile-up jets a cut on Jet Vertex Tagger

	$ZZ \to 4\ell$	$ZZ \to 2\ell 2\nu$
Identification	Likelihood Loose	Likelihood Medium ID
Kinematic cuts	$p_{\mathrm{T}} > 7$ GeV, $\lvert\eta\rvert < 2.47$	$p_{\mathrm{T}} > 20$ GeV, $\lvert\eta\rvert < 2.47$
Impact cuts	$\lvert d_0^{\mathrm{BL}}/d_{error}^{\mathrm{BL}}\rvert < 5.0$ $\lvert z_0^{\mathrm{BL}} \cdot \sin(\theta)\rvert < 0.5$ mm	$\lvert d_0^{\mathrm{BL}}/d_{error}^{\mathrm{BL}}\rvert < 5.0$ $\lvert z_0^{\mathrm{BL}} \cdot \sin(\theta)\rvert < 0.5$ mm
Isolation	FixedCutLoose	Loose

Table 6.2: Electron object definition.

(JVT) is applied on jets, which is introduced in section 4.3.3. The recommended 0.59 upper threshold on the JVT is used in the analyses to reject jets with $p_{\mathrm{T}} < 60$ GeV, $|\eta| < 2.4$, which corresponds to an efficiency of 92% and to an observed fake rate of 2%. The events are retained in the analyses only if all jets pass the Loose selection criteria for the Jet Cleaning, designed to provide an efficiency of selecting jets from proton-proton collisions above 99.5% for $p_{\mathrm{T}} > 20$ GeV.

In the ATLAS experiment, overlap removal is a procedure that deals with situations where multiple analysis objects are reconstructed from the same detector signals. After applying overlap removal, all but one such object are ignored[1]. In the standard overlap removal strategy, jets are preferred compared to leptons, which means if an object is identified as both jet and lepton the overlap removal will ignore the reconstructed lepton and the object will be counted as a jet. The overlap removal strategies used in this analysis are summarized in Table 6.3 and Table 6.4, for $llljj$ and $ll\nu\nu jj$ channels, respectively. The $ll\nu\nu jj$ channel follows the standard overlap removal procedure. In $llljj$ channel the leptons are given higher priority than jets, to keep leptons in case of overlapped with jets, to keep a high signal efficiency. With this lepton-favored overlap removal, the EWK signal increases 19 % with background only increasing 14 %.

6.2.2 Event selections in the $llll$ channel

Event selections in the $llll$ channel are shown in Table 6.5. The lepton and ZZ selections basically follow the on-shell ZZ cross-section measurements, and jet selections have been optimized to enhance the EWK VBS contributions in the forward region. Relatively loose m_{jj} and Δy_{jj} cuts have been applied to keep more events for further MVA studies.

[1] https://particle.wiki/wiki/Overlap_removal_(ATLAS)

	Reference objects	Criteria
Remove electrons	electrons	Share a track or have an overlapping calorimeter cluster. Keep higher p_{T} electron
Remove muons	electrons	Share track and muon is calo-tagged
Remove electrons	muons	Share track
Remove jets	electrons	$\Delta R_{e-jet} < 0.2$
	muons	$\Delta R_{\mu-jet} < 0.2$ OR muon track is ghost-associated to jet **AND** ($N_{Trk}(jet) < 3$ OR ($p_T^{jet}/p_T^{\mu} < 2$ and $p_T^{\mu}/\Sigma_{TrkPt} > 0.7$))

Table 6.3: Overlap removal criteria between pre-selection objects for the $lllljj$ channel. The overlap removal follows the order shown in this table. Once an object has been marked as removed, it does not participate in the subsequent stages of the overlap removal procedure.

	Reference objects	Criteria
Remove electrons	electrons	Share a track or have an overlapping calorimeter cluster. Keep higher p_{T} electron
Remove muons	electrons	Share track and muon is calo-tagged
Remove electrons	muons	Share track
Remove jets	electrons	$\Delta R_{e-jet} < 0.2$
	muons	$\Delta R_{\mu-jet} < 0.2$ OR muon track is ghost-associated to jet **AND** ($N_{Trk}(jet) < 3$ OR ($p_T^{jet}/p_T^{\mu} < 2$ and $p_T^{\mu}/\Sigma_{TrkPt} > 0.7$))
Remove electrons	jets	$\Delta R_{e-jet} < 0.4$
Remove muons	jets	$\Delta R_{\mu-jet} < 0.4$

Table 6.4: Overlap removal criteria between pre-selection objects for the $ll\nu\nu jj$ channel. The overlap removal follows the order shown in this table. Once an object has been marked as removed, it does not participate in the subsequent stages of the overlap removal procedure.

Category	Cut Name	Requirement
Event	Trigger	details in the trigger section
Pres-election	Vertex	At least one vertex reconstructed with 2 or more tracks
Electrons	p_{T}	> 7GeV
	η	$\lvert\eta\rvert < 2.47$
	ID	`VeryLooseLH` working point
	Object Quality	Not from a bad cluster
	$z_0 \sin\theta$	< 0.5mm
Muons	p_T	> 7GeV (15GeV if `CaloTagged`)
	η	$\lvert\eta\rvert < 2.7$
	ID	Loose
	$z_0 \sin\theta$	< 0.5mm if muon is not `StandAlone`
	$\lvert d_0\rvert$	< 1mm if muon is not `StandAlone`
Jets	Clustering	`AntiKt4EMTopo`
	p_{T}	> 30GeV (40GeV if $2.4 < \lvert\eta\rvert < 4.5$)
	Jet vtx. tag cut	JVT < 0.59
Overlap Removal		Lepton Favouring Working Point `HSG2` without ClusterMatch
Quadruplet Selection	p_{T}	$p^1_{\mathrm{T}} > 20$GeV, $p^2_{\mathrm{T}} > 20$GeV, $p^3_{\mathrm{T}} > 10$GeV
	Electron Quality	No more than two electrons fail `LooseLH` identification
	Muon Quality	Number of `StandAlone` or `CaloTagged` muons < 2
Quadruplet Ranking	Minimal Δm_Z	Select quadruplet with smallest $\lvert m_{12} - m_Z\rvert + \lvert m_{34} - m_Z\rvert$
Event Selection	Quarkonia Veto	$m_{12,34,14,23} > 10$GeV
	ΔR	> 0.2 between leptons in quadruplet
	Isolation	All leptons in quadruplet pass `FixedCutLoose` working point
	Electron ID	All e in quadruplet pass `LooseLH` working point
	Impact Parameter	$\lvert d_0/\sigma(d_0)\rvert < 5$ (3) for e (μ) in quadruplet
	Z Window	$66 < m_{12,34} < 116$GeV
Dijet Selection	Different detector sides	$y_{j_1} \times y_{j_2} < 0$
	Dijet Ranking	Select dijet with the highest scalar p_T sum
	Pseudo-rapidity separation	$\Delta y_{jj} > 2$
	Dijet mass	$m_{jj} > 300$

Table 6.5: Four lepton signal region definition. The object and event selection criteria were applied in the four-lepton channel of the analysis.

6.2.3 Event selections in the $llvv$ channel

The data used in the analyses is triggered with single lepton triggers (electron or muon). To satisfy the trigger threshold of these single lepton triggers, the p_{T} of two leading leptons should be 30 GeV and 20 GeV, respectively.

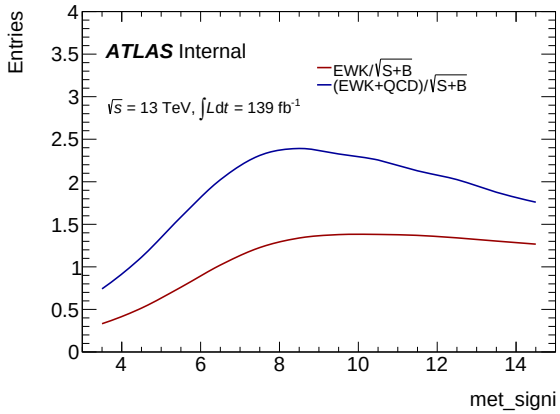


Figure 6.2: $S/\sqrt{(S+B)}$ as a function of E_T^{miss} significance.

Events are required to contain exactly two same flavor and opposite charged muons or electrons that pass the object selections described in Table 6.1 and 7.5, and furthermore, to reduce the $ZZ \to 4\ell$ and $WZ \to \ell\nu\ell\ell$ events, events with additional leptons are vetoed. The p_{T} thresholds and selection criteria for the additional leptons are set to be 7 GeV and "LooseLLH" selection criteria, for both muons and electrons. The invariant mass of the selected two leptons is required to be within the range of $80 < m(\ell^+\ell^-) < 100$ GeV. This requirement can significantly reduce the events that don't include the Z-boson, for example, $t\bar{t}$, $WW \to \ell\nu\ell\nu$, $Z \to \tau\tau$, etc.

Z jets is highly suppressed with a large E_T^{miss} significance cut: $E_T^{\mathrm{miss}} significance > 12$. $S/\sqrt{S+B}$ as a function of E_T^{miss} significance is shown in Figure 6.2.

At least two jets within $|\eta| < 4.5$ is pre-selected, with b-jets vetoed with an 85% working point (for $llvvjj$ channel only) to reduce the top background. We also impose the selection $p_{\mathrm{T}} > 60$ GeV for the leading jet and $p_{\mathrm{T}} > 40$ GeV for the second-leading jet. The large p_{T} threshold of jets is helpful to remove a large amount of contamination from Zjets and QCD ZZ+2jets which have relatively lower jets p_{T}.

To further improve the EWK ZZ+2jets sensitivity, typical back-to-back topology is used for the two p_{T} leading jets: $Y_{j1} \cdot Y_{j2} < 0$ and $\Delta y_{jj} > 2$, where Y_j stands for jet rapidity. Due to the large angle between these two jets, the invariant mass of them should be quite large: $m_{jj} > 400$ GeV.

The selections are summarized in Table 6.6.

Selections	Cut Value
veto events with additional leptons	("LooseLLH" quality, p_{T} >7 GeV) for the additional leptons
lepton p_{T} for (leading, sub-leading)	>(30 GeV, 20 GeV)
$m(\ell^+\ell^-)$	$80 < m(\ell^+\ell^-) < 100$ GeV
$E_{\mathrm{T}}^{\mathrm{miss}}$ significance	>12
Baseline Jet selection	$N_{\mathrm{jets}} >= 2$, No b-jets with 85% WP
Jet p_{T} for (leading, sub-leading)	>(60 GeV, 40 GeV)
Two leading p_{T} Jets	$Y_{j1} \cdot Y_{j2} < 0$, $\Delta Y(j1, j2) > 2$, $m_{jj} > 400$ GeV

Table 6.6: Summary of the event selection criteria for $ZZ \to 2\ell 2\nu$ measurement. The other object-level selections (lepton impact parameter) are the same as $lllljj$ channel, shown in Table 6.5.

6.2.4 Event yields in signal regions

After applying all selections in signal regions, the event yields in both channels are summarized in table 6.7. All the minor backgrounds are summed together as 'Others', and the $WWjj$ and $t\bar{t}$ processes are referred to as the non-resonant-ll backgrounds. Uncertainties in the predictions include both the statistical and systematic components of the predicted yields before fit.

Process	$lllljj$	$ll\nu\nu jj$
EW $qq \to ZZjj$	31.4 ± 3.5	15.0 ± 0.8
QCD $qq \to ZZjj$	77 ± 25	17.2 ± 3.5
QCD $gg \to ZZjj$	13.1 ± 4.4	3.5 ± 1.1
Non-resonant-ll	-	21.4±4.8
WZ	-	24.6 ± 1.1
Others	3.2 ± 2.1	1.2 ± 0.9
Total	124 ± 26	82.9 ± 6.4
Data	127	82

Table 6.7: Observed data and expected event yields in 139 fb^{-1} of data in the $lllljj$ and $ll\nu\nu jj$ signal regions.

6.3 Background estimation

6.3.1 Background estimation in the $4l$ channel

This section summarizes the background estimation for $4l$ channel.

For the $4l$ channel, there are very small contributions from non-prompt backgrounds, fake leptons from Z + jets, and top processes. Those contributions are estimated with fake

factor method. When looking at the EWK processes alone, the QCD component (from both qq and gg induced) becomes the major background, and a QCD-enriched control region is defined to constrain the contribution, by reverting either the m_{jj} or Δy_{jj} cuts, as $m_{jj} <$ 300 GeV or $\Delta y_{jj} < 2$.

The triboson background has a minor contribution and is estimated directly from simulation.

6.3.2 QCD background for 4l channel

The four lepton plus two jets contributions from qq initiated QCD processes are the dominant backgrounds when looking for the EW production mode. These contributions are constrained by a QCD-enriched control region, by reverting either the m_{jj} or Δy_{jj} cuts, as $m_{jj} < 300$ GeV or $\Delta y_{jj} < 2$. This control region is included in the final fit to constrain the QCD yield in the signal region, as well as the theoretical uncertainties from QCD processes, which are the dominant one in $4l$ channel 6.4.3, apart from statistic uncertainty. The shapes are taken from MC and the normalization factor is used as a floating parameter in the final fit. The plots are shown in section 6.5 for a side-by-side comparison of distributions in SR and QCD CR.

6.3.3 Fake background for 4l channel

The backgrounds from misidentified leptons are usually difficult to model in MC thus a data-driven (fake factor) method is used, to estimate the contribution in the signal region. The fake factor methods are developed as

- Define the dedicated background dominant region to derive the fake factor for the dedicated background. The fake factor is defined as $F = N_G/N_P$, by counting the number of additional leptons in those dedicated control regions. N_G refers to the number of good leptons (passing SR selections listed in Table 6.5) and N_P refers to the number of poor leptons (passing SR selections with certain cuts reverted)

- Define a four-lepton plus two jets fake control region, where one or two leptons pass the poor lepton definition, and the other leptons still pass SR selections.

- The poor electrons are defined as fail isolation or fail the Likelihood Loose electron ID (but still should pass the VeryLooseLH). The other requirements are the same as the good electrons.

- The poor muons are defined as fail isolation or fail the d0 significance cut (but still should pass |d0significance| < 10. The other requirements are the same as the good muons.

- Calculate the final estimation in SR, following the equation

$$N^{\text{fake}} = \left(N_{gggp} - N^{ZZ}_{gggp}\right) \times F - \left(N_{ggpp} - N^{ZZ}_{ggpp}\right) \times F^2 \,, \tag{6.1}$$

  where the contribution from ZZ is subtracted from the calculation. The second term is due to the double counting of N_{gggp} and N_{ggpp}.

The source of misidentified leptons (from now on called fake leptons) are mostly leptons in jets of leptonic decaying hadrons or other objects misidentified as leptons. In the $4l$ channel, the major contributions come from Z+jets and $t\bar{t}$. Thus dedicated Z+jets and $t\bar{t}$ dominant regions have been defined to derive the fake factor from each process. There are still smaller contributions from W+jets, W^+W^-+jets, and WZ, those are included in the fake estimation, but no dedicated fake factors have been derived.

Fake factor calculation for Z+jets

The fake factor for Z+jets is calculated from Z+jets dominant region from data, selected with one same-flavor opposite-charge dilepton around Z mass, and additional two jets. The detailed selections are summarized in Table 6.8.

The fake factor from Z+jets are calculated vs. different p_T and η, and the contribution from non-Z+jets (ZZ, WZ and $t\bar{t}$) has been subtracted. The calculated fake factors are shown in Figure 6.3 and Figure 6.4, for electrons and muons, respectively. The fake factors calculated from Z+jets MC are also shown for comparison.

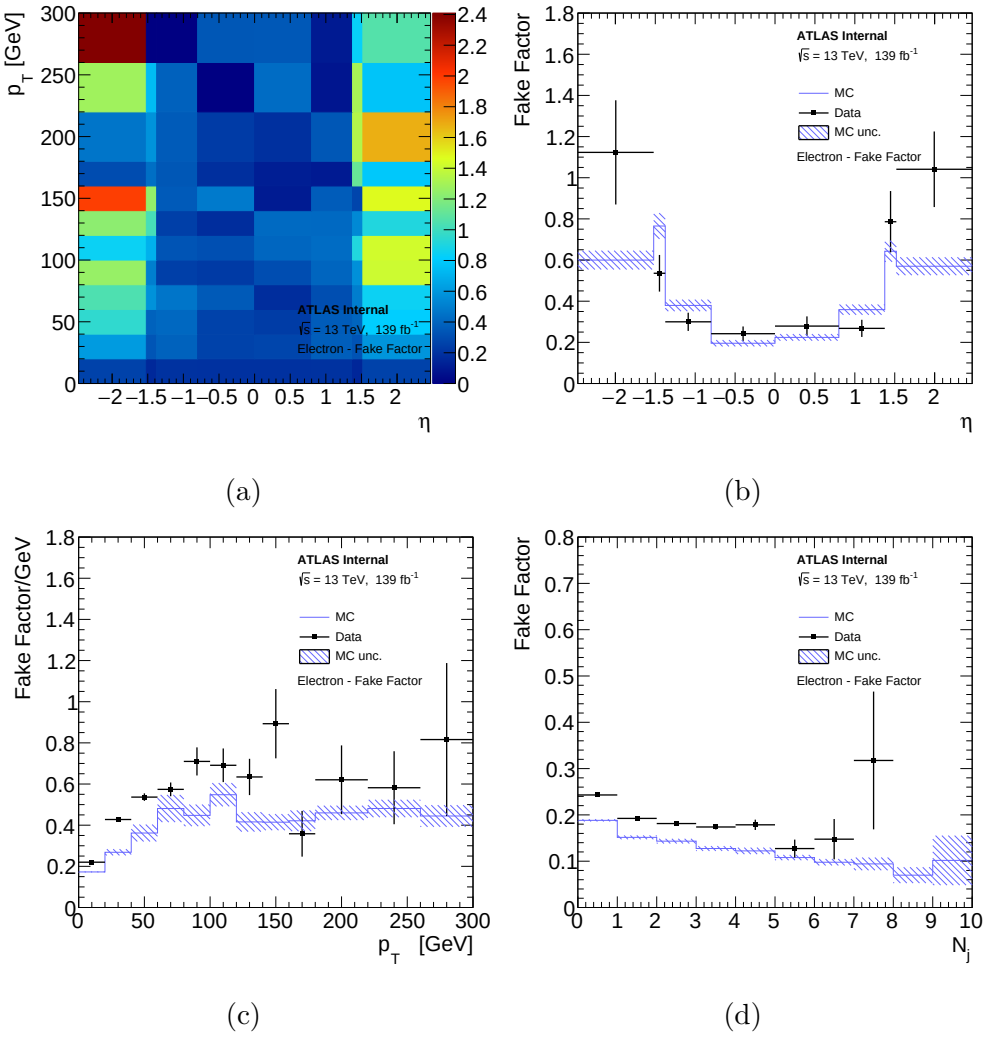


Figure 6.3: The Zjets fake factor constructed using the additional electrons. Shown as a function of η and p_{T} together in (a), and separately in (b) and (c), and as a function of the number of jets in (d).

Category	Cut Name	Requirement
		... (Object Selection)
Dilepton Selection	Pairing	Form same flavor opposite charge dileptons
	Isolation	Leptons in dilepton pass `FixedCutLoose` working point
	Impact Parameter	$\lvert d_0/\sigma(d_0)\rvert < 5$ (3) for e (μ) in dilepton
	Electron ID	Electrons in dilepton pass `LooseLH` working point
	p_{T}	> 25 GeV
Dilepton Ranking	Minimal Δm_Z	Select dilepton with smallest $\lvert m_{\ell\ell} - m_Z\rvert$
Event Selection	$E_{\mathrm{T}}^{\mathrm{miss}}$	< 25GeV
Additional Lepton Selection	Good Lepton	Definition is the same as in Table 6.5
	Poor Lepton	Definition is the same as in list 6.3.3
Overlap Removal		Remove jets with $\Delta R_{\ell\text{-}j} < 0.4$ to Poor leptons in the selected quadruplet

Table 6.8: Fake factor region definition for Z + jets. Since the region is supposed to have the same Good and Poor lepton definition as the signal region the object selection criteria is inherited from Tab 6.5 and then follows the Z tagging boson selection and evaluation of the additional lepton.

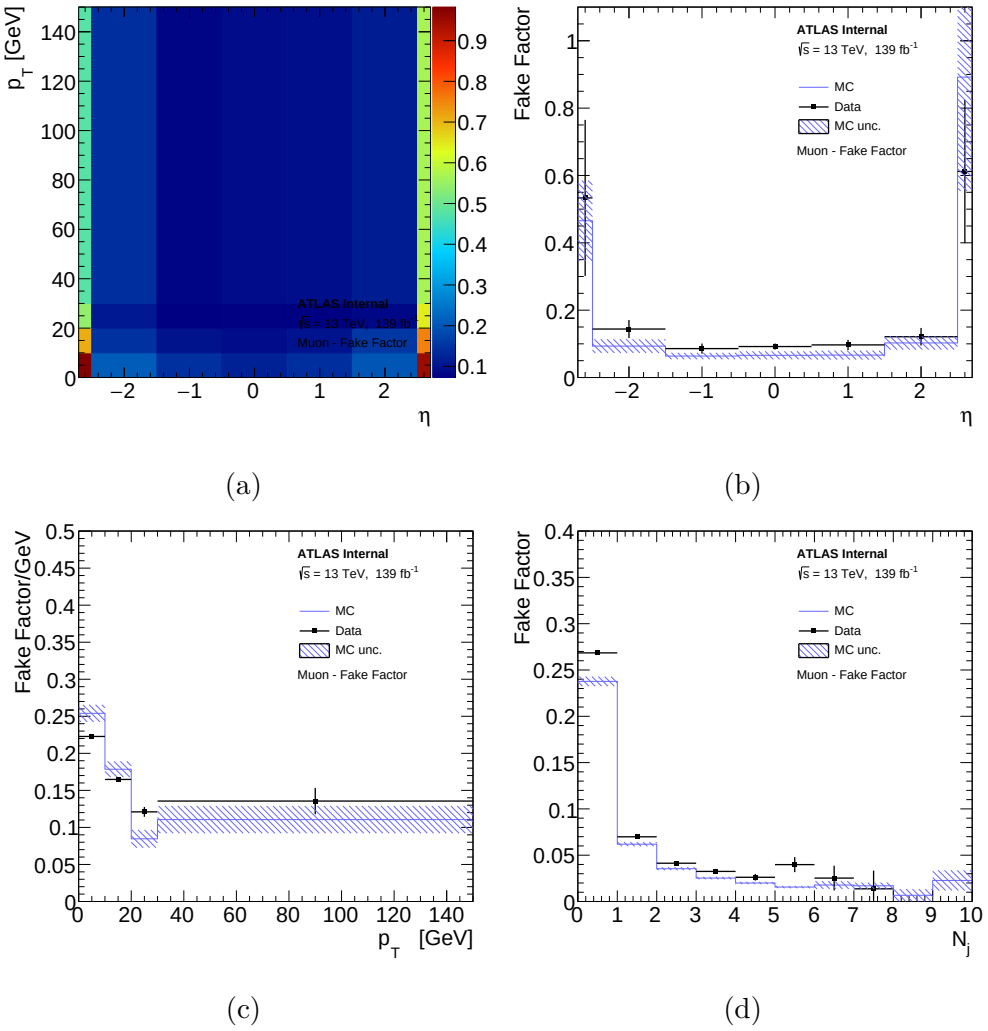


Figure 6.4: The Zjets fake factor constructed using the additional muons. Shown as a function of η and p_{T} together in (a), and separately in (b) and (c), and as a function of the number of jets in (d).

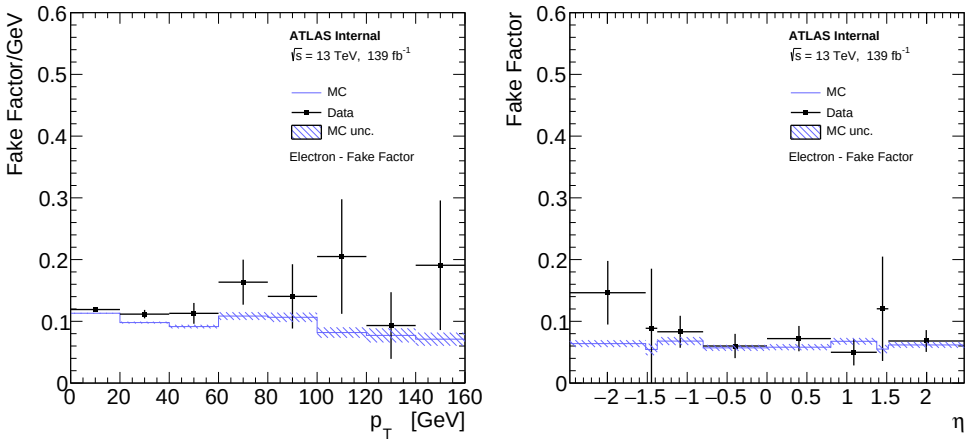
Fake factor calculation for $t\bar{t}$

The fake factors for $t\bar{t}$ are calculated from a $t\bar{t}$ dominant region, selected with one $e\mu$ lepton pair plus two additional jets. In the events with 3 leptons, the requirement on m_T^W is applied to reduce the contribution from $t\bar{t}$+W. The m_T^W is defined as

$$m_T^W = \sqrt{2P_T^{\ell 3} E_T^{\text{miss}} \left[1 - \cos\left(\Delta\phi\left(P_T^{\ell 3}, E_T^{\text{miss}}\right)\right)\right]} \tag{6.2}$$

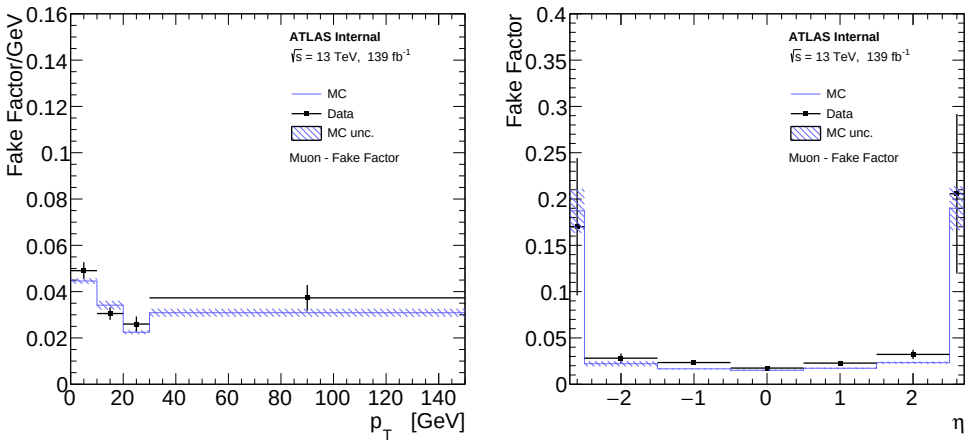
Also, at least one b-tagged jet is required to enhance the top contribution. The detailed selections are summarized in Table 6.9.

The fake factor from $t\bar{t}$ are calculated vs. different p_T and η, and the contribution from non-$t\bar{t}$ (ZZ, WZ and Z+jets) has been subtracted. The calculated fake factors are shown in Figure 6.5 and Figure 6.6, for electrons and muons, respectively. The fake factors calculated from $t\bar{t}$ MC are also shown for comparison. The 2D plots are shown in Figure 6.7.



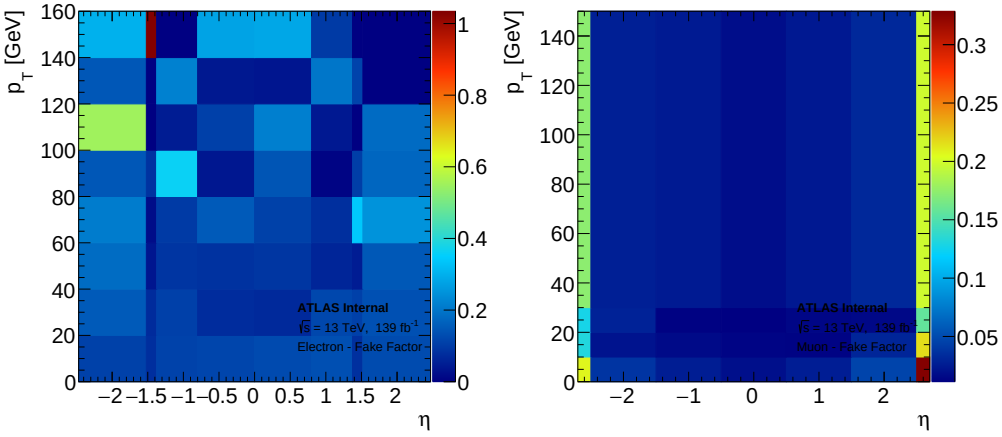
(a) (b)

Figure 6.5: The $t\bar{t}$ fake factor for electrons vs p_T(a) and η (b)distribution.



(a) (b)

Figure 6.6: The $t\bar{t}$ fake factor muons vs p_T(a) and η (b)distribution.

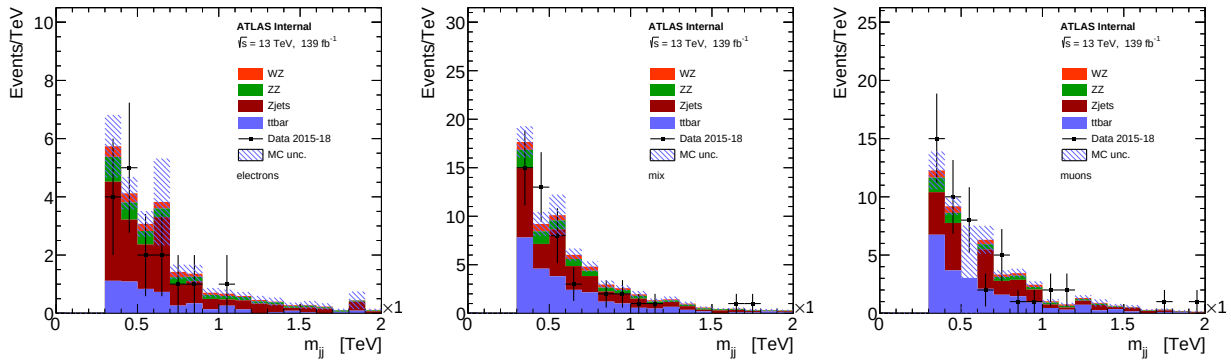


(a) (b)

Figure 6.7: The $t\bar{t}$ fake factor for electrons (a) and muons (b) 2-D distribution vs. η and p_T.

Combination of fake factors

Fake factors calculated from Z+jets and $t\bar{t}$ dedicated control regions (Table 6.8, Table 6.9) are combined accordingly to their contributions in the four-lepton plus two jets fake control region in the different channels, as defined in list 6.3.3. The individual ratios (Z+jets over $t\bar{t}$) are listed for electron 2.59, mixed 0.95, and muon 0.74 channel. The relation can be seen in the dijet mass distribution in Fig 6.8 from which the ratios were extracted. The raw event yields in the four-lepton fake control region are summarized in Table 6.10.



(a) (b) (c)

Figure 6.8: Dijet mass distribution in the fake control region split into electron (a), mixed (b), and muon (c) channel.

Systematics of the fake estimation

The usual way of estimating systematics for fake factor methods includes the variations of poor lepton definition, MC subtraction, the binning effect of fake factors, etc. In addition to those, in this specific $lllljj$ plus 2jets region, due to the very limited statistics even in the fake control region, a very conservative approach is used here, by including additionally the difference from fake factors calculated from data and MC. In detail, the following sources of systematics have been included for the fake estimation

- Variation of isolation cut for the poor lepton definition by a factor of two up and down
- Variation of the yield of other MC while subtraction by 30 % up and down, in the fake control regions
- Use fake factors derived directly from Z+jets and $t\bar{t}$ MCs
- Use one bin fake factor instead of p_T and η dependent ones
- The statistical uncertainties on fake factor and four-lepton fake control region

Estimation of the fake contribution in signal region

The final estimations from fake contribution in the signal region are summarized in Table 6.11, along with the systematic variations.

Closure test of the fake-factor method

A closure test of the fake-factor method is performed, by applying MC-derived fake factors, to the MC 4 lepton control region. This is done separately for Z+jets and $t\bar{t}$ samples and numbers are compared to the direct MC predictions to test biases due to the fake-factor method. The results are summarized in Table 6.12. The closure test is limited by MC statistics in the $lllljj$ channel, where 0 events are left after full event selections. Thus loose selections have been applied, only requiring ZZ plus 2jets (no dijet mass or η selections) but MC statistical uncertainty still dominates. Due to the above issues, the closure test should only serve as an additional check for the fake factor method, in this $lllljj$ plus 2 jets region. The numbers can not be compared to the data estimations due to different selections (No MC events to test with SR selections). The closure test is performed using data taken by ATLAS from 2015 to 2017 with an integrated luminosity of 80 fb^{-1}.

Category	Cut Name	Requirement
		... (Object Selection)
Dilepton Selection	Pairing	e-μ dileptons
	Isolation	Leptons in dilepton pass `FixedCutLoose` working point
	Impact Parameter	$\lvert d_0/\sigma(d_0)\rvert < 5$ (3) for e (μ) in dilepton
	Electron ID	Electrons in dilepton pass `LooseLH` working point
	p_{T}	> 25 GeV
Dilepton Ranking	Impact parameter	Select dilepton with smallest $\lvert d_0/\sigma(d_0)$
Event Selection	$E_{\mathrm{T}}^{\mathrm{miss}}$	> 50GeV
	W mass	$m_T^W < 60$GeV for 3 lepton events
	p_{T}	> 28GeV (at least one tagging lepton)
		Remove event with 4 Good leptons
		At least two jets, at least one b-tagged jet
Additional Lepton Selection	Good/Poor Lepton	Definition is the same as in Table 6.5
Overlap Removal		Remove jets with $\Delta R_{\ell\text{-}j} < 0.4$ to Poor leptons in the selected quadruplet

Table 6.9: Fake factor region definition for $t\bar{t}$. Since the region is supposed to have the same Good and Poor lepton definition as the signal region the object selection criteria is inherited from Tab 6.5 then follows the $t\bar{t}$ tagging leptons selection and evaluation of the additional lepton.

channel	electron	mix(fake e)	mix(fake μ)	mix(fake $e+\mu$)	muon
N_{gggp}	8.0	7.0	10.0	/	14.0
N_{gggp}^{ZZ}	3.3	3.7	3.7	/	4.7
N_{ggpp}	8.0	7.0	17.0	9.0	33.0
N_{ggpp}^{ZZ}	0.3	0.1	0.1	0.3	0.3

Table 6.10: Fake control region yields in four lepton channels.

channel	electron	mix	muon	inclusive
Nominal estimate	0.678± 0.652	1.023± 0.740	0.566± 0.240	2.268± 1.015
F stat. uncertainty varied down	0.698± 0.622	0.872± 0.652	0.509± 0.214	2.079± 0.926
F stat. uncertainty varied up	0.657± 0.685	1.173± 0.840	0.622± 0.267	2.452± 1.116
One bin F	0.653± 0.590	0.594± 0.558	0.646± 0.313	1.892± 0.870
MC F	0.534± 0.471	1.415± 0.993	0.439± 0.184	2.389± 1.114
Isolation varied down	0.938± 0.686	0.552± 0.466	0.215± 0.107	1.704± 0.837
Isolation varied up	0.723± 0.646	1.104± 0.739	0.559± 0.237	2.386± 1.010
MC corr. varied down	0.697± 0.695	1.048± 0.811	0.832± 0.385	2.577± 1.136
MC corr. varied up	0.660± 0.614	0.984± 0.687	0.316± 0.159	1.961± 0.935
Variations added in quadrature	0.304± 0.702	0.783± 0.803	0.535± 0.289	1.622± 1.106

Table 6.11: Fake background estimations in the SR. For the nominal value the 2D fake factor together with the Z+jets and $t\bar{t}$ combination is applied. The other lines show the estimations with different uncertainty variations. The differences between each variation and nominal value are summed quadratically for the final systematic estimation. The ± uncertainty numbers shown in the table are statistical uncertainties on each number.

Process	MC prediction	Closure test
Zjet	-0.88 ± 1.11	3.73 ± 0.46
Top	0.28 ± 0.28	0.172 ± 0.04

Table 6.12: MC closure test comparison for fake-factor method. Numbers are normalized to 80 fb^{-1}. Note a loose ZZ plus 2jets selections have been applied to gain more MC events.

6.4 Systematic uncertainties

This section summarizes the theoretical and experimental systematic uncertainties in this analysis, apart from the ones from data-driven-based background estimation, which is summarized in the background section. This analysis is still statistically dominant. Apart from that, the dominant systematics in the $4l$ channel comes from the theoretical modeling of QCD processes. The dominant systematics in the $2l2\nu$ channel comes from the data-driven background estimation, which is limited by either the theoretical modeling of WZ processes (for WZ background) or the data statistics in $e\mu$CR (for non-resonant background).

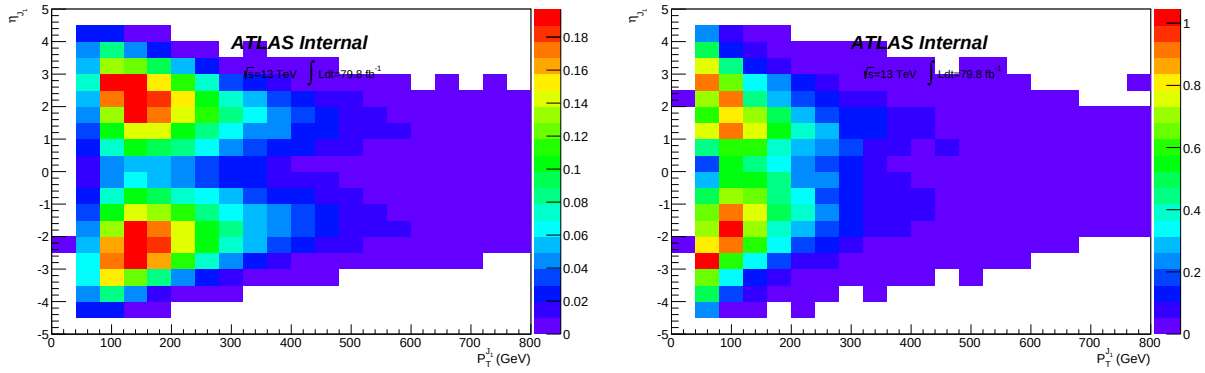
6.4.1 Experimental uncertainties

The experimental uncertainties in this section include the systematics from lepton/jets/MET energy scale and smearing and efficiency, trigger efficiency, and pileup. This is included for all the MC-based backgrounds and also data-driven ones depending on corresponding methods.

Experimental uncertainties in the $lllljj$ channel are studied following CP recommendation uncertainties. The experimental uncertainties are estimated for the EWK ZZjj, QCD qq$\rightarrow 4l$, and gg$\rightarrow$ZZ$\rightarrow 4l$ processes.

For EWK and QCD processes, large differences have been observed from jet uncertainties. QCD process has a larger (10%) jet uncertainty than the EWK process (2%). This has been checked and found to be due to that in the QCD process there are more jets in the low p_T region. Comparison of jet p_{T} and η distributions are shown in Figure 6.9.

JVT selection is only applied for jets with $20 < p_T < 60$ GeV and $|\eta| < 2.4$. For both QCD and EWK ZZ processes most jets' p_T are higher than 60 GeV so the uncertainty of the JVT scale factor is found to be very small (less than 1%). Since the QCD process has more soft jets, the QCD process has a larger JVT systematic uncertainty.



(a) (b)

Figure 6.9: Jet p_T η distribution in EWK (left) and QCD (right) process

6.4.2 Jet pileup uncertainty in 4l channel

In the final fit, we also include systematic uncertainty due to different pileup jets. First, we trained BDT using the whole signal sample (EWK ZZ→ $4l$) and QCD qqZZ sample. After that, we divided our signal sample (EWK ZZ→ $4l$) and QCD qqZZ sample into low μ and high μ subsets and compare their difference in gradient BDT response (both low μ and high μ subsets are normalized to 139 fb^{-1}). The average μ value in the signal sample is 34.5 and the average μ value in the QCD sample is 33 so we chose 33 as the low/high μ boundary. Difference performance of low μ and high μ QCD qqZZ events is shown in figure 6.10

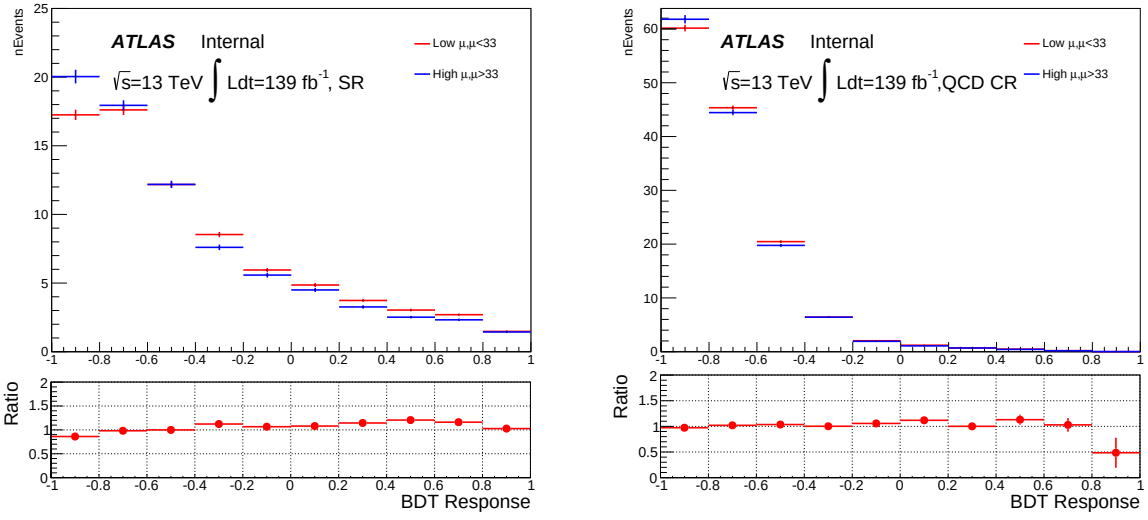


Figure 6.10: Gradient BDT response of low μ and high μ QCD qqZZ events in the signal region (left) and QCD control region (right)

6.4.3 Theoretical uncertainties

The theoretical uncertainties are estimated for the EW and QCD processes, including contributions from the different QCD scales and α_S, different parton density functions (PDF) choices, and different parton showers. The dominant one comes from different QCD scale choices.

The PDF uncertainty is estimated by comparing events in different PDF sets, as well as the uncertainty from the nominal PDF set itself. Envelop is chosen following the PDF4LHC [109] recommendations. The α_S uncertainty is estimated by comparing events generated with different α_S values with the nominal PDF set.

The QCD uncertainty is estimated by comparing events with different renormalization scales (μ_R) and factorization scale (μ_F) settings, where the largest deviation is chosen as the systematics.

The parton shower uncertainty is estimated by comparing events with different parton shower settings for the EW process in the $4l$ channel, with private samples. Due to the difficulty in producing an equivalently large number of QCD events, the same number has been used for the QCD process, and $2l2\nu$ EW and QCD processes.

Technically, the PDF, α_S, and QCD systematics are estimated via the reweighting procedure, following the PMG recommendations. The parton shower uncertainty is estimated with different samples. Details about estimation methods are summarized in Table 6.13. The results of EW and QCD qqZZ processes are shown in Table 6.15 for the FV cross-section. The results of the QCD $gg \to ZZ$ process as a function of the gradient BDT discriminant which was used in the final fitting are shown in figure 6.11.

Process	Nominal PDF	Alternative PDFs used	α_S setting	QCD scale ([μ_R, μ_F])	Parton shower used
EW $llljj$	NNPDF30lo	CT14lo		[0.5,0.5], [0.5,1], [1,0.5]	Pythia8
			0.118	[1,1], [1,2], [2,1]	Herwig7
				[2,2]	
QCD $llljj$	NNPDF30nnlo	MMHT2014nnlo68cl	0.117	[0.5,0.5], [0.5,1], [1,0.5]	
		CT14nnlo	0.118	[1,1], [1,2], [2,1]	
			0.119	[2,2]	
EW $ll\nu\nu jj$	NNPDF30lo	CT14lo		[0.5,0.5], [0.5,1], [1,0.5]	Pythia8
			0.118	[1,1], [1,2], [2,1]	
				[2,2]	
QCD $ll\nu\nu jj$	NNPDF30nnlo	MMHT2014nnlo68cl	0.117	[0.5,0.5], [0.5,1], [1,0.5]	
		CT14nnlo	0.118	[1,1], [1,2], [2,1]	
			0.119	[2,2]	

Table 6.13: Summary table of how the different theoretical uncertainties are estimated.

Process	PDF (%)	α_S (%)	QCD scale (%)	Parton shower (%)
EW $lllljj$	+5.9 -5.9		+6.1 -5.6	+3.3 -3.3
qqQCD $lllljj$	+2.0 -1.0	+2.6 -2.6	+34.2 -22.8	
EW $ll\nu\nu jj$	±2.0		±2.0	
qqQCD $ll\nu\nu jj$	±1.3	±2.6	±19.6	

Table 6.14: Summary table of theoretical uncertainties for the fiducial volume cross-section for both channels for the EW and qqZZ processes. The parton shower uncertainty is estimated with the EW sample in $lllljj$ channel, and being used for the other processes listed in this table.

Process	PDF (%)	α_S (%)	QCD scale (%)	Parton shower (%)
EW $lllljj$	+6.1 -6.1		+0.8 -1.1	+10.1 -10.1
qqQCD $lllljj$	+2.0 -1.0	+2.6 -2.6	+31.5 -22.0	

Table 6.15: Summary table of theoretical uncertainties for the QCD control region cross-section for the $4l$ channel for the EW and qqZZ processes. The parton shower uncertainty is estimated with the EW sample in the $4l$ channel and is used for the other processes listed in this table.

6.4.4 Modeling uncertainties

For the QCD qqZZ background, which is the largest background in the $lllljj$ channel, the nominal Monte Carlo sample that is used in this analysis is modeled by the Sherpa generator. And the normalization uncertainties as mentioned above are also estimated by Sherpa samples. In addition, a shape systematic has been evaluated to take into account the uncertainties in the modeling of the qqZZ background by using the alternative generator. The samples in the $lllljj$ channel are generated and showered by MG and PythiaEight. Then this modeling uncertainty is calculated from the envelope of shape difference between Sherpa theoretical components and the difference between Sherpa and MG on the BDTG discriminant used in the analysis. Figure 6.12 and 6.13 show the modeling systematics in 4ℓ channel.

For the $2l2\nu$ channel, the analysis uses the $2e2\mu$ samples to mimic the $2l2\nu$ fiducial volume definition, by treating the two muons in the events as neutrinos. The shape difference for qqZZ as the function of di-jet invariant mass in $2l2\nu$ signal region is shown in figure 6.14 (left). Due to the absence of MET significance in truth level, one cannot get BDTG distribution in fiducial volume. The final treatment of this uncertainty item in the $2l2\nu$ channel is then decided to apply in MJJ based on the reco-level MJJ value event by event, and transfer it to BDTG disctiminant as shown in figure 6.14 (right).

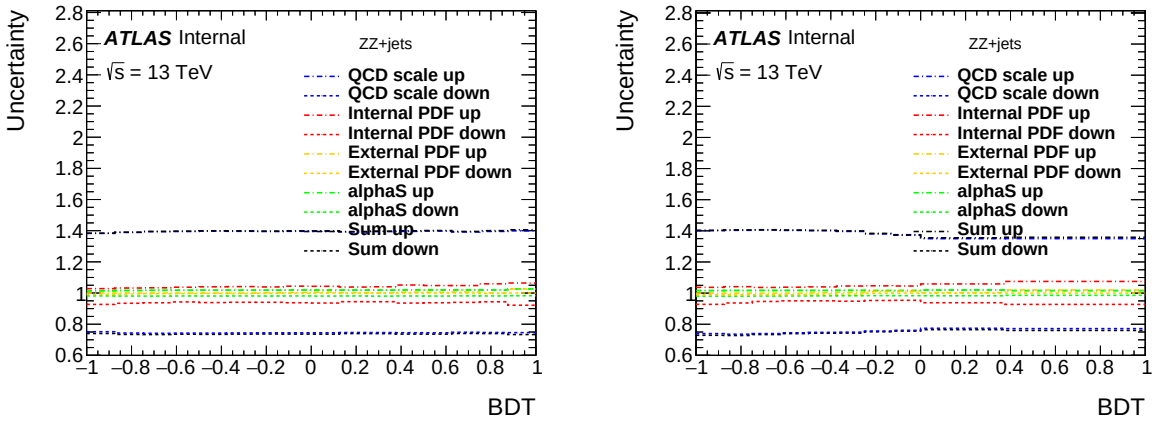


Figure 6.11: The theoretical uncertainties for ggZZ background in (a) signal region; (b) control region.

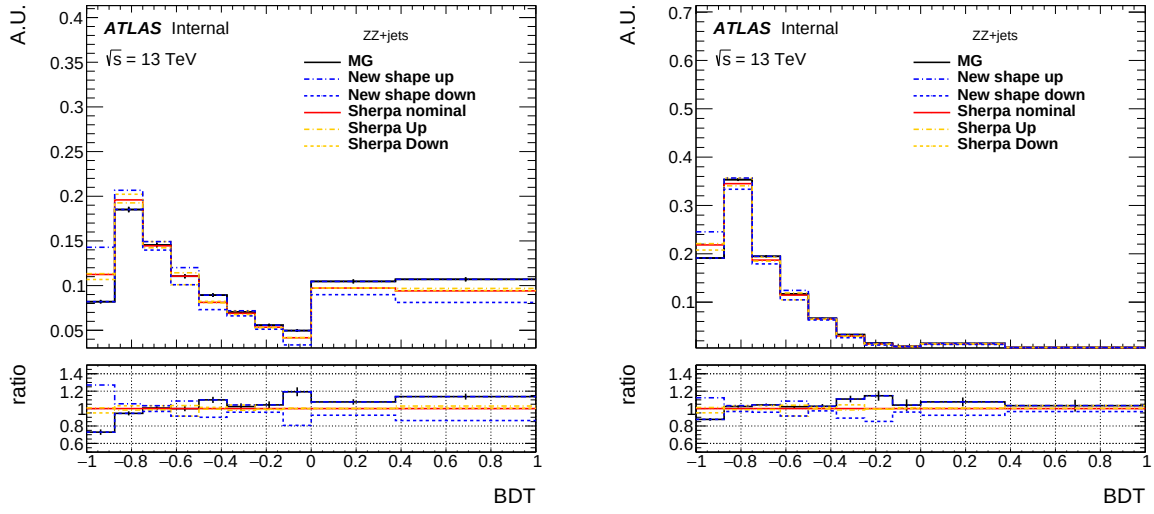


Figure 6.12: Comparison of BDT shape difference for $q\bar{q} \to ZZ$ background between different SHERPA theoretical uncertainties and MADGRAPH5 in 4ℓ channel of (a) signal region; (b) control region.

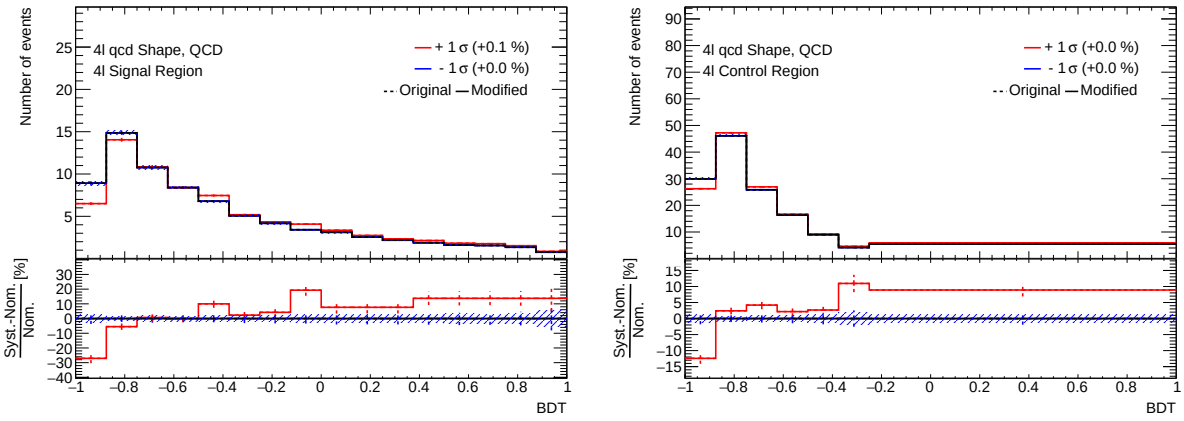


Figure 6.13: The shape uncertainties for qqZZ background in 4ℓ channel of (a) signal region; (b) control region. In the final fitting, to be more conservative, two nuisance parameters have been used in SR and CR separately for this shape uncertainty.

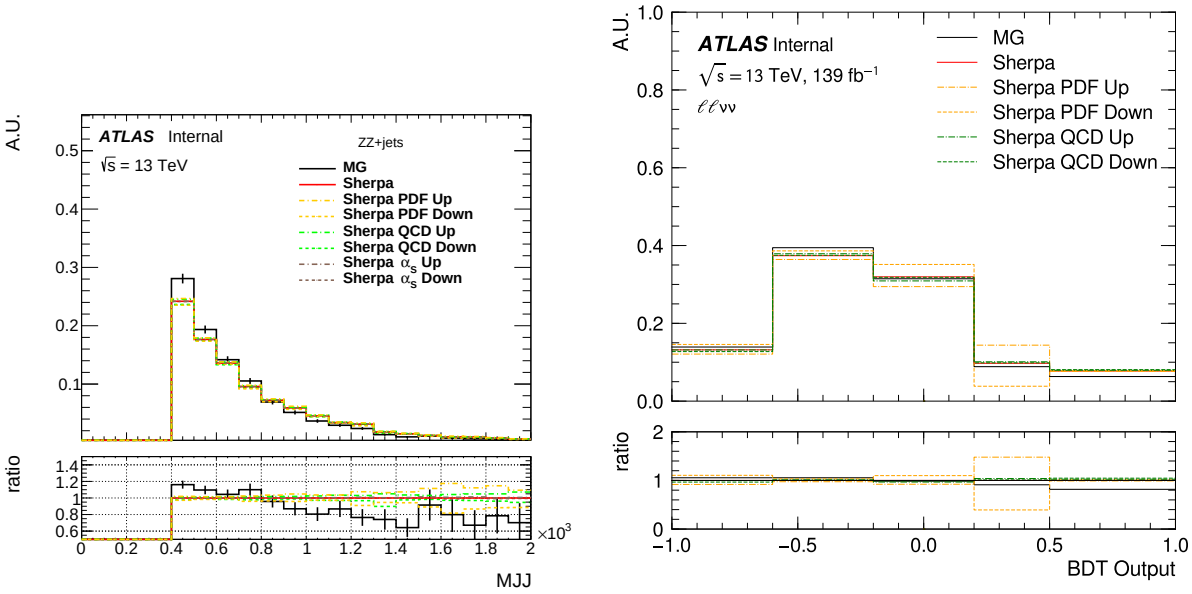


Figure 6.14: Comparison of shape difference for qqZZ background between different Sherpa theoretical uncertainties and MG in $llv\nu jj$ channel of (a) MJJ distribution; (b) BDTG distribution.

6.4.5 Interference between the EWK and QCD processes

The interference between the EWK and QCD processes is treated as an additional systematic on the EWK signal. It is estimated directly by producing the $ZZjj$ sample with only the electroweak and QCD interfering term at the leading order with MadGraph5 v2.6.0 generator.

The theoretical interpretation of the modeling recipe in method 2 is according to the orders of QCD and QED coupling at the Matrix-Element-Square (ME-Sq) level. At LO, the QCD-mediated $ZZjj$ and electroweak-mediated $ZZjj$ have different QCD and QED coupling orders at the ME level.

At the ME-Sq level, the squared amplitudes of the combined $ZZjj$ sample with both electroweak and QCD-mediated productions and their interference effects at LO are parameterized as equation 6.3.

$$|M|^2 = |M_{EWK} + M_{QCD}|^2 = |M_{EWK}|^2 + |M_{QCD}|^2 + 2 \times Re(M^*_{EWK} \cdot M_{QCD}) \tag{6.3}$$

Each of the EWK, QCD, and interference terms could be calculated separately, by requiring different orders of QED and QCD vertex.

In $llljj$ channel, the interference is included as a BDT-dependent systematics, by using truth level information (vs. m_{jj}) and transferred to the BDT score. The final BDT distributions of the nominal signal and signal with inference term in the SR are shown in Figure 6.15.

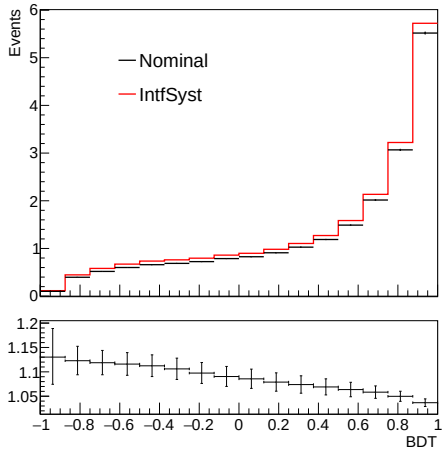
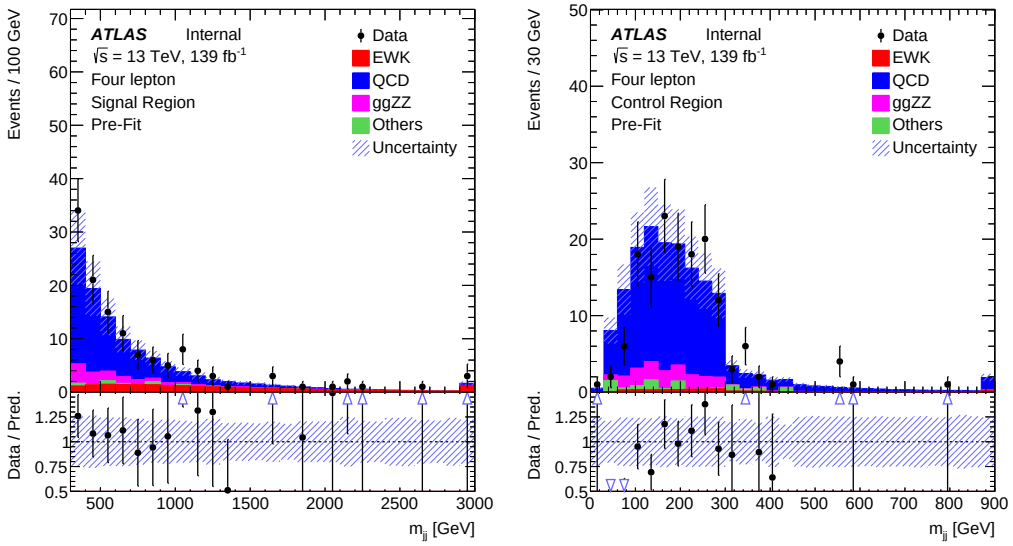


Figure 6.15: Comparison of interference systematic variation and nominal for the $llljj$ channel in SR. The bottom panel shows the ratio of those two.

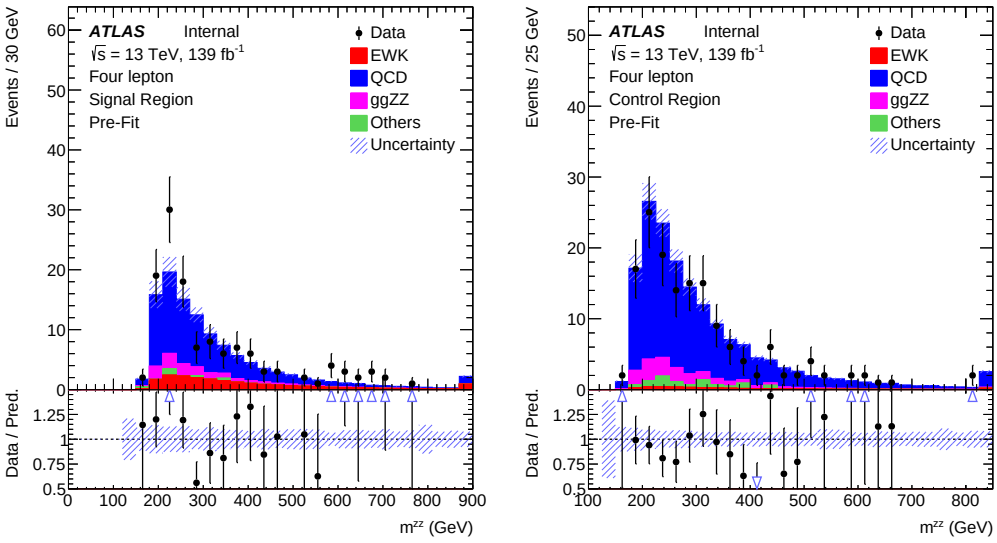
6.5 Data and Monte Carlo simulation comparison

This section shows plots of comparison of kinematics distributions between data and Monte Carlo simulation in the $llljj$ channel from figure 6.16 to figure 6.28. Distributions in the signal region and QCD control regions are provided together for comparison. Those kinematic variables are used in the multivariate analysis described in section 6.7. All the plots in this section are pre-fit ones but already include the data-driven backgrounds. The experimental systematics have been included in the uncertainty band for all plots. The theoretical systematics have been included in the m_{jj} plot. To check the multiplicity of jets in the $llljj$ channel, distributions of the number of jets in each event are shown in figure 6.29 for both the signal and QCD control regions.



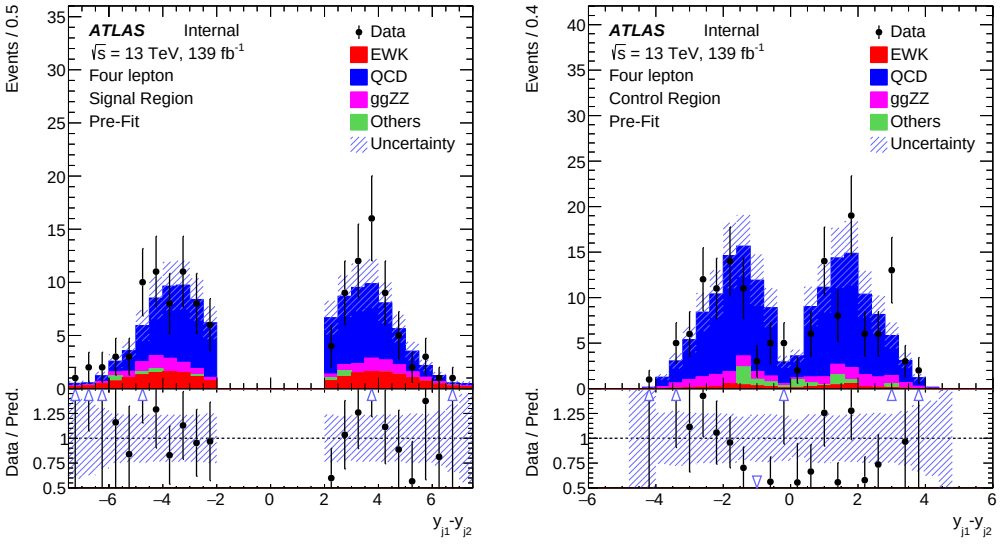
(a) (b)

Figure 6.16: D-jet invariant mass (m_{jj}) distributions in SR (left) and QCD CR (right) with experimental and theoretical systematics included.



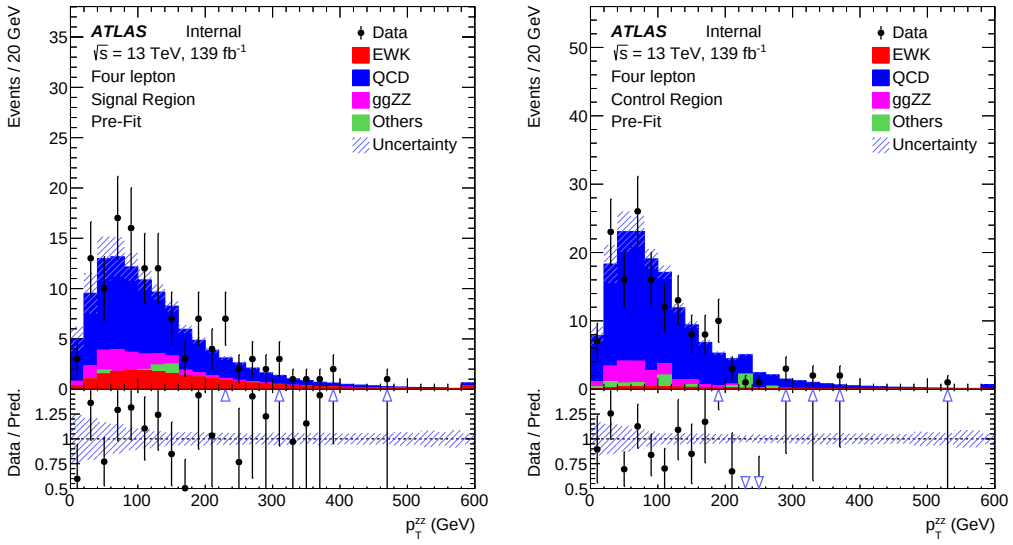
(a) (b)

Figure 6.17: Four lepton mass distributions in SR (left) and QCD CR (right) with experimental systematics included.



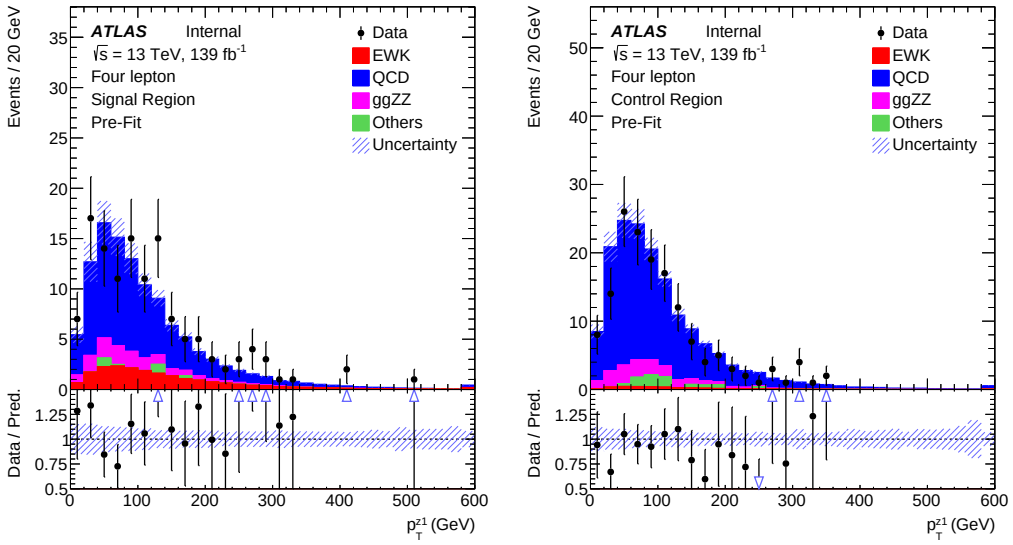
(a) (b)

Figure 6.18: Rapidity difference between the leading and sub-leading jets distributions in SR (left) and QCD CR (right) with experimental systematics included.



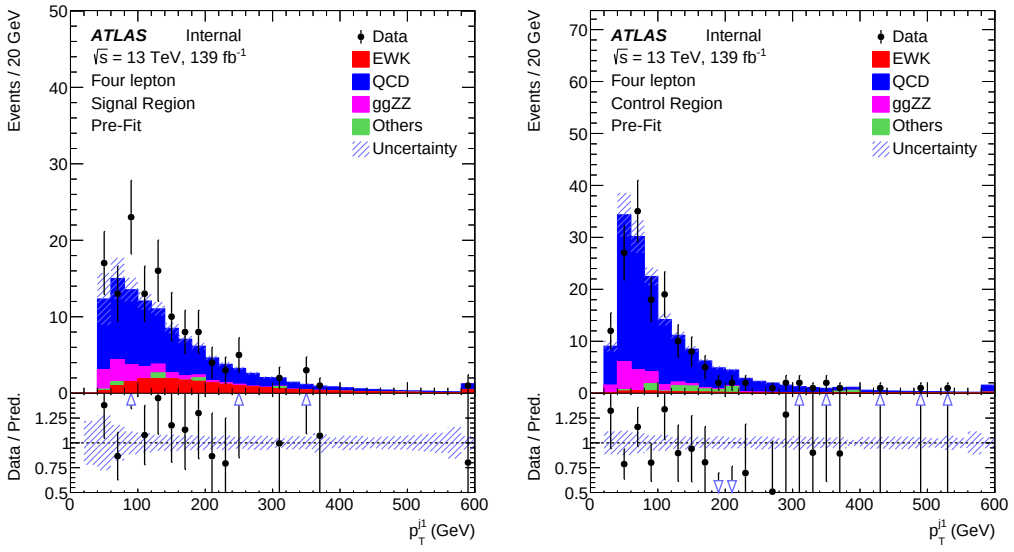
(a) (b)

Figure 6.19: Four lepton p_{T} distributions in SR (left) and QCD CR (right) with experimental systematics included.



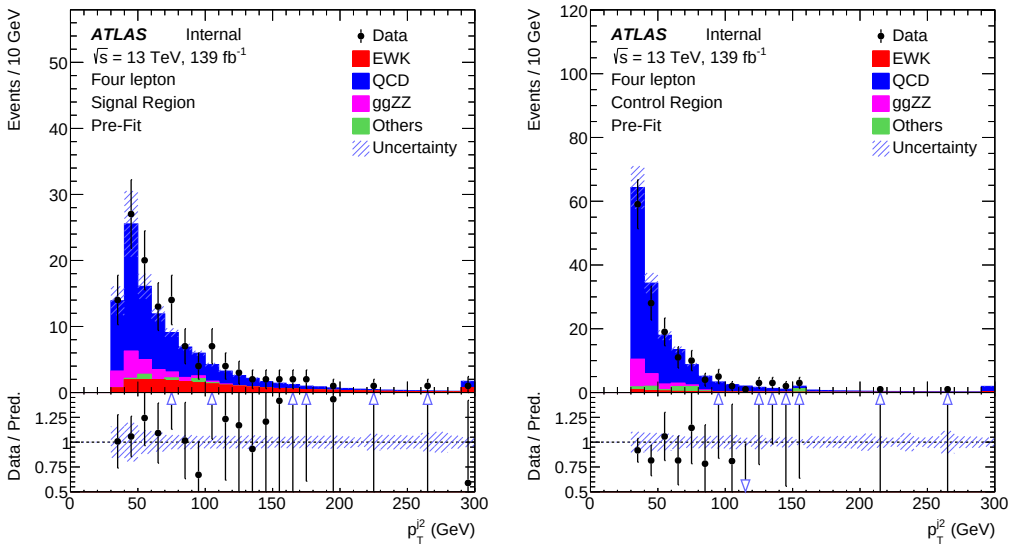
(a) (b)

Figure 6.20: Leading lepton pair distributions in SR (left) and QCD CR (right) with experimental systematic uncertainties included.



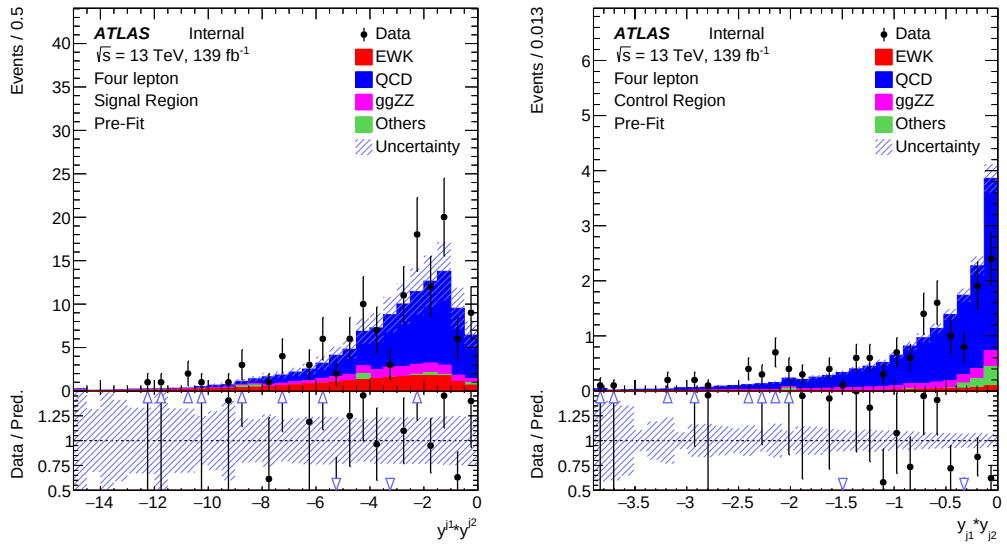
(a) (b)

Figure 6.21: Leading jet p_{T} distributions in SR (left) and QCD CR (right) with experimental systematic uncertainties included.



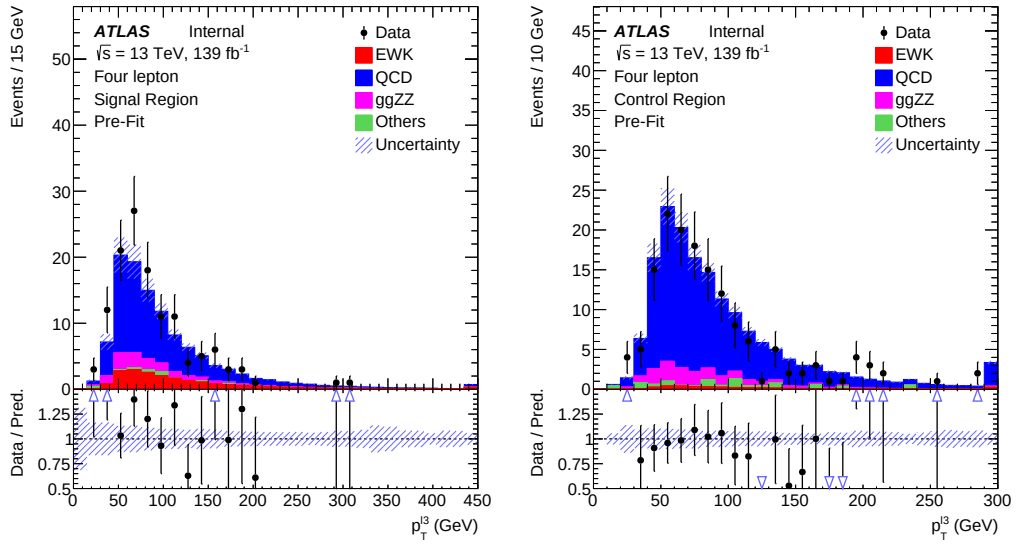
(a) (b)

Figure 6.22: Sub-leading jet p_{T} distributions in SR (left) and QCD CR (right) with experimental systematic uncertainties included.



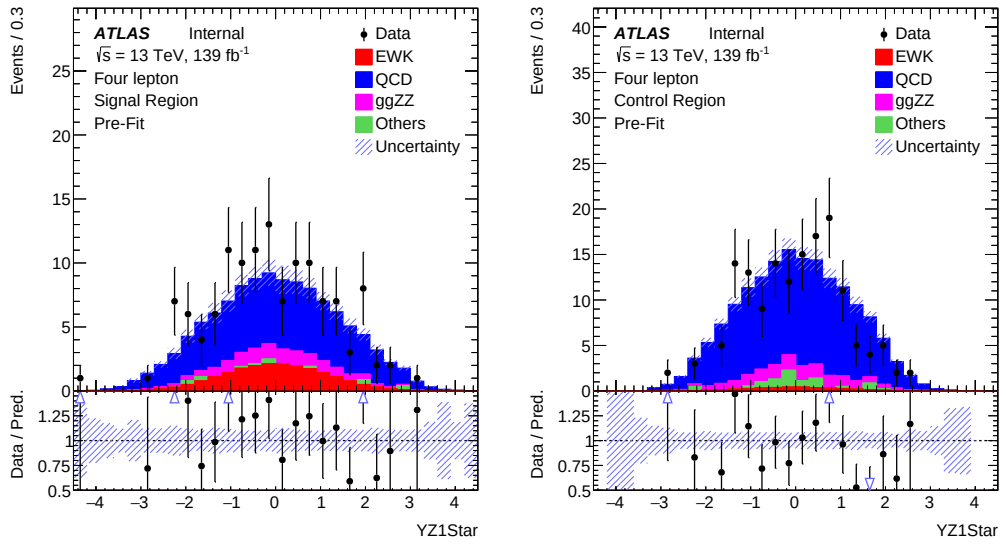
(a) (b)

Figure 6.23: Product of rapidity of leading and sub-leading jets distributions in SR (left) and QCD CR (right) with experimental systematic uncertainties included.



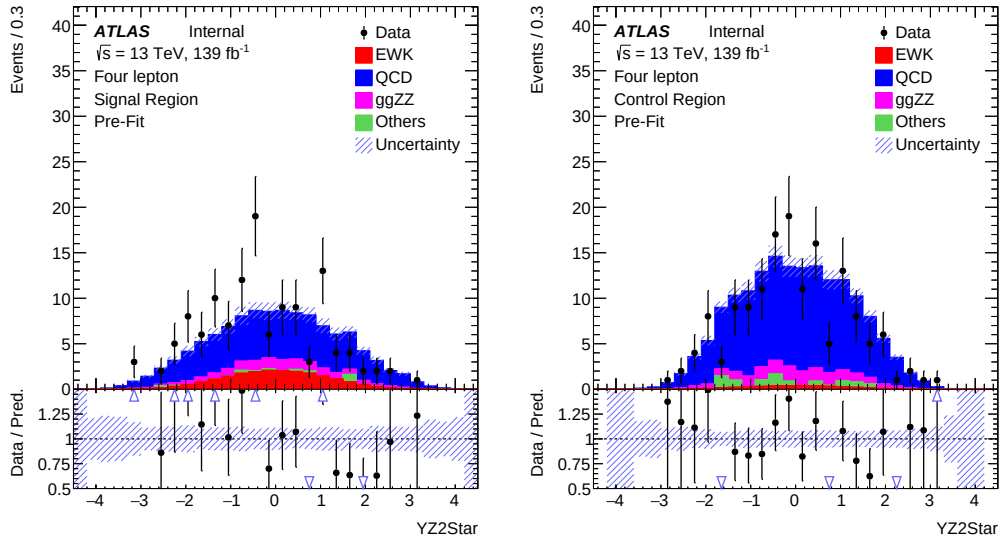
(a) (b)

Figure 6.24: p_{T} of leading lepton in the sub-leading lepton pair (p_T^{L3}) distributions in SR (left) and QCD CR (right) with experimental systematic uncertainties included.



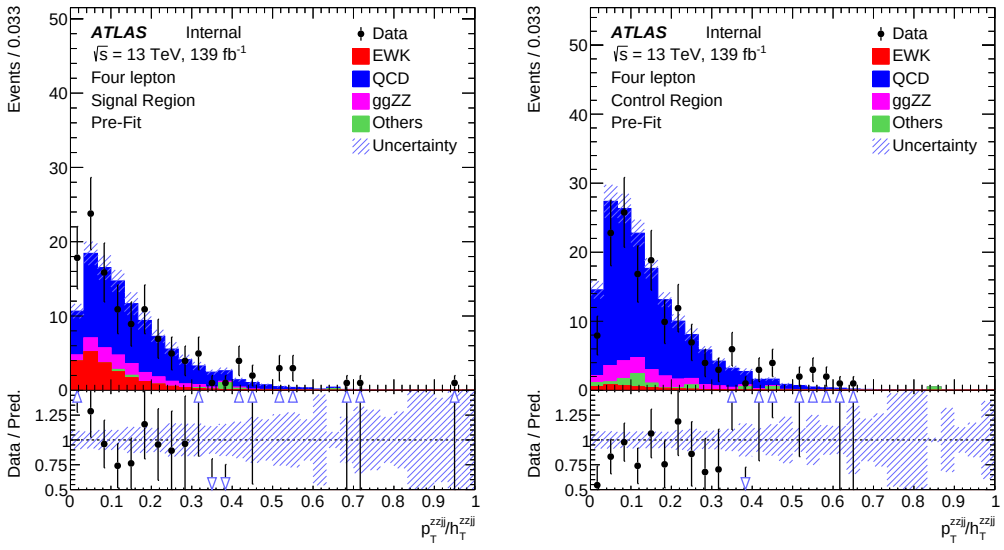
(a) (b)

Figure 6.25: y^*_{Z1} distributions in SR (left) and QCD CR (right) with experimental systematic uncertainties included. $y^*_{Z1} = y_{Z1} - (y_{j1} + y_{j2})/2$ and y refers to the rapidity of an object.



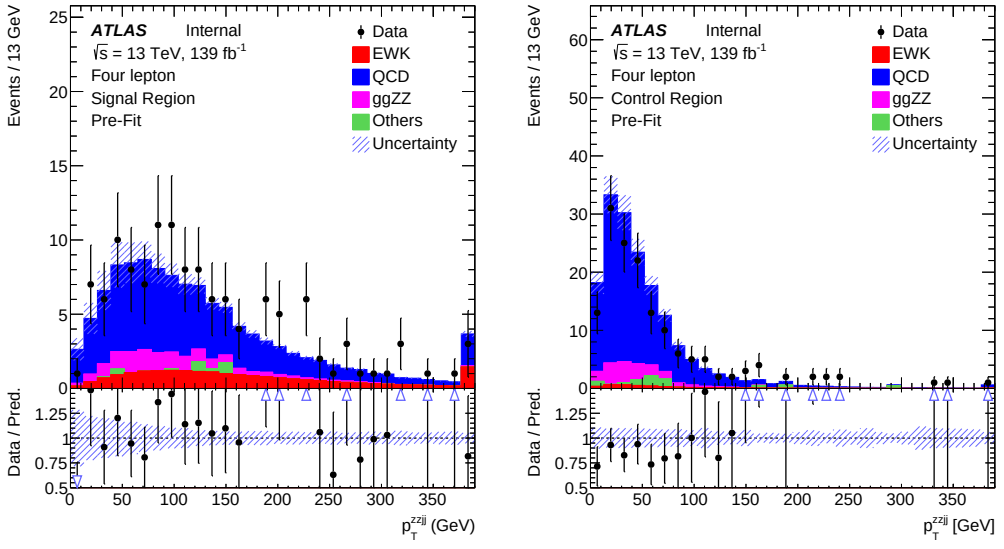
(a) (b)

Figure 6.26: y^*_{Z2} distributions in SR (left) and QCD CR (right) with experimental systematic uncertainties included. $y^*_{Z2} = y_{Z1} - (y_{j1} + y_{j2})/2$ and y refers to the rapidity of an object.



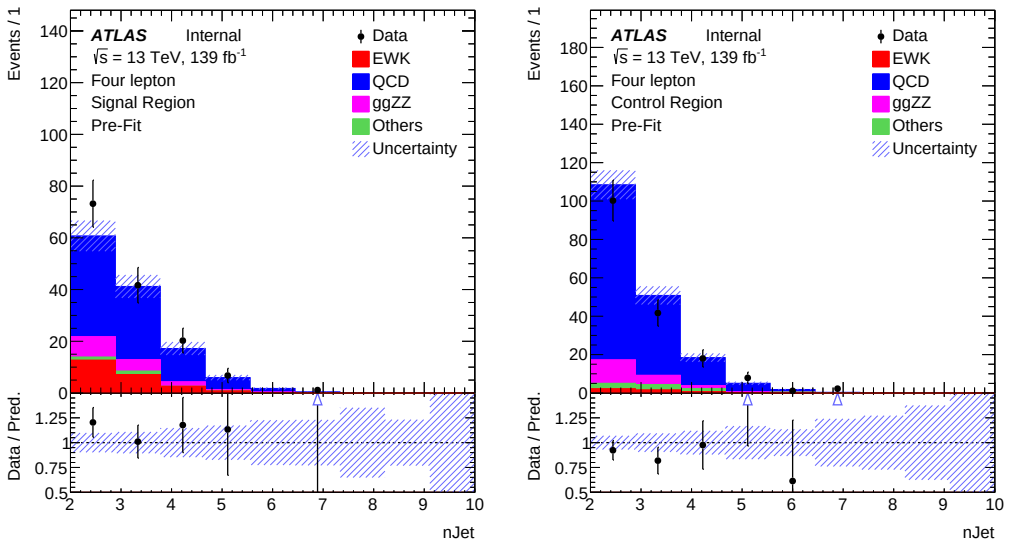
(a) (b)

Figure 6.27: p_T^{ZZjj}/h_T^{ZZjj} distributions in SR (left) and QCD CR (right) with experimental systematic uncertainties included. h_T^{ZZjj} refers to the scalar sum of p_{T} of two z boson and two jets while p_T^{ZZjj} refers to the vector sum.



(a) (b)

Figure 6.28: p_{T} of the $ZZjj$ system distributions in SR (left) and QCD CR (right) with experimental systematic uncertainties included.



(a) (b)

Figure 6.29: Number of jets distributions in SR (left) and QCD CR (right) with experimental systematic uncertainties included.

6.6 Inclusive $ZZjj$ production cross-section measurement

The cross-section of inclusive $4l$ plus two jets, and $2l2\nu$ plus two jets are measured and summarized in this section. The cross sections are measured in fiducial volumes as below. The detector correction factor (C factor) is estimated, as events passing full reconstruction level selections, divided by events passing truth level FV selections.

6.6.1 Definition of the fiducial volume

The fiducial volumes (FV) are defined for the cross-section measurements in the $llll$ and $ll\nu\nu$ channels accordingly. The event and object selections in the fiducial volumes are close to selections used in the reconstruction level signal regions. Some selections are only available in

For the $4l$ channel, the FV is defined as:

- Leptons are dressed with photons within $\Delta R(\ell, \gamma) < 0.1$;
- $|\eta^e| < 2.47$, $|\eta^\mu| < 2.7$;
- $p_T^\ell > 20$, 20, 10, 7 GeV for the leading, sub-leading, third and last lepton;

- $\Delta R(\ell, \ell') > 0.2$;
- Select two Z candidates, minimizing $|m_{z1} - m_z| + |m_{z2} - m_z|$. The one SFOS with dilepton mass closest to Z mass is labeled as Z1. The other one as Z2;
- $60 < m_{Z1} < 120$ GeV, $60 < m_{Z2} < 120$ GeV;
- $m_{\ell^+\ell^-} > 10$ GeV for all the SFOS dilepton;
- At least two AntiKt4TruthWZ jets and select one leading jet per η side;
- $p_T^j > 30$ (40) GeV for jets with $|\eta^j| < 2.4$ (4.5);
- $\Delta y_{jj} > 2$ and $m_{jj} > 300$ GeV;

For the $2l2\nu$ channel, the FV is defined as:

- Leptons are dressed with photons within $\Delta R(\ell, \gamma) < 0.1$;
- $|\eta^\ell| < 2.5$;
- $p_T^\ell > 30$, 20 GeV for the leading and sub-leading lepton;
- Veto events with a third lepton satisfying $p_T^\ell > 20$ GeV and $|\eta^\ell| < 2.5$;
- $80 < m_{\ell\ell} < 100$ GeV;
- Truth $E_T^{\text{miss}} > 130$ GeV, where truth E_T^{miss} is defined as the vector sum of all the neutrinos not from hadron decay;
- At least two AntiKt4TruthWZ jets;
- $p_T^j > 60$, 40 GeV for the leading and sub-leading jet, $Y_{j1} \times Y_{j2} < 0$;
- $\Delta y_{jj} > 2$ and $m_{jj} > 400$ GeV;

The FVs are used as the pre-VBS regions, where inclusive cross sections are measured in the $4l$ or $2l2\nu$ plus 2jets channels. Differences between $4l$ and $2l2\nu$ FV definitions are due to different event selections applied at reconstruction-level analyses. In general, tighter selections are applied in the $2l2\nu$ channel due to more background contributions.

6.6.2 C factor

The definition of the C factor is the ratio of the number of events passing reconstruction level selections and the number of events number in fiducial volume at the truth level, for the defined signal processes:

$$C = \frac{N_{Reco}}{N_{F.V-truth}}$$

C factors in the $4l$ channel are calculated as:

Sample	C	ΔC(stats)	ΔC(sys)	ΔC(theo)
EWK ZZjj	0.663	±0.002	$\pm^{0.032}_{0.031}$	NA
qq→ZZjj (QCD)	0.702	±0.003	$\pm^{0.061}_{0.051}$	$\pm^{0.015}_{0.018}$
gg→ZZjj	0.741	±0.021	$\pm^{0.143}_{0.072}$	±0.002

Table 6.16: C Factor of ZZjj→ $4l$ process

C factors in the $2l2\nu$ channel are calculated as:

	C	ΔC(stats)	ΔC(syst)	ΔC(theo)
EWK $ZZ \to \ell\ell\nu\nu + 2j$	0.2303	± 0.0015	± 0.0044	± 0.0009
$qqZZ \to \ell\ell\nu\nu + 2j$	0.2060	± 0.0055	± 0.0127	± 0.0141
$ggZZ \to \ell\ell\nu\nu + 2j$	0.1899	± 0.0056	± 0.0094	

Table 6.17: C Factor of ZZjj→ $\ell\ell\nu\nu$ process

When calculating the C factor, overlap removal between truth leptons and ANTIKT4WZ truth jets is applied, where jets within $\Delta R(l, j)$ <0.2 are removed.

The combined C factor in the $4l$ channel is calculated as follows:

$$C_{ZZjj\to 4l} = \sum_i \frac{N^i_{F.V-truth}}{\sum_j N^j_{F.V-truth}} \times C_i = 0.699 \pm 0.003(stats) \pm^{0.011}_{0.013} (theo) \pm 0.028(exp)$$

The combined C factor in the $2l2\nu$ channel is calculated with the same formula, as:

$$C_{ZZjj\to\ell\ell\nu\nu} = 0.2158 \pm 0.0032(stats) \pm 0.0080(theo) \pm 0.0076(exp) \quad (6.4)$$

Statistical uncertainty refers to the uncertainty from signal MC statistics. Systematic uncertainty refers to the experimental systematic uncertainty on signal processes, as discussed in Section 6.4.1.

6.6.3 Cross section of $4l$ and $2l2\nu$ channels

The cross-section of inclusive $ZZjj$ production in fiducial volume is calculated as:

$$\sigma^{F.V} = \frac{N_{data} - N_{bkg}}{C \times Lumi},$$

where N_{data} refers to the number of events selected from data passing SR reconstruction level selections, while N_{bkg} refers to the number of background events passing SR reconstruction level selections. The observed and expected yields are listed in table 6.7. The predicted and measured cross-section in the $llll$ and $ll\nu\nu$ channels are summarized in table 6.18.

Channel	Measured fiducial σ [fb]	Predicted fiducial σ [fb]
$llljj$	1.27±0.12(stat)±0.02(theo)±0.07(exp)±0.01(bkg)±0.02(lumi)	1.26±0.04(stat)±0.22(theo)
$ll\nu\nu jj$	1.13±0.28(stat)±0.04(theo)±0.06(exp)±0.15(bkg)±0.02(lumi)	1.11±0.01(stat)±0.12(theo)

Table 6.18: Measured and predicted fiducial cross-sections of inclusive $ZZjj$ production.

In table 6.18, uncertainties due to different sources are presented explicitly, including the one from the statistical uncertainty of the data and simulated samples (stat), the one from the theoretical predictions (theo), the experimental ones due to the lepton and jet calibrations (exp), the ones from background estimates (bkg), and the one from luminosity (lumi).

6.7 Statistical analysis and result

6.7.1 Multivariate analysis

The inclusive cross section calculated in the previous section includes $ZZjj$ productions via pure electroweak and QCD processes. Another important goal of this analysis is to calculate the significance of the $ZZjj$ production via pure electroweak processes. As shown in table 6.7, the signal and background ratio in signal regions of both the $llljj$ and $ll\nu\nu jj$ channels although very tight cuts already applied. To improve the sensitivity of the signal, the Gradient Boost Decision Tree (BDTG) method, introduced in section 5, is used to distinguish the EWK and other components. The BDTG method used in this analysis is based

on the TMVA package [2]. For the $4l$ channel, the dominant background comes from QCD ZZ plus two jets, and training is performed based on the EWK and QCD samples. During BDT training, EWK samples are assigned with label=1, while QCD samples are assigned with label=-1. Hyper-parameter settings for BDT in the $lllljj$ channel are summarized in table 6.19.

$llll$	NTrees=1000,MinNodeSize=2.5%,BoostType=Grad Shrinkage=0.10,UseBaggedBoost,BaggedSampleFraction=0.5 nCuts=30,MaxDepth=5

Table 6.19: Hyper-parameters of gradient BDT in the $lllljj$ channel

For the $ll\nu\nu$ channel, the dominant backgrounds come from WZ and non-resonant components. In BDT training EWK signal is assigned with label=1, while backgrounds including QCD (both qqZZ and ggZZ), non-resonant, WZ, and other samples are assigned with label=-1. After signal region selection, Z+jets has been highly suppressed. Thus, Z+jets is not used in the training background. Hyper-parameter settings for BDT in the $ll\nu\nu jj$ channel are summarized in table 6.20.

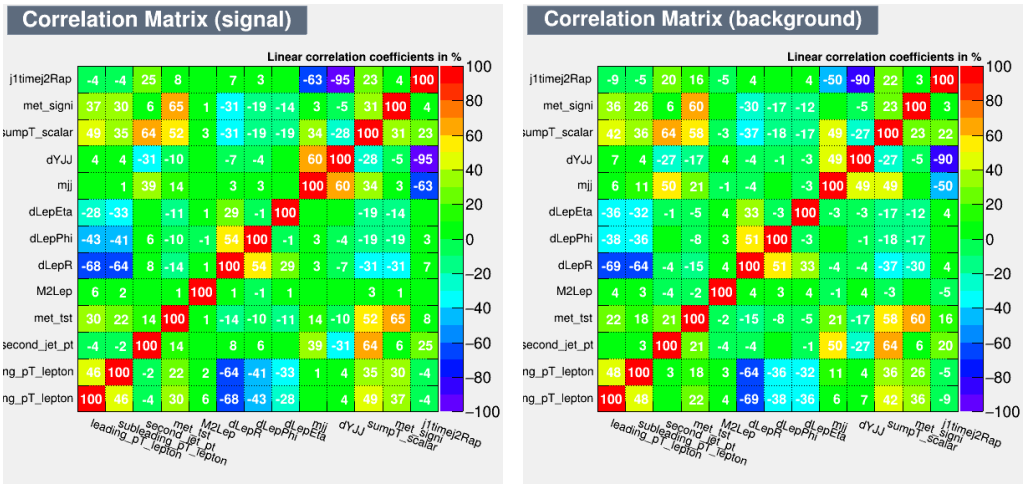
$ll\nu\nu$	NTrees=500, MinNodeSize=25%, BoostType=Grad Shrinkage=0.10, UseBaggedBoost, BaggedSampleFraction=0.5 nCuts=20, MaxDepth=3"

Table 6.20: Hyper-parameters of gradient BDT in the $ll\nu\nu jj$ channel

In the hyper-parameter table, the shrinkage rate represents the learning speed and bagged boost means the BDT is trained on a random sample of data selected with replacement from the original dataset to reduce the overtraining issue.

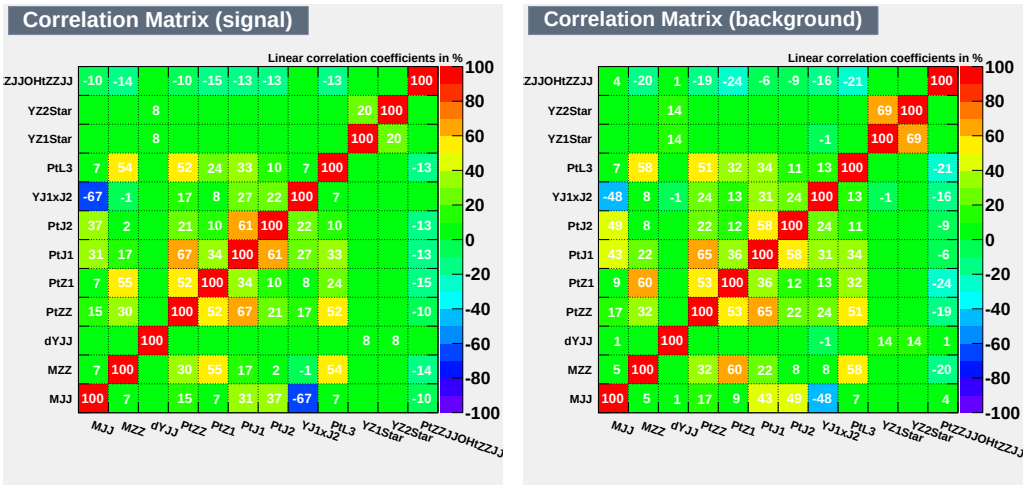
Input variables used for the training in both channels are summarized in Table 6.21, ordered by the variable importance in descending order. The variable importance is calculated by using the average gain across all splits where the variable was used when building each tree. Distributions of input variables are shown in Section 6.5, side-by-side comparing SR and QCD CR ones. Input variable linear correlation coefficients plots are shown in Figure 6.30. Considering the statistics of signal and background samples, $SpliitMode = Random$ and $NormMode = NumEvents$ are used in preparing the training and test trees step. The parameters used in BDTG are listed in Table 6.20.

[2]TMVA is the ROOT library that provides the interfaces and implementations of machine learning techniques such as neural network and boost decision tree. https://root.cern/manual/tmva.



(a) (b)

Figure 6.30: Input variable linear correlation coefficients in signal (left) and background (right) for the $2l2\nu$ channel.



(a) (b)

Figure 6.31: Input variable linear correlation coefficients in signal (left) and background (right) for the $4l$ channel.

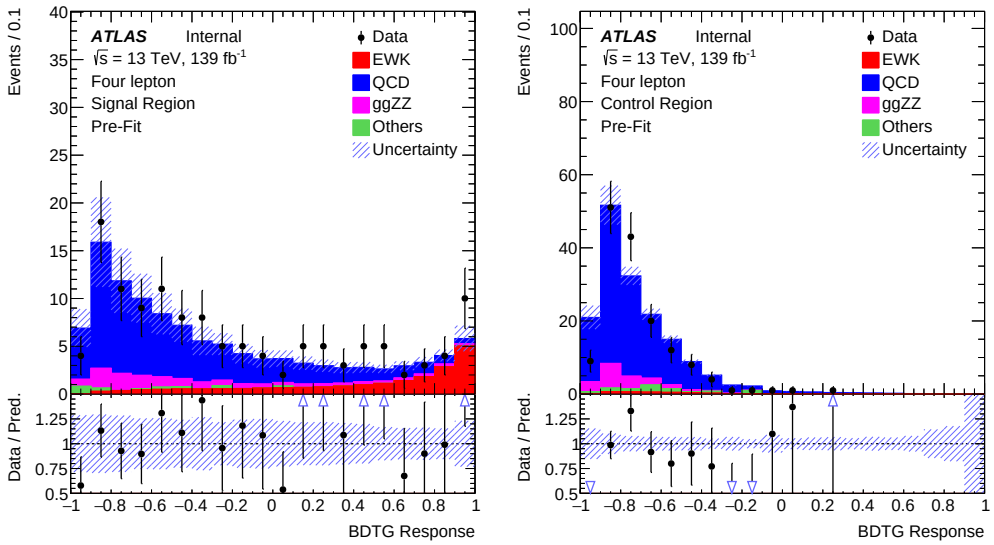
Rank	$llvvjj$ variables	$lllljj$ variables
1	$\Delta\eta(ll)$	m_{jj}
2	m_{ll}	leading p_T^j
3	$\Delta\phi(ll)$	sub-leading p_T^j
4	m_{jj}	$p_T(ZZjj)/H_T(ZZjj)$
5	$E_T^{\text{miss}} significance$	$Y(j1) \times Y(j2)$
6	$\Delta Y(jj)$	Δy_{jj}
7	$Y(j1) \times Y(j2)$	Y^*_{Z2}
8	HT	Y^*_{Z1}
9	$\Delta R(ll)$	p_T^{ZZ}
10	sub-leading p_T^j	m_{ZZ}
11	E_T^{miss}	p_T^{Z1}
12	sub-leading p_T^l	$p_T^{\ell 3}$
13	leading p_T^l	-

Table 6.21: Input variables for the $2l2\nu$ channel (left) and $4l$ channel (right) ordered by variable importance in descending order. The variable importance is calculated by using the average gain across all splits where the variable was used when building each tree

The final BDT score distribution in SR and QCD CR is shown in figure 6.32 for the $lllljj$ channel, and in figure 6.33 for $llvvjj$ SR and VR before statistical fitting. The BDT score shows a good separation between signal and backgrounds in the $lllljj$ channel.

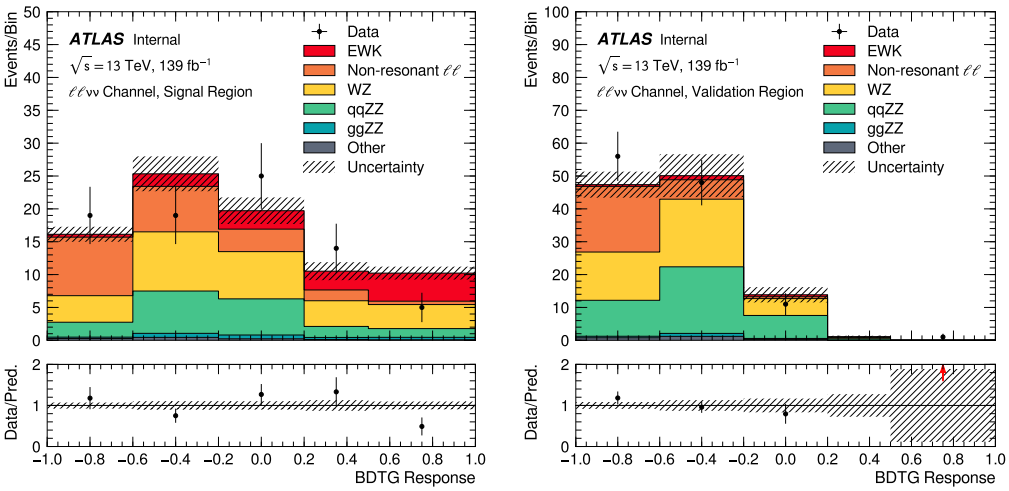
6.7.2 Statistical fitting and result

To calculate signal strength μ_{EW} and significance of the pure electroweak $ZZjj$ production, the statistical fitting based on the maximum likelihood method and profile likelihood ratio, which are introduced in chapter 5, are used. Since the BDT score distribution shows a good separation between signal and background, the likelihood is calculated based on the BDT score distribution for both $lllljj$ and $llvvjj$ channels. In this case, the BDT score distribution can be called the main discriminator (MD). In both channels, the binned likelihood functions are used, which is shown in 5.3. The signal strength of the EWK is treated as the parameter of interests and scale factors on backgrounds as well as uncertainties are treated



(a)

Figure 6.32: Pre-fit BDT score distributions in SR (a) and QCD CR (b) for the $llljj$ channel, with experimental systematics included in the plot.



(a) (b)

Figure 6.33: Pre-fit BDT score distributions in SR (a) and VR (b) for the $llννjj$ channel, with experimental systematics included in the plot.

as the nuisance parameters. In the $lllljj$ channel, the fitting is spontaneously performed in both the signal region and QCD control region, while in the $ll\nu\nu jj$ channel only the signal region is included. The statistical tests are performed in both the individual $lllljj$ and $ll\nu\nu jj$ channels and in the combined channels.

For the treatment of nuisance parameters, the experimental systematic uncertainties are considered as correlated in all the bins and regions whenever applicable. The theoretical uncertainties for $ZZjj$ productions are treated as uncorrelated between the $lllljj$ and $ll\nu\nu jj$ channels, due to the different fiducial volume definitions. The QCD scale uncertainty for QCD $ZZjj$ production can be assessed in various ways in terms of correlations between fitted regions and is conservatively treated as uncorrelated between the SR and QCD CR in the $lllljj$ channel. Furthermore, the generator modeling uncertainty for QCD $ZZjj$ production is treated as uncorrelated between the low and high BDT score regions.

After statistical fitting, the BDT score distribution in the $lllljj$ QCD CR, $lllljj$ SR, and $ll\nu\nu jj$ SR are shown in figure 6.34. The post-fit kinematic distributions are shown in figure 6.35

The fitting signal strength μ_{EW} and significance of pure electroweak $ZZjj$ production are shown in table 6.22.

Channel	μ_{EW}	$\mu_{\text{QCD}}^{lllljj}$	Significance Obs. (Exp.)
$lllljj$	0.97±0.27	0.98±0.22	5.5 (5.6) σ
$ll\nu\nu jj$	0.7±0.5	-	1.3 (2.1) σ
Combined	0.92±0.24	0.99±0.22	5.7 (5.9) σ

Table 6.22: Signal strength and significance of EW $ZZjj$ production.

From the combined channels, the observed μ_{EW} is 0.92±0.24 while the $\mu_{\text{QCD}}^{lllljj}$ is determined to be 0.99 ± 0.22. The statistical component accounts for 88% of the total uncertainty in μ_{EW}. The probability that the background can randomly fluctuate to produce a measured likelihood ratio at least as signal-like as the excess observed in data is 1.6×10^{-8}, leading to the observation of electroweak $ZZjj$ production. With a normalized Gaussian distribution, the background-only hypothesis is rejected at 5.7(5.9) σ from the data (expectation). The electroweak $ZZjj$ cross-section in the combined fiducial volume is found to be 0.70 ± 0.18 fb. The cross-section is calculated by multiplying μ_{EW} to the cross-section predicted by the SM, which equals 0.76 ± 0.04 fb.

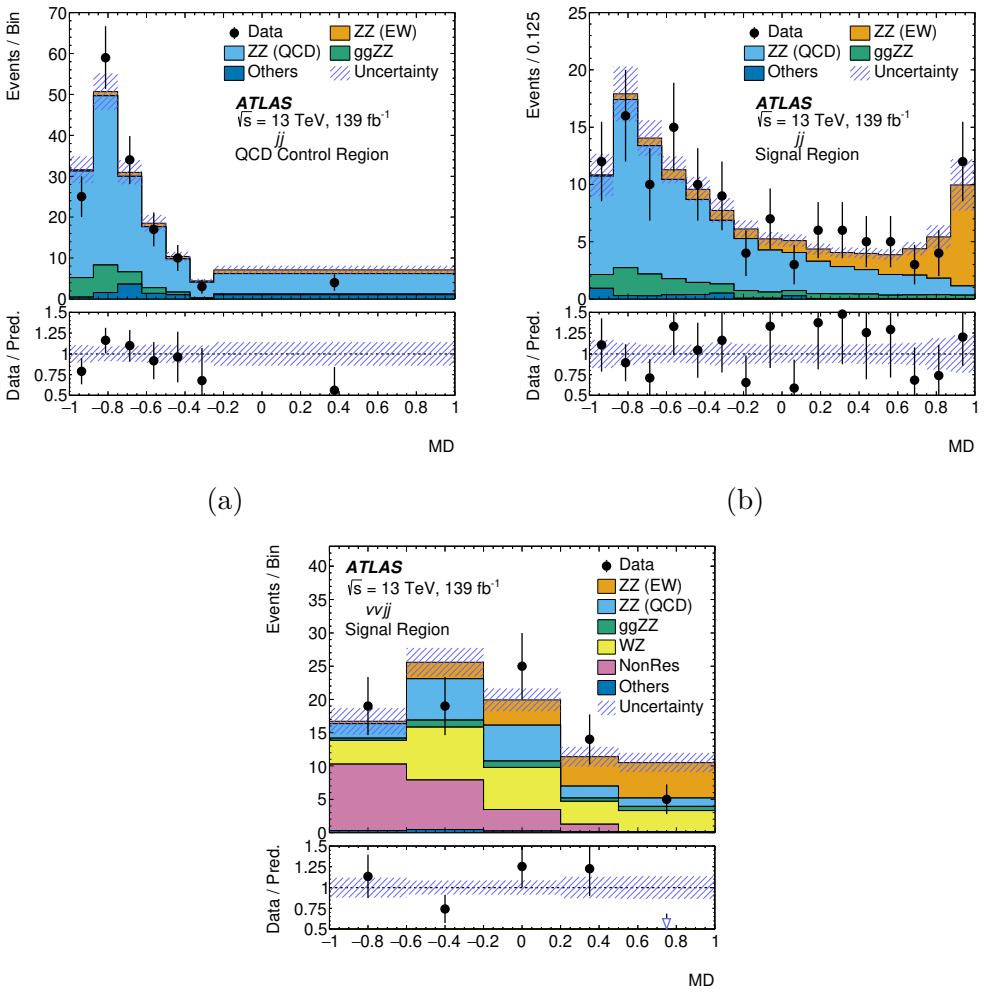


Figure 6.34: Observed and expected BDT score distributions. Distributions are shown after the statistical fit in the (a) $llljj$ QCD control region and in the (b) $llljj$ and (c) $ll\nu\nu jj$ signal regions.

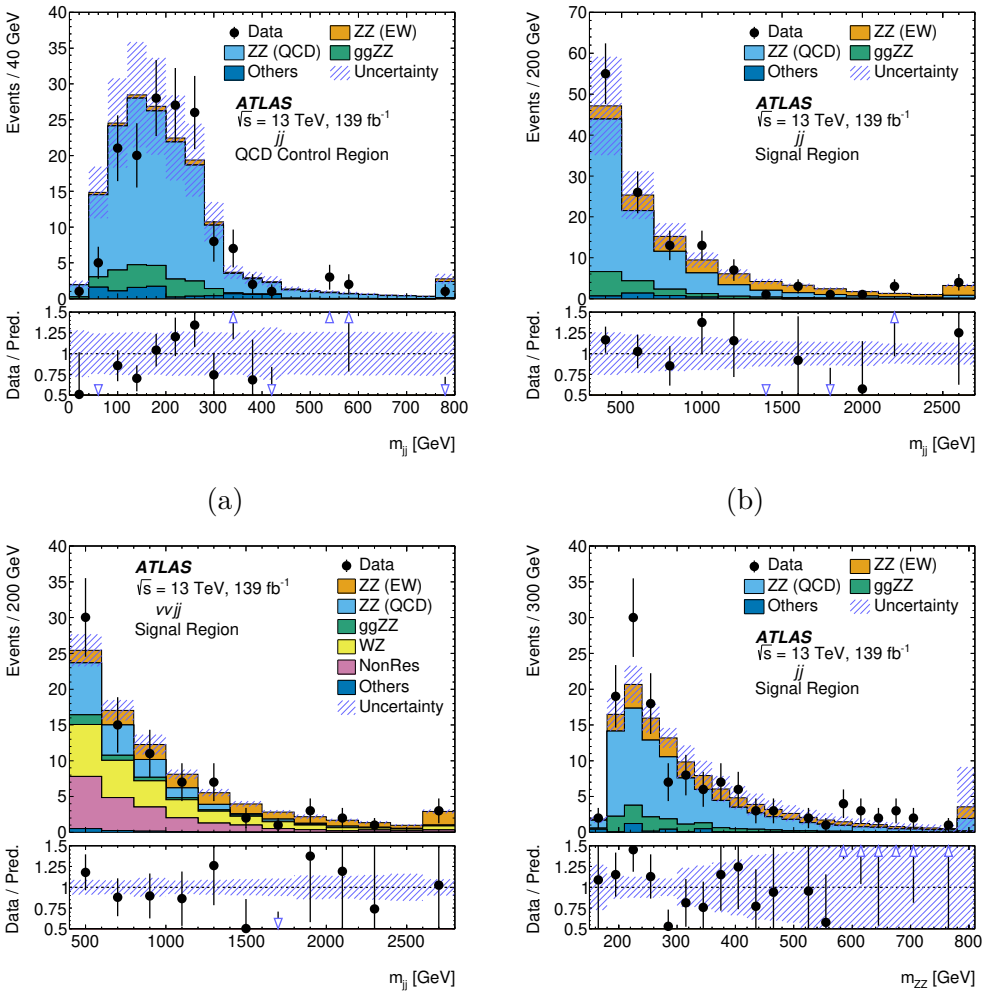


Figure 6.35: Observed and expected kinematic distributions after statistical fitting. The m_{jj} distributions in the (a) $lllljj$ QCD CR and the (b) $lllljj$ and (c) $ll\nu\nu jj$ signal regions. Four lepton mass distribution in the $lllljj$ SR (d) is also included.

6.8 Event display

The VBS-like events are selected, as described in section 6.2. In addition, strict cuts with BDTG greater than 0.25, di-jet invariant mass greater than 500 GeV, and 4-lepton invariant mass greater than 500 GeV are applied. Figure 6.36 and 6.37 show event displays of two VBS candidate events in the $ee\mu\mu$ jj final state.

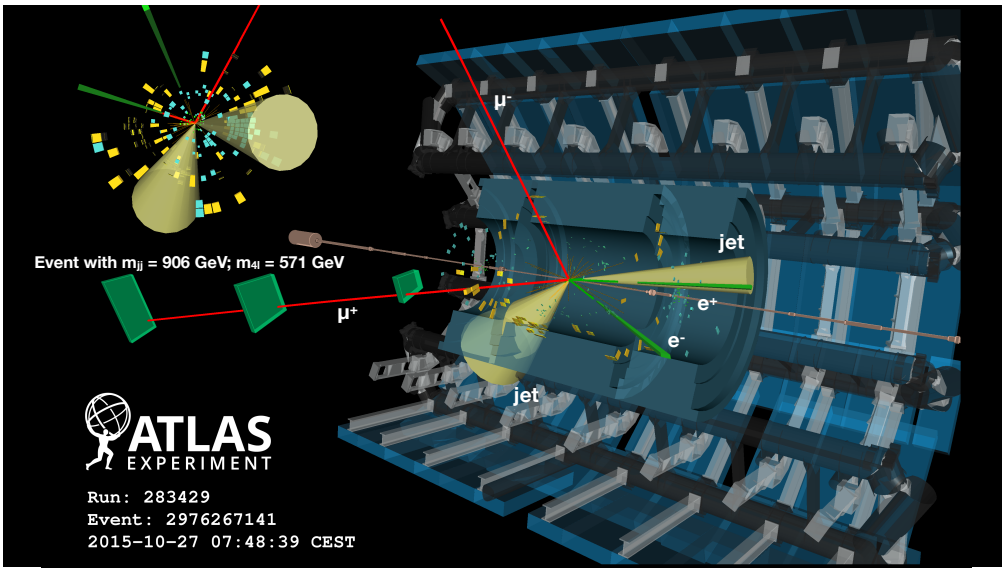


Figure 6.36: The event in data taken in 2015 with run and event number of 283429 2976267141.

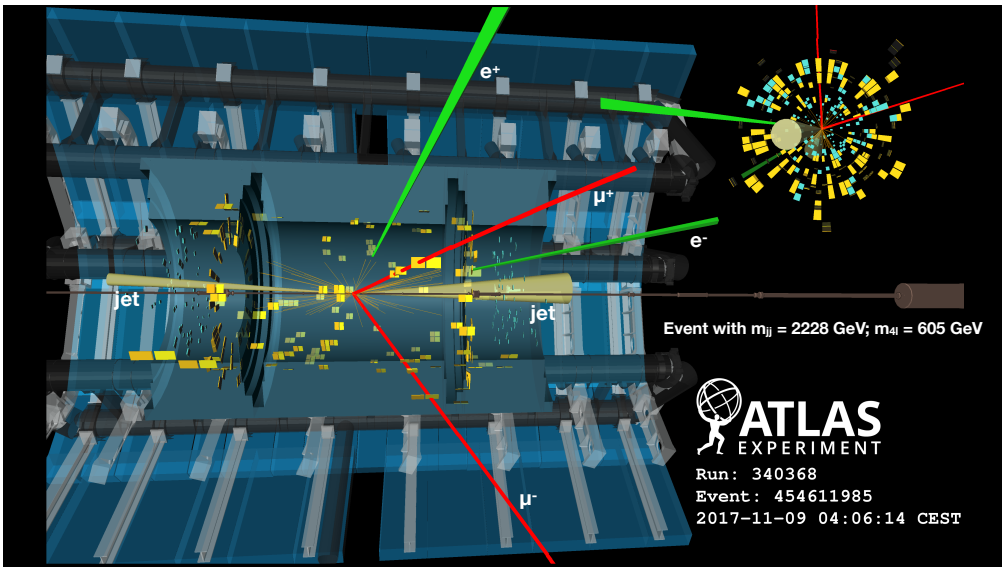


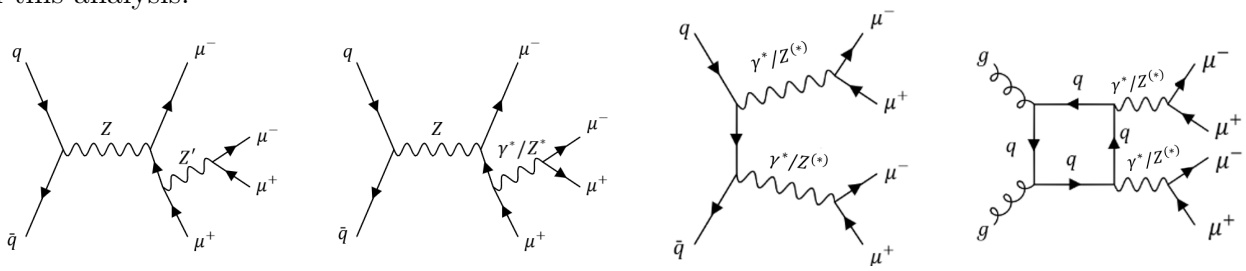
Figure 6.37: The event in data taken in 2017 with run and event number of 340368 454611985.

Chapter 7

Search for a New Gauge boson Z'

The second analysis presented in this thesis is a search for a new leptophilic vector boson Z' which is performed with the four-muon (4μ) final state. The search is focused on the 4μ invariant mass range of [80, 180] GeV, excluding the Higgs boson mass window [110, 130] GeV. The mass range of the Z' in this search is [5, 81] GeV.

In pp collisions at the LHC, the Z' could be produced from final state radiation of μ or τ pairs of the Drell-Yan (DY) process as shown in Figure 7.1a with a 4μ final state, which provides the cleanest signature to search for the Z'. For relatively low Z' mass, the most promising experimental signature would be an excess of 4μ events with the invariant mass of one $\mu^+\mu^-$ pair peaking around the Z' mass. The major background comes from the SM 4μ production processes shown in figure 7.1. The Z' could also be produced in W production through the DY process, $pp \to W \to Z'\mu\nu \to 3\mu + \nu$. The experimental signature would have a final state of 3μ plus large missing transverse energy. This final state is not included in this analysis.



(a) (b) (c) (d)

Figure 7.1: Feynman diagrams of Z' production through radiation in a Drell-Yan process (a), and of the corresponding SM background processes (b - d) with a 4μ final state.

Both the ATLAS and CMS Collaborations have measured the cross-sections of the SM $Z \to 4\mu$ process [111, 112]. The measurement by ATLAS was used by theorists to set limits

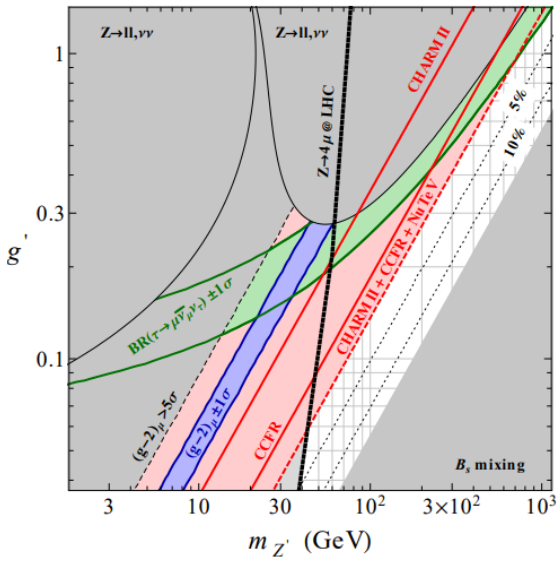


Figure 7.2: Constraints on the model parameter space from the different leptonic processes discussed in Section IV. The region in white is the allowed region. The anomaly in $B \to K^* \mu^+ \mu^-$ can be accommodated everywhere to the left of the bottom-right triangle. Note that the constraint from the neutrino trident production of muon pairs (red region) completely excludes the region favored by $(g-2)_\mu$. The dotted lines in the allowed region denote (5-10)% NP effects in B_s mixing [110].

in the parameter space of the $L_\mu - L_\tau$ model together with other experiments [110], as shown in Fig. 7.2. The CMS Collaboration has directly searched for the Z' boson in the mass region between 5 to 70 GeV with the 4μ final state using 77.3 fb^{-1} of data [113], and set upper limits on the Z' to muon coupling strength, g, of 0.004 – 0.3 at 95% confidence level, depending on the Z' mass. This search uses the full Run 2 dataset (139 fb^{-1}) and the MVA technique to improve the search sensitivity.

7.1 Dataset and Monte Carlo simulations

The data used for this analysis were recorded using single-muon and multi-muon triggers, corresponding to an integrated luminosity of 139 fb^{-1} after the application of data quality requirements [114]. The transverse momentum (p_{T}) thresholds for the single-muon trigger vary from 20 to 26 GeV, for the di-muon trigger from 10 to 14 GeV, and for the tri-muon trigger from 4 to 6 GeV, depending on the data-taking periods [115]. The overall trigger efficiency for the 4μ events selected in this analysis is higher than 98%.

7.1.1 Simulation of Z' production

The $L_\mu - L_\tau$ model is used for Monte Carlo (MC) Z' signal sample production, where the Z' couples to the left-handed (LH) muon or tau leptons and their corresponding neutrinos, and to the right-handed (RH) muon or tau leptons. In the model, the Z' couplings to the first lepton families (electron and its neutrino) and all quarks are set to zero. The couplings to second and third-generation leptons are assumed to be identical, therefore the branching fractions of the Z' decay to a pair of muons and a pair of muon neutrinos are set to $\frac{1}{3}$ and $\frac{1}{6}$, respectively. The signal from tau decays in the 4μ final state is found to be negligible in this analysis and is not included as the signal.

The signal events are generated with MadGraph5_aMC@NLO 2.7.3 [116] at leading-order (LO) accuracy in QCD by using the *Universal FeynRules Output* (UFO) format [117, 118]. The interactions [119] mediated by a resonance Z' which couples to the second and third-generation leptons are used for four-muon signal event generation. Based on theoretical calculations [120,121] for related processes, the appropriate NNLO/LO correction factor K of 1.3 is used to correct the MC LO signal cross-sections. In the signal production simulations, the NNPDF2.3nlo set [105] is used as the parton distribution function (PDF) for the pp collisions. Photos++ 3.61 [122] is used as the photon emissions from electroweak vertices and charged leptons via a QED process.

The MC simulated events are generated for a range of masses and coupling parameters of the $L_\mu - L_\tau$ model. The Z' mass ranges from 5 to 81 GeV, and the value of coupling constant g ranges from 0.008 to 0.316, as summarized in Table 7.1. The value of the g at each Z' mass was chosen to allow a sensitive search for a very small Z' signal in the full Run 2 dataset.

With the chosen g, the natural width Γ of the Z' (smaller than the experimental resolution) and the cross-section, σ ($pp \to Z'\mu^+\mu^- \to \mu^+\mu^-\mu^+\mu^-$), are calculated as listed in Table 7.1.

The Z' signal samples were simulated with the ATLAS fast simulation framework (Atlfast-II) [107] to produce predictions that can be directly compared with the data.

7.1.2 Simulation of background events

Dominant SM backgrounds in this analysis come from the SM $Z \to 4\mu$ processes where the four leptons have an invariant mass close to that of the Z boson. In the higher mass region the ZZ^* production contribute a sizable number of prompt 4μ events. In addition, there are very small contributions from the Higgs boson, $t\bar{t}V$ ($V = W, Z$), and tri-boson (VVV) production processes. These events are estimated with MC simulations. The non-

Table 7.1: Summary of the chosen Z' hypotheses and corresponding coupling, width, and cross-section (calculated at LO accuracy in QCD) at each mass point.

$m_{Z'}$ [GeV]	g	Γ [GeV]	σ [fb]	$m_{Z'}$ [GeV]	g	Γ [GeV]	σ [fb]
5	0.0080	2.45E-5	9.96	42	0.0900	2.71E-2	13.38
7	0.0085	3.99E-5	7.06	45	0.1000	3.58E-2	11.72
9	0.0090	5.78E-5	5.60	48	0.1100	4.62E-2	9.96
11	0.0095	7.89E-5	4.65	51	0.1200	5.84E-2	8.24
13	0.0100	1.03E-4	3.95	54	0.1600	1.10E-1	10.07
15	0.0120	1.72E-4	4.45	57	0.2000	1.81E-1	10.73
17	0.0140	2.65E-4	4.80	60	0.2665	3.39E-1	12.92
19	0.0160	3.87E-4	5.00	63	0.2680	3.60E-1	8.84
23	0.0240	1.05E-4	7.30	66	0.2780	4.06E-1	6.50
27	0.0320	2.20E-3	8.50	69	0.2890	4.59E-1	4.89
31	0.0400	3.95E-3	8.72	72	0.3000	5.15E-1	3.80
35	0.0600	1.00E-2	12.82	75	0.3000	5.37E-1	2.88
39	0.0800	1.99E-2	14.77	78	0.3080	5.89E-1	2.40
				81	0.3160	6.44E-1	2.08

prompt muon background events, mostly coming from Z + jets, $t\bar{t}$, and single-top-quark production processes, are estimated from data in this analysis, as described in section 7.3.1. The Z + jets and $t\bar{t}$ MC samples are also produced for background studies.

Samples of diboson final states ($q\bar{q} \to VV^{(*)}$), including the processes shown in Figures 7.1b and 7.1c, were simulated with the Sherpa [2.2.2] [123] generator, including off-shell effects and Higgs boson contributions, where appropriate. Fully leptonic final states and semileptonic final states, where one boson decays leptonically and the other hadronically, were generated using matrix elements at next-to-leading-order (NLO) accuracy in QCD for up to one additional parton and at LO accuracy for up to three additional parton emissions. Samples for the loop-induced processes $gg \to VV^{(*)}$, shown in Figure 7.1d, were generated using LO-accurate matrix elements for up to one additional parton emission for both the cases of fully leptonic and semileptonic final states. The matrix element calculations were matched and merged with the Sherpa parton shower based on Catani–Seymour dipole factorisation [102, 124] using the MEPS@NLO prescription [125–128]. The virtual QCD corrections were provided by the OpenLoops library [103,129,130]. The NNPDF3.0nnlo set of PDFs was used [95], along with the dedicated set of parton-shower parameters (tune) developed by the Sherpa authors.

The production of $t\bar{t}$ events was modelled using the Powheg-Box [v2] [131–134] gener-

ator at NLO in QCD with the NNPDF3.0NLO PDF set and the h_{damp} parameter[1] set to 1.5 [135]. The events were interfaced to PYTHIA [8.230] [136] to model the parton shower, hadronization, and underlying event, with parameters set according to the A14 tune [106] and using the NNPDF2.3LO set of PDFs. The associated production of top quarks with W bosons (tW) was modeled by the POWHEG-BOX [v2] [132–134, 137] generator at NLO in QCD using the five-flavor scheme and the NNPDF3.0NLO set of PDFs. The diagram removal scheme [138] was used to remove interference and overlap with $t\bar{t}$ production. The events were interfaced to PYTHIA [8.230] using the A14 tune and the NNPDF2.3LO set of PDFs.

The production of V+jets was simulated with the SHERPA [2.2.1] generator using NLO matrix elements for up to two partons, and LO matrix elements for up to four partons, calculated with the Comix and OPENLOOPS libraries. They were matched with the SHERPA parton shower using the MEPS@NLO prescription with the set of tuned parameters developed by the SHERPA authors. The NNPDF3.0NNLO set of PDFs was used and the samples were normalized to a next-to-next-to-leading-order (NNLO) prediction [139].

The production of events was modeled using the MADGRAPH5_aMC@NLO 2.3.3 generator at NLO with the NNPDF3.0NLO PDF. The events were interfaced to PYTHIA [8.210] using the A14 tune and the NNPDF2.3LO PDF set. The production of tri-boson (VVV) events was simulated with the SHERPA [2.2.2] generator. Matrix elements accurate to LO in QCD for up to one additional parton emission were matched and merged with the SHERPA parton shower based on Catani–Seymour dipole factorization using the MEPS@NLO prescription. Samples were generated using the NNPDF3.0NNLO PDF set, along with the dedicated set of tuned parton-shower parameters developed by the SHERPA authors.

The generated background MC samples were produced through the full ATLAS detector simulation based on GEANT4 [108]. The effect of multiple interactions in the same and neighboring bunch crossings (pileup) was modeled by overlaying the simulated hard-scattering event with inelastic pp events generated with PYTHIA [8.186] [104] using the NNPDF2.3LO set of PDF and the A3 set of tuned parameters [140]. Simulated events were reweighted to match the pile-up conditions in the data. All simulated events were processed using the same reconstruction algorithms and triggering requirements as used in data.

[1]The h_{damp} parameter is a resummation damping factor and one of the parameters that controls the matching of POWHEG matrix elements to the parton shower and thus effectively regulates the high-p_{T} radiation against which the $t\bar{t}$ system recoils.

7.2 Event selection

Proton-proton interaction vertices are reconstructed in events with at least two tracks, each with $p_{\mathrm{T}} > 0.5$ GeV. The primary hard-scatter vertex is defined as the one with the largest value of the sum of squared track transverse momenta.

Table 7.2 summarizes the selection cuts of muons in this analysis. Since this analysis includes the signal sample with $m_{Z'} = 5$ GeV, we lower the p_{T} cut on muon from 7 GeV to 3 GeV, which improves signal selection efficiency by 10%.

Identification	Loose
Kinematic cuts	$p_{\mathrm{T}} > 3$ GeV, $\lvert\eta\rvert < 2.7$
Impact cuts	$\lvert d_0^{\mathrm{BL}}/d_{error}^{\mathrm{BL}}\rvert < 3.0$ $\lvert z_0^{\mathrm{BL}} \cdot \sin(\theta)\rvert < 0.5$ mm
Isolation	FixedCutLoose

Table 7.2: Muon object definition.

Muons are identified by matching tracks reconstructed in the MS to tracks reconstructed in the ID (referred to as combined muons). To increase the muon reconstruction efficiency non-combined muon identification algorithms are also used in the analysis, including using the MS stand-alone tracks in the region $2.5 < |\eta| < 2.7$, and matching the ID tracks with calorimeter hit information within $|\eta| <0.1$, as well as using the ID tracks associated with at least one local track segment in the MS. In the 4μ event selection at most one of the selected muons can be a non-combined muon. Each muon is then required to satisfy the 'loose' identification criteria [141]. Muons are required to be isolated using a particle-flow algorithm [142] and associated with the primary hard-scatter vertex by satisfying $|\frac{d_0}{\sigma_{d_0}}| < 3$ and $|z_0 \times \sin\theta| < 0.5$ mm, where d_0 is the transverse impact parameter calculated with respect to the measured beam-line position, σ_{d_0} its uncertainty, and z_0 is the longitudinal distance between the point at which d_0 is measured and the primary vertex. The minimum muon p_{T} threshold is 3 GeV.

In addition to muons, electrons, jets, and missing transverse momentum ($E_{\mathrm{T}}^{\mathrm{miss}}$) are also used to select control samples for background estimation in this analysis. The reconstructions of these objects are discussed in chapter 4.

Each electron is required to satisfy the 'medium' likelihood identification criteria [143],

as well as similar vertex and isolation requirements as muons. The reconstructed electrons are required to have $p_{\mathrm{T}} > 7$ GeV and $|\eta| < 2.5$, excluding the transition region between the barrel and endcap calorimeters, $1.37 < |\eta| < 1.52$.

Jets are reconstructed with the anti-k_t algorithm introduced in section 4.3.1 with a radius parameter of $R = 0.4$. The jet clustering input objects are based on particle-flow [144] in the ID and the calorimeter. Jets are required to have $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.5$. To reduce the effect of pile-up an additional quality requirement based on the JVT [72] is applied in jet identification, which is introduced in section 4.3.3.

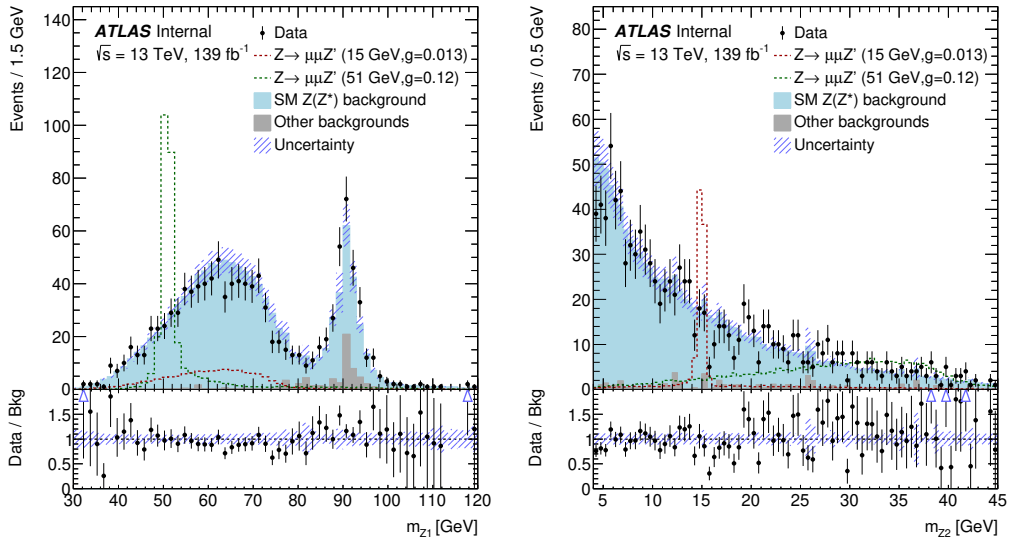
Events containing at least four muons with kinematics consistent with $Z(Z^*/\gamma^*) \to 4\mu$ production are then selected as follows. The four leading p_{T}-ordered muons are required to pass the p_{T} thresholds of 20, 15, 8, and 3 GeV, respectively. If a muon is selected as a non-combined muon, its p_{T} must be greater than 15 GeV. Any di-muon pair in the event must have an invariant mass $m_{\mu\mu}$ greater than 4 GeV and an angular separation ΔR larger than 0.2. To search for the $Z' \to \mu^+\mu^-$ signature, two muon pairs are selected based on their invariant mass values. The first pair (referred to as Z_1) is selected from all the possible $\mu^+\mu^-$ pairs to have the smallest mass difference between the Z_1 mass and the Z mass, $|m_Z - m_{Z_1}|$. The second $\mu^+\mu^-$ pair is selected from the remaining muons that have the highest invariant mass (referred to as Z_2). The correct signal di-muon pairing fraction varies with the Z' mass, where the selected di-muon pair that forms m_{Z1} or m_{Z2} originates from the Z'. For example, for $m_{Z'} = 5$, 42, 63, 72, and 81 GeV, the correct di-muon pairing fractions are about 78%, 50%, 88%, 82%, and 90%, respectively. Finally, the selected four muons must have an invariant mass in a range of 80 to 180 GeV, excluding the Higgs boson resonance mass region of 110 to 130 GeV.

The Z' signal efficiency at various stages of the event selection is shown in Table 7.3 for five representative mass points. The event selection efficiencies vary significantly depending on the Z' mass. At the generator level, an MC filter is applied, which requires at least four muons with $p_T > 2$ GeV and $|\eta| < 3.0$. The MC filter efficiencies of these representative Z' signal samples are listed in the table as well.

The Z' production signature is searched for in the Z_1 or Z_2 mass spectrum depending on the Z' mass. The relatively high-mass Z' signals mostly appear as a peak in the Z_1 spectrum while the relatively low-mass signals mostly appear as a peak in the Z_2 spectrum. Representative examples of the predicted signal over background, after further selection with a deep learning approach which will be described in Section 7.6, are shown in Figure 7.26. In this analysis, the Z_1 and Z_2 mass spectra are scanned to search for a Z' with a mass greater or smaller than 42 GeV, respectively. The boundary value of 42 GeV is chosen based on the studies to optimize the search sensitivity.

Table 7.3: The Z' signal event selection efficiencies compared to the events passing the previous cut level for several representative mass points. The overall signal efficiencies are the products of the 4μ MC filter and the combined event selection efficiencies.

$m_{Z'}$ [GeV]	5	42	63	72	81
MC filter efficiency	32.8%	57.7%	61.0%	65.3%	70.0%
Number of identified muons ≥ 4	47.3%	74.1%	70.8%	72.4%	75.4%
$p_{\mathrm{T}}^{i}(i = 1, 4) > 20,\ 15,\ 8,\ 3$ GeV	60.0%	82.6%,	90.3%	93.6%	98.2%
$\Delta R(\mu_i, \mu_j) > 0.2$ & vertex requirement	87.2%	95.4%	96.2%	96.6%	97.2%
Isolation	54.2%	76.9%	79.2%	84.1%	87.5%
$m_{4\mu}$ within [80, 110] or [130, 180] GeV	91.9%	88.8%	58.9%	33.5%	16.8%
Combined event selection efficiency	12.3%	39.9%	28.7%	18.4%	10.6%
Overall 4μ signal efficiency	4.1%	23.0%	17.5%	11.9%	7.4%



(a) (b)

Figure 7.3: Distributions of m_{Z1} (a) and m_{Z2} (b) with the pre-selected 4μ events. MC simulated signals and background are normalized by their production cross-sections and 139 fb^{-1} integrated luminosity together with the overall selection efficiencies.

The numbers of 4μ events in data and the estimated background yields are given in Table 7.4. More details about the estimation of the reducible backgrounds containing non-prompt muons can be found in Section 7.3. The total uncertainties of simulated backgrounds are also listed in the table. The evaluations of systematic uncertainties will be described in Section 7.5.

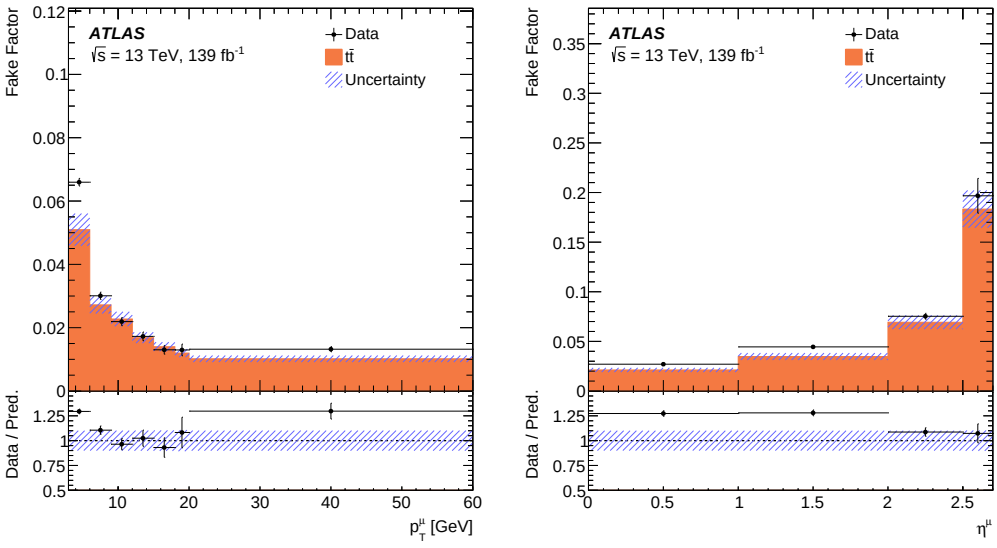
Table 7.4: The selected 4μ events in data and the estimated backgrounds and their combined statistical and systematic uncertainties.

Data	Total background	$qq \to ZZ^*$ from MC	$gg \to ZZ^*$ from MC	$ttV + VVV + H$ from MC	Reducible background from data
1131	1148 ± 70	1065^{+70}_{-69}	15.6 ± 2.5	6.2 ± 2.9	$61.1^{+8.3}_{-9.1}$

7.3 Background estimation

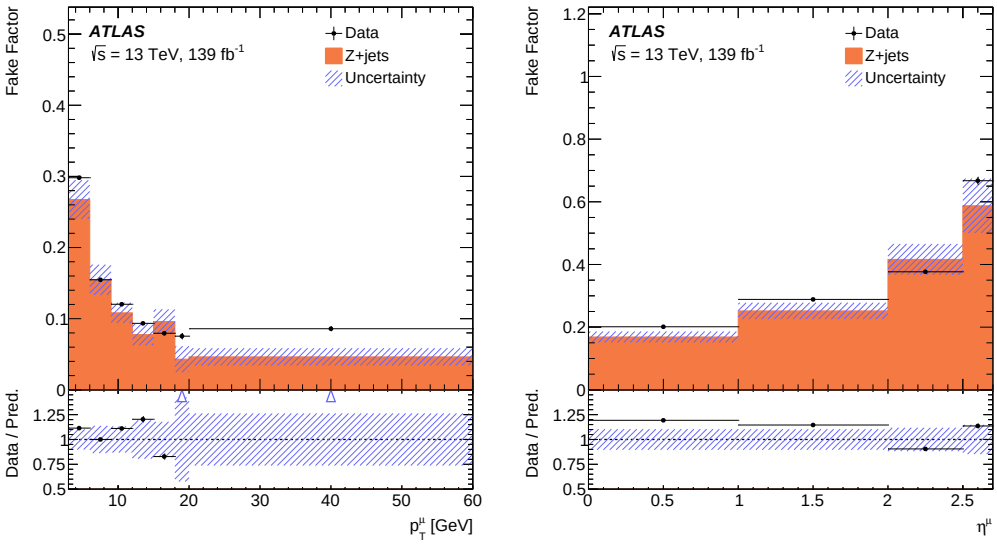
7.3.1 Fake background estimation

The reducible backgrounds from Z+jets and $t\bar{t}$ are estimated using a data-driven technique, the fake factors method described in section 6.3.3. Since both analyses require four-lepton final states and sources of reducible backgrounds are the same, in this analysis, we used similar definitions of the Z+jets and $t\bar{t}$ control regions to estimate the fake factors, which can be found in table 6.8 (Z+jets) and table 6.9 ($t\bar{t}$). Only fake factors of muons are calculated since this analysis only includes the 4μ channel. The same as 6.3, bad muons are defined as objects that pass basic muon selections but fail the isolation and d_0 significance cuts, and good muons are defined as objects that pass all SR muon cuts. The lowest p_T of the fake factor function and p_T cuts on bad and good muons are lower from 7 GeV to 3 GeV following the signal region cuts. The 1-D fake factor as a function of muon p_T and η are shown in figure 7.4 ($t\bar{t}$) and figure 7.5 (Z+jets).



(a) (b)

Figure 7.4: Fake factors as a function of muon p_T (a) and $|\eta|$ (b) derived using the $t\bar{t}$ control sample, which is selected by isolated high-p_{T} $e^{\pm}\mu^{\mp}$ pair associated with at least one b-jet and $E_T^{miss} > 50$ GeV in each event.

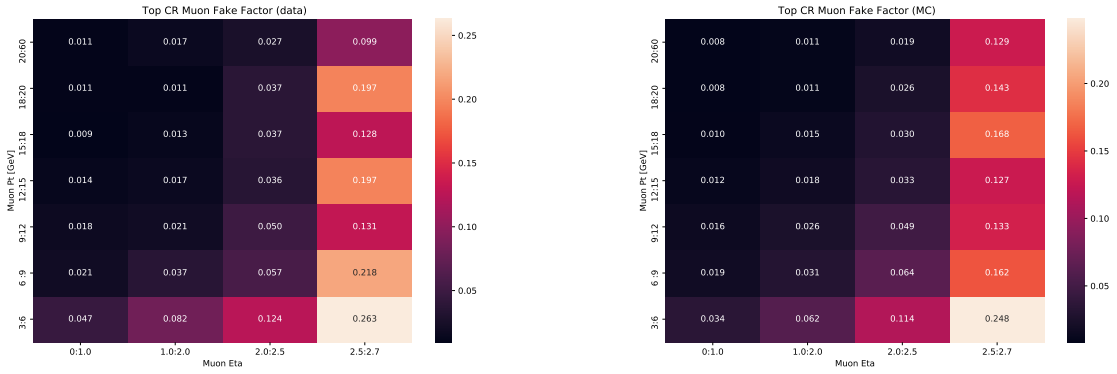


(a) (b)

Figure 7.5: Fake factors as a function of muon p_T (a) and $|\eta|$ (b) derived using the Z+jets control sample, which is selected by requiring an isolated high-p_{T} di-lepton pair ($e^+e^-, or\ \mu^+\mu^-$) decay from the Z boson (with their inv. mass within 10 GeV of the Z mass) and $E_T^{miss} < 25$ GeV in each event.

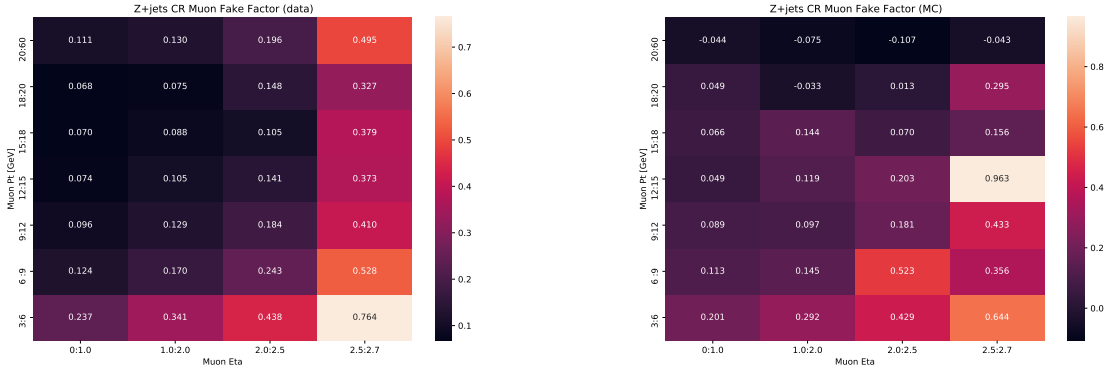
In the estimation of fake backgrounds, the 2-D fake factors as functions of both p_T and

η are used for better estimation. In plots of the 2-D fake factors, fake factors derived from data and Mento Carlo are put together for comparison.



(a) (b)

Figure 7.6: The 2-Dim fake factors as functions of muon p_T and η in the $t\bar{t}$ CR from data (left) and Monte-Carlo (right). We used 2D fake factor to calculate the final fake background yield.



(a) (b)

Figure 7.7: 2-Dim fake factor as functions of muon p_T and η in the Z+jets CR from data (left) and Monte-Carlo (right). We used the 2D fake factor to calculate the final fake background yield.

Then, the fake factors derived from two data control samples are combined using the m_{Z1} spectra obtained from simulated 4μ background events from Z+jets and $t\bar{t}$ processes.

A simultaneous fit of the m_{Z1} spectra of the MC events to data selected in the $\mu^+\mu^-\mu_j\mu_j$ control sample is performed to determine the fractions of the Z + jets and $t\bar{t}$ events in each

bin of the mass spectrum. The post-fit distribution of m_{Z1} in the control region is shown in figure 7.8.

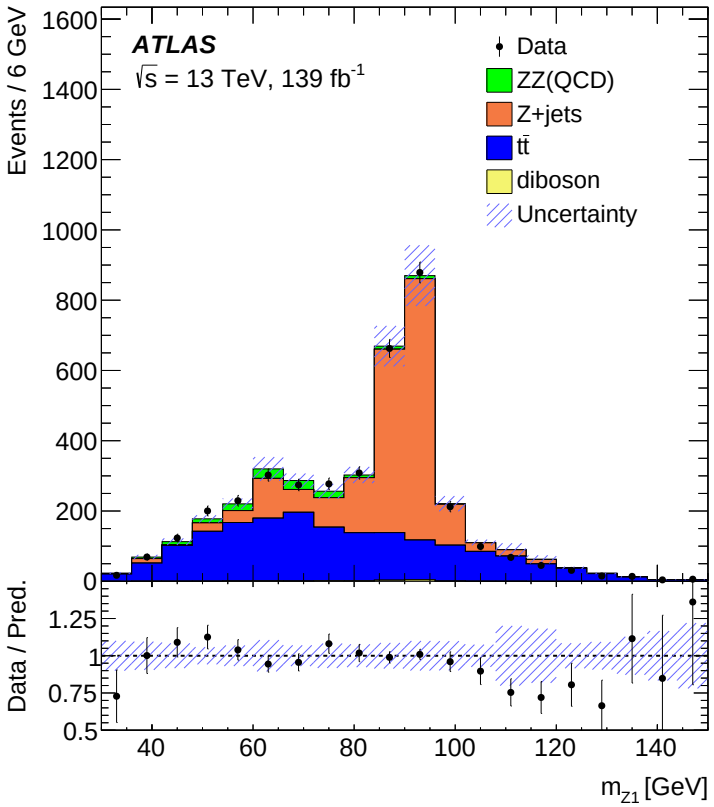


Figure 7.8: Post-fit data and MC comparison of the m_{Z1} mass spectrum in the background control region.

The overall systematic uncertainty of the fake factor is about 14%. It is determined with alternative non-prompt muon selections, such as changing the muon isolation criteria in data control samples (7.6%), and the uncertainties of the Z + jets and $t\bar{t}$ event fractions when combining the fake-factors derived from two control samples (8.6%). The statistical uncertainties of the control samples are also accounted as part of the systematic uncertainties (6.1%).

7.3.2 Estimation of reducible backgrounds

Additional reducible background contributions come from the WZ process. These events contain three prompt muons from W and Z decays and one non-prompt muon. This background is estimated by scaling a $3\mu + \mu_j$ control sample by the f derived from the Z + jets sample. The total estimated number of reducible background events, $61.1^{+8.3}_{-9.1}$, is listed in

Table 7.4.

7.4 Data and MC comparison in the four-electron control region

To better validate the physics modeling of the SM ZZ process in this low mass region, we use four-electron events from $ZZ^* \to 4e$ to compare data with MC simulations. The four-electron event selection and pairing method are the same as for the four-muon event selection. The electron selection criteria are shown in Table 7.5. The same fake-factor method is used to estimate the fake electron contributions to the four-electron control region. The fake background is included in the data and MC simulation comparison plots.

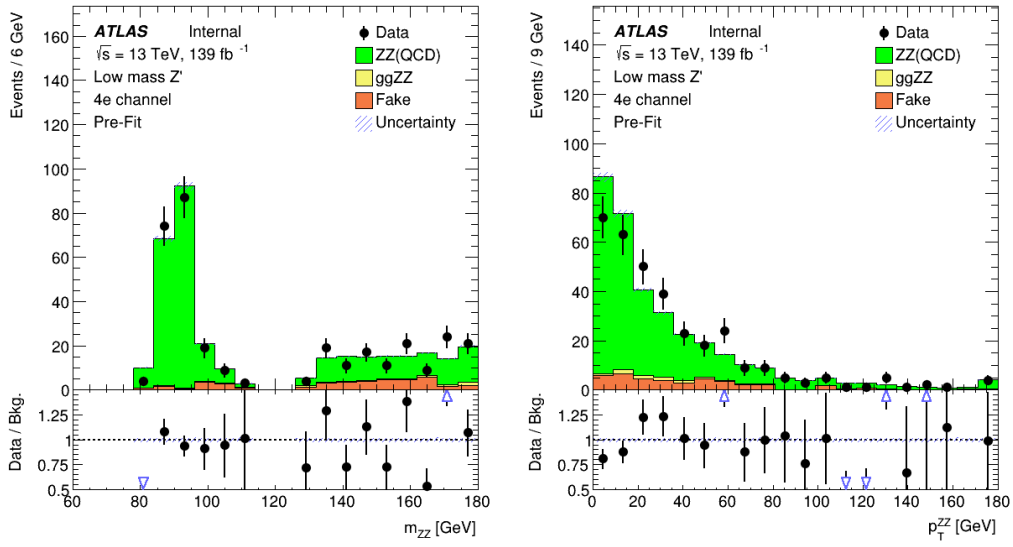
	Electron selection
Identification	Likelihood Loose
Kinematic cuts	$p_{\mathrm{T}} > 7$ GeV, $\lvert\eta\rvert < 2.47$
Impact cuts	$\lvert d_0^{\mathrm{BL}}/d_{error}^{\mathrm{BL}}\rvert < 5.0$ and $\lvert z_0^{\mathrm{BL}} \cdot \sin(\theta)\rvert < 0.5$ mm
Isolation	FixedCutLoose

Table 7.5: Electron selection criteria for $ZZ^* \to 4e$ control region event selection.

Since the fake factor of the electron is higher than the muon fake factor, the fake electron background is more visible in the four-electron plots shown in Figure 7.9, Figure 7.10, Figure 7.11, and Figure 7.12.

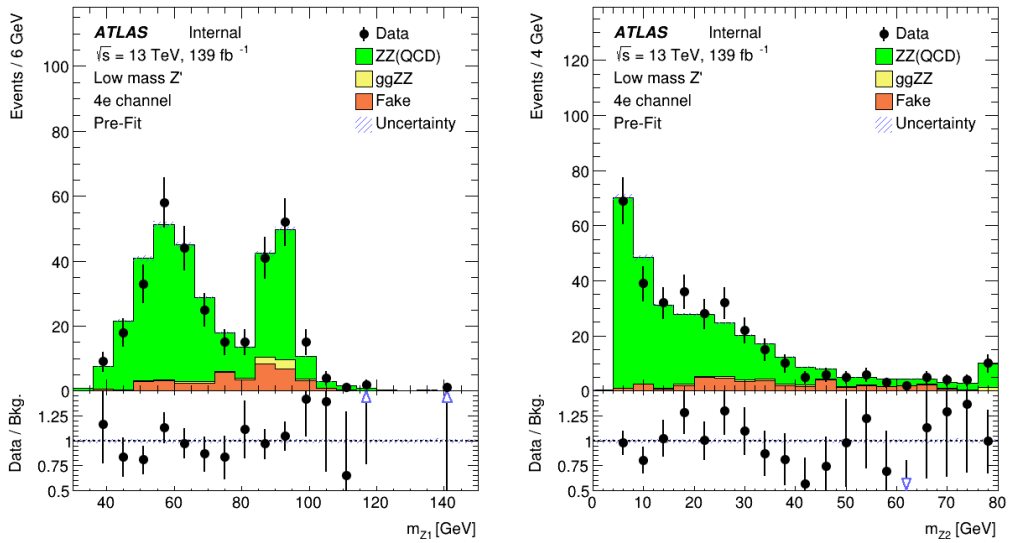
Individual electron p_T and η distributions in the selected $4e$ events are shown in Fig. 7.13, Fig. 7.14, Fig. 7.15, and Fig. 7.16.

In general, data and MC agree well in the selected $4e$ sample, indicating MC ZZ background modeling is good.



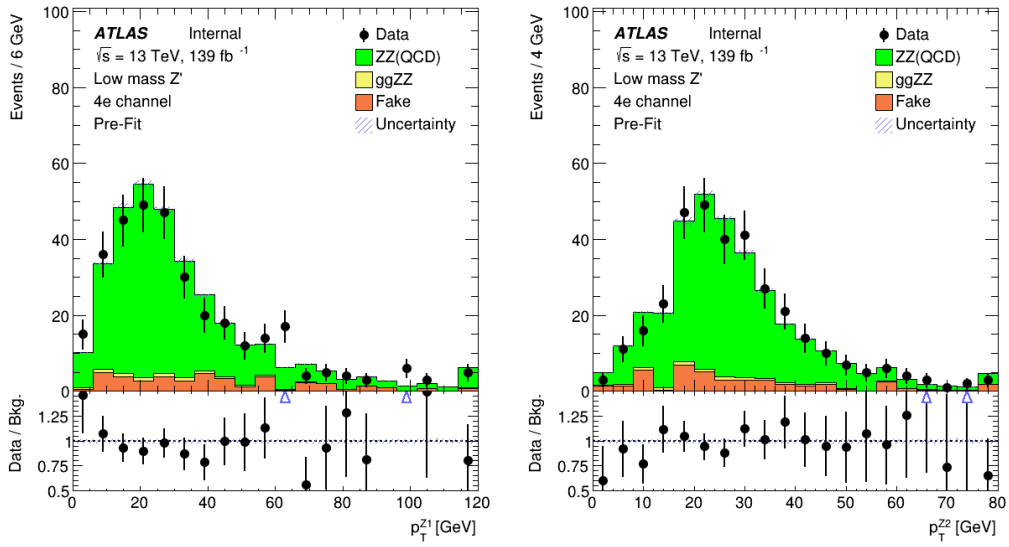
(a) (b)

Figure 7.9: Four lepton mass and p_T distribution in 4 electron channel.



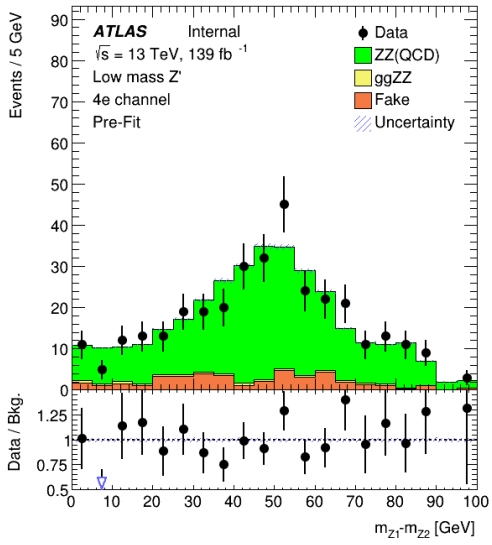
(a) (b)

Figure 7.10: Leading and sub-leading lepton pair mass distribution in 4 electron channels.



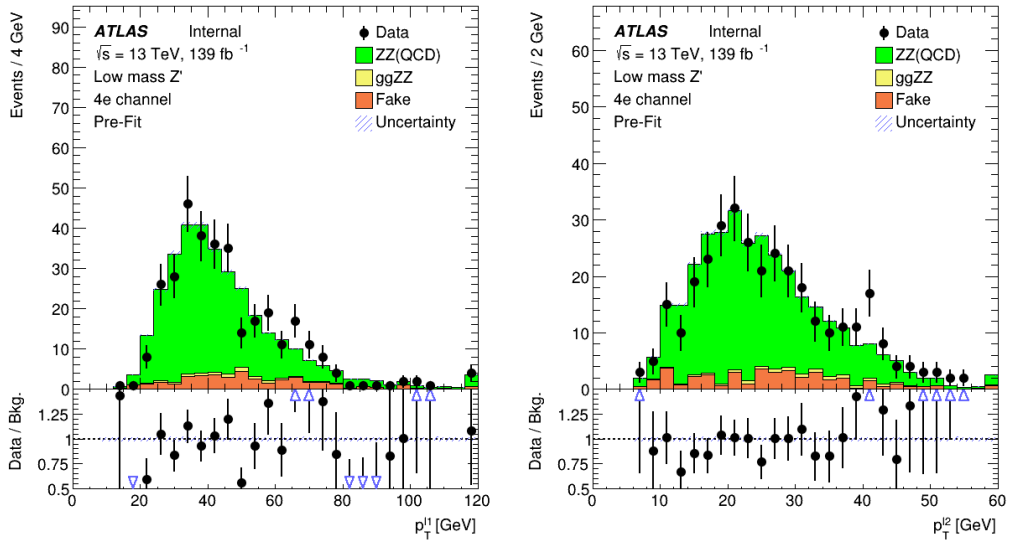
(a) (b)

Figure 7.11: Leading and sub-leading lepton pair p_T distribution in 4 electron channel.



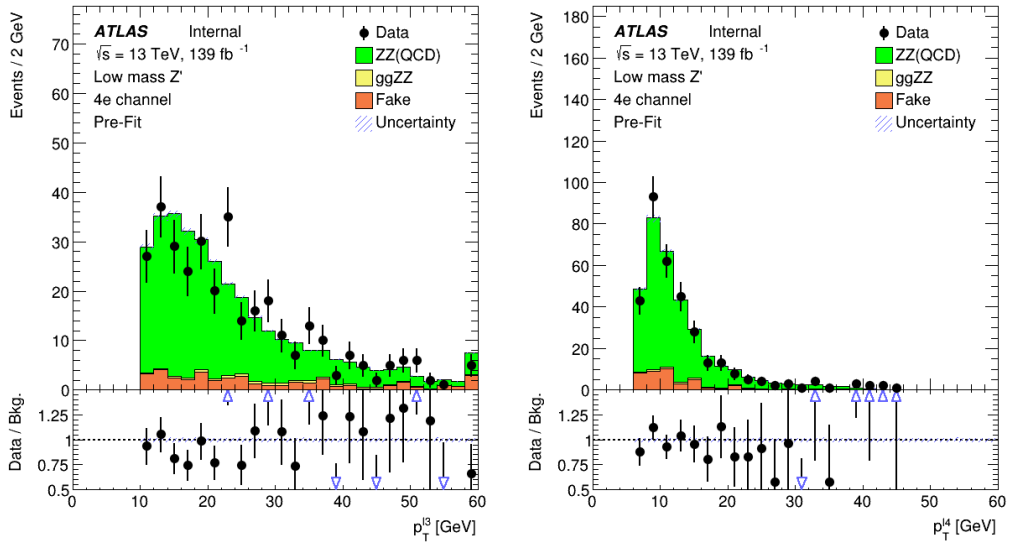
(a)

Figure 7.12: The mass difference between leading and sub-leading lepton pair in 4 electron channel.



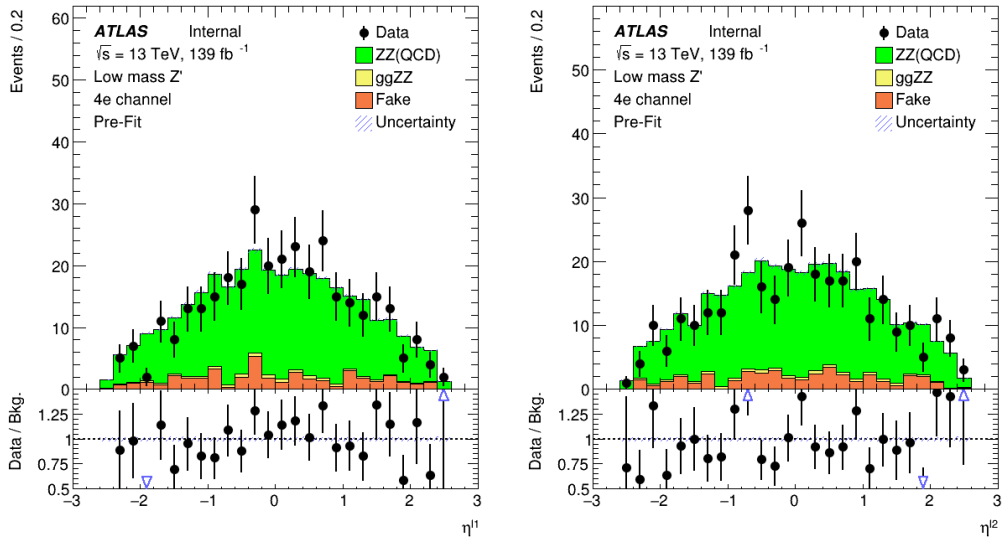
(a) (b)

Figure 7.13: p_T distribution of leading and sub-leading lepton in the leading lepton pair in 4 electron channel.



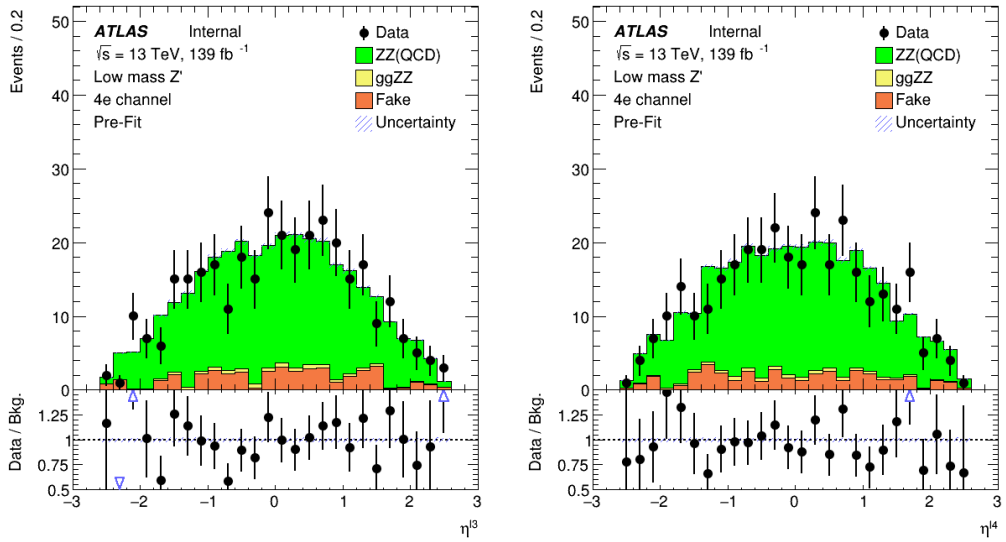
(a) (b)

Figure 7.14: p_T distribution of leading and sub-leading lepton in the sub-leading lepton pair in 4 electron channel.



(a) (b)

Figure 7.15: η distribution of leading and sub-leading lepton in the leading lepton pair in 4 electron channel.



(a) (b)

Figure 7.16: η distribution of leading and sub-leading lepton in the sub-leading lepton pair in 4 electron channel.

7.5 Systematic uncertainties

Systematic uncertainties in the simulated event yields and shapes, for both signal and background processes, may arise from the calibration of the physics objects and from the theoretical modeling used in the predictions.

The major experimental uncertainties come from the muon reconstruction, identification, and isolation requirement efficiencies. These efficiencies are corrected based on studies performed in data control regions. The energy and momentum scales and resolutions of the simulated objects are calibrated to reproduce data from $Z \to \mu^+\mu^-$ and $J/\psi \to \mu^+\mu^-$ decays [141]. The uncertainties on the 4μ detection efficiency are determined by varying the nominal calibrations in the MC samples by one standard deviation, including muon momentum resolutions and scales, and the trigger, reconstruction, identification, and isolation requirement efficiencies. The overall relative experimental uncertainties in the 4μ event selection efficiency are about 3.9%, dominated by the uncertainty of the isolation efficiency (2.9%) and the low p_{T} calibration uncertainty (2.0%). The signal event selection efficiency uncertainties vary from 8.3% to 3.9%, depending on the Z' mass. In addition, the uncertainty in the combined 2015-2018 integrated luminosity is 1.7% [55], obtained using the LUCID-2 detector [145] for the primary luminosity measurements.

Sources of theoretical uncertainties come from the choice of QCD scales (renormalization μ_R and factorization μ_F), strong coupling constant α_S, and PDFs, as well as the parton shower models. These uncertainties affect the signal and background event selection efficiencies, normalization, and the shape of their kinematic distributions. The scales are varied independently from 0.5 to 2.0 times the nominal values and the largest deviation is chosen as the systematic uncertainty. The PDF uncertainty is estimated by comparing events generated with different PDF sets, as well as the uncertainties from the nominal PDF set itself. The maximal variation (envelope) is accounted for as the systematic uncertainty, following the PDF4LHC [109] recommendations. The α_S uncertainty is estimated by comparing events generated with different α_S values using the nominal PDF set. The parton shower uncertainty is estimated by comparing events with different parton shower parameters in the Sherpa MC samples. For the $Z(Z^*) \to 4\mu$ process, the relative uncertainties of event yields for scale, PDF, α_s, and parton shower, are 4.6%, 1.8%, 1.0%, and 1.9%, respectively. The $t\bar{t}V$, VVV, and Higgs boson, processes contribute 5.2, 0.50, and 0.53 background events, respectively. The total relative theoretical uncertainty for these backgrounds is estimated at about 16%. The interference effect between the signal and background is estimated (using MC samples generated by MadGraph5_aMC@NLO) by evaluating the cross-section ratio, $\Delta\sigma/\sigma_{Z'}$, where $\Delta\sigma$ is the difference of the cross-section from the inclusive MC sample

and the sum of the cross-sections of the signal and background samples in the Z' detection phase space. The effect varies from 1.8 to 7.5% for the Z' mass from 5 - 81 GeV, which is accounted for as additional signal yield systematic uncertainties. In addition, the MC Z' signal filter acceptance uncertainty is estimated to be about 2%, which is calculated by varying the QCD scales, the PDF-sets, and the strong coupling constant using MC events at the generator level for different Z' mass points.

All the uncertainties (from both experiment and theory) on the final discriminant di-muon mass spectra are included as nuisance parameters in the signal extraction fitting process described in Section 7.7.1.

7.6 Event classification with deep learning approach

7.6.1 Deep learning model setup

The selected 4μ events are classified with a "Deep Learning" approach to further separate the Z' signal from the SM background. A parametrized deep neural network (pDNN) architecture [146], which is introduced in section 5.2, is used in the analysis. This algorithm allows the training of a single classifier for multiple signal mass hypotheses in the search range. In practice, the kinematic inputs together with the Z' mass parameters of signal and background are used for training. The MC generated Z' masses (listed in Table 7.1) are used as the multiple signal mass parameters, while the distribution of the mass parameter for background events is randomly drawn from the same distribution as for the signal class. The algorithm was implemented in the Keras [147] framework with the TensorFlow [148] backend. Two classifiers are trained for low (high) Z' mass searches using MC mass parameters lower (higher) than 40 GeV. During the training of the classifier, the training samples were composed of simulated signal and background 4μ events using the pre-selection described in Section 7.2. A set of kinematic distributions were used for pDNN training input features. They are the p_{T} and η of each muon, the p_{T} of the Z_1 and Z_2, the mass difference of the Z_1 and Z_2, ΔR and $\Delta\eta$ of each muon pair that forms the Z_1 and Z_2, and the p_{T} and mass of the 4μ system. Distributions of input variables, both for representative signals ($m_{Z'} = 15$, and 51 GeV) and background, are shown in figure 7.17, 7.18 and 7.21 to compare the predicted and observed transverse momenta ($p_{\mathrm{T}}^{Z_1}$ and $p_{\mathrm{T}}^{Z_2}$) distributions and the mass difference of the Z_1 and Z_2. In addition to the major background from the SM $Z(Z^*) \to 4\mu$ production, other backgrounds, including 4μ events containing non-prompt muons estimated from data, and from ttV, VVV, and Higgs boson production processes, are included in the plots, which are denoted as "others" in the plots.

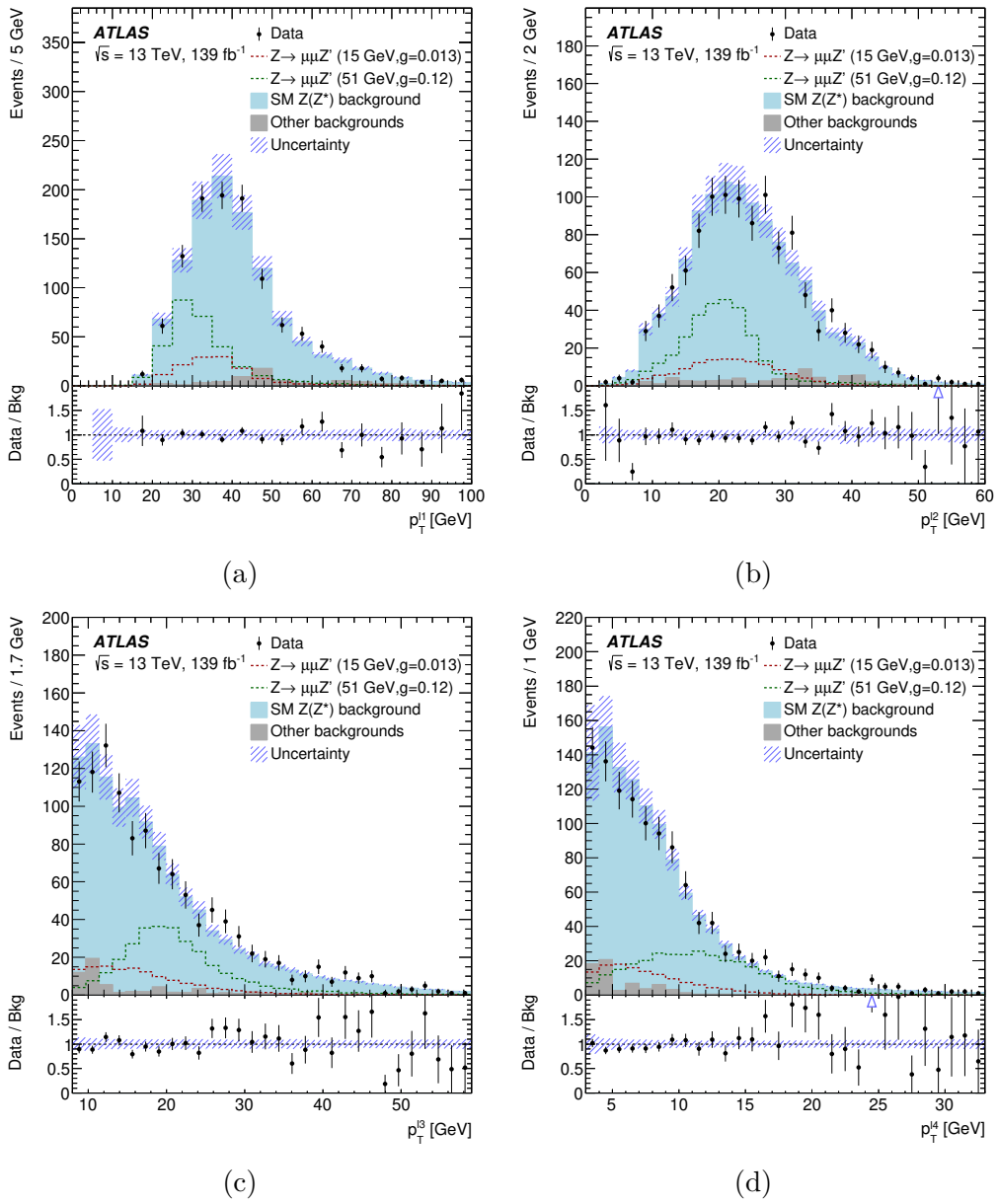


Figure 7.17: Distributions of the p_{T} for each muon (ordered with p_{T} from (a) to (d)).

The training results are a set of weights of the pDNN model which are applied to real data or MC samples to obtain the pDNN output discriminating variable. The m_{Z1} or m_{Z2} are used as the mass parameter when applying the model to data.

The input parameters used in the pDNN training are the kinematic variables with 4μ final state, which are summarized in table 7.6. The "Mass" parameter is used as a parameter input of the pDNN model and the standard kinematic inputs. Relevant kinematic distributions of the selected 4μ events of the Z' signal and the SM background are shown in Figure **??**

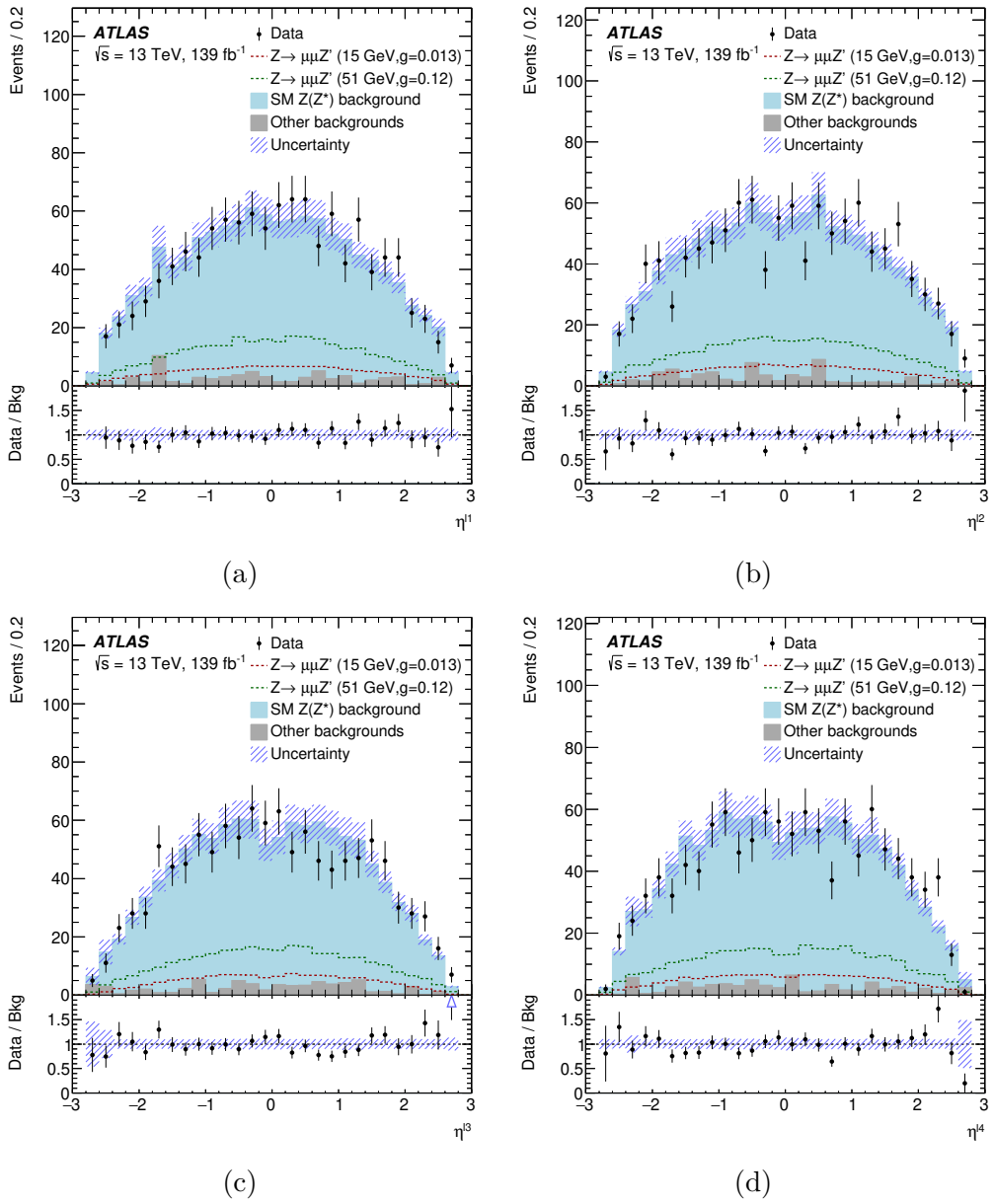


Figure 7.18: Distributions of the *Eta* for each muon (ordered with p_{T} from (a) to (d)).

To cross-check the background model, fully simulated MC events from the $Z(Z^*) \to 4e$ process are produced, and a dedicated $4e$ control sample is selected using the same 4μ event selection criteria and the background estimation technique. The $4e$ data agrees well with the predicted background distributions, indicating good background modeling in this analysis.

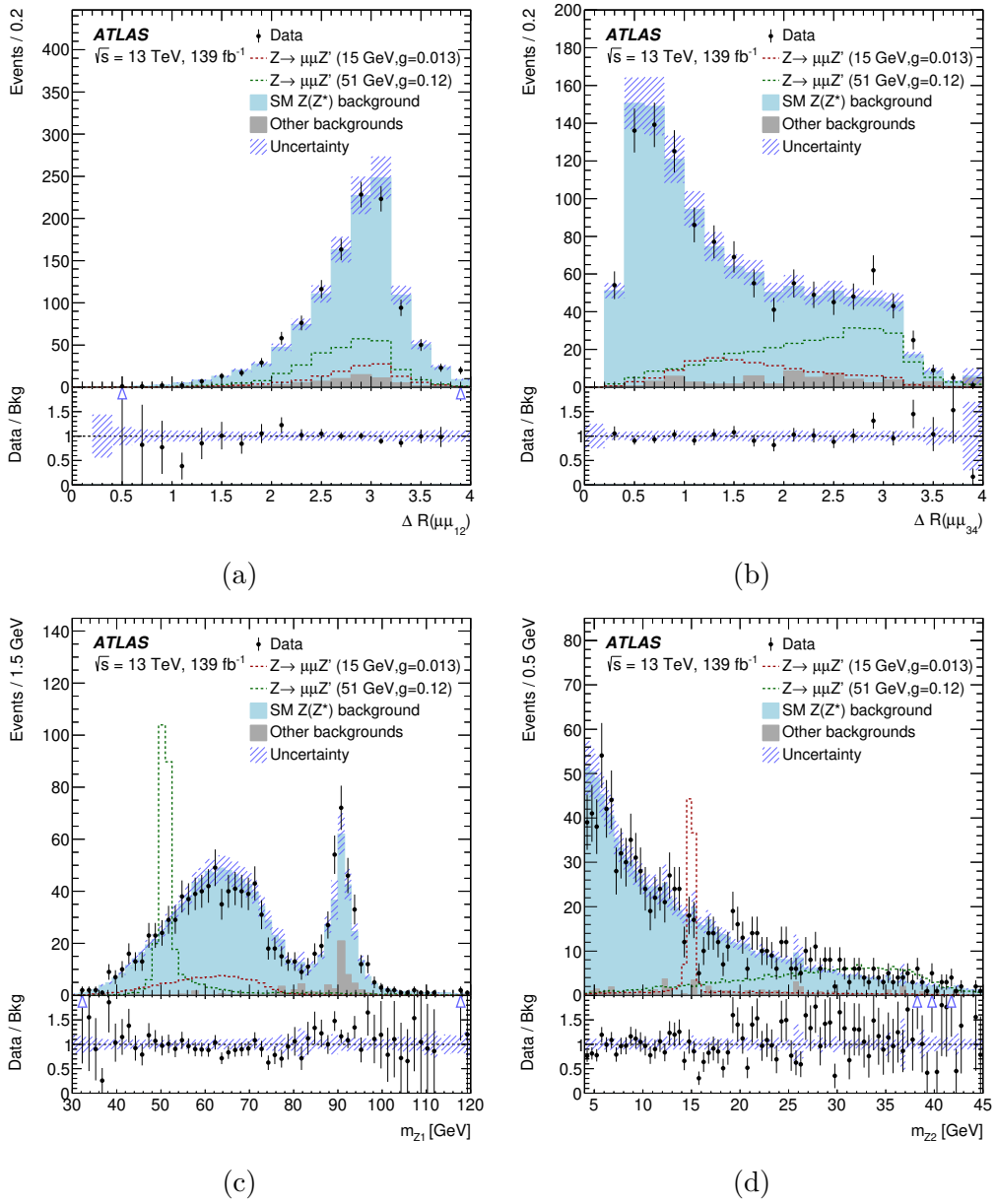
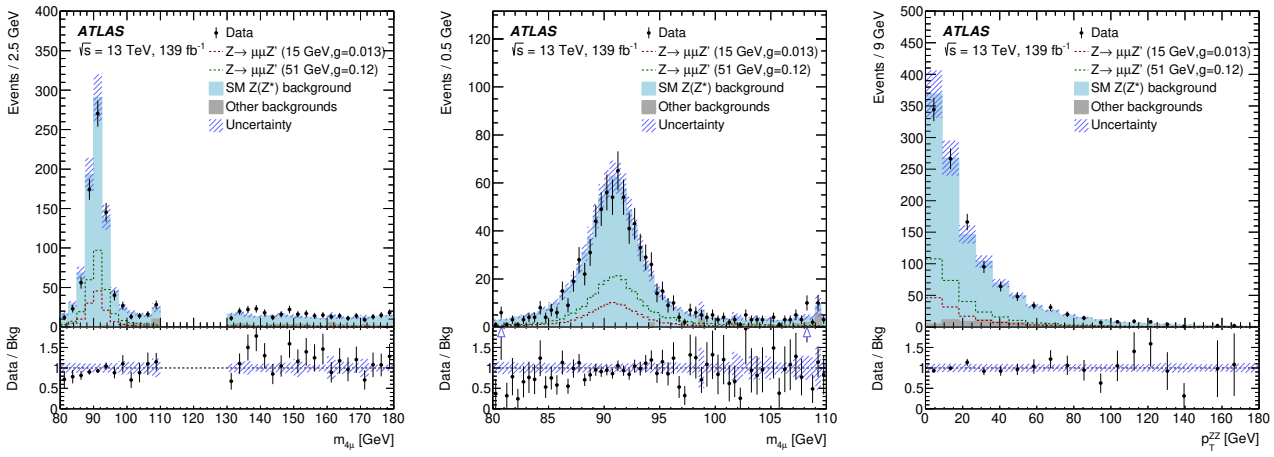
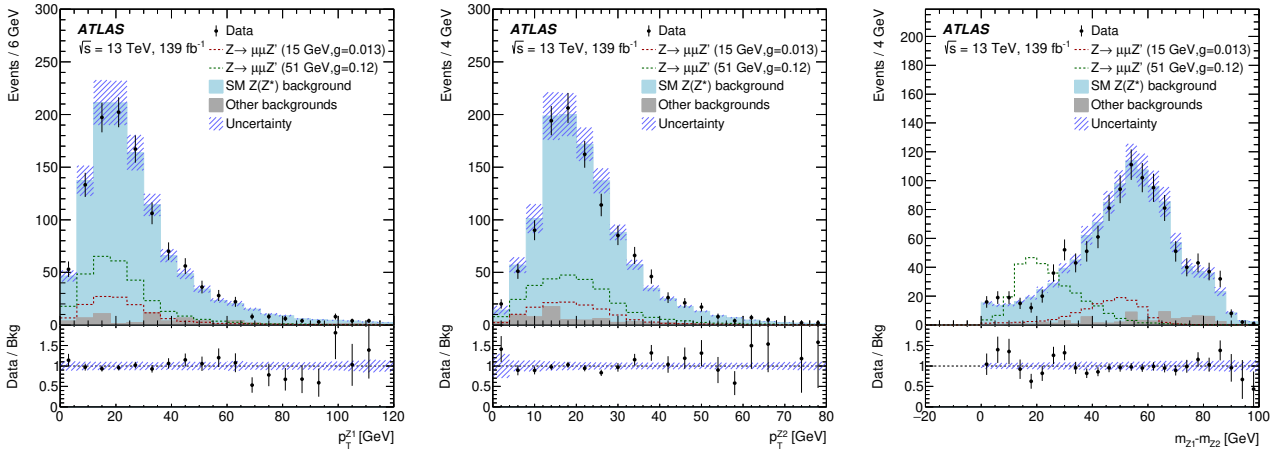


Figure 7.19: Kinematic distributions of the pre-selected 4μ events. Plots (a) to (d) are the angular separations, ΔR of the two muons that formed Z_1 and Z_2, and the mass distributions of Z_1 and Z_2.



(a) (b) (c)

Figure 7.20: Data and MC comparison of the $Z \to 4\mu$ invariant mass using pre-selected 4μ events. Figure (b) shows the distribution between 80 GeV and 110 GeV in Figure (a). Figure (c) shows the distribution of transverse momentum of 4μ system.



(a) (b) (c)

Figure 7.21: Distributions of $p_{\mathrm{T}}^{Z_1}$ (a), and $p_{\mathrm{T}}^{Z_2}$ (b), and the mass difference of the Z_1 and Z_2 candidates (c).

Table 7.6: Description of the variables used as inputs to the pDNN classifier

variable	description
Mass	parametrized signal mass feature
$p_{\mathrm{T}}^{i}, \eta^{i}$, (i =1,4)	transverse momentum p_{T} and η of μ_i (i=1,4)
p_{T,Z_i}, (i=1,2)	transverse momenta p_{T} of the Z_1 and the Z_2
$m_{Z1} - m_{Z2}$	subtraction between the mass of two di-muon pairs
$\Delta R(\mu_i^+, \mu_i^-), i = 1, 2$	ΔR of the di-muon formed Z_i (i=1,2)
$\Delta\eta(\mu_i^+, \mu_i^-), i = 1, 2$	$\Delta\eta$ of the di-muon formed Z_i (i=1,2)
$m_{4\mu}$, $p_{\mathrm{T},\ 4\mu}$	invariant mass and transverse momenta of 4 muons system

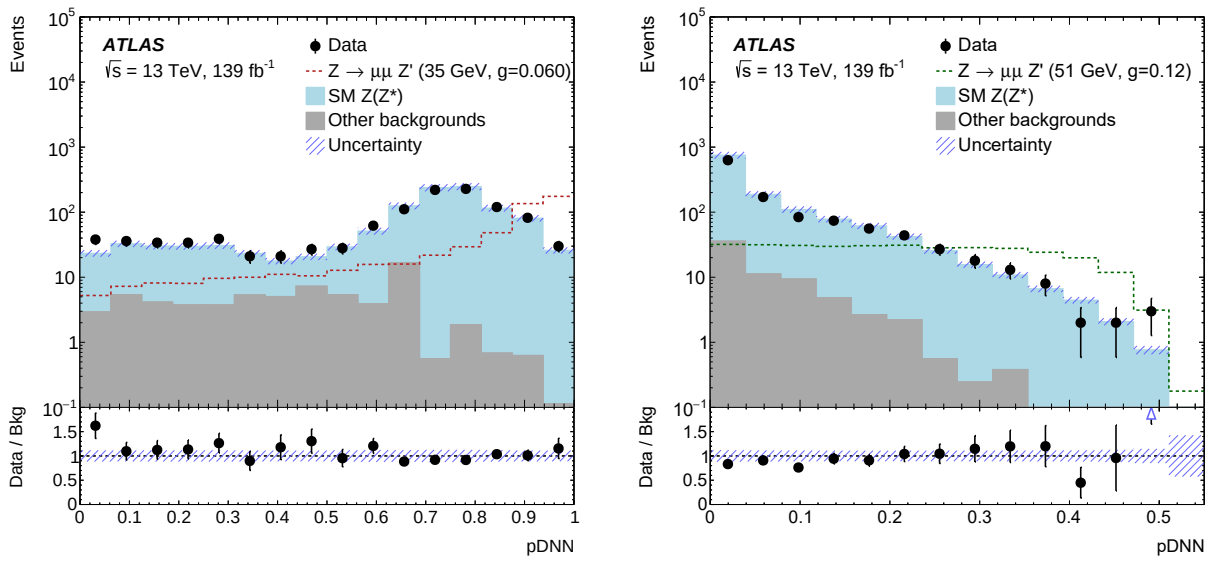
7.6.2 Hyper-parameter selection

To improve the power of the pDNN classifier the model is optimized through an automatic process developed with the package Tune [149] and HEBO Search algorithm [150]. The neural-network structure was chosen to have two fully connected hidden layers, each with 256 (32) nodes for the high (low) mass search in the analysis. Other training hyper-parameters, such as the *learning rate* with *decay* and *class weight*, for both high and low mass searches were also selected. These selected parameters and optimized values are given in Table 7.7.

Table 7.7: The DNN classifier hyper-parameters determined from an auto-optimization process.

	learning rate	learning rate decay	batch size	signal weight	background weight
high-mass	0.03914	0.008538	512	1.04	19.06
low-mass	0.004129	0.0003592	128	9.92	8.01

The learning rate determines how fast the pDNN moves along the gradient descent direction of the loss function, where the learning rate decay decreases the learning rate over epochs to help the pDNN converge more easily to the minimum and avoid oscillation. The batch size determines how many samples are to be propagated to the DNN every time when we estimate the gradient and update the weight. In the end, the class weight for the signal and background is also applied to decide how much attention the DNN pays to each class. After optimization, the pDNN score distributions for Z' mass at 35 GeV and 51 GeV are shown in figure 7.22.



(a) low mass: $m_{Z'} \leq 42$ GeV

(b) high mass: $m_{Z'} \geq 42$ GeV

Figure 7.22: The pDNN output discriminant variable distributions for low mass (a) and high mass (b) with a signal sample at 35 GeV and 51 GeV, respectively.

7.6.3 Input feature importance of pDNN model

To better understand which variables have more contribution to the discriminate power of the pDNN model, we studied the input feature importance. Unlike the tree-based models describe in section 5.2.1, where the feature importance can be calculated directly using average gain across all splits where each feature was used, there is no direct method to calculate feature importance in neural network models. The method we used in this analysis is based on permutation[2]. In a nutshell, to calculate the feature importance x_i, we randomly permute feature x_i across all samples and train a new neural network model f_i. The performance decrease between model f_i and the original model f_0 represents the gain including feature x_i, which shows the importance of that input feature. The reason we didn't directly remove feature x_i instead of permutation is that removing one input feature will change the weight matrix size of the input layer from $n_0 \times n_1$ to $(n_0 - 1) \times n_1$, where n_0 and n_1 are the numbers of input features and number of neurons in the following layer. The change in the neural network structure will also affect the performance of the model. The algorithm to calculate feature importance is shown in 5.

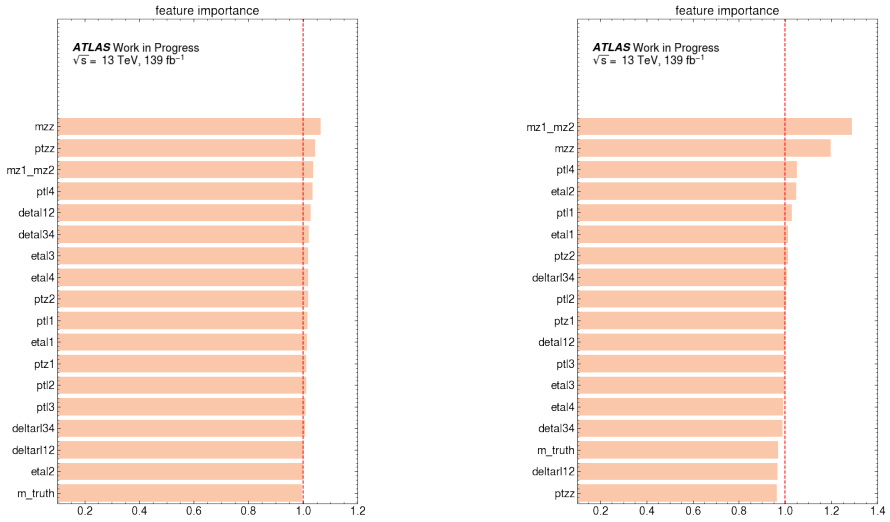
In practice, the quantity we used to represent the model's error is 1-AUC(f), in which AUC represents the area under ROC (receiver operating characteristic) curve. A higher

[2] `https://christophm.github.io/interpretable-ml-book/feature-importance.html`

Algorithm 5 Permutation method of feature importance

The original model is $f_0(x)$ and has n input features. For i=0:
repeat
Randomly permute feature x_i.
Train the alternative model $f_i(\hat{x})$ using the permuted dataset.
Calculated the decrease of model performance or increase in error compared with the original model using the original input dataset ($f_0(x)$-$f_i(x)$).
until $i = n$
Sort the input features in descending of their importance.

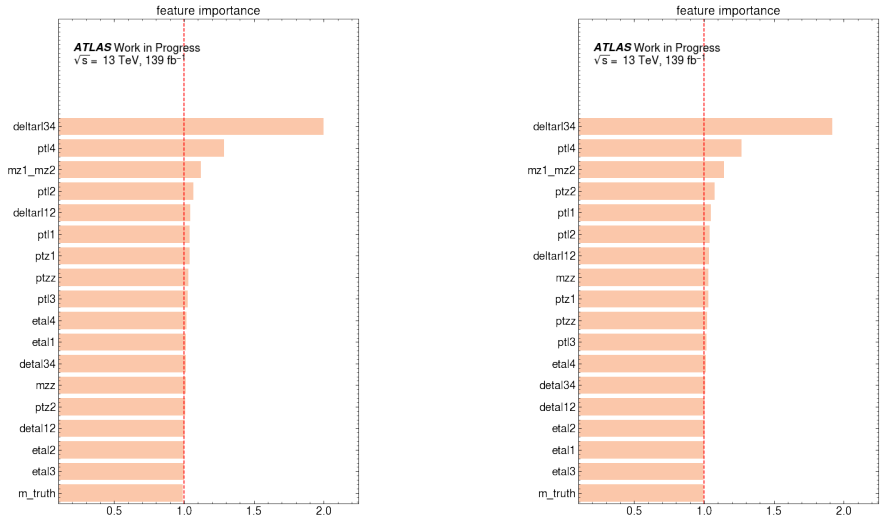
value of 1-AUC means the model performance is worse. To compare the performance of the model trained on permuted dataset $\bar{x}$ and the original dataset x, we used the ratio of 1-AUC value between model f_i and f_0 $\frac{1-\text{AUC}(f_i(x))}{1-\text{AUC}(f_0(x))}$. A higher value means the model performance is worse compared with the original model after training on the permuted dataset. Since input feature distributions for signals at different mass points are different, we performed the feature importance study at different mass points separately. Figures 7.23 and 7.24 show the input feature importance at different Z' mass points.



(a) (b)

Figure 7.23: Input feature importance at 19 GeV (a) and 39 GeV (b).

From the rank plots, in the low mass region, we can see that the major contributions are from $m_{Z1} - m_{Z2}$, $m_{4\mu}$ and p_{l1}^T, while the major contributions are from ΔR_{34}, $m_{Z1} - m_{Z2}$, p_l^T, p_{Z2}^T and $p_{4\mu}^T$ in the high mass region.

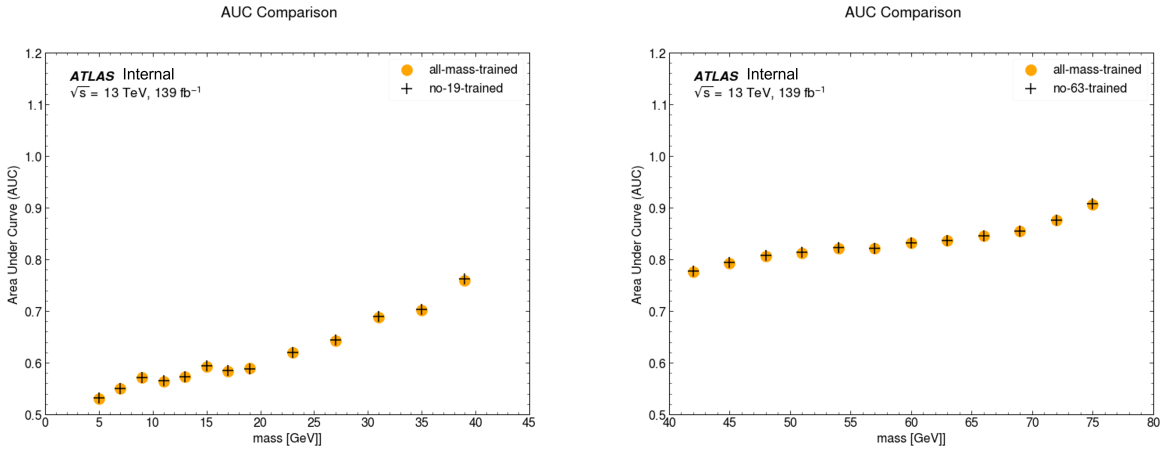


(a) (b)

Figure 7.24: Input feature importance at 57 GeV (a) and 69 GeV (b).

7.6.4 Z' mass interpolation of the pDNN model

As described in section 5.2.3, one major advantage of the parameterized neural network compared with the ordinary neural network model is that the pDNN model could achieve good performance at mass points not included in the training dataset by learning the relation between signals at different mass points, which provides us more flexibility to study the data at more mass points with one model. To validate this property of the pDNN model, we trained alternative test models with certain mass points excluded from the training dataset and compared the performance with the original model. Since the kinematic distributions and model performance are quite different in the high ($m_{Z'}$ >42 GeV) and low mass region ($m_{Z'}$ <42 GeV), this study is performed separately in both the high and low mass regions. In the low mass region, $m_{Z'}$ = 19 GeVsignal is excluded from the training dataset, while $m_{Z'}$ = 63 GeVis excluded in the high mass region. The comparisons are shown in figure 7.25, where test models have similar AUC compared with the original model at all mass points, which means pDNN indeed can achieve good performance in mass points not included in the training.



(a) low mass: $M_{Z'} < 40$ GeV

(b) high mass: $M_{Z'} > 40$ GeV

Figure 7.25: Comparison between AUC of pDNN model trained with all mass points (yellow dots) with pDNN model trained without signal of $m_{Z'}$ = 19 (a) or 63 (b) GeV.

7.6.5 Selection of DNN cuts

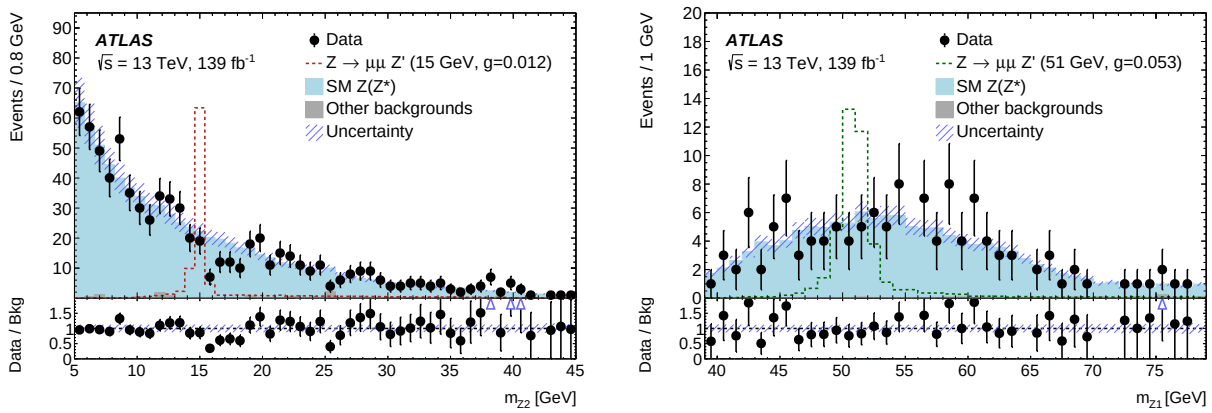
To maximize the search sensitivity, a scan of the pDNN output scores was performed to find the optimal cut values for each Z' mass hypothesis. These cut values vary from 0.42 to 0.74 (0.12 to 0.16) for the low (high) mass region. The cut values for different $m_{Z'}$ signals are given in Table 7.8.

Table 7.8: DNN cut values for different $m_{Z'}$ for optimal search sensitivity.

Mass [GeV]	5	7	9	11	13	15	17	19	23	27	31	35	39
DNN cut	0.42	0.48	0.48	0.44	0.48	0.48	0.48	0.54	0.54	0.60	0.68	0.70	0.74
Mass [GeV]	42	45	48	51	54	57	60	63	66	69	72	75	
DNN cut	0.16	0.14	0.14	0.14	0.14	0.14	0.12	0.12	0.12	0.14	0.14	0.14	

These cuts keep high signal efficiencies between 98% and 95% (90% and 50%), while the corresponding background reductions range from 10% to 50% (50% to 96%) for the low (high) mass region.

After the 4μ event selection with the pDNN classifier, the final discriminant to search for the Z' resonance signature is the Z_1 (for $m_{Z'} \geq 42$ GeV) or Z_2 (for $m_{Z'} \leq 42$ GeV) mass spectrum as shown in Figure 7.26. Data are compared to the estimated background together with two representative signals with masses of 15 and 51 GeV, shown in figure 7.26, respectively. The values of the gauge coupling strengths (g) for the two mass points are chosen for the purpose of illustration.



(a) (b)

Figure 7.26: Mass spectra of m_{Z2} (left) and m_{Z1} (right) for the pDNN-selected events with a signal sample at 15 GeV and 51 GeV, respectively.

7.7 Data interpretation and results

The statistical analysis is performed by comparing the data to the sum of the background prediction and the signal to search for the Z' signature and information about the $pp \to \mu^+\mu^- Z' \to 4\mu$ signal production cross-section and the associated coupling strength.

In case of no significant data excess over the background prediction, upper limits on the signal production cross-section times the decay branching fraction for different Z' masses are set at 95% confidence level (CL).

7.7.1 Statistical fitting

To set the 95% upper limits, the profile likelihood ratio and CLs methods introduced in section 5.1 are used in this analysis. A binned likelihood function as equation 5.3 is used in the statistical fitting. The final discriminator in this analysis is the leading lepton pair mass (m_{Z1}) and sub-leading lepton pair mass m_{Z2} after applying cuts on the DNN score. Each mass spectrum is divided into the signal region (SR) and the background control region (CR). For each Z' mass point the SR is defined in a mass window of $m_{Z'} \pm 3\sigma_{m_{\mu\mu}}$ of the di-muon mass spectrum. The mass windows at different Z' mass points are calculated by a fourth-order polynomial fitting using simulated Z' mass resolution. The fitted Z' mass window function and simulated resolution can be found in figure 7.27.

The Z' mass resolution $\sigma_{m_{\mu\mu}}$ is determined by the fully simulated muon pair mass distribution at the truth level using the double-Gaussian fitting, which combines the Z' natural width and the detector resolution, ranging from 0.10 to 1.75 GeV. The Z' mass resolution is

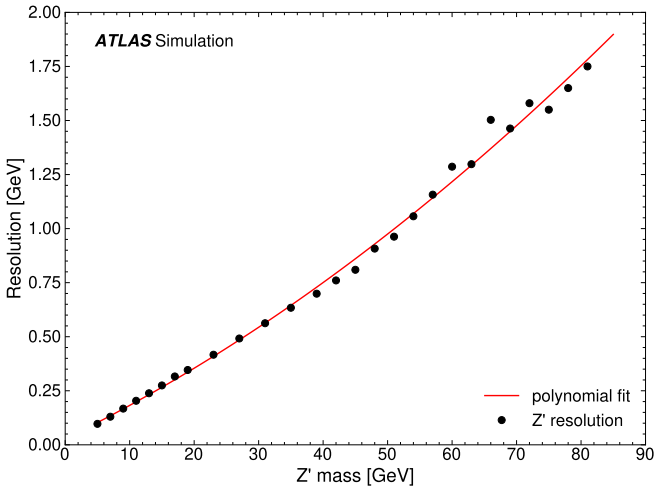
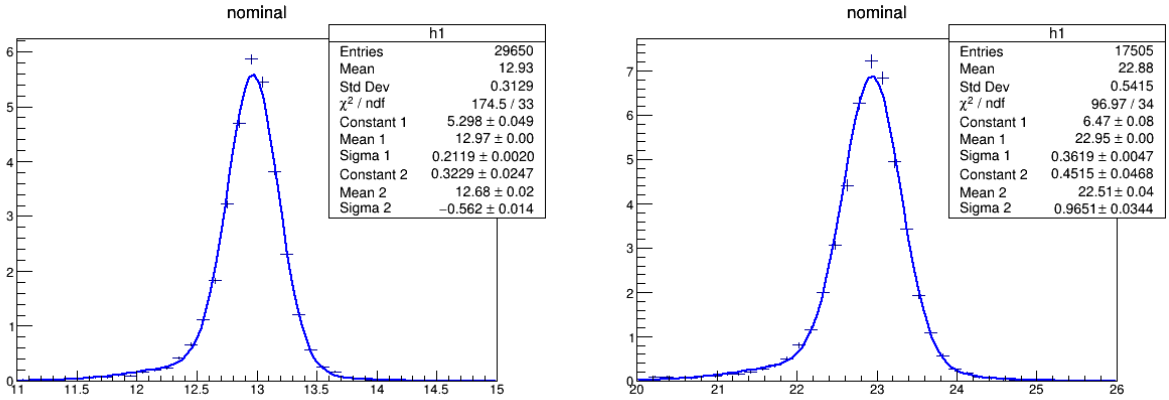


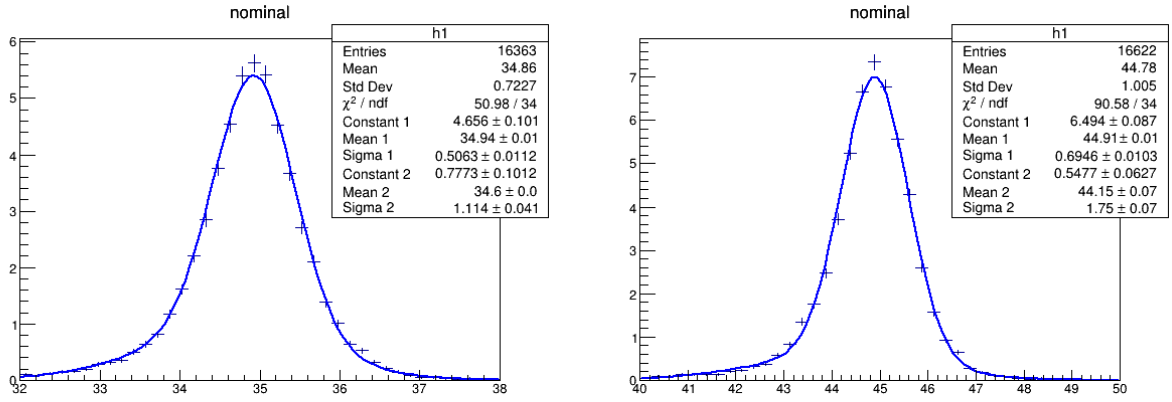
Figure 7.27: The parameterized mass resolution $\sigma_{m_{\mu\mu}}$ as a function of $m_{Z'}$ using fully simulated $Z' \to \mu^+\mu^-$ events.

the weighted average of standard deviations of the two Gaussian distributions. Representative Z' mass distributions and the fitting results are shown in figure 7.28 to figure 7.30. The mass resolution is mostly dominated by the detector resolution.



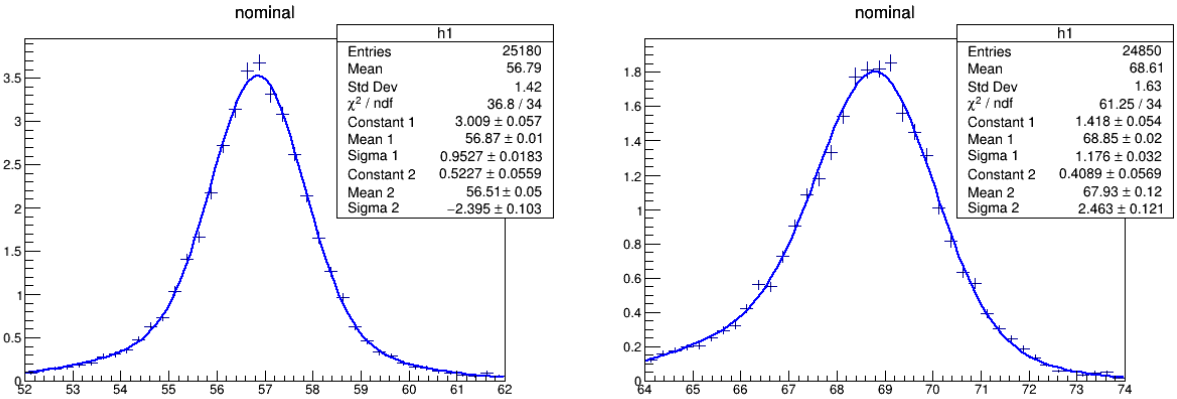
(a) (b)

Figure 7.28: Double Gaussian fitting result at 13 (a) and 23 (b) GeV.



(a) (b)

Figure 7.29: Double Gaussian fitting result at 35 (a) and 45 (b) GeV.



(a) (b)

Figure 7.30: Double Gaussian fitting result at 57 (a) and 69 (b) GeV.

The sidebands outside of the SR are defined as the CR. Finer binning is used in the SR to enhance sensitivity. The background CR is used to constrain the overall normalization for the background in the signal region. The shape of the major background from $Z(Z^*) \to 4\mu$ is fixed with prior uncertainties included in the fitting process, but the normalization (or

strength) floats in the fit. Other background normalizations and shapes are fixed with prior uncertainties included in the fitting.

7.7.2 The p_0 Scan Results

The p_0-values corresponding to the background-only hypothesis is scanned in the mass range of this analysis. A binned profile-likelihood fit [73] is performed simultaneously across the Z' signal region and the background control region using the predicted and observed mass spectrum as inputs. Data are fit to the m_{Z1} and m_{Z2} distributions for $m_{Z'} \geq 42$ GeV and $m_{Z'} \leq 42$ GeV, respectively, with a "sliding" mass window as the defined SR changes for different Z' mass points. The chosen bin size inside the SR is around 0.3 $\sigma_{m_{\mu\mu}}$, for each mass point. The total number of bins in the CR is 20. The fit mass range of m_{Z1} (m_{Z2}) is [30, 85] GeV ([0, 45] GeV).

The p_0-values at different Z' mass hypothesis points are computed and transformed into Gaussian standard deviations to indicate the significance as shown in Figure 7.31. The smallest p_0-value is at 39.6 GeV, corresponding to a local 2.65σ deviation from the background-only hypothesis, while the global deviation [151] is found to be 0.52σ, indicating that no significant data excess over the expected background is observed.

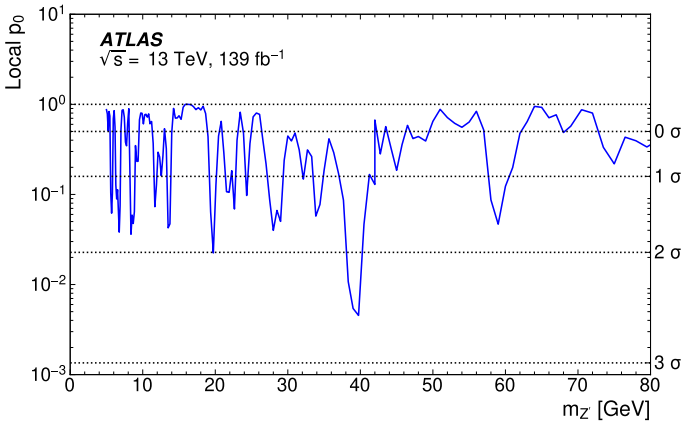


Figure 7.31: The p_0-value scan across the Z' mass signal regions.

7.7.3 Upper Limits

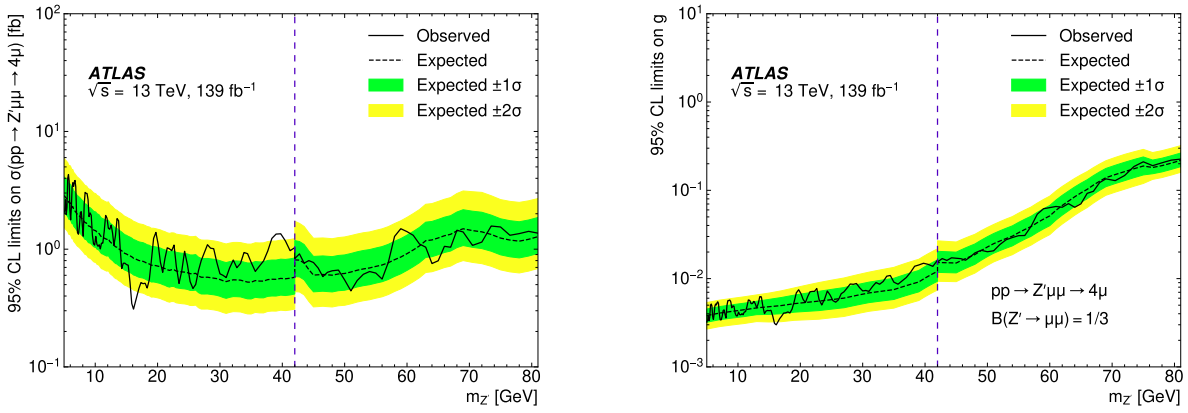
The upper limits on the production cross-section times branching fraction of the $pp \to \mu^+\mu^- Z' \to 4\mu$ process are calculated using a similar fitting procedure described in Section 7.7.2. Confidence intervals are computed based on the profile-likelihood-ratio test

statistics [73]. The observed and expected upper limits at 95% CL on the cross-section times branching fraction, $\sigma(pp \to Z'\mu\mu \to 4\mu)$, are shown in Figure 7.32a.

Assuming the same coupling strength g of the Z' boson to the second and third lepton families and to the left- and right-handed fermions, the upper limits on the coupling parameter g are extracted from the limits of the Z' production cross-section times branching fraction using the $L_\mu - L_\tau$ model, where the branching fraction of $\mathcal{B}(Z' \to \mu^+\mu^-) = \frac{1}{3}$, which is determined by counting all the possible Z' decay modes in this model.

At each generated Z' mass point, a limit on the coupling strength g has been obtained from the cross-section limit.

The observed and expected upper limits on the coupling parameter g are shown in figure 7.32b. The limits on the coupling g are in the range of 0.003 (for $m_{Z'} = 5$ GeV) to 0.2 (for $m_{Z'} = 81$ GeV) depending on the Z' mass ranging from 5 to 81 GeV. This ensures that the ratio of the Z' natural width and mass, $\Gamma(Z')/m_{Z'}$, is well below 1% in this mass range.



(a) (b)

Figure 7.32: 95% CL upper limits (expected and observed) on the cross-sections times branching fraction (a) and coupling parameter (b). The discontinuity at 42 GeV represents the border of the low/high mass classifiers.

Motivated by theoretical interpretations in Ref. [152], a 2-dimensional exclusion contour at 95% CL in the parameter-space of $(m_{Z'}, g)$ of the $L_\mu - L_\tau$ model from this analysis is produced and shown in Figure 7.33. The parameter space exclusion regions were calculated by theorists using data from the Neutrino Trident experiment [153] and the B_s mixing measurements by a global analysis performed in Ref. [152] are also shown in Figure 7.33. This had left a large gap in the parameter space not yet excluded, that was allowed to explain the LHCb $b \to s\mu^+\mu^-$ anomalies [44, 154]. This gap is now largely excluded by this analysis.

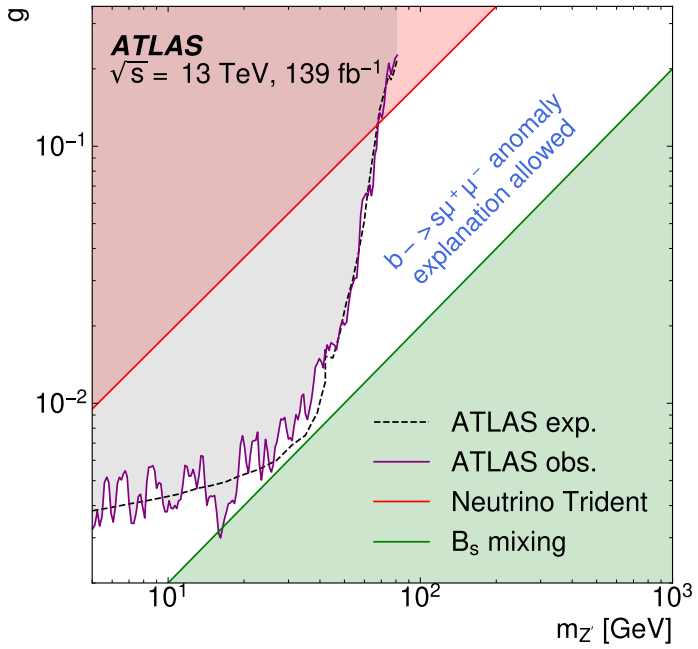


Figure 7.33: The coupling parameter g limits from this search as a function of the Z' mass compared to the limits [152] from the Neutrino Trident (red) and the B_s mixing (green) experimental results.

Chapter 8

Summary

Physics results from two analyses based on proton-proton data collected by the ATLAS detector on the LHC are presented in this dissertation. The data used in this thesis work are collected from 2015 to 2018 during the LHC Run 2 period with the integrated luminosity of 139 fb^{-1}. The four-lepton final states are used in both analyses. The four-lepton final states are clean and sensitive to searches of the VBS ZZ production and the new gauge boson Z' detection. Dominate background in the four-lepton final states comes from the $qq \to ZZ^*$ process including the QCD vertex. The reducible backgrounds from fake leptons have small contributions in the underlying signal regions. In both analyses, we used data-driven methods to estimate the contributions from fake backgrounds. Machine learning algorithms (MVA), such as BDT, DNN, and pDNN are used in the analyses to improve the detection sensitivity.

The electroweak vector boson scattering (VBS) in the $ZZjj$ production is observed for the first time at the LHC, which is an important probe for studying the dynamics of electroweak symmetry breaking. Two final states, containing either four charged leptons or two charged leptons and two neutrinos, were used in the analysis. The technique of the Boosted Decision Trees was employed in the analysis to reduce the QCD $ZZjj$ background, resulting in an overall signal significance of 5.7σ over the background-only hypothesis model. Both inclusive and the electroweak VBS $ZZjj$ production cross sections are measured, consisting of the Standard Model predictions. This analysis completes the last piece in the observation of vector boson scattering, following the observation in the WW, WZ, and $Z\gamma$ [155] by the ATLAS experiment. It is an important milestone in the study of the Standard Model and Higgs mechanism.

A search for a new vector boson Z' with the four-muon (4μ) final state is conducted. The Z' is predicted by $L_\mu - L_\tau$ models to address observed phenomena that the Standard Model cannot explain. A deep learning neural network classifier is used to separate the Z' signal from Standard Model background events. No significant excess of events was observed over

the predicted background, and upper limits at a 95% confidence level on the Z' production cross-section times the decay branching fraction were set for the Z' mass ranging from 5 to 81 GeV. The corresponding common coupling strengths, $g_{Z'}$, of the Z' boson to the second and third-generation leptons above 0.003 – 0.2 have been excluded.

The measurements on the vector-boson scattering processes and searching for new gauge boson Z' will be continued at the LHC in Run 3 (and beyond) programs. With much-increased luminosity as well as increased energy provided by the LHC the potential of discovering the breakdown of the SM and new physics is great in the next decade.

Bibliography

[1] The ATLAS Collaboration and Others. The atlas experiment at the cern large hadron collider. *Journal of Instrumentation*, 3(08):S08003, aug 2008.

[2] Oliver Sim Brüning, Paul Collier, P Lebrun, Stephen Myers, Ranko Ostojic, John Poole, and Paul Proudlock. *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004.

[3] CERN. CERN Rapport annuel 2021. Technical report, CERN, Geneva, 2022.

[4] A. Pais and S. B. Treiman. How many charm quantum numbers are there? *Phys. Rev. Lett.*, 35:1556–1559, Dec 1975.

[5] The CMS collaboration. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716(10):P31:P36, Sep 2012.

[6] The ATLAS collaboration. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716(10):P1:P29, Sep 2012.

[7] M. Tanabashi et al. Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018.

[8] S. Mandelstam. Determination of the pion-nucleon scattering amplitude from dispersion relations and unitarity. general theory. *Phys. Rev.*, 112:1344–1360, Nov 1958.

[9] The ATLAS Collaboration. Observation of electroweak production of two jets and a z-boson pair. *Nature Physics*, Feb 2023.

[10] ATLAS Collaboration. Search for a new z' gauge boson in 4 events with the atlas experiment, 2023.

[11] C. N. Yang and R. L. Mills. Conservation of isotopic spin and isotopic gauge invariance. *Phys. Rev.*, 96:191–195, Oct 1954.

[12] S. Tomonaga. On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields*. *Progress of Theoretical Physics*, 1(2):27–42, 08 1946.

[13] Julian Schwinger. Quantum electrodynamics. i. a covariant formulation. *Phys. Rev.*, 74:1439–1461, Nov 1948.

[14] Julian Schwinger. On quantum-electrodynamics and the magnetic moment of the electron. *Phys. Rev.*, 73:416–417, Feb 1948.

[15] R. P. Feynman. The theory of positrons. *Phys. Rev.*, 76:749–759, Sep 1949.

[16] R. P. Feynman. Space-time approach to quantum electrodynamics. *Phys. Rev.*, 76:769–789, Sep 1949.

[17] R. P. Feynman. Mathematical formulation of the quantum theory of electromagnetic interaction. *Phys. Rev.*, 80:440–457, Nov 1950.

[18] F. J. Dyson. The radiation theories of tomonaga, schwinger, and feynman. *Phys. Rev.*, 75:486–502, Feb 1949.

[19] D. Hanneke, S. Fogwell, and G. Gabrielse. New measurement of the electron magnetic moment and the fine structure constant. *Phys. Rev. Lett.*, 100:120801, Mar 2008.

[20] M Gell-Mann. The eightfold way: A theory of strong interaction symmetry. 3 1961.

[21] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.

[22] Abdus Salam and J. C. Ward. Weak and electromagnetic interactions. *Il Nuovo Cimento (1955-1965)*, 11(4):568–577, Feb 1959.

[23] Sheldon L. Glashow. The renormalizability of vector meson interactions. *Nuclear Physics*, 10:107–117, 1959.

[24] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[25] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

[26] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Christopher S. Deans, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, José I. Latorre, Juan Rojo, Maria Ubiali, and The NNPDF collaboration. Parton distributions for the lhc run ii. *Journal of High Energy Physics*, 2015(4):40, Apr 2015.

[27] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental test of parity conservation in beta decay. *Phys. Rev.*, 105:1413–1415, Feb 1957.

[28] Tadao Nakano and Kazuhiko Nishijima. Charge Independence for V-particles*. *Progress of Theoretical Physics*, 10(5):581–582, 11 1953.

[29] M. Gell-Mann. The interpretation of the new particles as displaced charge multiplets. *Il Nuovo Cimento (1955-1965)*, 4(2):848–866, Apr 1956.

[30] M. Goldhaber, L. Grodzins, and A. W. Sunyar. Helicity of neutrinos. *Phys. Rev.*, 109:1015–1017, Feb 1958.

[31] Nicola Cabibbo. Unitary symmetry and leptonic decays. *Phys. Rev. Lett.*, 10:531–533, Jun 1963.

[32] Makoto Kobayashi and Toshihide Maskawa. CP-Violation in the Renormalizable Theory of Weak Interaction. *Progress of Theoretical Physics*, 49(2):652–657, 02 1973.

[33] Particle Data Group. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2020(8), 08 2020. 083C01.

[34] A DJOUADI. The anatomy of electroweak symmetry breaking tome II: The higgs bosons in the minimal supersymmetric model. *Physics Reports*, 459(1-6):1–241, apr 2008.

[35] Jung Chang, Kingman Cheung, Chih-Ting Lu, and Tzu-Chiang Yuan. Ww scattering in the era of post-higgs-boson discovery. *Phys. Rev. D*, 35:072003, 2013.

[36] Michał Szleper. The higgs boson and the physics of ww scattering before and after higgs discovery. 2014.

[37] Carsten Bittrich. Study of polarization fractions in the scattering of massive gauge bosons $w^{\pm}z \to w^{\pm}z$ with the atlas detector at the large hadron collider. 2015.

[38] Xiao-Gang He, G. C. Joshi, H. Lew, and R. R. Volkas. Simplest $Z^{'}$ model. *Phys. Rev. D*, 44:2118–2132, 1991.

[39] Muon g 2 Collaboration. Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm. *Phys. Rev. Lett.*, 126(14):141801, 2021.

[40] Alexander Keshavarzi. The muon g-2 experiment at fermilab. *EPJ Web of Conferences*, 212:05003, 2019.

[41] Graziano Venanzoni. The Fermilab Muon g-2 Experiment. *PoS*, EPS-HEP2015:568, 2016.

[42] Muon g-2 Collaboration. Final report of the e821 muon anomalous magnetic moment measurement at bnl. *Phys. Rev. D*, 73:072003, 2006.

[43] LHCb Collaboration. Test of lepton universality in beauty-quark decays. 3 2021.

[44] LHCb Collaboration. Measurement of form-factor-independent observables in the decay $b^0 \to k^{*0}\mu^+\mu^-$. *Physical Review Letters*, 111(19), 2013.

[45] LHCb Collaboration. Test of lepton universality using $b^+ \to k^+\ell^+\ell^-$ decays. *Physical Review Letters*, 113(15), Oct 2014.

[46] LHCb Collaboration. Measurement of *cp*-averaged observables in the $B^0 \to K^{*0}\mu^+\mu^-$ decay. *Phys. Rev. Lett.*, 125:011802, Jul 2020.

[47] Wolfgang Altmannshofer, Stefania Gori, Stefano Profumo, and Farinaldo S. Queiroz. Explaining dark matter and b decay anomalies with an $l_\mu - l_\tau$ model. *Journal of High Energy Physics*, 2016(12), Dec 2016.

[48] Ernest Ma, D.P. Roy, and Sourov Roy. Gauged $l_\mu - l_\tau$ with large muon anomalous magnetic moment and the bimaximal mixing of neutrinos. *Physics Letters B*, 525(1-2):101–106, 2002.

[49] Paul Langacker. The physics of heavy z' gauge bosons. *Reviews of Modern Physics*, 81(3):1199–1228, Aug 2009.

[50] Fatemeh Elahi and Adam Martin. Constraints on $l_\mu - l_\tau$ interactions at the lhc and beyond. *Physical Review D*, 93(1), Jan 2016.

[51] Fatemeh Elahi and Adam Martin. Using the modified matrix element method to constrain $l_\mu - l_\tau$ interactions. *Physical Review D*, 96(1), 2017.

[52] Andreas Crivellin, Giancarlo D'Ambrosio, and Julian Heeck. Explaining $h \to \mu^\pm\tau^\mp$, $b \to K^*\mu^+\mu^-$, and $b \to k\mu^+\mu^-/b \to ke^+e^-$ in a two-higgs-doublet model with gauged $L_\mu - L_\tau$. *Phys. Rev. Lett.*, 114:151801, Apr 2015.

[53] Renilde Vanden Broeck. THE CERN ACCELERATOR COMPLEX. Complexe des accelerateurs du CERN. 2019.

[54] Mike Lamont. Status of the lhc. *Journal of Physics: Conference Series*, 455(1):012001, aug 2013.

[55] Luminosity determination in pp collisions at $\sqrt{s}$ = 13 TeV using the ATLAS detector at the LHC. Technical report, CERN, Geneva, 2019. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2019-021.

[56] Jorg Wenninger. Operation and Configuration of the LHC in Run 2. 2019.

[57] 2010–2013: the LHC's first long run. *CERN Courier*, 53(7):25–28, 2013.

[58] The ATLAS collaboration. Atlas pixel detector electronics and sensors. *Journal of Instrumentation*, 3, Jul 2008.

[59] A.L. Schorlemmer. Monitoring radiation damage in the ATLAS Pixel Detector. *JINST*, 8:C01045, 2013.

[60] H. Pernegger. The pixel detector of the ATLAS experiment for LHC run-2, 2015.

[61] Y. Unno. ATLAS silicon microstrip detector system (SCT). *Nucl. Instrum. Meth. A*, 511:58–63, 2003.

[62] Hua-Qiao Zhang. The ATLAS Liquid Argon calorimeter: Overview and performance. *J. Phys. Conf. Ser.*, 293:012044, 2011.

[63] A. Henriques. The ATLAS Tile Calorimeter. In *4th International Conference on Advancements in Nuclear Instrumentation Measurement Methods and their Applications*, IEEE Nucl.Sci.Symp.Conf.Rec., 2015.

[64] ATLAS/Tile Calorimeter Collaboration. *ATLAS Tile Calorimeter : Technical Design Report*. CERN, 1996.

[65] E Diehl, S Goldfarb, D Levin, S McKee, H Neal, H Schick, G Tarle, R Thun, C Weaverdyck, Q Xu, Z Zhao, and B Zhou. Long Precision Drift Tube Production at Michigan. Technical report, CERN, Geneva, 2002.

[66] ATLAS Collaboration. Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV. *The European Physical Journal C*, 79(8), Aug 2019.

[67] ATLAS Collaboration. Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s}$=13 TeV. *The European Physical Journal C*, 76(5), May 2016.

[68] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-kt jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, apr 2008.

[69] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev. D*, 48:3160–3166, Oct 1993.

[70] Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189 – 1232, 2001.

[71] ATLAS Collaboration. Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run. 2016.

[72] Tagging and suppression of pileup jets with the ATLAS detector. Technical report, CERN, Geneva, 2014. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2014-018.

[73] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011.

[74] Alexander L. Read. Presentation of search results: the CL_S technique. *J. Phys. G*, 28:2693, 2002.

[75] S. S. Wilks. The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *The Annals of Mathematical Statistics*, 9(1):60 – 62, 1938.

[76] J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, Mar 1986.

[77] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. *CoRR*, abs/1603.02754, 2016.

[78] J J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci U S A*, 79(8):2554–2558, April 1982.

[79] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, page 807–814, Madison, WI, USA, 2010. Omnipress.

[80] Andrew L. Maas. Rectifier nonlinearities improve neural network acoustic models. 2013.

[81] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, Oct 1986.

[82] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.

[83] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015.

[84] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In Yee Whye Teh and Mike Titterington, editors, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, pages 249–256, Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010. PMLR.

[85] The CMS Collaboration. Observation of electroweak production of same-sign w boson pairs in the two jet and two same-sign lepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 120:081801, Feb 2018.

[86] The ATLAS Collaboration. Observation of electroweak production of a same-sign w boson pair in association with two jets in pp collisions at $\sqrt{s} = 13$ TeV with the atlas detector. *Phys. Rev. Lett.*, 123:161801, Oct 2019.

[87] The ATLAS Collaboration. Observation of electroweak w±z boson pair production in association with two jets in pp collisions at s=13 tev with the atlas detector. *Physics Letters B*, 793:469–492, 2019.

[88] The CMS Collaboration. Measurements of production cross sections of wz and same-sign ww boson pairs in association with two jets in proton-proton collisions at s=13tev. *Physics Letters B*, 809:135710, 2020.

[89] The CMS Collaboration. Measurement of vector boson scattering and constraints on anomalous quartic couplings from events with four leptons and two jets in proton–proton collisions at s=13 tev. *Physics Letters B*, 774:682–705, 2017.

[90] The CMS Collaboration. Evidence for electroweak production of four charged leptons and two jets in proton-proton collisions at s=13tev. *Physics Letters B*, 812:135992, 2021.

[91] The ATLAS collaboration. Observation of electroweak production of two jets and a z-boson pair with the atlas detector at the lhc. *arXiv:2004.10612 [hep-ex]*, 2020.

[92] Barbara Jäger, Alexander Karlberg, and Giulia Zanderighi. Electroweak zzjj production in the standard model and beyond in the powheg-box v2. *Journal of High Energy Physics*, 2014(3):141, Mar 2014.

[93] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7):79, Jul 2014.

[94] T. Gleisberg, Stefan. Höeche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert, and J. Winter. Event generation with SHERPA 1.1. *JHEP*, 02:007, 2009.

[95] Richard D. Ball et al. Parton distributions for the LHC run II. *JHEP*, 04:040, 2015.

[96] Fabrizio Caola, Kirill Melnikov, Raoul Röntsch, and Lorenzo Tancredi. Qcd corrections to zz production in gluon fusion at the lhc. *Phys. Rev. D*, 92:094028, Nov 2015.

[97] Hung-Liang Lai et al. New parton distributions for collider physics. *Phys. Rev. D*, 82:074024, 2010.

[98] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.

[99] Paolo Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.

[100] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *JHEP*, 11:070, 2007.

[101] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph5: going beyond. *JHEP*, 06:128, 2011.

[102] Tanju Gleisberg and Stefan Höche. Comix, a new matrix element generator. *JHEP*, 12:039, 2008.

[103] Fabio Cascioli, Philipp Maierhofer, and Stefano Pozzorini. Scattering Amplitudes with Open Loops. *Phys. Rev. Lett.*, 108:111601, 2012.

[104] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Z. Skands. A brief introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852, 2008.

[105] Richard D. Ball et al. Parton distributions with LHC data. *Nucl. Phys. B*, 867:244, 2013.

[106] ATLAS Collaboration. ATLAS Pythia 8 tunes to 7 TeV data. ATL-PHYS-PUB-2014-021, 2014.

[107] ATLAS Collaboration. The ATLAS simulation infrastructure. *Eur. Phys. J. C*, 70:823, 2010.

[108] S. Agostinelli et al. GEANT4 – a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250, 2003.

[109] Jon Butterworth et al. PDF4LHC recommendations for LHC Run II. *J. Phys. G*, 43:023001, 2016.

[110] Wolfgang Altmannshofer, Stefania Gori, Maxim Pospelov, and Itay Yavin. Quark flavor transitions in $L_\mu - L_\tau$ models. *Phys. Rev. D*, 89:095033, May 2014.

[111] CMS Collaboration. Observation of Z Decays to Four Leptons with the CMS Detector at the LHC. *JHEP*, 12:034, 2012.

[112] ATLAS Collaboration. Measurements of Four-Lepton Production at the Z Resonance in pp Collisions at $\sqrt{s}$ =7 and 8 TeV with ATLAS. *Phys. Rev. Lett.*, 112(23):231806, 2014.

[113] CMS Collaboration. Search for an $l_\mu - l_\tau$ gauge boson using z $\rightarrow 4\mu$ events in proton-proton collisions at $\sqrt{s} = 13$ tev. *Phys. Lett. B*, 792:345, 2019.

[114] ATLAS Collaboration. ATLAS data quality operations and performance for 2015-2018 data-taking. *JINST*, 15(04):P04003–P04003, apr 2020.

[115] ATLAS Collaboration. Performance of the ATLAS muon triggers in Run 2. *JINST*, 15:P09015, 2020.

[116] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[117] Celine Degrande, Claude Duhr, Benjamin Fuks, David Grellscheid, Olivier Mattelaer, and Thomas Reiter. UFO - The Universal FeynRules Output. *Comput. Phys. Commun.*, 183:1201–1214, 2012.

[118] Adam Alloul, Neil D. Christensen, Céline Degrande, Claude Duhr, and Benjamin Fuks. FeynRules 2.0 - A complete toolbox for tree-level phenomenology. *Comput. Phys. Commun.*, 185:2250–2300, 2014.

[119] Francisco del Aguila, Mikael Chala, Jose Santiago, and Yasuhiro Yamamoto. Collider limits on leptophilic interactions. *JHEP*, 03:059, 2015.

[120] Massimiliano Grazzini, Stefan Kallweit, and Dirk Rathlev. ZZ production at the LHC: Fiducial cross sections and distributions in NNLO QCD. *Phys. Lett. B*, 750:407–410, nov 2015.

[121] Ajjath A. H., Goutam Das, M. C. Kumar, Pooja Mukherjee, V. Ravindran, and Kajal Samanta. Resummed drell-yan cross-section at n^3ll, 2020.

[122] N. Davidson, T. Przedzinski, and Z. Was. PHOTOS Interface in C++: Technical and physics documentation. *Comput. Phys. Commun.*, 199:86–101, 2016.

[123] Enrico Bothmann et al. Event generation with Sherpa 2.2. *SciPost Phys.*, 7(3):034, 2019.

[124] Steffen Schumann and Frank Krauss. A parton shower algorithm based on Catani–Seymour dipole factorisation. *JHEP*, 03:038, 2008.

[125] Stefan Höche, Frank Krauss, Marek Schönherr, and Frank Siegert. A critical appraisal of NLO+PS matching methods. *JHEP*, 09:049, 2012.

[126] Stefan Höche, Frank Krauss, Marek Schönherr, and Frank Siegert. QCD matrix elements + parton showers. The NLO case. *JHEP*, 04:027, 2013.

[127] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber. QCD Matrix Elements + Parton Showers. *JHEP*, 11:063, 2001.

[128] Stefan Höche, Frank Krauss, Steffen Schumann, and Frank Siegert. QCD matrix elements and truncated showers. *JHEP*, 05:053, 2009.

[129] Federico Buccioni, Jean-Nicolas Lang, Jonas M. Lindert, Philipp Maierhöfer, Stefano Pozzorini, Hantian Zhang, and Max F. Zoller. OpenLoops 2. *Eur. Phys. J. C*, 79(10):866, 2019.

[130] Ansgar Denner, Stefan Dittmaier, and Lars Hofer. COLLIER: A fortran-based complex one-loop library in extended regularizations. *Comput. Phys. Commun.*, 212:220–238, 2017.

[131] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction. *JHEP*, 09:126, 2007.

[132] Paolo Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004.

[133] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with parton shower, simulations: the POWHEG method. *JHEP*, 11:070, 2007.

[134] Simone Alioli, Paolo Nason, Carlo Oleari, and Re Emanuele. A general framework for implementing NLO calculations in shower monte carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010.

[135] ATLAS Collaboration. Studies on top-quark Monte Carlo modelling for Top2016. (ATL-PHYS-PUB-2016-020), Sep 2016.

[136] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159, 2015.

[137] Emanuele Re. Single-top Wt-channel production matched with parton showers using the POWHEG method. *Eur. Phys. J. C*, 71:1547, 2011.

[138] Stefano Frixione, Eric Laenen, Patrick Motylinski, Chris White, and Bryan R. Webber. Single-top hadroproduction in association with a W boson. *JHEP*, 07:029, 2008.

[139] Charalampos Anastasiou, Lance J. Dixon, Kirill Melnikov, and Frank Petriello. High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at next-to-next-to leading order. *Phys. Rev. D*, 69:094008, 2004.

[140] ATLAS Collaboration. The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie-Landshoff diffractive model. 8 2016.

[141] ATLAS Collaboration. Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13\,\text{TeV}$. *Eur. Phys. J. C*, 81(7):578, Jul 2021.

[142] ATLAS Collaboration. Jet reconstruction and performance using particle flow with the atlas detector. *Eur. Phys. J. C*, 77(7), Jul 2017.

[143] ATLAS Collaboration. Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data. *JINST*, 14(12):P12006–P12006, dec 2019.

[144] ATLAS Collaboration. Jet reconstruction and performance using particle flow with the ATLAS Detector. *Eur. Phys. J. C*, 77:466, 2017.

[145] G. Avoni, M. Bruschi, G. Cabras, D. Caforio, N. Dehghanian, A. Floderus, B. Giacobbe, F. Giannuzzi, F. Giorgi, P. Grafström, V. Hedberg, F. Lasagni Manghi, S. Meneghini, J. Pinfold, E. Richards, C. Sbarra, N. Semprini Cesari, A. Sbrizzi, R. Soluk, G. Ucchielli, S. Valentinetti, O. Viazlo, M. Villa, C. Vittori, R. Vuillermet, and A. Zoccoli. The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS. *JINST*, 13(07):P07017–P07017, jul 2018.

[146] Pierre Baldi, Kyle Cranmer, Taylor Faucett, Peter Sadowski, and Daniel Whiteson. Parameterized neural networks for high-energy physics. *Eur. Phys. J. C*, 76(5), apr 2016.

[147] François Chollet et al. Keras. `https://keras.io`, 2015.

[148] Mart'ın Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Man'e, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Vi'egas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[149] Richard Liaw, Eric Liang, Robert Nishihara, Philipp Moritz, Joseph E. Gonzalez, and Ion Stoica. Tune: A research platform for distributed model selection and training. 2018.

[150] Alexander I. Cowen-Rivers, Wenlong Lyu, Rasul Tutunov, Zhi Wang, Antoine Grosnit, Ryan Rhys Griffiths, Alexandre Max Maraval, Hao Jianye, Jun Wang, Jan Peters, and Haitham Bou Ammar. Hebo pushing the limits of sample-efficient hyperparameter optimisation, 2020.

[151] Eilam Gross and Ofer Vitells. Trial factors for the look elsewhere effect in high energy physics. *Eur. Phys. J. C*, 70(1-2):525–530, oct 2010.

[152] Wolfgang Altmannshofer, Stefania Gori, Stefano Profumo, and Farinaldo S. Queiroz. Explaining dark matter and B decay anomalies with an $L_\mu - L_\tau$ model. *JHEP*, 12:106, 2016.

[153] S. R. Mishra, S. A. Rabinowitz, C. Arroyo, K. T. Bachmann, R. E. Blair, C. Foudas, B. J. King, W. C. Lefmann, W. C. Leung, E. Oltman, P. Z. Quintas, F. J. Sciulli, B. G. Seligman, M. H. Shaevitz, F. S. Merritt, M. J. Oreglia, B. A. Schumm, R. H. Bernstein, F. Borcherding, H. E. Fisk, M. J. Lamm, W. Marsh, K. W. B. Merritt, H. Schellman, D. D. Yovanovitch, A. Bodek, H. S. Budd, P. de Barbaro, W. K. Sakumoto, P. H. Sandler, and W. H. Smith. Neutrino tridents and w-z interference. *Phys. Rev. Lett.*, 66:3117–3120, Jun 1991.

[154] LHCb Collaboration. Test of lepton universality in beauty-quark decays. *Nat. Phys.*, 18(3):277–282, Mar 2022.

[155] ATLAS Collaboration. Measurement of electroweak $z(\nu\bar{\nu})jj$ production and limits on anomalous quartic gauge couplings in pp collisions at $\sqrt{s} = 13$ tev with the atlas detector, 2022.