

**Search for the Lepton Flavor Violating
decay $Z \rightarrow e\mu$ with the ATLAS detector at the LHC**

by

Hai Jiang

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: High Energy Physics

Program of Study Committee:
Chunhui Chen, Major Professor
Soeren Prell
Kerry Whisnant
James Vary
Cindy Yu

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2019

Copyright © Hai Jiang, 2019. All rights reserved.

DEDICATION

I would like to dedicate this thesis to my parents, Tianchao Jiang and Min Li, without whose support I would not have been able to complete this work. I would also like to thank my friends and family for their effort and encouragement during the research work.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	vii
ABSTRACT	viii
CHAPTER 1. INTRODUCTION AND THEORY	1
1.1 Introduction	1
1.2 Why We Need High Energy Physics	3
1.3 The Standard Model	4
1.4 Fermions	6
1.4.1 Quarks	7
1.4.2 Leptons	8
1.5 Bosons and Forces	9
1.5.1 Electromagnetic force with photon γ	10
1.5.2 Weak force with $W^\pm$ and Z^0 boson	11
1.5.3 Strong force with gluon g	12
1.5.4 Higgs	13
1.6 Lepton Flavor	14
1.7 Lepton Flavor Violation and Searches	15
CHAPTER 2. THE LHC AND ATLAS DETECTOR	19
2.1 CERN and LHC	21
2.2 ATLAS Detector	25
2.3 Inner Detector	29

2.3.1	Pixel Detector	34
2.3.2	Semi-Conductor Tracker (SCT)	34
2.3.3	Transition Radiation Tracker (TRT)	35
2.4	Calorimeter	36
2.4.1	Electromagnetic Calorimeter	38
2.4.2	Hadronic Calorimeter	38
2.5	Muon Spectrometer	40
2.6	Trigger System and Data Acquisition	41
2.7	Pileup	42
CHAPTER 3. ANALYSIS STRATEGY		45
CHAPTER 4. OBJECT RECONSTRUCTION		48
4.1	Introduction	48
4.2	Primary vertex	48
4.3	Electron	49
4.4	Muon	51
4.5	Jet	54
4.5.1	Jets reconstruction	57
4.5.2	b-tagging	59
4.6	Missing Transverse Momentum E_T^{miss}	60
4.7	Object overlap removal	61
CHAPTER 5. EVENT PRE-SELECTION		62
5.1	Data and Good Runs List	62
5.2	Main Backgrounds	63
5.3	Monte Carlo Samples	63
5.4	Event cleaning	65

5.5	Trigger Selection	66
5.6	Event Reweighting	66
5.7	Event Pre-selection	67
CHAPTER 6. EVENT FINAL-SELECTION		74
6.1	Multivariate Technique	74
6.2	Gradient Boosted Decision Tree (BDTG)	77
6.3	Event final selection based on a multivariate technique	82
6.4	Data and MC comparison after final selection	86
6.5	Determination of the signal excess by a fit to the $m_{e\mu}$ spectrum	87
6.5.1	Description of the $Z \rightarrow \tau\tau$ background	87
6.5.2	Description of the $Z \rightarrow \mu\mu$ background	91
6.5.3	Description of the combinatorial background	94
6.5.4	Data-driven fitting method	96
CHAPTER 7. SYSTEMATIC UNCERTAINTIES		99
7.1	Sideband fitting	99
7.2	Number of $Z \rightarrow ee, \mu\mu$ events	99
7.3	Pileup Reweighting	99
7.4	Luminosity	100
7.5	Electron	100
7.6	Muon	101
7.7	Missing transverse energy E_T^{miss}	101
7.8	Jet Energy Scale and Resolution Uncertainty	102
7.9	B-tagging systematics	102
7.10	Other systematic uncertainties	102
CHAPTER 8. UPPER LIMIT CALCULATIONS		104

8.1 Upper Limit and CL_s method	104
8.2 Upper Limit on branching ratio	105
CHAPTER 9. SUMMARY	108
BIBLIOGRAPHY	110
APPENDIX A. THE LIST OF TRIGGERS USED IN EACH DATA-TAKING PERIOD . .	121
APPENDIX B. SENSITIVITY USING CUT-BASED METHOD	122
APPENDIX C. STUDY ON VETOING b -jets	123
APPENDIX D. MC SAMPLES	125

ACKNOWLEDGMENTS

I would like to take this opportunity to express my thanks to those who helped me with various aspects of conducting research and the writing of this thesis.

Firstly, I'd like to thank Professor Chunhui Chen for his guidance, patience and support throughout this research. His insights and words of encouragement have often inspired me and renewed my hopes for completing my graduate education. I would also want to thank the ISU ALTAS group for their assistance and friendship, and I have learned a lot from them. And I want to thank my committee members for their efforts and contributions to this work. Moreover, thank you to my friends and family, for providing unconditional and unending support, especially when I was at the lows of life. Many thanks to the difficulties through my doctoral journey, they would be a great fortune for my whole life. Finally, a special thanks to Physics, the most elegant masterpiece in the world, that I am lucky to experience the ultimate beauty of the universe.

ABSTRACT

A search for the lepton flavor violating process $Z \rightarrow e\mu$ is conducted with the ATLAS detector at the Large Hadron Collider in pp collisions using $140 \pm 2.9 \text{ fb}^{-1}$ of data collected at $\sqrt{s} = 13 \text{ TeV}$ in Run2. An enhancement in the $e\mu$ invariant mass spectrum is searched for at the Z boson mass. The number of Z bosons produced in the data sample is estimated using events of similar topology, $Z \rightarrow ee$ and $\mu\mu$ channels, significantly reducing the object reconstruction systematic uncertainty in the measurement. For the current blinding stage, an improved expected upper limit on the branching fraction $\text{Br}(Z \rightarrow e\mu) < 1.52 \times 10^{-7}$ at the 95% confidence level has been achieved.

CHAPTER 1. INTRODUCTION AND THEORY

1.1 Introduction

Protons, electrons, neutrons, neutrinos and quarks are often featured in news of scientific discoveries. All of these, and a whole zoo of others, are tiny sub-atomic particles too small to be seen even with microscopes. While molecules and atoms are the basic elements of familiar substances that we can see and feel, we have to "look" within atoms in order to learn about the "elementary" sub-atomic particles and to understand the nature of our Universe. The science of this study is called Particle Physics, Elementary Particle Physics or sometimes High Energy Physics (HEP).

Atoms were postulated long ago by the Greek philosopher Democritus, and until the beginning of the 20th century, atoms were thought to be the fundamental indivisible building blocks of all forms of matter. People soon realized that they could categorize atoms into groups that shared similar chemical properties as in the Periodic Table of the Elements. This indicated that atoms were made up of simpler building blocks, and that it was these simpler building blocks in different combinations that determined which atoms had which chemical properties. Protons, neutrons and electrons came to be regarded as the fundamental particles of nature when we learned in the 1900's through the experiments of Rutherford and others that atoms consist of mostly empty space with electrons surrounding a dense central nucleus made up of protons and neutrons.

The science of particle physics surged forward with the invention of particle accelerators that could accelerate protons or electrons to high energies and smash them into nuclei – to the surprise of scientists, a whole host of new particles were produced in these collisions.

By the early 1960s, as accelerators reached higher energies, a hundred or more types of particles were found. Could all of these then be new fundamental particles? Confusion reigned until it became clear late in the last century, through a long series of experiments and theoretical studies, that there existed a very simple scheme of two basic sets of particles: the quarks and leptons

(among the leptons are electrons and neutrinos), and a set of fundamental forces that allow these to interact with each other. These "forces" themselves can be regarded as being transmitted through the exchange of particles called gauge bosons. An example of these is the photon, the quantum of light and the transmitter of the electromagnetic force.

Together these fundamental particles form various combinations that are observed today as protons, neutrons and the zoo of particles seen in accelerator experiments. And all these sets of particles also include their anti-particles, or might be called their complementary opposites. These make up matter and anti-matter.

Today, almost everyone has become used to the idea that all matter is a collection of atoms, and that those atoms have nuclei with electrons circling around them. And the nuclei are composed of protons and neutrons, which contain the quarks. As is illustrated in the Figure 1.1.

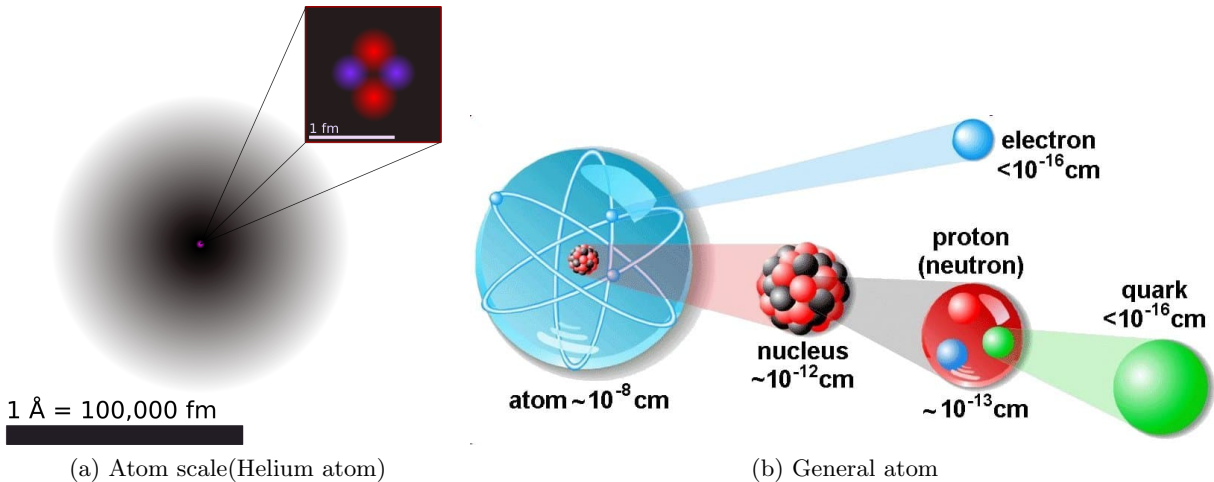


Figure 1.1: Atom structure

As far as we know, quarks and leptons are like points in geometry. They're fundamental particles, not made up of anything else. To be more specific, hadrons (protons, neutrons, pions, ...) are built up from quarks bound by gluons. The color force between particles with color charge binds them into hadrons. The residual color force outside color-neutral hadrons is the nuclear

force, which binds stable hadrons into nuclei. The electrically charged nuclei and stable electrically charged leptons (only the electron) are bound into atoms by the electric force, mediated by photons. The residual electromagnetic force outside electrically neutral atoms binds them into molecules. Thus is the hierarchy of structures in nature built. The details will be explained in later sections.

The Standard Model is the theory that describes the role of these fundamental particles and interactions between them. It is a beautiful scheme with well defined calculational rules, agreeing well with experiment results. It still contains many secrets including the graviton and so called "beyond standard model" which can't be adequately explained. Even though, the Standard Model represents an enormous body of knowledge of Nature that can be seen as the cumulation of 400 years of Physics. And the role of Particle Physics is to test this model in all conceivable ways, and seeking to discover something more lies beyond it, which is called the Beyond Standard Model physics.

1.2 Why We Need High Energy Physics

Our present understanding of the universe in a nutshell is we believe that the Universe started off with a "Big Bang", with enormously high energy and temperature concentrated in an infinitesimally small volume. The Universe immediately started to expand at a furious rate and some of the energy was converted into pairs of particles and antiparticles with mass – remember Einstein's famous $E = mc^2$. In the first tiny fraction of a second, only a mix of radiation (photons of pure energy) and quarks, leptons and gauge bosons existed. During the very dense phase, particles and antiparticles collided and annihilated each other into photons, leaving just a tiny fraction of matter to carry on in the Universe. As the Universe expanded rapidly, in about a hundredth of a second it cooled to a temperature of about 100 billion degrees, and quarks began to clump together into protons and neutrons which swirled around with electrons, neutrinos and photons in a grand soup of particles. From this point on, there were no free quarks to be found. In the next three minutes or so, the Universe cooled to about a billion degrees, allowing protons and neutrons to clump together to form the nuclei of light elements such as deuterium, helium and lithium. After about

three hundred thousand years, the Universe cooled enough (to a few thousand degrees) to allow the free electrons to become bound to light nuclei and thus formed the first atoms. Free photons and neutrinos continue to stream throughout the Universe, meeting and interacting occasionally with the atoms in galaxies, stars and us.

We see now that to understand how the Universe evolved, which is illustrated in Figure 1.2, we really need to understand the behavior of the elementary particles: the quarks, leptons and gauge bosons. Physicists constantly look for new particles. When we find them, we categorize them and try to find patterns that tell us about how the fundamental building blocks of the universe interact. These elementary particles make up all the known recognizable matter in our Universe.

Below we will describe this Standard Model and its salient features.

1.3 The Standard Model

Particle physicists now believe we can describe the behavior of all known subatomic particles within a single theoretical framework called the Standard Model(SM), incorporating quarks and leptons and their interactions through the strong, weak and electromagnetic forces. Gravity is the one force not described by the Standard Model.

One guiding principle that led to current ideas about the nature of elementary particles was the concept of Symmetry. Nature points the way to many of its underlying principles through the existence of various symmetries. The Standard Model (SM) is a gauge theory based on the symmetry group $U(1) \times SU(2) \times SU(3)$, which can be used to describe the three fundamental forces covered by the SM - electromagnetic, weak, and strong forces. Gravity is the fourth fundamental force, but it hasn't been included in SM yet.

Based on the spin character (spin is the intrinsic angular momentum of a particle), all the fundamental particles are categorized as fermions (spin = $1/2$), gauge bosons (spin = 1), and scalar boson (spin = $1/2$). All ordinary matter consists of spin $1/2$ particles known as fermions, while fermions are further separated as quarks and leptons, which will be explained in Chapter 1.4. The interaction between these fermions are mediated by spin 1 particles known as gauge bosons. Finally,

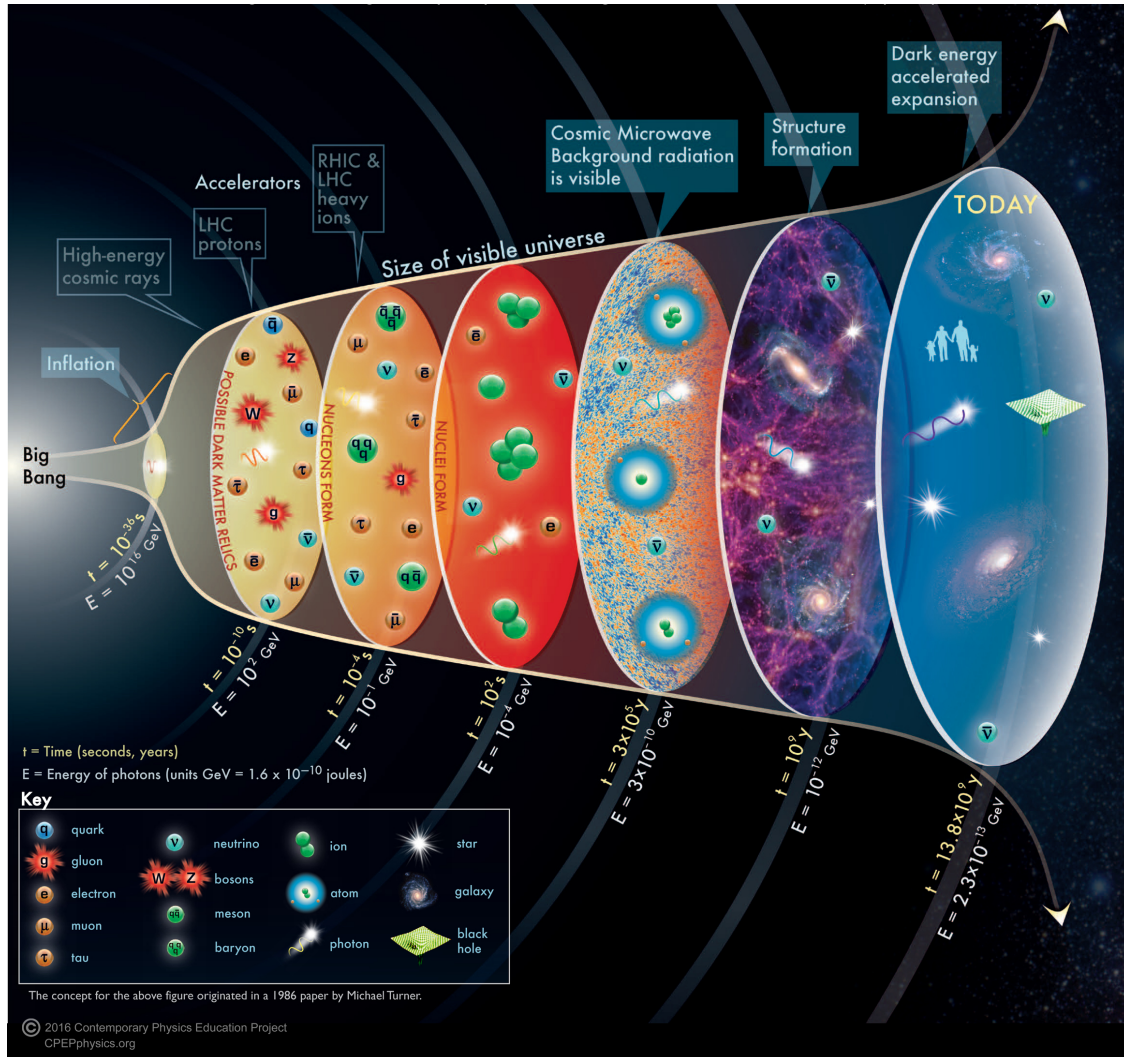


Figure 1.2: The history and fate of the Universe

there exists the spin 0 particle known as the Higgs boson which interacts with other particles and give them mass through its Higgs field. Both gauge bosons and scalar boson will be explained in Chapter 1.5.

The elementary particle fermions and the interaction bosons are included in the scheme of Standard Model are summarized in Figure 1.3.

Standard Model of Elementary Particles

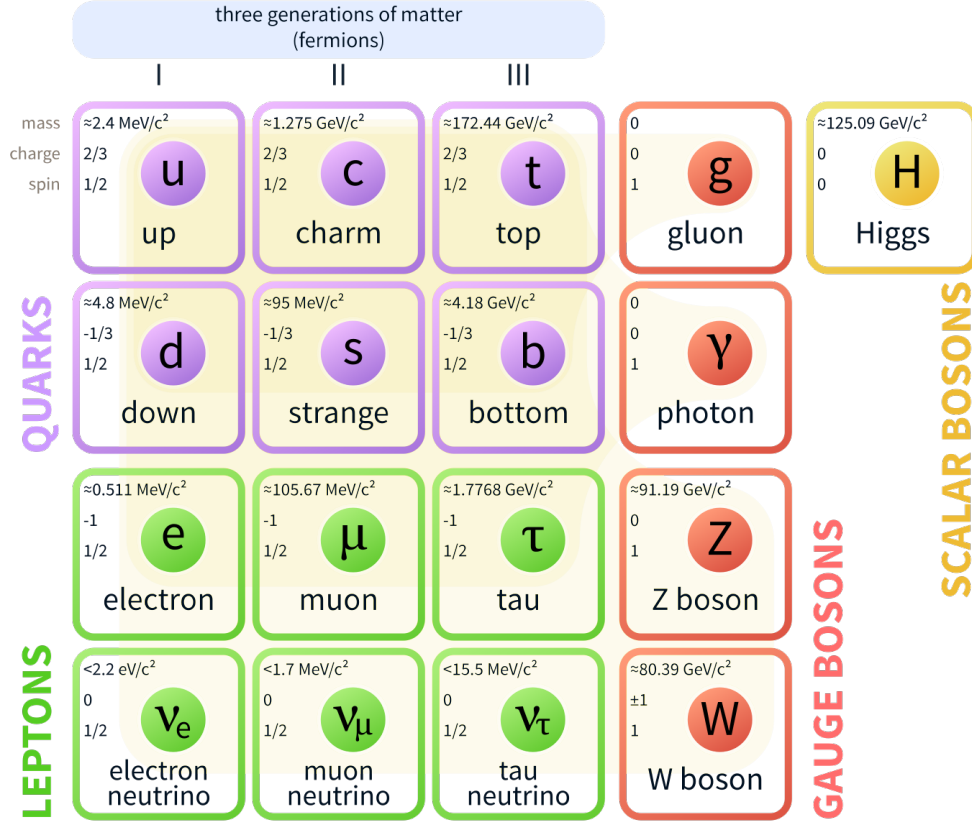


Figure 1.3: Standard Model

1.4 Fermions

Fermions are spin $1/2$ particles and can be broken into two groups: leptons and quarks. Each group can be further separated into three generations (I II III columns in the Figure 1.3) of particles based on their mass and stability. The lightest and most stable particles are the first generation (Up and Down quark in quark sector, along with electron and electron neutrino in lepton sector), on which our matter world are based. And as we move through the generations, the particles are becoming heavier and less stable, because higher generation fermions can decay into lower generation fermions by emitting weakforce gauge boson W . Besides, each fermion also has an associated anti-

particle with the opposite sign electrical charge/color charge, which results from the conjugation of CP symmetry.

1.4.1 Quarks

The quark scheme was suggested by the symmetries in the way the many mesons and baryons seemed to be arranged in families. Theorists Gell-Mann and Zweig independently proposed in 1964 that just three fundamental "constituents" (and their anti-particles) combined in different ways according to the rules of mathematical symmetries could explain the whole zoo. Gell-Mann called these constituents quarks, and the three types were named up, down and strange quarks. Evidence for quark-like constituents of protons and neutrons became clear in the late 1960s and 1970s. In 1974, a new particle was unexpectedly discovered at SLAC (Stanford Linear Accelerator Center). It was given the unwieldy dual name J/Ψ , because of its simultaneous discovery by two groups of experimenters. The J/Ψ was later shown to be a bound state of a completely new quark-antiquark pair, which had been predicted on the basis of a subtle phenomenon. The new fourth quark was named charm.

The four-quark scheme was extended to its present state of six quarks by the addition of a new pair, bottom and top. In 1977, the bottom quark was observed by a team at Fermilab led by Leon Lederman. A new heavy meson called the Υ was discovered at Fermilab and later shown to be the bound state of the bottom and anti-bottom quark pair. The B meson, containing an anti-b quark and a u or d quark was discovered by the CLEO experiment at Cornell in 1983. This was a strong indicator of the top quark's existence: without the top quark, the bottom quark would have been without a partner. However, it was not until 1995 that the top quark was finally observed, also by the teams at Fermilab. It had a mass much larger than had been previously expected. So now we have the six quarks: Up, Down, Strange, Charm, Bottom and Top quarks and they each have their partner anti-quarks. The quarks are usually labeled by their first letters: u, d, s, c, b and t. In various combinations they make up all the mesons and baryons that have been seen.

Quarks have fractional charge $\pm 2/3$ for Up, Charm, and Top quarks, $\pm 1/3$ for Down, Strange, and Bottom quarks. In addition, each of these quarks also has an anti-particle making for a total of 12 quarks. Quarks also have a color charge associated-red, green, or blue. Due to a phenomenon known as quark color confinement, no quarks can exist freely after a second after the Big Bang. Instead, quarks group together in color neutral combinations known as hadrons. Three quarks, one of each color in the combination, group to form baryons, while two quarks, color and anti-color, group to form mesons. The quarks in a given hadron constantly exchange gluons. For this reason, physicists talk about the color-force field which consists of the gluons holding the bunch of quarks together. If one of the quarks in a given hadron is pulled away from its neighbors, the color-force field "stretches" between that quark and its neighbors. In so doing, more and more energy is added to the color-force field as the quarks are pulled apart. At some point, it is energetically cheaper for the color-force field to "snap" into a new quark-antiquark pair. In so doing, energy is conserved because the energy of the color-force field is converted into the mass of the new quarks, and the color-force field can "relax" back to an unstretched state. Since quarks have both color charge and electric charge, they interact via the strong and electroweak forces.

As we know now, the up and down quark are the two constituents of the proton and neutron: the proton contains two up quarks and one down quark, while the neutron contains one up quark and two down quarks. And because this is the baryon system, each quark comes in three colors. And we may wonder what if one of the down quarks in neutron changes to up quark, the neutron will become a proton, associated with a W boson. This is known as the neutron decay. We will talk about it in [Chapter 1.5](#).

1.4.2 Leptons

The leptons, known as electrons, muons, and taus, each have a charge of ± 1 and each has a neutrally charged neutrino associated. Charged leptons interact via the electroweak(electromagnetic and weak) force while neutral leptons interact only via the weak force, making them much more difficult to detect.

Only the electron, muon and neutrino were known before the 1960s. These behave differently from the mesons and baryons. First, they are much less massive. The mass of the electron is almost 2,000 times smaller than the mass of the proton, and the muon appears to be just a heavier version of the electron, its mass being nine times smaller than that of the proton. The neutrino has almost no mass at all, and up until recently, its mass was thought to be truly zero. Hence the name "leptons" or light particles. Second, the electron and muon interact with matter mainly through their electric charges; the neutrino being neutral, hardly at all. They all have a weak interaction with the matter in nuclei and, in high energy collisions, they do not produce the profusion of new mesons and baryons that protons and neutrons do when colliding with nuclei. In 1962, the first experiment using a high-energy neutrino beam showed that the electron has its own electron-neutrino, and the muon has its own distinct muon-neutrino. This was the very first evidence that there could be families or generations of pairs of fundamental particles. This notion was dramatically extended in 1975, when shortly after the discovery of the J/Ψ , a new heavy lepton was discovered by Stanford Linear Accelerator Center (SLAC), called the tau, almost twice as massive as the proton, but behaving like the other leptons, sharing the weak interaction property! This was the first evidence that three pairs or families of leptons existed: the electron and electron-neutrino, the muon and muon-neutrino and the tau and tau-neutrino.

Like the quarks, leptons also have their own anti-particles. The antiparticle of electron is the positron, which was the first antiparticle observed.

However, the fundamental questions for the fermions still remain: why are there quarks and leptons, with different charges and interaction characteristics? Why are there three generations, and so many different masses?

1.5 Bosons and Forces

There are four fundamental forces in the Standard Model: the electromagnetic force, the weak force, the strong force and the gravitational force. In our macroscopic world, the gravitational and electromagnetic forces are obvious, from the falling of an apple, or the attraction between

magnets. The reason is that the two are long-ranged forces, here long means much larger than the size of nucleus. However the strong and weak forces are not noticeable in daily life, because of the short-ranged characteristic.

It was a long process that the concept of force has grown from the object exerting force upon each other into the concept of field, which is the exchange of particles(gauge bosons). Each of the four different kinds forces has an associated force carrier particle accordingly. One important thing to know about force carriers is that a particular force carrier particle can only be absorbed or produced by a matter particle which is affected by that particular force. As we know about the quarks and leptons, all quarks can interact via the electromagnetic, weak, and strong interactions, while all leptons can interact through the weak force, among them the charged leptons can additionally interact through the electromagnetic force.

The properties of the four interactions can be seen in Figure 1.4.

Properties of the Interactions				
The strengths of the interactions (forces) are shown relative to the strength of the electromagnetic force for two u quarks separated by the specified distances.				
Property	Gravitational Interaction	Weak Interaction (Electroweak)	Electromagnetic Interaction	Strong Interaction
Acts on:	Mass – Energy	Flavor	Electric Charge	Color Charge
Particles experiencing:	All	Quarks, Leptons	Electrically Charged	Quarks, Gluons
Particles mediating:	Graviton (not yet observed)	W^+ W^- Z^0	γ	Gluons
Strength at $\begin{cases} 10^{-18} \text{ m} \\ 3 \times 10^{-17} \text{ m} \end{cases}$	10^{-41} 10^{-41}	0.8 10^{-4}	1 1	25 60
© 2016 Contemporary Physics Education Project CPEPphysics.org				

Figure 1.4: Properties of the interactions

To be specific, the fundamental forces of the SM are mediated by gauge bosons – the photon(γ), $W^\pm$, Z^0 , and the gluon(g) as following.

1.5.1 Electromagnetic force with photon γ

The electromagnetic force causes like-charged particles to repel and oppositely-charged particles to attract. Many forces in everyday life, such as friction, and even magnetism, are caused by the

electromagnetic force. Besides, the electromagnetic force also contribute to bind atoms together. Atoms usually have the same numbers of protons and electrons. They are electrically neutral, therefore, because the positive protons cancel out the negative electrons. Since they are neutral, what causes them to stick together to form stable molecules? The answer is that we've discovered that the charged parts of one atom can interact with the charged parts of another atom. This allows different atoms to bind together, which is the effect of residual electromagnetic force. So the electromagnetic force is what allows atoms to bond and form molecules, allowing the world to stay together and create the matter you interact with all of the time.

Electromagnetic interactions always involves a photon that's either absorbed or emitted. This is the first discovered spin 1 particle. It's thought that the mass of photon is 0 (but an extremely small mass, less than $6 \times 10^{-16} \text{eV}$, is still possible experimentally). The photon couples to any non-zero charged particles, including the fermions and the charged gauge bosons.

1.5.2 Weak force with $W^\pm$ and Z^0 boson

There are six kinds of quarks and six kinds of leptons. But all the stable matter of the universe appears to be made of just the two least-massive quarks (up quark and down quark), the least-massive charged lepton (the electron), and the neutrinos. Weak interactions are responsible for the decay of massive quarks and leptons into lighter quarks and leptons. The only matter around us that is stable is made up of the smallest quarks and leptons, which cannot decay any further. When a quark or lepton changes type, it is said to change flavor. All flavor changes are due to the weak interaction.

The weak interaction always involves the so-called vector boson $W^\pm$ and Z^0 boson. There are three of them, two charged (W^+ and W^-) and one neutrally charged (Z^0). They all are very heavy and have the mass 80.39 GeV for charged boson and 91.19 GeV for the neutral boson respectively. W^+ and W^- are each other's antiparticle while Z^0 is its own antiparticle. Besides coupling with the fermions, those vector bosons also couple to each other, and the charged boson W^+ and W^- also couple to the photon.

1.5.3 Strong force with gluon g

To understand what is happening inside the nucleus, we need to understand more about the quarks that make up the protons and neutrons in the nucleus. Quarks have electromagnetic charge, and they also have an altogether different kind of charge called color charge. The force between color-charged particles is very strong, so this force is called "strong" force.

The strong force holds quarks together to form hadrons, so its carrier particles are intuitively called gluons because they so tightly "glue" quarks together. Color charge behaves differently than electromagnetic charge. Gluons, themselves, have color charge, which is weird and not at all like photons which do not have electromagnetic charge. And while quarks have color charge, composite particles made out of quarks have no net color charge (they are color neutral). For this reason, the strong force only takes place within the short distance of quark interactions, which is why we are not aware of the strong force in everyday life. Because quarks and gluons are color-charged particles, and just like electrically-charged particles interact by exchanging photons in electromagnetic interactions, color-charged particles exchange gluons in strong interactions. When two quarks are close to one another, they exchange gluons and create a very strong color force field that binds the quarks together. The force field gets stronger as the quarks get further apart. Quarks constantly change their color charges as they exchange gluons with other quarks.

There are eight kinds of gluons according to the color charge combinations and the interactions are complicated via the Quantum Chromodynamics(QCD) process. Each gluon is characterized by one color and an anti-color. As we know that there are three different colors, so in total there should be 9 different gluons. But there is one superposition, known as the white gluon, does not exist. Hence in total we have 8 different gluons. Besides the strong interactions with quarks, those gluons can also couple with each other. But they are electrically neutral, so there is no interaction with photon.

So now we know that the strong force binds quarks together because quarks have color charge. But that still does not explain what holds the nucleus together, since positive protons repel each other with electromagnetic force, and protons and neutrons are color-neutral. The reason is the

strong force between the quarks in one proton and the quarks in another proton is strong enough to overwhelm the repulsive electromagnetic force. This is called the residual strong interaction, and it is what "glues" the nucleus together.

1.5.4 Higgs

The Higgs boson was the last particle of the Standard Model to be discovered. It is a critical component of the Standard Model. Its discovery helps confirm the mechanism by which fundamental particles (quarks, leptons, and force-carrier particles) get mass.

The basic symmetry of the electroweak theory (the unification of electromagnetic and weak interaction) is broken as the temperature drops and the forces separate in strength as the bosons gain mass. The reason that causes this is actually the Higgs field. It is possible to visualize how this works. We can treat mass as a manifestation of inertia or resistance to acceleration. If a Higgs field suddenly permeates all of space as the Universe cools, it can act as a drag on every particle moving in space, the drag depending on how well each interacts with the Higgs field. This drag shows up as inertia and thus a measurable mass of the particles that were originally massless. And this is the boson in standard model that carries this field – the Higgs boson.

Higgs interaction always involve the neutral spin 0 scalar boson, known as Higgs boson, which is first predicted in 1964 [58] and then discovered in year 2012 at the Large Hadron Collider (LHC), with a mass of 125 GeV [16] [38] and Francois Englert and Peter W. Higgs were awarded the Nobel Prize of year 2013 for their proposed Higgs theory.

In gauge theory, the gauge fields of the Z and W bosons are required to be massless. Physically, the Higgs mechanism is responsible for the spontaneous symmetry breaking, leading to Higgs' non trivial vacuum expectation values, without which all bosons would be massless and identical just like at the beginning of our universe. All massive particles (including Higgs itself) obtain their mass by interacting with the Higgs field. And the discovery of Higgs is the big success of the Standard Model and the ATLAS and CMS collaboration. The strength of Higgs interaction with any particle is proportional to the mass of that particle and its strength is very weak when compared with other

forces. The Higgs can couple to both bosons and fermions, with "Higgs coupling" governing the Higgs interaction with bosons, while the couplings to quarks and leptons are known as "Yukawa couplings".

1.6 Lepton Flavor

And we can easily find there is a big difference between the mass of the proton and that of the electron, actually the proton is 1800 times heavier. So energetically it's much easier for a proton to decay to a positron, but in fact it doesn't happen(luckily), because of some basic conservation rules, including the "Lepton Number Conservation", are obeyed(or not violated) by the Nature.

The Lepton Number Conservation is about the lepton number should be conserved in every process. As we know in the Figure 1.3, there are the quarks and leptons. And those leptons are assigned the following numbers:

- $L = +1$ for leptons(e^- , μ^- , τ^- , ν)
- $L = -1$ for anti-leptons(e^+ , μ^+ , τ^+ , $\bar{\nu}$)
- $L = 0$ for all non-leptons(like quarks and bosons)

So this lepton number conservation is the reason why the process that proton decay to the much lighter lepton directly is prohibited.

Another example is the neutron decay process, as shown in Figure 1.5.

The beta decay is the process that neutron may decay into a proton, electron and an anti-electron-neutrino. So we can see that both initial and final states have the total lepton number 0. Actually the lepton number conserved here is specifically the electron lepton number.

We know from the Standard Model that there are three generations for leptons. Besides the electron in the first generation, there also exist the muon and tau as the second and third generation. It is thought that not only the total lepton number should be conserved, but also the lepton number of each specific generation should be conserved as well, and we use the term "flavor" to distinguish each lepton pair as electron flavor, muon flavor and tau flavor for those three generations. Obviously,

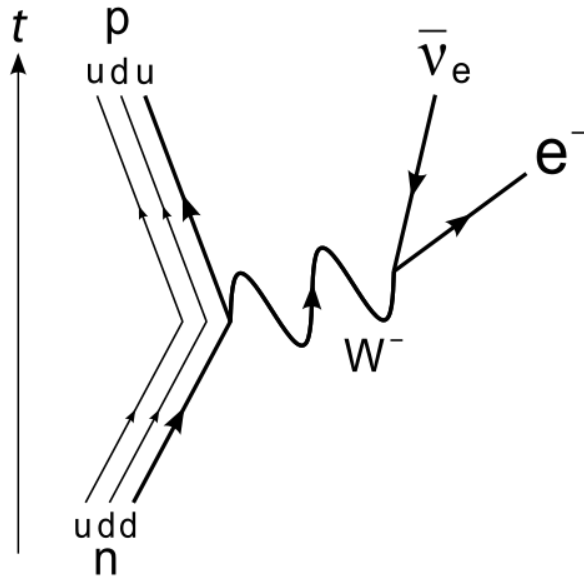


Figure 1.5: Beta decay

total lepton number conservation is only a direct consequence of the conservation of each lepton flavor (electron, muon and tau number).

Interestingly, the beta decay is the first experiment that we discover the existence of neutrino, which in return may indicate the lepton flavor can be violated in some way. If neutrinos have masses, that probably implies a breakdown of the conservation of individual leptonic quantum number (lepton flavors) while not affecting the total lepton number. And this is the topic of our analysis, Lepton Flavor Violation.

1.7 Lepton Flavor Violation and Searches

Isidor Isaac Rabi's famous question about the muon's existence "Who ordered that?" was prescient and deep. His question, in modern terms, asked why are there flavors and generations? Why are there muons and taus in addition to the electron? The same question applies to the quark and neutrino sectors. We believe there are three generations in each sector, and that the number in each sector must be the same. We see quarks changing generations, as codified in the Cabibbo-Kobayashi-Maskawa (CKM) matrix, and neutrinos changing from muon to electron to tau neutrinos

according to the Pontecorvo-Maki-Nakagawa-Saka(PMNS) matrix. Lepton Flavor Violation is an established fact, but only in the neutral neutrinos. What about their charged partners? Is there Charged Lepton Flavor Violation?" [33]

As we already know that there are three generations of fermions in the Standard Model, and we have already observed the flavor changing effects in both quark sector(CKM matrix) and neutral lepton sector(neutrino oscillation, PMNS matrix, which indicates flavor violation in loops is allowed in the Standard Model). Lepton flavor number conservation is an assumption of the Standard Model(LFV), so is the total lepton number conservation. Specific lepton flavor conservation in the charged lepton sector is a fundamental assumption of the Standard Model but there is no associated symmetry. Searching for the lepton flavor violation phenomenon may lead to the discovery of physics beyond Standard Model. Moreover, in the neutrino sector, we have already observed the neutrino oscillation which indicate that the lepton flavor can be violated, however, the similar mechanism hasn't been found in the charged lepton sector yet [70]. All searches for charged LFV have produced null results so far [71]. The current best upper limit of $Br(Z \rightarrow e\mu)$ by the direct search was set by the ATLAS collaboration at 7.5×10^{-7} at 95% CL with Run1 data(at $\sqrt{s} = 8 \text{ TeV}$) corresponds to 7.8×10^8 Z bosons produced [14]. A summary of limits on various Lepton Flavor Violating processes is shown in Table 1.1.

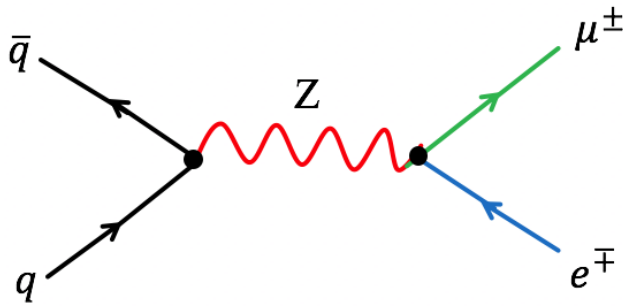
There are stringent experimental limits on other lepton flavor violating processes which can be used to derive an upper limit on $Z \rightarrow e\mu$ decay with some theoretical assumptions. For example, the upper limits on $\mu \rightarrow 3e$ yields $Br(Z \rightarrow e\mu) < 10^{-12}$ [47] and $\mu \rightarrow e\gamma$ yields $Br(Z \rightarrow e\mu) < 10^{-10}$ [44]. There are stringent limits on other LFV processes which can be used to derive an upper limit on $Z \rightarrow e\mu$ albeit with some theoretical assumptions [32] [69] [76].

Although the origin behind the neutrino lepton flavor violation may be different from the charged lepton sector, the search for this effect can also be conducted to set an upper limit constraint on theories beyond Standard Model [46] [54] [59]. LFV Z boson decays are predicted in models with heavy neutrinos [62], extended gauge models [64] and supersymmetry [49], which are able to increase the branching fraction of LFV Z decays to observable levels up to $\sim 10^{-8}$ [62] [86]

Table 1.1: Summary of limits on various LFV processes.

Decay	Upper Limit	Confidence Level(%)	Experiment
$\mu \rightarrow e\gamma$	1.2×10^{-12}	90	MEGA [23]
$\mu \rightarrow e\gamma\gamma$	7.2×10^{-11}	90	LAMPF [55]
$\mu \rightarrow 3e$	1×10^{-12}	90	SINDRUM [32]
$\tau \rightarrow \mu\gamma$	4.4×10^{-8}	90	BABAR [29]
$\tau \rightarrow e\gamma$	3.3×10^{-8}	90	BABAR [29]
$\tau \rightarrow 3e$	3.6×10^{-8}	90	BELLE [72]
$\tau \rightarrow 3\mu$	3.2×10^{-8}	90	BELLE [72]
$Z \rightarrow e\mu$	1.7×10^{-6}	95	OPAL [24]
$Z \rightarrow e\mu$	7.5×10^{-7}	95	ATLAS [14]
$Z \rightarrow e\tau$	9.8×10^{-6}	95	OPAL [24]
$Z \rightarrow \mu\tau$	1.2×10^{-5}	95	DELPHI [45]
$K_L^0 \rightarrow e\mu$	4.7×10^{-12}	90	E391a [75]
$D^0 \rightarrow e\mu$	8.1×10^{-7}	90	BABAR [28]
$B^0 \rightarrow e\mu$	6.4×10^{-9}	90	CDF [21]

Several BSM scenarios raise the branching ratios to levels reachable at the LHC, and because their Standard Model branching ratios are far too tiny for possible detection, any observation of LFV in the charged lepton sector is a discovery. That's what makes such sensitive searches potentially transformative [50]. (A "branching ratio" is the probability that a particle will decay via a given decay channel. These ratios are predicted by the Standard Model.) And the analysis is shown in the Figure 1.6.

Figure 1.6: $Z \rightarrow e\mu$ decay

For this analysis, the $140 \pm 2.9 \text{ fb}^{-1}$ of data collected at $\sqrt{s} = 13 \text{ TeV}$ by the ATLAS detector starting from 2015 to 2018, corresponds to a total of 8.4×10^9 Z bosons have been produced. Despite

the large background at the LHC, a better expected sensitivity for $Z \rightarrow e\mu$ process is achieved when compared with Run1 $Z \rightarrow e\mu$ analysis and the OPAL analysis.

CHAPTER 2. THE LHC AND ATLAS DETECTOR

Throughout the history of Physics, experimental discoveries and theoretical ideas and explanations have moved forward together, always drawing inspiration from the another.

Starting from Rutherford's table-top experiment, physicist began to explore the atomic and sub-atomic structures with different probes. The problem with using waves to detect the physical world is that the quality of your image is limited by the used wavelength. The wavelength of visible light is too long to analyze anything smaller than a biological cell. To observe objects with higher magnification, you must use waves with smaller wavelengths. That's why people turn to scanning electron microscopes when studying sub-microscopic objects like viruses. However, even the best scanning electron microscope can only show a fuzzy picture of an atom. Things with shorter wavelengths can provide you with more detailed information about what they hit. The shorter the probe's wavelength is, the more information you can get about the target. As all particles have wave properties, so when using a particle as a probe, we need to use particles with short wavelengths to get detailed information about small things. Roughly a particle can only probe down to distances equal to the particle's wavelength. To probe down to smaller scales, the probe's wavelength has to be made smaller. A particle's momentum and its wavelength are inversely related. High-energy physicists apply this principle when they use particle accelerators to increase the momentum of a probing particle, thus decreasing its wavelength.

The collision of particles at high energy, either with other particles or with a stationary target, allows physicists not only to look at what's inside these particles, but also to use the energy of their collisions to create different, more massive and more exotic particles of matter. When physicists want to use particles with low mass to produce particles with greater mass, all they have to do is put the low-mass particles into an accelerator, give them a lot of kinetic energy, and then collide them together. During this collision, the particle's kinetic energy is converted into the formation

of new massive particles according to Einstein's famous equation that $E = mc^2$. It is through this process that we can create massive unstable particles and study their properties (In nature, objects tend to find the state of lowest energy, so the high energetic massive particle can not last long. If they have excess energy, they tend to disintegrate or decay to more stable particles). Each collision is very complicated since lots of particles are produced. Most of these particles have lifetimes so short that they travel only an extremely short distance before decaying into other particles, and therefore leave no detectable tracks. Modern physicists will look at particles' decay products, from which to deduce the particles' existence.

To create such high-energy collisions, scientists must use very powerful particle accelerators. Modern versions of Rutherford's table-top experiment on the scattering of alpha particles occupy a large-scale of land, with massive and costly apparatuses in underground tunnels from hundred meters to kilometers long. These are the particle accelerators that speed protons, anti-protons, electrons, or positrons to near the speed of light and then make them collide head-on with each other or with stationary targets.

The quest has mostly been for higher and higher collision energies. On the other hand, to look for rare phenomena, it is necessary to increase the intensity of particle beams and the collision rates. So accelerators have proceeded along parallel paths of ever higher energies and ever higher intensities.

To observe and interpret the results of collisions, particle detectors have to be developed that can track and analyze the particles that decay and disappear in picoseconds or femtosecond. The detector consists of many different types of complex apparatuses and electronics, requiring a cadre of experts in every corresponding technology. Collider experiments use large detectors completely surrounding the interaction point where high energy particles collide head-on.

The art and science of particle accelerators and detectors has depended heavily on technology. The technology of solid state devices, superconducting magnets, electronics, computers and exotic materials, all have been utilized in experimental particle physics, sometimes driving and sometimes being driven by the inventions of particle physicists.

All these very complex detectors are built and operated by large numbers of physicists, in collaborations ranging from hundreds to thousands. The collaborations extend across boundaries of countries and continents, in a typical illustration of science extending the hand of cooperation and friendship across national and political barriers. And CERN is the biggest High Energy Physics organization in the world.

2.1 CERN and LHC

The European Organization for Nuclear Research (CERN) was founded in 1954. It was designed to enable physicists and engineers to probe extreme particle energy ranges at the very limit of what was possible to reach with current technologies. One of CERN's primary goals is to provide particle accelerators for high energy research. These accelerators are used in a wide variety of experiments, with a major focus currently being the LHC. CERN is a multinational organization, which has a total of 22 member states, and there are more than 10,000 scientists working and contributing to the organization.

At CERN, the Large Hadron Collider (LHC) is currently the largest accelerator in the world and it is around 27 kilometers in length and sits between 50 and 175 meters below the surface, as shown in Figure 2.1 and Figure 2.2. It is located in the tunnel originally dug for the Large Electron Positron Collider (LEP) which operated from 1989 until 2000, when it was dismantled. The accelerator complex at CERN is a succession of machines that accelerate particles to increasingly higher energies. Each machine boosts the energy of a beam of particles, before injecting the beam into the next machine in the sequence. The Large Hadron Collider (LHC) is the last element in this chain. The protons are injected into the LHC at an energy of 450 GeV and are then boosted to approximately 6.5 TeV. Those protons will be accelerated or decelerated into discrete bunches of particles. There are two beams of proton bunches which move in opposite directions along the beam-line allowing for a total center-of-mass energy of 13 TeV at the interaction points, which is shown in Figure 2.3. The design collision rate is roughly 40 million bunch crossings per second [83]. Each of the beams travels in its own beam pipe. The beam pipes themselves are held in

vacuum, and superconducting magnets are used to turn the proton beams around the loop. Most of the other accelerators in the chain have their own experimental halls where beams are used for experiments at lower energies.

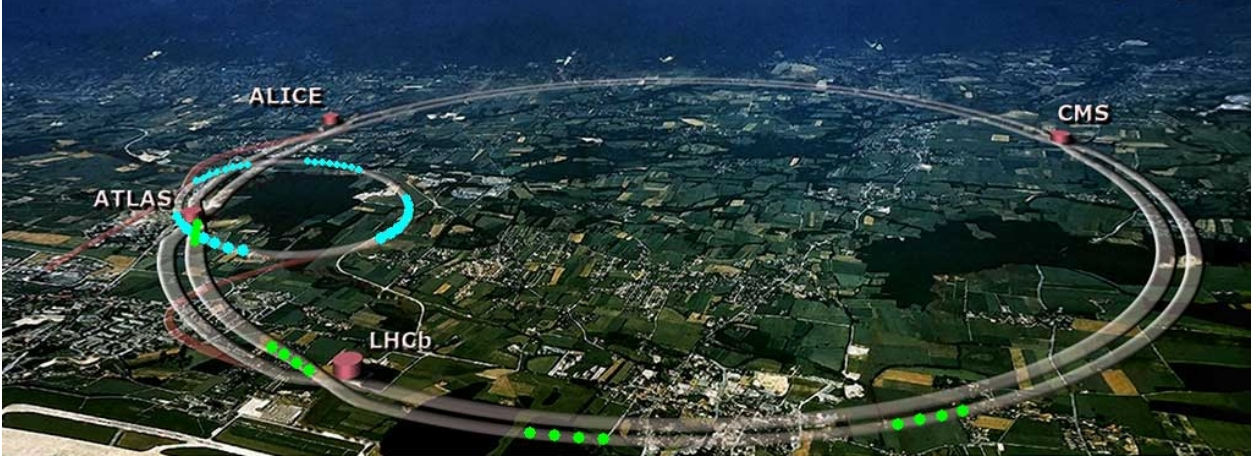


Figure 2.1: Overall view of the LHC experiment

The LHC started proton-proton (pp) collisions in 2011 at a center-of-mass energy of $\sqrt{s} = 7$ TeV. In 2012 the center-of-mass energy was increased to 8 TeV. This was called Run 1 and lasted until 2013. At the date of this publication, the LHC has completed Run 2, which began in 2015 and lasted until the end of 2018. The analysis described in this document has been performed using the data collected during the whole Run2 period. The increase in available energy in the center-of-mass frame increases the cross-section for many physics processes, including Z boson production in our analysis, as shown in Figure 2.4. The cross-section is a measure of the probability for that process to occur during any proton-proton collision. Processes with larger cross-sections occur more often than processes with small cross-sections.

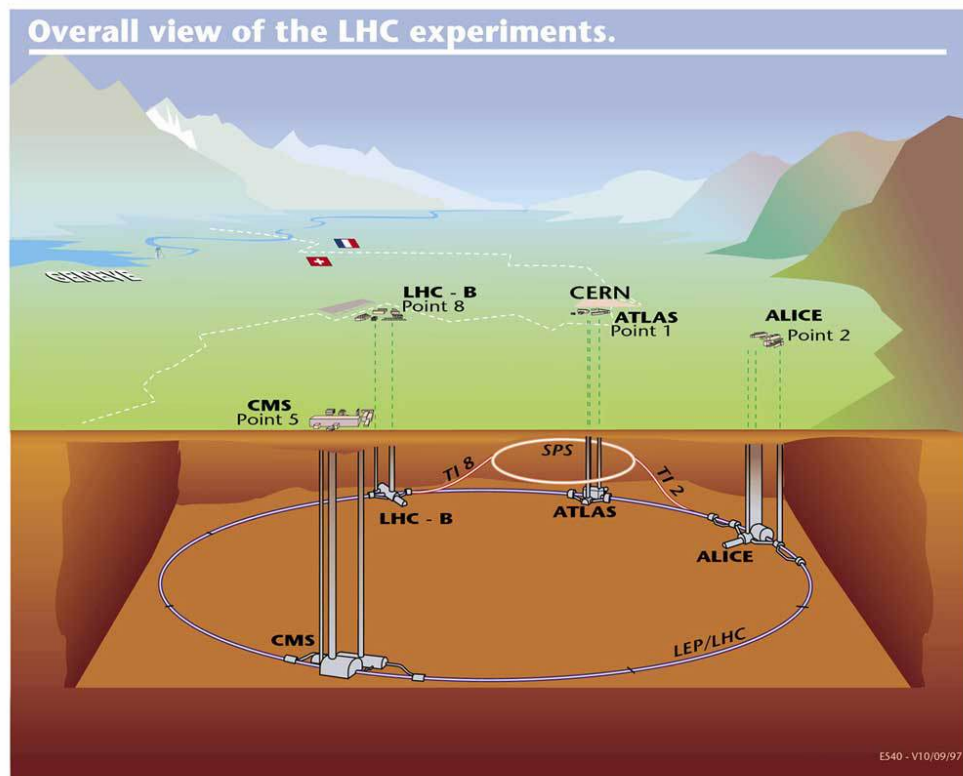


Figure 2.2: Overall view of the LHC experiment

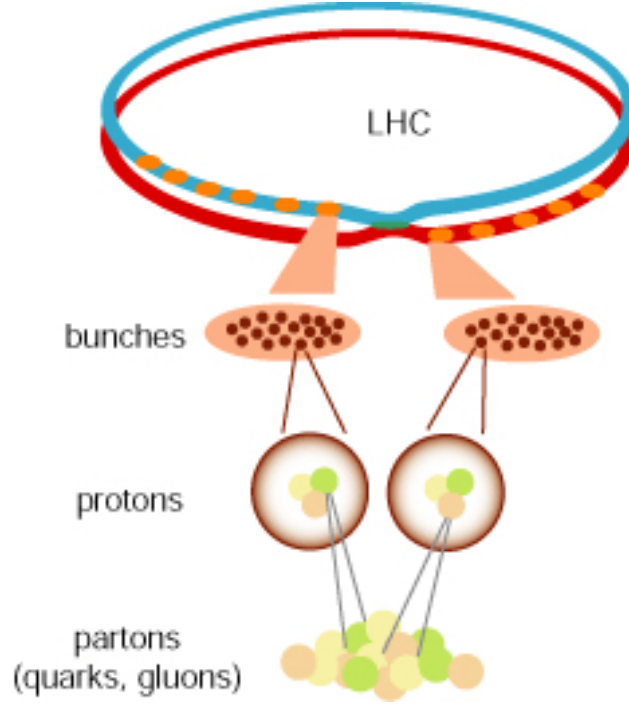


Figure 2.3: Proton-Proton collision

The target energies and integrated luminosities of the LHC project, as well as the scheduled data-taking and upgrade phases are shown in Figure 2.5. Luminosity is a number of events per unit time divided by the cross section. It can be calculated by the following Equation 2.1.

$$\mathcal{L}_{total} = \frac{\langle \mu \rangle \cdot n_b \cdot f_r}{\sigma_{total}} \quad (2.1)$$

Luminosity at the LHC is obtained by measuring [13] the number of inelastic interactions per bunch crossing $\langle \mu \rangle$ which is shown in Figure 2.6, the number of bunches crossing per unit time $n_b \cdot f_r$, and the proton-proton inelastic total cross-section σ_{total} , which is shown in Figure 2.4 and it is calculated by summing up all the accessible process in the Feynman graphs. The probability for any given processes to occur is in proportion to its cross section. Thus, specifying the cross section σ for a given reaction can be treated as a proxy for stating the probability that

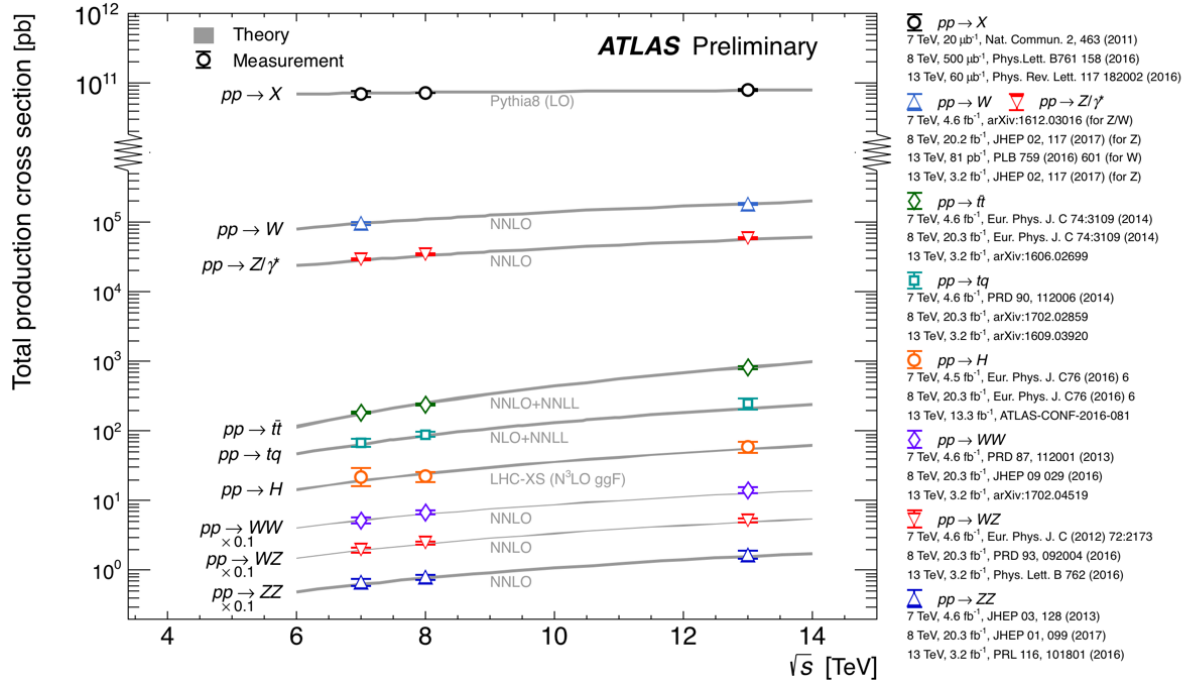


Figure 2.4: Production Cross-Section along center-of-mass energy $\sqrt{s}$

a given scattering process will occur (e.g. about once in every 10 billion collisions to produce a Higgs boson, which is $pp \rightarrow H$ in Figure 2.4 along with specific center-of-mass energy $\sqrt{s}$). And finally luminosity is integrated over the total number of bunches n_b or over time to give the total luminosity over a data collection period.

There are four main detectors associated with the LHC as seen in Figure 2.2: ATLAS, ALICE, CMS, and LHCb. Next we are going to introduce our detector, ATLAS.

2.2 ATLAS Detector

Unveiling the tiniest constituents of matter with accelerators is only half the battle. Physicists also need extraordinary particle detectors to observe what happens in high-energy collisions.

To look for these various particles and decay products, physicists have designed multi-component detectors that test different aspects of an event. Each component of a modern detector is used for

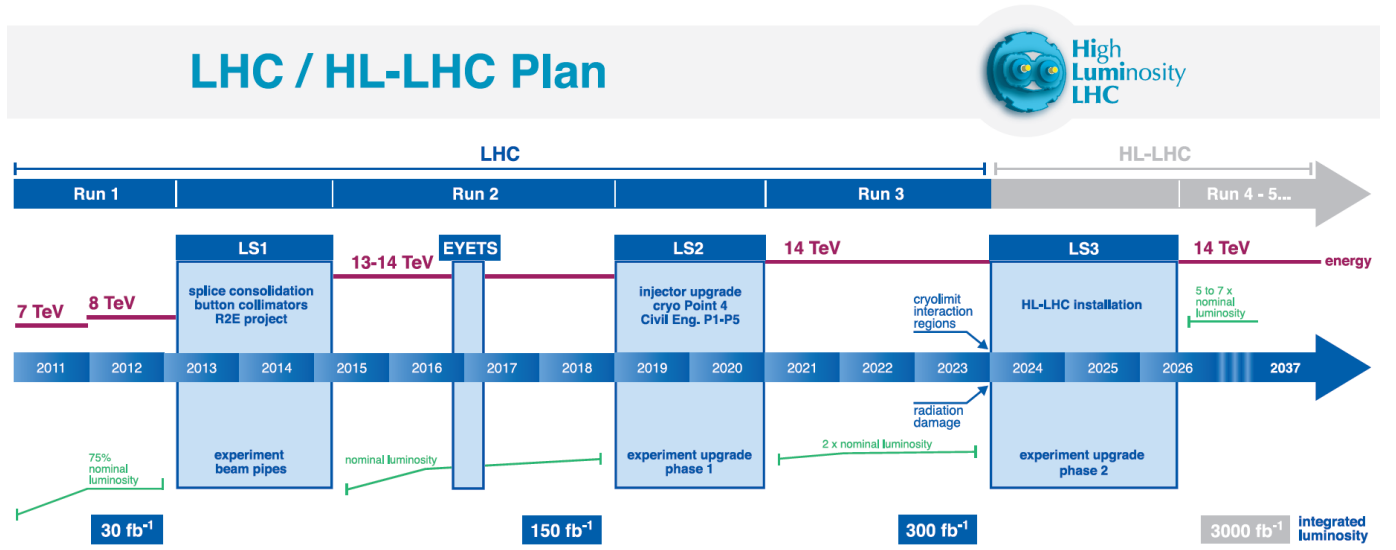


Figure 2.5: LHC time-line

measuring particle energies and momenta, and/or distinguishing different particle types. When all these components work together to reconstruct an event, individual particles can be singled out from the multitudes for analysis. Specifically, detectors are instruments that count particles, visualize tracks, measure particle energies, record time of flight and identify different particle types. Following each event, computers collect and interpret the vast quantity of data from the detectors and present the extrapolated results to the physicist.

Depending on the type of accelerator and the particles and forces to be studied, physicists combine various detection devices arranged in intricate configurations. In the case of colliding beams, physicists build a detector surrounding the point at which the two beams collide. Like the layers of an onion, such a detector contains successive layers of detection devices with different functions. Close to the center, physicists place precision tracking instruments such as silicon detectors and wire chambers. Those instruments are usually surrounded by calorimeters that measure the energy of particles passing through. The outer shell of a detector, farthest from the collision area, is devoted to detecting muons, heavy electron-like particles that has a mean lifetime of 2.2×10^{-6} s. And due to their greater mass, muons are not as sharply accelerated when they encounter electromagnetic fields, and do not emit as much bremsstrahlung (deceleration radiation). This allows muons of

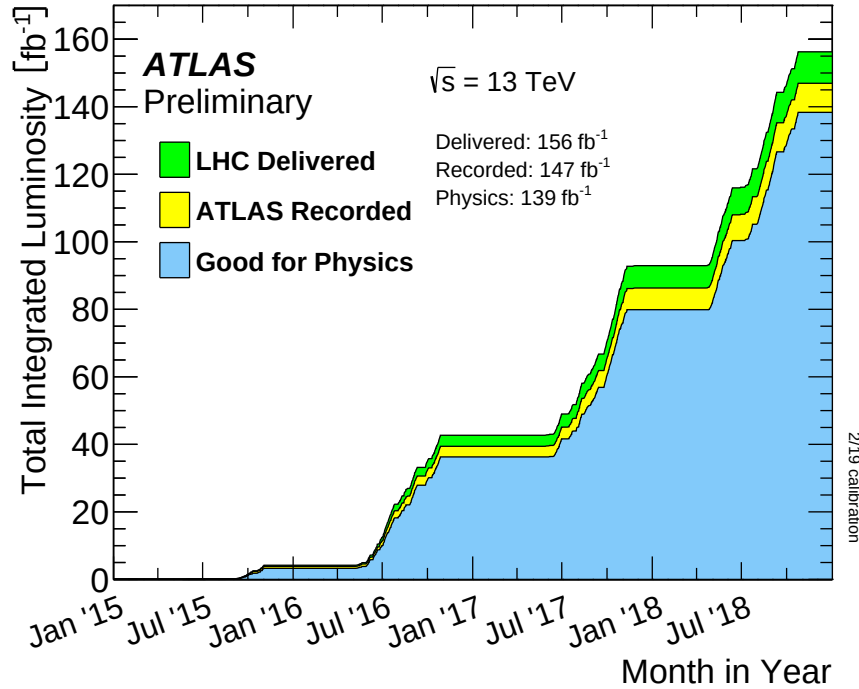


Figure 2.6: Collected luminosity along time

a given energy to penetrate far more deeply into matter than electrons since the deceleration of electrons and muons is primarily due to energy loss by the bremsstrahlung mechanism

One of the most important tasks of the detector electronics, called triggering, is the selection of collision signals that are interesting enough to be recorded permanently for later examination. It is unnecessary and impossible to keep the ten thousands of detector signals created every millionth of a second. Physicists build and program the detector hardware to perform several levels of go or no-go decisions before passing the data to the next, more sophisticated level of processing.

A Toroidal LHC ApparatuS (ATLAS) shown in Fig 2.7 is a one of the major experiments at the Large Hadron Collider (LHC) at The European Organization for Nuclear Research (CERN). The detector is constructed from three major sections each with the task of detecting some aspect of a collision event. The reason that detectors are divided into many components is that each component tests for a special set of particle properties. These components are stacked so that all particles will go through the different layers sequentially. A particle will not be evident until it

either interacts with the detector in a measurable fashion, or decays into detectable particles. The three sections of the ATLAS detector are the Inner Detector (ID), the Calorimeter system, and the Muon Spectrometer (MS) as shown in Figure 2.8. The ID is placed in a magnetic field generated by a solenoidal magnet and is responsible for tracking and identifying particles as well as finding decay vertices. There are two calorimeter systems, the electromagnetic and hadronic calorimeters, whose purposes are to absorb and measure the energies of the particles with which they interact. The outermost layer is the MS which is immersed in a magnetic field generated by toroidal magnets. The MS primary function is to detect muons. The geometry of each section is a cylindrical barrel extending radially from the beam-line with closed disk-shaped end-caps on each end perpendicular to the beam-line. Finally, there is a hardware based Level 1(L1) trigger system and a software based High-Level Trigger(HLT) consisting of Level 2(L2) and Event Filter(EF) triggers that are necessary to cope with the extremely high collision rate. We will introduce them separately in the following subsections.

Before describing the Inner Detector, we introduce the coordinate system of the detector.

A common coordinate system is used throughout ATLAS. The interaction point is defined as the origin of the coordinate system. The z-axis runs along the beam line. The x-y plane is perpendicular to the beam line and is referred to as the transverse plane. Particle momenta measured in the transverse plane are referred to as transverse momenta p_T while transverse energy is E_T . The transverse plane is often described in terms of $r - \phi$ cylindrical coordinates, where the azimuthal angle ϕ is measured from the x-axis, around the beam. The radial dimension, r , measures the distance from the beam line. The polar angle θ is defined as the angle from the positive z-axis. The transverse momentum is defined as $p_T = p \cdot \sin \theta$ and similarly the energy in the transverse plane is defined as $E_T = E \cdot \sin \theta$. The polar angle θ is often reported in terms of pseudorapidity η instead, defined as $\eta = -\ln \tan(\theta/2)$ and shown in Figure 2.9. The angular distance between objects ΔR is defined in η - ϕ space as $\Delta R = \sqrt{\Delta \eta^2 + \Delta \phi^2}$. Due to the fact that practically all of the momentum

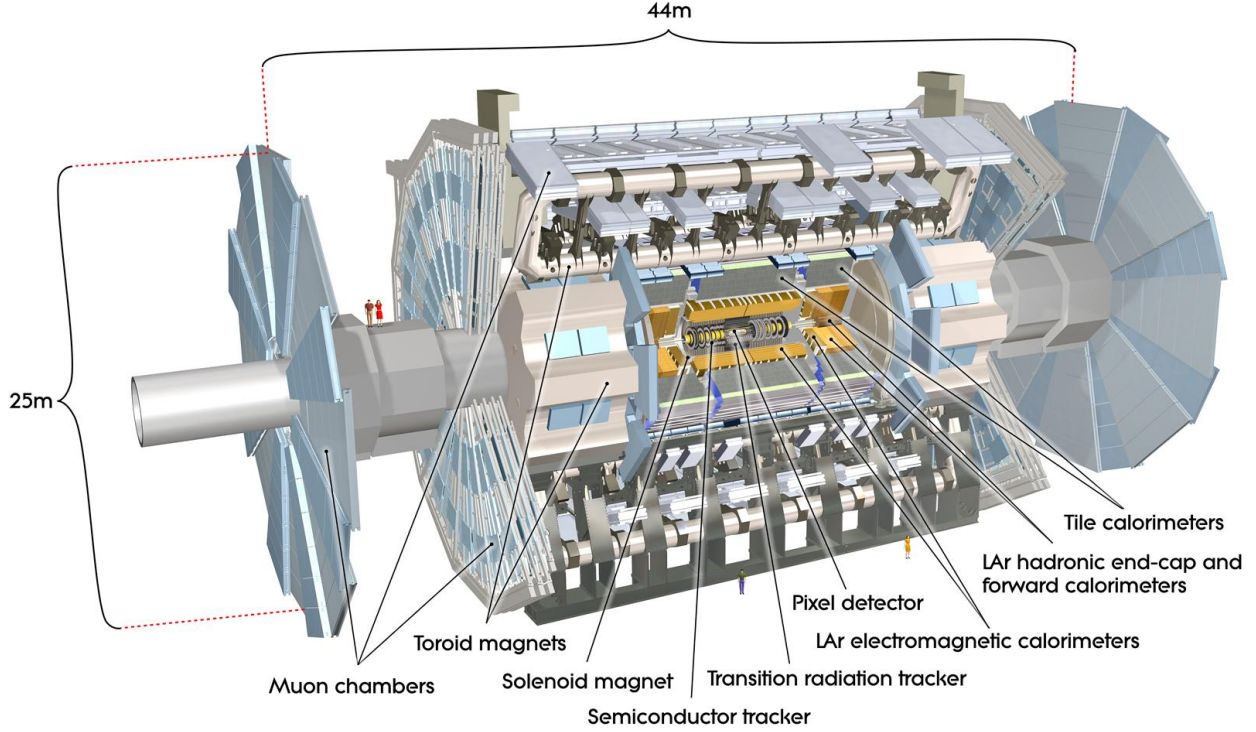
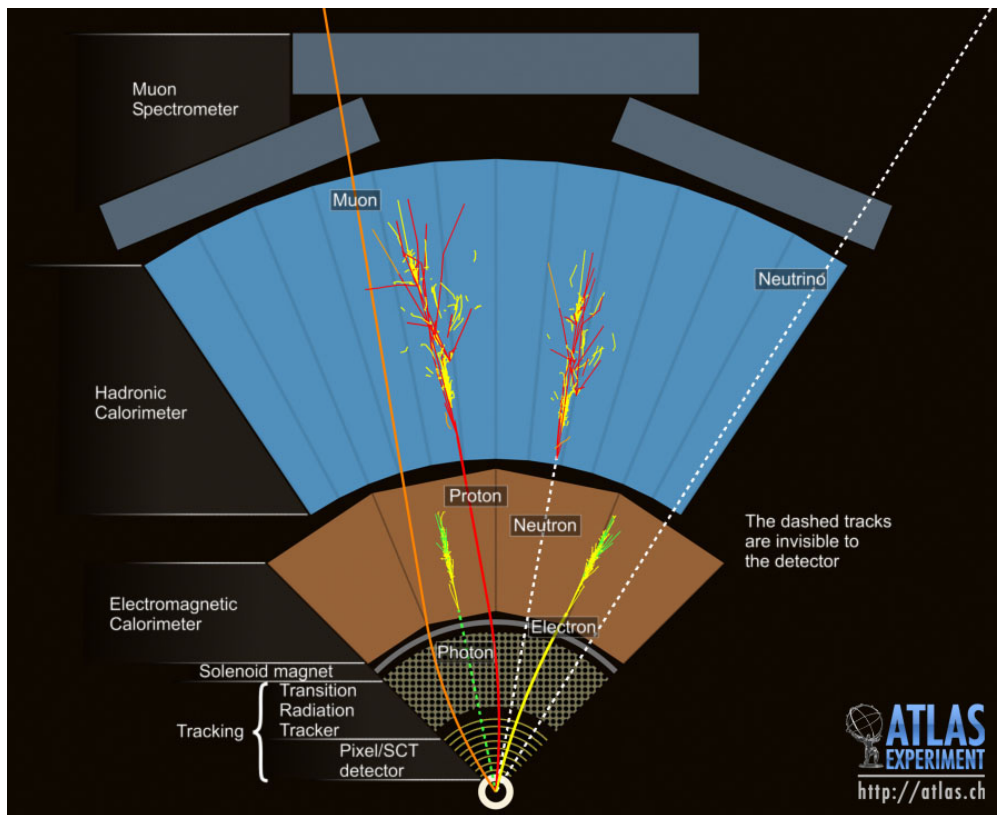


Figure 2.7: The four major components of the ATLAS detector are the Inner Detector, the Calorimeter, the Muon Spectrometer and the Magnet System. Integrated with the detector components are: the Trigger and Data Acquisition System, a specialized multi-level computing system, which selects physics events with distinguishing characteristics; and the Computing System, which develops and improves computing software used to store, process and analyse vast amounts of collision data at 130 computing centres worldwide.

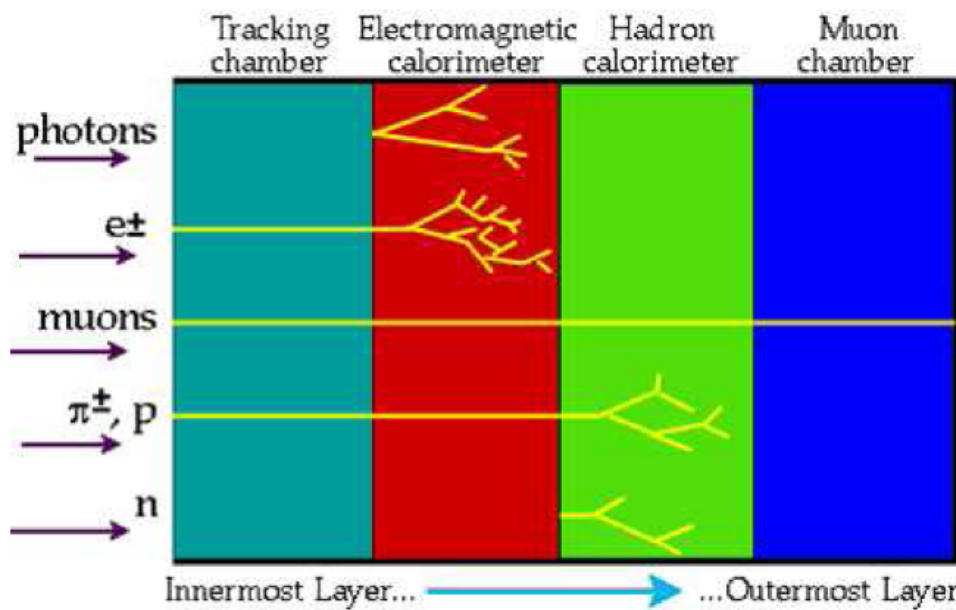
of the beams lies in the z -direction the momentum in the transverse plane is well constrained and is therefore well suited for analysis. Such quantities are often indicated by a capital T subscript (e.g E_T, p_T, m_T , etc)

2.3 Inner Detector

The Inner Detector is the first part of ATLAS to see the decay products of the collisions, so it is very compact and highly sensitive. As shown in Figure 2.10 and Figure 2.11, the Inner Detector is comprised of three separate parts:



(a) The ordering of the ATLAS sub detectors and the particles to which they are sensitive.



(b) The ordering of the ATLAS sub detectors and the particles to which they are sensitive.

Figure 2.8: Detector structure and particles identification.

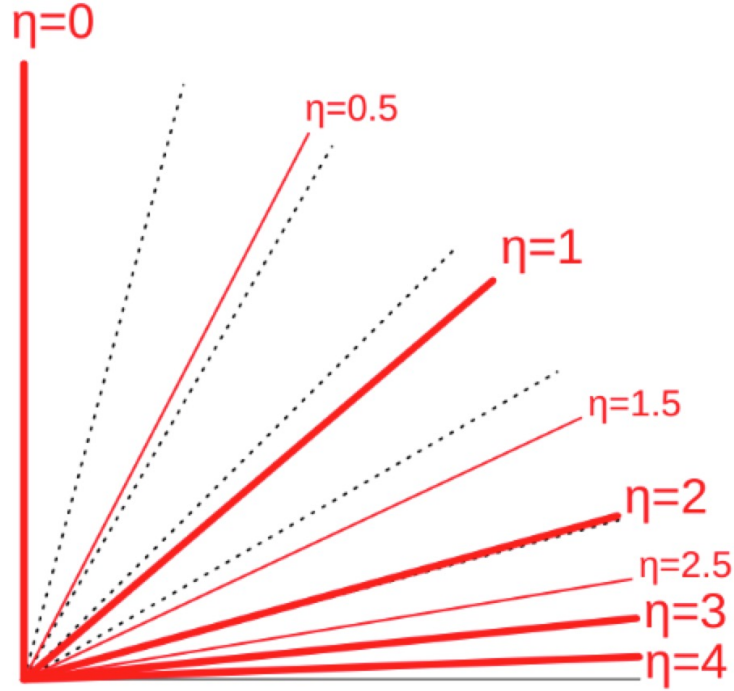


Figure 2.9: Values for pseudorapidity η between 0° and 90° . Dotted lines are shown at 15° intervals as a means of comparison.

- the Silicon Pixel Detector
- the SemiConductor Tracker (SCT)
- the Transition Radiation Tracker (TRT)

All of them are immersed in a 2 T magnetic field which is oriented parallel to the beam line. The magnetic field is generated by a superconducting solenoidal magnet which is placed between the ID and the Electromagnetic Calorimeter(EM Calo). This magnetic field is meant to curve the path of particles in the plane transverse to the beam.

The precision pixel and silicon microstrip (SCT) trackers with a very fine segmentation cover the pseudorapidity range up to $|\eta| < 2.5$. The precision tracking detectors are arranged on concentric cylinders around the beam axis while in the end-caps are located on disks perpendicular to the beam axis. The first layer of the pixel detector with highest granularity, so-called B-layer, is

very important for an excellent vertexing (to reconstruct the vertex which indicates tracks of all associated particles coming from the same interaction point). Typically three pixel layers and eight SCT layers are crossed by a good quality track.

Using reconstruction algorithms the trajectory of each charged particle in a given event is calculated. The trajectory is used to create an object called a track. The momenta of these tracks can then be calculated based on the curvature of their paths according to Equation 2.2, where p_T is the transverse momentum, B is the magnetic field, Q is the charge, and r is the radius of the curvature of the track. Then knowing this curvature provides information about each particles transverse momentum and charge which is crucial to particle identification.

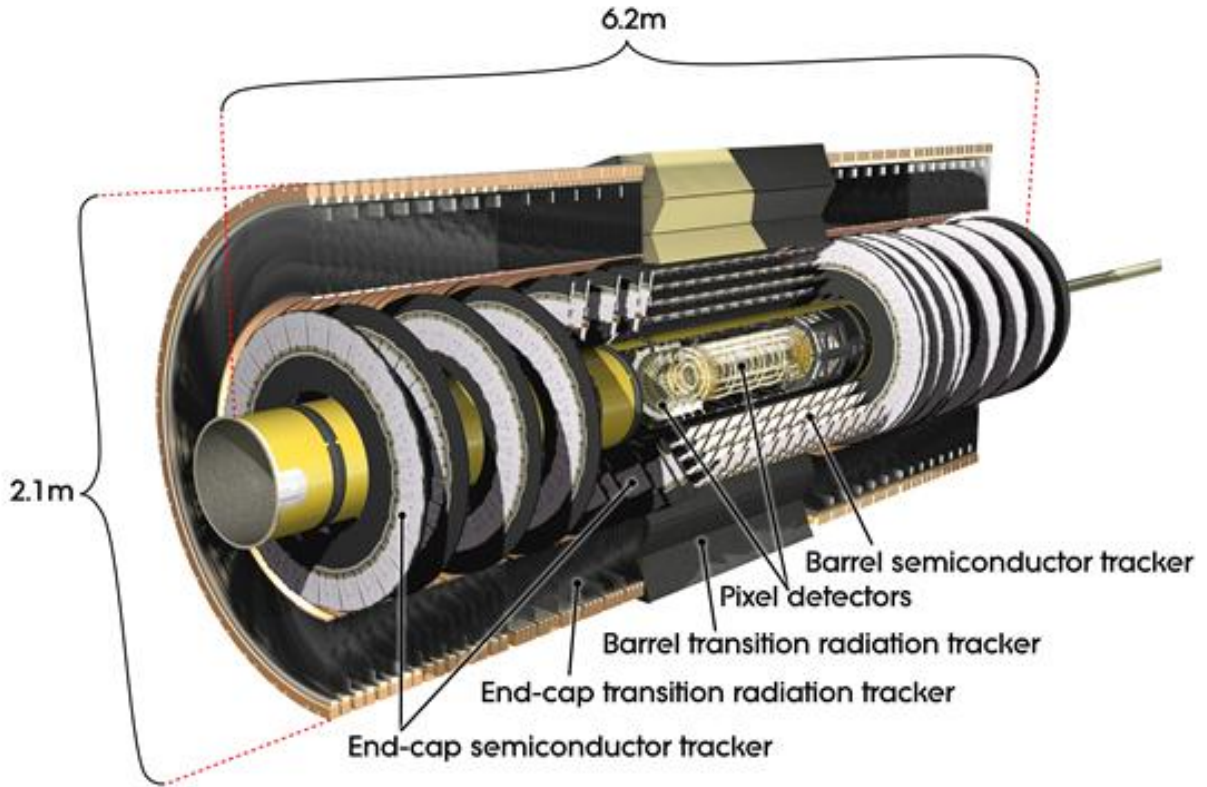


Figure 2.10: The main components of the Inner Detector are: Pixel Detector, Semiconductor Tracker (SCT), and Transition Radiation Tracker (TRT).

$$p_T = B \cdot Q \cdot r \quad (2.2)$$

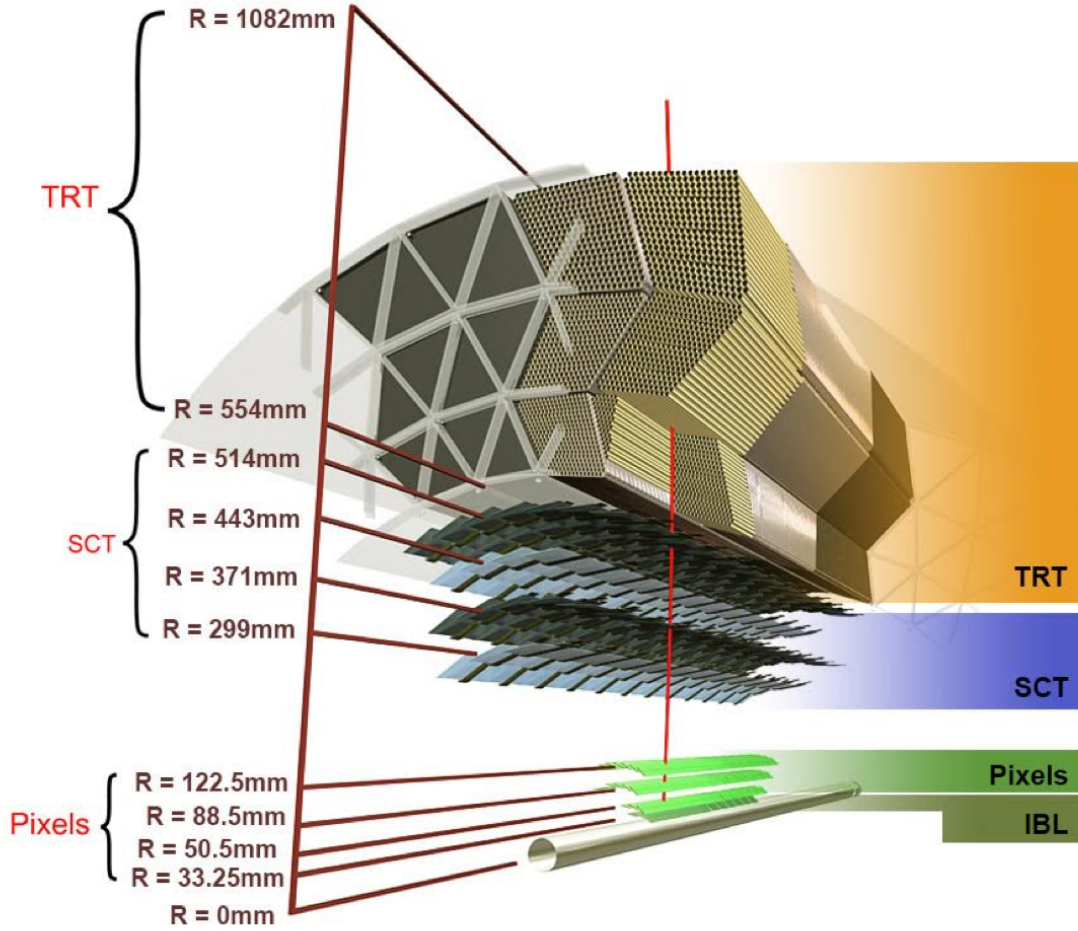


Figure 2.11: The cross-section of the Inner Detector.

A large number of hits, typically 36 per track, is measured with straw tubes of the Transition Radiation Tracker (TRT) which covers the pseudorapidity region up to $|\eta| < 2.0$ and creates the outermost part of the tracking detector. The TRT detector enables also electron vs. pion identification through the detection of transition radiation photons in the xenon-based gas mixture of its straw tubes. The tracking system has an expected resolution of $\sigma_{p_T} / p_T = 0.05 \%$ in the whole pseudorapidity coverage $|\eta| < 2.5$.

So, the Inner Detector overall measures the direction, momentum, and charge of electrically-charged particles produced in each proton-proton collision, which we have seen in Figure 2.8.

2.3.1 Pixel Detector

The innermost part of the ID is the pixel detector. The pixel detector is made of about 80 million pixel channels that are spread across three barrel sections and six disk sections (three on each side). Each pixel is a $50 \times 400 \mu m^2$ silicon strip. The barrels are arranged surrounding the beam at radii of 4, 10, and 13 cm. Pixel detectors are a form of semiconductor detector. By doping a narrow piece of silicon, one creates a diode which is then reverse biased. When a charged particle passes through the pixel (diode), an ionization current can be measured. Thus, with millions of extremely small pixels surrounding an interaction point one can track the path of a charged particle with high precision.

The setup of the Pixel Detector is:

- 80 million pixels (80 million channels). Area $1.7m^2$. 15 kW power consumption.
- Barrel has 1,744 modules ($10cm^2$) with 46,080 readout channels per module
- Pixel size $50 \times 400\mu m^2$. Resolution $14 \times 115\mu m^2$
- Three Pixel disks (in each endcap) have 6.6 million channels
- 3 barrel layers: 1,456 modules
- 3 disks in each end-cap: 288 modules

2.3.2 Semi-Conductor Tracker (SCT)

The middle ID subdetector is the semiconductor tracker. This detector is very similar to the pixel detector, but instead of using pixels, four layers of narrow silicon strips are used in the barrel section and nine disks of strips are used in the endcap. This subdetector also covers the region $|\eta| < 2.5$. This system is designed to measure the momentum, impact parameter, and vertex

position of particles and decays. There are double layers of strips in each barrel/disk and the strips on either side of the barrel/disk are oriented at a relative angle of 40 mrad. This allows reconstruction of both the ϕ and η of a charged particle traversing the detector.

The setup of the Semi-Conductor Tracker is:

- A silicon microstrip tracker consisting of 4,088 two-sided modules and over 6 million implanted readout strips (6 million channels)
- $60m^2$ of silicon distributed over 4 cylindrical barrel layers and 18 planar endcap discs
- Readout strips every $80\mu m$ on the silicon, allowing the positions of charged particles to be recorded to an accuracy of $17\mu m$ per layer (in the direction transverse to the strips)

2.3.3 Transition Radiation Tracker (TRT)

The outermost part of the ID is the transition radiation detector (TRT) which covers the region $|\eta| < 2.0$. Due to the high cost of silicon, semiconductor trackers are not used for this region of the detector. Instead, the TRT uses gaseous wire drift detectors called straw tubes. A straw tube is a 4 mm diameter metalized tube with a fine wire in the center, immersed in a gas mixture of 70% Xe, 27% CO₂, and 3% O₂. There are 73 planes, each composed of a stack of 144 cm long tubes in the barrel and 160 planes of tubes in the endcaps. The straws are aligned longitudinally in the barrel and radially in the endcaps, providing a measurement of the ϕ . A large potential (1500 V) is maintained between the wall of the tube and the sense wire. When a charged particle passes through the tube the gas becomes ionized which creates free electrons. These electrons drift towards the wire thereby producing a current when in contact with the wire. By measuring the timing of the current in the wire the distance from the wire to the traversing particle can be inferred. Thus, by utilizing many straws the trajectory of the particle can be measured.

The TRT plays another role in identifying particles. When a highly relativistic charged particle passes through materials with varying dielectric constants the particle will radiate x-rays. The number of radiated photons is proportional to the relativistic boost of the particle. The layers

in the TRT are made from alternating layers of foil and foam, two materials with very different dielectric constants. An electron (a particle with a large relativistic boost) passing through these layers will radiate photons that will interact with the Xe in the straw tubes. This interaction will create pulses in the sense wires that are much larger than the other types of particles. Thus, a particle with a large boost (e.g. an electron) can be discriminated from a particle with a lower boost (e.g. a hadron).

The setup of the Transition Radiation Tracker is:

- 350,000 read-out channels
- Volume $12m^3$
- Basic detector element: straw tube with 4mm diameter, in the centre a 0.03mm diameter gold-plated tungsten wire
- 50,000 straws in Barrel, each straw 144 cm long. The ends of a straw are read out separately
- 250,000 straws in both endcaps, each straw 39 cm long
- Precision measurement of 0.17 mm (particle track to wire)
- Provides additional information on the particle type that traveling through the detector, i.e. if it is an electron or pion

2.4 Calorimeter

Calorimeters measure the energy a particle loses as it passes through the detector. It is designed to stop the particle entirely and absorb most of the particles' energy coming from a collision, forcing them to deposit all of their energy within the detector. Calorimeters typically consist of layers of passive or absorbing high-density material— for example, lead — interleaved with layers of an active medium such as solid lead-glass or liquid argon.

Electromagnetic calorimeters measure the energy of electrons and photons as they interact with matter. Hadronic calorimeters sample the energy of hadrons (particles that contain quarks, such

as protons and neutrons) as they interact with atomic nuclei. Calorimeters can stop most known particles except muons and neutrinos.

The components of the ATLAS calorimeter system are: the Liquid Argon (LAr) Calorimeter and the Tile Hadronic Calorimeter.

The ATLAS calorimeter system, shown in Figure 2.12, consists of different types of sampling calorimeters covering the pseudorapidity range $|\eta| < 4.9$. The fine granularity of the electromagnetic calorimeter in the region matched to the Inner Detector is necessary for precision measurements of electrons and photons. The hadronic calorimeters are dedicated for the jet reconstruction and missing transverse energy measurement for which a coarser granularity is sufficient.

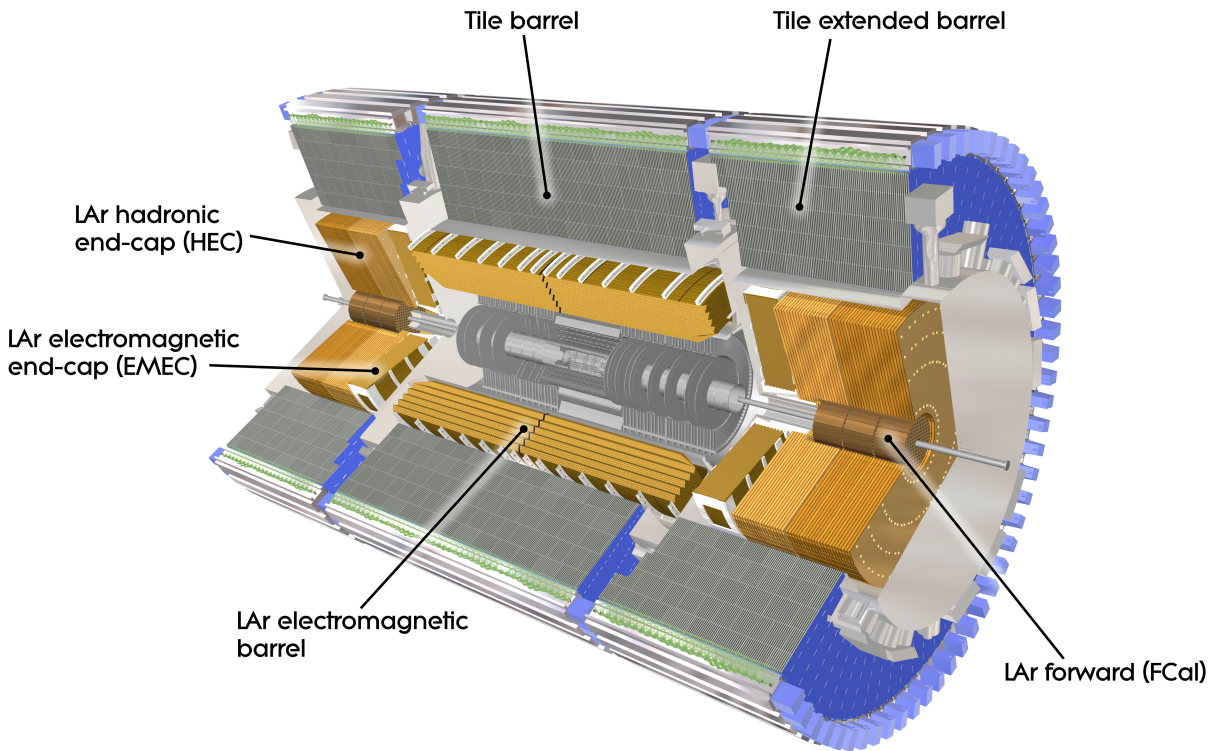


Figure 2.12: ATLAS calorimeter

2.4.1 Electromagnetic Calorimeter

The electromagnetic (EM) system consists of two parts - a presampler and an EM calorimeter (EMC). The EM calorimeter with liquid argon (LAr) as an active material has a typical structure of an accordion-geometry with kapton electrodes and lead absorber plates. The calorimeter is symmetric in the azimuthal angle without any azimuthal cracks. The calorimeter is built of three longitudinal layers. Most of the EM shower energy for high E_T electrons and photons is collected in the middle layer which has a fine granularity of 0.025×0.025 in $\eta - \phi$ space. The first layer, so-called strip layer, offers an excellent γ and π^0 discrimination. The last layer with coarser granularity collects the energy deposited in the tail of very energetic EM showers. The EM calorimeter is divided into a barrel part ($|\eta| < 1.475$) and two end-caps ($1.375 < |\eta| < 3.2$). The presampler detector is located in front of the EM calorimeter in the region $|\eta| < 1.8$. It is developed to correct for the energy lost in the material before the calorimeter. It consists of an active LAr layer of thickness 1.1 cm in the barrel and 0.5 cm in the end-cap.

The setup is listed as below:

- Barrel 6.4m long, 53cm thick, 110,000 channels.
- Works with Liquid Argon at -183°C
- LAr endcap consists of the forward calorimeter, electromagnetic (EM) and hadronic endcaps
- EM endcaps each have thickness 0.632m and radius 2.077m
- Hadronic endcaps consist of two wheels of thickness 0.8m and 1.0m with radius 2.09m
- Forward calorimeter has three modules of radius 0.455m and thickness 0.450m each

2.4.2 Hadronic Calorimeter

Hadrons penetrate the detector more deeply than electrons and photons. Thus, as the name suggests, the hadronic calorimeter is responsible for detecting the energy deposition of the hadrons

that pass into the calorimeter. Hadronic showers occur through a succession of inelastic hadronic interactions. When a hadron enters the calorimeter, the hadron interacts with a nucleus within the material via the strong force to create other particles. The excitation of nucleons from the hadron can also lead to nuclear decay. These processes form a hadronic shower whose energy deposition is proportional to the energy of the incident hadron. The hadronic calorimeters are much larger (and more dense) than the EMCs to prevent hadrons from entering the muon system.

The Tile Calorimeter is a hadronic calorimeter covering the range $|\eta| < 1.7$ with steel used as an absorber and scintillating tiles as an active material. This system consists of two barrel sections: a central section that covers $|\eta| < 1.0$ and an extended section on each side that covers $0.8 < |\eta| < 1.7$. Each section of the detector is made of 64 modules each covering $\Delta\phi \approx 0.1$. When a hadron hits a layer of steel it produces a hadronic shower. The secondary particles from the shower enter the scintillating polystyrene tiles doped with 2% wavelength-shifting fluorescents as active material which produces light. The light is detected by photomultiplier tubes (PMT) outside of the detector. The current produced by the PMT is proportional to the hadronic energy deposition in a cell.

Other than the Tile Calorimeter, the forward hadronic calorimeters use LAr technology. The Hadronic End-cap Calorimeter (HEC) covers pseudorapidity range from 1.5 to 3.2 using copper as the absorber. Finally, the Forward Calorimeter (FCal) covers the most forward region up to $3.1 < |\eta| < 4.9$. The FCal consists of three modules in each end-cap: The first module is made of copper and is optimized for electromagnetic measurements, the other two are made of tungsten and are used primarily for measurements of the hadronic showers.

The setup is listed as below:

- Central barrel made of 64 wedges, each 5.6m long and weighing 20,000 kg
- Two extended barrels each with 64 wedges, each 2.6m long and weighing 9,600 kg
- 500,000 plastic scintillator tiles

2.5 Muon Spectrometer

The muon spectrometer measures the deflection of the muon tracks in the magnetic field produced by large superconduction air-core toroid magnets (one in the barrel and two in the end-caps) in the region $|\eta| < 2.7$. The spectrometer chambers are arranged in three cylindrical layers around the beam axis while in the transition region and in the end-caps the chambers are installed in three planes perpendicular to the beam axis. The layout of the muon chambers is shown in Figure 2.13.

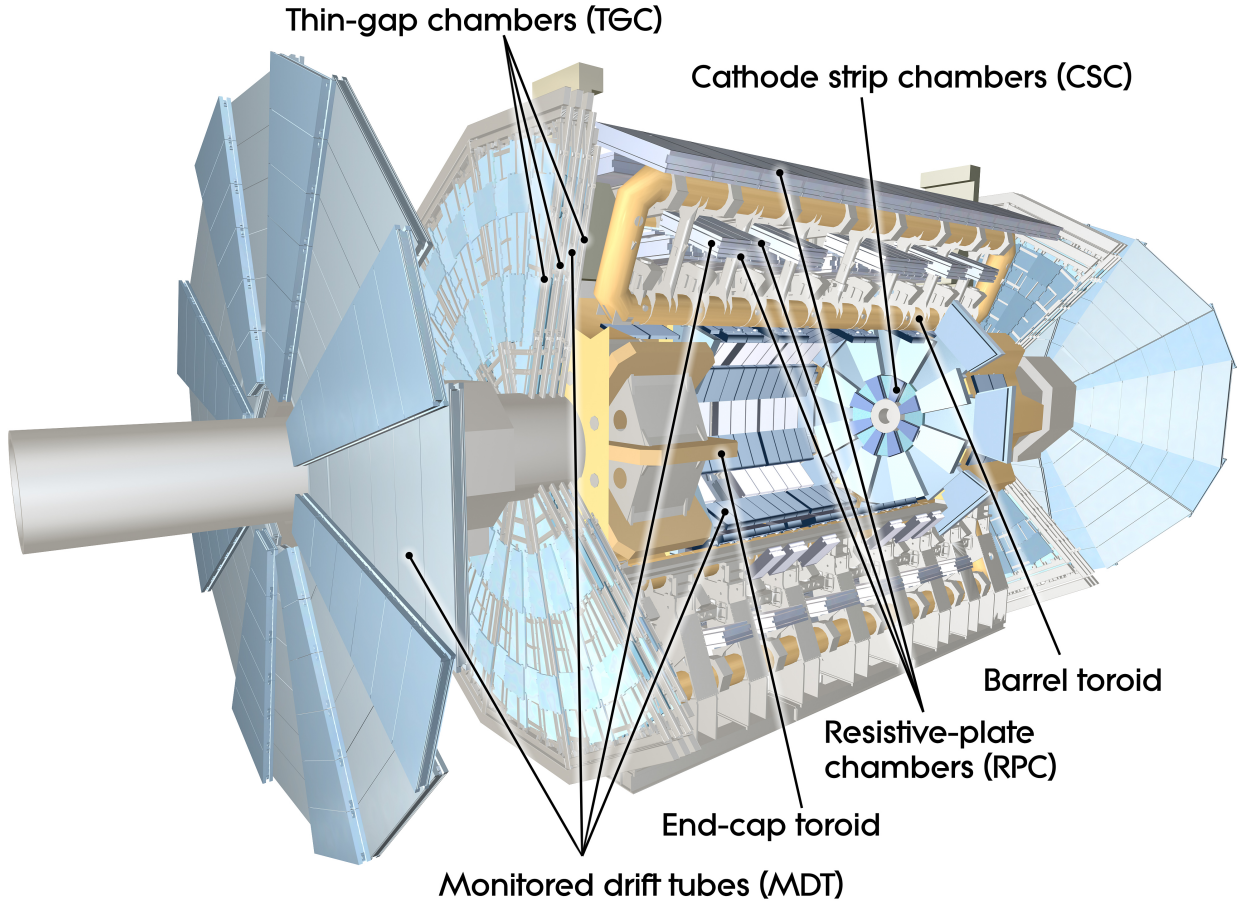


Figure 2.13: ATLAS Muon Spectrometer

The Monitored Drift Tubes (MDTs) cover most of the pseudorapidity range of the muon system and provide a precision measurement of the muon tracks. Cathode Strip Chambers (CSCs) with

higher granularity are used in a large pseudorapidity ($2.0 < |\eta| < 2.7$). The CSCs are radiation resistant and can be used in a region with an increased particle rate.

The muon trigger system covers the pseudorapidity range up to 2.4. Resistive Plate Chambers (RPCs) are used in the barrel and Thin Gap Chambers (TGCs) in the end-caps. These chambers are used to measure the muon coordinate in the direction orthogonal to the precision-tracking chambers and also for triggering.

The resolution of the muon spectrometer is $\sigma_{p_T} / p_T = 10 \%$ at $p_T = 1 \text{ TeV}$ [40].

2.6 Trigger System and Data Acquisition

The LHC collides bunches of protons at a rate of 40 MHz. Saving the results of each of these interactions would require writing 60 TB of data every second - which is impractical in terms of the storage space. Because of this limit, the detector can only record a small fraction of events, and special hardware (Trigger System) is needed to pick out events most useful for analysis.

The ATLAS trigger system carries out the selection process in two stages:

The Level-1 hardware trigger, constructed with custom-made electronics, works on a subset of information from the calorimeter and muon detectors. The decision to keep the data from an event is made less than 2.5 microseconds after the event occurs, and the event is then retrieved from pipelined storage buffers. The Level-1 trigger can save at most 100,000 events each second for the High-Level Trigger (HLT).

The HLT is a large farm of CPUs - i.e. a software based trigger - which refines the analysis of the hardware-based Level-1 trigger. It conducts a very detailed analysis either by performing overall examination of the whole event for selected layers of the detector (for example calorimeters, trackers, muon detectors), or, by utilizing the data in smaller and isolated regions of the detector. About 1000 events per second are selected by the HLT analysis and are fully assembled into an event record. These events are passed on to a data storage system for offline analysis.

The whole scheme is summarized in the Figure 2.14 [87].

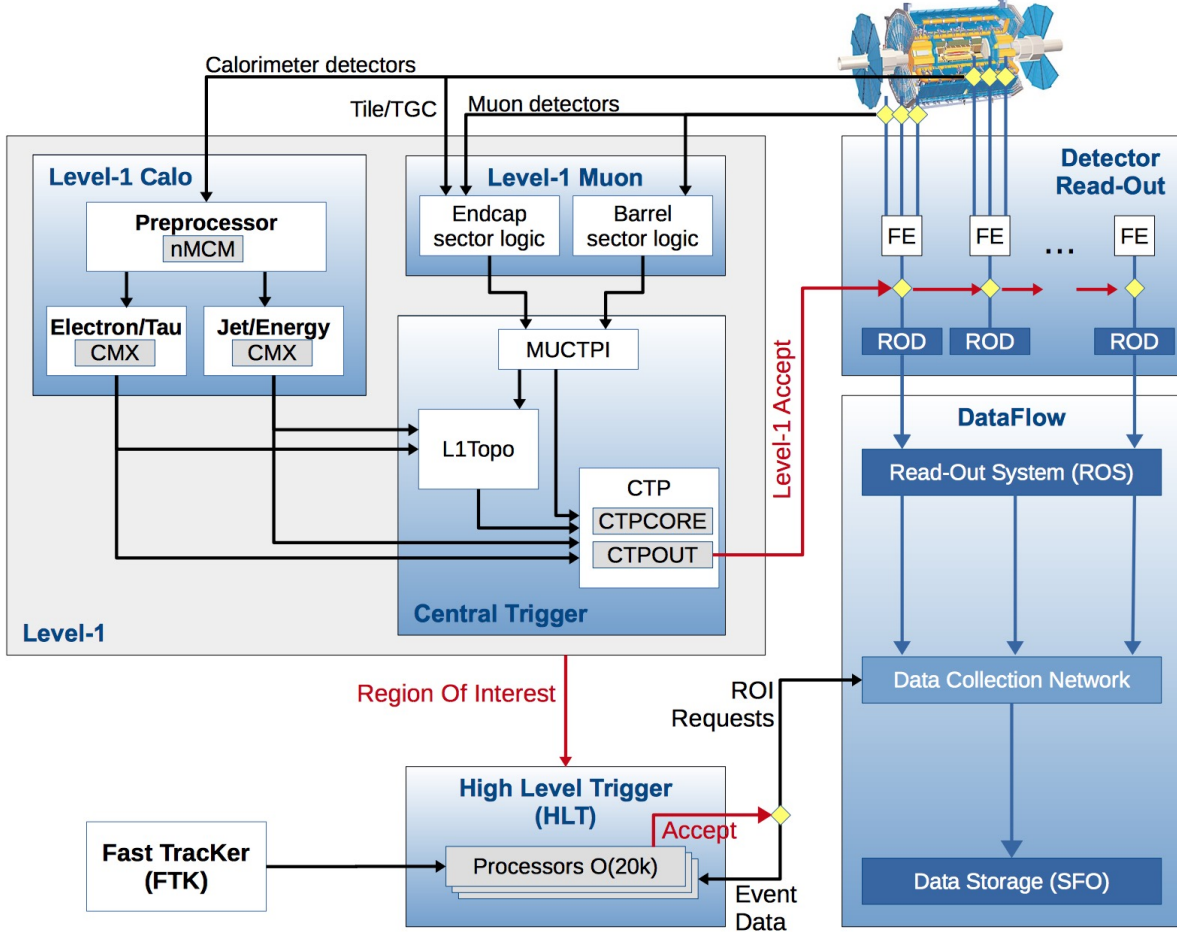


Figure 2.14: Schematic layout of the ATLAS trigger and data acquisition system in Run-2

2.7 Pileup

Due to the fact that there are as many as 10^{11} protons per bunch, multiple protons can interact in each bunch crossing, so-called pileup events need to be separated from the events of interest. The majority of proton-proton collisions result in low energy hadronic jets. Several measures are taken to properly account for pileup in events. The tracking system can be used to determine the vertex with the largest transverse momentum (largest $\sum p_T^2$) and jets without tracks coming from this vertex may be presumed to be pileup jets and therefore ignored (see Section 4 for more details).

The rate of pileup events is defined in Equation 2.3, which is directly related to the luminosity(explained previously in Equation 2.1) of the beam and the known cross-section of inelastic collisions($\sigma_{inelastic} = 80\text{mb}$ at $\sqrt{s} = 13\text{TeV}$ [6])

$$R_{inelastic} = L \cdot \sigma_{inelastic} \quad (2.3)$$

From $R_{inelastic}$, the average number of interactions per bunch crossing, called $\langle\mu\rangle$, can be computed as in Equation 2.4:

$$\langle\mu\rangle = \frac{R_{inelastic}}{n_{bunch}f_{cross}} \quad (2.4)$$

where $n_{bunch}f_{cross}$ is the number of bunches over a period. The mean number of interactions per bunch crossing from 2015 to 2018 are shown in Figure 2.15. From this figure it is clear that the instantaneous luminosity of the beam was increased along time, resulting in a much larger number of pileup events.

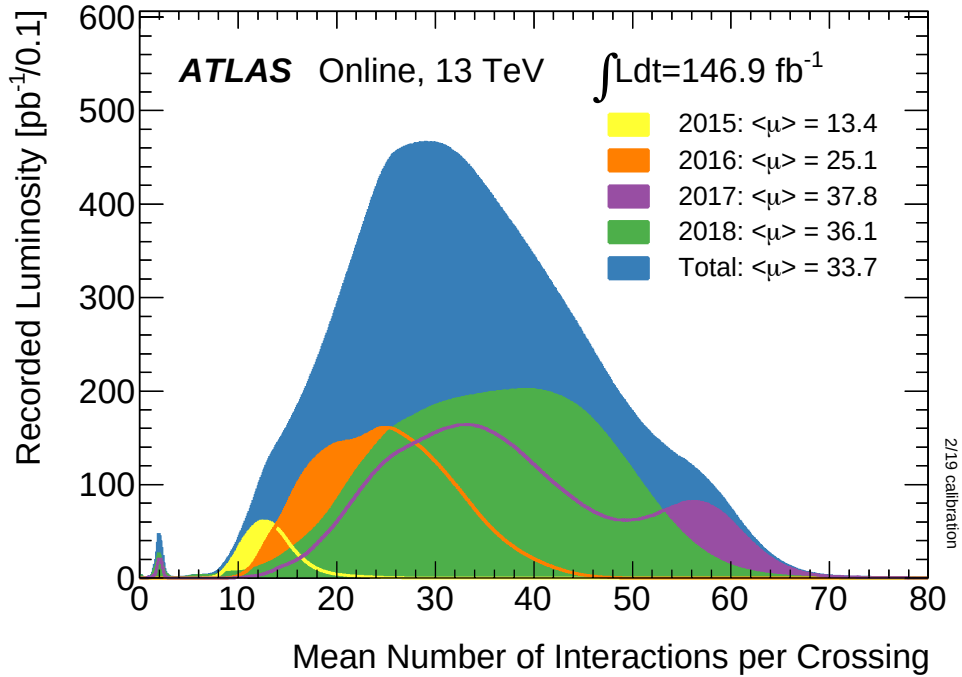


Figure 2.15: Collected luminosity along $\langle\mu\rangle$

To account for situations where pileup is not distinguishable from true signal events, pileup information is included in physics simulations and thus must be subsequently reweighted to match the pileup distribution in data [35]. This reweighting procedure will be discussed further in Section [5.6](#).

CHAPTER 3. ANALYSIS STRATEGY

The $Z \rightarrow e\mu$ signal is characterized by two isolated energetic, oppositely charged leptons with little jet activity and small missing energy. Charge conjugation is implied. In the pre-selection (preliminary selection cuts used for reducing the backgrounds), we require there to be an isolated high- p_T $e^\pm$ and $\mu^\mp$ pair. The main backgrounds which can survive this selection consist of $t\bar{t} \rightarrow e\mu\nu\bar{\nu}b\bar{b}$, the diboson process $WW \rightarrow e\mu\nu\bar{\nu}$, the $Z \rightarrow \tau\tau \rightarrow e\mu\nu\bar{\nu}\nu\bar{\nu}$, whose invariant mass spectrum of the two leptons ($m_{e\mu}$) extends into the Z signal region, and $Z \rightarrow \mu\mu$ where one muon is faked as an electron.

The $t\bar{t}$ background is characterized by the existence of energetic jets. We further reduce it by vetoing events with tagged b-jets, and applying a loose cut on the transverse momentum of the leading jet ($p_T^{\text{leading jet}}$). The diboson background has the neutrinos as the final particles then it can be suppressed by a cut on the missing transverse momentum (E_T^{miss}).

After pre-selection, the remaining backgrounds have quite similar topology as the signal. To further reduce them, a machine learning technique is explored using the Toolkit for Multivariate Analysis (TMVA) [61]. The training set and the cut on the TMVA discriminant is optimized by gaining the maximum sensitivity (FOM), which will be explained in details in 6.3.

Candidate events surviving all the above selection criteria form a smooth $m_{e\mu}$ spectra in the [70, 110] GeV window. The invariant mass $m_{e\mu}$ is the rest mass of the mother particle (in $Z \rightarrow e\mu$ process, the Z boson is the mother particle) which is invariant under the Lorentz transformation in Relativity Theory. The invariant mass $m_{e\mu}$ is calculated by the Equation 3.1, here the $\sum_{e,\mu}$ is summing over the electron and muon pair, and E is the energy and $\vec{p}$ is the momentum of electron and muon. Currently the signal region (SR) [85, 95] GeV is blinded in this thesis. We estimate the

background around Z boson mass(signal region is $[85, 95]$ GeV) by fitting the sidebands (SB) $[70, 85]$ and $[95, 110]$ GeV.

$$m_{e\mu} = \sqrt{\sum_{e,\mu} E^2 - (\sum_{e,\mu} \vec{p})^2} \quad (3.1)$$

The Upper Limit (UL) of $Br(Z \rightarrow e\mu)$ is calculated using Eq.(3.2), while the details of UL will be explained in Chapter 8. We take advantage of the reference channels $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ in the normalization so that most uncertainties arising from electron, muon and jet identification and reconstruction cancel out. Otherwise they would have been the largest sources of systematic uncertainty. The $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ events are selected in the same way as for the $Z \rightarrow e\mu$ signals.

$$Br(Z \rightarrow e\mu) < \frac{N_{95\%}^{\text{UL}}}{\varepsilon_{Z \rightarrow e\mu} \cdot N_Z^{\text{avg}}}. \quad (3.2)$$

The nominator $N_{95\%}^{\text{UL}}$ is the upper limit of the observed $Z \rightarrow e\mu$ events in the SR at the 95% confidence level (C.L.). It is extracted with the signal MC and background fit from fitting the sideband $[70, 85]$ and $[95, 110]$ GeV, using the HistFitter Frequentist Calculator with one-sided Profile Likelihood as test statistic. In the denominator, $\varepsilon_{Z \rightarrow e\mu}$ denotes the signal efficiency obtained from signal MC, and N_Z^{avg} the estimated number of Z bosons produced in data is calculated using the observed $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ events.

So let's explain more details of each term in denominator as below:

The efficiency $\varepsilon_{Z \rightarrow e\mu}$ in the denominator is calculated base on MC simulations with the following equation,

$$\varepsilon_{Z \rightarrow e\mu} = \frac{N_f^{MC}}{N_i^{MC}}, \quad (3.3)$$

Here N_f^{MC} is the number of the selected signal MC events after the pre-selection and final selection in the signal region(invariant mass region of $[85,95]$ GeV), and N_i^{MC} is the number of generated $Z \rightarrow e\mu$ events.

In addition, the number of produced Z bosons N_Z^{avg} is the average value calculated from $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ channels using the following equations:

$$N_Z^{\text{avg}} = \frac{1}{2} \cdot (N_{Z \rightarrow ee} + N_{Z \rightarrow \mu\mu}), \quad (3.4)$$

$$N_{Z \rightarrow ee} = \frac{N_{Z \rightarrow ee}^{\text{obs}}}{\varepsilon_{Z \rightarrow ee}} / Br(Z \rightarrow ee), \quad (3.5)$$

$$N_{Z \rightarrow \mu\mu} = \frac{N_{Z \rightarrow \mu\mu}^{\text{obs}}}{\varepsilon_{Z \rightarrow \mu\mu}} / Br(Z \rightarrow \mu\mu). \quad (3.6)$$

Here $\varepsilon_{Z \rightarrow ee}$ and $\varepsilon_{Z \rightarrow \mu\mu}$ are defined similarly as $\varepsilon_{Z \rightarrow e\mu}$ in Equation 3.3.

Combining Equations 3.2-3.6 together, we can write the upper limit of $Br(Z \rightarrow e\mu)$ as below:

$$Br(Z \rightarrow e\mu) < \frac{N_{95\%}^{UL}}{\varepsilon_{Z \rightarrow e\mu}^{MC} \cdot \frac{1}{2} \cdot \left(\frac{N_{Z \rightarrow ee}^{\text{obs}}}{\varepsilon_{Z \rightarrow ee}^{MC}} / Br(Z \rightarrow ee) + \frac{N_{Z \rightarrow \mu\mu}^{\text{obs}}}{\varepsilon_{Z \rightarrow \mu\mu}^{MC}} / Br(Z \rightarrow \mu\mu) \right)}. \quad (3.7)$$

From it we can see more clearly that the $\varepsilon_{Z \rightarrow e\mu}$ can cancel out with $\varepsilon_{Z \rightarrow ee}$ and $\varepsilon_{Z \rightarrow \mu\mu}$, to reduce the efficiency systematic uncertainties, which is the reason we choose the reference channels $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ in the normalization.

CHAPTER 4. OBJECT RECONSTRUCTION

4.1 Introduction

As we are searching for the Z boson decay to one electron and one muon, good reconstruction of electron and muon objects are crucial to the analysis. Besides, to gain maximum sensitivity in the Signal Region, we also use E_T^{miss} and leading p_T^{jet} requirements to suppress the backgrounds like $t\bar{t}$ and diboson processes. The object reconstruction of E_T^{miss} and jets are necessary as well.

These object definitions are also shared by the $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ channels.

4.2 Primary vertex

There are many tracks associated with any given bunch crossing of protons. Using an algorithm, these tracks are used to identify primary vertex candidates. A primary vertex is the vertex which originated the objects which caused triggers to fire and the event to be saved. Typically the primary vertex candidates are ordered according to the sum of p_T of tracks in the vertex. Usually the highest $\sum p_T$ vertex is assumed to be the primary vertex.

Events are required to have at least one primary vertex that has at least two associated tracks, each with transverse momentum $p_T > 400\text{MeV}$. The primary vertex is subsequently used for calculation of the main physics objects in the analysis: electrons, muons, jets and missing transverse energy E_T^{miss} . The definitions of those physics objects are summarized in the following sections.

Objects have some kind of requirement to ensure that they come from the chosen primary vertex. This is determined using two inner tracking criteria : the transverse and longitudinal impact parameters.

The default transverse (d_0) and longitudinal (z_0) impact parameters are referred to the beamline. Such choice also simplifies applying selections based on impact parameters with respect to a vertex

position. The definition of longitudinal and transverse impact parameters is illustrated in Figure 4.1.

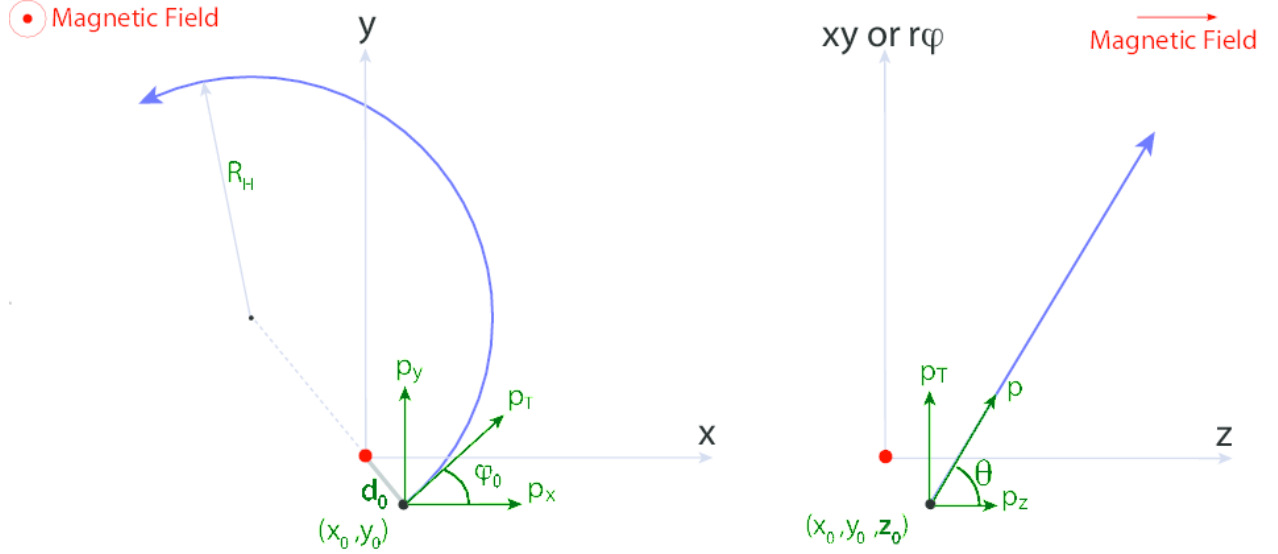


Figure 4.1: Impact parameters (z_0 , d_0) sketch graph

4.3 Electron

Electrons are identified by the Inner Detector and EM Calorimeter energy depositions to reconstruct the momentum and energy of electron. Electrons produced at the LHC at a center-of-mass energy of 13 TeV typically are fully captured within the Electromagnetic calorimeter. They will not pass through the Electromagnetic calorimeter. Because the penetration of electrons can clearly be an issue since the total energy of the electron would not be fully represented by the deposits in the Electromagnetic calorimeter, causing problems reliably reconstructing events.

The energy clusters reconstructed by the electromagnetic calorimeter are determined using a sliding window algorithm [65], which clusters calorimeter cells within fixed-size rectangles, like 5×7 cell window. After calibrations and corrections are applied, the four momentum is calculated using the energy measurement of the cluster and the η and ϕ measurements from the track.

This procedure does not completely exclude other physics objects from being mis-reconstructed as prompt(originated from primary vertex) electrons, including hadronic jets, non-prompt(not orig-

inated from primary vertex) electrons from photon conversions, and semi-leptonic decays of hadrons with heavy quarks. A multivariate analysis (MVA) technique takes into account several cluster and track variables to create a likelihood (LH) identification for each candidate as electron or not. The likelihood identification criteria is described in Ref. [85]. Different working points balancing signal efficiency with background rejection are provided. The levels of identification are categorized as LooseLH, MediumLH and TightLH corresponding respectively to 96%, 94% and 90% identification efficiencies for signal electrons at $E_T = 100$ GeV.

An electron from a W or Z boson decay will be produced relatively isolated while jets, converted photons, and other fake electron objects will have additional nearby energy deposits. Isolation is measured for each candidate in both the calorimeter and the tracking system. The calorimeter isolation ($E_T^{cone0.2}$) calculates the sum of E_T within a cone with $\Delta R < 0.2$ around the center of the cluster, subtracting the 5×7 cell window contribution from the electron. Track isolation is determined similarly from the transverse momentum of tracks $p_T^{varcone0.2}$ within a cone of variable size $\Delta R < \min(0.2, 10\text{GeV}/E_T)$.

The identification and isolation working points are optimized for electrons coming from the PV, and in this analysis the leptons are crucial for the proper identification of the PV. For these reasons we additionally impose cuts on the transverse (d_0) and longitudinal (z_0) impact parameters to verify that the electron indeed comes from the PV. Electrons are required to have $|\eta| < 2.47$, excluding the crack region (transition region between the barrel and endcap calorimeters) $1.37 < |\eta| < 1.52$ in which uncertainties are known to be very large.

Finally, a minimum pT threshold is set above trigger threshold to avoid large uncertainties from the trigger turnover region at low electron pT region, which can be seen in 4.2. We choose the 27 GeV as the pT threshold to ensure the electron efficiency is on the plateau region.

The candidate signal electron should meet the following requirements:

- Trigger matching, with the lowest unprescaled single lepton triggers explained in Chapter 5.5.
- Identification: "TightLH"

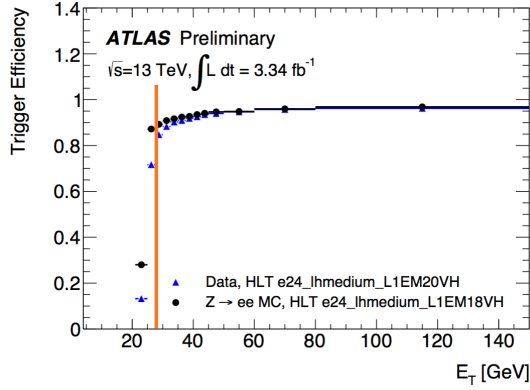
- Isolation: "gradient" working point using the IsolationSelectionTool.
- $p_T > 27$ GeV. This cut value is chosen at 1 GeV higher than the cut in the single-lepton trigger. As can be seen in Figure 4.2, electron trigger efficiency in MC and data are well consistent above 27 GeV.
- $|\eta| < 2.47$, excluding $1.37 < |\eta| < 1.52$
- $|d_0/\sigma(d_0)| < 5$, where d_0 is the transverse impact parameter of the electron candidate with respect to the measured beam line position at the point of the closest approach of the track, $\sigma(d_0)$ is the corresponding uncertainty.
- $|z_0 \sin \theta| < 0.5$ mm, where z_0 and θ are the longitudinal impact parameter and the polar angle of the electron candidate at the point of the closest approach of the track, respectively

We veto events if they include additional LooseLH electrons besides the signal candidate. Compared to the TightLH candidate signal electron, LooseLH electrons have no isolation requirement.

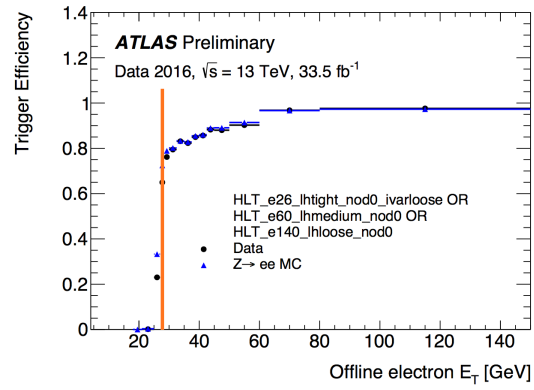
Corrections to the electron energy scale and resolution are applied to data and the simulated samples as recommended in [12]. The scale factors recommended by the Egamma group in the AseElectronEfficiencyCorrectionTool [12] are used to correct for the identification, reconstruction and isolation efficiencies of the simulation in order to match the data.

4.4 Muon

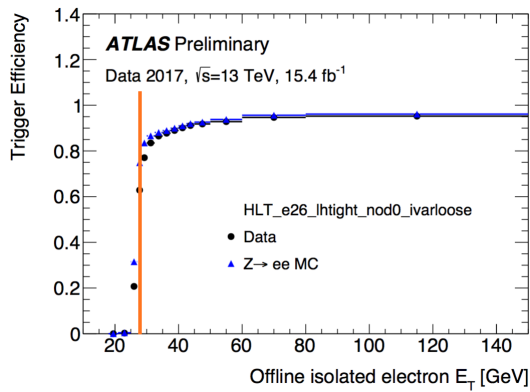
Muons are identified in the ATLAS detector not only in Inner Detector, but also in the additional outermost layer Muon Spectrometer of our detector. Muons pass through the Inner Detector, causing hits which can be used to reconstruct a track. They then pass through the Electromagnetic calorimeter, but depositing very little energy, according to the large suppression on bremsstrahlung from its large mass(The bremsstrahlung radiation power is inversely proportional to $mass^4$). Finally a muon will reach the muon spectrometer and pass through, leaving hits to be reconstructed as tracks to match the ones reconstructed in Inner Detector. The combined tracks will be used to improve the identification of muons.



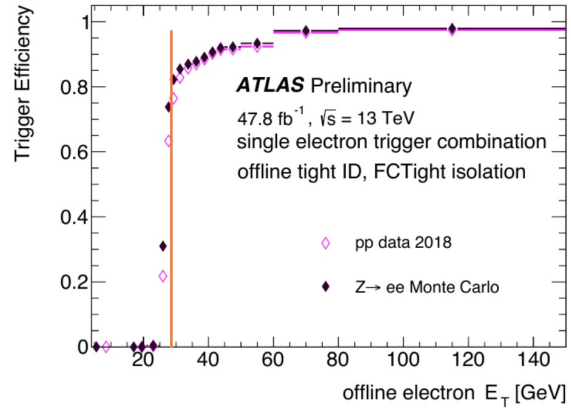
(a) 2015 electron trigger efficiency



(b) 2016 electron trigger efficiency



(c) 2017 electron trigger efficiency



(d) 2018 electron trigger efficiency

Figure 4.2: Efficiency of different triggers: (a) 2015 Efficiency of the combined L1 and HLT e24.lhmedium_L1EM20VH trigger as a function of the offline electron candidates transverse energy (E_T), (b) 2016 Efficiency of the logical OR between HLT_e26.lhtight_nod0_ivarloose, HLT_e60.lhmedium_nod0 and HLT_e140.lhloose_nod0 triggers as a function of the offline electron candidate's transverse energy (E_T), (c) 2017 efficiency of the HLT_e26.lhtight_nod0_ivarloose trigger as a function of the offline electron candidate's transverse energy (E_T). (d) 2018 efficiency of the single electron trigger combination (logical OR of HLT_e26.lhtight_nod0_ivarloose, HLT_e60.lhmedium_nod0 and HLT_e140.lhloose_nod0) as a function of the offline electron candidate's transverse energy (E_T)

In the MS, hits collected by the MDTs [74] are checked for a trajectory along a straight line, forming segments. Segments are determined separately by the CSC detectors using a specialized search algorithm. Muon track candidates are then constructed by combining the segments from multiple layers, removing some hits as needed to improve the quality of the fit. At least two matching segments are required to construct a track (except in the barrel-endcap transition region where 1 high quality segment will suffice).

Muon objects are reconstructed by matching track candidates independently created in the MS and ID [39]. The candidates are identified using the recommendations described in Ref. [18] and the tools described in Ref. [18]. To be within the detector acceptance, muon candidates are required to have $|\eta| < 2.5$. Four identification quality levels are available offering different background rejection, namely "VeryLoose", "Loose", "Medium" and "Tight", which are conducted by using the MuonSelectionTool [18]. In this analysis the Medium working point is used. The Medium muons include only muons identified by the combination of ID and MS (so-called "combined muons"). "Combined muons" are required to have 3 hits in at least 2 MDT layers, except for tracks in the $|\eta| < 0.1$ region.

Muon isolation is determined similarly to electron isolation: both tracking and calorimeter based isolation measurements are used to create working points. The track based muon isolation uses a larger cone size than the electron case, $\Delta R < 0.3$.

The candidate signal muon should meet the following requirements:

- Trigger matching, with the lowest unprescaled single lepton triggers explained in Chapter 5.5.
- Identification: "Medium" and "Tight"
- $p_T > 27$ GeV. This cut value is chosen at 1 GeV higher than the cut in the single-lepton trigger. Shown in Figure 4.3 and Figure 4.4, muon trigger efficiency in MC and data are consistent above 27 GeV.
- Isolation: "gradient" working point using the IsolationSelectionTool.

- $|\eta| < 2.5$
- $|d_0/\sigma(d_0)| < 3$
- $|z_0 \sin \theta| < 0.5 \text{ mm}$

We veto events if they include additional muons besides the signal candidate, namely "loose muon". Compared to the candidate signal muon, loose muons only pass "gradientLoose" isolation criterion.

Corrections to the muon momentum scale and resolution are applied to the simulation as recommended in Ref. [18]. The scale factors recommended by the Muon Combined performance group in MuonTriggerScaleFactors [18] are used to correct for the identification and isolation efficiencies in simulation to match those obtained in data.

4.5 Jet

Jets are the experimental signatures of quarks and gluons produced in high-energy processes such as head-on proton-proton collisions. As quarks and gluons have a net color charge and cannot exist freely due to color-confinement, they are not directly observed in Nature. Instead, they come together to form color-neutral hadrons, a process called hadronization that leads to a collimated spray of hadrons called a jet. Quarks and gluons produced from the hard interaction that shower into a dense cluster of many particles can be called a jet. Figure 4.5 shows an illustration of the process of jet formation in pp collisions.

Ideally, we would like the total momentum of all of the particles in the jet to be close to that of the initial quark or gluon, even though that is not realistic given the detector resolution, radiation etc. Jet reconstruction essentially defines which cells in the calorimeter belong to a given jet. It is important when clustering the jet that the kinematic properties of the original quark or gluon match as closely as possible to those of the reconstructed jet. Figure 4.6 shows a typical simulated di-jet event as seen by the ATLAS detector.

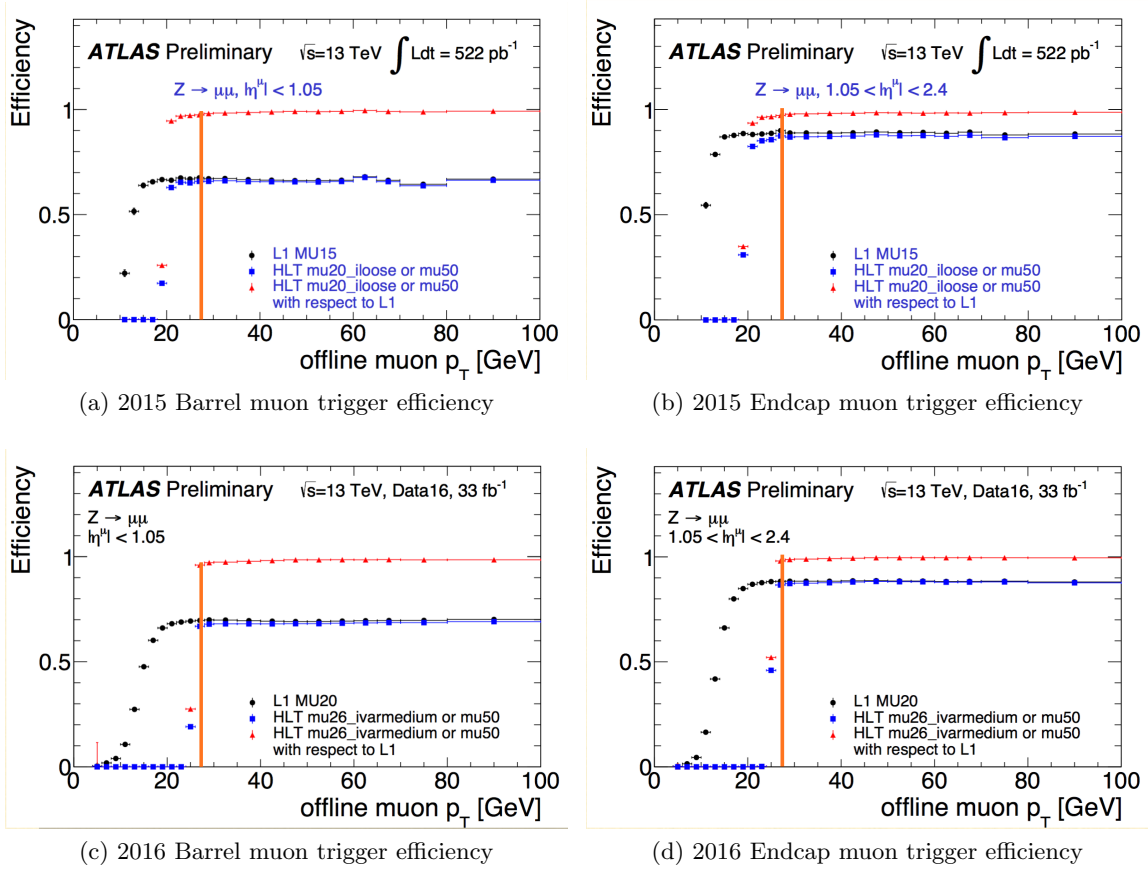


Figure 4.3: Efficiency of different triggers: (a) 2015 Barrel muon trigger efficiency as a function of muon p_T , (b) 2015 Endcap muon trigger efficiency as a function of muon p_T , (c) 2016 Barrel muon trigger efficiency as a function of muon p_T , (d) 2016 Endcap muon trigger efficiency as a function of muon p_T ,

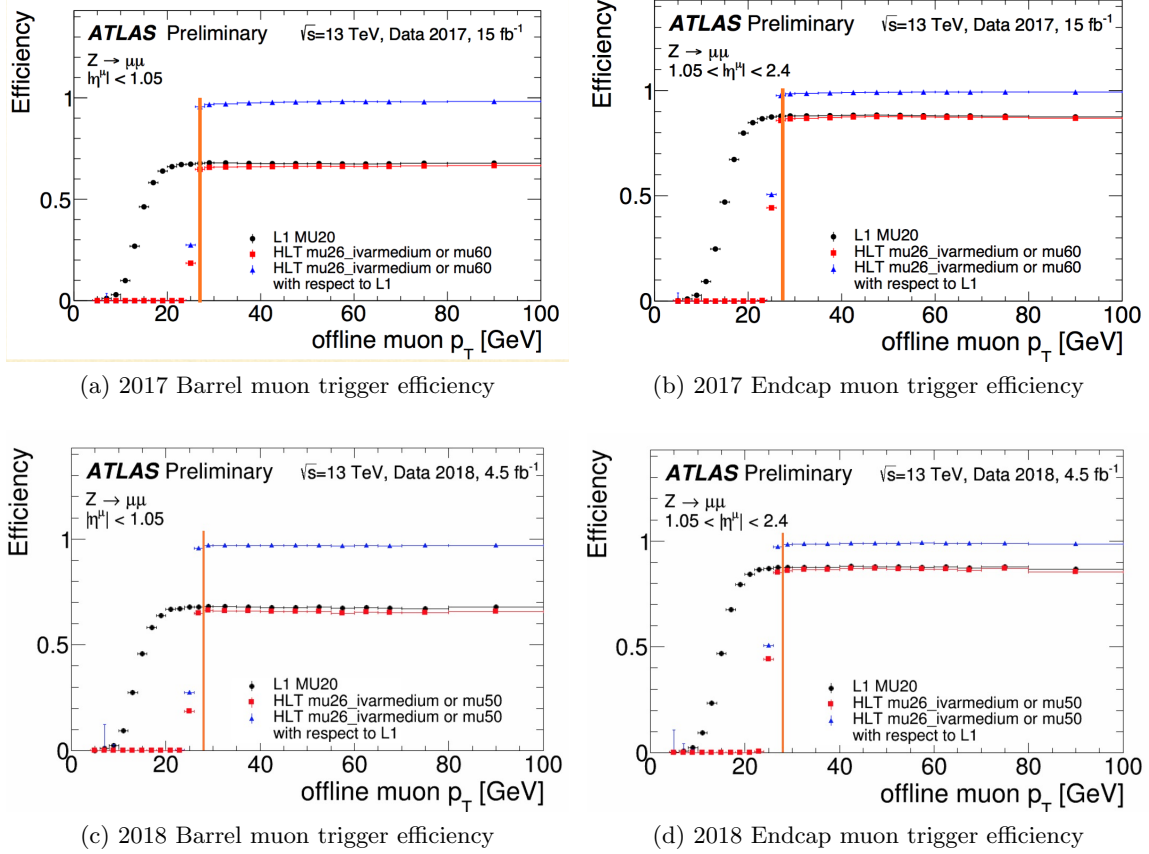


Figure 4.4: Efficiency of different triggers: (a) 2017 Barrel muon trigger efficiency as a function of muon p_T , (b) 2017 Endcap muon trigger efficiency as a function of muon p_T , (c) 2018 Barrel muon trigger efficiency as a function of muon p_T , (d) 2018 Endcap muon trigger efficiency as a function of muon p_T .

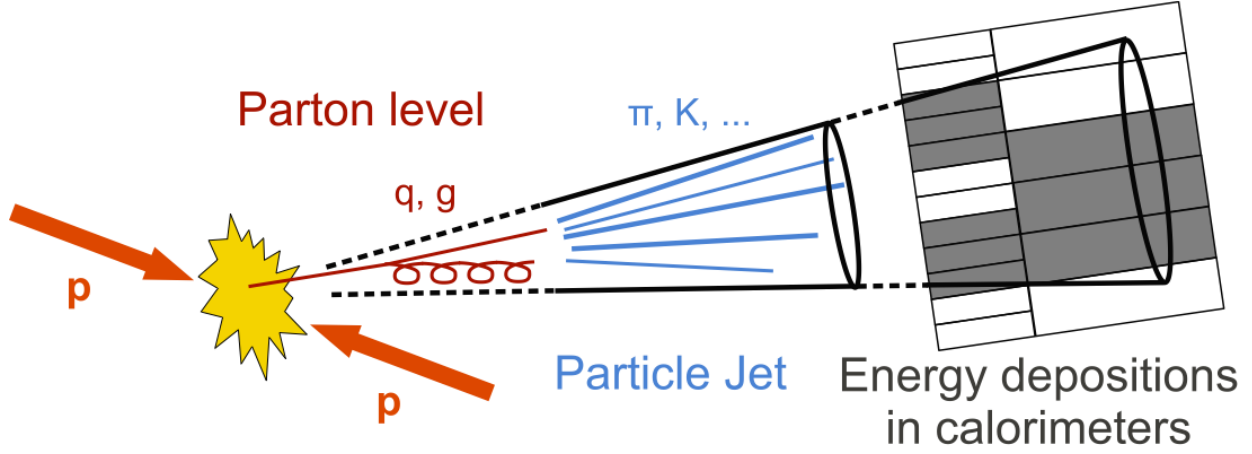


Figure 4.5: sketch of pp-collision and resulting collimated spray of particles, a jet

4.5.1 Jets reconstruction

We use jets to reconstruct the MET and to suppress the $t\bar{t}$ background. Good reconstruction of jets improve the good quality of our signal sensitivity.

To ensure the quality of jet reconstruction, several criteria are applied.

Firstly, jet are reconstructed using anti- k_t algorithm [36] to collect energy deposits in the calorimeter together in a way such that the objects contained within a given jet are likely to have come from the same jet. The typical size of jets used at ATLAS is a distance parameter of $\Delta R = 0.4$ in η and ϕ space, where ΔR is defined as $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

In addition, the reconstructed jet should pass the minimum pT requirement of 20 GeV, which is recommended by the event cleaning tool to remove jets arising from non-collision backgrounds or noise in the calorimeters. A jet should be in the central region of detector, $|\eta| < 2.5$, rather than the forward segments, or the jet will not be within the tracker acceptance.

Jet energies are corrected [17] for detector inhomogeneities, the non-compensating nature of the calorimeter, and the impact of multiple overlapping pp interactions. Correction factors are derived using test beam, cosmic ray, pp collision data, and a detailed Geant4 detector simulation.

Besides, the majority of jets from pile-up are rejected using the jet-vertex-tagger(JVT) parameter [17], a likelihood discriminant combining information from several track-based variables. The

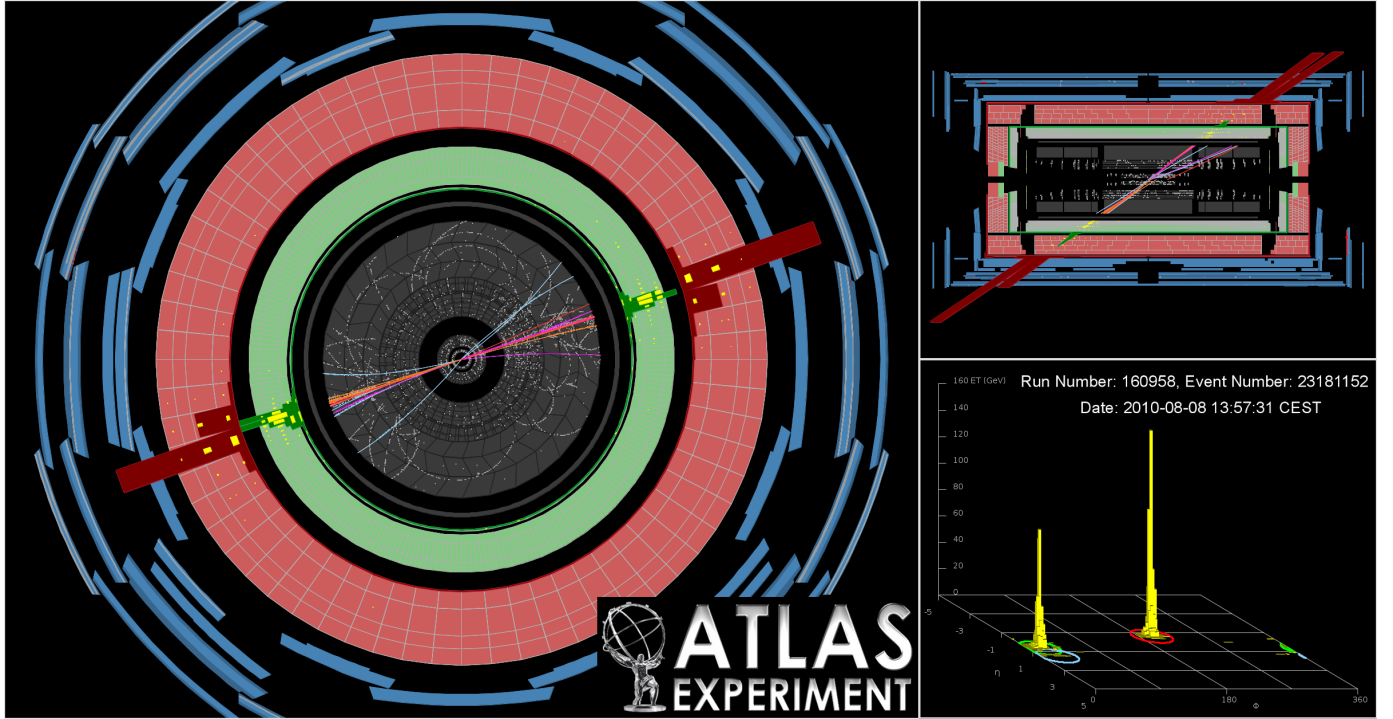


Figure 4.6: A high mass dijet event: two high- p_T jets with invariant mass 2.8 TeV. A track p_T cut of 2.5 GeV has been applied for the display to remove the pile-up jets

JVT is a discriminant constructed using Jet Vertex Fraction(JVF), which represents the fraction of the charged particles (i.e. tracks) inside the jet (i.e. within the cone of the jet) that are associated with the primary vertex. More precisely, JVF is the ratio of the sum of p_T of matched tracks which originate from the chosen primary vertex to the sum of p_T of all matched tracks in the jet, but independent of their origin. According to the results of the working point evaluation of JVT, for the jets of which the p_T is between 20 GeV and 60 GeV, and its $|\eta| < 2.4$, the medium JVT cut that $JVT > 0.59$ [3] is chosen in the analysis. The medium selection working point with a value of 0.59 has an average efficiency of 92 percent.

Overall, the following criteria are implied to ensure good jet quality:

- Passing the requirement `DFCommonJets_jetClean_LooseBad`, which is implemented by the `JetCleaningTool`

- AntiKt jets with distance parameter $R < 0.4$
- $p_T > 20$ GeV
- $|\eta| < 2.5$
- JVT > 0.59 for jets with $p_T < 60$ GeV and $|\eta| < 2.4$

4.5.2 b-tagging

One of the most important selection criteria for the analysis of events containing top or top-like quarks is the identification of jets containing b-quarks. Events which contain a top quark are expected to produce a b quark as the top quark decays, and in general, $Z \rightarrow e\mu$ events are not expected to have any b quarks in the event. So for the present Z boson flavor violation analysis it is important to identify jets originating from b-hadrons and to veto events containing any b-hadrons as to suppress the $t\bar{t}$ and other backgrounds. The discrimination of b jet from light-quark jets originates mainly in the relatively long lifetime of b-flavored hadrons, resulting in a significant flight path length L . This leads to measurable secondary vertices and impact parameters, combined with several other variables, that can be used to distinguish b-jets from light quark or gluon jets. Figure 4.7 shows a displaced secondary vertex for a b-jet.

The b-jets are identified with a b-tagging algorithm based on a multivariate discriminant technique [26]. This algorithm uses a set of tracks with loose impact parameter constraints in a region of interest around each jet axis to enable the reconstruction of the b-hadron decay vertex. Both the efficiency to correctly identify b-jets and the rate with which light jets are misidentified as b-jets, the mis-tag rate, have to be carefully measured in data and compared to the predictions in simulated events. This is typically done at one or more working points of the algorithm. The b-tagging working points are defined by a specific b-tagging efficiency using an inclusive $t\bar{t}$ sample [11]. The MV2c10 tagger is used [20]. We employ the working point with the b-tagging efficiency of 85%..

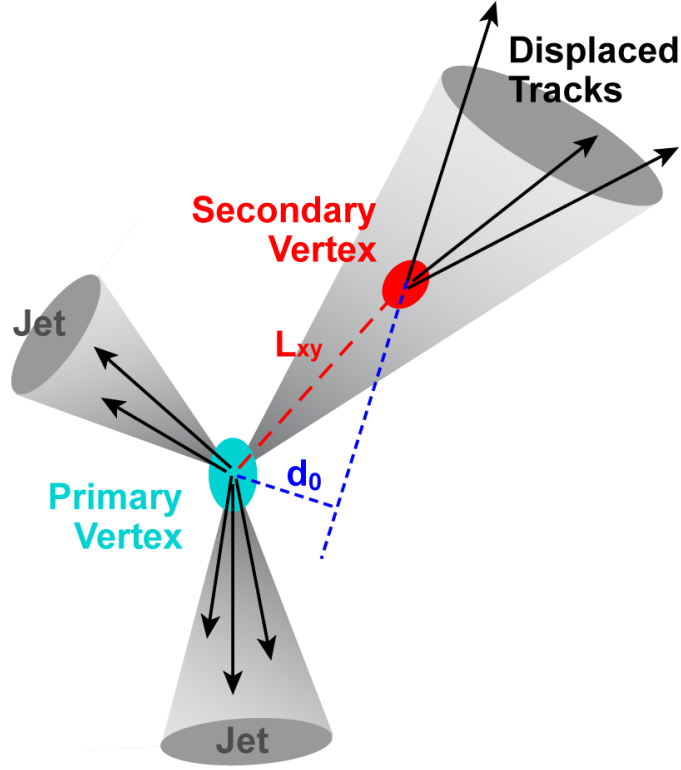


Figure 4.7: Secondary vertex such as shown here is useful to distinguish a b-hadron jet from other jets which typically don't have such displaced vertices.

4.6 Missing Transverse Momentum E_T^{miss}

When neutrinos are produced in an collision, they pass through the detector without leaving any tracks or any energy depositions in the detector so their presence can only be inferred. Neutrinos can be detected by using the conservation of momentum in the x-y plane, considering that the total transverse momentum of pp collision in x-y plane is essentially 0, relative to the magnitude of the momentum involved in the experiment. If a neutrino is produced in an event, we can observe that the total transverse momentum p_T of all detected particles in this event will not sum to 0, as the neutrino will not be detected. Then the net p_T can be treated as the neutrino, with its direction opposite. So from the net p_T , the ϕ direction of neutrino can be determined, while E_T^{miss} is the magnitude.

Because the E_T^{miss} can be helpful to reduce the diboson background, which produces neutrinos. The missing transverse momentum E_T^{miss} is defined in Equation 4.1 as below:

$$\begin{aligned}
 E_T^{miss} &= \sqrt{(E_x^{miss})^2 + (E_y^{miss})^2} \\
 E_x^{miss} &= -\left(\sum_{electrons} p_x + \sum_{muons} p_x + \sum_{jets} p_x + \sum_{soft\ items} p_x \right) \\
 E_y^{miss} &= -\left(\sum_{electrons} p_y + \sum_{muons} p_y + \sum_{jets} p_y + \sum_{soft\ items} p_y \right)
 \end{aligned} \tag{4.1}$$

The magnitude of the negative vector sum of the transverse momenta of preselected electrons, muons, photons and jets, and soft items to account for charged-particle tracks compatible with the primary vertex and not associated with any of these selected objects [27]. E_T^{miss} is calculated from inner detector tracks matched to the primary vertex to make it more resilient to contamination from pile-up interactions.

4.7 Object overlap removal

Sometimes one or more of these objects will overlap with each other in the detector. To select only good events, a procedure called "Overlap Removal" is used to isolate the objects of interest.

After the reconstruction of electrons, muons and jets, electrons are discarded if they share ID tracks with the selected muon candidates. Jets are discarded if they are within a cone of size $\Delta R < 0.2$ around the direction of the electron or muon candidate. If the distance between a jet and an electron candidate is within $0.2 < \Delta R < 0.4$, the jet is retained and the nearby electron is rejected. However, a jet within $0.2 < \Delta R < 0.4$ of a muon candidate is retained only when it has at least three associated tracks, otherwise the muon is retained. The overlap removal is done by the package `OverlapRemovalTool` [79], corresponding to the standard configuration.

CHAPTER 5. EVENT PRE-SELECTION

5.1 Data and Good Runs List

The analysis is based on the full Run2 dataset from 2015-2018 proton-proton collisions with center-of-mass energy $\sqrt{s} = 13\text{TeV}$ collected by the ATLAS experiment.

Good Runs List(GRL) are used to define a set of data-taking runs and luminosity blocks for which the data was found to be of good enough quality for further analysis. The time unit in which ATLAS luminosity data is recorded is called luminosity block. It is a roughly 2min. interval during which the luminosity is supposed to remain constant. Luminosity block is the smallest time interval for which integrated luminosity or cross section can be calculated. A good run list is formed by applying data quality criteria to the list of all valid physics runs and luminosity blocks. Only those data events in GRL can be considered for final analyses for which the full detector was properly functioning.

In the analysis, the datasets used are listed in Table 5.1. The events are required to be taken during stable beam status and have good data-quality, which are recorded in the Good Run List taken from the official Data Preparation Group webpage [42].

The integrated luminosity is calculated by the ATLAS luminosity calculation tool. It is derived, following a methodology similar to that detailed in Ref. [9], from calibrations of the luminosity scale using x-y beam-separation scans.

Table 5.1: Datasets in the analysis.

Year	$\int \mathcal{L} dt$ (fb^{-1})	average $\langle\mu\rangle$
2015	3.22 ± 0.068	13.4
2016	32.99 ± 0.726	25.1
2017	44.31 ± 1.063	37.8
2018	59.94 ± 1.198	36.1

5.2 Main Backgrounds

The major backgrounds for the $Z \rightarrow e\mu$ process are $t\bar{t}$, WW and $Z \rightarrow \tau\tau$.

Backgrounds to this search comprise all those processes predicted by the Standard Model which may result in a pair of different flavor leptons. This includes processes which produce $e\mu$ directly, e.g. top-quark pair production, as well as indirect processes where one or more particles are misidentified by the detector.

The dominant SM processes directly producing lepton pairs are as follows:

- $t\bar{t}$ production, where the top quarks subsequently decay leptonically to produce e and μ .
- top quark in association with a W boson, where both decay leptonically.
- Diboson (WW, WZ, ZZ) production.
- $Z \rightarrow \tau\tau$, where the taus subsequently decay to produce e and μ .

Of these processes, $t\bar{t}$ and tW are dominant if significant E_T^{miss} is required. $Z \rightarrow \tau\tau$ has a large cross-section, but produces events with lower E_T^{miss} .

Indirect $e\mu$ events arise from fake leptons, for example from a W + jets event with one jet falsely identified as a lepton, or from a multi-jet event with two jets misidentified. And those contributions are not easy to be estimated, which is also the reason that in this analysis we use the data driven fitting method to estimate the backgrounds, which will be explained in details in Chapter 6.5.

5.3 Monte Carlo Samples

Monte-Carlo (MC) simulation is needed to understand the background composition, optimize the selection criteria, calculate the signal selection efficiency, and estimate the systematic uncertainty. The MC samples used in this analysis are the official MC16a, MC16d and MC16e production. They are summarized in Appendix D.

The signal events are simulated with Pythia 8.186 [81]. The A14 set of tuned parton shower parameters [4] is used together with the NNPDF2.3LO parton distribution function (PDF) set [31].

W and Z bosons are created primarily via quark anti-quark interactions in the pp collisions at the LHC. Due to the strong force, W and Z bosons are often produced in association with gluons and quarks which produce jets at the LHC. The $W, Z + \text{jet(s)}$ process (with W and Z decay leptonically) is modeled using Sherpa 2.2.1 [51] event generator. Matrix elements are calculated for up to 2 partons at NLO and 4 partons at LO using Comix [52] and OpenLoops [37] and merged with the Sherpa parton shower [80] according to the ME+PS@NLO prescription [60]. The CT10nlo PDF set is used in conjunction with dedicated parton shower tuning developed by the Sherpa authors. The $Z + \text{jet(s)}$ events are normalized with the NNLO cross sections.

The diboson processes with 4 charged leptons, 3 charged leptons + 1 neutrino or 2 charged leptons and 2 neutrinos are simulated with the Sherpa 2.2.2 event generator [51]. Matrix elements contain all diagrams with four electroweak vertices. They are calculated for up to 1 parton at NLO and up to 3 partons at LO using Comix [52] and OpenLoops [37], and merged with the Sherpa parton shower [80] according to the ME+PS@NLO prescription [60]. The NNPDF3.0nnlo PDF set is used in conjunction with dedicated parton shower tuning developed by the Sherpa authors. The event generator cross sections are used in this case (already at NLO).

For the generation of $t\bar{t}$ events, Powheg-Box v2 [5] is used with the CT10 PDF set in the matrix element calculations. Electroweak t-channel, s-channel and Wt -channel single top-quark events are generated with Powheg-Box v1. This event generator uses the 4-flavor scheme for the NLO matrix element calculations together with the fixed four-flavor PDF set CT10f4. For all top processes, top-quark spin correlations are preserved (for t-channel, top quarks are decayed using MadSpin [25]). The parton shower, hadronisation, and the underlying event are simulated using Pythia 6.428 [67] with the CTEQ6L1 PDF set and the corresponding Perugia 2012 set of tuned parameters (P2012) [82]. The top mass is set to 172.5 GeV. The EvtGen 1.2.0 program [41] is used for the properties of b- and c-hadron decays.

A single event recorded by the ATLAS detector consists of the superposition of a hard-scattering proton-proton collision and several additional proton-proton interaction vertices referred to as pile-up which is mentioned in Chapter 2.7. The effect of the pile-up was included by overlaying collisions,

simulated by Pythia 8.186 with a set of tuned parameters referred to as the A2 tune [10] and the MSTW2008LO PDF [68], on each generated signal and background event. The average number of interactions per pp bunch crossing μ of the simulated samples are re-weighted to reproduce the distribution observed in data.

The detector response was simulated within a framework [15] based on GEANT4 [22]. 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.

5.4 Event cleaning

The event cleaning is applied as the first filter to select the good events. We apply the general event quality criteria following [43] and reject an event in either of the two cases:

- Not passing the data quality criteria (GRL)
- Marked as corrupt due to malfunction of LAr, TileCal or SCT subdetectors

Events are required to have a primary vertex with at least two associated tracks. The primary vertex is selected as the one with the largest $\sum p_T^2$, where the sum is over all tracks with transverse momentum $p_T > 0.5$ GeV that are associated with the vertex.

We keep the events passing the event cleaning requirement, which is implemented by Event-CleaningTool. Only the events in which the jets pass the criteria needed for cleaning consideration (passing p_T , η cuts, along with JVT and OR requirements mentioned in Chapter 4) are kept. As the official jet cleaning recommendation, the event level decision is taken as a logic AND of all jet cleaning decisions; if any event has an unclean jet, the whole event is flagged as unclean and rejected.

5.5 Trigger Selection

After the event cleaning, we will choose the events passing the triggers which are properly chosen for the analysis.

The trigger system has been previously introduced in Chapter 2.6. There are two kinds of triggers: unprescaled and prescaled. Some events are so rare that every single one can be saved (for example events with a very high energy muon). Triggers for such events are said to be unprescaled. On contrary, prescaled triggers are ones which are saved only some of the events, e.g. a trigger with a prescaled of 3 will save 1 out of every 3 events fire it.

The triggers used in this analysis are the lowest threshold unprescaled single-lepton triggers [56] in every data-taking period, as summarized in Appendix A.

As luminosities become higher along with period, the rates at which triggers fire also increase. In order for a trigger to remain unprescaled, the E_T threshold has to be increased. The offline object reconstruction criteria(details shown in Section 4) were chosen to be 1GeV above the trigger E_T thresholds to ensure that the corresponding efficiencies of leptons are at the plateau region and any effects from the calibration of lepton pT do not result in widely different efficiencies which may cause the large systematic uncertainties.

5.6 Event Reweighting

In a high-luminosity pp collider such as the LHC, there is a non-negligible probability that one single bunch crossing may produce several pp collisions called pileup events. The pileup concept has been explained previously in Chapter 2.7. While the Monte Carlo simulated events are generated assuming constant beam conditions, but the conditions vary during real data-taking, resulting in a different amount of soft interactions overlapping with the hard interaction. This results in different amount of pileup events in Monte Carlo compared to data.

The true pileup, average number of interactions per bunch crossing $\langle\mu\rangle$ distribution from real data events included in the GRL, is used to reweight the events in MC to account for different levels of pileup so that they properly match the real collected data.

In addition, other kinds of weights are applied on MC simulated events to correct for the inaccuracies in object selections. Electrons and Muons have weights applied to correct for reconstruction and isolation efficiencies, as well as the efficiencies of their corresponding triggers (see Section 4 for details).

Finally, MC events are scaled to match the total integrated luminosity of the real data used in the analysis.

5.7 Event Pre-selection

Besides the event cleaning and trigger firing requirements and according to the analysis strategy described in Chapter 3, we need to apply additional cuts on the events in order to select out our ideal events candidates for the $Z \rightarrow e\mu$ signal analysis. Those cuts described below together is called pre-selection. And the same criteria are implied in pre-selection on both $Z \rightarrow e\mu$ and the reference channels $Z \rightarrow ee, \mu\mu$ except for the flavor constraint of the lepton pair.

All the criteria in pre-selection are listed below:

- Pass event cleaning and GRL.
- Implement pass Trigger requirement. At least one of the lowest unprescaled single lepton triggers fired. If both Egamma and Muon triggers fired, remove the event from the Egamma stream, which.
- Implement Primary Vertex and Overlap Removal requirement.
- Contain no more than 2 loose leptons who pass the "loose electron" or "loose muon" requirements in events as the additional lepton veto cut.
- Require exactly two loose leptons passing the tight selection criteria(tight signal lepton) with consistent flavor for $Z \rightarrow ee, \mu\mu, e\mu$ channels.

- Implement the trigger matching. One offline reconstructed lepton with $p_T > 27$ GeV must match an online reconstructed lepton, according to the lowest unscaled single lepton trigger p_T threshold.
- Require the two selected leptons to be oppositely charged.
- Require the invariant mass of the lepton pair to be within the window $70 < m_{ll} < 110$ GeV.
- Veto events who have the leading jet with transverse momentum $p_T^{\text{leading jet}} > 60$ GeV, shown in the left plot of Figure 5.1, the first bin of 0 Jets_leading_pT in the left plot is the events without any qualified jet.
- Veto events with large transverse missing energy $E_T^{\text{miss}} > 50$ GeV to remove backgrounds with neutrinos, shown in the right plot of Figure 5.1.
- Veto events containing b -jets to remove $t\bar{t}$ and single top backgrounds. Figure C.1 shows the distribution of the number of b -jets at different Working Points (WP). We choose the tightest WP with 85% b -tagging efficiency to remove more backgrounds.

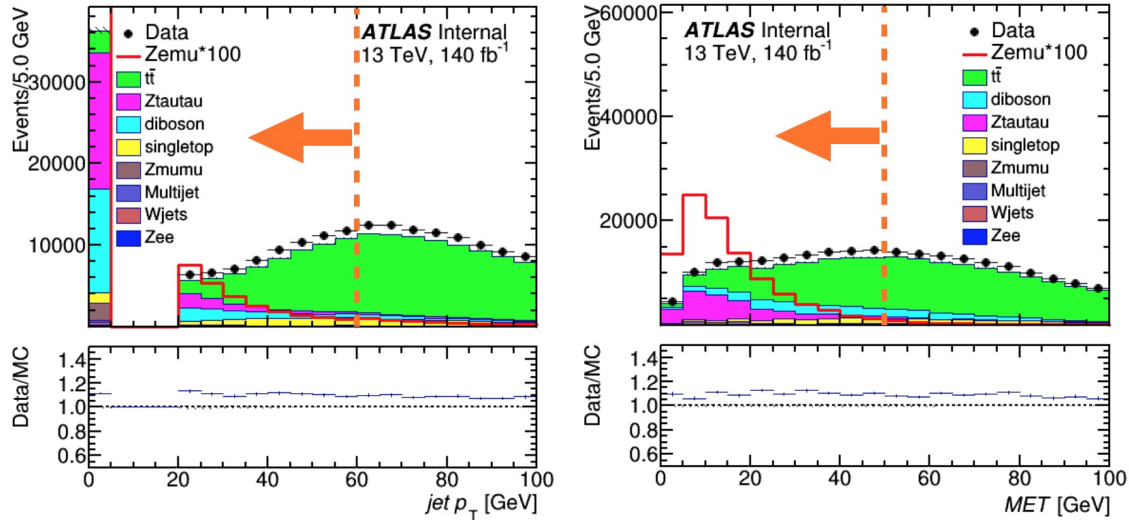


Figure 5.1: Cuts applied on the transverse momentum of the leading jet (left) and the transverse missing energy (right) in pre-selection. Signal MC is scaled to the run1 branching ratio upper limit 7.5×10^{-6} .

The cut efficiencies are summarized in Tables 5.2, showing the efficiency of each selection criterion.

After implementing the pre-selection, the comparisons of the kinematic variables between data and MC are illustrated in Figures 5.2 - 5.4. The QCD background in the $Z \rightarrow e\mu$ channel is estimated by using same-sign data, i.e., events with same-charge lepton pairs in the final states. The contribution from SM backgrounds in the same-sign events are subtracted based on MC. Both statistic and systematic errors are included for the $Z \rightarrow e\mu$ channel, while for the $Z \rightarrow ee, \mu\mu$ channels only statistical errors are included.

The background composition after pre-selection can be understood from Figure 5.4. The remaining backgrounds mainly include the $Z \rightarrow \tau\tau \rightarrow e\mu\nu\nu\nu$ events, the diboson process $WW \rightarrow e\mu\nu\nu$, the $Z \rightarrow \mu\mu$ events with one muon faked as an electron, and the $t\bar{t} \rightarrow e\mu\nu\bar{\nu}b\bar{b}$ decays. We explore a multivariate discriminant technique to further suppress these backgrounds.

Table 5.2: Summary of $Z\mu/Zee/Z\mu\mu$ channels outflow table

Channel:	Zee channel				Zee channel				Zee channel			
	Zee MC16a		Data15+16		Zee MC16d		Data17		Zee MC16d		Data17	
	#Entries	each cut eff %	#Entries	each cut eff %	#Entries	each cut eff %	#Entries	each cut eff %	#Entries	each cut eff %	#Entries	each cut eff %
Event Cuts												
Initial	2.58E+07	100.00%	6.35E+09	1.27%	3.10E+07	100.00%	4.61E+09	1.89%	3.10E+07	100.00%	4.61E+09	1.89%
DAOD level	2.58E+07	100.00%	8.05E+07	78.81%	3.10E+07	100.00%	8.71E+07	78.16%	3.10E+07	100.00%	8.71E+07	78.16%
passTrigger	2.49E+07	96.43%	6.35E+07	100.00%	2.87E+07	100.00%	6.81E+07	100.00%	2.87E+07	100.00%	6.81E+07	100.00%
PrimaryVertex	2.49E+07	100.00%	6.35E+07	38.52%	2.87E+07	100.00%	6.81E+07	39.30%	2.87E+07	100.00%	6.81E+07	39.30%
exact two loose leptons	2.25E+07	90.46%	2.44E+07	48.42%	2.59E+07	49.38%	2.68E+07	45.77%	2.59E+07	49.38%	2.68E+07	45.77%
exact two tight leptons	1.15E+07	51.32%	1.18E+07	100.00%	1.28E+07	100.00%	1.23E+07	100.00%	1.28E+07	100.00%	1.23E+07	100.00%
Trigger Match	1.15E+07	100.00%	1.18E+07	99.40%	1.28E+07	99.28%	1.23E+07	99.34%	1.28E+07	99.28%	1.23E+07	99.34%
Opposite Charge	1.15E+07	96.75%	1.18E+07	95.67%	1.27E+07	96.43%	1.16E+07	95.22%	1.27E+07	96.43%	1.16E+07	95.22%
b-veto	1.11E+07	95.91%	1.13E+07	95.44%	1.22E+07	95.80%	1.11E+07	95.40%	1.22E+07	95.80%	1.11E+07	95.40%
70 < m < 110	1.06E+07	98.50%	1.07E+07	98.46%	1.17E+07	97.45%	1.08E+07	97.39%	1.17E+07	97.45%	1.08E+07	97.39%
MET < 50GeV	1.05E+07	95.09%	1.06E+07	94.77%	1.14E+07	94.78%	1.02E+07	94.59%	1.14E+07	94.78%	1.02E+07	94.59%
pTmaxjet < 60GeV	9.97E+06		1.00E+07		1.08E+07				1.08E+07			
Channel:	Zmumu channel				Zmumu channel				Zmumu channel			
	Zmumu MC16a		Data15+16		Zmumu MC16d		Data17		Zmumu MC16d		Data17	
	#Entries	each cut eff %	#Entries	each cut eff %	#Entries	each cut eff %	#Entries	each cut eff %	#Entries	each cut eff %	#Entries	each cut eff %
Event Cuts												
Initial	3.07E+07	100.00%	6.35E+09	1.27%	3.72E+07	100.00%	4.61E+09	1.89%	3.72E+07	100.00%	4.61E+09	1.89%
DAOD level	3.07E+07	100.00%	8.05E+07	78.81%	3.72E+07	100.00%	8.71E+07	78.16%	3.72E+07	100.00%	8.71E+07	78.16%
passTrigger	2.94E+07	95.84%	6.35E+07	100.00%	3.45E+07	100.00%	6.81E+07	100.00%	3.45E+07	100.00%	6.81E+07	100.00%
PrimaryVertex	2.94E+07	100.00%	6.35E+07	45.73%	3.45E+07	100.00%	6.81E+07	45.30%	3.45E+07	100.00%	6.81E+07	45.30%
exact two loose leptons	2.70E+07	91.65%	2.90E+07	59.17%	3.17E+07	55.80%	3.09E+07	54.89%	3.17E+07	55.80%	3.09E+07	54.89%
exact two tight leptons	1.62E+07	59.92%	1.72E+07	100.00%	1.77E+07	100.00%	1.69E+07	100.00%	1.77E+07	100.00%	1.69E+07	100.00%
Trigger Match**	1.58E+07	97.95%	1.72E+07	99.99%	1.77E+07	99.99%	1.69E+07	99.99%	1.77E+07	99.99%	1.69E+07	99.99%
Opposite Charge	1.58E+07	100.00%	1.72E+07	95.90%	1.77E+07	96.54%	1.62E+07	95.44%	1.77E+07	96.54%	1.62E+07	95.44%
b-veto	1.53E+07	96.91%	1.65E+07	95.46%	1.71E+07	95.78%	1.54E+07	95.39%	1.71E+07	95.78%	1.54E+07	95.39%
70 < m < 110	1.47E+07	95.87%	1.57E+07	98.41%	1.63E+07	97.41%	1.50E+07	97.35%	1.63E+07	97.41%	1.50E+07	97.35%
MET < 50GeV	1.45E+07	98.48%	1.55E+07	94.99%	1.59E+07	94.99%	1.42E+07	94.80%	1.59E+07	94.99%	1.42E+07	94.80%
pTmaxjet < 60GeV	1.38E+07	95.36%	1.47E+07		1.51E+07				1.51E+07			
Channel:	Zemu channel				Zemu channel				Zemu channel			
	Zemu MC16a		Data15+16		Zemu MC16d		Data17		Zemu MC16d		Data17	
	#Entries	each cut eff %	#Entries	each cut eff %	#Entries	each cut eff %	#Entries	each cut eff %	#Entries	each cut eff %	#Entries	each cut eff %
Event Cuts												
Initial	388,390	100.00%	6.35E+09	1.27%	468,086	100.00%	4.61E+09	1.89%	468,086	100.00%	4.61E+09	1.89%
DAOD level	388,390	100.00%	8.05E+07	78.81%	468,086	100.00%	8.71E+07	78.16%	468,086	100.00%	8.71E+07	78.16%
passTrigger	376,285	96.88%	6.35E+07	100.00%	436,469	100.00%	6.81E+07	100.00%	436,469	100.00%	6.81E+07	100.00%
PrimaryVertex	376,285	100.00%	6.35E+07	35.41%	436,469	91.91%	6.81E+07	38.82%	436,469	91.91%	6.81E+07	38.82%
exact two loose leptons	345,998	91.95%	2.25E+07	1.27%	401,151	52.40%	2.64E+07	1.11%	401,151	52.40%	2.64E+07	1.11%
exact two tight leptons	193,327	55.88%	285,914	99.97%	210,204	100.00%	292,884	99.98%	210,204	100.00%	292,884	99.98%
Trigger Match	184,949	95.67%	285,838	97.72%	210,204	96.42%	285,871	97.63%	210,204	96.42%	285,871	97.63%
Opposite Charge	184,147	99.57%	279,309	26.04%	209,202	97.52%	73,641	25.76%	209,202	97.52%	73,641	25.76%
b-veto	178,257	96.80%	72,740		201,719	97.52%			201,719	97.52%		
70 < m < 110	174,077	97.66%			196,724	97.50%			196,724	97.50%		
MET < 50GeV	171,569	98.56%			191,802	95.34%			191,802	95.34%		
pTmaxjet < 60GeV	164,127	95.66%			182,866				182,866			

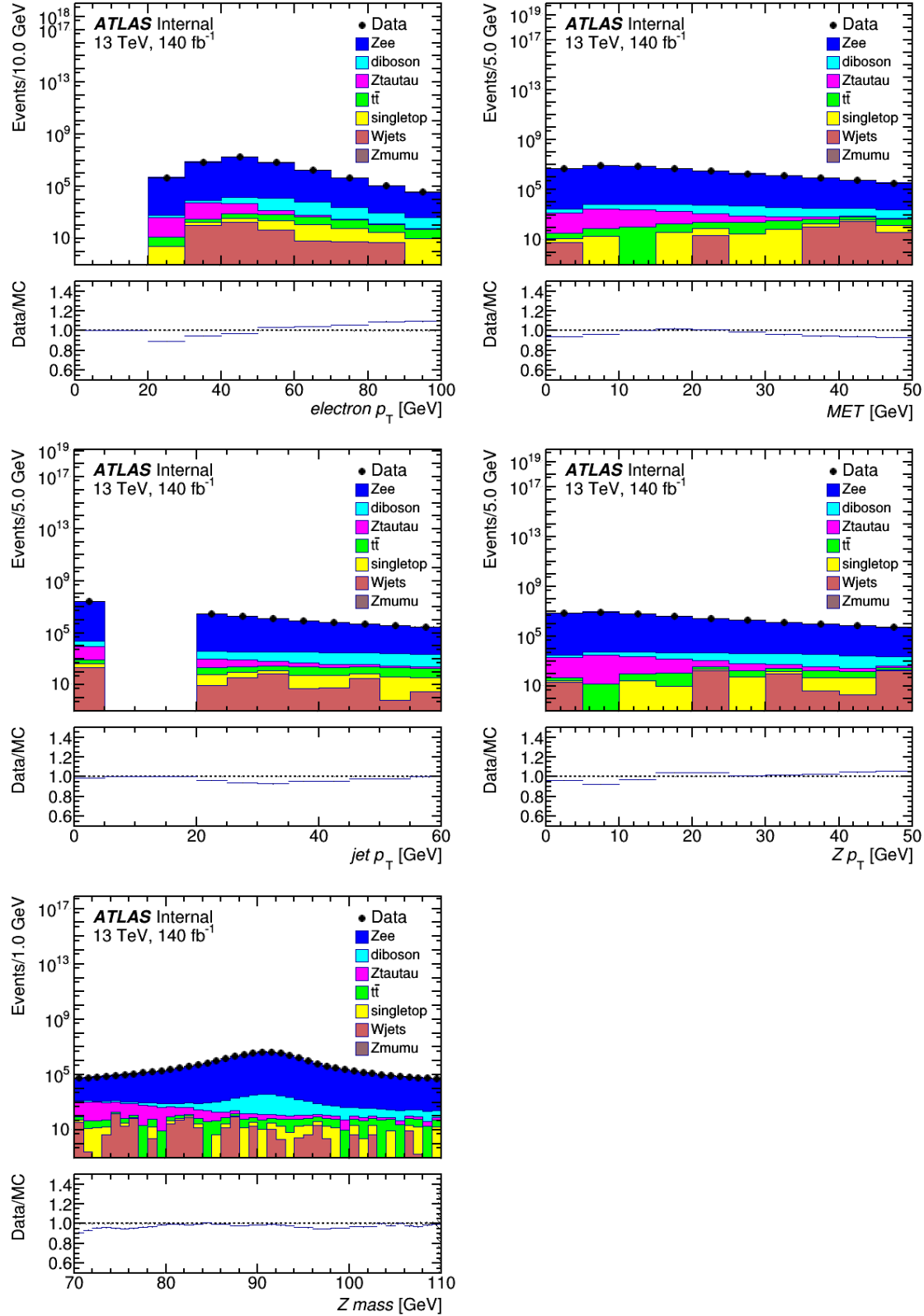


Figure 5.2: Data and MC comparison of kinematic variables in $Z \rightarrow ee$ after pre-selection. Top left: electron transverse momentum; Top right: missing transverse momentum; Middle left: transverse momentum of the leading jet; Middle right: transverse momentum of the reconstructed Z boson; Bottom left: invariant mass spectrum of the reconstructed Z boson.

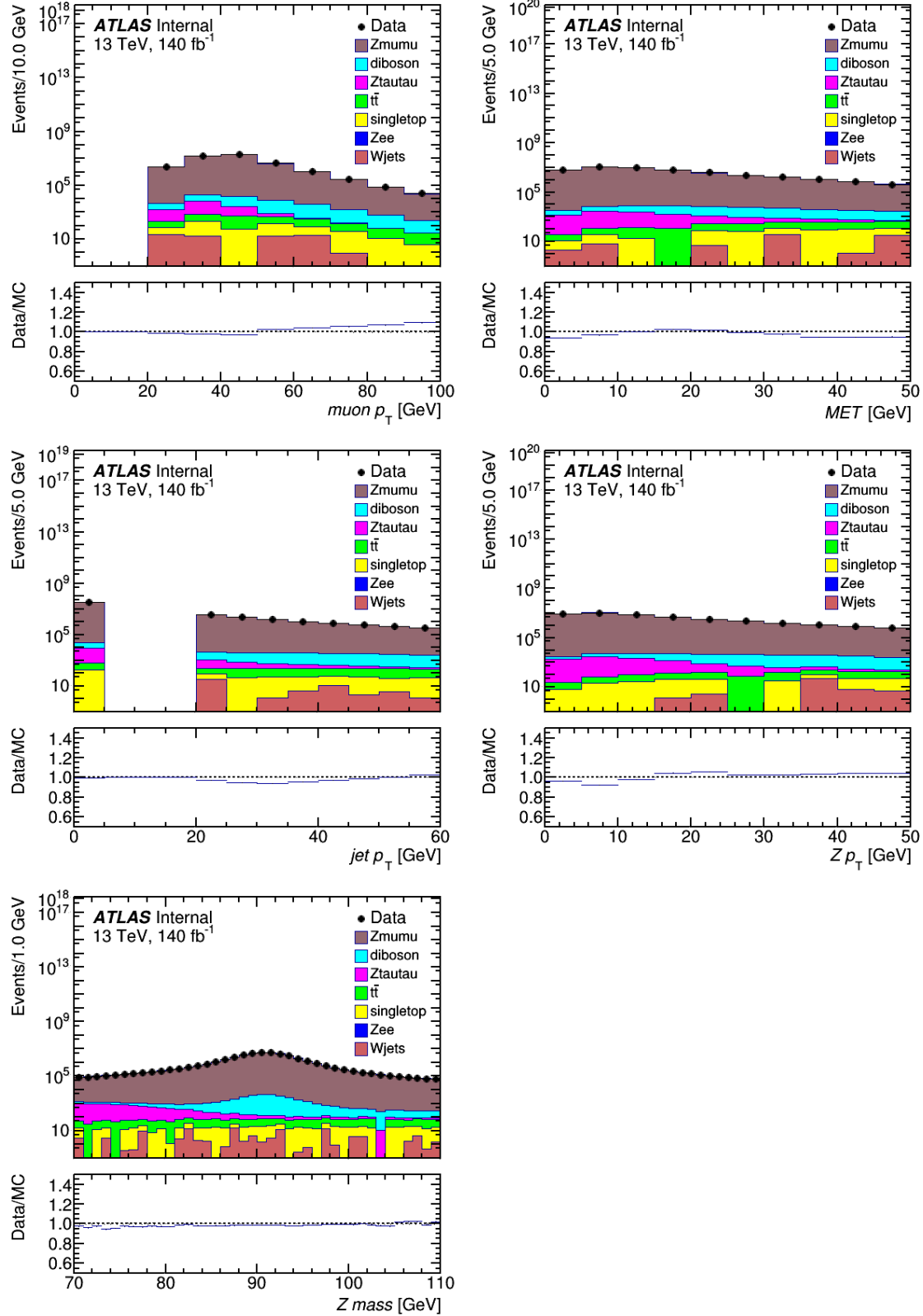


Figure 5.3: Data and MC comparison of kinematic variables in $Z \rightarrow \mu\mu$ after pre-selection. Top left: muon transverse momentum; Top right: missing transverse momentum; Middle left: transverse momentum of the leading jet; Middle right: transverse momentum of the reconstructed Z boson; Bottom left: invariant mass spectrum of the reconstructed Z boson.

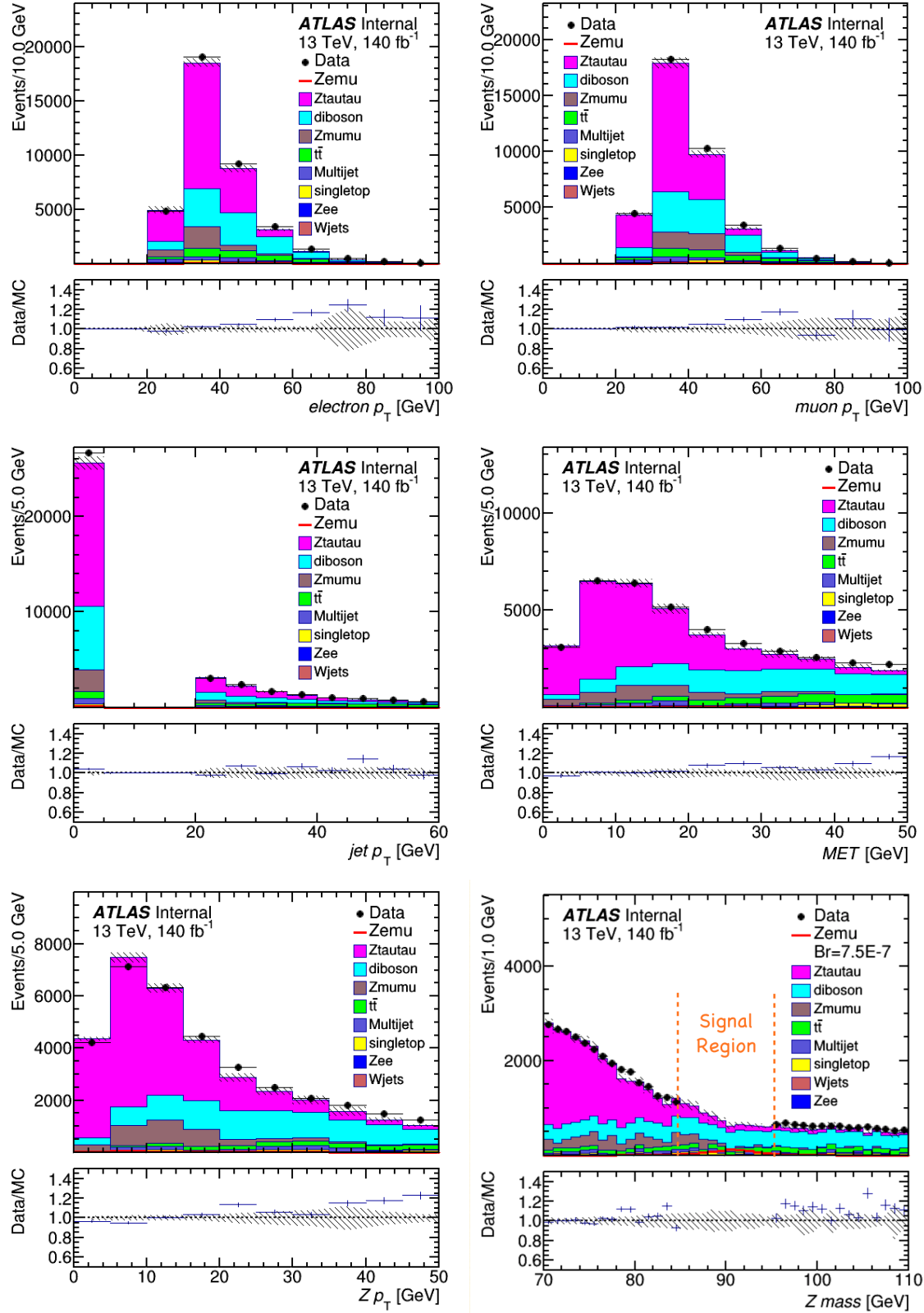


Figure 5.4: Data and MC comparison of kinematic variables in $Z \rightarrow e\mu$ after pre-selection. Top left: electron transverse momentum; Top right: muon transverse momentum; Middle left: leading jet transverse momentum; Middle right: missing transverse momentum; Bottom left: reconstructed Z boson transverse momentum; Bottom right: reconstructed Z boson invariant mass spectrum.

CHAPTER 6. EVENT FINAL-SELECTION

A multivariate technique is implemented to enhance the signal relative to the backgrounds using the TMVA tool [61]. It will potentially perform better than the traditional cut-based method because it can take advantage of the correlation between variables. In the Run I analysis, the selection criteria is optimized by scanning the two-dimensional $(E_T^{miss}, p_T^{\text{leading jet}})$ space. We compare the sensitivities obtained from the multivariate technique and the two-dimensional cut optimization, and observe a 7% improvement by the former. More details about the cut-based method are given in Appendix B.

6.1 Multivariate Technique

The advantage of multivariate analysis is that it's good at finding the local optimal boundary between signal and background, which can not be achieved by the traditional cut-based selection if the discriminant power of single cut variables are not strong enough, or especially the variables are correlated with each other nonlinearly.

There is an example which can illustrate this difference between multivariate analysis and traditional cut based selection, Figure 6.1 and Figure 6.2.

We can see in Figure 6.1, if var0 and var1 are nonlinearly correlated for separating the signal and background, then it's hard for the traditional cut-based selection to have strong discriminant power. The reason is the cut-based selection is only good at handling the linear rectangular relation at the boundary between signal and background, and the performance will deteriorate when there exist the correlations among the variables.

On the other hand, in Figure 6.2, we can see the multivariate analysis is a good choice for the nonlinear situation. Because the multivariate analysis can learn and reconstruct the boundary gradually based on the training sample. And the higher dimensions(variables, where the name

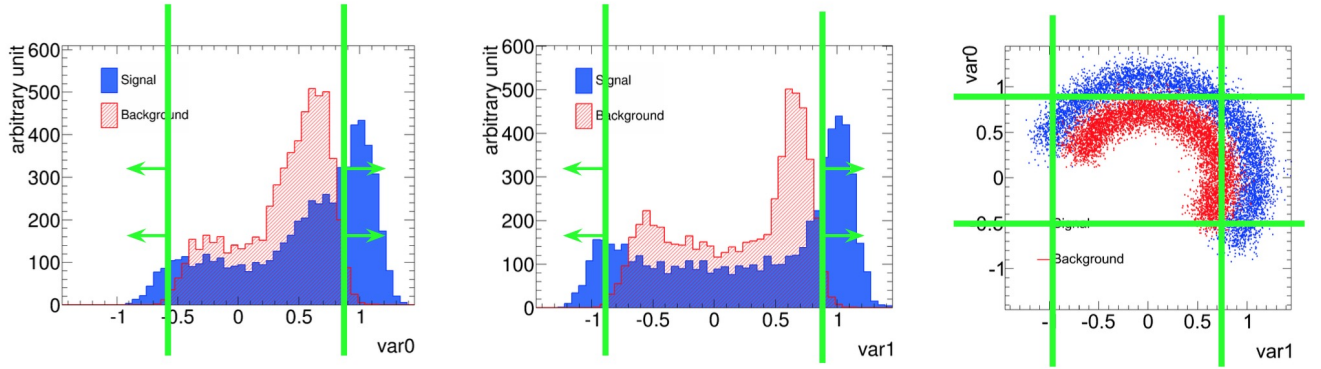


Figure 6.1: Cut based selection

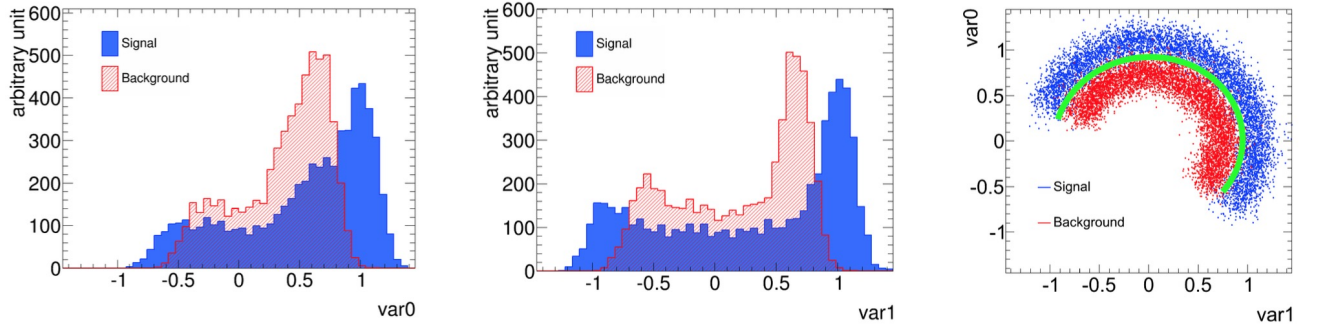


Figure 6.2: Multivariate analysis selection

multivariate is originated), the more the points are spread in the high dimensional space, then the better the possible separation power of the multivariate model.

In brief, we can treat the cut-based selection as the box boundary to separate the signal and background in the high dimension space, while the multivariate analysis has the non-linear hyper-plane boundary between signal and background, which can have better discriminant performance when the signal and background are mixed together and hard to be separated linearly.

But multivariate analysis has its own weaknesses, the nonlinear discriminant power is always with the sacrifice of interpretation when we use the complicated machine learning or deep learning models, like neural network. Besides the lack of interpretation ability, the overfitting which results

from trade off between bias and variance is also the issue of multivariate analysis, which is shown in Figure 6.3.

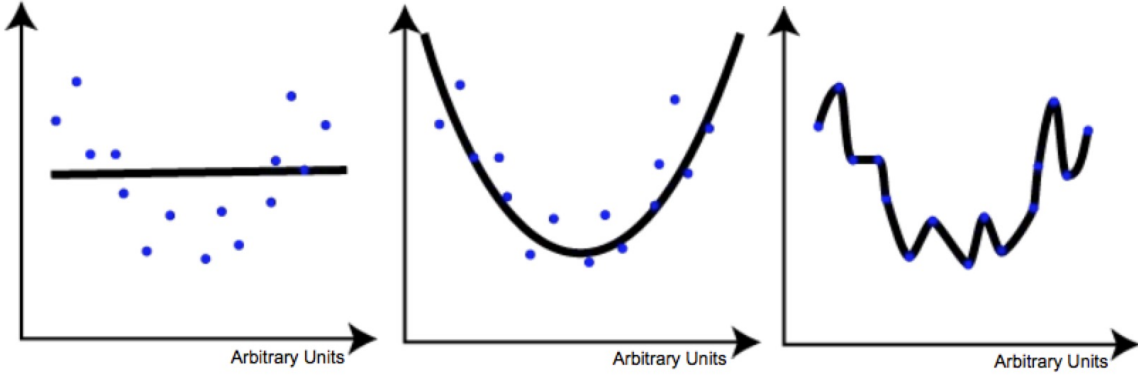


Figure 6.3: The blue points have a parabolic dependence. The black lines are attempts to represent the pattern of the data. The left most figure shows under-fitting, the middle line shows a good fit, and the last line is overfitting.

The three graphs above show the same dataset depicted with blue dots, where the true data has a parabolic dependence. The data points are smeared around the true parabolic values by Gaussian noise. The black lines in the figures are attempts to fit the data points. On the left the line is under-fit, it does not fit the shape of the data. The central figure is a good fit. The right figure is the result of over-fitting. It represents this particular data very well, but additional data will not follow this pattern and will result in poor performance overall.

So we need to avoid overfitting of training of Multivariate model in the analysis, and at the same time ensure the good discriminant power.

Based on this overfitting concern, we need to divide both the signal MC and background MC into two parts:

(1) One is the training sample, which is used to train the Multivariate model to have a good discriminant power, with the signal MC and background MC explicitly tagged.

(2) The other one is the test sample, which is given to the Multivariate model obtained from training sample, to test the performance on the test sample dataset and ensure the Multivariate model is not overfitting.(if the model is overfitting, it will perform good on training sample but poorly on the test sample)

We normally set the fraction of training sample and test sample to be half/half and they are chosen randomly, to avoid the manual selection bias of input training sample.

Next, we need to choose the certain Multivariate model. A comparison of models with strengths and weaknesses are listed in the Figure 6.4 [84]:

		MVA METHOD									
CRITERIA		Cuts	Likeli- hood	PDE- RS / k-NN	PDE- Foam	H- Matrix	Fisher / LD	MLP	BDT	Rule- Fit	SVM
Perfor- mance	No or linear correlations	★	★★	★	★	★	★★	★★	★	★★	★
	Nonlinear correlations	○	○	★★	★★	○	○	★★	★★	★★	★★
Speed	Training Response	○ ★★	★★ ★★	★★ ○	★★ ★	★★ ★★	★★ ★★	★ ★★	★ ★	★ ★★	○ ★
	Overtraining Weak variables	★★ ★★	★ ★	★ ○	★ ○	★★ ★★	★★ ★★	★ ★	★ ³⁹ ★★	★ ★	★★ ★
Curse of dimensionality		○	★★	○	○	★★	★★	★	★	★	
Transparency		★★	★★	★	★	★★	★★	○	○	○	○

Figure 6.4: Assessment of MVA method properties. The symbols stand for the attributes "good" (★★), "fair" (★) and "bad" (○). Curse of dimensionality refers to the burden of required increase in training statistics and processing time when adding more input variables.

6.2 Gradient Boosted Decision Tree (BDTG)

For the multivariate technique we choose the Gradient Boosted Decision Tree (BDTG) classifier. The BDTG classifier is commonly used in data analysis because of its fast calculation speed and

its strong discriminant ability especially when non-linear correlation exists among variables. It has an advantage over other methods especially when the training sample is not very large or the input variables individually do not have strong discriminant power. Besides, little tuning on hyperparameters is needed to obtain reasonably good results for BDTG.

A decision tree is a binary tree structured classifier like the one sketched in Figure 6.5. An algorithm is used to go through all the available variables and find the optimal splits to make for each of the variables. This split is called a node. Starting from the root node, a sequence of binary splits using the discriminating variables x_i is performed. Each split uses the variable that at this node gives the best separation power between signal and background with its algorithm described in [77]. The same variable may thus be used at several nodes, while others might not be used at all. The leaf nodes at the bottom end of the tree are labeled S for signal and B for background depending on the majority of events that end up in the respective nodes.

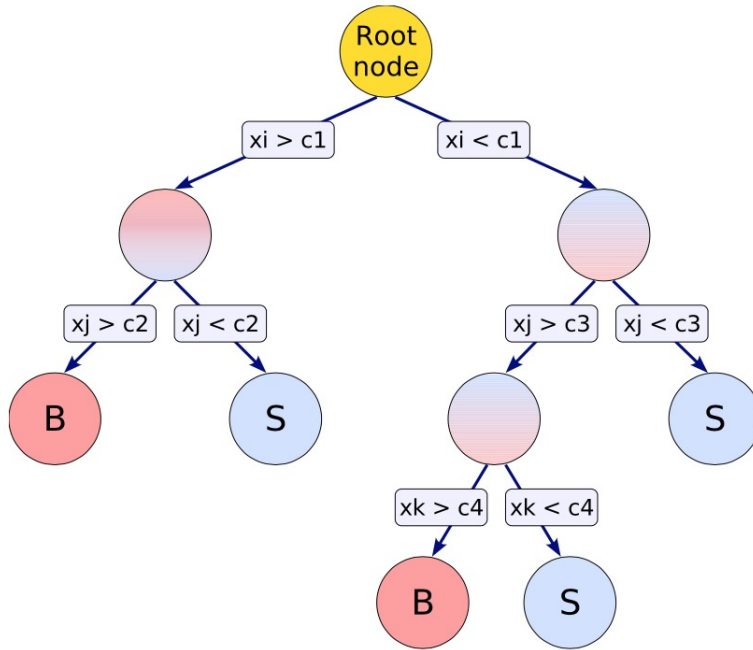


Figure 6.5: Schematic view of a decision tree.

Repeated binary left/right(yes/no) decisions are performed on a single variable(the same variable can be repeatedly used in different tree nodes) at a time until some stop criterion of minimized Gini impurity is reached. Gini impurity is defined in Equation 6.1.

$$Gini = \left(\sum_{i=1}^n W_i \right) P(1 - P). \quad (6.1)$$

Where W_i is the weight of the data point, P is the probability of correctly classifying the point, and $1 - P$ is the probability of mis-classifying the point. $Gini$ reaches 0 when all points fall into a single classification.

Like this the high dimension space is divided into regions that are eventually classified as signal or background, depending on the majority of training events that end up in this region. So with those variables cuts together, we can classify the events as signal or background according to which regions those events are in. And because the decision tree is only based on the binary decisions, we can interpret this model easily.

Decision trees are well known classifiers that allow straightforward interpretation as they can be visualized by a simple two dimensional tree structure. They are in this respect similar to rectangular cuts. However, whereas a cut-based analysis is able to select only one hypercube as region of phase space, the decision tree is able to split the phase space into a large number of hypercubes, each of which is identified as either signal-like or background-like. The path down the tree to each leaf node represents an individual cut sequence that selects signal or background depending on the type of the leaf node.

But the weakness of single decision tree is sensitive to the statistical fluctuations or outliers points of the training dataset, as the outliers will cause the Gini purity getting worse. To solve this problem, the boosting of a decision tree (BDT) is introduced. It represents an extension of a single decision tree. Several decision trees (a forest), derived from bootstrapping(a randomly sampling method) the same training sample by reweighting events, are combined to form a classifier which is given by a (weighted) majority vote of the individual decision trees. Boosting stabilizes the response of the decision trees with respect to fluctuations in the training sample, but with the sacrifice of

the interpretability as we have generated so many trees. Its weakness is if we have number of trees far more than needed, there is the possibility that the BDT model is overfitting.

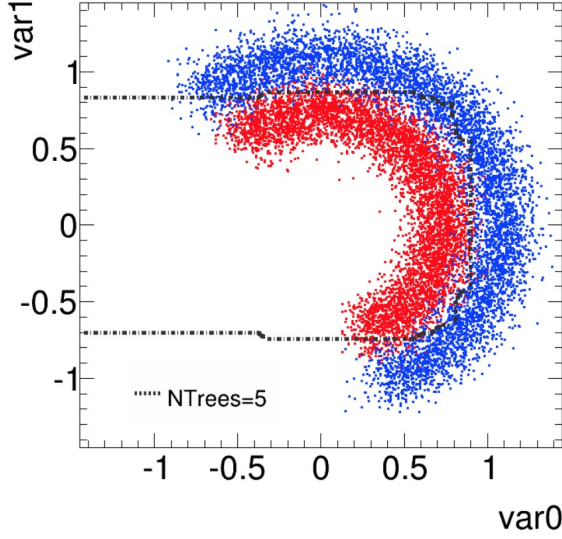
As we can see in the following Figures 6.6, the number of trees in BDT will have great influence on the level of overfitting and underfitting. The proper number of trees in boosting method should be chosen to avoid overfitting. Normally 500 trees is used as default setup, to balance the overfitting and the underfitting. More information about the hyperparameters setup will be shown in next Section 6.3.

Besides the boosting method being introduced in the BDTG model, we can combine gradient descent optimization method with boosting method, which as a whole is called "gradient boosting" method.

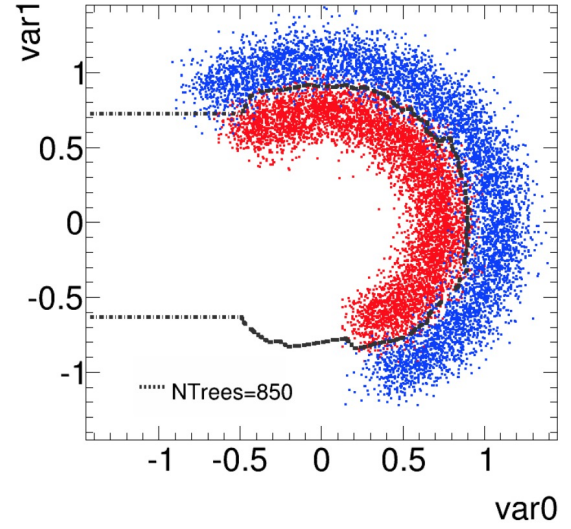
Gradient descent is an optimization algorithm used to minimize some function by iteratively moving in the direction of steepest descent as defined by the negative of the gradient. In machine learning, we use gradient descent to update the parameters of our model. So in the "gradient boosting" method, gradient descent can help boosting method to optimize the parameters more efficient.

Gradient boosting is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function. Each gradient descent step is fitted with one tree, and all the trees are all the steps down all the way to the minimum of loss function, and each tree will be assigned its gain of reduction of loss function(or like the effective step length towards the minimum) as the weight. Then the final BDTG result is the ensemble results of all the trees.

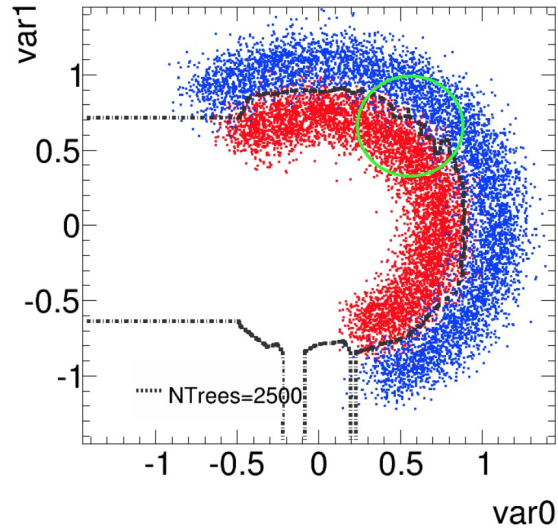
More details about gradient boosting algorithms are explained in Ref. [48], and more details of BDTG model implementation in TMVA toolkit are described in the Ref. [61].



(a) Boosting Decision Tree with 5 trees.



(b) Boosting Decision Tree with 850 trees.



(c) Boosting Decision Tree with 2500 trees.

Figure 6.6: Boosting Decision Tree performance with different number of trees (a) shows not enough trees and sub-optimal separation between signal and background, (b) shows close to best performance, (c) shows BDT is learning fluctuations and there is the tendency towards overfitting,

6.3 Event final selection based on a multivariate technique

We use the signal and background MC events falling in the SR [85,95] GeV in the machine learning process. The MC backgrounds used in machine learning include all those mentioned in Sec. 5.3 except for the $Z \rightarrow \mu\mu$ background. We exclude it because $Z \rightarrow \mu\mu$ MC events have large weights ($Z \rightarrow \mu\mu$ MCs has been generated with much fewer events in the Z lower mass region so the luminosity normalization weights are large for those low mass region events, the reweighting process is explained in Chapter 5.6), causing the training to be unstable. The event samples are split into two halves, even-numbered events for training and odd-numbered ones for test. The signal events used in training are not removed from the efficiency estimation in the later chapter. As we are not optimizing the efficiency but the FOM significance in the TMVA model training, there should be no bias introduced into the efficiency calculation.

The candidate variables we tried in machine learning are summarized in Table 6.1. Based on the importance ranking of all the variables calculated by the TMVA tool [84], we abandon the variables with poor discriminant ability and only keep the first three most powerful variables in machine learning, i.e., $p_T^{\text{leading jet}}$, E_T^{miss} and p_T^Z .

Table 6.1: Summary of the candidate variables tried in machine learning.

Input variable	Description
$p_T^{\text{leading jet}}$	Transverse momentum of the leading jet
E_T^{miss}	Missing transverse momentum
p_T^Z	Transverse momentum of the reconstructed Z boson
η_Z	Pseudo-rapidity of the reconstructed Z boson
p_T^e	Transverse momentum of the electron
p_T^μ	Transverse momentum of the muon
N_{jet}	Number of jets
$\Delta\eta$	Difference of the pseudo-rapidity between the lepton pairs
$\Delta\phi$	Difference of the azimuthal angle between the lepton pairs

The comparison between the signal and backgrounds are shown in Figure 6.7, the first bin of 0 Jets_leading_pT in the left plot is the events without any qualified jet. Obvious differences between signal and background are seen in the variable E_T^{miss} , making it have the best discriminant power.

The correlation between the input variables are given in Figure 6.8 for signal and background events, respectively. Stronger correlation between the variables $p_T^{\text{leading jet}}$ and p_T^Z exists in signal events. But for the BDTG model, the strong correlation among the variables do not affect the model discriminant performance [61].

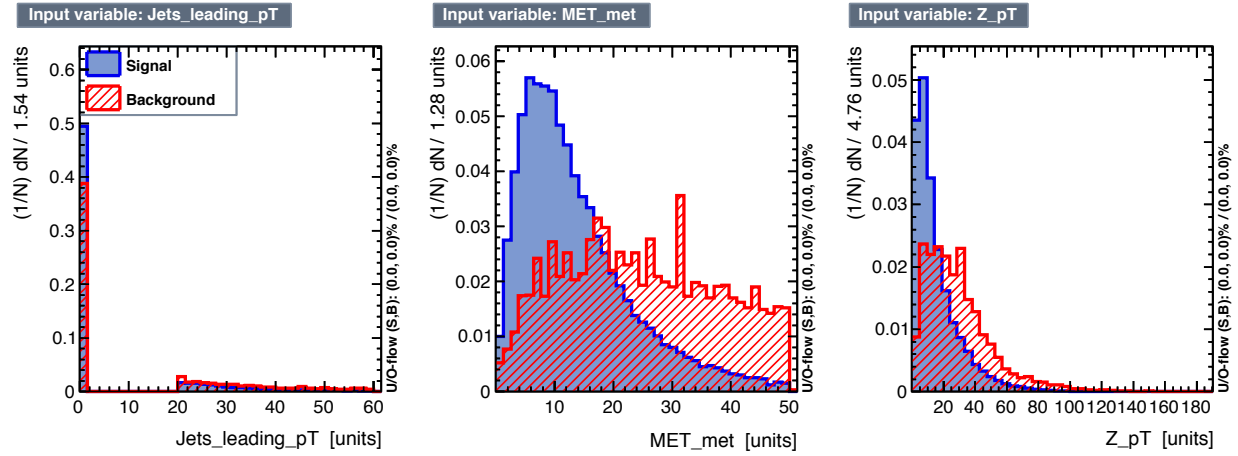


Figure 6.7: Input variables used in the Boosting Decision Tree

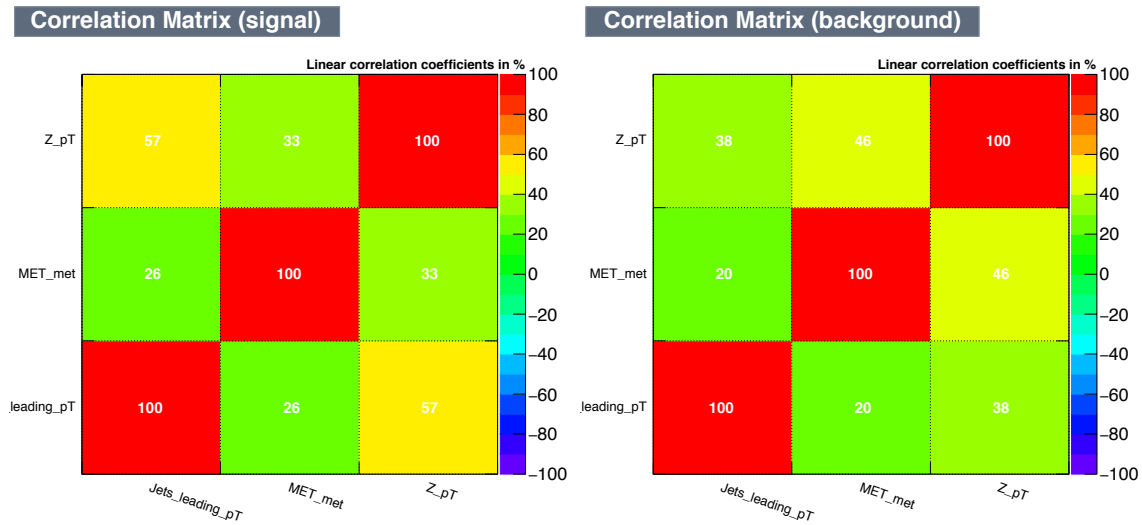


Figure 6.8: Input variables correlation matrix for the signal (left) and background (right).

The BDTG model hyperparameters configuration in the analysis is described in Table 6.2:

Table 6.2: Hyperparameters configuration of BDTG.

Hyperparameter	Values
Boost Type	GradientBoost
N trees	500
Minimum node size	2.5%
Shrinkage	0.1
Bagged Sample Fraction	0.5
ncuts	40
Max depth	2

"N Trees" is number of trees in the BDTG, "Minimum node size" is minimum percentage of training events required in a leaf node, "Shrinkage" is the learning rate for GradBoost algorithm. A technique to slow down the learning in the gradient boosting model is to apply a weighting factor for the corrections by new trees when added to the model. This weighting is called the learning rate. "BaggedSampleFraction" is relative size of bagged event sample to original size of the data sample, while bagged event sample is the sample drawn by the bagging technique which is a sampling method to increase the training dataset. "ncuts" is number of grid points in variable range used in finding optimal cut in node splitting, and "Max depth" is the max depth of the decision tree allowed, while the maximum depth is the number of nodes along the longest path from the root node down to the farthest leaf node..

With the configuration above, the distribution of the discriminant scores from the trained BDTG model are shown in Figure 6.9. We can see that signal and background are well separated. The good consistency between the training and testing samples indicate that there is no overfitting problem. Figure 6.10 gives the comparison between data and MC events in the SB region. The

different MC backgrounds are illustrated in the plot. Most of the combinatorial backgrounds can be distinguished by the machine learning.

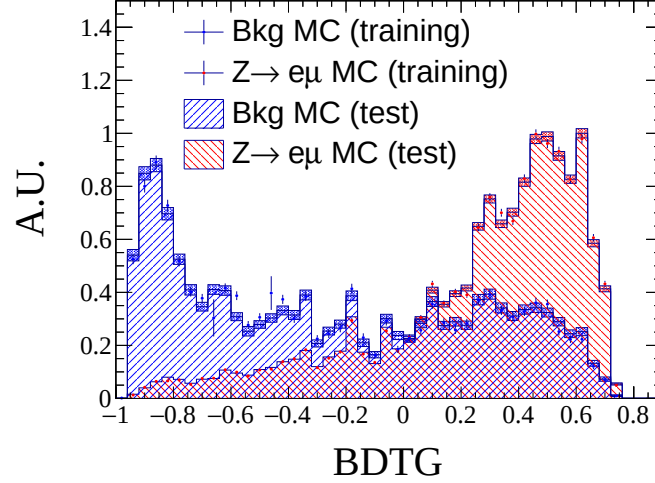


Figure 6.9: Comparison of the normalized distributions of the discriminant output between training and test samples. Only statistical uncertainties are included in the plots.

The cut on the discriminant score is optimized based on the figure of merit (FOM), defined as

$$\text{FOM} = \frac{N_{\text{sig}}}{\sqrt{N_{\text{bkg}}}}. \quad (6.2)$$

Here N_{sig} denotes events sum of weights of the signal MC events in the SR, and N_{bkg} the estimated number of backgrounds in the SR by extrapolation from fits to the SB. Details about the study on the fitting method will be given in Sec. 6.5.

We apply the trained machine learning to real data and signal MC sample and retrieve the discriminant score of each events. Then we scan the cut on the discriminant score by an increment of 0.005 to ensure the optimal cut can be achieved. For each value we calculate N_{sig} , N_{bkg} and FOM, and obtain the variation of FOM along with the cut value, as shown in Figure 6.11. The optimal cut on the discriminant output is set at 0.18.

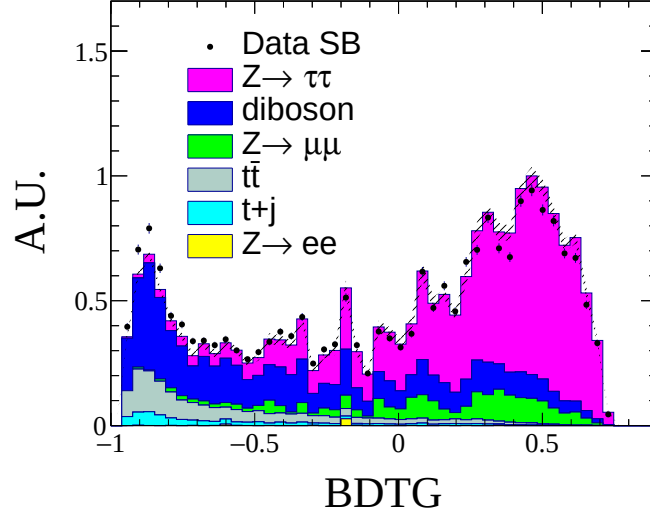


Figure 6.10: Comparison of the normalized distributions of the BDTG discriminant output between data and MC. Only statistical uncertainties are included in the plots.

6.4 Data and MC comparison after final selection

After the final selection, comparison of the kinematic variables between data and MC are given in Figure 6.12, 6.13 and 6.14 for the $Z \rightarrow e\mu, ee, \mu\mu$ channels, respectively. Data and MC agree well within one sigma standard deviation. However, one should note that in extracting the upper limit on the branching fraction of $Z \rightarrow e\mu$, the background is estimated from the data instead of MC (see Sec. 6.5). But good consistency between data and MC is necessary for the signal efficiency estimation and MC shape extraction from histogram in the backgrounds fitting, which will be explained in details in Chapter 6.5.

In Figure 6.12 we can see that the $m_{e\mu}$ spectrum has a smooth shape and no obvious peak enhancement is produced after applying the BDTG method. Both statistical and systematical uncertainties included in error band.

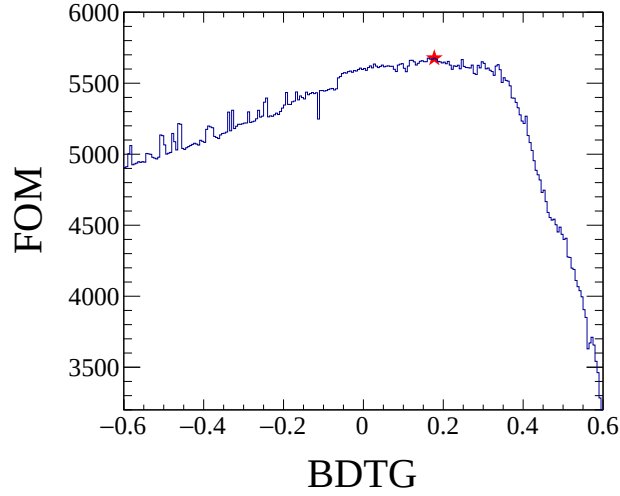


Figure 6.11: Variation of FOM versus the discriminant output from machine learning.

6.5 Determination of the signal excess by a fit to the $m_{e\mu}$ spectrum

The sensitivity is estimated using a data-driven method. We use a maximum likelihood fit to the $m_{e\mu}$ spectrum in the mass window $[70, 110]$ GeV and search for the possible signal excess. The lower limit of 70 GeV is chosen to avoid the turnover coming from the $Z \rightarrow \tau\tau$ background at the lower mass region. Currently the SR $[85, 95]$ GeV is fixed and blinded.

In RunI analysis, several fitting functions were employed. The resulting uncertainty caused by fitting function choice was the dominant systematic error. In this analysis, any simple analytical function won't work any more for the much larger dataset. Thus we analyze the background composition based on MC simulation and produce a combined probability density function (PDF). It manages to describe the SB well and reduces the resulting systematic error from fitting. In the following subsections we explain the construction of the PDF and provide the results from the data-driven method.

6.5.1 Description of the $Z \rightarrow \tau\tau$ background

The largest background after the event selection comes from the $Z \rightarrow \tau\tau \rightarrow e\mu\nu\nu\nu\nu$ process. To describe this background, we especially generate a large MC sample and study the events that

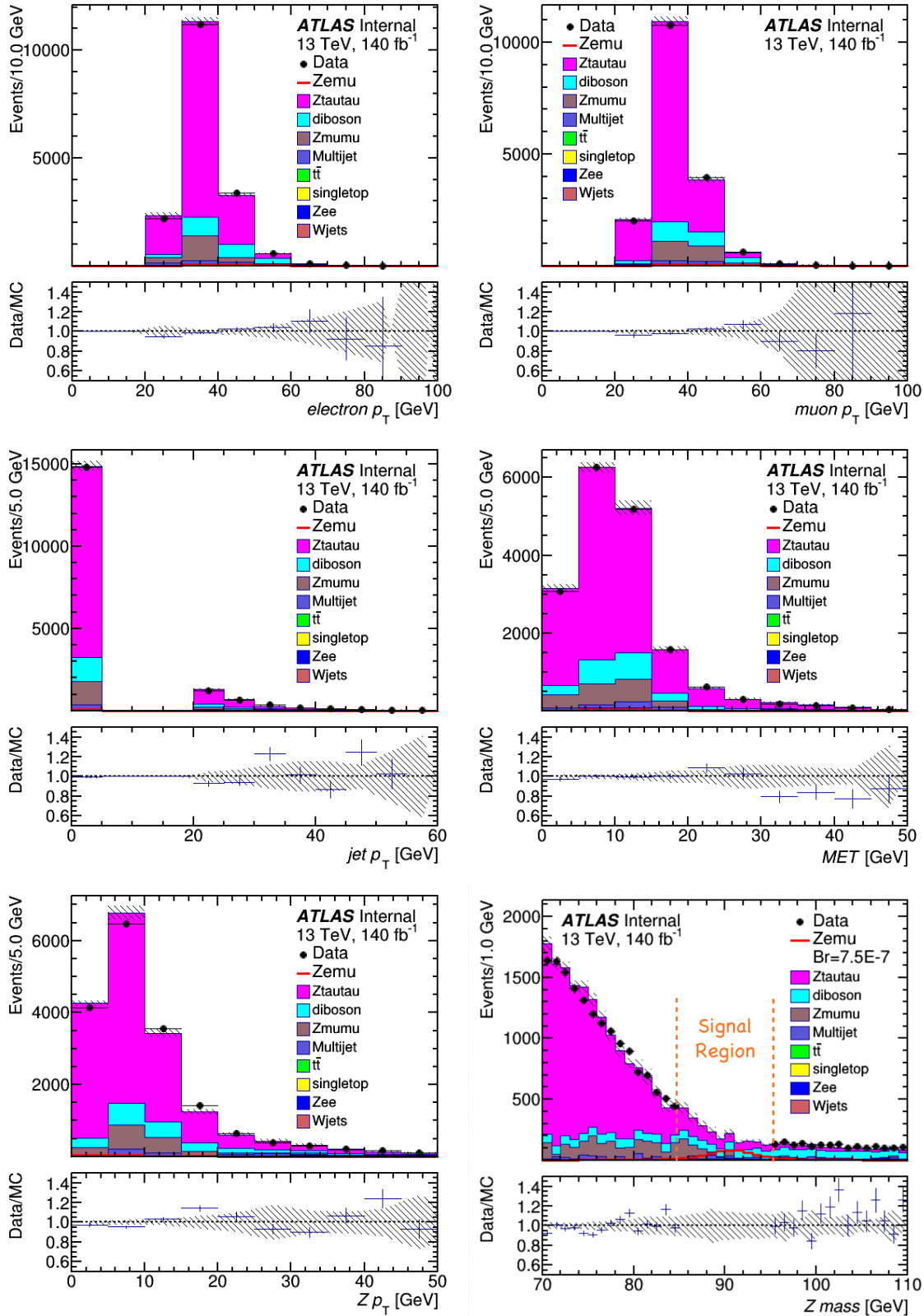


Figure 6.12: Data and MC comparison of kinematic variables in $Z \rightarrow e\mu$ after final selection. Top left: electron p_T ; Top right: muon p_T ; Middle left: the leading jet p_T ; Middle right: missing transverse momentum; Bottom left: p_T of reconstructed Z boson; Bottom right: invariant mass spectrum of reconstructed Z boson.

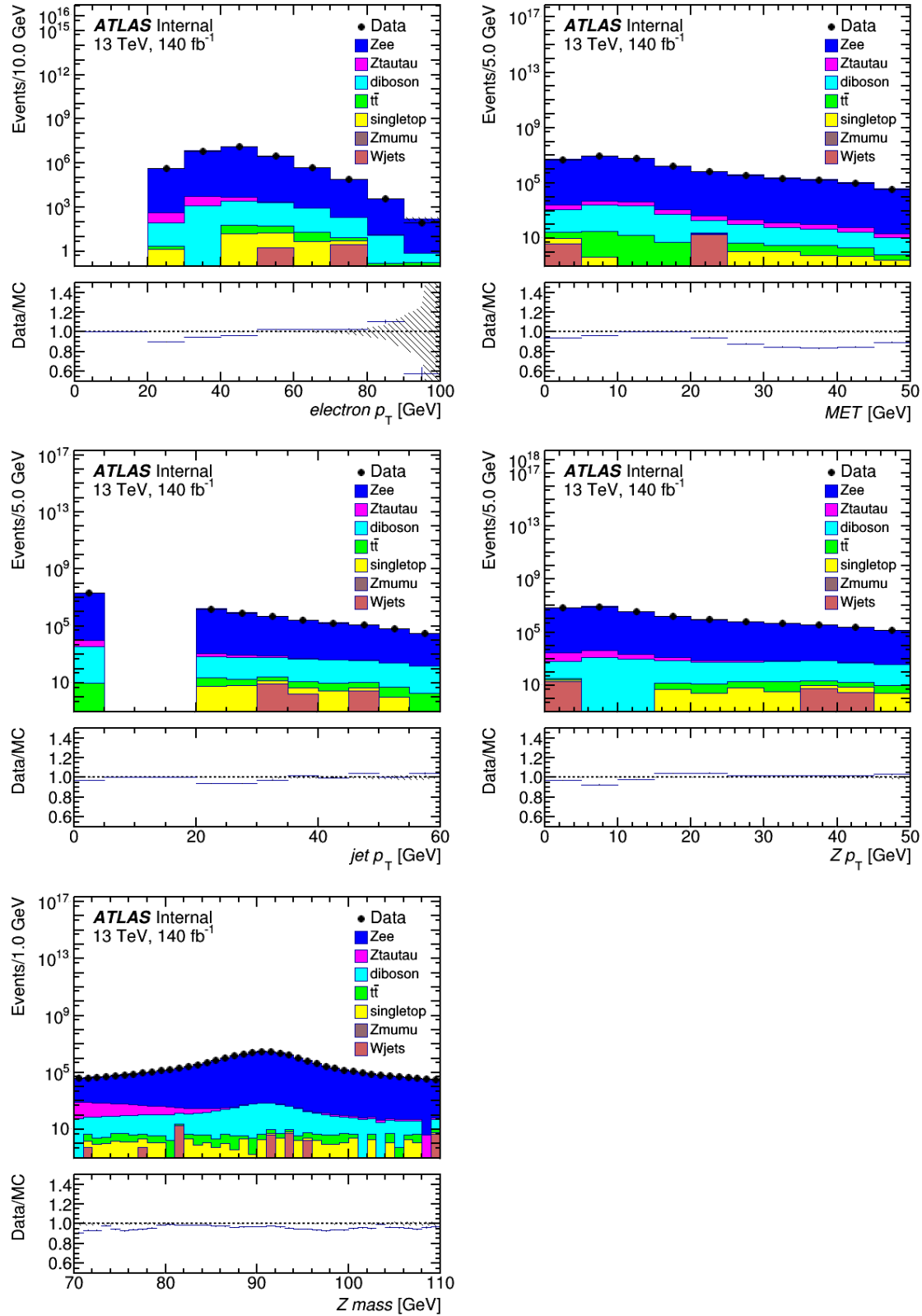


Figure 6.13: Data and MC comparison of kinematic variables in $Z \rightarrow ee$ after final selection(only statistical uncertainties included in error band). Top left: electron p_T , Top right: missing transverse momentum, Middle left: p_T of the leading jet, Middle right: p_T of the reconstructed Z boson, Bottom left: invariant mass spectrum of reconstructed Z boson.

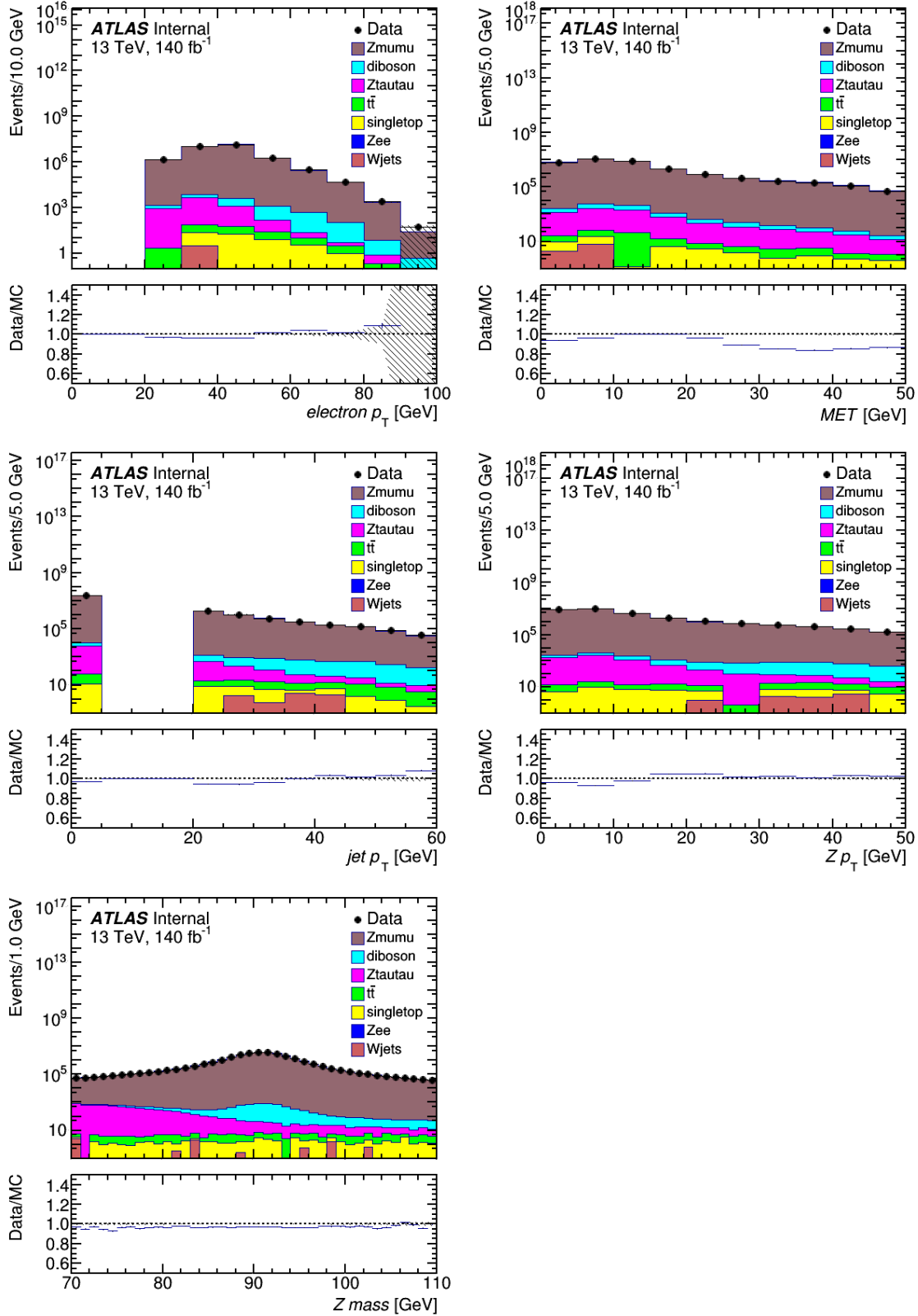


Figure 6.14: Data and MC comparison of kinematic variables in $Z \rightarrow \mu\mu$ after final selection(only statistical uncertainties included in error band). Top left: electron p_T , Top right: missing p_T , Middle left: transverse momentum of the leading jet, Middle right: p_T of the reconstructed Z boson, Bottom left: invariant mass spectrum of reconstructed Z boson.

pass the event selection. Its distribution of the $m_{e\mu}$ spectrum(100 bins) is shown in Figure 6.15, with the statistical uncertainty included. The four missing neutrinos in the final states make the invariant mass of the lepton pair accumulate below the Z pole and the tail extends into the signal window $[70, 110]$ GeV. The requirement of the transverse momentum of the leptons to be larger than 27 GeV creates a turnover around 70 GeV. Thus we choose 70 GeV as the lower limit of the fitting range.

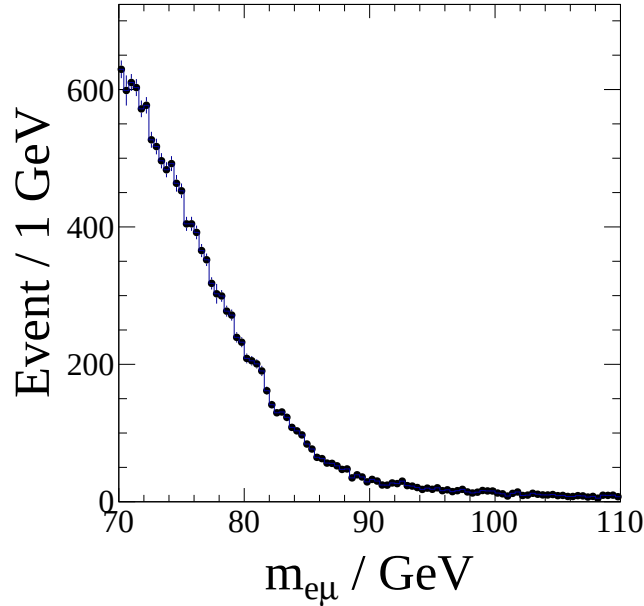


Figure 6.15: Description of $Z \rightarrow \tau\tau$ background.

6.5.2 Description of the $Z \rightarrow \mu\mu$ background

Another peaking background is from the $Z \rightarrow \mu\mu$ decay. The $m_{e\mu}$ spectrum of this MC forms a bump below the Z pole, as shown in Figure 6.16, with the statistical uncertainty included.

To understand how this background survives from the selection criteria, we investigate into the truth information of the MC events that pass the selection. It shows a strong correlation between the fake electron and the truth muon which is not the one successfully reconstructed as muon. Figure 6.17 shows the ΔR distribution between the reconstructed fake electron and the

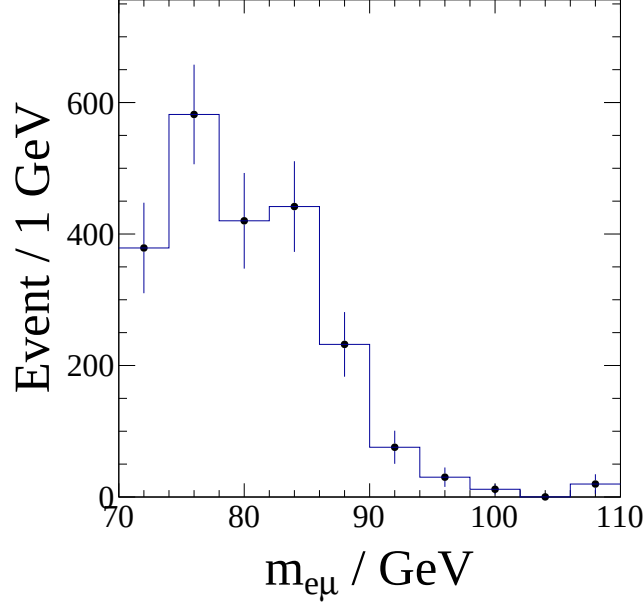


Figure 6.16: Description of $Z \rightarrow \mu\mu$ background.

truth track. Mostly they are close to each other. This favors the idea that one muon is faked as an electron.

Figure 6.18 illustrates several possible cases. In Figure 6.18a one muon particle decays into an electron plus two neutrinos before reaching the ECal, thus having both electron and muon in the final state like a signal event. We make a rough estimation of the number of such events using the below equation:

$$\begin{aligned}
 N &= 2N_{Z \rightarrow \mu\mu} \cdot \varepsilon_{Z \rightarrow \mu\mu} \cdot (1 - e^{-\frac{L}{c\tau}}) \\
 &\simeq 2N_Z \cdot Br(Z \rightarrow \mu\mu) \cdot \varepsilon_{Z \rightarrow \mu\mu} \cdot L/(ct) \\
 &\simeq 2N_Z \cdot Br(Z \rightarrow \mu\mu) \cdot \varepsilon_{Z \rightarrow \mu\mu} \cdot L/(c\tau) \cdot m_\mu/E_\mu
 \end{aligned} \tag{6.3}$$

where N_Z is the number of generated Z bosons in the 140 fb^{-1} of data (8.7×10^9), $Br(Z \rightarrow \mu\mu)$ the branching ratio (3.36%), $\varepsilon_{Z \rightarrow \mu\mu}$ the reconstruction efficiency (~ 0.1), L the radius of the ECal (2.0 m), $c\tau$ the intrinsic mean free path (660 m), E_μ the energy of the muon particle ($\sim 45 \text{ GeV}$)

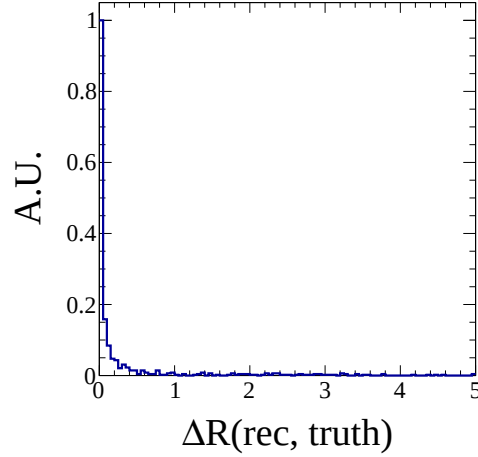


Figure 6.17: Angular distance between the reconstructed fake electron and the truth track of the closest muon.

and m_μ its mass (0.105 GeV). Put all these numbers in the equation we get about 415 $Z \rightarrow \mu\mu$ events which have one muon decaying into an electron. This makes more than 20% of the $Z \rightarrow \mu\mu$ backgrounds based on MC study.

Besides the case discussed above, one muon can also radiate an energetic photon before reaching the ECal or within it, as illustrated in Figure 6.18b. If the photon carries a significant part of the whole energy of the muon, the event will also represent similar signature as the signal.

Figure 6.18c shows the case that the muon which is not energetic deposits a large part of its energy in the ECal and not being captured by the muon spectrometer, so it is mis-identified as an electron.

In all the cases given in Figure 6.18, the fake electron will have smaller energy than the truth muon, resulting in $m_{e\mu}$ smaller than the Z boson mass. Because of lacking statistics in $Z \rightarrow \mu\mu$ MCs especially at low invariant mass region, the shape can not be easily described by any analytical function. But the contribution from $Z \rightarrow \mu\mu$ makes a significant part of about 10%. We use a MC shape to describe it in the fitting. The available truth information can not tell us the contribution from each case in Figure 6.18. But we suppose the MC simulation is reliable, and the MC shape is good enough to describe this process in the fitting.

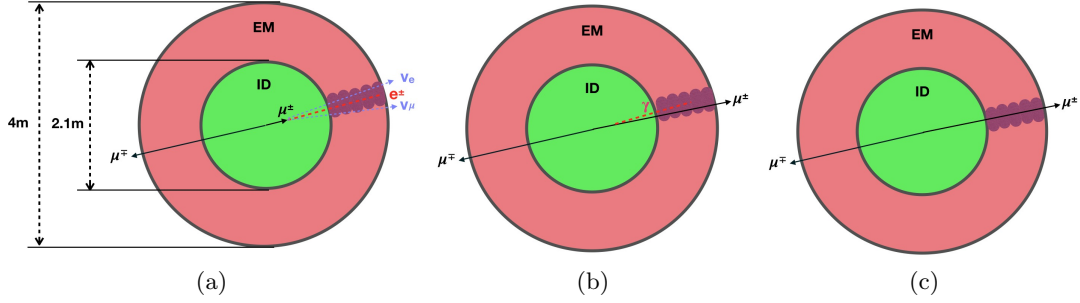


Figure 6.18: Illustration of several cases of the $Z \rightarrow \mu\mu$ events that can be mis-identified as $Z \rightarrow e\mu$ signals: (a) one muon decays to an electron, (b) one muon radiates one photon, and (c) one muon deposits much energy in ECal.

6.5.3 Description of the combinatorial background

The remaining backgrounds, mainly the $WW \rightarrow e\mu\nu\nu$ process, where the electron and muon particles don't originate from one Z boson, yield a linear shape in the $m_{e\mu}$ spectrum. Besides that diboson process, there also might exist QCD backgrounds like meson decay semi-leptonically, or the events can pass the selection criteria with quark or gluon jets faked as an electron. As we know, the lepton pair have no charge constraint, so we estimate them using the same-sign (the electron and muon have the same charge) data. There are also possibilities that other backgrounds can pass the selection as same-sign final state by charge flipping, such as $Z \rightarrow \mu\mu$ etc. These backgrounds are excluded from the same-sign data using MC simulation. The QCD background and the whole combinatorial background are shown in Figure 6.19, with the statistical uncertainty included. We have researched on multiple choices of the fitting functions, like 2nd polynomial and exponential functions, it turns out the best choice is linear function(see Eq.(6.5)), according to the goodness of fitting criteria $\chi^2/\text{n.d.f}$, which will be explained in details in next Chapter 6.5.4. Another advantage of linear function is it will only introduce 1 free parameter into the fitting, which reduces the risk of overfitting.

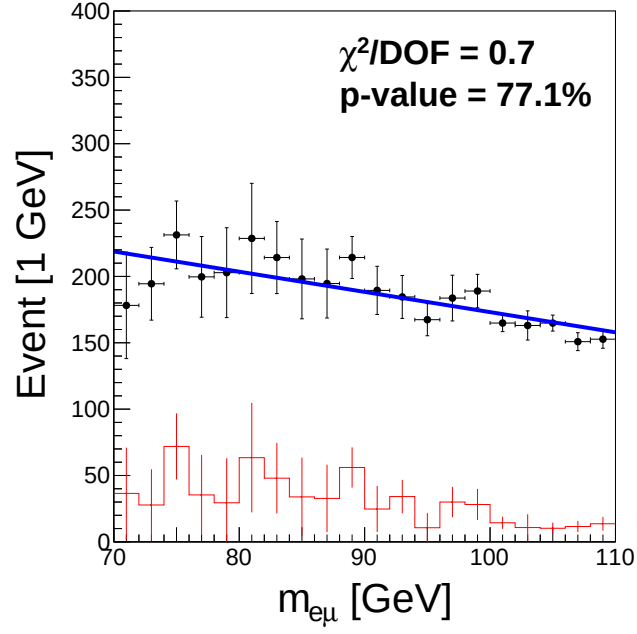


Figure 6.19: Description of the combinatorial background. The QCD background is estimated using same-sign data and simulation events, marked by the red solid line. Other backgrounds excluding the $Z \rightarrow \tau\tau$ and $Z \rightarrow \mu\mu$ are from MC simulation. The whole combinatorial background is fitted with a linear function.

6.5.4 Data-driven fitting method

We extract the background event number in the $m_{e\mu}$ spectrum by performing a maximum likelihood fit to the events in the SB that pass the selection. A combined PDF is used to describe the backgrounds. The likelihood of the fit is defined as

$$\mathcal{L} = \prod_{i=1}^N \left[(n_1 \cdot f_{\tau\tau} + n_2 \cdot f_{\mu\mu}) \otimes g(\mu, \sigma) + n_3 \cdot f_{\text{cmb}}(a) \right]. \quad (6.4)$$

Here N is the number of events passing the selection criteria, and $f_{\tau\tau}$ and $f_{\mu\mu}$ denote the PDFs of the $Z \rightarrow \tau\tau$ and $Z \rightarrow \mu\mu$ peaking backgrounds, respectively. They are derived from MC shapes and are convoluted with the same Gaussian function. The Gaussian resolution function is introduced to smear the $Z \rightarrow \tau\tau$ and $Z \rightarrow \mu\mu$ MC PDF shapes to take into account possible difference in the invariant mass resolution between data and MC. The mean of the Gaussian function μ is fixed at zero to ensure the stability of the fit while the width σ is floating. The PDF of the combinatorial background f_{cmb} is described by a linear function as validated by simulation, whose slope a is left float in the fit. The numbers of each background n_{1-3} are determined by the fit.

The goodness of the fit is estimated using $\chi^2/\text{n.d.f.}$, where n.d.f. denotes the number of degrees of freedom, equal to the number of bins N_{bin} minus the number of fit parameters minus 1. The χ^2 is calculated from the difference between the data and fit result using the following formula:

$$\chi^2 = \sum_i^{N_{\text{bin}}} \left(\frac{n_i^{\text{data}} - n_i^{\text{fit}}}{\sigma_i^{\text{fit}}} \right)^2, \quad (6.5)$$

where n_i^{data} denotes the measured content of the i th bin, and n_i^{fit} the value predicted by the fitted PDF.

The nominal fit to the SB is shown in Figure 6.20. We take $N_{\text{bin}} = 30$ bins when calculating the $\chi^2/\text{n.d.f.}$ and the derived p -value. The number of events in the signal region ($N_{\text{ext bkg}}$) is calculated by extrapolating the fitting function into the blinded region. The propagated error is taken as a systematic uncertainty of the predicted number of background events in the SR. The fit result is summarized in Table 6.3.

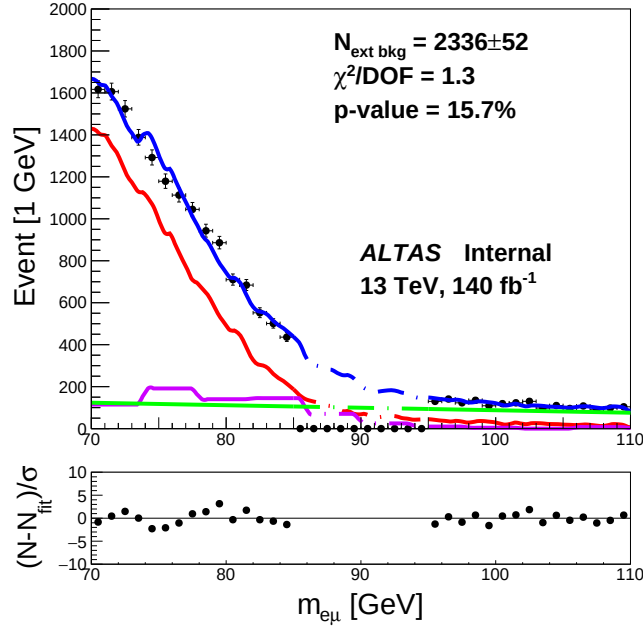


Figure 6.20: Fit to the $m_{e\mu}$ spectrum with the SR blinded. The blank dots with error bars represent the real data passing event selection. The red and purple lines describe the $Z \rightarrow \tau\tau$ and $Z \rightarrow \mu\mu$ components using MC shapes, while the green line shows the linear combinatorial background. The number of events in the SR extrapolated from fitting the SB is provided in the plot.

Table 6.3: Summary of the nominal fit result

Parameter	Description	Fit result
μ	mean of the Gaussian function convoluted to the MC shapes	0
σ	width of the Gaussian function convoluted to the MC shapes	$(5.64 \pm 0.02) \times 10^{-3}$
a	slope of the linear function	$(-5.8 \pm 1.3) \times 10^{-3}$
n_1	number of the $Z \rightarrow \tau\tau$ component	12744 ± 280
n_2	number of the $Z \rightarrow \mu\mu$ component	2813 ± 333
n_3	number of the combinatorial component	3987 ± 382

To verify this fitting method, a toy MC study is performed. We generate 100 k pseudo experiments by sampling from the real data in the SB region. In each bin of the pseudo datasets, the number of events fluctuates following the gaussian distribution whose mean and width are determined by the real data. In the signal region, we inject events sampled from MC distribution and scale it to the fixed number obtained by fitting real data (shown in Figure 6.20). Using the same function as when fitting the real data, we fit each pseudo dataset and calculate the relative difference between the extrapolated number of events and the counted one divided by the fitting error. The pull distribution of invariant mass spectrum is shown in Figure 6.21, close to a standard normal distribution. The N_{fit} is the estimation of background from sideband fitting for each pseudo experiment, N_{count} is the counted backgrounds from MCs, while $\sigma_{fitting}$ is the extrapolated error from fitting. The sigma of pull distribution is 1.02 indicating a fitting uncertainty, so we take this into account by scaling the fitting error with a factor of 1.02.

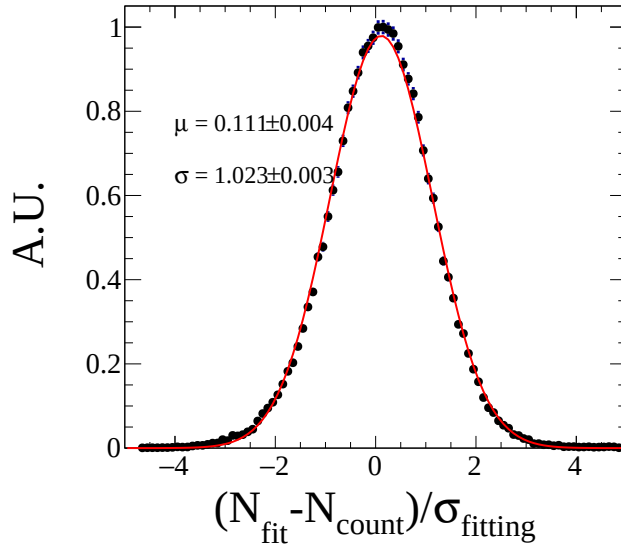


Figure 6.21: Pull distribution of 100 k pseudo experiments.

CHAPTER 7. SYSTEMATIC UNCERTAINTIES

The UL of $BrZ \rightarrow e\mu$ is calculated using Eq.(3.2-3.6). The systematic uncertainties included the uncertainties in the electron and muon reconstruction and trigger efficiencies, the absolute scale and resolution of the electron energy and muon p_T , the imperfect simulation of E_T^{miss} and $p_T^{\text{leading jet}}$. These uncertainties mostly cancel out in the ratio by using the $Z \rightarrow ee, \mu\mu$ reference channels to calculate the number of Z bosons, becoming negligible compared to the uncertainty coming from the overall fitting. The details are given in the following subsections.

7.1 Sideband fitting

The uncertainty induced by the fitting are considered in the calculation of $N_{95\%}^{\text{UL}}$. As discussed in Sec. 6.5, the discrepancy in mass resolution between data and MC is absorbed in the Gaussian function. The error of the number of events in the SR propagated from the SB fitting is 52. The pull distribution from a toyMC study indicates the underestimation of the fitting uncertainty. So we correct the fitting error with a scale factor of 1.1, resulting an total systematic uncertainty of 57.2.

7.2 Number of $Z \rightarrow ee, \mu\mu$ events

The number of the observed $Z \rightarrow ee, \mu\mu$ events are obtained by counting those falling in the SR minus the background from MC simulation. The background in the SR is less than 0.1%. The resulting systematic uncertainty is thus negligible.

7.3 Pileup Reweighting

The PileupReweightingTool of ATLAS provides an event pileup weight as a function of μ that makes the μ distribution of the MC match that of the data sample. The number of reconstructed

vertices is directly related to how much pileup there was in a given event. Studies comparing the number of vertices in data and MC for the same μ have shown a mis-modeling in MC that has to be corrected using a scale factor for μ between data and MC. The variation of the scale factors by the errors give variations for the pileup weight used to determine the systematic uncertainties associated with pileup.

7.4 Luminosity

The uncertainties in the integrated luminosity for different datasets are given in Table 5.1, varying from 2.1-2.4%. It is derived following a methodology detailed in Ref. [13] from a calibration of the luminosity scale using x-y beam-separation scans.

7.5 Electron

The lepton trigger systematics are provided by Combined Performance group(ATLAS). The trigger systematics for electron are returned by Combined Performance tool. It contains the systematic and statistical error on the trigger SF respectively. Reconstruction (Reco) and identification (ID) efficiency and its systematic uncertainties are provided by the Combined Performance group(ATLAS) through the AseElectronEfficiencyCorrectionTool [1]. The electron efficiency scale factors have been calculated using the full data of 2015 and 2016 [66] from J/ψ and Z measurements at low and high p_T respectively (using a tag and probe method). The ID efficiency scale factor includes the combined cuts of impact parameter significance and on $|z_0 \sin \theta|$ and is available from $p_T > 7$ GeV. The isolation efficiency scale factor is also provided. Three independent systematic sources (identification systematic of electron efficiency, reconstruction systematic of electron efficiency, isolation systematic of electron efficiency) are considered as electron efficiency systematic uncertainties.

Energy scale and resolution systematic uncertainties have been provided by the Combined Performance group, however these are not dominant systematic sources. This analysis is very

weakly sensitive to the energy scale and resolution systematics. All of the aforementioned electron systematics result in variations $< 0.1\%$ and are thus negligible compared to the other known sources.

7.6 Muon

For the muons, the trigger CP tool [19] returns two components : the systematics error (MUON_EFF_TrigSystUncertainty) and the statistical error (MUON_EFF_TrigStatUncertainty) on the trigger SF. Systematic uncertainties are obtained by a variation of $\pm 1\sigma$ of these errors.

Reconstruction, isolation and track-to-vertex association (d_0 significance and $|z_0 \sin \theta|$) scale factors are calculated using $Z \rightarrow \mu\mu$ and $J/\Psi \rightarrow \mu\mu$ (tag and probe) events in the full data of 2015, which corresponds to 3.2 fb^{-1} . Due to the fact that J/ψ measurement is valid below 15 GeV and Z measurement is more accurate above 15 GeV, separate systematic uncertainties are used in the low- p_T and high- p_T regions (below / above 15 GeV). This analysis relies only on those made from the Z measurement (MUON_EFF_STAT, MUON_EFF_SYS).

The isolation scale factor and its systematic uncertainties (MUON_ISO_STAT, MUON_ISO_SYS) are supported in the range of $10 < p_T < 500 \text{ GeV}$. The scale factor of the combined cuts on the d_0 significance and the $|z_0 \sin \theta|$ are also provided through (MUON_TTVA_STAT, MUON_TTVA_SYS).

All muon associated systematic uncertainties have an effect $< 0.01\%$ and are thus negligible compared to the other known sources.

7.7 Missing transverse energy E_T^{miss}

The missing transverse energy is calculated using physics objects as described in Sec. 4.6. Thus all of the systematic errors on the reconstructed components, e.g. the jet energy scale, result in an uncertainty on E_T^{miss} . These are the dominant sources of uncertainty on E_T^{miss} . In addition, the uncertainty arising from the "Soft term" accounting for the unassociated tracks is also considered. The resolution and scale of this soft term are varied within their errors to evaluate their contribution to the total uncertainty using METUtilities-00-02-46 [8].

7.8 Jet Energy Scale and Resolution Uncertainty

The jet energy scale and resolution of the small-R jets are measured in situ by calculating the response between MC and data in various bins of kinematic phase space using JetUncertainties-00-09-63 [7]. We use globally-reduced parameter configuration, which introduces 21 nuisance parameters in total (nuisance parameter is any parameter which is not of immediate interest but which must be accounted for in the analysis of those parameters which are of interest). They also enter the boosted analysis because they are used in the calculation of the missing transverse energy. We also consider the uncertainty on JVT efficiency, using JetJvtEfficiency-00-00-13 [3].

7.9 B-tagging systematics

The systematic uncertainties associated to the b-tagging are considered [2]. They are evaluated as uncertainties on the scaling factor to take account for possible disagreement of the b-tag efficiency between data and MC.

7.10 Other systematic uncertainties

Other systematic uncertainties include the running constant α_S (The coupling of the strong force, α_S , is deemed to be a fundamental parameter of Nature, and the running of the s coupling with Q, the energy-momentum-transfer scale, which in particle physics is the amount of momentum that one particle gives to another particle), the renormalization μ_R (the renormalization scale in Quantum Chromodynamics) [34] and the factorization μ_F factors (of which the parton distribution and fragmentation functions will become a function of, to address infrared (IR) divergence which appear because either a virtual or a real particle can reach a zero momentum, or because a massless particle radiates another massless particle) [53], the PDF uncertainty (the parton density function (PDF) of the proton is an essential component of Monte Carlo simulations and it can be determined using cross-section data through a number of different approaches [57], the difference between these PDFs can be treated as an additional systematic uncertainty), etc. They have impacts on the

calculations of the signal efficiencies for $Z \rightarrow e\mu, ee$ and $\mu\mu$, as well as the shape of the peaking background from $Z \rightarrow \tau\tau$. However, those effects on the analysis are found to be negligible.

CHAPTER 8. UPPER LIMIT CALCULATIONS

With the background estimate made, it is possible to determine the experimental sensitivity to our signal model. Limit setting is an important part of an experimental science, and is performed with HistFitter [30] which utilizes Roostats [73] in the analysis. The Roostats package provides high-level statistical tools for confidence interval estimation and hypothesis testing. It was used in this analysis to test the background+signal hypothesis to provide a limit on the branching ratio $Br(Z \rightarrow e\mu)$.

8.1 Upper Limit and CL_s method

In the absence of any significant data excess, the $m_{e\mu}$ invariant mass spectrum shown in Figure 8.1 is used to derive 95% CL upper limits on the branching ratio $Br(Z \rightarrow e\mu)$ using the CL_s method [63] [78].

The Poisson distribution is an appropriate model in circumstances where the occurrence of one event does not affect the probability of another (independence), the rate of occurrence is a constant, and the number of events is an integer. For an event counting experiment, these are all guaranteed and thus we can assume a Poisson probability density function (pdf) of Equation 8.1 and Equation 8.2.

L_{s+b} is a binned likelihood function (product of Poisson probabilities of each bin) to observe the data under the signal-plus-background (background-only) hypothesis, while L_b is a binned likelihood function (product of Poisson probabilities of each bin) to observe the data under the background-only hypothesis.

$$L_b = Poisson(N_{bkg}) = e^{-N_{bkg}} \frac{(N_{bkg})^k}{k!} \quad (8.1)$$

$$L_{s+b} = Poisson(N_{bkg} + \mu \cdot N_{sig}) = e^{-N_{bkg} + \mu \cdot N_{sig}} \frac{(N_{bkg} + \mu \cdot N_{sig})^k}{k!} \quad (8.2)$$

Using the two pdf, we can assume that in an experiment where we observe N_{obs} events, expect N_{bkg} background events and $\mu \cdot N_{sig}$ signal events (here μ is called signal strength), we can compute the quantities in Equation 8.3 and Equation 8.4 listed below. CL_b represents the probability to obtain a result less compatible with the signal than the observed one in the background-only hypothesis, while CL_{s+b} represents the probability to obtain a result less compatible with the signal than the observed result, assuming the signal hypothesis.

$$CL_b = \sum_{k=0}^{N_{obs}} L_b = \sum_{k=0}^{N_{obs}} e^{-N_{bkg}} \frac{(N_{bkg})^k}{k!} \quad (8.3)$$

$$CL_{s+b} = \sum_{k=0}^{N_{obs}} L_{s+b} = \sum_{k=0}^{N_{obs}} e^{-(N_{bkg} + \mu \cdot N_{sig})} \frac{(N_{bkg} + \mu \cdot N_{sig})^k}{k!} \quad (8.4)$$

Pseudo-experiments are generated for both hypotheses, taking into account per-bin statistical fluctuations of the total predictions according to Poisson statistics, as well as Gaussian fluctuations describing the effect of systematic uncertainties. Signal events for which $CL_s = CL_{s+b}/CL_b < 0.05$ are deemed to be excluded at 95% Confidence Level. Dividing by CL_b avoids the possibility of mistakenly excluding a signal we are not sensitive to because of downward fluctuation of the background. In this way, the signal strength μ is retrieved from the hypothesis test, and $\mu \cdot N_{sig}$ is the upper limit on the signal events.

8.2 Upper Limit on branching ratio

The Upper Limit(UL) of the number of signal events in the SR, $N_{95\%}^{UL}$, is calculated using the HistFitter tool [30] with the one-sided Profile Likelihood as test statistic. The UL extraction is conducted with the 10-bin signal and background fit in the signal region [85,95] GeV, with the signal MC and the prediction of background histogram from fitting sideband as the inputs. The Asymptotic calculators is used. A binned likelihood function is constructed from the product of

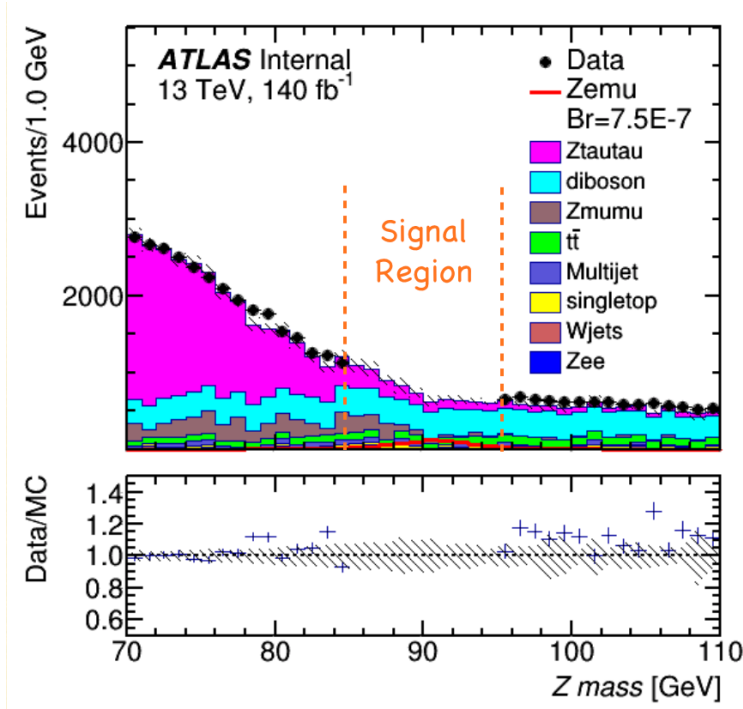


Figure 8.1: Invariant mass of dilepton $m_{e\mu}$ after the final-selection.

the Poisson probabilities of the observed and expected numbers of events in each mass bin. To get the expected UL, we generate pseudo data in the blinded region by prediction from fitting the SB. The median value of the pseudo-experiment distribution of the 95% CL UL is taken as the expected limit. The one- and two-standard deviation intervals of the expected limit are obtained by finding the 68% and 95% intervals of the pseudo-experiment upper limit distribution, respectively.

The UL result given by Asymptotic Calculator is shown in the Figure 8.2. The p -value equal to 0.05 is marked out by the red line. The expected UL of the signal events corresponding to CL_s Median, $\pm 1\sigma$ and $\pm 2\sigma$ are provided in Table 8.1.

Table 8.1: Summary of 95% UL result from Asymptotic Calculator for the Blinded data

Blinding Status	Method	UL(Events)
Blinded	Expected(Median)	102.34
	Expected(-1σ)	73.58
	Expected($+1\sigma$)	143.38
	Expected(-2σ)	55.14
	Expected($+2\sigma$)	194.43

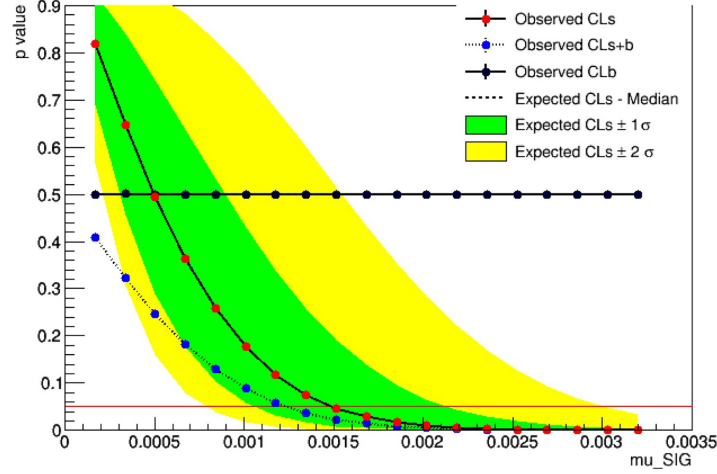


Figure 8.2: Upper Limit result of Asymptotic Calculator given by the HistFitter tool based on the data taken during the years from 2015-2018 with the SR blinded.

The formula of the UL calculation is given in Eq. 3.2. The measurements of efficiencies and the number of the observed Z bosons are provided in Table 8.2.

Table 8.2: The selection efficiencies after TMVA final selection for $Z \rightarrow ee, \mu\mu, e\mu$ decays are shown. Also provided are the produced number of Z bosons estimated from $Z \rightarrow ee, \mu\mu$, together with the weighted average.

Z decay	$140 \text{ fb}^{-1} @13 \text{ TeV}$	
	ε (%)	N_Z (10^9)
ee	6.71	8.44
$\mu\mu$	8.13	8.43
$\langle ee, \mu\mu \rangle$		8.43
$e\mu$	8.01	

With measurements in Table 8.1 and 8.2 as input, the 95% UL of $Br(Z \rightarrow e\mu)$ is calculated as below:

$$Br(Z \rightarrow e\mu) < \frac{N_{95\%}^{\text{UL}}}{\varepsilon_{Z \rightarrow e\mu} \cdot N_Z^{\text{avg}}} = \frac{102.34}{8.01\% \times 8.43 \times 10^9} = 1.52 \times 10^{-7}. \quad (8.5)$$

CHAPTER 9. SUMMARY

In the standard model (SM), lepton flavor violating (LFV) processes are strongly suppressed. Thus, searches for LFV processes are good candidates for probing new physics. In this note, we present a search for the LFV decay $Z \rightarrow e\mu$ using a 140 fb^{-1} of ATLAS Run 2 data. A much better expected upper limit of 1.52×10^{-7} at the 95% confidence level has been obtained on the branching fraction of the process $Z \rightarrow e\mu$, when compared with run 1 result 7.5×10^{-7} . The search is done by examining the invariant mass distribution ($m_{e\mu}$) of selected $e\mu$ pairs in the events and search for a possible enhancement in the signal region ($85 < m_{e\mu} < 95 \text{ GeV}$) around the Z boson pole mass. The signal extraction is performed with a combined signal and background fit in the signal region, where the total number of the background in the signal region is constrained to an estimated number from a fit to the $m_{e\mu}$ distribution of the events in the mass sidebands ($70 < m_{e\mu} < 85 \text{ GeV}$ or $95 < m_{e\mu} < 110 \text{ GeV}$). The final determination of the decay branching fraction of $Z \rightarrow e\mu$ is then computed using a normalization based on the observed number of $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ events, which significantly reduces many common experimental systematic uncertainties in the measurement.

Comparing to previous search for the $Z \rightarrow e\mu$ decay in Run 1, several analysis improvements have been implement besides a much larger data sample. The current analysis uses a b -jet veto during the preselection to reject a significant fraction of the background from the SM top production. A multivariable analysis based on the boosted decision tree algorithm is subsequently carried out to further suppress the background contributions. This approach yields a 7% gain of the expected search sensitivity comparing to a cut based analysis approach used in the previous Run 1 measurement. In addition, a more dedicated treatment is done to model individual background probability density function (PDF) in the fit, instead of the usage of a simple analytical function in the past. Both the dominant background from $Z \rightarrow \tau\tau$, where both τ leptons decay to e and μ ,

and the peaking background from the SM $Z \rightarrow \mu\mu$ events, where one of the muons is misidentified as an electron, are modeled using a histogram PDF from MC simulated events, respectively. The remaining combinatorial background distribution from all the other background sources (diboson, $t\bar{t}$, single top, W +jets, $Z \rightarrow ee$ and multi-jet) is described using a first order polynomial function in the fit. The improvement of the analysis technique results in a $\sim 40\%$ improvement of the expected search sensitivity on top of the improvement due to the increase of the data events comparing to the analysis in Run 1.

A complete analysis that includes final event selection, background estimation, signal extraction and systematic uncertainties has been carried out and described in detail in this note. The signal window is still blinded so only expected upper limit of the $Z \rightarrow e\mu$ branching fraction at the 95% confidence level is given.

There are a number of things that can be improved in the future for this analysis to achieve a better significance(FOM) or reduce the systematic uncertainties, which both improvements can result in a better upper limit result on the branching ratio. One is the fitting methodology can be further improved with more statistics of $Z \rightarrow \tau\tau$ and $Z \rightarrow \mu\mu$ backgrounds MC productions, which can help to achieve a more accurate fitting results on the backgrounds in Signal Region from the sideband data and reduce the extrapolated fitting systematic error. In addition, the Multivariate methods can be expanded to more complicated models like neural network and more variables which may potentially enhance the discriminating power, and conduct some transformations on those variables to make the machine learning models more effective, which overall would help to improve the significance of the analysis.

BIBLIOGRAPHY

- [1] Combined Performance Tools for Physics Analysis.
- [2] Flavour tagging recommendations in Release 21.
<https://twiki.cern.ch/twiki/bin/view/AtlasProtected/BTagCalib2015/>.
- [3] Jet Vertex Tagger.
<https://twiki.cern.ch/twiki/bin/view/AtlasProtected/JetVertexTagger/>.
- [4] (2014). ATLAS Run 1 Pythia8 tunes. Technical Report ATL-PHYS-PUB-2014-021, CERN, Geneva.
- [5] (2016). Simulation of top quark production for the ATLAS experiment at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2016-004, CERN, Geneva.
- [6] Aaboud, M. and Aad, G. (2016). Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV with the atlas detector at the lhc. *Phys. Rev. Lett.*, 117:182002.
- [7] Aaboud, M., Boyko, I. R., Bracinik, J., and Brandt, A. (2017a). Jet reconstruction and performance using particle flow with the atlas detector. *The European Physical Journal C*, 77(7):466.
- [8] Aaboud, M. and Bratzler, U. (2018). Performance of missing transverse momentum reconstruction with the atlas detector using proton–proton collisions. *The European Physical Journal C*, 78(11):903.
- [9] Aaboud, M. et al. (2016). Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC. *Eur. Phys. J.*, C76(12):653.

- [10] Aaboud, M. et al. (2017b). Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties. *JINST*, 12(05):P05002.
- [11] Aaboud, M. et al. (2018). Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV. *JHEP*, 08:089.
- [12] Aaboud, M. et al. (2019). Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV. *Submitted to: Eur. Phys. J.*
- [13] Aaboud, M. and etal (2016). Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC. *The European Physical Journal C*, 76(12):653.
- [14] Aad, G., Abbott, B., Abdallah, J., Abdel Khalek, S., Abdinov, O., Aben, R., Abi, B., Abolins, M., AbouZeid, O. S., Abramowicz, H., Abreu, H., Abreu, R., Abulaiti, Y., Acharya, B. S., Adamczyk, L., Adams, D. L., Adelman, J., Adomeit, S., Adye, T., Agatonovic-Jovin, T., Aguilar-Saavedra, J. A., Agostoni, M., Ahlen, S. P., Zhuang, X., Zhukov, K., Zibell, A., Zieminska, D., Zimine, N. I., Zimmermann, C., Zimmermann, R., Zimmermann, S., Zimmermann, S., Zinonos, Z., Ziolkowski, M., Zobernig, G., Zoccoli, A., zur Nedden, M., Zurzolo, G., Zutshi, V., and Zwalinski, L. (2014). Search for the lepton flavor violating decay $z \rightarrow e\mu$ in pp collisions at $\sqrt{s} = 8$ TeV with the atlas detector. *Phys. Rev. D*, 90:072010.
- [15] Aad, G. et al. (2010). The ATLAS Simulation Infrastructure. *Eur. Phys. J.*, C70:823–874.
- [16] Aad, G. et al. (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29.
- [17] Aad, G. et al. (2013). Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s} = 7$ TeV. *Eur. Phys. J.*, C73(3):2304.

- [18] Aad, G. et al. (2016a). Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s}=13$ TeV. *Eur. Phys. J.*, C76(5):292.
- [19] Aad, G. et al. (2016b). Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s}=13$ TeV. *Eur. Phys. J.*, C76(5):292.
- [20] Aad, G. et al. (2016c). Performance of b -Jet Identification in the ATLAS Experiment. *JINST*, 11(04):P04008.
- [21] Aaltonen, T., Adelman, J., Akimoto, T., Álvarez González, B., Amerio, S., Amidei, D., Anastassov, A., Annovi, A., Antos, J., Apollinari, G., Apresyan, A., Arisawa, T., Artikov, A., Ashmanskas, W., Attal, A., Aurisano, A., Azfar, F., Azzurri, P., Badgett, W., Barbaro-Galtieri, A., Barnes, V. E., Barnett, B. A., Bartsch, Wallny, R., Wang, S. M., Warburton, A., Waters, D., Weinberger, M., Weinelt, J., Wenzel, H., Wester, W. C., Whitehouse, B., Whiteson, D., Wicklund, A. B., Wicklund, E., Wilbur, S., Williams, G., Williams, H. H., Wilson, P., Winer, B. L., Wittich, P., Wolbers, S., Wolfe, C., Wright, T., Wu, X., Würthwein, F., Xie, S., Yagil, A., Yamamoto, K., Yamaoka, J., Yang, U. K., Yang, Y. C., Yao, W. M., Yeh, G. P., Yoh, J., Yorita, K., Yoshida, T., Yu, G. B., Yu, I., Yu, S. S., Yun, J. C., Zanello, L., Zanetti, A., Zhang, X., Zheng, Y., and Zucchelli, S. (2009). Search for the decays $B_{(s)}^0 \rightarrow e^+ \mu^-$ and $B_{(s)}^0 \rightarrow e^+ e^-$ in cdf run ii. *Phys. Rev. Lett.*, 102:201801.
- [22] Agostinelli, S., Allison, J., Amako, K., Apostolakis, J., Araujo, H., Arce, P., Asai, M., and Axen, D. (2003). Geant4 simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303.
- [23] Ahmed, M., Amann, J. F., Barlow, D., Black, K., Bolton, R. D., Brooks, M. L., Carius, S., Chen, Y. K., Chernyshev, A., Concannon, H. M., Cooper, M. D., Cooper, P. S., Crocker, J., Dittmann, J. R., Dziedzic, M., Empl, A., Fisk, R. J., Fleet, E., Foreman, W., Gagliardi, C. A.,

- Haim, D., Hallin, A., Hoffman, C. M., Hogan, G. E., Hughes, E. B., Hungerford, E. V., Jui, C. C. H., Kim, G. J., Knott, J. E., Koetke, D. D., Kozlowski, T., Kroupa, M. A., Kunselman, A. R., Lan, K. A., Laptev, V., Lee, D., Liu, F., Manweiler, R. W., Marshall, R., Mayes, B. W., Mischke, R. E., Nefkens, B. M. K., Nickerson, L. M., Nord, P. M., Othoudt, M. A., Otis, J. N., Phelps, R., Piilonen, L. E., Pillai, C., Pinsky, L., Ritter, M. W., Smith, C., Stanislaus, T. D. S., Stantz, K. M., Szymanski, J. J., Tang, L., Tippens, W. B., Tribble, R. E., Tu, X. L., Van Ausdeln, L. A., von Witch, W. H., Whitehouse, D., Wilkinson, C., Wright, B., Wright, S. C., Zhang, Y., and Ziock, K. O. H. (2002). Search for the lepton-family-number nonconserving decay $\mu^+ \rightarrow e^+ \gamma$. *Phys. Rev. D*, 65:112002.
- [24] Akers, R. et al. (1995). A Search for lepton flavor violating Z^0 decays. *Z. Phys.*, C67:555–564.
- [25] Artoisenet, P., Frederix, R., Mattelaer, O., and Rietkerk, R. (2013). Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations. *JHEP*, 03:015.
- [26] ATLAS Collaboration (2015a). Expected performance of the ATLAS b -tagging algorithms in Run-2. ATL-PHYS-PUB-2015-022.
- [27] ATLAS Collaboration (2015b). Performance of missing transverse momentum reconstruction with the ATLAS detector in the first proton–proton collisions at $\sqrt{s} = 13$ TeV. ATL-PHYS-PUB-2015-027.
- [28] Aubert, B., Barate, R., Boutigny, D., Couderc, F., Gaillard, J.-M., Hicheur, A., Karyotakis, Y., Lees, J. P., Tisserand, V., Zghiche, A., Palano, A., Pompili, A., Chen, J. C., Qi, N. D., Rong, G., Wang, P., Zhu, Y. S., Eigen, G., Ofte, I., Stugu, B., Abrams, G. S., Borgland, A. W., Breon, A. B., Brown, D. N., Button-Shafer, J., Cahn, R. N., Charles, E., Day, C. T., Gill, M. S., Gritsan, A. V., Groysman, Y., Jacobsen, R. G., Kadel, R. W., Kadyk, J., Kerth, L. T., Kolomensky, Y. G., Kukartsev, G., Lynch, G., Mir, L. M., Oddone, P. J., Orimoto, T. J., Vuagnin, G., Panvini, R. S., Banerjee, S., Brown, C. M., Fortin, D., Jackson, P. D., Kowalewski, R., Roney, J. M., Sobie,

- R. J., Band, H. R., Cheng, B., Dasu, S., Datta, M., Eichenbaum, A. M., Graham, M., Hollar, J. J., Johnson, J. R., Kutter, P. E., Li, H., Liu, R., Mihalyi, A., Mohapatra, A. K., Pan, Y., Prepost, R., Tan, P., von Wimmersperg-Toeller, J. H., Wu, J., Wu, S. L., Yu, Z., Greene, M. G., and Neal, H. (2004). Search for flavor-changing neutral current and lepton-flavor violating decays of $D^0 \rightarrow l^+ l^-$. *Phys. Rev. Lett.*, 93:191801.
- [29] Aubert, B., Barate, R., Boutigny, D., Couderc, F., Karyotakis, Y., Lees, J. P., Poireau, V., Tisserand, V., Zghiche, A., Grauges, E., Palano, A., Pappagallo, M., Pompili, A., Chen, J. C., Qi, N. D., Rong, G., Wang, P., Zhu, Y. S., Eigen, G., Ofte, I., Stugu, B., Abrams, G. S., Battaglia, M., Breon, A. B., Brown, D. N., Kowalewski, R., Roney, J. M., Sobie, R. J., Back, J. J., Harrison, P. F., Latham, T. E., Mohanty, G. B., Band, H. R., Chen, X., Cheng, B., Dasu, S., Datta, M., Eichenbaum, A. M., Flood, K. T., Graham, M., Hollar, J. J., Johnson, J. R., Kutter, P. E., Li, H., Liu, R., Mellado, B., Mihalyi, A., Pan, Y., Pierini, M., Prepost, R., Tan, P., Wu, S. L., Yu, Z., and Neal, H. (2006). Search for lepton flavor violation in the decay $\tau^\pm \rightarrow e^\pm \gamma$. *Phys. Rev. Lett.*, 96:041801.
- [30] Baak, M., Besjes, G. J., Côté, D., Koutsman, A., Lorenz, J., and Short, D. (2015). HistFitter software framework for statistical data analysis. *The European Physical Journal C*, 75(4):153.
- [31] Ball, R. D. et al. (2013). Parton distributions with LHC data. *Nucl. Phys. B*, 867:244.
- [32] Bellgardt, U., Otter, G., Eichler, R., Felawka, L., Niebuhr, C., Walter, H., Bertl, W., Lordong, N., Martino, J., Egli, S., Engfer, R., Grab, C., Grossmann-Handschin, M., Hermes, E., Kraus, N., Muheim, F., Pruys, H., Schaaf, A. V. D., and Vermeulen, D. (1988). Search for the decay $e^+e^+e^-$. *Nuclear Physics B*, 299(1):1 – 6.
- [33] Bernstein, R. and Cooper, P. S. (2013). Charged lepton flavor violation: An experimenters guide. *Physics Reports*, 532(2):27 – 64. Charged Lepton Flavor Violation: An Experimenter’s Guide.

- [34] Brodsky, S. J. and Di Giustino, L. (2012). Setting the Renormalization Scale in QCD: The Principle of Maximum Conformality. *Phys. Rev.*, D86:085026.
- [35] Buttinger, W. (2015). Using Event Weights to account for differences in Instantaneous Luminosity and Trigger Prescale in Monte Carlo and Data. Technical Report ATL-COM-SOFT-2015-119, CERN, Geneva.
- [36] Cacciari, M., Salam, G. P., and Soyez, G. (2008). The anti-kt jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063.
- [37] Cascioli, Fabio and Maierhofer, Philipp and Pozzorini, Stefano (2012). Scattering Amplitudes with Open Loops. *Phys. Rev. Lett.*, 108:111601.
- [38] Chatrchyan, S. et al. (2012). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61.
- [39] Collaboration, A. (2016). Muon reconstruction performance of the atlas detector in proton–proton collision data. *The European Physical Journal C*, 76(5):292.
- [40] Collaboration, T. A. and .etc, G. A. (2008). The atlas experiment at the cern large hadron collider. *Journal of Instrumentation*, 3(08):S08003.
- [41] D. J. Lange (2001). The EvtGen particle decay simulation package. *Nucl. Instrum. Meth. A*, 462:152.
- [42] D. P. Group. Good Runs List for Analysis Run2 at ATLAS.
- [43] DataPrep. Data Preparation Check List For Physics Analysis in Run2 at ATLAS.

- [44] Davidson, S., Lacroix, S., and Verdier, P. (2012). Lhc sensitivity to lepton flavour violating z boson decays. *Journal of High Energy Physics*, 2012(9):92.
- [45] DELPHI Collaboration (1997). Search for lepton flavour number violating z^0 - decays. *Zeitschrift für Physik C Particles and Fields*, 73(2):243–251.
- [46] Ding, G.-J. and Yan, M.-L. (2008). Lepton flavor violating $\mu \rightarrow e\gamma$ and $\mu - e$ conversion in unparticle physics. *Phys. Rev. D*, 77:014005.
- [47] Dinh, D. N., Ibarra, A., Molinaro, E., and Petcov, S. T. (2012). The $\mu - e$ Conversion in Nuclei, $\mu \rightarrow e\gamma$, $\mu \rightarrow 3e$ Decays and TeV Scale See-Saw Scenarios of Neutrino Mass Generation. *JHEP*, 08:125. [Erratum: JHEP09,023(2013)].
- [48] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Ann. Statist.*, 29(5):1189–1232.
- [49] Gabbiani, F., Kim, J. H., and Masiero, A. (1988). SUSY: Are they observable? *Physics Letters B*, 214(3):398 – 402.
- [50] Glashow, S. L. (2013). Particle Physics in The United States, A Personal View. In *Proceedings, 2013 Community Summer Study on the Future of U.S. Particle Physics: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, July 29-August 6, 2013*.
- [51] Gleisberg, T. and Höche, Stefan. and Krauss, F. and Schönherr, M. and Schumann, S. and others (2009). Event generation with SHERPA 1.1. *JHEP*, 02:007.
- [52] Gleisberg, Tanju and Höche, Stefan (2008). Comix, a new matrix element generator. *JHEP*, 12:039.

- [53] Gomez, C. and Panchenko, M. (2016). Asymptotic dynamics, large gauge transformations and infrared symmetries.
- [54] Gripaio, B. (2010). Composite Leptoquarks at the LHC. *JHEP*, 02:045.
- [55] Grosnick, D. et al. (1986). Search for the Rare Decay $\mu^+ \rightarrow e^+ \gamma \gamma$. *Phys. Rev. Lett.*, 57:3241.
- [56] Group, A. T. (2017). Lowest un-prescaled triggers per data-taking period.
- [57] Harland-Lang, L. A., Martin, A. D., Motylinski, P., and Thorne, R. S. (2015). Parton distributions in the lhc era: Mmht 2014 pdfs. *The European Physical Journal C*, 75(5):204.
- [58] Higgs, P. W. (1964). Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509.
- [59] Hirsch, M., Vicente, A., Meyer, J., and Porod, W. (2009). Publisher’s note: Majoron emission in muon and tau decays revisited [phys. rev. d 79, 055023 (2009)]. *Phys. Rev. D*, 79:079901.
- [60] Höche, Stefan and Krauss, Frank and Schönherr, Marek and Siegert, Frank (2013). QCD matrix elements + parton showers: The NLO case. *JHEP*, 04:027.
- [61] Hoecker, A., Speckmayer, P., Stelzer, J., Therhaag, J., von Toerne, E., and Voss, H. (2007). TMVA: Toolkit for Multivariate Data Analysis. *PoS, ACAT*:040.
- [62] Illana, J. I. and Riemann, T. (2001). Charged lepton flavor violation from massive neutrinos in z decays. *Phys. Rev. D*, 63:053004.
- [63] Junk, T. (1999). Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth.*, A434:435–443.

- [64] Kuo, T. K. and Nakagawa, N. (1985). Lepton-flavor-violating decays of Z^0 and τ . *Phys. Rev. D*, 32:306–307.
- [65] Lampl, W., Laplace, S., Lelas, D., Loch, P., Ma, H., Menke, S., Rajagopalan, S., Rousseau, D., Snyder, S., and Unal, G. (2008). Calorimeter Clustering Algorithms: Description and Performance. Technical Report ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva.
- [66] Lohwasser, K. (2017). LatestRecommendationsElectronIDRun2.
- [67] Mangano, M. L. et al. (2003). ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP*, 07:001.
- [68] Martin, A. D., Stirling, W. J., Thorne, R. S., and Watt, G. (2009). Parton distributions for the LHC. *Eur. Phys. J.*, C63:189–285.
- [69] MEG Collaboration (2016). Search for the lepton flavour violating decay $\mu^+ \rightarrow e^+ \gamma$ with the full dataset of the MEG experiment. *Eur. Phys. J.*, C76(8):434.
- [70] Mihara, S., Miller, J., Paradisi, P., and Piredda, G. (2013a). Charged lepton flavorviolation experiments. *Annual Review of Nuclear and Particle Science*, 63(1):531–552.
- [71] Mihara, S., Miller, J., Paradisi, P., and Piredda, G. (2013b). Charged lepton flavorviolation experiments. *Annual Review of Nuclear and Particle Science*, 63(1):531–552.
- [72] Miyazaki, Y., Adachi, I., Aihara, H., Arinstein, K., Aulchenko, V., Aushev, T., Bakich, A., Balagura, V., Barberio, E., Bay, A., Bitenc, U., Bondar, A., Bozek, A., Brako, M., Browder, T., Chen, A., Chen, K.-F., Chen, W., Cheon, B., Chistov, R., Cho, I.-S., Choi, Y., Dash, M., Drutskoy, A., Eidelman, S., Ueno, D. K., Unno, Y., Uno, S., Urquijo, P., Usov, Y., Varner, G., Villa, S., Vinokurova, A., Wang, C., Wang, P., Wang, X., Watanabe, Y., Won, E., Yamashita,

- Y., Zhang, Z., Zhilich, V., Zupanc, A., and Zyukova, O. (2008). Search for lepton flavor violating decays into three leptons. *Physics Letters B*, 660(3):154 – 160.
- [73] Moneta, L., Cranmer, K., Schott, G., and Verkerke, W. (2010). The RooStats project. In *Proceedings of the 13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research. February 22-27*, page 57.
- [74] Morgenstern, M. and Owen, M. (2017). Muon trigger efficiency in early 2018 data taking. Technical Report ATL-COM-DAQ-2018-047, CERN, Geneva.
- [75] Nix, J., Ahn, J. K., Akune, Y., Baranov, V., Doroshenko, M., Fujioka, Y., Hsiung, Y. B., Inagaki, T., Ishibashi, S., Ishihara, N., Ishii, H., Iwai, E., Iwata, T., Kobayashi, S., Komatsubara, T. K., Kurilin, A. S., Kuzmin, E., Lednev, A., Lee, H. S., Lee, S. Y., Lim, G. Y., Matsumura, T., Moiseenko, A., Morii, H., Morimoto, T., Nakano, T., Nomura, T., Nomachi, M., Okuno, H., Omata, K., Perdue, G. N., Perov, S., Podolsky, S., Porokhovoy, S., Sakashita, K., Sasao, N., Sato, H., Sato, T., Sekimoto, M., Shinkawa, T., Sugaya, Y., Sugiyama, A., Sumida, T., Tajima, Y., Tsamalaidze, Z., Tsukamoto, T., Wah, Y., Watanabe, H., Yamaga, M., Yamanaka, T., Yoshida, H. Y., and Yoshimura, Y. (2007). First search for $K_L \rightarrow \pi^0 \pi^0 \nu \bar{\nu}$. *Phys. Rev. D*, 76:011101.
- [76] Nussinov, S., Peccei, R. D., and Zhang, X. M. (2000). Constraints on two-body lepton flavor violating decay processes. *Phys. Rev. D*, 63:016003.
- [77] Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- [78] Read, A. L. (2002). Presentation of search results: The CL(s) technique. *J. Phys.*, G28:2693–2704. [11(2002)].
- [79] S. Farrell (2018). AssociationUtils OverlapRemovalTool in Run2 at ATLAS.

- [80] Schumann, S. and Krauss, F. (2008). A Parton shower algorithm based on Catani-Seymour dipole factorisation. *JHEP*, 03:038.
- [81] Sjostrand, T. et al. (2015). An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159.
- [82] Skands, P. Z. (2010). Tuning Monte Carlo Generators: The Perugia Tunes. *Phys. Rev. D*, 82:074018.
- [83] Stockton, M. and the ATLAS collaboration (2011). The atlas level-1 central trigger. *Journal of Physics: Conference Series*, 331(2):022041.
- [84] Therhaag, J. (2010). TMVA Toolkit for multivariate data analysis in ROOT. *PoS, ICHEP2010*:510.
- [85] Weinberg, S. (1967). A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266.
- [86] Yang, J. (2010). Lepton flavor violating Z-boson decays at GigaZ as a probe of supersymmetry. *Science China Physics, Mechanics and Astronomy*, 53(11):1949–1952.
- [87] zur Nedden, M. (2017). The lhc run 2 atlas trigger system: design, performance and plans. *Journal of Instrumentation*, 12(03):C03024.

APPENDIX A. THE LIST OF TRIGGERS USED IN EACH DATA-TAKING PERIOD

Table A.1: The list of triggers used in each data-taking period.

Data period	Electron trigger	Muon Trigger
2015	HLT_e24_lhmedium_L1EM20VH OR HLT_e60_lhmediumm OR HLT_e120_lhloose	HLT_mu20_loose_L1MU15 OR HLT_mu40
2016 A	HLT_e24_lhvloose_L1EM20VH OR HLT_e60_lhmedium_nod0 OR HLT_e60_medium OR HLT_e140_lhloose_nod0 OR HLT_e300_etcut	HLT_mu24_ivarloose OR HLT_mu40 OR HLT_mu50
B-D3	HLT_e24_lhvloose_L1EM20VH OR HLT_e60_lhmedium_nod0 OR HLT_e60_medium OR HLT_e140_lhloose_nod0 OR HLT_e300_etcut	HLT_mu24_ivarmedium OR HLT_mu26_ivarmedium OR HLT_mu50
D4-F	HLT_e26_lhtight_nod0_ivarloose OR HLT_e60_lhmedium_nod0 OR HLT_e60_medium OR HLT_e140_lhloose_nod0 OR HLT_e300_etcut	HLT_mu26_ivarmedium OR HLT_mu50
G-onward	HLT_e26_lhtight_nod0_ivarloose OR HLT_e60_lhmedium_nod0 OR HLT_e60_medium OR HLT_e140_lhloose_nod0 OR HLT_e300_etcut	HLT_mu26_ivarmedium OR HLT_mu50
2017 B-onward	HLT_e26_lhtight_nod0_ivarloose OR HLT_e60_lhmedium_nod0 OR HLT_e140_lhloose_nod0 OR HLT_e300_etcut	HLT_mu26_ivarmedium OR HLT_mu50
2018 B-onward	HLT_e26_lhtight_nod0_ivarloose OR HLT_e60_lhmedium_nod0 OR HLT_e140_lhloose_nod0 OR HLT_e300_etcut	HLT_mu26_ivarmedium OR HLT_mu50

APPENDIX B. SENSITIVITY USING CUT-BASED METHOD

We also follow the traditional method employed by RunI analysis, i.e., to scan the sensitivity in the two-dimensional space $(E_T^{\text{miss}}, p_T^{\text{leading jet}})$, as shown in Figure B.1. The optimal cut is marked out. The resulting best sensitivity is 7% worse than the one obtained from the multivariate technique.

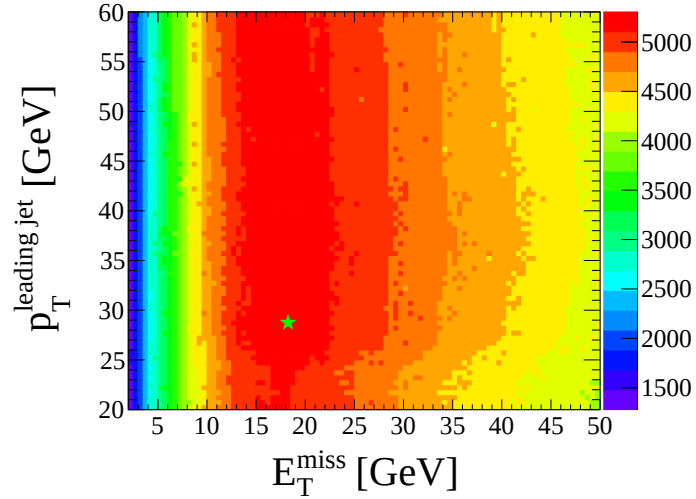
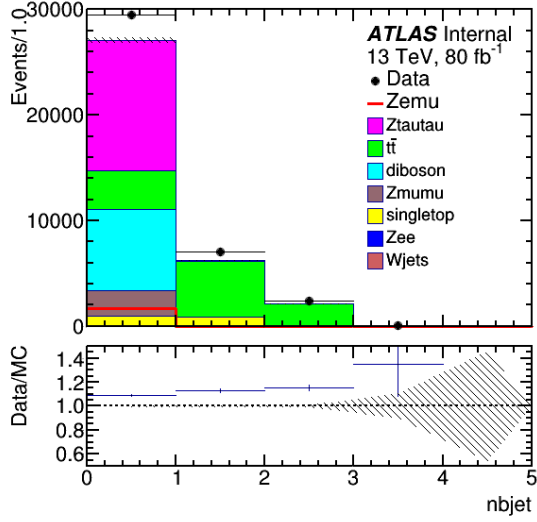


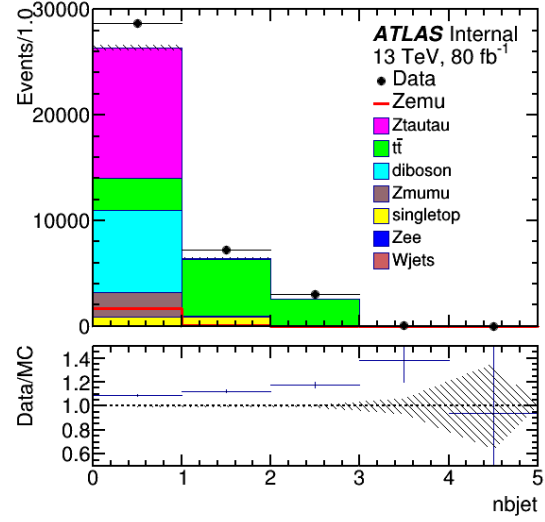
Figure B.1: Variation of FOM versus the discriminant output from TMVA.

APPENDIX C. STUDY ON VETOING b -jets

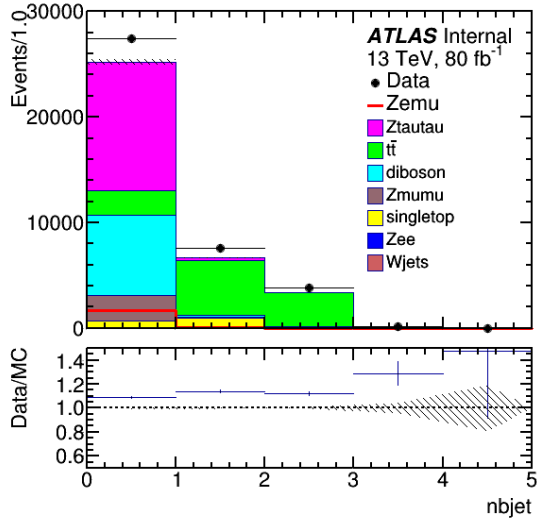
We study the distribution of the number of b -jets using different b -tagging working points, as shown in Figure C.1. We take the working point with 85% b -tagging efficiency trying to remove most top backgrounds.



(a) 70% WP



(b) 77% WP



(c) 85% WP

Figure C.1: bjets distribution with 70/77/85 working points after pre-selection: (a) 70% working point, (b) 77% working point, (c) 85% working point.

APPENDIX D. MC SAMPLES

Table D.1: Summary of Signal and SM background MC Samples used in the analysis.

MCID	Name	# of events	Cross-section [pb]	k-factor	ϵ_{filter}
304998	Pythia8EvtGen_A14NNPDF23LO_Zemu	100000.0	2864	1.0	0.51265
364100	Sh_221_Zmumu_M0_70_CvBv	7970000.0	1983.0	0.9751	8.2210E-01
364101	Sh_221_Zmumu_M0_70_CfBv	4982000.0	1978.4	0.9751	1.1308E-01
364102	Sh_221_Zmumu_M0_70_Bf	7983000.0	1982.2	0.9751	6.4161E-02
364103	Sh_221_Zmumu_M70_140_CvBv	5983000.0	108.92	0.9751	6.8873E-01
364104	Sh_221_Zmumu_M70_140_CfBv	1996800.0	109.42	0.9751	1.8596E-01
364105	Sh_221_Zmumu_M70_140_Bf	5981600.0	108.91	0.9751	1.1375E-01
364106	Sh_221_Zmumu_M140_280_CvBv	5000000.0	39.878	0.9751	6.0899E-01
364107	Sh_221_Zmumu_M140_280_CfBv	3000000.0	39.795	0.9751	2.3308E-01
364108	Sh_221_Zmumu_M140_280_Bf	12458900.0	39.908	0.9751	1.4618E-01
364109	Sh_221_Zmumu_M280_500_CvBv	2000000.0	8.5375	0.9751	5.5906E-01
364110	Sh_221_Zmumu_M280_500_CfBv	999600.0	8.5403	0.9751	2.6528E-01
364111	Sh_221_Zmumu_M280_500_Bf	1999400.0	8.4932	0.9751	1.7559E-01
364112	Sh_221_Zmumu_M500_1000	2996500.0	1.7881	0.9751	1.0000E+00
364113	Sh_221_Zmumu_M1000_E_CMS	997000.0	0.14769	0.9751	1.0000E+00
364114	Sh_221_Zee_M0_70_CvBv	8000000.0	1981.8	0.9751	8.2106E-01
364115	Sh_221_Zee_M0_70_CfBv	4999000.0	1980.8	0.9751	1.1295E-01
364116	Sh_221_Zee_M0_70_Bf	7995600.0	1981.7	0.9751	6.3809E-02
364117	Sh_221_Zee_M70_140_CvBv	5956000.0	110.5	0.9751	6.9043E-01
364118	Sh_221_Zee_M70_140_CfBv	1999200.0	110.63	0.9751	1.8382E-01
364119	Sh_221_Zee_M70_140_Bf	5340000.0	110.31	0.9751	1.1443E-01
364120	Sh_221_Zee_M140_280_CvBv	5000000.0	40.731	0.9751	6.1452E-01
364121	Sh_221_Zee_M140_280_CfBv	3000000.0	40.67	0.9751	2.3044E-01
364122	Sh_221_Zee_M140_280_Bf	12379600.0	40.694	0.9751	1.4927E-01
364123	Sh_221_Zee_M280_500_CvBv	1999800.0	8.6743	0.9751	5.6134E-01
364124	Sh_221_Zee_M280_500_CfBv	999900.0	8.6711	0.9751	2.6294E-01
364125	Sh_221_Zee_M280_500_Bf	1999850.0	8.6766	0.9751	1.7223E-01
364126	Sh_221_Zee_M500_1000	2999000.0	1.8081	0.9751	1.0000E+00
364127	Sh_221_Zee_M1000_E_CMS	1000000.0	0.14857	0.9751	1.0000E+00

Table D.2: Summary of Signal and SM background MC Samples used in the analysis continued.

MCID	Name	# of events	Cross-section [pb]	k-factor	ϵ_{filter}
344772	Sh_221_Ztautau_M0_70_l13l7	10491000.0	246.053	0.9751	0.203784
344776	Sh_221_Ztautau_M70_140_l13l7	1999500.0	13.7241	0.9751	0.035526
344780	Sh_221_Ztautau_M140_280_l13l7	1000000.0	5.04426	0.9751	0.040383
364137	Sh_221_Ztautau_M280_500_CvBv	2000000.0	8.5502	0.9751	5.6036E-01
364138	Sh_221_Ztautau_M280_500_CfBv	1000000.0	8.6707	0.9751	2.6245E-01
364139	Sh_221_Ztautau_M280_500_Bf	1983950.0	8.6804	0.9751	1.7313E-01
364140	Sh_221_Ztautau_M500_1000	2999800.0	1.8096	0.9751	1.0000E+00
364141	Sh_221_Ztautau_M1000_E.CMS	1000000.0	0.14834	0.9751	1.0000E+00
363356	Sh_221_ZqqZll	5400000.0	15.563	1.00	1.3961E-01
363358	Sh_221_WqqZll	5400000.0	3.437	1.00	1.0000E+00
364250	Sh_222_llll	17762900.0	1.251	1	1
364253	Sh_222_lllv	15447900.0	4.571	1	1
364254	Sh_222_llvv	14996000.0	12.502	1	1
410503	PwPy8EG_A14_ttbar_hdamp258p75_dil	19999000.0	831.76	1.0	0.105
410011	PwPyEG_P2012_singleTop_tchan_lept_top	4454000.0	43.739	1.0094	1.0
410012	PwPyEG_P2012_singleTop_tchan_lept_antitop	4998000.0	25.778	1.0193	1.0
410015	PwPyEG_P2012_Wt_dilepton_top	500000.0	3.5835	1.054	1.0
410016	PwPyEG_P2012_Wt_dilepton_antitop	500000.0	3.5814	1.054	1.0
410025	PwPyEG_P2012_SingleTopSchan_noAllHad_top	1000000.0	2.0517	1.0046	1.0
410026	PwPyEG_P2012_SingleTopSchan_noAllHad_antitop	1000000.0	1.2615	1.0215	1.0
364156	Sh_221_Wmunu_M0_70_CvBv	24966000.0	19143.0	0.9702	8.2380E-01
364157	Sh_221_Wmunu_M0_70_CfBv	9993000.0	19121.0	0.9702	1.3040E-01
364158	Sh_221_Wmunu_M0_70_Bf	17485200.0	19135.0	0.9702	4.4118E-02
364159	Sh_221_Wmunu_M70_140_CvBv	14964000.0	944.85	0.9702	6.7463E-01
364160	Sh_221_Wmunu_M70_140_CfBv	9975200.0	937.78	0.9702	2.3456E-01
364161	Sh_221_Wmunu_M70_140_Bf	19915000.0	944.63	0.9702	7.5648E-02
364162	Sh_221_Wmunu_M140_280_CvBv	10000000.0	339.54	0.9702	6.2601E-01
364163	Sh_221_Wmunu_M140_280_CfBv	7500000.0	340.06	0.9702	2.8947E-01
364164	Sh_221_Wmunu_M140_280_Bf	25000000.0	339.54	0.9702	1.0872E-01
364165	Sh_221_Wmunu_M280_500_CvBv	4999000.0	72.067	0.9702	5.4647E-01
364166	Sh_221_Wmunu_M280_500_CfBv	2989000.0	72.198	0.9702	3.1743E-01
364167	Sh_221_Wmunu_M280_500_Bf	2939500.0	72.045	0.9702	1.3337E-01
364168	Sh_221_Wmunu_M500_1000	5997500.0	15.01	0.9702	1.0000E+00
364169	Sh_221_Wmunu_M1000_E.CMS	3998000.0	1.2344	0.9702	1.0000E+00
364170	Sh_221_Wenu_M0_70_CvBv	24997000.0	19127.0	0.9702	8.2447E-01
364171	Sh_221_Wenu_M0_70_CfBv	9910500.0	19130.0	0.9702	1.3030E-01
364172	Sh_221_Wenu_M0_70_Bf	17482400.0	19135.0	0.9702	4.4141E-02
364173	Sh_221_Wenu_M70_140_CvBv	14815500.0	942.58	0.9702	6.6872E-01
364174	Sh_221_Wenu_M70_140_CfBv	9940400.0	945.67	0.9702	2.2787E-01
364175	Sh_221_Wenu_M70_140_Bf	9895900.0	945.15	0.9702	0.10341
364176	Sh_221_Wenu_M140_280_CvBv	10000000.0	339.81	0.9702	5.9691E-01
364177	Sh_221_Wenu_M140_280_CfBv	7498000.0	339.87	0.9702	2.8965E-01
364178	Sh_221_Wenu_M140_280_Bf	24860800.0	339.48	0.9702	1.0898E-01
364179	Sh_221_Wenu_M280_500_CvBv	4998800.0	72.084	0.9702	5.4441E-01
364180	Sh_221_Wenu_M280_500_CfBv	2999400.0	72.128	0.9702	3.1675E-01
364181	Sh_221_Wenu_M280_500_Bf	2998000.0	72.113	0.9702	1.3391E-01
364182	Sh_221_Wenu_M500_1000	5998600.0	15.224	0.9702	1.0000E+00
364183	Sh_221_Wenu_M1000_E.CMS	3999000.0	1.2334	0.9702	1.0000E+00
364184	Sh_221_Wtaunu_M0_70_CvBv	24921000.0	19152.0	0.9702	8.2495E-01
364185	Sh_221_Wtaunu_M0_70_CfBv	9964600.0	19153.0	0.9702	1.2934E-01
364186	Sh_221_Wtaunu_M0_70_Bf	17483200.0	19163.0	0.9702	4.4594E-02
364187	Sh_221_Wtaunu_M70_140_CvBv	14998500.0	947.65	0.9702	6.7382E-01
364188	Sh_221_Wtaunu_M70_140_CfBv	9999000.0	946.73	0.9702	2.2222E-01
364189	Sh_221_Wtaunu_M70_140_Bf	9189000.0	943.3	0.9702	0.10391
364190	Sh_221_Wtaunu_M140_280_CvBv	5194000.0	339.36	0.9702	5.9622E-01
364191	Sh_221_Wtaunu_M140_280_CfBv	7500000.0	339.63	0.9702	2.9025E-01
364192	Sh_221_Wtaunu_M140_280_Bf	24999900.0	339.54	0.9702	1.1799E-01
364193	Sh_221_Wtaunu_M280_500_CvBv	4999200.0	72.065	0.9702	5.4569E-01
364194	Sh_221_Wtaunu_M280_500_CfBv	2998400.0	71.976	0.9702	3.1648E-01
364195	Sh_221_Wtaunu_M280_500_Bf	2998100.0	72.026	0.9702	1.3426E-01
364196	Sh_221_Wtaunu_M500_1000	5960000.0	15.046	0.9702	1.0000E+00
364197	Sh_221_Wtaunu_M1000_E.CMS	3950000.0	1.2339	0.9702	1.0000E+00