UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
PROGRAMA DE PÓS GRADUAÇÃO EM FÍSICA
UNIDADE ACADÊMICA DE FÍSICA

Maria Gabriela Ferreira Siqueira Amaral Gomes

Measurement of the Invisible Width of the Z Boson Using Open Data from the CMS Experiment

Campina Grande, Paraíba, Brazil

June 10, 2025

Maria Gabriela Ferreira Siqueira Amaral Gomes

Measurement of the Invisible Width of the Z Boson Using Open Data from the CMS Experiment

A dissertation submitted to the Federal University of Campina Grande, Academic Unit of Physics, in partial fulfillment of the requirements for the degree of Master of Science in Physics, under the supervision of Prof. Dr. Diego Alejandro Cogollo Aponte and co-supervision of Prof. Dr. Gilvan Augusto Alves.

Campina Grande, Paraíba, Brazil

June 10, 2025

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
POS-GRADUACAO EM FISICA
Rua Aprigio Veloso, 882, - Bairro Universitario, Campina Grande/PB, CEP 58429-900

REGISTRO DE PRESENÇA E ASSINATURAS

ATTENDANCE LETTER OF REGISTER AND SIGNATURES
ATTENDANCE LETTER OF REGISTER AND SIGNATURES OF THE DEFENSE FOR AWARDING THE DEGREE OF MASTER IN PHYSICS, HELD ON JUNE 6, 2025

On the sixth day of the month of June in the year two thousand twenty-five, the members of the committee joined to evaluate **Maria Gabriela Ferreira Siqueira Amaral Gomes** for the degree of Master in Physics, area of research Physics, convened in a public solemn session. The Examining Committee was composed of the following specialists: Professor **Diego Alejandro Cogollo Aponte** (Advisor) – PhD in Physics; **Gilvan Augusto Alves** (Co-advisor) – PhD in Physics; Professor **Luiz Augusto Stuani Pereira** – PhD in Physics; and Professor **Patrícia Rebello Teles** – PhD in Physics. The first and third members are faculty of the Post-Graduate Program in Physics at the Federal University of Campina Grande, and the second and fourth are researchers at the Brazilian Center for Physics Research (CBPF). START TIME: 10:00 AM – LOCATION: Virtual Room. Opening the proceedings, the Committee President, Professor **Diego Alejandro Cogollo Aponte**, after stating the purpose of the meeting, introduced the candidate **Maria Gabriela Ferreira Siqueira Amaral Gomes**, who was then granted the floor to present succinctly and orally her dissertation entitled "**Measurement of the Invisible Width of the Z Boson Using CMS Open Data**." After expounding on her topic, the candidate was questioned by the examiners in accordance with the regulations. Thereafter, the Committee retired in private to evaluate and judge the work, concluding with the decision to assign the grade Approved. In view of this approval, the President declared the candidate legally qualified to receive the Degree of Master in Physics, and directed the Federal University of Campina Grande to take the necessary steps to issue the corresponding diploma. There being no further business, this letter was drawn up and will be submitted for approval by the Examining Committee. Campina Grande, June 6, 2025.

Signatures:

Diego Alejandro Cogollo Aponte
President of the Committee and Advisor

Gilvan Augusto Alves
Co-advisor

Luiz Augusto Stuani Pereira
Internal Examiner

Patrícia Rebello Teles
External Examiner

Maria Gabriela Ferreira Siqueira Amaral Gomes
Candidate

João Rafael Lucio dos Santos
Program Coordinator

2 – APPROVAL

2.1. The present Minutes of the Master's Thesis Defense of the candidate **Maria Gabriela Ferreira Siqueira Amaral Gomes** follow, electronically signed by the Examining Committee members identified above.

2.2. In the case of external examiners who do not hold an active external user credential in SEI for electronic signature, the internal examiners hereby certify that the external examiners named above participated in the thesis defense and have taken note of the contents of this document.

Documento assinado eletronicamente por **DIEGO ALEJANDRO COGOLLO APONTE**, **PROFESSOR(A) DO MAGISTERIO SUPERIOR**, em 06/06/2025, às 11:56, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da Portaria SEI nº 002, de 25 de outubro de 2018.

Documento assinado eletronicamente por **JOAO RAFAEL LUCIO DOS SANTOS**, **COORDENADOR(A)**, em 06/06/2025, às 12:00, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da Portaria SEI nº 002, de 25 de outubro de 2018.

Documento assinado eletronicamente por **Patricia Rebello Teles**, **Usuário Externo**, em 06/06/2025, às 12:03, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da Portaria SEI nº 002, de 25 de outubro de 2018.

Documento assinado eletronicamente por **Maria Gabriela Ferreira Siqueira Amaral Gomes**, **Usuário Externo**, em 06/06/2025, às 12:16, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da Portaria SEI nº 002, de 25 de outubro de 2018.

Documento assinado eletronicamente por **Gilvan Augusto Alves**, **Usuário Externo**, em 06/06/2025, às 12:20, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da Portaria SEI nº 002, de 25 de outubro de 2018.

Documento assinado eletronicamente por **Luiz Augusto Stuani Pereira**,

Usuário Externo, em 06/06/2025, às 12:32, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da Portaria SEI nº 002, de 25 de outubro de 2018.

A autenticidade deste documento pode ser conferida no site https://sei.ufcg.edu.br/autenticidade, informando o código verificador **5500181** e o código CRC **8B675146**.

Acknowledgements

This dissertation is the result of work I carried out over the course of two years. None of it would have been possible without the support, guidance, and encouragement I received from many people—both before and during my Master's— for which I am deeply grateful.

I feel truly privileged to have had the guidance of Professors Diego Cogollo, Gilvan Alves, and Patricia Telles throughout my Master's degree. I am sincerely thankful for their mentorship, constant support, and the trust they placed in me. The opportunities they provided were decisive for the development of my academic path. They are, to me, great inspirations—as people, educators, and researchers. I genuinely hope to one day give back by helping guide other students, just as they have done for me.

I also wish to thank the Graduate Program in Physics at the Federal University of Campina Grande (UFCG), the outstanding faculty, and the good friends I made along the way. I am especially grateful to Professor Luiz Stuani for serving on my thesis committee and for offering valuable feedback. I am equally thankful to the professors at the State University of Paraíba (UEPB), who were instrumental during my undergraduate years and continue to be significant in my journey.

My gratitude extends to FAPESQ for the Master's scholarship, and to the "Paraíba Sem Fronteiras" program, a partnership between FAPESQ and SECTIES, for the sandwich Master's scholarship.

I would like to thank the Open Data CMS and RPC-CMS groups for their guidance, support, and for all they taught me during my time at CERN.

To my many dear friends, even those who are far away, thank you for always being by my side, keeping me company, and bringing joy to my daily life. Many times, it was because of you that I found reasons to smile through the anxiety.

To my family—Maria, Vilanilza, Carlos, Rayanny, Jefferson, Elizabete, Cleyton, Marizabel, Zequinha, and Zara—my favorite people in the world, I am endlessly thankful for the love, encouragement, and dedication you have always given me. In the moments when I doubted myself, I thought of you. I dedicate this work to each of you, with all my affection.

Finally, I thank everyone who, in any way, contributed to the completion of this chapter. Every gesture, no matter how small, made a difference.

"É justo que muito custe o que muito vale" (Santa Teresa D'Ávila)

Abstract

This work presents a measurement of the invisible width of the Z boson, a fundamental parameter of the Standard Model, using CMS Open Data from proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$, corresponding to an integrated luminosity of $16{,}290\,\text{fb}^{-1}$. The invisible width of the Z boson, associated with the decay $Z \to \nu\bar{\nu}$, is inferred from the missing transverse momentum observed in $Z+$jets events. This approach provides insight into the coupling of the Z boson to neutrinos. The result obtained is compared with previous measurements published by the CMS Collaboration. In addition, the study includes an exploratory investigation of event selection using machine learning techniques, employing Graph Neural Networks (GNNs) as an alternative to traditional cut-based selection methods. The research also involves technical contributions to the upgrade of the CMS muon system, focusing on the assembly and testing of the Improved Resistive Plate Chambers (iRPCs). These gaseous detectors, designed for high-precision timing, will be installed in the endcap regions of the detector to extend pseudorapidity coverage and significantly enhance muon detection and tracking efficiency, particularly in the high-luminosity environment expected in future LHC upgrades.

Keywords: Z Boson, Standard Model, CMS Open Data, Invisible Width, GNN, Machine Learning, Deep Learning, iRPCs, Particle Physics.

List of Figures

List of Tables

Contents

1 Introduction

Data preservation is one of the key priorities of CERN (the European Organization for Nuclear Research). In this context, CMS Open Data, launched in 2014, stands out as a significant initiative by making proton-proton and heavy-ion collision data collected by the CMS (Compact Muon Solenoid) experiment publicly available. These data serve various purposes, such as the reinterpretation and reanalysis of results, the promotion of education and science outreach, and the development of technical and algorithmic tools [9].

This dissertation aims to perform a direct measurement of the invisible width of the Z boson, a key observable predicted by the Standard Model, and to compare the result with the measurement previously published by the CMS Collaboration [8]. According to the Standard Model, approximately 20% of all Z boson decays result in the production of a neutrino–antineutrino pair [10]. These so-called invisible decays cannot be directly observed in the CMS detector but can be inferred from the missing transverse momentum (p_T^{miss}) caused by the presence of neutrinos.

Measuring the invisible width of the Z boson provides a critical test of the Standard Model and offers insights into the number of neutrino species that couple to the Z boson. Deviations from the Standard Model prediction may indicate the presence of new particles beyond the Standard Model that contribute to the invisible decay width of the Z.

The invisible width of the Z boson is a well-established observable within the scientific community, with measurements performed by major experiments such as LEP, CMS, and ATLAS. Although this observable is well known, conducting the analysis in the context of CMS Open Data is highly valuable. It is presented here as a complete and educational example of a full particle physics analysis workflow: from event selection and pileup corrections, through data visualization and exploration, to the simultaneous fit required for a precise estimate of the Z boson's invisible width.

The analysis strategy considers signal events in which the Z boson is produced in association with jets ($Z+$jets) and decays into neutrinos, which are identified by the detector as missing transverse momentum (p_T^{miss}). In these cases, the p_T^{miss} balances the transverse momentum of the jets, reflecting momentum conservation in the interaction. The invisible width of the Z boson is inferred using a reference channel: the decay of the Z boson into charged leptons, which share similar kinematic features with the invisible decay. This similarity allows the extrapolation of the leptonic decay properties to accurately estimate the contribution from neutrinos.

$$\Gamma(Z \to \text{inv}) = \frac{\sigma(Z + \text{jets})\mathcal{B}(Z \to \text{inv})}{\sigma(Z + \text{jets})\mathcal{B}(Z \to \ell\bar{\ell})}\Gamma(Z \to \ell\bar{\ell}) \tag{1.1}$$

In this expression, $\sigma(Z + \text{jets})$ represents the cross section of the $Z + \text{jets}$ process, scaled by either the invisible branching ratio[1] $\mathcal{B}(Z \to \text{inv})$, or the branching ratio to dileptons $\mathcal{B}(Z \to \ell\bar{\ell})$.

For this study, we used data available from the CMS Open Data portal, corresponding to proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$, recorded in 2016, with an integrated luminosity of $16{,}290\,\text{fb}^{-1}$. The analysis was conducted using the Python programming language, and the code is publicly available in a GitHub repository at: <https://github.com/Mariaggomes>.

This work also includes a complementary study on event selection using machine learning techniques. Specifically, the use of Graph Neural Networks (GNNs) was explored to identify events compatible with the $Z \to \nu\bar{\nu}$ signal region. This type of neural network, which operates directly on graph-based representations, has shown promising results for classification tasks in particle physics, especially in scenarios involving multiple reconstructed objects per event.

In addition to the analysis, technical activities were conducted during a sandwich Master's internship at CERN, under the Paraíba Sem Fronteiras program. These activities were related to the assembly and testing of Improved Resistive Plate Chambers (iRPCs). RPCs are gaseous detectors with excellent time resolution, located both in the barrel and endcap regions of CMS[2]. These detectors play a crucial role in the CMS trigger system for muon detection, providing precise timing information essential for distinguishing closely spaced collision events.

To prepare for the expected increase in luminosity in the coming years, the CMS muon system is undergoing an upgrade. New RPC chambers—called iRPCs—will be installed in the endcap region, where the current muon detection coverage is limited. These new detectors will extend the pseudorapidity range from 1.9 to 2.4 [11]. Furthermore, the iRPCs offer significant improvements over the original RPCs, such as thinner electrodes and narrower gas gaps, which lead to more accurate muon tracking.

As part of the CMS muon system upgrade, in response to the expected increase in luminosity in the coming years, new RPC chambers—called iRPCs—will be installed in the endcap region, where the current detector coverage is limited for muon detection. These iRPCs will extend the muon system coverage by increasing the pseudorapidity range

[1] The invisible branching ratio of a process represents the fraction of all particle decays that occur via a specific (invisible) channel.

[2] The barrel region corresponds to the central cylindrical part of the detector surrounding the proton beam line, while the endcaps are the detector's ends, positioned perpendicularly to the beam, covering regions of higher pseudorapidity (η).

from 1.9 to 2.4 [11]. In addition, iRPCs include significant improvements over the original RPCs, such as thinner electrodes and narrower gas gaps, resulting in more precise muon tracking.

In the following chapters, the Standard Model of Particle Physics, the Large Hadron Collider (LHC), and the CMS experiment will be discussed, along with the obtained results, including the study using machine learning, the extraction of the invisible width of the Z boson, and the contribution to the upgrade of the CMS muon system.

2 The Standard Model of Particle Physics

The Standard Model (SM) is the theory that describes the elementary particles and the fundamental interactions—except for gravity. It encompasses three of the four fundamental forces: the strong, weak, and electromagnetic interactions. The theory is grounded in the gauge principle and based on the symmetry group [12]:

$$SU(3)_C \times SU(2)_L \times U(1)_Y. \tag{2.1}$$

The fundamental interactions in the SM are described through gauge symmetries, which dictate the form of the interactions between fields. To understand the structure of the Standard Model, we begin with Quantum Electrodynamics (QED) as a basic example of the gauge principle.

2.1 Quantum Electrodynamics (QED) and the Gauge Principle

In QED, the gauge symmetry is associated with the group $U(1)$. The initial Lagrangian for a free fermion with mass m is:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi, \tag{2.2}$$

where ψ is the fermion field, γ^μ are the Dirac matrices, and ∂_μ is the ordinary derivative [13]. This Lagrangian is invariant under global phase transformations:

$$\psi \to e^{i\alpha}\psi, \tag{2.3}$$

with α being a constant.

When the phase transformation becomes dependent on spacetime coordinates ($\alpha = \alpha(x)$), we have a local transformation:

$$\psi(x) \to e^{i\alpha(x)}\psi(x). \tag{2.4}$$

In this case, the ordinary derivative $\partial_\mu \psi$ is no longer invariant:

$$\partial_\mu \psi \to e^{i\alpha(x)}(\partial_\mu \psi + i(\partial_\mu \alpha)\psi). \tag{2.5}$$

To restore invariance, we introduce the gauge field A_μ, which transforms as:

$$A_\mu \to A_\mu + \partial_\mu \alpha. \tag{2.6}$$

The ordinary derivative is replaced by the covariant derivative:

$$D_\mu = \partial_\mu + ieA_\mu, \tag{2.7}$$

where e is the charge of the fermion. The Lagrangian invariant under local transformations is then:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{2.8}$$

with the field strength tensor defined as:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{2.9}$$

This Lagrangian describes the interaction between charged fermions (such as electrons) and the photon (A_μ), which is the gauge boson of QED. Therefore, QED is a gauge theory based on the group $U(1)$. The first term represents the dynamics of the fermion and its interaction with the electromagnetic field through the covariant derivative D_μ, while the second term corresponds to the energy of the electromagnetic field, expressed by the field strength tensor $F_{\mu\nu}$.

2.2 The Electroweak Theory

The electroweak theory unifies the electromagnetic and weak interactions and is described by the gauge group $SU(2)_L \times U(1)_Y$. Fundamental particles are organized into doublets and singlets, reflecting their symmetry properties. Left-handed fermions belong to the fundamental representation—also known as a doublet—of the $SU(2)_L$ group, while right-handed fermions are in the singlet representation [13].

$$f_L^a = \begin{pmatrix} \nu^a \\ \ell^a \end{pmatrix}_L, \quad \ell_R^a, \tag{2.10}$$

$$\nu^1 = \nu_e, \quad \nu^2 = \nu_\mu, \quad \nu^3 = \nu_\tau,$$

$$\ell^1 = e^-, \quad \ell^2 = \mu^-, \quad \ell^3 = \tau^-.$$

Meanwhile,

$$Q_L^a = \begin{pmatrix} u^a \\ d^a \end{pmatrix}_L, \quad u_R^a, \quad d_R^a, \tag{2.11}$$

$$u^1 = u, \quad u^2 = c, \quad u^3 = t,$$

$$d^1 = d, \quad d^2 = s, \quad d^3 = b.$$

The covariant derivative for doublets (Q_L, L_L) is given by:

$$D_\mu Q_L = \partial_\mu Q_L + igW^a_\mu T^a Q_L + ig'\frac{Y}{2}B_\mu Q_L, \tag{2.12}$$

and for singlets (u_R, d_R, e_R), it is:

$$D_\mu \psi_R = \partial_\mu \psi_R + ig'\frac{Y}{2}B_\mu \psi_R. \tag{2.13}$$

In these expressions, Q_L and L_L represent the quark and lepton doublets, which transform under the $SU(2)_L$ symmetry group. The fields W^a_μ (with $a = 1, 2, 3$) and B_μ are the gauge bosons associated with the $SU(2)_L$ and $U(1)_Y$ groups, respectively. The T^a are the generators of $SU(2)_L$, given by $T^a = \tau^a/2$, where τ^a are the Pauli matrices, and Y denotes the weak hypercharge of the field.

2.3 The Theory of the Strong Interaction: Quantum Chromodynamics (QCD)

Within the context of the Standard Model, the group $SU(2)_L \times U(1)_Y$ describes the unification of the electromagnetic and weak interactions in the electroweak theory, while the group $SU(3)_C$ is associated with the strong interaction [13]. Quantum Chromodynamics (QCD) is the component of the Standard Model that governs the strong interaction and describes how quarks and gluons interact through color charges. The QCD Lagrangian can be written as:

$$\mathcal{L}_{\text{QCD}} = -\frac{1}{4}G^a_{\mu\nu}G^{a\mu\nu} + \sum_q \bar{\psi}_q \left(i\gamma^\mu D_\mu - m_q\right)\psi_q, \tag{2.14}$$

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu \tag{2.15}$$

Here, $G^a_{\mu\nu}$ is the field strength tensor, G^a_μ are the gluon fields, g_s is the strong coupling constant, and f^{abc} are the structure constants of the $SU(3)_C$ group.

- ψ_q represents the quark fields of different flavors.
- $D_\mu = \partial_\mu - ig_s T^a G^a_\mu$ is the covariant derivative, with T^a being the generators of the $SU(3)_C$ group.
- m_q are the quark masses.

A characteristic phenomenon of QCD is color confinement: when two color-charged particles are pulled apart, the energy of the field between them increases until it becomes energetically more favorable to create a new quark–antiquark pair than to continue the separation. As a result, quarks cannot be observed in isolation in nature; they are always confined within color-neutral combinations known as hadrons [12]. These hadrons can be classified into:

- Mesons: combinations of one quark and one antiquark ($q\bar{q}$).
- Baryons: combinations of three quarks (qqq).

For example, the proton is a baryon composed of three quarks (uud).

Another important phenomenon in QCD is asymptotic freedom. At high energies (or equivalently, short distances), the interaction between quarks becomes weaker, and quarks can be treated as nearly free particles. This behavior is described by the running of the strong coupling constant α_s, which decreases with increasing energy scale Q^2. The energy dependence of α_s is given by:

$$\alpha_s(Q^2) = \frac{1}{\beta_0 \ln(Q^2/\Lambda^2_{\mathrm{QCD}})}, \tag{2.16}$$

where:

- β_0 is the coefficient of the QCD beta function, which depends on the number of active quark flavors.
- Λ_{QCD} is the characteristic energy scale of QCD.

2.4 The Standard Model Lagrangian

In this way, the Standard Model is capable of describing the dynamics of particles and their interactions, including both the electroweak theory and the theory of strong interactions. This theoretical framework provides a comprehensive description of the fundamental interactions of elementary particles, organized under the gauge symmetries $SU(3)_C \times SU(2)_L \times U(1)_Y$.

The terms that compose the dynamics of the Standard Model are as follows:

- $\mathcal{L}_{\text{leptons}}$: describes the dynamics of leptons and their interactions.
- $\mathcal{L}_{\text{quarks}}$: describes the dynamics of quarks and their interactions.
- $\mathcal{L}_{\text{scalar}}$: includes the contribution from the Higgs scalar field, responsible for electroweak symmetry breaking.

- $\mathcal{L}_{\text{Yukawa}}$: defines the Yukawa couplings, which are responsible for the masses of the fundamental particles.
- $\mathcal{L}_{\text{gauge}}$: governs the dynamics of the gauge fields that mediate the fundamental interactions.

These components combine to form the complete Lagrangian of the Standard Model, expressed as:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{leptons}} + \mathcal{L}_{\text{quarks}} + \mathcal{L}_{\text{scalar}} + \mathcal{L}_{\text{Yukawa}} + \mathcal{L}_{\text{gauge}}. \tag{2.17}$$

2.5 The Fermion Sector

The first two terms of the Standard Model Lagrangian represent the fermionic sector, describing the interactions of fermions with the physical gauge bosons, both neutral and charged. The fermionic Lagrangian is expressed as:

$$\mathcal{L}_{\text{fermions}} = \sum_{a=1}^{3} \left[\bar{\psi}_L^a i\gamma^\mu D_\mu^L \psi_L^a + \bar{\psi}_R^a i\gamma^\mu D_\mu^R \psi_R^a \right]. \tag{2.18}$$

In this expression:

- When $\psi_L^a = f_L^a$ and $\psi_R^a = \ell_R^a$, we refer to the leptonic Lagrangian.
- When $\psi_L^a = Q_L^a$ and $\psi_R^a = u_R^a, d_R^a$, we refer to the quark Lagrangian.

The fields and covariant derivatives transform under the gauge group $SU(2)_L \times U(1)_Y$ as follows:

$$\psi_L(x) \to \psi'_L(x) = e^{it^a\alpha^a(x)+i\frac{Y}{2}\beta(x)}\psi_L(x) \tag{2.19}$$

$$\bar{\psi}_L(x) \to \bar{\psi}'_L(x) = \bar{\psi}_L(x)e^{-it^a\alpha^a(x)-i\frac{Y}{2}\beta(x)} \tag{2.20}$$

$$\psi_R(x) \to \psi'_R(x) = e^{i\frac{Y}{2}\beta(x)}\psi_R(x) \tag{2.21}$$

$$\bar{\psi}_R(x) \to \bar{\psi}'_R(x) = \bar{\psi}_R(x)e^{-i\frac{Y}{2}\beta(x)} \tag{2.22}$$

The covariant derivatives transform as:

$$D_\mu^L \psi_L \to D_\mu^{L\prime} \psi'_L = e^{it^a\alpha^a(x)+i\frac{Y}{2}\beta(x)} D_\mu^L \psi_L, \tag{2.23}$$

$$D_\mu^R \psi_R \to D_\mu^{R\prime} \psi'_R = e^{i\frac{Y}{2}\beta(x)} D_\mu^R \psi_R. \tag{2.24}$$

These transformations ensure that the fermionic Lagrangian remains invariant under the gauge group $SU(2)_L \times U(1)_Y$ [12].

The fermions in the Standard Model can be classified into two groups:

- Quarks:
  - Participate in all fundamental interactions (strong, weak, and electromagnetic).
  - Possess three color states.
  - There are six flavors: up (u), down (d), charm (c), strange (s), top (t), and bottom (b).
- Leptons:
  - Carry no color charge and therefore do not participate in the strong interaction.
  - Include the electron (e), muon (μ), tau (τ), and their respective neutrinos (ν_e, ν_μ, ν_τ).

Fermions are organized into three generations or families:

- 1st generation (ordinary matter):

$$\left\{ \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, e_R, \begin{pmatrix} u \\ d \end{pmatrix}_L, u_R, d_R \right\}.$$

- 2nd generation:

$$\left\{ \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \mu_R, \begin{pmatrix} c \\ s \end{pmatrix}_L, c_R, s_R \right\}.$$

- 3rd generation:

$$\left\{ \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L, \tau_R, \begin{pmatrix} t \\ b \end{pmatrix}_L, t_R, b_R \right\}.$$

These generations share the same quantum numbers and interactions, differing only in mass. Particles from the second and third generations decay rapidly into first-generation particles, which make up the ordinary matter observed in the universe [12].

2.6 The Gauge Sector

This part of the Lagrangian gives rise to the triple and quartic interactions among the gauge bosons themselves, as well as their dynamics, described by:

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \tag{2.25}$$

where:

$$G^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + g\epsilon^{abc} W^b_\mu W^c_\nu, \tag{2.26}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{2.27}$$

with $a, b, c = 1, 2, 3$.

In this expression, the gauge bosons W_μ^a and B_μ are referred to as symmetric gauge bosons, as they belong to a Lagrangian that is invariant under the gauge group $SU(2)_L \times U(1)_Y$ [13]. These gauge bosons are initially massless. Adding mass terms directly, such as $m_W^2 W_\mu^a W^{a\mu}$, would explicitly break the $SU(2)_L \times U(1)_Y$ symmetry.

However, it is experimentally known that the mediators of the weak interaction ($W^\pm$ and Z^0) are quite massive, while the photon (γ), associated with electromagnetic interactions, remains massless [12]. This highlights the need for a mechanism that generates mass only for the gauge bosons mediating the weak interaction, while preserving gauge invariance for the photon.

To ensure gauge invariance, the infinitesimal gauge transformations of the gauge fields are given by:

$$W_\mu^a \to W_\mu^{a\prime} = W_\mu^a - \frac{1}{g}\partial_\mu \alpha^a - \epsilon^{abc}\alpha^b W_\mu^c, \tag{2.28}$$

$$B_\mu \to B'_\mu = B_\mu + \partial_\mu \beta. \tag{2.29}$$

When applying these transformations, one finds that mass terms such as $m_W^2 W_\mu^a W^{a\mu}$ are not invariant, since they transform as:

$$W_\mu^a W^{a\mu} \to \left(W_\mu^a - \frac{1}{g}\partial_\mu \alpha^a - \epsilon^{abc}\alpha^b W_\mu^c\right)\left(W^{a\mu} - \frac{1}{g}\partial^\mu \alpha^a - \epsilon^{abc}\alpha^b W^{\mu c}\right). \tag{2.30}$$

As a result, it is not possible to introduce mass terms directly for the symmetric gauge bosons without breaking the $SU(2)_L \times U(1)_Y$ symmetry. The mechanism responsible for generating masses for the $W^\pm$ and Z^0 bosons—while keeping the photon massless—will be discussed in the following section. It is essential to explain the extremely short range of the weak interaction and the dynamics of the electroweak interactions.

2.7 The Yukawa Sector

A direct mass term for charged leptons, such as $M\bar{\ell}\ell = M[\bar{\ell}_L\ell_R + \bar{\ell}_R\ell_L]$, would explicitly violate gauge invariance. To circumvent this, it is required that charged leptons be massless prior to spontaneous symmetry breaking, and that their masses arise only after the symmetry is broken. In the Standard Model, the inclusion of mass terms for charged leptons in a gauge-invariant way is achieved through the coupling of the lepton doublet f_L^a with a scalar doublet Φ and the right-handed singlet ℓ_R^b:

$$\mathcal{L}_{Y,\text{leptons}} = -G_{ab}\left[(\bar{f}_L^a\Phi)\ell_R^b\right] + \text{H.C.}, \tag{2.31}$$

where G_{ab} are the Yukawa coupling constants, and the term H.C. (Hermitian Conjugate) refers to the hermitian conjugate of the preceding expression, ensuring that the total Lagrangian is real. After spontaneous symmetry breaking, this Lagrangian becomes:

$$\mathcal{L}_{Y,\text{leptons}} = -\left[\bar{\ell}_L^a\left(\frac{v_\Phi G_{ab}}{\sqrt{2}}\right)\ell_R^b\right] + \text{H.C.}, \tag{2.32}$$

with the mass matrix for the charged leptons given by:

$$M_\ell = \frac{v_\Phi G_{ab}}{\sqrt{2}}. \tag{2.33}$$

2.8 Spontaneous Symmetry Breaking

The generation of masses for the gauge bosons in the Standard Model is achieved through spontaneous symmetry breaking via the Higgs mechanism. This mechanism is implemented by introducing a scalar field Lagrangian that is invariant under transformations of the $SU(2)_L \times U(1)_Y$ group. The scalar Lagrangian is constructed from a complex scalar doublet Φ, defined as:

$$\Phi = \begin{pmatrix}\Phi^+ \\ \Phi^0\end{pmatrix}, \quad Y_\Phi = 1. \tag{2.34}$$

The scalar Lagrangian of the Standard Model is given by:

$$\mathcal{L}_{\text{scalar}} = (D_\mu^L\Phi)^\dagger(D_L^\mu\Phi) - V(\Phi), \tag{2.35}$$

where the scalar potential is written as:

$$V(\Phi) = \mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2. \tag{2.36}$$

Spontaneous symmetry breaking occurs when the complex scalar field Φ^0 acquires a nonzero vacuum expectation value ($\langle\Phi^0\rangle_0 \neq 0$). This breaks the $SU(2)_L \times U(1)_Y$ symmetry down to $U(1)_{\text{EM}}$, preserving the electromagnetic gauge invariance [13].

The field configuration that minimizes $V(\Phi)$, denoted $\langle\Phi\rangle_0$, depends on the sign of μ^2:

- For $\mu^2 > 0$, the potential has a minimum at the origin:

$$\langle\Phi\rangle_0 = \begin{pmatrix}0 \\ 0\end{pmatrix}.$$

In this case, the $SU(2)_L \times U(1)_Y$ symmetry remains unbroken.

- For $\mu^2 < 0$, the potential acquires a nontrivial minimum:

$$\langle\Phi\rangle_0^\dagger\langle\Phi\rangle_0 = \frac{-\mu^2}{2\lambda}.$$

Parametrizing the scalar field as:

$$\Phi = \begin{pmatrix} \phi^+ \\ (v + H + iG^0)/\sqrt{2} \end{pmatrix},$$

the minimum condition is satisfied for $\phi^+ = G^0 = H = 0$, resulting in:

$$\langle\Phi\rangle_0 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix},$$

where $v = \sqrt{-\mu^2/\lambda}$ is the vacuum expectation value (VEV) of the scalar field ϕ^0. The electric charge operator is defined such that:

$$\exp(iQ)\langle\Phi\rangle_0 = \langle\Phi\rangle_0, \tag{2.37}$$

which implies:

$$(1 + iQ)\langle\Phi\rangle_0 = \langle\Phi\rangle_0,$$
$$\langle\Phi\rangle_0 + iQ\langle\Phi\rangle_0 = \langle\Phi\rangle_0.$$

Therefore:

$$Q\langle\Phi\rangle_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ v_\Phi/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{2.38}$$

This demonstrates that the vacuum has no electric charge, i.e., the electric charge operator annihilates the vacuum state [13]. This condition ensures that the photon remains massless after symmetry breaking, while the other gauge bosons ($W^\pm$ and Z^0) acquire mass via the Higgs mechanism.

2.9 The Physical Gauge Bosons

The covariant derivative of the Standard Model can be written in matrix form as:

$$D_\mu^L = \partial_\mu + \frac{ig}{2}\begin{pmatrix} W_\mu^3 & \sqrt{2}W_\mu^+ \\ \sqrt{2}W_\mu^- & -W_\mu^3 \end{pmatrix} + \frac{ig'}{2}YB_\mu, \tag{2.39}$$

where the fields $W_\mu^\pm$ are defined as:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2).$$

For the Higgs field with $Y_\phi = 1$, we have:

$$D_\mu^L \langle\Phi\rangle_0 = \begin{pmatrix} \partial_\mu + \frac{ig}{2}W_\mu^3 + \frac{ig'}{2}B_\mu & \frac{ig}{\sqrt{2}}W_\mu^+ \\ \frac{ig}{\sqrt{2}}W_\mu^- & \partial_\mu - \frac{ig}{2}W_\mu^3 + \frac{ig'}{2}B_\mu \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_\Phi \end{pmatrix}. \tag{2.40}$$

By eliminating the kinetic terms of the Higgs boson, we are left with the terms responsible for the masses of the gauge bosons:

$$D_\mu^L \langle\Phi\rangle_0 = \begin{pmatrix} \frac{ig}{\sqrt{2}}W_\mu^+ v_\Phi \\ -\frac{iv_\Phi}{\sqrt{2}}\left(W_\mu^3 - tB_\mu\right) \end{pmatrix}, \tag{2.41}$$

with $t = g'/g$.

The quadratic term $|D_\mu^L \langle\Phi\rangle_0|^2$ leads to the following expression:

$$|D_\mu^L \langle\Phi\rangle_0|^2 = \frac{g^2 v_\Phi^2}{4} W_\mu^- W_\mu^+ + \frac{g^2 v_\Phi^2}{8}\left(W_\mu^3 W_3^\mu + t^2 B_\mu B^\mu - 2tB_\mu W_3^\mu\right). \tag{2.42}$$

From this expression, the $W_\mu^\pm$ bosons emerge as mass eigenstates—known as the charged gauge bosons. Meanwhile, the W_μ^3 and B_μ bosons, associated with the diagonal generators, mix [13]. The mixing matrix in the $\{B_\mu, W_\mu^3\}$ basis is given by:

$$\frac{g^2 v_\Phi^2}{8}\begin{pmatrix} t^2 & -t \\ -t & 1 \end{pmatrix}. \tag{2.43}$$

The eigenvectors of this matrix correspond to the photon A_μ and the neutral boson Z_μ^0:

$$\begin{pmatrix} A_\mu \\ Z_\mu^0 \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix}\begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix}, \tag{2.44}$$

where the Weinberg angle (θ_W) is defined as:

$$\cos\theta_W = \frac{g}{\sqrt{g^2 + g'^2}}, \quad \sin\theta_W = \frac{g'}{\sqrt{g^2 + g'^2}}. \tag{2.45}$$

The corresponding eigenvalues are:

$$\lambda_{A_\mu} = 0, \quad \lambda_{Z_\mu^0} = \frac{g^2 v_\Phi^2}{8}(t^2 + 1). \tag{2.46}$$

Thus, the photon A_μ is the only massless gauge boson, while Z_μ^0 and $W_\mu^\pm$ acquire mass through the Higgs mechanism [13].

2.10 Lifetime and Decay Width of Particles

In the Standard Model, each unstable particle has an associated lifetime determined by its decay into lighter particles, in accordance with conservation laws such as energy, momentum, electric charge, and color charge [14]. The decay width (Γ) describes the total decay rate of a particle into all possible final states and is directly related to the particle's lifetime (τ) through the relation:

$$\tau = \frac{\hbar}{\Gamma}. \tag{2.47}$$

The total decay width (Γ) is the sum of the partial widths (Γ_i) for each decay channel:

$$\Gamma = \sum_i \Gamma_i. \tag{2.48}$$

The fraction of decays occurring in a specific channel is described by the branching ratio ($\mathcal{B}_i$):

$$\mathcal{B}_i = \frac{\Gamma_i}{\Gamma}. \tag{2.49}$$

As discussed in the previous chapter, the invisible width of the Z boson is determined using a reference signal: the decay width of the Z into leptons. The similarity between the decays of the Z boson into neutrinos and into charged leptons is exploited to calculate the invisible width. This relationship is expressed as:

$$\Gamma(Z \to \text{inv}) = \frac{\sigma(Z + \text{jets})\mathcal{B}(Z \to \text{inv})}{\sigma(Z + \text{jets})\mathcal{B}(Z \to \ell\bar{\ell})}\Gamma(Z \to \ell\bar{\ell}). \tag{2.50}$$

where:

- $\Gamma(Z \to \text{inv})$: decay width of the Z boson into neutrinos;
- $\sigma(Z + \text{jets})$: cross section of the $Z + \text{jets}$ process;
- $\mathcal{B}(Z \to \text{inv})$: branching ratio of the Z boson decaying into neutrinos;
- $\mathcal{B}(Z \to \ell\bar{\ell})$: branching ratio of the Z boson decaying into charged dileptons ($\ell\bar{\ell}$);
- $\Gamma(Z \to \ell\bar{\ell})$: decay width of the Z boson into lepton pairs.

2.11 The Number of Neutrinos

The invisible width of the Z boson plays a crucial role in determining the number of neutrino families that couple to the Z. This number is directly related to the invisible decay width of the Z into neutrinos (Γ_{inv}). In the Standard Model, it is assumed that each neutrino flavor (ν_e, ν_μ, ν_τ) contributes equally to Γ_{inv} [15]. Thus, the effective number of neutrinos (N_ν) can be calculated as:

$$N_\nu = \frac{\Gamma_{\text{inv}}}{(\Gamma_{\nu\bar{\nu}})_{\text{SM}}}. \tag{2.51}$$

where $(\Gamma_{\nu\bar{\nu}})_{\text{SM}}$ is the theoretical decay width of the Z boson into a neutrino–antineutrino pair in the Standard Model.

To relate the invisible width of the Z to the observable decay width into dileptons ($\Gamma_{\ell\bar{\ell}}$), the experimental ratio R^{exp} is used, defined as:

$$R^{\text{exp}} = \frac{\Gamma_{\text{inv}}}{\Gamma_{\ell\bar{\ell}}}, \tag{2.52}$$

Combining this information, the effective number of neutrinos can be rewritten as:

$$N_\nu = R^{\text{exp}} \left(\frac{\Gamma_{\ell\bar{\ell}}}{\Gamma_{\nu\bar{\nu}}}\right)_{\text{SM}}, \tag{2.53}$$

where $\left(\frac{\Gamma_{\nu\bar{\nu}}}{\Gamma_{\ell\bar{\ell}}}\right)_{\text{SM}} = 1.991 \pm 0.001$ is the Standard Model prediction for the ratio of decay widths into neutrinos and charged leptons [15].

Using the experimental value [15]:

$$R^{\text{exp}} = 5.942 \pm 0.016, \tag{2.54}$$

we obtain:

$$N_\nu = \frac{5.942}{1.991} = 2.984 \pm 0.008. \tag{2.55}$$

This value is in agreement with the existence of three fermion families in the Standard Model [15].

3 The Large Hadron Collider

3.1 Introduction

The Large Hadron Collider (LHC) is a particle accelerator located at CERN, on the border between Switzerland and France, designed to collide protons or lead ions at high energies. It was instrumental in the discovery of the Higgs boson by the ATLAS and CMS experiments and plays a key role in the search for physics beyond the Standard Model by setting constraints on the parameters of new theoretical scenarios. The accelerator consists of two rings with a circumference of 27 km, situated approximately 100 meters underground [16].

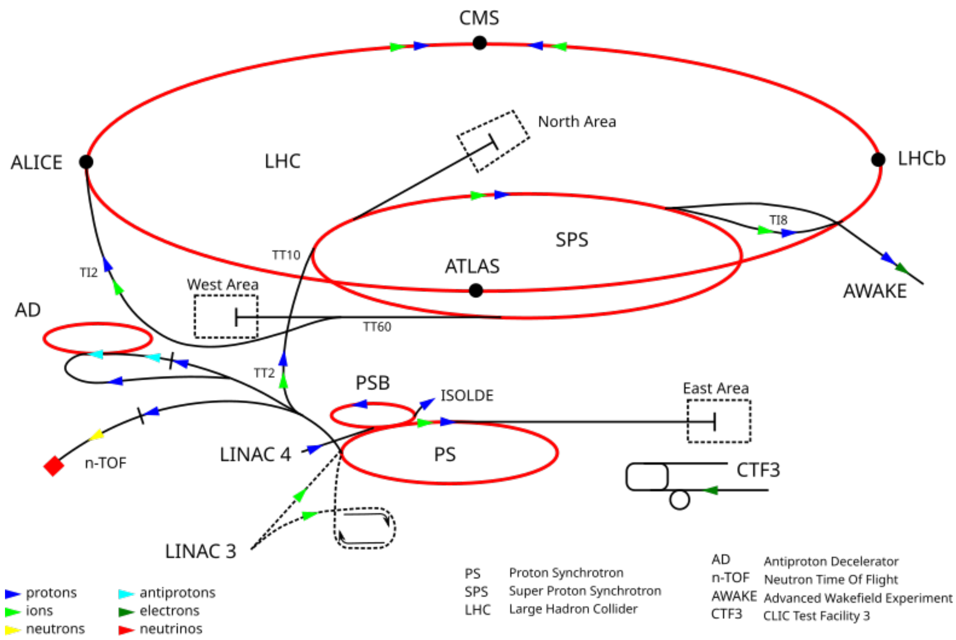

Figure 1 – Diagram of the CERN accelerator complex. Source: [1].

The acceleration process begins with the ionization of hydrogen gas, which separates protons from their bound atomic states [17]. These protons are accelerated in stages through a complex of accelerators, composed of:

- Linac2: A linear accelerator that increases the proton energy to 50 MeV.
- PS Booster (PSB): A set of four synchrotron rings that boost the energy to 1.4 GeV and intensify the proton beam.
- Proton Synchrotron (PS): A synchrotron with a radius of 72 meters that accelerates the protons to 25 GeV and arranges them into 4 ns bunches spaced 25 ns apart.

- Super Proton Synchrotron (SPS): A synchrotron with a 6.9 km circumference that accelerates the protons to 450 GeV before injecting them into the LHC.

In the LHC, protons are organized into 2808 bunches, separated by 25 ns, and reach energies of up to 6.5 TeV per beam, totaling 13 TeV in center-of-mass energy during collisions [17].

The performance of the LHC is primarily characterized by the beam energy and the luminosity. The luminosity (L) is a key parameter that describes the number of collisions per unit time and area. For a Gaussian beam, it is given by:

$$L = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi \epsilon_n \beta^*} F, \tag{3.1}$$

where:

- N_b: number of protons per bunch,
- n_b: number of bunches per beam,
- f_{rev}: revolution frequency,
- γ_r: relativistic gamma factor,
- ϵ_n: normalized transverse emittance,
- β^*: beta function at the collision point,
- F: geometric reduction factor due to the crossing angle.

The event rate (N) is directly proportional to the integrated luminosity and the cross section (σ):

$$N = L \cdot \sigma. \tag{3.2}$$

During its 2016 run, the LHC achieved instantaneous luminosities of $10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$ [18].

3.2 The CMS Experiment

The CMS (*Compact Muon Solenoid*) is one of the two largest general-purpose detectors at the LHC, located at “Point 5” of the accelerator ring. With dimensions of 21.6 meters in length, 14.6 meters in diameter, and a total weight of 12,500 tons, it was designed to detect particles produced in proton-proton collisions at energies of up to 14 TeV [19].

In 2016, CMS recorded collisions with a center-of-mass energy of 13 TeV and an average of 25 interactions per bunch crossing, producing approximately 1000 charged particles every 25 ns [18]. These data were used in the analysis presented in this dissertation.

The CMS detector is composed of four main subsystems:

- Silicon Tracker: responsible for reconstructing the trajectories of charged particles.
- Electromagnetic Calorimeter (ECAL): measures the energy of electromagnetic particles such as electrons and photons.
- Hadronic Calorimeter (HCAL): responsible for detecting and measuring the energy of hadrons.
- Muon System: essential for the identification and precise measurement of muons.

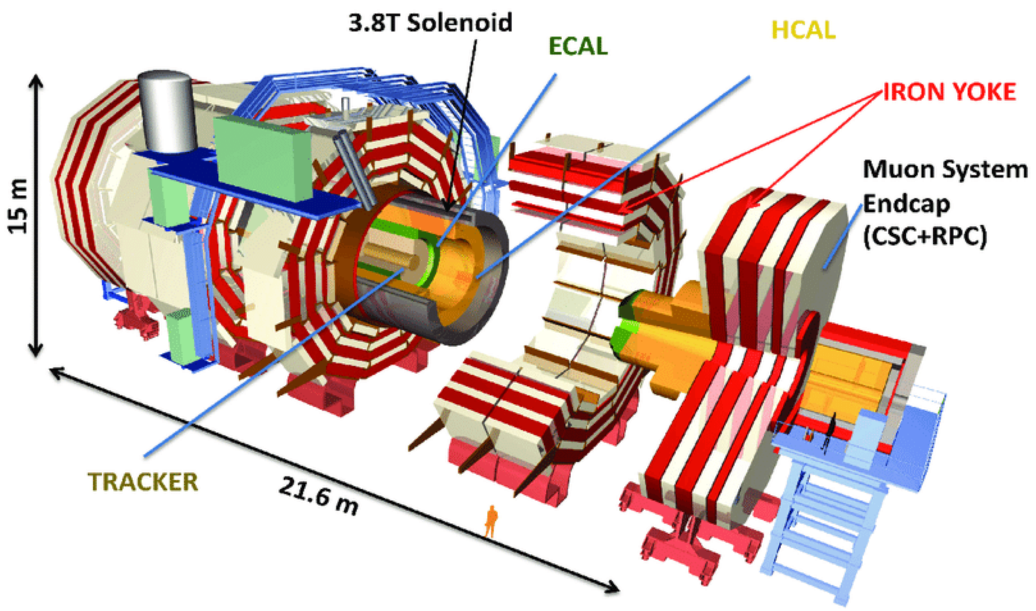

Figure 2 – Schematic view of the CMS detector. Source: [2].

In addition to these subsystems, CMS features advanced electronics for data readout and acquisition, enabling the recording and analysis of collision events.

3.2.1 Superconducting Solenoid

The central component of the CMS detector is a superconducting solenoid that generates a magnetic field of 3.8 T, surrounded by an iron yoke[1] that extends the field to the outer region, where the muon subdetectors are located [19]. This magnetic field causes charged particles to follow curved trajectories, allowing the determination of their electric charge and the precise measurement of their momentum.

[1] The yoke is an iron structure used to contain and redirect the magnetic field lines generated by the solenoid.

Together with the tracker and calorimeters, the magnetic field plays a crucial role in the reconstruction of the transverse momentum (p_T) of particles, contributing to CMS's high performance in complex event analyses.

3.2.2 Silicon Tracker

The Silicon Tracker is the innermost subdetector of CMS, designed to measure the position of charged particles with high precision and covering a pseudorapidity range of $|\eta| < 2.5$ [19]. It is essential for reconstructing primary and secondary vertices, as well as the trajectories of charged particles. The tracker is composed of two main subsystems:

- Pixel Modules: Located closest to the interaction point, these consist of three cylindrical barrel layers and two disks in each endcap. With 66 million pixels, they provide high granularity and spatial resolution, which are crucial in high particle flux environments.

- Silicon Strips: Positioned around the Pixel Tracker, these consist of 9.6 million silicon strips arranged in barrel layers and endcap disks. They cover an active area of $198\,\mathrm{m}^2$ and provide complementary information for three-dimensional reconstruction.

3.2.3 Electromagnetic Calorimeter (ECAL)

The ECAL 3 is a homogeneous calorimeter made of lead tungstate crystals, designed to measure electromagnetic energy deposits from electrons and photons [19]. It covers up to $|\eta| = 3.0$ and is divided into three sections:

- Barrel (EB): Crystals are organized into supermodules, covering up to $|\eta| = 1.479$.

- Endcaps (EE): Extend the coverage up to $|\eta| = 3.0$.

- Preshower (ES): Installed in front of the endcaps, it helps to distinguish photons originating from π^0 decays.

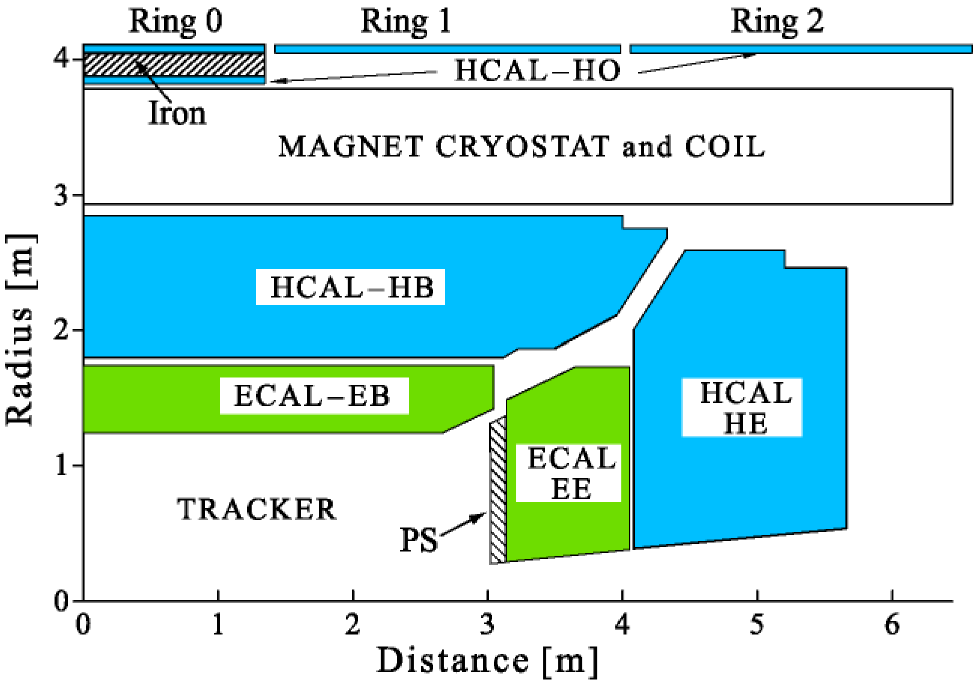

Figure 3 – Location of the ECAL and HCAL detectors in a quarter longitudinal cross-section of CMS. Source: [3]

3.2.4 Hadronic Calorimeter (HCAL)

The HCAL 3 detects hadronic particles and provides full coverage up to $|\eta| = 5.0$ [19]. It is divided into four subdetectors:

- HB (Barrel): Located inside the solenoid.
- HE (Endcaps): Complements the barrel in the $|\eta| > 1.4$ regions.
- HF (Forward): Detects particles in the most forward region of the detector, $|\eta| > 3.0$.
- HO (Outer): Located outside the solenoid to capture residual energy.

3.2.5 Muon System

Due to their high penetration power, muons are detected in the outermost part of the CMS detector. The muon system consists of three types of gaseous detectors 4:

- Drift Tubes (DT): Located in the barrel region, providing excellent spatial resolution.
- Cathode Strip Chambers (CSC): Used in the endcaps, offering precise three-dimensional reconstruction.
- Resistive Plate Chambers (RPC): With high time resolution, used primarily for triggering in the Level-1 system.

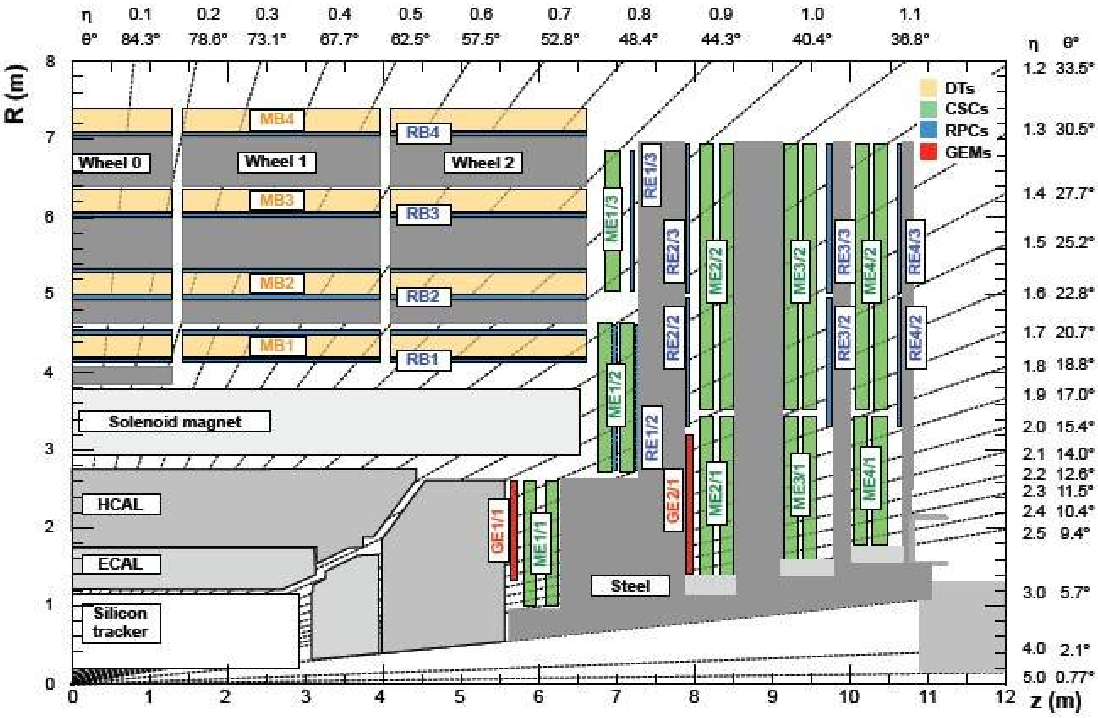

Figure 4 – Schematic of the CMS subdetectors highlighting the muon detectors (DTs, CSCs, RPCs, and GEMs) in a longitudinal view. Source: [4].

3.2.6 Trigger and Data Acquisition System

Due to the LHC's 40 MHz collision rate, CMS generates up to 40 TB of data per second. To manage this volume, the trigger system performs event selection in two levels:

- Level-1 Trigger (L1): Implemented in hardware, it reduces the event rate to 100 kHz.
- High-Level Trigger (HLT): Implemented in software, it applies complex algorithms and reduces the rate to 100 events per second, storing only the most relevant events for analysis.

3.2.7 Data Acquisition System (DAQ)

The CMS Data Acquisition (DAQ) system is responsible for collecting, processing, and storing the events selected by the trigger system, ensuring that only the most relevant data are retained for further analysis. Each CMS subsystem has readout electronics that record the information of an individual event. This information is initially stored in raw format (RAW) and goes through several processing stages until it reaches the *mini-AOD* format[2], which is widely used by the particle physics community.

[2] The Mini Analysis Object Data is a reduced and optimized data format developed by the CMS collaboration. It contains the essential information for most physics analyses, such as reconstructed particles, vertices, and associated variables.

3.2.8 Event Reconstruction

Event reconstruction in the CMS experiment combines tracking information from the tracker, energy deposits in the calorimeters, and signals from the muon detectors. Using the *Particle Flow* (PF) algorithm [20], individual particles such as muons, electrons, photons, and charged and neutral hadrons are identified. From these reconstructed particles, composite objects such as jets, tau leptons, and missing transverse momentum (p_T^{miss}) are computed 5.

The p_T^{miss}, defined as the negative vector sum of the transverse momenta of all reconstructed particles, is crucial for detecting neutral particles that escape the detector without interacting, such as neutrinos, and for identifying possible signs of physics beyond the Standard Model.

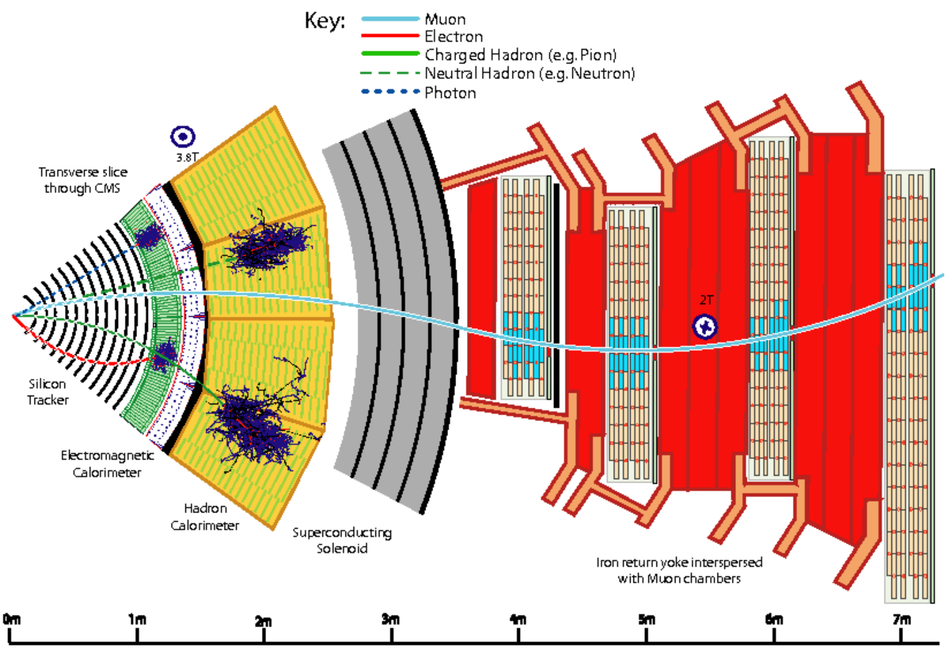

Figure 5 – Transverse cross-section of the CMS detector illustrating event reconstruction. Source: [5].

4 Results and Discussion

To determine the invisible width of the Z boson, we analyze signal events of the type $Z(\to \nu\nu) + \text{jets}$, characterized by jets recoiling against an invisible system (p_T^{miss}). Figure 6 illustrates this type of process with a simplified Feynman diagram. Additionally, control regions are defined to estimate the contributions from processes with signatures similar to the signal.

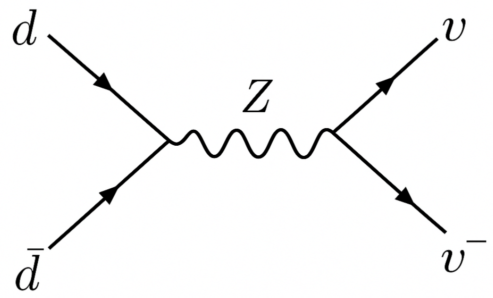

Figure 6 – **Feynman diagram** illustrating a typical process in proton–proton (pp) collisions, in which a down quark d and an anti-down quark $\bar{d}$, originating from the incident protons, annihilate to produce a Z boson. The Z boson then decays invisibly into a neutrino–antineutrino pair ($\nu\bar{\nu}$). Although not depicted in this diagram, hadronic jets ($Z + \text{jets}$) usually accompany the process, produced via gluon radiation (QCD) emitted by the initial-state quarks. The presence of jets, combined with the missing transverse momentum (p_T^{miss}) carried away by the neutrinos, characterizes the experimental signature of this process. *Source: author.*

The decay of the Z boson into charged leptons ($Z \to \ell^+\ell^-$) is used as a reference signal to compute the invisible width of the Z, while the decay of the W boson into a lepton and a neutrino ($W \to \ell\nu$) constitutes the main background.

The following control regions are defined:

- Single Muon, Single Electron, and Single Tau: used to estimate the $W \to \ell\nu$ background.
- Double Muon and Double Electron: used to estimate the reference signal $Z \to \ell^+\ell^-$.
- QCD Region: used to estimate the multijet QCD background.

These control regions are used to compute the *transfer factors*, which quantify the relationship between the observed events in the control regions and the expected number of events in the signal region. These factors allow the background contributions to be extrapolated to the signal region based on observations in the control regions.

Finally, a simultaneous fit is performed, combining information from both the control and signal regions. This procedure enables a precise determination of the invisible width of the Z boson.

4.1 Datasets

The .root files used in this analysis were obtained from the CERN Open Data portal, available at: <[https://opendata.cern.ch>. The data correspond to proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$, recorded by the CMS experiment during 2016.

Only data from eras G and H (specific data-taking periods) are available on the CERN Open Data portal, representing a subset of the total data collected that year. These eras correspond to run numbers in the range 278820–284044. The data are provided in the extttNanoAOD format, a compact and optimized structure that includes only high-level physics objects (such as muons, electrons, photons, jets, MET, etc.). This format is ideal for large-scale analyses, allowing for efficient data reading and seamless use with tools such as extttuproot [21] and extttawkward-array [22].

The proton-proton collision datasets selected for this analysis include: extitSingle Muon (which ensures the presence of at least one muon in the event), extitSingle Electron (at least one electron), and MET (*Missing Transverse Energy*), as shown in Table 2. Table 1 presents the control regions defined for each dataset.

Table 1 – Primary datasets used for the selection of different analysis regions.

Primary Datasets	**Regions**
MET	Jets + p_{miss} Single Muon Double Muon Single Tau QCD
Single Muon	Single Muon Double Muon
Single Electron	Single Electron Double Electron

During CMS data-taking, run quality information is recorded in a JSON-format file. This file allows analysts to select only the events from validated runs, ensuring the integrity and reliability of the data used in the analysis.

To certify the data used in this study, the JSON file `Cert_278820-284044_13TeV_Legacy2016_Collisions16GH_JSON.txt` was employed. This file was pre-edited to include only information corresponding to eras G and H.

In addition to real data, this analysis also employed Monte Carlo (MC) simulations to model signal and background processes, available from the CERN Open Data portal. The following processes were included:

- Signal: $Z \to \nu\bar{\nu}$ (Table 3),
- Reference signal: $Z \to \ell^+\ell^-$ (Table 3),
- Main background: $W +$ jets (Table 4),
- Backgrounds: Diboson, $t\bar{t}$ (Table 5),
- QCD multijet (Table 5),
- Less relevant backgrounds: VBS, single top (Table 5).

The Monte Carlo (MC) samples used in this analysis were generated using the `MadGraph5_aMC@NLO` event generator[1], interfaced with `Pythia8` for hadronization. The `CP5` tune (`TuneCP5_13TeV`) was employed, developed using LHC collision data at 13 TeV and incorporating the NNPDF3.1 NNLO parton distribution functions (PDFs) [23]. In contrast, the samples used in the CMS-published analysis [8] were generated using the `CUETP8M1_13TeV` tune, based on NNPDF2.3 LO PDFs and calibrated with 7 TeV data [23]. The `CUETP8M1` tune was widely used by the CMS collaboration during the early stages of Run 2.

Both tunes are validated and recognized by the CMS collaboration. However, differences—particularly in the PDF sets, treatment of multiple interactions, and radiation modeling [23]—may result in subtle variations in kinematic distributions and event rates, although these typically remain within the systematic uncertainties assigned to theoretical modeling.

The samples used in this analysis were generated as part of the `RunIISummer20UL16 - NanoAODv9-106X_mcRun2_asymptotic_v17-v2` simulation campaign, which belongs to the CMS Ultra Legacy reprocessing for the year 2016. This campaign represents the final complete reprocessing of simulated Run 2 data, incorporating all available corrections and updates for object reconstruction [24]. The `asymptotic` suffix indicates that the

[1] *MadGraph* is an event generator used in particle physics to simulate collisions based on any model that can be expressed in the form of a Lagrangian.

calibration settings and operating conditions reflect the ideal and stable performance of the detector throughout the data-taking period.

Regarding the signal Monte Carlo simulation, the `DYJetsToNuNu` dataset—commonly used in CMS analyses—is not available through the CERN Open Data portal. However, the `Z2JetsToNuNu` sample is provided, which also describes the $Z \to \nu\bar{\nu}$ process with accompanying jets. This sample was generated in *MadGraph* using the following process:

$$\texttt{generate} \quad \texttt{pp > } \nu \ \bar{\nu} \ \texttt{j j [QCD]}$$

The main difference between the two datasets lies in how additional jets are included. The `DYJetsToNuNu` sample is enriched with jets using the following *MadGraph* command sequence:

$$\begin{aligned} \texttt{generate} \quad & pp \to \nu\bar{\nu}\,[\mathrm{QCD}]\ @0, \\ \texttt{add process} \quad & pp \to \nu\bar{\nu}j\,[\mathrm{QCD}]\ @1, \\ \texttt{add process} \quad & pp \to \nu\bar{\nu}jj\,[\mathrm{QCD}]\ @2. \end{aligned}$$

In contrast, the `Z2JetsToNuNu` sample applies a cut on the invariant mass of the leptonic system (`lepton mass` > 50 GeV), which does not affect this analysis since it focuses on the mass region of the Z boson. Therefore, the `Z2JetsToNuNu` dataset was adopted as a suitable substitute for simulating the $Z \to \nu\bar{\nu}$ signal.

Cross section information for the MC samples is available on the corresponding `MINIAOD` dataset pages of the CERN Open Data portal. While this analysis uses `NANOAOD` files, cross section values are not provided for some datasets in the `NANOAOD` section of the portal.

Table 2 – List of primary datasets used in this analysis.

Dataset
/MET/Run2016G-UL2016_MiniAODv2_NanoAODv9-v1/NANOAOD /MET/Run2016H-UL2016_MiniAODv2_NanoAODv9-v1/NANOAOD
/SingleMuon/Run2016G-UL2016_MiniAODv2_NanoAODv9-v1/NANOAOD /SingleMuon/Run2016H-UL2016_MiniAODv2_NanoAODv9-v1/NANOAOD
/SingleElectron/Run2016G-UL2016_MiniAODv2_NanoAODv9-v1/NANOAOD /SingleElectron/Run2016H-UL2016_MiniAODv2_NanoAODv9-v1/NANOAOD

Table 3 – Simulated event samples for signal processes

Dataset	**Cross section [fb]**
/Z1JetsToNuNu_M-50_LHEFilterPtZ-50To150_[1]/[2]_v17-v1/[3]	577600
/Z1JetsToNuNu_M-50_LHEFilterPtZ-150To250_[1]/[2]_v17-v1/[3]	17410
/Z1JetsToNuNu_M-50_LHEFilterPtZ-250To400_[1]/[2]_v17-v1/[3]	1978
/Z1JetsToNuNu_M-50_LHEFilterPtZ-400ToInf_[1]/[2]_v17-v1/[3]	217.1
/Z2JetsToNuNu_M-50_LHEFilterPtZ-50To150_[1]/[2]_v17-v2/[3]	315100
/Z2JetsToNuNu_M-50_LHEFilterPtZ-150To250_[1]/[2]_v17-v2/[3]	28830
/Z2JetsToNuNu_M-50_LHEFilterPtZ-250To400_[1]/[2]_v17-v2/[3]	4976
/Z2JetsToNuNu_M-50_LHEFilterPtZ-400ToInf_[1]/[2]_v17-v2/[3]	815.1
/DYJetsToLL_LHEFilterPtZ-100To250_[1]/[2]_v17-v2/[3]	96620
/DYJetsToLL_LHEFilterPtZ-250To400_[1]/[2]_v17-v2/[3]	3738
/DYJetsToLL_LHEFilterPtZ-400To650_[1]/[2]_v17-v2/[3]	505
/DYJetsToLL_LHEFilterPtZ-650ToInf_[1]/[2]_v17-v2/[3]	47.63

[1] `_TuneCP5_13TeV-amcatnloFXFX-pythia8`

[2] `RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic`

[3] `NANOAODSIM`

Table 4 – Simulated event samples for dominant background processes

Dataset	**Cross section [fb]**
/WJetsToLNu_Pt-100To250_[1]/[2]_v17-v1/[3]	757800
/WJetsToLNu_Pt-250To400_[1]/[2]_v17-v1/[3]	27490
/WJetsToLNu_Pt-400To600_[1]/[2]_v17-v1/[3]	3501
/WJetsToLNu_Pt-600ToInf_[1]/[2]_v17-v1/[3]	542.7
/WZTo2Q2L_mllmin4p0_[1]/[2]_v17-v2/[3]	6429
/WZTo1L1Nu2Q_4f_[1]/[2]_v17-v1/[3]	9159
/WZTo1L3Nu_4f_[1]/[2]_v17-v1/[3]	3415
/WZTo3LNu_[1]/[2]_v17-v1/[3]	5218
/WWTo1L1Nu2Q_4f_[1]/[2]_v17-v1/[3]	50890
/WWTo2L2Nu_[1]/[2]_v17-v1/[3]	11090
/WWTo4Q_4f_[1]/[2]_v17-v1/[3]	51570
/ZZTo2Nu2Q_5f_[1]/[2]_v17-v1/[3]	4557
/ZZTo2L2Nu_[1]/[2]_v17-v1/[3]	973.8
/ZZTo4L_[1]/[2]_v17-v1/[3]	1325
/TTJets_[1]/[2]_v17-v1/[3]	756100

[1] `_TuneCP5_13TeV-amcatnloFXFX-pythia8`

[2] `RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic`

[3] `NANOAODSIM`

Table 5 – Simulated event samples for subdominant background processes

Dataset	**Cross section [fb]**
/QCD_Pt_15to30_[1]/[2]_v17-v1/[3]	1.245×10^{12}
/QCD_Pt_30to50_[1]/[2]_v17-v1/[3]	1.069×10^{11}
/QCD_Pt_50to80_[1]/[2]_v17-v1/[3]	1.571×10^{10}
/QCD_Pt_80to120_[1]/[2]_v17-v1/[3]	2.341×10^{9}
/QCD_Pt_120to170_[1]/[2]_v17-v1/[3]	4.074×10^{8}
/QCD_Pt_170to300_[1]/[2]_v17-v1/[3]	1.037×10^{8}
/QCD_Pt_300to470_[1]/[2]_v17-v1/[3]	6.832×10^{6}
/QCD_Pt_470to600_[1]/[2]_v17-v1/[3]	5.519×10^{5}
/QCD_Pt_600to800_[1]/[2]_v17-v1/[3]	1.565×10^{5}
/QCD_Pt_1000to1400_[1]/[2]_v17-v1/[3]	7.475×10^{3}
/QCD_Pt_1400to1800_[1]/[2]_v17-v1/[3]	6.482×10^{2}
/QCD_Pt_1800to2400_[1]/[2]_v17-v1/[3]	8.746×10^{1}
/QCD_Pt_2400to3200_[1]/[2]_v17-v1/[3]	5.235
/QCD_Pt_3200toInf_[1]/[2]_v17-v1/[3]	0.1352
/WGToLNuG_01J_5f_[1]/[2]_v17-v1/[3]	1.912×10^{5}
/ZGToLLG_01J_5f_[1]/[2]_v17-v1/[3]	5.144×10^{4}
/ZGTo2NuG_[1]/[2]_v17-v1/[3]	3.022×10^{4}
/ST_tW_antitop_5f_[1]/[2]_v17-v1/[3]	3.251×10^{4}
/ST_tW_top_5f_[1]/[2]_v17-v1/[3]	3.245×10^{4}
/ST_t-channel_antitop_4f_[1]/[2]_v17-v1/[3]	6.793×10^{4}
/ST_t-channel_top_4f_[1]/[2]_v17-v1/[3]	1.134×10^{5}
/EWKWMinus2Jets_[1]/[2]_v17-v1/[3]	3.210×10^{4}
/EWKWPlus2Jets_[1]/[2]_v17-v1/[3]	3.908×10^{4}
/EWKZ2Jets_ZToLL_[1]/[2]_v17-v1/[3]	6.206×10^{3}
/EWKZ2Jets_ZToNuNu_[1]/[2]_v17-v1/[3]	1.065×10^{4}

`[1] _TuneCP5_13TeV-amcatnloFXFX-pythia8`

`[2] RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic`

`[3] NANOAODSIM`

4.2 *Trigger* and Luminosity

For the selection of primary datasets, unprescaled *triggers* are used. The prescale of a trigger is a factor that determines how often the Data Acquisition System (DAQ) records events that satisfy certain selection criteria. When the prescale factor of a trigger is equal to 1, all events that meet the trigger criteria are recorded. Therefore, an unprescaled trigger is one for which the prescale factor is 1, ensuring that every qualifying event is stored for analysis. The prescale values for the *triggers* were obtained following the CMS Open Data instructions [25]. Table 6 lists the *triggers* used for each primary dataset.

Table 6 – Primary datasets and their corresponding *triggers*.

Primary Datasets	***Triggers***
MET	HLT_PFMET170_HBHECleaned OR HLT_PFMET170_HBHE_BeamHaloCleaned OR HLT_PFMETNoMu120_PFMHTNoMu120_IDTight OR HLT_MET75_IsoTrk50
Single Muon	HLT_IsoMu24 OR HLT_IsoTkMu24
Single Electron	HLT_Ele27_WPTight_Gsf

According to the CMS Open Data guidelines [26], using the aforementioned *triggers* and the BRILCalc tool[2], the integrated luminosity for the selected primary datasets was calculated to be $16.290\,\mathrm{fb}^{-1}$.

4.3 Event Selection

The goal of this analysis is to select events consistent with the production of a Z boson decaying invisibly, as well as to characterize the dominant background contributions. The typical signature of this signal consists of high-energy jets recoiling against the missing transverse momentum vector ($\vec{p}_T^{\,\text{miss}}$). To describe this kinematic configuration, the **recoil vector** is introduced, defined as:

$$\vec{U} = \vec{p}_T^{\,\text{miss}} + \sum_{\ell} \vec{p}_T^{\,\ell}$$

In events without charged leptons, $\vec{U}$ coincides with $\vec{p}_T^{\,\text{miss}}$, and thus $U = |\vec{p}_T^{\,\text{miss}}|$.

Event selection is guided by a baseline set of requirements valid across all analysis regions:

- $U > 200\,\text{GeV}$;
- The leading jet must have $p_T > 200\,\text{GeV}$;
- Events containing jets with $p_T > 40\,\text{GeV}$ and pseudorapidity $|\eta| > 2.4$ are vetoed to suppress vector boson scattering processes.

To ensure reconstruction quality, consistency between p_T^{miss} and U is required using the variable:

$$\Delta = \frac{|p_{T calo}^{\text{miss}} - U|}{U}$$

Events with $\Delta > 0.5$ are rejected. Additionally, the following vetoes are applied:

[2] BRILCalc is a tool developed by CMS to compute luminosity and monitor data quality in particle physics experiments, based on beam and *trigger* information.

- Isolated photons with $p_T > 25\,\text{GeV}$ and $|\eta| < 2.5$ are vetoed to suppress initial state radiation and multijet production;
- Jets identified as originating from b-quarks (b-tagged) are rejected using the CSVv2 discriminator with a medium working point (CSVv2 > 0.8), following [27];
- The minimum azimuthal separation between the four leading jets and $\vec{p}_T^{\text{miss}}$, denoted $\Delta\phi_{\text{min}}$, must satisfy $\Delta\phi_{\text{min}} > 0.5$ to reduce the multijet QCD background.

After applying the baseline criteria, the selected events are categorized into different analysis regions, each optimized to target specific signal signatures and control background contributions. The following outlines the selection criteria for each of these regions:

- **p_T^{miss} + jets region:** targets events without charged leptons, characteristic of invisible Z boson decays. The following vetoes are applied:
  - Isolated muons or electrons with $p_T > 10\,\text{GeV}$ and $|\eta| < 2.5$;
  - Taus with $p_T > 20\,\text{GeV}$ and $|\eta| < 2.3$.
- **Dilepton region ($\ell^+\ell^-$):** aims to identify events with two isolated leptons in the final state, consistent with $Z \to \ell^+\ell^-$ decays. Selection criteria:
  - A pair of isolated muons or electrons with invariant mass $71 < m_{\ell\ell} < 111\,\text{GeV}$;
  - Muons with $p_T > 25\,\text{GeV}$ and electrons with $p_T > 30\,\text{GeV}$.
- **μ+jets and e+jets regions:** select events with a single isolated lepton in the final state, compatible with $W \to \ell\nu$ decays.
  - **μ+jets region:** requires one isolated muon with $p_T > 25\,\text{GeV}$ and $|\eta| < 2.4$;
  - **e+jets region:**
    * One electron with $p_T > 30\,\text{GeV}$;
    * $p_T^{\text{miss}} > 100\,\text{GeV}$ to suppress QCD multijet background.
  - Both regions require transverse mass in the range $30 \leq M_T(\ell, p_T^{\text{miss}}) < 125\,\text{GeV}$.
- **Tau + jets region:** focuses on selecting events with a single isolated hadronically decaying tau, compatible with the process $W \to \tau\nu$.
  - One isolated tau with $p_T > 40\,\text{GeV}$, $|\eta| < 2.3$;
  - Transverse mass in the range $30 < m_T < 125\,\text{GeV}$.
- **QCD region:** used as a control region for multijet backgrounds, defined by:

– Inversion of the azimuthal angle cut between the four leading jets and p_T^{miss}:

$$\min\left[\Delta\phi(j_1, j_2, j_3, j_4, p_T^{\text{miss}})\right] \leq 0.5$$

Each of the described analysis regions includes, in addition to their specific criteria, a set of vetoes designed to suppress unwanted background processes. These cuts are implemented to ensure the purity of the selected sample and the robustness of the analysis. A complete description of all selection criteria, including vetoes and kinematic requirements for each region, is provided in Table 20.

4.3.1 Cut Flow

The following Tables present the cut flow for the events selected in each analysis region.

Table 7 – Cut flow for the p_T^{miss} region – Part 1

Cuts	**Z2JetsToNuNu**	**Z1JetsToNuNu**	**WJetsToLNu**	**WZTo1L3Nu**	**WWTo2L2Nu**
All events	98,474,753	35,910,730	273,949,797	1,229,946	2,900,000
Luminosity	98,474,753	35,910,730	273,949,797	1,229,946	2,900,000
Trigger	84,053,704	27,049,364	99,232,003	136,417	183,010
$U > 200\,\text{GeV}$	52,424,077	13,140,991	62,785,171	61,743	82,492
$\frac{\lvert\text{calo\_MET}-U\rvert}{U} < 0.5$	51,495,863	12,993,015	27,800,151	41,376	28,778
$N_{\text{jets}} \geq 4$	37,037,380	4,072,441	21,917,439	30,120	23,803
Leading jet $p_T > 200\,\text{GeV}$	24,765,355	3,039,123	15,334,976	17,981	12,974
Photon veto	17,711,990	2,586,726	8,238,960	9,660	5,179
b-jet veto	15,311,766	2,356,936	7,183,083	8,551	4,510
Jet $\lvert\eta\rvert > 2.4$ veto	2,895,832	208,024	1,557,071	1,651	1,094
Muon veto	2,845,723	205,413	908,096	870	313
Electron veto	2,322,725	189,616	600,304	565	149
Tau veto	2,093,093	181,799	341,316	340	61
$\Delta\phi(j_1, j_2, j_3, j_4, U) > 0.5$	2,093,093	181,799	341,316	340	61

Table 8 – Cut flow for the p_T^{miss} region – Part 2

Cuts	**ZZTo2L2Nu**	**ZGTo2NuG**	**EWKZ2Jets**	**Total MC**	**MET**	**DATA/MC**
All events	15,928,000	1,183,548	1,500,000	431,076,774	66,747,616	0.154
Luminosity mask	15,928,000	1,183,548	1,500,000	431,076,774	63,250,532	0.146
Trigger selection	592,175	54,959	159,654	211,461,286	26,666,974	0.126
$U\, p_T > 200$	194,712	14,559	54,644	128,758,389	4,330,577	0.033
$(\lvert \text{calo\_MET} - U \rvert / U) < 0.5$	106,156	14,074	52,694	92,532,107	2,057,038	0.022
$N_{\text{jets}} \geq 4$	80,824	8,810	34,665	63,205,482	1,456,182	0.023
Leading jet $p_T > 200\,\text{GeV}$	26,935	4,167	18,893	43,220,404	677,544	0.015
Photon veto	12,741	1,562	12,534	28,579,352	370,927	0.012
b-jet veto	10,967	1,399	10,727	24,887,939	261,535	0.010
Jet veto $\lvert\eta\rvert > 2.4$	2,497	241	1,382	4,667,792	45,595	0.009
Muon veto	996	238	1,351	3,963,000	33,770	0.008
Electron veto	648	178	1,086	3,115,271	24,954	0.008
Tau veto	276	167	878	2,617,930	19,598	0.007
$\Delta\phi(j_{1,2,3,4}, U) > 0.5$	276	167	878	2,617,930	19,598	0.007

Table 9 – Cut flow for the single muon region – Part 1

Cut	**WJetsToLNu**	**WZTo1L1Nu2Q**	**WZTo1L3Nu**	**WZTo3LNu**	**WWTo1L1Nu2Q**	**WWTo2L2Nu**	**ZZTo2L2Nu**	**TTJets**	**WGToLNuG**
All events	273,949,797	3,690,271	1,229,946	10,441,724	19,976,139	2,900,000	15,928,000	89,003,534	55,939,475
Luminosity mask	273,949,797	3,690,271	1,229,946	10,441,724	19,976,139	2,900,000	15,928,000	89,003,534	55,939,475
Trigger selection	119,805,807	867,035	336,468	2,738,145	4,438,327	1,050,841	3,702,528	16,594,784	6,485,523
$U > 200\,\text{GeV}$	69,560,780	191,522	67,594	414,892	816,164	106,698	390,411	4,653,907	866,112
$\frac{\lvert \text{calo\_MET} - U \rvert}{U} < 0.5$	28,305,602	63,536	43,346	88,832	245,985	31,171	208,463	1,339,763	199,339
$N_{\text{jets}} \geq 4$	22,327,618	58,657	31,227	79,868	221,142	25,455	151,944	1,332,262	162,354
Leading jet $p_T > 200\,\text{GeV}$	15,649,273	31,295	18,004	30,755	107,177	13,047	38,984	599,563	92,230
Photon veto	8,407,480	14,401	9,674	11,413	50,784	5,202	18,490	206,877	20,984
b-jet veto	7,329,936	12,049	8,565	9,742	42,215	4,528	15,933	32,970	18,324
Jet veto $\lvert\eta\rvert > 2.4$	1,589,041	3,286	1,655	2,480	11,164	1,098	3,634	7,540	4,443
Single muon	676,892	1,427	753	720	4,903	450	851	3,007	1,930
Transverse mass	447,179	960	390	426	3,269	246	435	1,899	1,242
Electron veto	359,297	710	315	257	2,480	159	317	1,076	899
Tau veto	104,835	219	98	53	685	21	105	307	159
$\Delta\phi(j_{1,2,3,4}, U) > 0.5$	104,835	219	98	53	685	21	105	307	159

Table 10 – Cut flow for the single muon region – Part 2

Cut	**EWKWMinus**	**EWKWPlus**	**ST_tW_antitop**	**ST_tW_top**	**ST_t_top**	**ST_t_antitop**	**Total MC**	**Single Muon**	**MET**	**DATA/MC**
All events	2,202,000	2,033,000	840,000	2,491,000	25,457,000	12,557,000	518,638,886	323,952,013	66,747,616	0.753
Luminosity mask	2,202,000	2,033,000	840,000	2,491,000	25,457,000	12,557,000	518,638,886	318,036,559	63,250,532	0.735
Trigger selection	526,954	508,895	152,809	453,515	2,122,523	1,076,464	160,860,618	211,419,030	27,259,274	1.483
$U > 200$	120,908	127,545	40,649	122,154	252,236	109,251	77,840,823	4,575,084	4,461,942	0.116
$\frac{\lvert \text{calo\_MET} - U \rvert}{U} < 0.5$	35,205	47,168	10,137	30,522	83,385	30,959	30,763,413	249,061	2,062,626	0.075
$N_{\text{jets}} \geq 4$	29,114	38,681	9,801	29,527	75,092	27,803	24,600,545	200,493	1,461,404	0.067
Leading jet $p_T > 200\,\text{GeV}$	17,323	23,900	4,538	13,837	40,012	15,200	16,695,138	129,210	678,981	0.048
Photon veto	7,837	11,117	1,672	5,161	19,891	7,669	8,798,652	73,705	371,649	0.050
b-jet veto	5,121	7,572	427	1,249	5,168	1,983	7,495,782	42,406	261,958	0.040
Jet veto $\lvert\eta\rvert > 2.4$	912	1,178	96	328	961	364	1,628,180	8,400	45,679	0.033
Single muon	418	509	34	145	432	175	692,646	4,725	14,195	0.027
Transverse mass	268	343	21	91	279	129	457,177	2,822	7,549	0.022
Electron veto	203	269	16	58	212	98	366,366	2,318	5,871	0.022
Tau veto	54	51	2	17	70	22	106,698	287	2,937	0.030
$\Delta\phi(j_{1,2,3,4}, U) > 0.5$	54	51	2	17	70	22	106,698	287	2,937	0.030

Table 11 – Cut flow for the single electron region – Part 1

Cut	WJetsToLNu	WZTo1L1Nu2Q	WZTo1L3Nu	WZTo3LNu	WWTo1L1Nu2Q	WWTo2L2Nu	ZZTo2L2Nu	TTJets	WGToLNuG
All events	273,949,797	3,690,271	1,229,946	10,441,724	19,976,139	2,900,000	15,928,000	89,003,534	55,939,475
Luminosity	273,949,797	3,690,271	1,229,946	10,441,724	19,976,139	2,900,000	15,928,000	89,003,534	55,939,475
Trigger selection	143,142,995	785,836	277,421	2,216,313	8,906,344	823,242	2,945,541	13,545,655	5,187,209
$U > 200\,\mathrm{GeV}$	89,518,570	247,978	73,194	486,032	1,072,498	122,364	382,927	4,936,168	1,130,876
$\frac{\lvert\mathrm{calo\_MET}-U\rvert}{U} < 0.5$	28,617,456	64,293	42,033	79,563	248,451	30,119	160,069	1,148,117	201,915
$N_{\mathrm{jets}} \geq 4$	22,576,163	59,378	30,579	71,966	223,494	24,900	121,979	1,141,891	164,716
Leading jet $p_T > 200\,\mathrm{GeV}$	15,755,914	31,682	18,048	30,979	108,216	13,147	39,088	515,507	93,175
Photon veto	8,419,151	14,350	9,662	11,301	50,654	5,181	18,421	173,612	20,938
b-jet veto	7,340,523	12,002	8,553	9,658	42,106	4,512	15,878	27,777	18,285
Jet veto $\lvert\eta\rvert > 2.4$	1,590,694	3,274	1,651	2,455	11,142	1,094	3,623	6,360	4,431
Single Electron	499,546	1,142	514	812	3,734	389	1,004	2,318	1,489
$p_T^{\mathrm{miss}} > 100$	498,161	1,135	514	808	3,703	387	1,002	2,297	1,483
Transverse mass	310,964	702	321	501	2,249	248	572	1,460	841
Muon veto	213,518	446	195	156	1,471	86	290	687	516
Tau veto	125,336	252	119	45	777	36	116	343	256
$\Delta\phi(j_{1,2,3,4}, U) > 0.5$	125,336	252	119	45	777	36	116	343	256

Table 12 – Cut flow for the single electron region – Part 2

Cut	EWKWMinus	EWKWPlus	ST_tW_antitop	ST_tW_top	ST_t_top	ST_t_antitop	Total MC	Single Electron	Data/MC
All events	2,202,000	2,033,000	840,000	2,491,000	25,457,000	12,557,000	518,638,886	282,385,002	0.544
Luminosity Mask	2,202,000	2,033,000	840,000	2,491,000	25,457,000	12,557,000	518,638,886	279,642,412	0.539
Trigger selection	479,060	462,004	147,156	436,522	1,523,264	760,800	181,639,362	179,699,658	0.989
$U\,p_T > 200$	159,399	160,905	53,159	158,217	288,563	126,113	98,916,963	4,121,552	0.041
$\frac{\lvert\mathrm{calo\_MET}-U\rvert}{U} < 0.5$	35,650	47,725	10,172	30,525	84,633	31,499	30,832,220	132,195	0.004
$N_{\mathrm{jets}} \geq 4$	29,488	39,165	9,859	29,599	76,273	28,332	24,627,782	116,421	0.004
Leading jet $p_T > 200\,\mathrm{GeV}$	17,510	24,138	4,598	14,015	40,721	15,460	16,722,198	60,488	0.003
Photon veto	7,814	11,088	1,662	5,121	19,751	7,598	8,776,304	3,113	0.0003
b-jet veto	5,110	7,546	423	1,241	5,145	1,968	7,500,727	1,880	0.0002
Jet veto $\lvert\eta\rvert > 2.4$	908	1,172	93	326	959	363	1,628,545	391	0.0002
Single Electron	298	373	29	122	300	117	512,187	268	0.0005
$p_T^{\mathrm{miss}} > 100$	297	371	28	122	298	116	510,722	263	0.0005
Transverse mass	179	228	18	71	179	79	318,612	176	0.0005
Muon veto	119	137	10	35	114	43	217,823	170	0.0007
Tau veto	51	71	3	14	66	29	127,514	58	0.0004
$\Delta\phi(j_{1,2,3,4}, U) > 0.5$	51	71	3	14	66	29	127,514	58	0.0004

Table 13 – Cut flow for the single tau region – Part 1

Cut	WJetsToLNu	WZ1L1Nu2Q	WZ1L3Nu	WZ3LNu	WW1L1Nu2Q	WW2L2Nu	ZZ2L2Nu	WGToLNuG	TTJets
All events	273,949,797	3,690,271	1,229,946	10,441,724	19,976,139	2,900,000	15,928,000	55,939,475	89,003,534
Luminosity Mask	273,949,797	3,690,271	1,229,946	10,441,724	19,976,139	2,900,000	15,928,000	55,939,475	89,003,534
Trigger selection	100,402,976	307,302	136,417	586,746	1,307,094	183,010	859,797	1,296,937	7,988,708
$U\,p_T > 200$	63,841,036	163,317	61,743	296,565	681,836	82,492	282,479	705,768	3,933,392
$\frac{\lvert\mathrm{calo\_MET}-U\rvert}{U} < 0.5$	28,396,478	62,927	41,376	74,198	243,390	28,778	153,826	195,869	1,334,158
$N_{\mathrm{jets}} \geq 4$	22,389,086	58,085	30,120	66,970	218,776	23,803	116,942	159,676	1,326,863
Leading jet $p_T > 200\,\mathrm{GeV}$	15,694,967	31,191	17,981	30,404	106,874	12,974	38,871	91,978	606,228
Photon veto	8,435,439	14,348	9,660	11,299	50,648	5,179	18,420	20,932	209,414
b-jet veto	7,354,928	12,001	8,551	9,656	42,102	4,510	15,877	18,280	33,476
Jet veto $\lvert\eta\rvert > 2.4$	1,593,793	3,274	1,651	2,455	11,140	1,094	3,623	4,429	7,667
Isolated tau selection	920,921	1,690	998	892	6,072	520	1,388	2,822	3,735
Transverse mass	558,543	1,033	495	486	3,605	286	508	1,718	2,262
Electron veto	421,064	711	371	324	2,563	181	353	1,184	1,352
Muon veto	111,192	217	97	73	692	31	135	295	326
$\Delta\phi(j_{1,2,3,4}, U) > 0.5$	111,192	217	97	73	692	31	135	295	326

Table 14 – Cut flow for the single tau region – Part 2

Cut	**TTJets**	**EWKWMinus**	**EWKWPlus**	**ST_tW_antitop**	**ST_tW_top**	**ST_t_top**	**ST_antitop**	**Total MC**	**MET**	**Data/MC**
All events	89,003,534	2,202,000	2,033,000	840,000	2,491,000	25,457,000	12,557,000	518,638,886	66,747,616	0.128
Luminosity Mask	89,003,534	2,202,000	2,033,000	840,000	2,491,000	25,457,000	12,557,000	518,638,886	63,250,532	0.121
Trigger selection	7,988,708	171,579	189,620	63,580	189,577	460,209	205,924	114,349,476	26,666,974	0.233
$U\,p_T > 200$	3,933,392	104,125	112,339	33,803	101,092	193,225	81,672	70,674,884	4,330,577	0.061
$\frac{\lvert \text{calo\_MET}-U \rvert}{U} < 0.5$	1,334,158	34,761	46,657	9,789	29,418	80,579	30,082	30,762,286	2,057,038	0.066
$N_{\text{jets}} \geq 4$	1,326,863	28,742	38,264	9,494	28,517	72,582	27,023	24,594,943	1,456,182	0.059
Leading jet $p_T > 200\,\text{GeV}$	606,228	17,272	23,837	4,499	13,730	39,732	15,060	16,745,598	677,544	0.040
Photon veto	209,414	7,812	11,087	1,662	5,120	19,750	7,598	8,828,368	370,927	0.042
b-jet veto	33,476	5,109	7,545	423	1,241	5,145	1,968	7,520,812	261,535	0.034
Jet veto $\lvert\eta\rvert > 2.4$	7,667	908	1,172	93	326	959	363	1,632,947	45,595	0.027
Isolated tau	3,735	544	695	51	169	520	205	941,222	14,564	0.015
Transverse mass	2,262	307	411	35	88	362	148	570,287	7,637	0.013
Electron veto	1,352	221	310	21	59	256	114	429,084	5,593	0.013
Muon veto	326	53	64	7	12	41	18	113,253	1,722	0.015
$\Delta\phi(j_{1,2,3,4}, U) > 0.5$	326	53	64	7	12	41	18	113,253	1,722	0.015

Table 15 – Cut flow for the QCD region – Part 1

Cut	**QCD_Pt15–30**	**QCD_Pt30–50**	**QCD_Pt50–80**	**QCD_Pt80–120**	**QCD_Pt120–170**	**QCD_Pt170–300**	**QCD_Pt300–470**	**QCD_Pt470–600**	**QCD_Pt600–800**
All events	19,298,000	19,036,000	19,776,000	29,938,000	27,348,000	29,758,000	55,264,000	52,408,000	64,584,000
Luminosity Mask	19,298,000	19,036,000	19,776,000	29,938,000	27,348,000	29,758,000	55,264,000	52,408,000	64,584,000
Trigger selection	24	21	54	373	3,395	38,485	516,147	1,329,877	2,886,144
$U\,p_T > 200$	3	0	1	32	497	8,553	203,486	689,407	1,610,696
$\frac{\lvert \text{calo\_MET}-U \rvert}{U} < 0.5$	2	0	0	5	68	2,969	96,511	350,346	807,511
$N_{\text{jets}} \geq 4$	2	0	0	3	58	2,481	81,928	306,042	718,128
Leading jet $p_T > 200\,\text{GeV}$	0	0	0	2	3	1,500	73,536	291,751	692,426
Photon veto	0	0	0	0	2	796	41,030	166,291	400,050
b-jet veto	0	0	0	0	1	451	25,188	107,821	267,298
Jet veto $\lvert\eta\rvert > 2.4$	0	0	0	0	0	92	6,133	28,493	73,958
Muon veto	0	0	0	0	0	58	4,758	23,402	61,607
Electron veto	0	0	0	0	0	39	3,425	17,676	47,756
Tau veto	0	0	0	0	0	31	2,911	16,056	43,692
$\Delta\phi(j_{1,2,3,4}, U) \leq 0.5$	0	0	0	0	0	0	0	0	0

Table 16 – Cut flow for the QCD region – Part 2

Cut	**QCD_Pt800–1000**	**QCD_Pt1000–1400**	**QCD_Pt1400–1800**	**QCD_Pt1800–2400**	**QCD_Pt2400–3200**	**QCD_Pt3200–Inf**	**Total MC**	**MET**	**DATA/MC**
All events	37,698,000	19,892,000	10,722,000	5,236,000	2,848,000	996,000	394,802,000	66,747,616	0.169
Luminosity Mask	37,698,000	19,892,000	10,722,000	5,236,000	2,848,000	996,000	394,802,000	63,250,532	0.160
Trigger selection	2,939,925	2,525,004	2,286,684	1,540,684	1,104,619	483,667	15,655,103	26,666,974	1.703
$U\,p_T > 200$	1,698,392	1,532,194	1,530,432	1,129,104	879,340	406,701	9,688,838	4,330,577	0.446
$\frac{\lvert \text{calo\_MET}-U \rvert}{U} < 0.5$	797,698	654,181	570,680	372,986	255,871	110,125	4,018,953	2,057,038	0.511
$N_{\text{jets}} \geq 4$	720,288	594,866	519,472	337,432	228,593	96,803	3,606,096	1,456,182	0.403
Leading jet $p_T > 200\,\text{GeV}$	695,101	571,330	494,860	318,708	211,878	86,303	3,437,398	677,544	0.197
Photon veto	410,920	346,905	310,945	205,407	140,453	58,948	2,081,747	370,927	0.178
b-jet veto	281,258	242,692	220,723	144,998	97,557	39,522	1,427,509	261,535	0.183
Jet veto $\lvert\eta\rvert > 2.4$	79,983	70,930	67,071	45,707	32,150	14,068	418,585	45,595	0.108
Muon veto	67,588	60,701	57,946	39,473	27,256	11,738	354,527	33,770	0.095
Electron veto	53,393	48,599	47,249	32,530	22,690	9,896	283,253	24,954	0.088
Tau veto	49,003	44,509	42,588	28,841	19,956	8,717	256,304	19,598	0.076
$\Delta\phi(j_{1,2,3,4}, U) \leq 0.5$	0	0	0	0	0	0	0	0	0

Table 17 – Cut flow for the double muon region

Cut	**DYJetsToLL_PtZ**	**WZTo2Q2L**	**ZZTo4L**	**ZGToLLG**	**EWKZ2Jets**	**Total MC**	**Single Muon**	**Double Muon**	**MET**	**DATA/MC**
All events	55,104,205	13,526,954	52,104,000	31,562,465	453,000	152,750,624	323,952,013	94,148,416	66,747,616	3.174
Luminosity Mask	55,104,205	13,526,954	52,104,000	31,562,465	453,000	152,750,624	319,828,002	92,825,116	63,250,532	3.115
Trigger selection	21,847,293	2,583,629	10,020,776	8,584,899	130,892	43,167,489	213,210,473	19,107,239	27,259,274	6.013
$U > 200\,\mathrm{GeV}$	17,408,037	608,493	505,208	465,248	27,484	19,014,470	4,575,084	894,326	4,461,942	0.522
$\frac{\lvert\mathrm{calo\_MET}-U\rvert}{U} < 0.5$	1,649,929	32,493	35,304	17,691	1,392	1,736,809	249,061	57,209	2,062,626	1.374
$N_{\mathrm{jets}} \geq 4$	1,430,675	30,883	31,553	15,214	1,235	1,509,560	221,796	53,991	1,461,404	1.150
Leading jet $p_T > 200\,\mathrm{GeV}$	1,085,163	17,205	6,790	7,440	675	1,117,273	129,210	37,906	678,981	0.757
Photon veto	491,958	7,211	2,361	1,701	323	503,554	73,715	20,648	371,649	0.925
b-jet veto	397,902	5,613	1,942	1,418	246	407,121	42,416	8,059	261,958	0.767
Jet veto $\lvert\eta\rvert > 2.4$	96,877	1,485	493	356	43	99,254	8,400	1,650	45,679	0.561
Number of muons = 2	14,321	267	139	48	7	14,782	2,155	613	5,965	0.590
Double Muon selection	2,632	35	35	7	1	2,710	534	183	982	0.626
Invariant mass window	640	14	23	1	1	679	79	36	97	0.312
Electron veto	506	10	13	0	0	529	65	27	79	0.323
Tau veto	27	0	0	0	0	27	4	3	17	0.888
$\Delta\phi(j_{1,2,3,4}, U) > 0.5$	27	0	0	0	0	27	4	3	17	0.888

Table 18 – Cut flow for the double electron region

Cut	**DYJetsToLL**	**WZTo2Q2L**	**ZZTo4L**	**ZGToLLG**	**EWKZ2Jets**	**Total MC**	**Single Electron**	**DATA/MC**
All events	55,104,205	13,526,954	52,104,000	31,562,465	453,000	152,750,624	282,385,002	1.848
Luminosity mask	55,104,205	13,526,954	52,104,000	31,562,465	453,000	152,750,624	279,647,775	1.830
Trigger selection	32,806,331	2,770,413	7,455,388	6,674,844	130,187	49,837,163	180,991,183	3.631
$U > 200\,\mathrm{GeV}$	26,999,884	964,306	592,054	623,973	40,859	29,221,076	4,086,293	0.139
$\frac{\lvert\mathrm{calo\_MET}-U\rvert}{U} < 0.5$	1,667,219	35,086	29,048	17,297	1,409	1,750,059	131,019	0.074
$N_{\mathrm{jets}} \geq 4$	1,445,834	33,373	26,331	14,913	1,253	1,521,704	115,392	0.075
Leading jet $p_T > 200\,\mathrm{GeV}$	1,096,793	18,826	6,819	7,526	696	1,130,660	59,988	0.053
Photon veto	492,566	7,640	2,305	1,682	319	504,512	3,098	0.006
b-jet veto	398,507	5,945	1,898	1,403	243	407,996	1,890	0.004
Jet veto $\lvert\eta\rvert > 2.4$	96,960	1,554	480	351	43	99,388	381	0.003
Number of electrons = 2	8,249	181	65	34	2	8,531	82	0.009
Double Electron selection	53	1	0	0	0	54	0	0
Invariant mass window	16	0	0	0	0	16	0	0
Muon veto	7	0	0	0	0	7	0	0
Tau veto	0	0	0	0	0	0	0	0
$\Delta\phi(j_{1,2,3,4}, U) > 0.5$	0	0	0	0	0	0	0	0

After applying the event selection criteria (cuts), a significant reduction in event statistics is observed in the final stages of the cut flow, particularly in regions with two leptons in the final state. This statistical limitation arises from the fact that, for the real proton-proton collision data used in this analysis, only the datasets corresponding to **eras G and H** of the year 2016 are available through the CMS Open Data portal. As a result, it was not possible to use the full dataset collected in 2016—only a fraction—thereby restricting the number of events available for analysis. Additionally, the Monte Carlo samples employed, although representative, do not exactly match those used in the official CMS publication [8], which may impact the accuracy of background process estimations.

Table 19 – Baseline selection and criteria for each analysis region.

Region	**Selection Criteria**
Baseline	$U > 200\,\text{GeV}$ $\lvert p_{Tcalo}^{\text{miss}} - U\rvert / U < 0.5$ Leading jet $p_T > 200\,\text{GeV}, \lvert\eta\rvert < 2.4, 0.1 <$ Ch. Had. EF < 0.95 Veto jets with $p_T > 40\,\text{GeV}, \lvert\eta\rvert > 2.4$ Photon veto $p_T > 25\,\text{GeV}, \lvert\eta\rvert < 2.5$ Medium CSVv2 b-jet veto $p_T > 40\,\text{GeV}, \lvert\eta\rvert < 2.4$
Jets + p_T^{miss}	Baseline Muon veto $p_T > 10\,\text{GeV}, \lvert\eta\rvert < 2.5$ Electron veto $p_T > 10\,\text{GeV}, \lvert\eta\rvert < 2.5$ Tau veto $p_T > 20\,\text{GeV}, \lvert\eta\rvert < 2.3$ $\min[\Delta\phi(j_{1,2,3,4}, U)] > 0.5$
Single Muon	Baseline 1 muon $p_T > 25\,\text{GeV}, \lvert\eta\rvert < 2.4$ Electron veto $p_T > 10\,\text{GeV}, \lvert\eta\rvert < 2.5$ Tau veto $p_T > 20\,\text{GeV}, \lvert\eta\rvert < 2.3$ $30 \leq M_T(\mu, p_T^{\text{miss}}) < 125\,\text{GeV}$ $\min[\Delta\phi(j_{1,2,3,4}, U)] > 0.5$
Double Muon	Baseline 2 muons $p_T > 25\,\text{GeV}, \lvert\eta\rvert < 2.4$ Electron veto $p_T > 10\,\text{GeV}, \lvert\eta\rvert < 2.5$ Tau veto $p_T > 20\,\text{GeV}, \lvert\eta\rvert < 2.3$ $71 < M_{\mu\mu} < 111\,\text{GeV}$ $\min[\Delta\phi(j_{1,2,3,4}, U)] > 0.5$
Single Electron	Baseline 1 electron $p_T > 30\,\text{GeV}, \lvert\eta\rvert < 2.4$ Muon veto $p_T > 10\,\text{GeV}, \lvert\eta\rvert < 2.5$ Tau veto $p_T > 20\,\text{GeV}, \lvert\eta\rvert < 2.3$ $p_T^{\text{miss}} > 100\,\text{GeV}$ $30 \leq M_T(e, p_T^{\text{miss}}) < 125\,\text{GeV}$ $\min[\Delta\phi(j_{1,2,3,4}, U)] > 0.5$
Double Electron	Baseline 2 electrons $p_T > 30\,\text{GeV}, \lvert\eta\rvert < 2.4$ Muon veto $p_T > 10\,\text{GeV}, \lvert\eta\rvert < 2.5$ Tau veto $p_T > 20\,\text{GeV}, \lvert\eta\rvert < 2.3$ $71 < M_{ee} < 111\,\text{GeV}$ $\min[\Delta\phi(j_{1,2,3,4}, U)] > 0.5$
Single Tau	Baseline 1 tau $p_T > 40\,\text{GeV}, \lvert\eta\rvert < 2.3$ Muon veto $p_T > 10\,\text{GeV}, \lvert\eta\rvert < 2.5$ Electron veto $p_T > 10\,\text{GeV}, \lvert\eta\rvert < 2.5$ $\min[\Delta\phi(j_{1,2,3,4}, U)] > 0.5$
QCD Sideband	Baseline Muon veto $p_T > 10\,\text{GeV}, \lvert\eta\rvert < 2.5$ Electron veto $p_T > 10\,\text{GeV}, \lvert\eta\rvert < 2.5$ Tau veto $p_T > 20\,\text{GeV}, \lvert\eta\rvert < 2.3$ $\min[\Delta\phi(j_{1,2,3,4}, U)] \leq 0.5$

4.3.2 Event Selection Study Using Machine Learning

In addition to the traditional cut-based selection approach, a complementary study was conducted using machine learning techniques to identify signal region events. Classical supervised models such as Boosted Decision Trees (BDTs) may suffer from *overfitting*[3] in scenarios where the events of interest have low diversity — as is the case in the signal region, characterized by high missing transverse momentum (p_T^{miss}), a high-p_T leading jet, and strict lepton vetoes that reduce the variability in the dataset.

Here, the problem is framed as a binary classification task, aiming to train a model capable of identifying events that fulfill the predefined criteria for the signal region.

In this context, deep learning methods offer greater robustness and generalization capabilities. In particular, we opted for **Graph Neural Networks (GNNs)**, which have proven effective in particle physics event classification, especially in jet identification and substructure analysis [28, 29]. Since the events analyzed here involve multiple jets, GNNs are a natural fit.

Graph Neural Networks are deep learning models specifically designed to operate on graph-structured data — that is, sets of *nodes* and *edges* representing entities and their relationships. Unlike convolutional neural networks (CNNs), which assume Euclidean data structure, GNNs can process information in non-Euclidean domains. This makes them well-suited to capture the topology of collision events, where multiple particles interact in patterns that do not follow a regular geometry [28, 29].

In high-energy physics, each node in a graph can represent a reconstructed object (such as a jet, muon, or photon), while the edges link pairs of particles. The resulting graph captures the underlying physics of the event, allowing the GNN to learn interaction patterns relevant for distinguishing signal from background.

The training dataset consisted of **58,363 class 1 events** (signal region) and **486,097 class 0 events** (non-signal). This strong class imbalance can degrade model performance by favoring the majority class during learning.

To mitigate this, we applied a mixed strategy of *under-sampling*[4] the majority class and *over-sampling*[5] the minority class. Over-sampling was performed using the SMOTE (Synthetic Minority Over-sampling Technique) algorithm, which generates synthetic samples by interpolating between real, nearby class 1 events in feature space — thus increasing diversity without simple duplication.

[3] Overfitting occurs when a machine learning model becomes too closely fitted to the training data, capturing noise and statistical fluctuations specific to that data, which undermines its ability to generalize to new data.

[4] Under-sampling reduces the number of majority-class examples to balance the dataset.

[5] Over-sampling increases the number of minority-class examples by duplicating or synthetically generating new instances.

Each event was represented as a graph with four nodes, corresponding to the leading jet's kinematic variables: `lead_jet_pt`, `lead_jet_eta`, `lead_jet_phi`, and `lead_jet_mass`. Nodes were connected in a linear topology, linking consecutive pairs. This compact representation was sufficient to encode key features of the event while maintaining low computational cost.

More complex graph structures (e.g., including all reconstructed particles as nodes) would significantly increase computational demands. In this context, computational cost refers primarily to training and inference time, which scales with the number of nodes and edges. Larger graphs require more convolutions and memory, limiting practical applications on resource-constrained systems such as local machines.

The model was trained for 10 epochs, using accuracy, F1-score, and confusion matrix as performance metrics. Additionally, a decision threshold of 0.6 (rather than the default 0.5) was applied to the class 1 probability, meaning an event was classified as signal only if the predicted probability exceeded 60%. This choice aimed to reduce the false positive rate — i.e., prevent background events from being misclassified as signal.

Figure 7 shows the evolution of accuracy and F1-score across the 10 training epochs. The metrics improve steadily, peaking in the final epoch. An abrupt drop at epoch 7 likely results from statistical fluctuations in the validation set, a common issue when the validation sample is small or contains outliers. The rapid recovery in subsequent epochs indicates the drop was not due to model limitations, but rather sampling instability. The model ultimately achieves around 85% in both accuracy and F1-score.

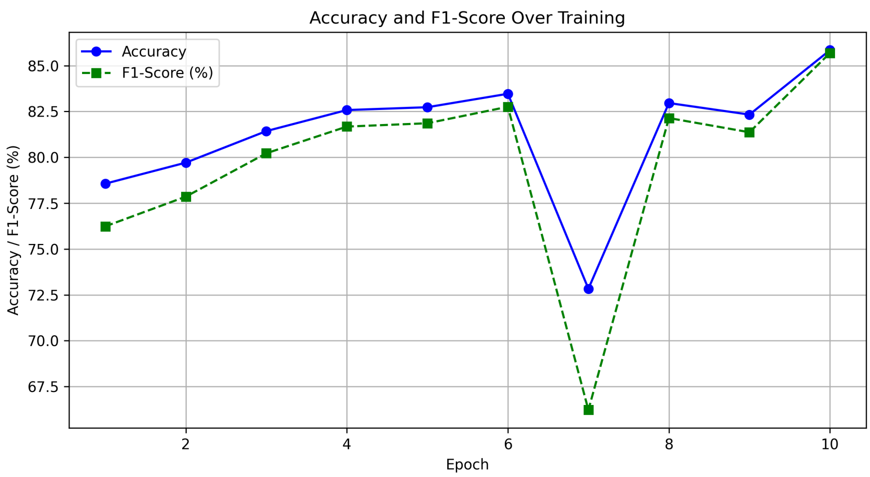

Figure 7 – Accuracy and F1-score over the training epochs.

Figure 8 further supports this analysis by showing the training loss curve. The loss decreases consistently in early epochs, reaching a minimum around epoch 7 — coinciding with the lowest performance in other metrics. This suggests a temporary mismatch between loss minimization and generalization capacity. From epoch 9 onward, both loss and performance metrics stabilize.

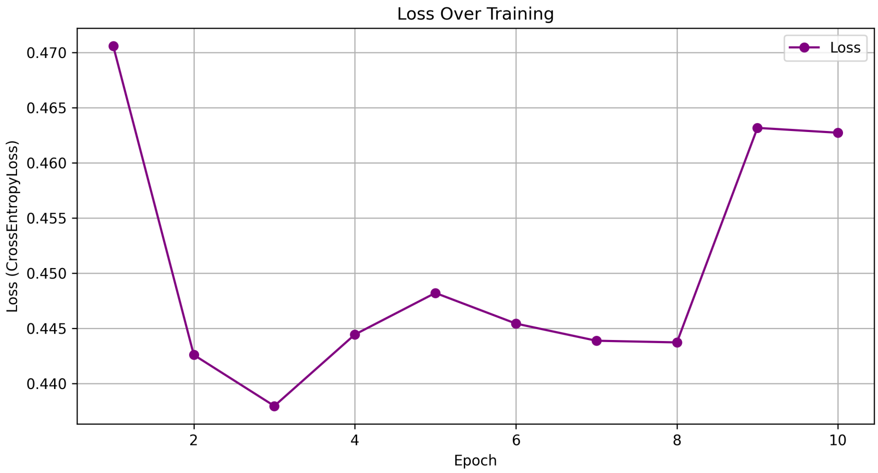

Figure 8 – Training loss (CrossEntropyLoss) during the epochs.

The final confusion matrix (Figure 9) shows that the model correctly classifies all class 1 events (signal), with no false negatives — in line with the analysis goal of efficiently capturing signal-compatible events. Although 6,611 class 0 events were misclassified as signal (false positives), this is mitigated by a physical post-selection cut: only events with recoil magnitude $u > 200$ GeV are retained. This final filter removes many of the false positives.

This setup effectively prioritizes signal retention while keeping computational cost low, enabling full local execution without high-performance resources.

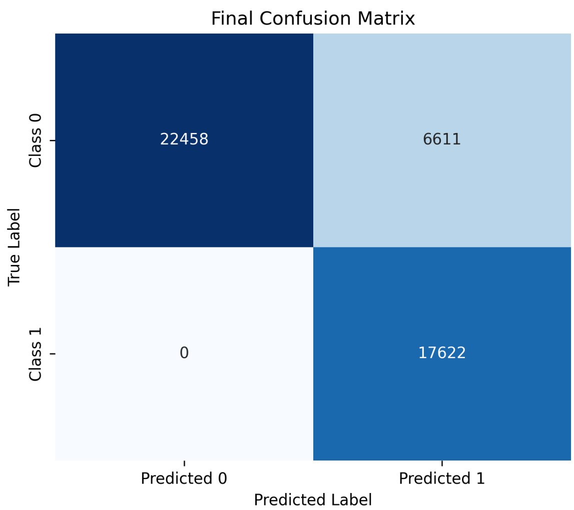

Figure 9 – Final confusion matrix after training the GNN model.

After training and validation, a new dataset was classified using both the traditional cut-based criteria and the GNN model. The results were compared to assess selection efficiency.

Figures 10 and 11 show that the GNN selects events satisfying key physical cuts, such as recoil magnitude u and leading jet p_T. Furthermore, it selects significantly more events than the traditional approach.

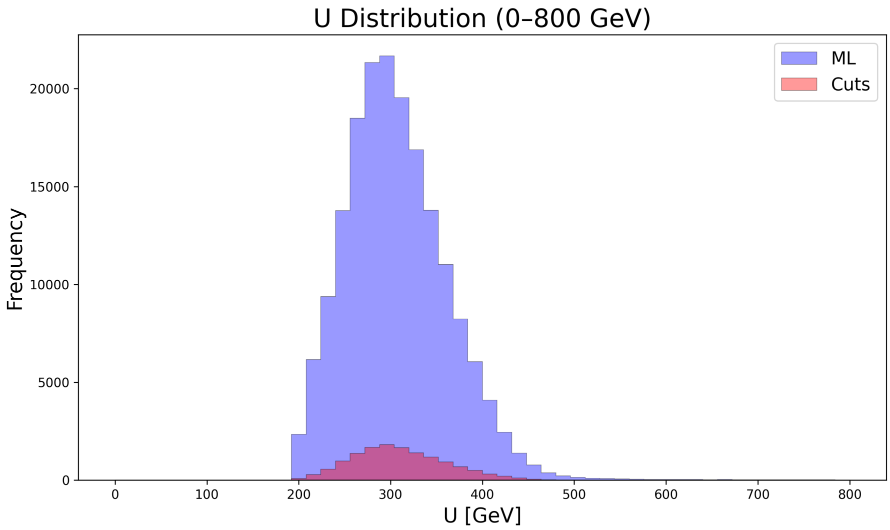

Figure 10 – Distribution of the recoil magnitude u for ML- vs. cut-based selected events.

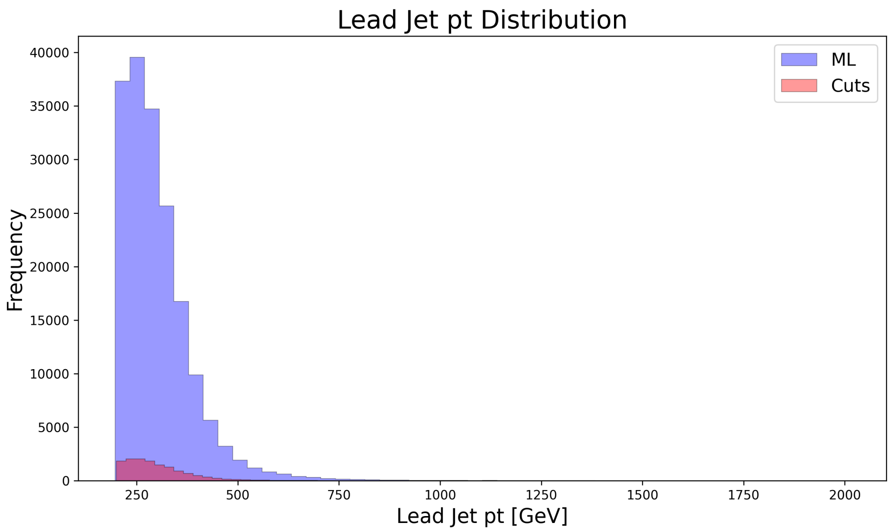

Figure 11 – Distribution of leading jet p_T for ML- vs. cut-based selected events.

However, Figures 12 and 13 reveal limitations in veto application. The presence of photons with forbidden p_T and η values suggests the model did not fully learn the veto logic — likely explaining the surplus of selected events relative to cut-based selection.

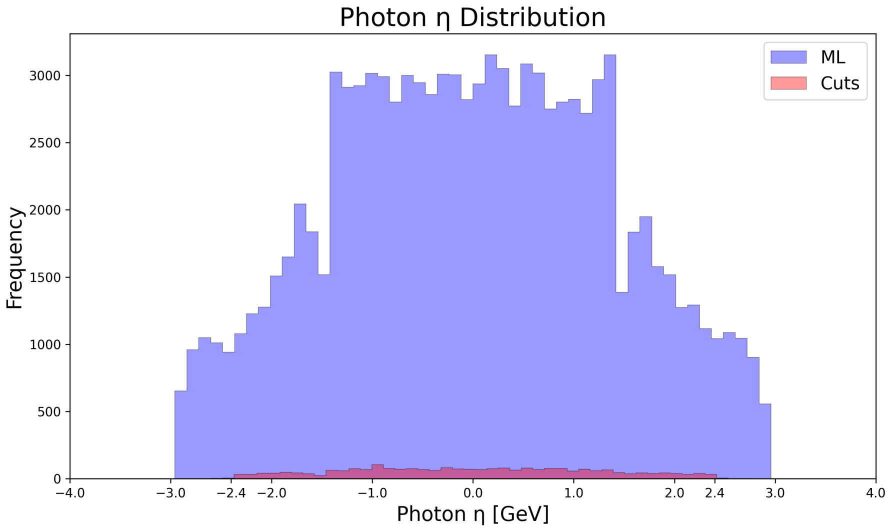

Figure 12 – Photon η distribution for ML- vs. cut-based selected events.

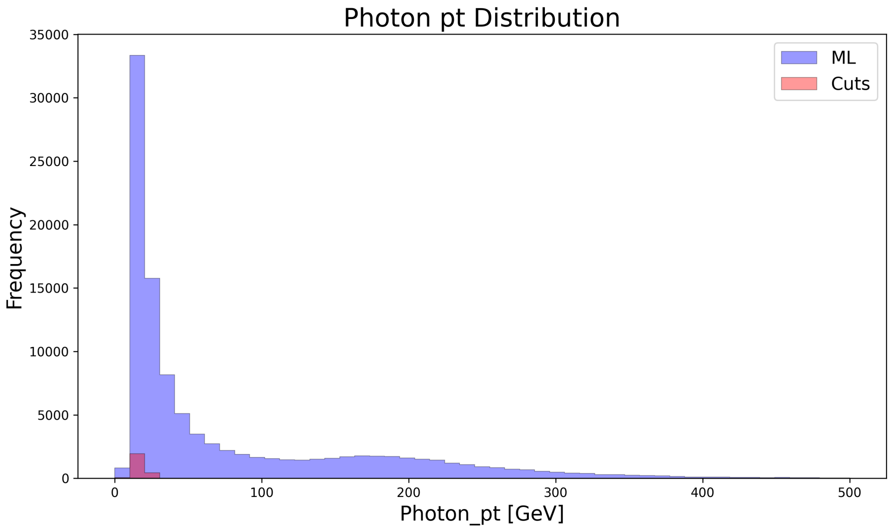

Figure 13 – Photon p_T distribution for ML- vs. cut-based selected events.

To address this, two improvements are proposed: (i) enrich the training set with more examples containing vetoed particle types, and (ii) explicitly represent those particles as nodes in the input graphs, enabling the network to learn their impact and apply proper rejection.

Nevertheless, the model shows strong performance in selection cuts, meeting its primary goal efficiently and with low computational cost.

4.4 Event Weighting for Normalization

Event normalization is a crucial step when using samples generated via Monte Carlo simulation, as it ensures that the simulated distributions reflect the behavior observed in experimental data. This normalization is performed by applying weights to each event. According to the CMS Open Data guidelines [30], event weights should be computed using the following expression:

$$w = \frac{\sigma \times \mathcal{L}}{n_{\text{gen}}}, \tag{4.1}$$

where n_{gen} denotes the effective number of generated events, defined as $n_{\text{gen}} = \text{gw\_pos} - \text{gw\_neg}$, with gw_pos being the sum of all positive weights ($\sum(\text{gen\_weights} > 0)$), and gw_neg the sum of all negative weights ($\sum(\text{gen\_weights} < 0)$).

In some Monte Carlo samples, events may carry positive or negative weights due to reweighting techniques applied to enhance simulation accuracy. This occurs, for example, in *event unweighting* procedures [31], which aim to produce unit-weight events by probabilistically accepting or rejecting events drawn from distributions with variable weights. Additionally, negative weights may arise in higher-order QCD calculations (e.g., NLO and NNLO[6]), where virtual corrections and interference terms require assigning negative weights to preserve theoretical consistency in the simulation [33].

For collision data, the applied event weight is equal to 1.

4.5 Correction of the γ^* Contribution in the DYJetsToLL Sample

In this analysis, the `DYJetsToLL` dataset is used to model the production of the Z boson followed by its decay into charged leptons. However, this dataset also includes contributions from the virtual photon (γ^*) and the interference between Z and γ^*. Therefore, it is necessary to apply a correction to isolate the pure $Z \to \ell^+\ell^-$ contribution.

To achieve this, the cross sections of the $Z \to \ell\ell$ and $\gamma^* \to \ell\ell$ processes were used to calculate the correction factors k_Z and k_γ, as shown in Table 21.

Process	**Cross Section (pb)**	**Ratio relative to DY**
DY($\ell\ell$)	5316	1
Z($\ell\ell$)	5203	0.9787 (k_Z)
$\gamma^*(\ell\ell)$	79	0.0149 (k_γ)

Table 20 – Cross sections and ratios of various processes relative to DY.

With these factors, the corrected cross sections for the Z and γ^* processes can be calculated and summed to obtain the corrected cross section of DYToLL:

[6] Next-to-Leading Order (NLO) and Next-to-Next-to-Leading Order (NNLO) refer to successive levels in the perturbative expansion of QCD. NLO includes first-order corrections in α_s, such as gluon emissions and virtual diagrams. NNLO incorporates second-order terms and multiple-loop contributions, improving the precision of cross-section predictions and observable distributions [32].

$$\sigma_{Z\text{ corrected}} = \sigma_Z \times k_Z = 5092.2\,\text{pb}, \tag{4.2}$$

$$\sigma_{\gamma^*\text{ corrected}} = \sigma_{\gamma^*} \times k_{\gamma^*} = 1.2\,\text{pb}, \tag{4.3}$$

$$\sigma_{\text{DY corrected}} = \sigma_{Z\text{ corrected}} + \sigma_{\gamma^*\text{ corrected}} = 5093.4\,\text{pb}, \tag{4.4}$$

Then, the interference and its percentage contribution to the DYToLL process can be estimated:

$$\text{Interference} = \sigma_{\text{DY}} - (\sigma_Z + \sigma_{\gamma^*}) = 34\,\text{pb}, \tag{4.5}$$

$$\frac{\text{Interference}}{\sigma_{\text{DY}}} \times 100 = \frac{34}{5316} \times 100 = 0.64\%, \tag{4.6}$$

Finally, the weight to be applied to the DYToLL dataset to exclude the interference and isolate the pure Z contribution can be computed:

$$\text{Interference} = \sigma_{\text{DY total corrected}} \times 0.0064 = 32.6\,\text{pb}, \tag{4.7}$$

$$\sigma_{\text{DY final}} = \sigma_{\text{DY total corrected}} - 32.6 = 5060.8\,\text{pb}, \tag{4.8}$$

$$w = \frac{\sigma_{\text{DY final}}}{\sigma_{\text{DY simulated}}} = \frac{5060.8}{5316} = 0.9519. \tag{4.9}$$

By applying a weight of $w = 0.9519$ to the events in the `DYJetsToLL` dataset, one ensures that only the contribution from the $Z \to \ell^+\ell^-$ process is included in the analysis, with the virtual photon (γ^*) and interference contributions properly corrected.

4.6 Pileup Correction

The term *pileup* refers to the multiple proton-proton interactions that occur within a single bunch crossing. Pileup correction is applied to Monte Carlo (MC) simulated events in order to match the real collision conditions observed in the CMS experiment data. This is achieved through a *reweighting* technique, in which each simulated event is assigned a weight that depends on the number of pileup interactions associated with it.

The true pileup distribution is obtained following the official CMS recommendations [34], using the `pileupCalc.py` tool in combination with recorded luminosity information and the recommended inelastic cross section value for the 2016 data-taking period [34]. The result is a normalized histogram in ROOT format, representing the number of pileup interactions per event in the real data. This distribution is then compared to the corresponding pileup distribution from the MC simulation. For each pileup value n, a correction weight is calculated as the ratio between the data and simulation distributions:

$$w(n) = \frac{P_{\text{data}}(n)}{P_{\text{MC}}(n)}, \tag{4.10}$$

where $P_{\text{data}}(n)$ represents the normalized probability of observing n interactions in real data, and $P_{\text{MC}}(n)$ is the corresponding probability in the simulation. These weights are applied on an event-by-event basis so that the final pileup distribution in the MC reproduces the one observed in the real data. This ensures that variables sensitive to pileup—such as the number of primary vertices and energy deposits in different regions of the detector—are properly modeled.

Figure 14 illustrates this procedure. The shaded gray area represents the pileup distribution from real data. The red dashed line shows the original pileup distribution in the MC simulation before reweighting, while the blue line corresponds to the reweighted distribution. After applying the weights, the simulated distribution aligns well with the data.

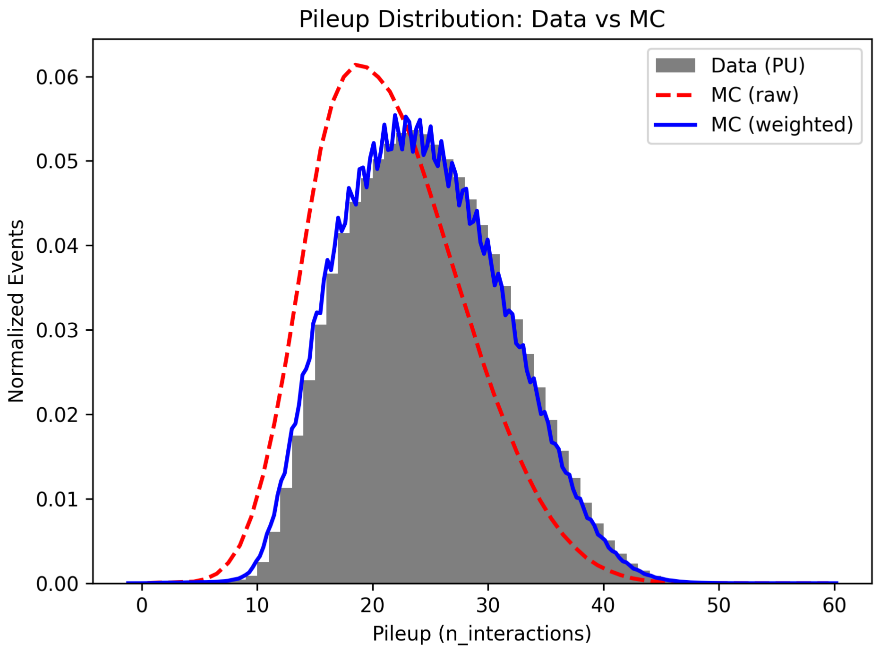

Figure 14 – Normalized pileup distributions in real data, MC simulation before reweighting, and MC simulation after reweighting.

4.7 Simultaneous Fit and Extraction of the Invisible Width of the Z Boson

The preprocessing steps — including event normalization, correction for the virtual photon contribution (γ^*) in the `DYJetsToLL` sample, and pileup correction — were applied to the events selected in the three analysis regions defined in this study: the **signal region** (p_T^{miss} + jets), used to estimate the process $Z \to \nu\bar{\nu}$; the **single lepton region**, consisting of events with exactly one muon, electron, or tau, used to model the dominant background $W \to \ell\nu$; and the **double lepton region**, composed of events with two muons or two electrons, used to estimate the reference process $Z \to \ell^+\ell^-$.

Figure 15 presents the distribution of the recoil vector U, comparing real collision data with Monte Carlo predictions. The left panel shows the distribution for the $Z \to \nu\bar{\nu}$ process, referred to as p_T^{miss} + jets, which includes events from the signal and single lepton regions. The right panel displays the distribution for the reference channel $Z \to \ell^+\ell^-$, with events taken from the double lepton region — denoted $\mu^+\mu^-$ + jets — since no events passed the selection criteria for the double electron category.

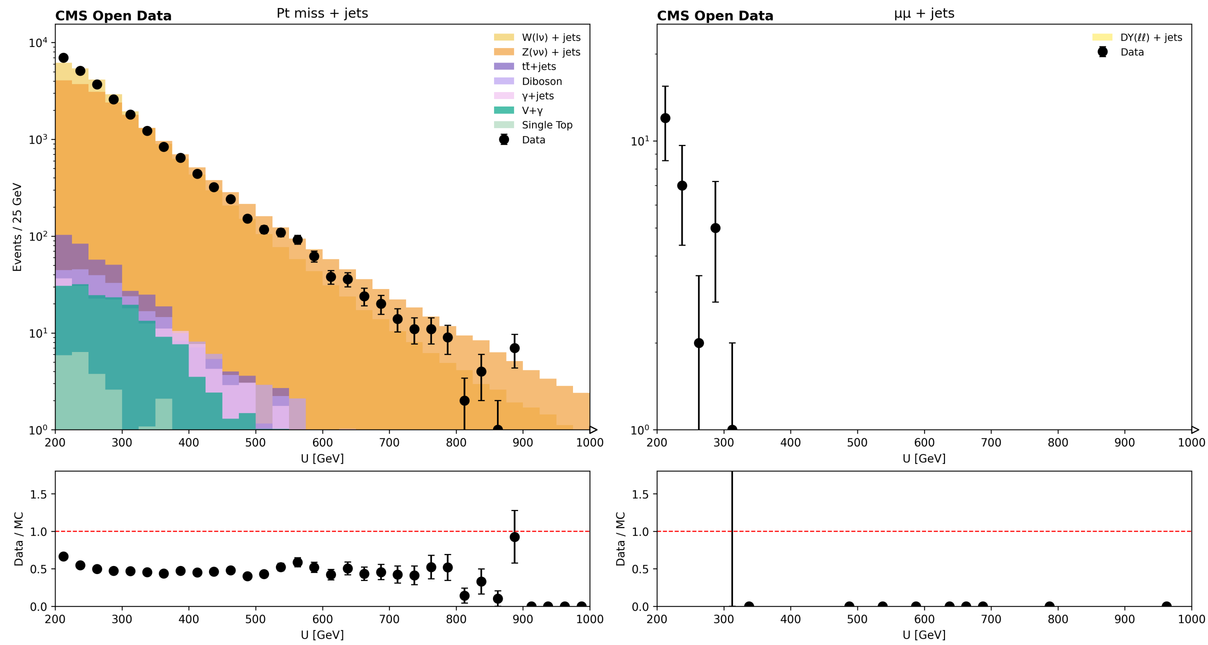

Figure 15 – Recoil vector U distributions.

The lower panels show the Data/MC ratio evaluated bin by bin, allowing a direct comparison between real data and the predictions from Monte Carlo simulated samples across the U spectrum.

In the p_T^{miss}+jets region, the Data/MC ratio remains consistently below 1 throughout most of the range, indicating that the simulation overestimates the number of events compared to what is observed in the experimental data. In contrast, the $\mu^+\mu^-$ + jets region exhibits significant fluctuations in the ratio, which are attributed to limited event statistics. This statistical limitation arises from the fact that the CERN Open Data portal provides only a subset of the 2016 proton-proton collision data, corresponding exclusively to data-taking periods G and H.

The recoil vector U distribution shown in Figure 15 serves as the primary observable upon which the simultaneous fit for extracting the invisible width of the Z boson is based. This fit is performed by maximizing a likelihood function that simultaneously accounts for the observed events in the three experimental regions — signal, single lepton, and double lepton — as well as the various physical processes contributing to each.

The likelihood function used is defined as:

$$\begin{aligned}\mathcal{L}(n_j, n_\ell, n_{\ell\ell} \mid r, r_Z, r_W, \theta) = &\text{Poisson}\left(n_j \mid r \cdot r_Z \cdot s_{Z,j}(\theta) + r_W \cdot b_{j,W}(\theta) + b_{\text{bkg.},j}(\theta)\right) \\ &\cdot \text{Poisson}\left(n_\ell \mid r_W \cdot b_{\ell,W}(\theta) + b_{\text{bkg.},\ell}(\theta)\right) \\ &\cdot \text{Poisson}\left(n_{\ell\ell} \mid r_Z \cdot s_{Z,\ell\ell}(\theta) + \sqrt{r_Z} \cdot s_{\text{int.},\ell\ell} + s_{\gamma^*,\ell\ell}(\theta) + b_{\text{bkg.},\ell\ell}(\theta)\right) \\ &\cdot p(\tilde{\theta}, \theta)\end{aligned} \tag{4.11}$$

where the terms correspond to:

- $s_{Z,j}(\theta)$: expected yield of the $Z \to \nu\bar{\nu}$ process in the signal region.
- $b_{j,W}(\theta)$: expected yield of the $W+$jets process in the signal region.
- $b_{\ell,W}(\theta)$: expected yield of the $W+$jets process in the single lepton region.
- $s_{Z,\ell\ell}(\theta)$: expected yield of the $Z \to \ell^+\ell^-$ process in the double lepton region.
- $s_{\text{int.},\ell\ell}$: interference term between Z and γ^* in the double lepton region.
- $s_{\gamma^*,\ell\ell}(\theta)$: contribution from the γ^* process in the double lepton region.
- $b_{\text{bkg.},j}(\theta)$, $b_{\text{bkg.},\ell}(\theta)$, $b_{\text{bkg.},\ell\ell}(\theta)$: expected yields of other background processes in the signal, single lepton, and double lepton regions, respectively.
- $r_Z = \frac{N_{\text{MC}}^{\text{signal}}(Z\text{+jets})}{N_{\text{MC}}^{\text{control}}(Z\text{+jets})}$: transfer factor for the $Z+$jets process.
- $r_W = \frac{N_{\text{MC}}^{\text{signal}}(W\text{+jets})}{N_{\text{MC}}^{\text{control}}(W\text{+jets})}$: transfer factor for the $W+$jets process.
- $r = \frac{\Gamma(Z\to\nu\bar{\nu})}{\Gamma_{\text{SM}}(Z\to\nu\bar{\nu})}$: scaling factor of the invisible width of the Z boson relative to the Standard Model prediction.

Based on the ratio between simulated events in the signal and control regions, the following transfer factors were obtained: $r_Z = (1.09 \pm 2.55) \times 10^5$ for the $Z+$jets process, and $r_W = 1.4050 \pm 0.0180$ for the $W+$jets process.

The value obtained for the r_Z factor is abnormally large and exhibits a relative uncertainty greater than 200%. This is due to the limited number of events from the `DYJetsToLL` sample that pass the selection criteria in the double lepton control region, undermining the reliability of a direct estimation of r_Z.

Given this, it is appropriate to treat r_Z as a free parameter[7] in the simultaneous fit, allowing its value to be inferred directly from the data. This choice avoids propagating an unstable and statistically unreliable estimate.

The simultaneous fit across the analysis regions was performed using the Markov Chain Monte Carlo (MCMC) sampling technique. This approach enables numerical exploration of the posterior distribution of the model parameters based on the constructed likelihood function and the priors defined for each parameter.

The set of fitted parameters includes the scale factor r, the transfer factor r_Z, and the global systematic uncertainty parameter θ. The transfer factor r_W was kept fixed at the previously obtained value.

To incorporate experimental uncertainties and theoretical constraints into the simultaneous fit, Gaussian penalties (priors) were applied to specific free parameters in the model:

- **Global parameter** θ: models the global systematic uncertainty shared across all analysis regions. A Gaussian prior with $\sigma_\theta = 0.032$ was applied, consistent with the combined systematic uncertainty of 3.2%, as detailed in Table 21.

- **Scale factor** r: directly related to the invisible width of the Z boson. An informative prior centered at $r = 1.0$ with $\sigma_r = 0.1$ was used, reflecting the Standard Model expectation.

- **Transfer factor** r_Z: applied to the Z+jets process, with a prior centered at $r_Z = 0.7$ and $\sigma_{r_Z} = 0.2$.

[7] A free parameter is a numerical quantity whose best estimate is determined directly from the fit to the experimental data, rather than being fixed in advance or computed from other quantities.

Table 21 – Sources of systematic uncertainty considered in the analysis. Source: [8]

Source	**Uncertainty (%)**
Muon identification efficiency (syst.)	2.1
Jet energy scale	1.8–1.9
Electron identification efficiency (syst.)	1.6
Electron identification efficiency (stat.)	1.0
Pileup	0.9–1.0
Electron trigger efficiency	0.7
Tau veto efficiency	0.6–0.7
p_T^{miss} trigger efficiency (jets + MET)	0.7
p_T^{miss} trigger efficiency ($Z/\gamma^* \to \mu\mu$)	0.6
QCD correction dependent on boson p_T	0.5
Jet energy resolution	0.3–0.5
Muon + jets trigger efficiency	0.4
Muon identification efficiency (stat.)	0.3
Electron reconstruction efficiency (syst.)	0.3
EWK corrections dependent on boson p_T	0.3
Parton distribution functions (PDFs)	0.2
Renormalization/factorization scale	0.2
Electron reconstruction efficiency (stat.)	0.2
Total (combined)	**3.2**

Figure 16 displays the correlation diagram of the parameters fitted via MCMC. The marginal distributions show the posterior density of each individual parameter, while the 2D panels reveal the correlations between parameter pairs.

The fitted parameters and their respective uncertainties, estimated from the posterior distribution obtained through MCMC, are:

- $r = 1{,}272^{+0{,}076}_{-0{,}073}$,
- $r_Z = 1{,}469^{+0{,}093}_{-0{,}088}$,
- $\theta = -0{,}6723^{+0{,}0052}_{-0{,}0051}$.

The results indicate that the model adjusted the parameters to compensate for the high value of r_Z, which results from the limited statistics in the double lepton control region. To maintain consistency with the observed data, the simultaneous fit increased the values of $r = 1.272^{+0.076}_{-0.073}$ and $\theta = -0.6723^{+0.0052}_{-0.0051}$. This behavior reflects the correlation between the parameters: by limiting the increase of $r_Z = 1.469^{+0.093}_{-0.088}$, the model redistributes the adjustments between r and θ to better match the observed distributions.

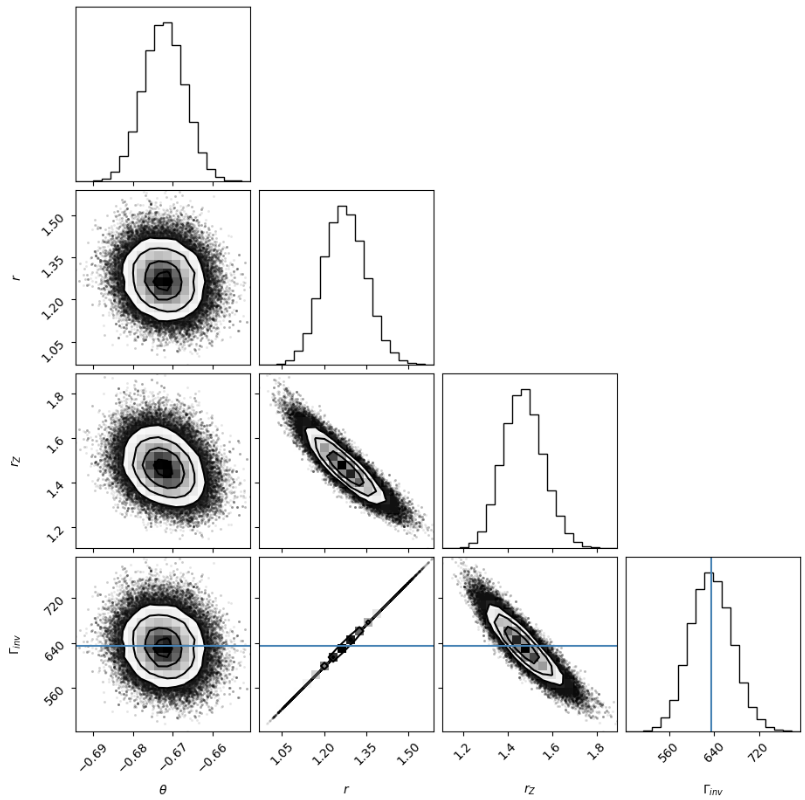

Figure 16 – Marginal distributions and correlations between the fitted parameters: θ (global systematic), r (scaling factor for the Z invisible width), r_Z (transfer factor for $Z+$ jets), and Γ_{inv} (invisible width).

With the fitted value of r, it is possible to compute the invisible width of the Z boson using the Standard Model theoretical prediction for this decay channel: $\Gamma_{\text{SM}}(Z \to \nu\bar{\nu}) = 499.0\,\text{MeV}$. The relation used is:

$$\Gamma(Z \to \nu\bar{\nu}) = r \cdot \Gamma_{\text{SM}}(Z \to \nu\bar{\nu}). \tag{4.12}$$

Substituting the values, we obtain:

$$\Gamma(Z \to \nu\bar{\nu}) = 1{,}272 \times 499{,}0 = 634{,}6\,\text{MeV}. \tag{4.13}$$

Taking into account the propagation of the statistical uncertainties associated with the parameter r, the final result for the invisible width of the Z boson is:

$$\Gamma(Z \to \nu\bar{\nu}) = 634{,}6^{+37{,}9}_{-36{,}5}\,\text{MeV}. \tag{4.14}$$

In addition to the statistical uncertainty arising from the fit of the parameter r, it is necessary to account for two additional sources of uncertainty that affect the final estimate of the invisible width:

- The uncertainty in the transfer factor $r_W = 1.4050 \pm 0.0180$;
- The global systematic uncertainty, represented by θ, with a standard deviation of $\sigma_\theta = 0.032$.

The absolute contribution of these uncertainties to $\Gamma(Z \to \nu\bar{\nu})$ was propagated as follows:

- Uncertainty due to r_W:

$$\Delta\Gamma_{r_W} = \Gamma_{\text{SM}} \cdot \langle r \rangle \cdot \sigma_{r_W} = 499.0 \cdot 1.272 \cdot 0.018 \approx 11.4\,\text{MeV};$$

- Global systematic uncertainty:

$$\Delta\Gamma_{\text{syst}} = \langle \Gamma \rangle \cdot \sigma_\theta = 634.6 \cdot 0.032 \approx 20.3\,\text{MeV}.$$

The total upper and lower uncertainties were then calculated by the quadratic combination of the statistical and systematic components:

$$\Delta\Gamma^{+}_{\text{total}} = \sqrt{37{,}9^2 + 11{,}4^2 + 20{,}3^2} \approx 44{,}5\,\text{MeV}, \tag{4.15}$$

$$\Delta\Gamma^{-}_{\text{total}} = \sqrt{36{,}5^2 + 11{,}4^2 + 20{,}3^2} \approx 43{,}3\,\text{MeV}. \tag{4.16}$$

Thus, the final value of the invisible width of the Z boson, incorporating all sources of uncertainty considered in this analysis, is:

$$\Gamma(Z \to \nu\bar{\nu}) = 634{,}6^{+44{,}5}_{-43{,}3}\,\text{MeV}. \tag{4.17}$$

Approximately, this result can be expressed as:

$$\Gamma(Z \to \nu\bar{\nu}) \approx 635 \pm 44 \text{ MeV}, \tag{4.18}$$

Figure 17 presents a direct comparison between the result obtained in this analysis and other experimental measurements of the invisible width of the Z boson. The vertical dashed line indicates the value predicted by the Standard Model, $\Gamma_{\text{SM}}(Z \to \nu\bar{\nu}) = 499.0\,\text{MeV}$.

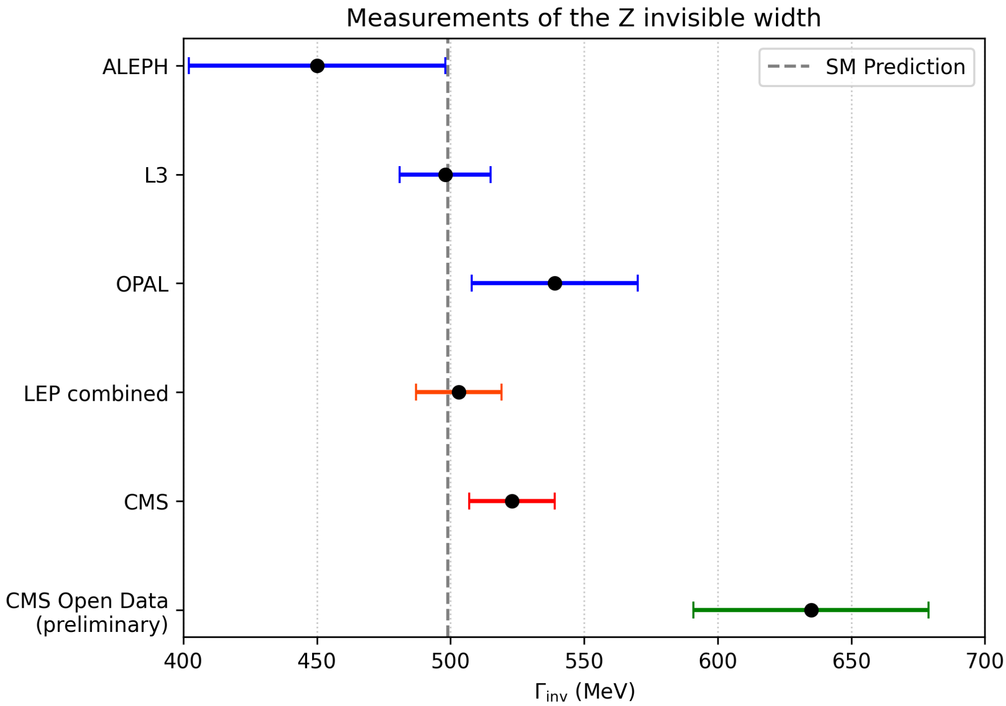

Figure 17 – Comparison between different experimental measurements of the invisible width of the Z boson. The dashed line represents the Standard Model prediction. The bottom point (in green) corresponds to the result obtained in this analysis using preliminary CMS Open Data.

While the combined results from the LEP experiments (ALEPH, L3, OPAL) measured values compatible with the Standard Model within relatively small uncertainties, the CMS collaboration analysis [8] already shows a modest upward deviation. The result obtained in this analysis — $\Gamma(Z \to \nu\bar{\nu}) = 634.6^{+44.5}_{-43.3}$ MeV — is significantly higher, with a central value approximately 27% above the SM prediction.

This discrepancy can largely be attributed to the statistical limitations of the samples available through the CERN Open Data portal, which include only the G and H eras of the 2016 proton-proton collision data, in addition to uncertainties associated with the transfer factor r_Z, which are exacerbated by the limited Monte Carlo statistics in the control region. The value obtained in this analysis, $\Gamma(Z \to \nu\bar{\nu}) = 634.6^{+44.5}_{-43.3}$ MeV, is considerably higher than the most precise estimate from the CMS collaboration, $\Gamma = 523 \pm 16$ MeV.

To quantify this discrepancy, we compute the statistical separation between the two results using the following equation:

$$\Delta = \frac{|\Gamma_{\text{CMS Open Data}} - \Gamma_{\text{CMS}}|}{\sqrt{\sigma^2_{\text{Open Data}} + \sigma^2_{\text{CMS}}}} = \frac{|634{,}6 - 523|}{\sqrt{44^2 + 16^2}} = \frac{111{,}6}{46{,}82} \approx 2{,}38\,\sigma. \tag{4.19}$$

The difference between the two results thus corresponds to approximately 2.4σ. Although statistically significant, this deviation can be attributed to the uncertainties arising from the limited statistics available in this analysis.

4.8 Contribution to the Upgrade of the CMS Muon System

With the upgrade to the High-Luminosity LHC (HL-LHC), the instantaneous luminosity is expected to increase to $5 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, approximately five times the original design value, while the integrated luminosity could reach $450\,\mathrm{fb}^{-1}$ per year. This will result in an average pileup of up to 100 proton-proton interactions per bunch crossing, posing challenges for accurate event reconstruction and accelerating detector component aging [35].

Resistive Plate Chambers (RPCs), positioned in both the barrel and endcap regions, are gaseous detectors known for their excellent time resolution of about 1.5 ns, which is essential for muon detection and triggering. With a total of 1,056 chambers, they operate in avalanche mode under a 9.8 kV electric field. The standard gas mixture includes 95.2% $C_2H_2F_4$, 4.5% iC_4H_{10}, and 0.3% SF_6, where each component plays a specific role such as initial ionization, amplification control, and discharge suppression [11].

To meet the demands of the HL-LHC, the RPC system will undergo significant upgrades. The electronics will be modernized to handle the increased data flow, enabling faster signal processing and improving the time resolution from 25 ns to the intrinsic 1.5 ns of the detectors. In addition, a new generation of RPCs, called *iRPCs* (improved RPCs), will be installed to extend angular coverage from $|\eta| = 1.9$ to $|\eta| = 2.4$, a region experiencing radiation rates of up to $700\,\mathrm{Hz/cm}^2$ [11].

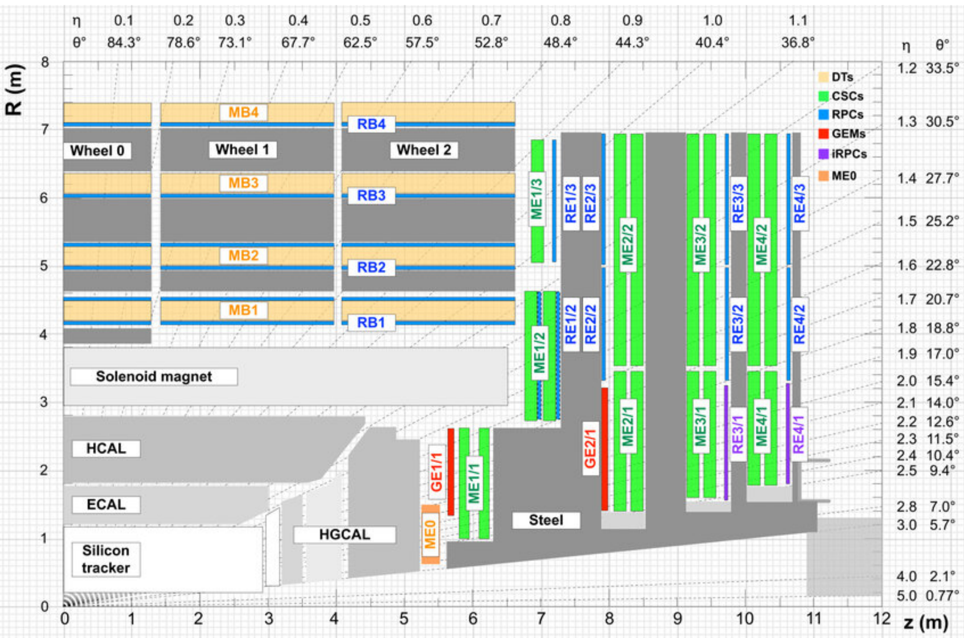

Figure 18 – A quadrant of the CMS detector. In purple, the locations of the two iRPC stations in the endcaps—RE3/1 and RE4/1. Source: [6].

The *iRPCs* introduce significant advances in noise rejection, high-rate detection capability, and improved spatial resolution. Using High Pressure Laminate (HPL) panels and narrower gas gaps of 1.4 mm, the *iRPCs* achieve spatial resolutions of $0.3 - 0.6$ cm in the ϕ direction and 1.5 cm in the η direction [11]. These improvements allow the *iRPCs* to operate efficiently in the high-luminosity HL-LHC environment, with enhanced time resolution reaching 500 ps—crucial for accurately reconstructing muon trajectories and distinguishing temporally close events.

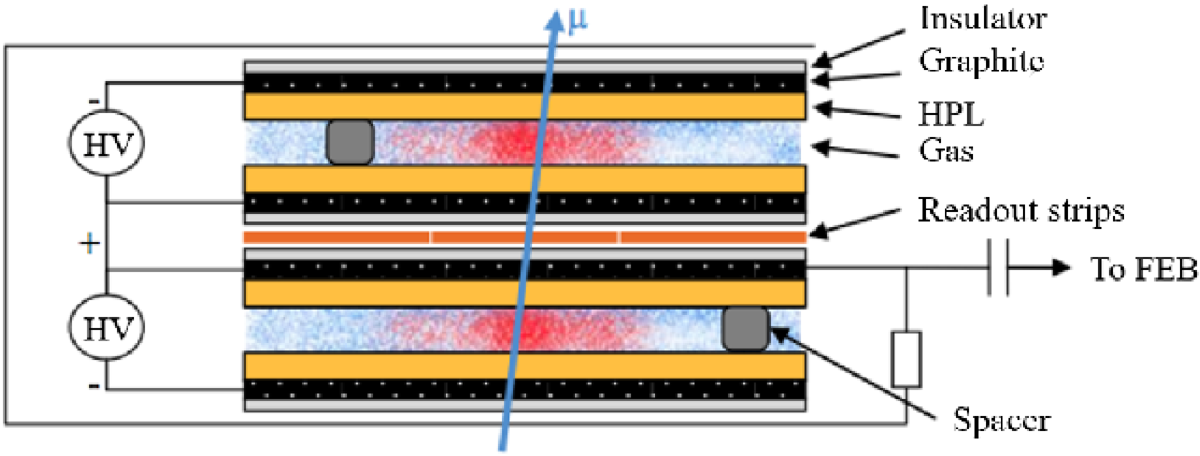

Figure 19 – Detailed diagram of an RPC showing its main components: HPL resistive plates, readout strips, gas, graphite layer, insulation, and spacers. Source: [7].

The assembly of the new detectors is taking place at the CERN 904 laboratory in Prevessin and at Ghent University. To ensure proper performance, the iRPC chambers undergo a series of quality control (QC) tests at each stage of the assembly chain, as outlined below:

- **QC1: Verification of the basic chamber components**
  - HPL: National Institute for Nuclear Physics (INFN)
  - Strip Detector PCBs: Lyon
  - Front-End Board (FEB): Lyon
  - Cooling System: University of Georgia
- **QC2: Testing of individual chamber elements**
  - Gap testing at the Korean Detector Laboratory:
    * Gas leak testing
    * Spacer fixation check
    * Dark current scans
  - At assembly sites:
    * Repetition of the above tests

- **QC3: Full performance evaluation of the chamber after assembly**
  - QC3.1: Chamber component testing after assembly
  - QC3.2: Cosmic ray data acquisition using a portable Front-End Board (FEB)
- **QC4: Final validation testing**
  - QC4.1: Final chamber tests
    * Cooling system leak test
    * Gas leak test
    * Dark current scans
  - QC4.2: Long-term high-voltage (HV) stability test
  - QC4.3: Final cosmic ray test with final FEBs

During my sandwich master's scholarship at CERN, I directly participated in chamber assembly and performed the QC4.3 phase, which involves calibrating the final Front-End Boards (FEBs) and conducting efficiency tests of the *iRPCs*. In these tests, the chambers are positioned between two scintillator detectors that serve as a reference system to identify cosmic muon passage. This setup allows for the precise determination of whether the muon was detected by the RPC chamber, enabling an accurate evaluation of detection efficiency under different operating conditions.

The first plot (Figure 20) shows a 3D reconstruction of muon hit positions in an iRPC chamber, highlighting the detector's uniform response across its active area. The second plot (Figure 21) shows the correlation between muon detection efficiency and the effective high voltage (HV_{eff}), demonstrating that efficiency reaches a plateau at 99% for high voltages.

The efficiency curve is obtained from the ratio between the number of events detected by the chamber and the number of events recorded by the trigger system, allowing correction for noise and false detections. In scenarios where noise may significantly affect the measurement, a Bayesian approach is used to provide more robust estimates of the true efficiency. The data collected at various voltages are fitted with a sigmoid curve, representing the expected behavior of gaseous detectors operating in avalanche mode [36].

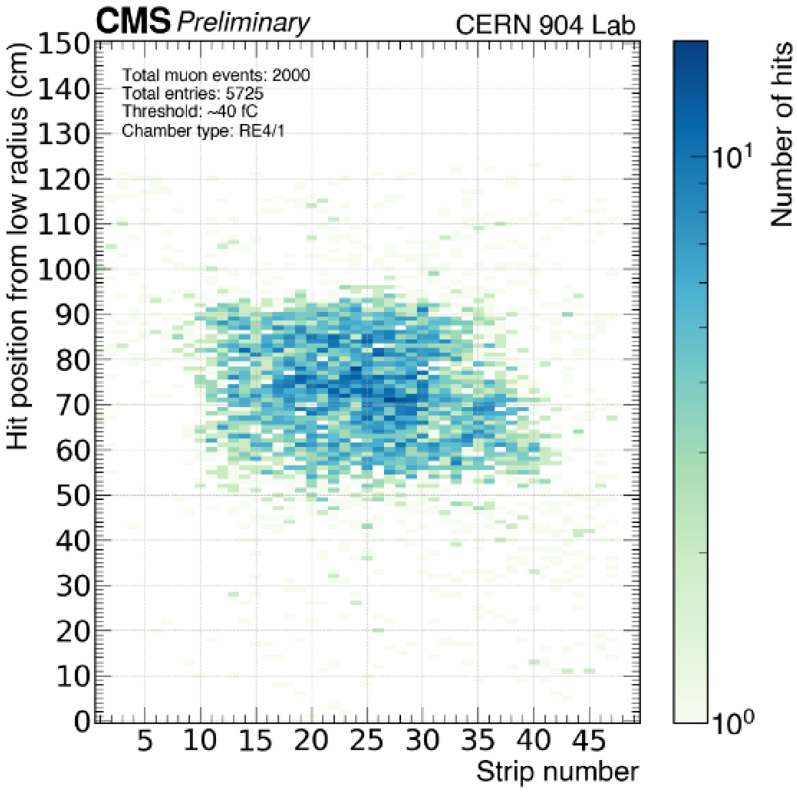

Figure 20 – 3D reconstruction of muon impact positions on an iRPC chamber at an operating point with a 40 fC charge threshold. Data were collected using a triple scintillator coincidence trigger with a 30×40 cm^2 area.

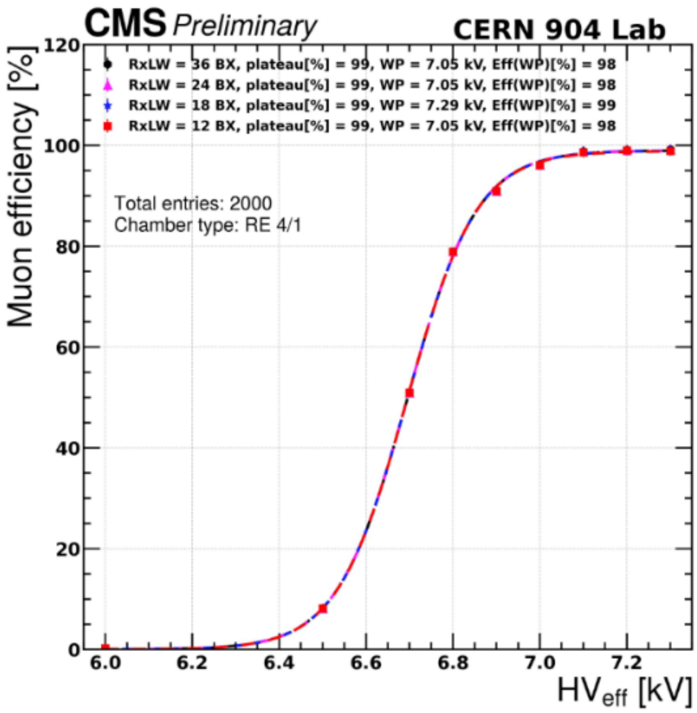

Figure 21 – Muon detection efficiency as a function of effective high voltage (HV_{eff}). The efficiency reaches a plateau of 99%, with the operating point defined as the voltage where efficiency is 95% (HV_{knee}) plus 150 V.

4.8.1 Gas Studies for the *iRPCs*

The study of gas mixtures for CMS *iRPCs* combines the technical demands of the HL-LHC with environmental responsibility. Traditional gases such as $C_2H_2F_4$ and SF_6 have high Global Warming Potentials (GWP), contributing significantly to greenhouse gas emissions. To mitigate this impact, CMS is exploring sustainable alternatives aligned with CERN's goals of reducing high-GWP gas emissions by 28% by the end of Run III and achieving net-zero emissions by 2050, in accordance with European Union targets.

Tests are being conducted at CERN's Gamma Irradiation Facility (GIF++) using alternative gas mixtures, including compositions with 30–40% CO_2 and reduced concentrations of SF_6 combined with $C_2H_2F_4$, as shown in Table 22.

Table 22 – Comparison of gas mixtures used in the CMS iRPC system and their Global Warming Potentials (GWP).

Component	**Standard Mixture**	**30% CO_2 + 1% SF_6**	**30% CO_2 + 0.5% SF_6**	**40% CO_2 + 1% SF_6**
Freon ($C_2H_2F_4$)	95.2%	64%	64.5%	54%
Isobutane (iC_4H_{10})	4.5%	5%	5%	5%
CO_2	0%	30%	30%	40%
SF_6	0.3%	1%	0.5%	1%
GWP	1486	1529	1337	1353

GIF++ uses a cesium-137 source combined with a high-energy muon beam (100–150 GeV/c) to simulate the extreme conditions expected at the HL-LHC. Tests are conducted with the gamma source both enabled and disabled, using ABS filters with values of 1, 2.2, 3.3, 4.6, 6.9, and 10. Each ABS value corresponds to a specific radiation level, which varies depending on the chamber's distance from the source.

The relationship between the ABS filter value and radiation level is inversely proportional: the lower the ABS value, the higher the radiation reaching the detector. During testing, the highest observed background radiation rate was $2.2\,\mathrm{kHz/cm^2}$ with the ABS filter set to 1, approximately three times higher than the expected rate at the HL-LHC.

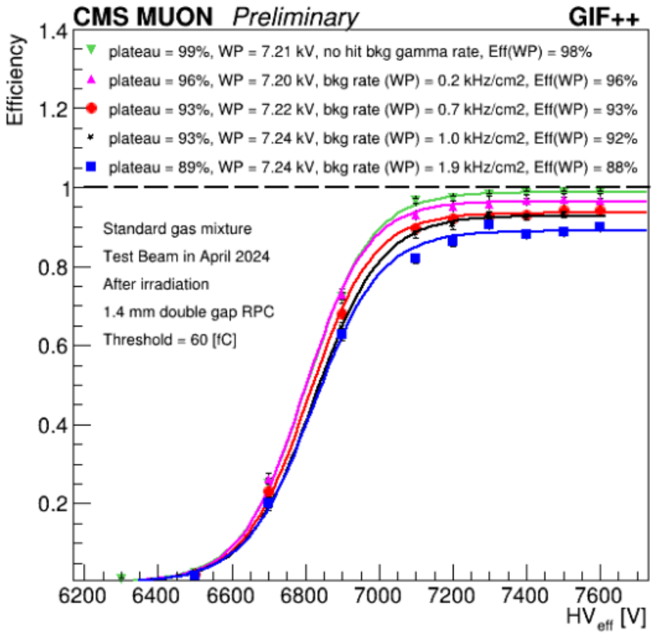

Figure 22 – Detection efficiency for the standard gas mixture with sources: OFF (green), 10 (pink), 3.3 (red), 2.2 (black), and 1 (blue), during the 2024 beam test.

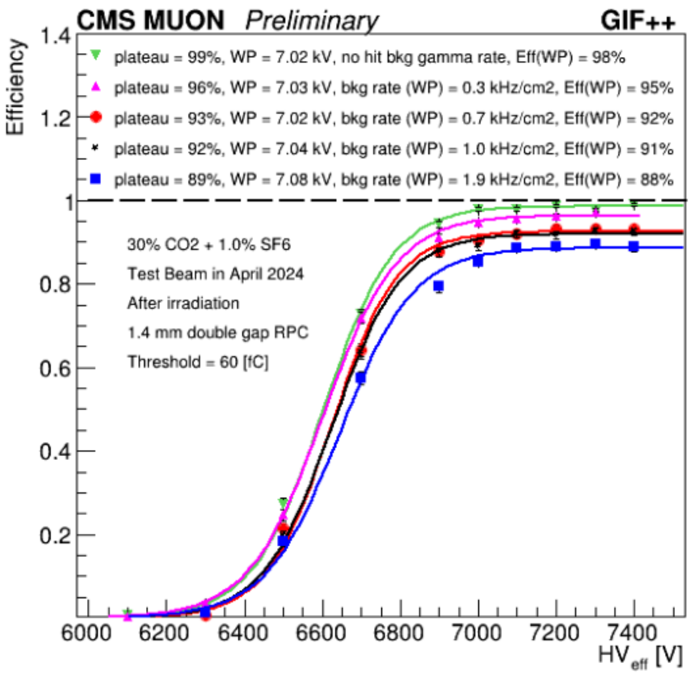

Figure 23 – Detection efficiency for the 30% CO_2 + 1.0% SF_6 gas mixture with sources: OFF (green), 10 (pink), 3.3 (red), 2.2 (black), and 1 (blue), during the 2024 beam test.

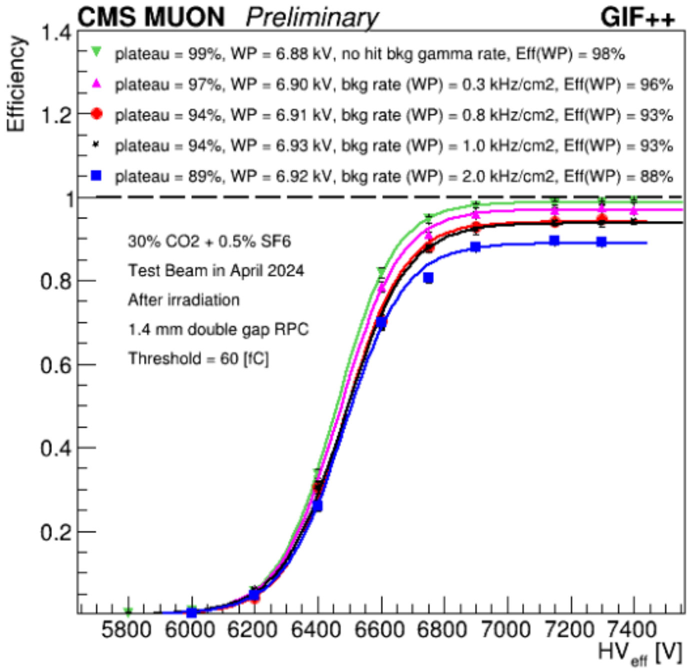

Figure 24 – Detection efficiency for the 30% CO_2 + 0.5% SF_6 gas mixture with sources: OFF (green), 10 (pink), 3.3 (red), 2.2 (black), and 1 (blue), during the 2024 beam test.

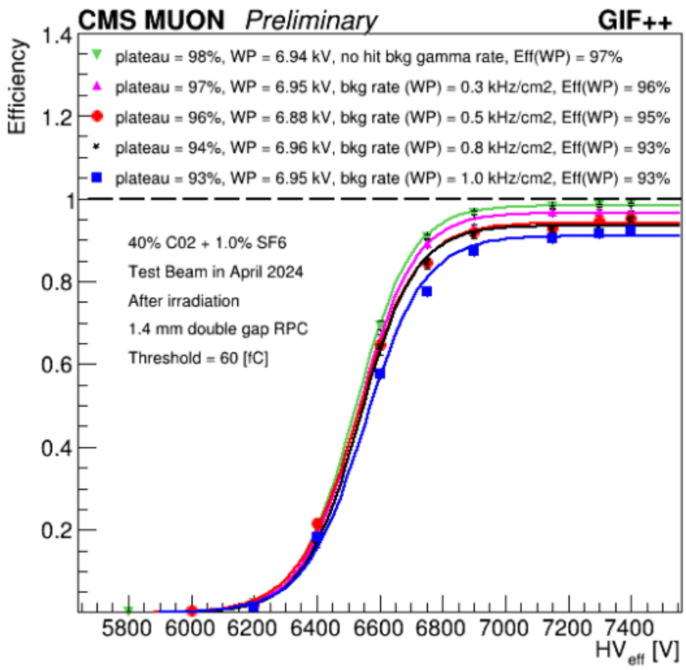

Figure 25 – Detection efficiency for the 40% CO_2 + 1.0% SF_6 gas mixture with sources: OFF (green), 10 (pink), 3.3 (red), 2.2 (black), and 1 (blue), during the 2024 beam test.

During the tests at GIF++, the RPC chambers are installed upstream in the test area, placed between two scintillator detectors that serve as the trigger system. The detection efficiency is determined by calculating, for each applied high voltage value, the ratio between the number of events recorded by the RPC chambers and the total number of events registered by the trigger system. The resulting points are then fitted with a sigmoid curve, characteristic of gaseous detectors operating in avalanche mode, reflecting the efficiency growth until saturation at high voltage regimes.

The efficiency curves (Figures 10–13) indicate that the mixture containing 30% CO_2 and 0.5% SF_6 provides the best balance between efficiency and low GWP, ensuring suitable performance for the demands of the HL-LHC.

5 Conclusions

In this work, a direct measurement of the invisible width of the Z boson was performed using public data provided by the *CERN Open Data portal*, corresponding to data-taking eras G and H of the year 2016, at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. The analysis focused on events in which the Z boson decays to neutrinos, inferred through missing transverse momentum (p_T^{miss}), in association with jets. The adopted method relied on extrapolation from a reference channel—the decay of the Z boson into charged leptons—and on the application of a simultaneous fit across multiple control regions, using the Markov Chain Monte Carlo (MCMC) sampling technique.

The resulting value for the invisible width is:

$$\Gamma(Z \to \nu\bar{\nu}) = 634{,}6^{+44{,}5}_{-43{,}3}\,\text{MeV},$$

or, approximately, $635 \pm 44\,\text{MeV}$. This result shows a deviation of about 2.4σ from the measurement reported by the CMS collaboration, which found $\Gamma = 523 \pm 16\,\text{MeV}$. This discrepancy can be largely attributed to the statistical limitations of the data available from the CMS Open Data, as well as to uncertainties in the transfer factor r_Z, which is affected by low Monte Carlo statistics in the double lepton control region.

In addition to the traditional cut-based approach, a complementary study was conducted using machine learning techniques to select signal region events. Specifically, *Graph Neural Networks* (GNNs) were employed, leveraging graph-based event representations to identify complex patterns in kinematic variables. The trained model achieved robust performance, with accuracy and F1-score exceeding 85%, demonstrating its ability to generalize to new data and efficiently identify signal-like events. Despite certain limitations in reproducing veto-related cuts, the GNN-based study proved to be a promising alternative to the cut-based selection strategy.

Despite its limitations, this analysis demonstrates the feasibility of performing sophisticated measurements using open data. In addition to validating a complete analysis workflow — from event selection, experimental corrections, and visualizations to the final statistical fit — this study highlights the pedagogical and scientific potential of CMS Open Data as a valuable tool for training and research in particle physics.

Additionally, this dissertation presents a practical contribution related to technical development within the CMS experiment, through participation in the characterization and testing of the iRPC detectors during the master's exchange internship conducted at CERN.

As future prospects, this analysis can be refined through the inclusion of additional datasets. The application of more advanced approaches, such as machine learning models — including more complex GNN architectures involving multiple reconstructed objects — represents a natural and promising extension for the study of invisible observables, for precision tests of the Standard Model, and for exploring possible signs of physics beyond the Standard Model using open data.

Bibliography

1 Wikimedia Commons contributors. *Complexo de Aceleradores do CERN*. 2023. <https://pt.m.wikipedia.org/wiki/Ficheiro:Cern-accelerator-complex.svg>. Acesso em: 15 maio 2025. Citado 2 vezes nas páginas 8 and 28.

2 CERN Courier. *CMS starts underground*. 2008. Acesso em: 15 maio 2025. Disponível em: <https://cerncourier.com/a/cms-starts-underground/>. Citado 2 vezes nas páginas 8 and 30.

3 CONNOR, P. L.; CONNOR, P. L. The large hadron collider and the compact muon solenoid. *Inclusive b Jet Production in Proton-Proton Collisions: Precision Measurement with the CMS experiment at the LHC at s= 13 TeV*, Springer, p. 65–93, 2019. Citado 2 vezes nas páginas 8 and 32.

4 ABBRESCIA, M. *The upgrade of the CMS muon system*. [S.l.], 2013. Citado 2 vezes nas páginas 8 and 33.

5 WALTENBERGER, W. *Development of vertex finding and vertex fitting algorithms for CMS*. [S.l.], 2004. Citado 2 vezes nas páginas 8 and 34.

6 PELLECCHIA, A. et al. The upgrade of the cms muon system for the high luminosity lhc. *Journal of Instrumentation*, IOP Publishing, v. 19, n. 02, p. C02077, 2024. Citado 2 vezes nas páginas 9 and 64.

7 KUMARI, P. et al. Improved-rpc for the cms muon system upgrade for the hl-lhc. *Journal of Instrumentation*, IOP Publishing, v. 15, n. 11, p. C11012, 2020. Citado 2 vezes nas páginas 9 and 65.

8 CMS collaboration. Precision measurement of the Z boson invisible width in pp collisions at $\sqrt{s} = 13\,\text{TeV}$. *arXiv preprint*, 2022. Citado 6 vezes nas páginas 10, 13, 37, 47, 60, and 63.

9 CERN Open Data Portal. *CERN Open Data Policy for LHC Experiments*. 2021. Accessed: 2025-01-20. Disponível em: <https://opendata.cern.ch/docs/cern-open-data-policy-for-lhc-experiments>. Citado na página 13.

10 HAGIWARA, K. et al. Review of particle physics: Particle data group. *Physical Review D*, American Institute of Physics, v. 66, n. 1 I, p. 100011–10001958, 2002. Citado na página 13.

11 SAMALAN, A. et al. Upgrade of the cms resistive plate chambers for the high luminosity lhc. *Journal of Instrumentation*, IOP Publishing, v. 17, n. 01, p. C01011, 2022. Citado 4 vezes nas páginas 14, 15, 64, and 65.

12 GRIFFITHS, D. *Introduction to elementary particles*. [S.l.]: John Wiley & Sons, 2020. Citado 5 vezes nas páginas 16, 19, 20, 21, and 22.

13 MANDL, F.; SHAW, G. *Quantum field theory*. [S.l.]: John Wiley & Sons, 2013. Citado 7 vezes nas páginas 16, 17, 18, 22, 23, 24, and 25.

14 BERTULANI, C. A. *Física Nuclear*. 2011. Citado na página 26.

15 COGOLLO, D. et al. Fermion family number and the z–z′ mixing in the 3-3-1 model with right-handed neutrinos. *Modern Physics Letters A*, World Scientific, v. 23, n. 40, p. 3405–3410, 2008. Citado na página 27.

16 COLLABORATION, C. et al. The cms experiment at the cern lhc. *Journal of instrumentation*, IOP Publishing Ltd., v. 3, n. August 2008, p. 1–334, 2008. Citado na página 28.

17 BRÜNING, O. et al. Lhc design report: Volume i, the lhc main ring. *CERN Reports*, CERN (European Organization for Nuclear Research), 2004. Citado 2 vezes nas páginas 28 and 29.

18 SIRUNYAN, A. M. et al. Precision luminosity measurement in proton–proton collisions at s= 13\, tev s= 13 tev in 2015 and 2016 at cms. *The European Physical Journal C*, Springer, v. 81, p. 1–42, 2021. Citado 2 vezes nas páginas 29 and 30.

19 ABBRESCIA, M. et al. The cms experiment at the cern lhc. *JOURNAL OF INSTRUMENTATION*, v. 3, 2008. Citado 4 vezes nas páginas 29, 30, 31, and 32.

20 COLLABORATION, C. et al. Particle-flow reconstruction and global event description with the cms detector. *Journal of Instrumentation*, IOP Publishing, v. 12, n. 10, p. P10003, 2017. Citado na página 34.

21 PIVARSKI, J.; CONTRIBUTORS. *uproot: A ROOT I/O in pure Python and Numpy*. 2024. <https://pypi.org/project/uproot/>. Version 5 (latest as of 2024). Citado na página 36.

22 PIVARSKI, J.; CONTRIBUTORS. *awkward: Manipulate nested, variable-sized data in Python*. 2024. <https://pypi.org/project/awkward/>. Version 2 (latest as of 2024). Citado na página 36.

23 SIRUNYAN, A. M. et al. Extraction and validation of a new set of cms pythia8 tunes from underlying-event measurements. *Eur. Phys. JC*, v. 80, n. 1, p. 4, 2020. Citado na página 37.

24 CMS Collaboration. *Info for MC production for Ultra Legacy Campaigns 2016, 2017, 2018*. 2020. <https://cms-pdmv.gitbook.io/project/mccontact/info-for-mc-production-for-ultra-legacy-campaigns-2016-2017-2018>. Acesso em: 17 de maio de 2025. Citado na página 37.

25 CMS. *Introduction to CMS Trigger - Getting trigger prescales - CMS Open Data Workshop 2021*. 2021. <https://cms-opendata-workshop.github.io/workshop2021-lesson-introtrigger/03_prescales/index.html>. Accessed: 2025-01-27. Citado na página 40.

26 CMS. *Triggers & Luminosity - Brilcalc Exercises - CMS Open Data Workshop 2024*. 2024. Acessado em: 27 janeiro 2025. Disponível em: <https://cms-opendata-workshop.github.io/workshop2024-lesson-triggers-lumi/instructor/05-challenge.html#brilcalc-exercises>. Citado na página 41.

27 CMS. *B-Tagging Recommendations for 80X*. 2025. Accessed: 2025-01-29. Disponível em: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation80X>. Citado na página 42.

28 SHLOMI, J.; BATTAGLIA, P.; VLIMANT, J.-R. Graph neural networks in particle physics. *Machine Learning: Science and Technology*, IOP Publishing, v. 2, n. 2, p. 021001, 2020. Citado na página 49.

29 THAIS, S. et al. Graph neural networks in particle physics: Implementations, innovations, and challenges. *arXiv preprint arXiv:2203.12852*, 2022. Citado na página 49.

30 CMS. *Modeling Backgrounds - CMS Open Data Workshop 2024*. 2024. <https://cms-opendata-workshop.github.io/workshop2024-lesson-background-modeling/instructor/index.html>. Accessed: 2025-01-28. Citado na página 54.

31 DOBBS, M.; LEFEBVRE, M. Unweighted event generation in hadronic wz production at order (α s). *Physical Review D*, APS, v. 63, n. 5, p. 053011, 2001. Citado na página 54.

32 FRIXIONE, S.; WEBBER, B. R. Matching nlo qcd computations and parton shower simulations. *Journal of High Energy Physics*, IOP Publishing, v. 2002, n. 06, p. 029, 2002. Citado na página 54.

33 DANZIGER, K.; HÖCHE, S.; SIEGERT, F. Reducing negative weights in monte carlo event generation with sherpa. *arXiv preprint arXiv:2110.15211*, 2021. Citado na página 54.

34 CMS. *Utilities for Accessing Pileup Information for Data*. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupJSONFileforData>. Accessed: 2025-01-28. Citado na página 55.

35 COLLABORATION, C. The phase-2 upgrade of the cms tracker. *CERN, Geneva, Tech. Rep. CERN-LHCC-2017-009. CMS-TDR-014*, 2017. Citado na página 64.

36 PINHEIRO, J. P. G. et al. Performance of new co2 based mixture in cms improved resistive plate chambers in hl-lhc environment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Elsevier, v. 1065, p. 169474, 2024. Citado na página 66.