Measurements on the production of a Z boson in association with a photon in proton-proton collisions and a search for anomalous neutral triple gauge couplings at the CERN Large Hadron Collider

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Mohammad Abrar Wadud

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Professor Roger Rusack

July, 2023

Acknowledgements

I began my journey in experimental particle physics knowing nothing about the tools of the trade. Nearly 8 years later, I now have measurements on the two of the most fundamental particles. This achievement would not have been possible without the unwavering support and guidance of my advisor, Professor Roger Rusack. The complexity of the CMS detector and the challenges associated with analyzing its data are truly staggering, and Roger has been by my side throughout the entire endeavor, providing invaluable guidance.

I extend my heartfelt gratitude to Dr. Shilpi Jain, with whom I collaborated most closely for the analysis presented in this thesis. Without her extensive professional experience, profound knowledge, and remarkable patience with me, reaching this point would have been difficult. I am deeply thankful for the opportunity to learn from her.

During my early days in CMS, Dr. Rajdeep Mohan Chattterjee served as a powerful motivator, inspiring me to uphold the highest standards and approach this field with a critical mindset.

I extend my deepest gratitude to my cherished fellow graduate students at the University, particularly Alex Gilman, Aritra De, Protyush Sahu, Natalia Calixto, and Rohith Saradhy, for their friendship and camaraderie.

Lastly, I want to take a moment to thank my family members, who reside 7,700 miles away. My father has been a constant source of love and inspiration, nurturing my passion for physics from the earliest days. I am profoundly grateful for his unwavering belief in me. I am grateful to my mother for her unconditional love throughout this journey.

Abstract

Within the framework of the Standard Model of particle physics, p-p collisions produce the final state of a photon and missing momentum from a Z boson solely through quark exchange in the t and u channels, since direct couplings between the neutral electroweak bosons Z and γ are not permitted at the tree level. However, the presence of new physics can induce the couplings $ZZ\gamma$ and $Z\gamma\gamma$, and enable the s channel, resulting in an enhancement in the rate of production. This enhancement would be particularly pronounced at the high-energy tail of the photon transverse momentum distribution. In this thesis, I present a search for these anomalous neutral triple gauge couplings (aNTGCs) under an effective vertex model and a measurement of the cross-section of the process $pp \rightarrow Z(\rightarrow \nu\nu) + \gamma$. The analysis is based on data collected by the CMS detector at the LHC between 2016 and 2018 at a center of mass energy of 13 TeV, corresponding to an integrated luminosity of 138 fb^{-1}. Moreover, two machine-learning techniques are introduced to enhance the sensitivity of the search. The first is an identification method designed specifically for photons with transverse momentum exceeding 200 GeV. The second is a tagger that suppresses background photons at high pseudorapidities of the detector that originate from Bremsstrahlung radiation emitted by muons within a diffuse halo produced by interactions of the proton beams with the beampipe material. The photon identification method improves the signal selection efficiency by 12% compared to a cut-based approach, while the beam halo tagger extends the pseudorapidity coverage of the photons in the search, allowing for the inclusion of photons in the CMS endcaps that were previously dominated by beam halo background. From the analysis of the distribution of the photon transverse momentum, limits on the strengths of the aNTGCs have been measured.

Blessed is he who learns how to engage in inquiry, with no impulse to harm his countrymen or to pursue wrongful actions, but perceives order of immortal and ageless nature, how it is structured.

— Euripides

Fragments from an unnamed play
Fifth century B.C.

Contents

Acknowledgements **i**

Abstract **iii**

Contents **v**

List of Tables **viii**

List of Figures **x**

1 Introduction **1**

2 Theory **4**

2.1 The Standard Model of particle physics 4

2.1.1 The Gauge Invariance Principle 7

2.1.2 The Higgs Field 9

2.2 Deficiencies of the Standard Model 12

2.3 Effective Field Theories 13

2.4 Anomalous Neutral Triple Gauge Couplings (aNTGCs) 15

2.5 Probing aNTGCs in *pp* collisions 17

3 The experimental apparatus **21**

3.1 The Large Hadron Collider 21

3.2 The Compact Muon Solenoid Experiment 24

3.2.1 Coordinate system 25

3.2.2 The Tracker 25

3.2.3 The Electromagnetic Calorimeter (ECAL) 27
3.2.4 The Hadron Calorimeter (HCAL) 30
3.2.5 The Muon System 31
3.2.6 The Triggering System 34

4 Event reconstruction 35
4.1 The particle flow algorithm 36
4.1.1 Tracks and vertices 37
4.1.2 Calorimeter clusters 38
4.1.3 The link algorithm 39
4.2 Muons 40
4.3 Photons and electrons 40
4.4 Jets 42
4.5 Missing transverse momentum 44

5 Photon identification 46
5.1 Shower Shape BDT 48
5.2 Isolation variables 54
5.3 Cut Optimization 57
5.4 Efficiency Scale Factors 59

6 Analysis of monophoton events 63
6.1 Analysis Strategy 64
6.2 Datasets 68
6.3 Simulations 68
6.4 Event Selection Criteria 69
6.5 Estimation of $e \to \gamma$ background 73
6.6 Estimation of $W + \gamma$ background 78
6.7 Estimation of Jet $\to \gamma$ background 81
6.8 Beam halo background 86
6.9 Estimation of spike background in the ECAL barrel 91
6.10 Statistical Analysis 94
6.10.1 Choice of bin boundaries 97

6.10.2 Signal modeling for cross-section measurement 97
6.10.3 aNTGC signal modeling 101
6.11 Systematic Uncertainties 103
6.12 Summary of estimated contributions 114

7 Results 120
7.1 Cross-section measurements 120
7.2 Differential cross-section measurements 122
7.3 Search for aNTGCs 124

8 Summary 127
8.1 Note on contributions 128

References 130

List of Tables

2.1 Electroweak charges of the fermions of the SM 6

5.1 Number of events in the training samples used for photon BDT training 52

5.2 Pileup corrections to ECAL isolation 56

5.3 Pileup corrections to tracker isolation 56

5.4 p_T corrections to ECAL isolation 56

5.5 Cuts on the discriminating variables for photon identification 57

6.1 List of simulation samples used in this analysis 69

6.2 List of criteria applied to select the events in the signal region 70

6.3 Dedicated fake ratios for dead regions of the pixel detector 77

6.4 List of criteria applied to select the events in the $W(\rightarrow \mu\nu) + \gamma$ control region (CR$_\mu$) and $W(\rightarrow e\nu) + \gamma$ control region (CR$_e$) 80

6.5 Size of the beam halo background with and without the use of the BDT-based tagger 91

6.6 Estimates of spike background from timing fit for the three years of data-taking 94

6.7 Truth-level criteria defining the fiducial region for the cross-section measurement 98

6.8 Simulated fiducial and non-fiducial events in the barrel and endcap regions 99

6.9 Relative impact of the experimental uncertainties on the signal efficiency. The uncertainties are given in percent. 108

6.10 Relative impact of the theoretical uncertainties on the predicted number of SM signal events. The uncertainties are given in percent. 112

6.11 Breakdown of event counts in the signal region in the barrel in ranges of photon p_T 116

6.12 Breakdown of event counts in the signal region in the endcaps in ranges of photon p_T 117
6.13 Breakdown of event counts in the $W(\to \mu\nu) + \gamma$ control region in the barrel in ranges of photon p_T 117
6.14 Breakdown of event counts in the $W(\to \mu\nu) + \gamma$ control region in the endcaps in ranges of photon p_T 118
6.15 Breakdown of event counts in the $W(\to e\nu) + \gamma$ control region in the barrel in ranges of photon p_T 118
6.16 Breakdown of event counts in the $W(\to e\nu) + \gamma$ control region in the endcaps in ranges of photon p_T 119
7.1 Measured and predicted cross-sections in the barrel, the endcaps and the combined fiducial phase spaces. The predicted cross-sections are calculated at NLO and NNLO in QCD. 122

List of Figures

2.1 Particle content of the Standard Model 6
2.2 Feynman diagram of beta decay 13
2.3 The anomalous vertices $Z\gamma\gamma$ and $ZZ\gamma$ 15
2.4 Anomalous vertices induced by a heavy fermion F 16
2.5 $Z(\nu\nu) + \gamma$ production in p-p collisions in the t/u-channels 17
2.6 $Z(\nu\nu) + \gamma$ production in p-p collisions in the s-channel 17
2.7 Excess in the number of $Z + \gamma$ events due to aNTGC for arbitrary integrated luminosity 18
2.8 Limits on the anomalous coupling parameters measured by the CMS collaboration using the final state of $Z(\rightarrow \nu\nu) + \gamma$ [1] 19
2.9 Limits on the anomalous coupling parameters measured by the ATLAS collaboration using the final state of $Z(\rightarrow \nu\nu) + \gamma$ [2] 19
3.1 An overview of the CERN accelerator complex. The protons are accelerated progressively in the order 50 MeV (LINAC 2) $\rightarrow$ 1.4 GeV (Booster) $\rightarrow$ 25 GeV (PS) $\rightarrow$ 450 GeV (SPS) $\rightarrow$ 6.5 TeV (LHC). 22
3.2 Integrated luminosity of p-p collisions versus time recorded by the CMS between 2016 and 2018 23
3.3 Overview of the CMS Detector 24
3.4 Coordinate system of the CMS detector 25
3.5 Schematic of the tracker shown as a vertical slice along the beam . . . 26
3.6 Schematic of a quadrant of the Electromagnetic Calorimeter (ECAL) shown as a vertical slice along the beam 28
3.7 Monitoring of crystal transparency with laser injection 29
3.8 Schematic of the Hadron Calorimeter 30

3.9 Schematic of the muon system 32
3.10 Layout of a DT cell, showing the electric field lines in the gas volume . . 32
3.11 Structure of the CSC chamber showing the anode and cathode grid . . . 33
3.12 Schematic of an RPC 33
4.1 Transverse slice of the CMS detector showing interactions of different particles with the sub-detectors 36
5.1 Deposits produced in the subdetectors by a jet and a genuine photon . . 46
5.2 Typical deposits in the ECAL produced by a genuine photon and a jet . 47
5.3 Photon ID scheme 48
5.4 Distributions of shower shape variables for genuine (prompt) and jet-induced (jet fake) photons 50
5.5 Distributions of shower shape variables for genuine (prompt) and jet-induced (jet fake) photons 51
5.6 ROC curves for the BDT model trained for 2017 data. The test samples are independent of the training samples, and they show similar performance. 53
5.7 BDT distributions for true (prompt) and jet-induced (fake) photons using the model trained for 2017 data. The overlap between signal and background distributions is lower here than that in any single shower shape variable. 53
5.8 Pileup and p_T corrections for photon isolations are estimated from 2D distributions in simulation of true photons. The points represent the 92nd percentile of the 1D isolation distributions in the x-bins 55
5.9 True and fake photon selection efficiency in simulated data as a function of photon p_T under our identification criteria (BDT+Iso). For comparison with purely cut-based methods, three other identification criteria are also shown (EGM Tight, EGM Medium and EGM Loose). 58
5.10 Fits of the invariant mass distributions of $Z \rightarrow ee$ pairs where the probe electron passes and fails the photon selection criteria. The probe electron is required have $0 < \eta < 0.6$ and 250 GeV $< p_T <$ 300 GeV. 59
5.11 Extrapolation uncertainties assigned to scale factors for photons simulated in $0 < \eta < 0.6$ and 250 GeV $< p_T <$ 300 GeV. 61

5.12 Scale factors to correct for mismodeling of photon selection efficiency in simulated data. The corrections are derived separately for different ranges of pseudorapidity (η_{SC}) and photon transverse momentum (p_T) 62
6.1 Overview of background sources 67
6.2 Fake ratio method for estimating $e \to \gamma$ background 73
6.3 Fit of the invariant mass distributions of regions A and B for the pseudorapidity range $|\eta| < 0.55$ corresponding data collected in 2018 76
6.4 Fake ratio for the 2017 data-taking period in bins of pseudorapidity . . 77
6.5 Construction of the joint likelihood function for the estimation of $W + \gamma$ background 79
6.6 Regions used to estimate the background from jets misidentified as photons 82
6.7 Example of a fit of the BDT score distribution of the photons observed in Region A. These photons are in the barrel of the ECAL in the 2017 data-taking period and have a p_T between 380 GeV and 430 GeV. . . . 83
6.8 Subtraction of contamination from genuine photons in the fake photon template. 84
6.9 Variation of the fake ratios with photon transverse momentum for the barrel region of the 2017 data-taking period 85
6.10 Bremsstrahlung radiation by beam halo muons in the ECAL barrel . . . 87
6.11 Lack of coverage of the cathode strip chambers (CSCs) limiting the identification of beam halo muons 88
6.12 Deposits produced by beam halo muons and photons from proton collisions 88
6.13 Distribution of the BDT score of simulated real photons (red) and photons from beam halo (blue). The BDT score of beam halo photons in real data in a control sample is also shown. 89
6.14 Beam halo background in the endcap signal region of data collected in 2017 91
6.15 Fit of the timing distribution of signal photons in the ECAL barrel . . . 92
6.16 Energy sharing between crystals in the ECAL described by eta wing . . 93
6.17 Response matrix for the measurement of the differential cross-section . . 101
6.18 Predicted number of events as a function of the coupling parameters (h_3^Z, h_4^Z) 103
6.19 Corrections to the SM signal to account for QCD NNLO effects 109

6.20 Corrections to the SM signal to account for electroweak NLO effects . . 110
6.21 Observed and predicted photon p_T distributions in the signal region in the barrel and endcaps, with breakdown of contributions from individual processes 114
6.22 Observed and predicted photon p_T distributions in the $W(\to e\nu) + \gamma$ control region in the barrel and endcaps, with breakdown of contributions from individual processes 115
6.23 Observed and predicted photon p_T distributions in the $W(\to \mu\nu) + \gamma$ control region in the barrel and endcaps, with breakdown of contributions from individual processes 115
7.1 The measured cross-section in separate photon pseudorapidity ranges. . 121
7.2 The measured differential cross-section in separate photon pseudorapidity ranges. Comparison is shown with theoretical predictions calculated at NNLO and NLO in QCD. 123
7.3 Scan of $-2\ln\lambda(h_3^V, h_4^V)$ over the aNTGC parameters. Best fit values and contours corresponding to 68%, 95% and 99% confidence levels are shown. 124
7.4 Observed and predicted distributions of photon transverse momentum at the best-fit values of (h_3^V, h_4^V) 125
7.5 One-dimensional scan of $-2\ln\lambda(h)$ over each aNTGC parameter. Confidence intervals at 68%, 95% and 99% confidence levels are shown. . . . 126

Chapter 1

Introduction

Particle physics is a reductionist approach to explain all physical phenomena in terms of fundamental interactions between fundamental particles. According to our current understanding, particles are excitations of quantum fields that span all of space, and their interactions are mediated by gauge bosons. This approach has given us the Standard Model (SM) of particle physics, which successfully explains the electromagnetic, weak, and strong interactions. The SM has been tested to a high precision and, following the discovery of the Higgs boson at the LHC in 2012, all of its predicted particles have been experimentally observed. However, the SM is not a complete description of nature, as it does not incorporate gravity, explain the nature of dark matter, address the origin of matter-antimatter asymmetry, or account for neutrino oscillation.

The electroweak sector of the SM consists of four gauge bosons: W^+, W^-, Z and the photon (γ). Probing the interactions between these bosons may provide insights into how the SM can be extended to incorporate the phenomena it does not explain. This thesis investigates the interactions between the Z boson and the photon, γ. We analyze proton-proton collisions at a center-of-mass energy of 13 TeV recorded by the CMS detector at the LHC to understand how the pair of bosons $Z + \gamma$ is produced in the final state of the collisions. The focus is specifically on collision events where the Z boson decays into a pair of neutrinos. Since neutrinos do not interact with the detector, the final state is identified in the detector only through the presence of a photon and missing transverse momentum. This final state is known as a *monophoton* event.

According to the SM, the $Z+\gamma$ final state is produced through a specific mechanism without direct interactions between the Z and γ. However, there are theories beyond the SM that predict the interaction of three of these bosons. In particular, the existence of the interactions $ZZ\gamma$ and $Z\gamma\gamma$ would affect the production rate of $pp \rightarrow Z+\gamma$. These interactions are known as *anomalous neutral triple gauge couplings* (aNTGCs). It is predicted that aNTGCs would manifest as an excess in the cross-section of $Z+\gamma$ production compared to the SM prediction. This excess would concentrate at the high-energy tail of the photon's transverse momentum distribution. Our analysis aims to examine this distribution and search for any compatible excess that could be attributed to an aNTGC.

To enhance the sensitivity of the search, we developed an algorithm based on machine-learning to increase the efficiency of identifying high-energy photons in the CMS detector. In addition, we developed an algorithm, also based on machine-learning, to reject a prominent background, known as *beam halo*, that is large in the vicinity of the proton beams. This allowed us to increase the geometric coverage of the photons selected for the study over previous previous searches conducted by the CMS collaboration in this final state [1].

In addition to the search for aNTGCs, we measure the cross-section of $pp \rightarrow Z(\rightarrow \nu\nu)+\gamma$ and the differential cross-section as a function of photon transverse momentum. This allows a comparison of the accuracy of perturbative calculations for theoretical cross-section predictions with observation.

This thesis is organized as follows:

- Chapter 2 introduces the theoretical elements of the thesis. It starts with a review of the SM and proceeds to show how effective field theories can be used to describe aNTGCs
- Chapter 3 describes the LHC and the components of the CMS detector used for the reconstruction of the final state particles.
- Chapter 4 outlines the algorithms employed by the CMS experiment to reconstruct the final-state particles using information from the subdetectors.
- Chapter 5 describes the machine-learning based algorithm we developed for this analysis to efficiently identify high-energy photons.

- Chapter 6 presents the main analysis. It describes how events are selected from the data, how the various background processes are measured, how the signal process is modeled, and the systematic uncertainties that affect the measurements.
- Chapter 7 presents the results of the measurements of the cross-section, the differential cross-section, and the search for aNTGCs.
- Chapter 8 summarizes the thesis and acknowledges contributions from the researchers involved in the analysis.

Chapter 2

Theory

This chapter provides an overview of the theoretical framework underlying the thesis. We begin by presenting an outline of the Standard Model (SM) of particle physics. Subsequently, we summarize the phenomena that the SM is unable to explain. We then show how new particles or interactions added in extensions of the SM can induce interactions among SM particles, which can be described using effective field theories at low energies in a model-independent way. Lastly, we discuss anomalous neutral triple gauge couplings, which represent effective interactions between the neutral gauge bosons of the SM. These effective interactions are the focus of this thesis.

2.1 The Standard Model of particle physics

The Standard Model (SM) [3] is the foundation of our current understanding of the behavior of particles and their interactions in the universe. The model assumes that particles are discrete excitations of quantum fields and that a general, renormalizable Lagrangian with these fields can be constructed based on a set of underlying symmetries. These symmetries are described by the local gauge transformation group

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$$

and govern three of the four fundamental forces: the strong, weak, and electromagnetic interactions. By imposing these gauge symmetries, bosonic fields are introduced, which serve as mediators for the interactions. The strong interaction is described by the gauge group $SU(3)_C$, also known as quantum chromodynamics (QCD), while the electroweak interaction is a unification of the electromagnetic and weak interactions and is represented by the group $SU(2)_L \otimes U(1)_Y$.

The matter sector of the SM consists of 12 fermions with a spin of 1/2. Among them, six are known as quarks, which possess two types of charge: color charge due to the $SU(3)_C$ group and electroweak charge due to the $SU(2)_L \otimes U(1)_Y$ group. The quarks are further divided into up-type (u, c, and t) and down-type (d, s, and b) based on their weak quantum numbers. The other six fermions, called leptons, have electroweak charge but no color charge. While three of the leptons (e, μ, and τ) are electrically charged, the other three (ν_e, ν_μ, and ν_τ) are neutral. The quarks and leptons are arranged into three generations, with the following structure:

$$\begin{array}{lcccc} & & & \textbf{Generation} \\ & & & 1 \quad\; 2 \quad\; 3 \\ \text{Leptons } (l) & : & \begin{pmatrix} \nu_l \\ l^- \end{pmatrix} = & \begin{pmatrix} \nu_e & \nu_\mu & \nu_\tau \\ e^- & \mu^- & \tau^- \end{pmatrix} \\ \text{Quarks } (q) & : & \begin{pmatrix} q_u \\ q_d \end{pmatrix} = & \begin{pmatrix} u & c & t \\ d & s & b \end{pmatrix} \end{array}$$

Here l and ν_l repesent leptons with and without electric charge, respectively, while q_u and q_d are the up and down type quarks, respectively.

The mediators of the strong and electroweak interactions have spin 1. They are:

- 8 massless gluons (g) for the strong interaction
- 3 massive gauge bosons (W^+, W^- and Z) for the weak interaction
- 1 massless photon (A) for the electromagnetic interaction

While the color symmetry, $SU(3)_C$, is conserved, the electroweak symmetry, $SU(2)_L \otimes U(1)_Y$, is broken spontaneously by the Brout-Englert-Higgs-Kibble (BEHK) mechanism. This generates mass terms for the massive vector bosons by the spin-0 Higgs field (Φ). The quantum excitation of the Higgs field in turn gives rise to the Higgs boson (h).

Fermion		Y	T_3	Q
Quarks (3 colors)	$\begin{pmatrix} q_u \\ q_d \end{pmatrix}_L$	+1/6	+1/2	+2/3
		+1/6	-1/2	-1/3
	u_R	+2/3	0	+2/3
	d_R	-1/3	0	-1/3
Leptons	$\begin{pmatrix} \nu \\ l \end{pmatrix}_L$	-1/2	+1/2	0
		-1/2	-1/2	-1
	l_R	-1	0	-1

Table 2.1: Electroweak charges of the fermions of the SM

The weak interaction is chiral, i.e. the left-handed (L) and right-handed (R) chiral fermionic fields transform differently under the $SU(2)_L$ group. Because of the chirality, each of the three fermion generations come with five different representations: q_L, q_{uR}, q_{dR}, l_L, and l_R (charges are listed in Table 2.1). The left-handed fermions transform as doublet, while the right-handed fields transform as singlets under the $SU(2)_L$ group. The quarks have three color degrees of freedom and transform as triplets under $SU(3)_C$ while the leptons remain singlets under $SU(3)_C$.

The particle content of the SM is summarized in Figure 2.1.

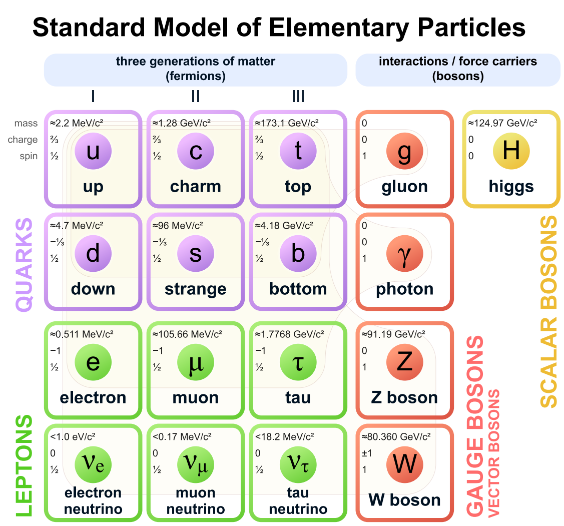


Figure 2.1: The particles of the Standard Model [4]

2.1.1 The Gauge Invariance Principle

The mediator fields of the interactions arise naturally as a result of the symmetries imposed on the fermionic fields. The classic example for demonstrating the idea is quantum electrodynamics (QED). Consider the Dirac lagrangian of a free fermionic field $\Psi(x)$:

$$\mathcal{L}_0 = i\bar{\Psi}(x)(\partial\!\!\!/ - m)\Psi(x)$$

$\mathcal{L}_0$ is invariant under the local $U(1)$ transformation

$$\Psi(x) \xrightarrow{U(1)} \Psi'(x) = \exp[iQ\theta(x)]\Psi(x)$$

if a new spin-1 field $A_\mu(x)$ is introduced such that the spacetime derivative ∂_μ is replaced with the *covariant derivative* $\mathcal{D}_\mu$

$$\partial_\mu \to \mathcal{D}_\mu = \partial_\mu - ieQA_\mu(x)$$

and an offset term is added to $U(1)$ transformation law for $A_\mu(x)$:

$$A_\mu(x) \xrightarrow{U(1)} A'_\mu(x) = A_\mu(x) + \frac{1}{e}\partial_\mu\theta(x)$$

The new spin-1 field $A_\mu(x)$ is realized as the photon. This demonstrates that the local $U(1)$ invariance of a fermionic Lagrangian requires a gauge field.

To allow $A_\mu(x)$ to be a freely propagating field, a kinetic term is also added to the Lagrangian :

$$\mathcal{L}_{\text{kin}} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the electromagnetic field strength tensor. The full QED Lagrangian is then

$$\mathcal{L}_{\text{QED}} = i\bar{\Psi}(x)(\mathcal{D}\!\!\!\!/ - m)\Psi(x) - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}$$

which gives the Maxwell equations:

$$\partial_\mu F^{\mu\nu} = j^\nu \text{ where } j^\nu = -eQ\bar{\Psi}\gamma^\nu\Psi$$

The electromagnetic (EM) current j^ν couples to the photon field A^ν via the interaction term in the QED Lagrangian:

$$\mathcal{L}_{\text{int}} = A_\nu j^\nu = eQA_\nu(x)\bar{\Psi}(x)\gamma^\nu\Psi(x)$$

To implement the full $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ symmetry group, three sets of gauge fields are introduced:

- 1 field B_μ with generator Y (hypercharge) for the $U(1)_Y$ group
- 3 fields W_μ^a with generators $T^a = \sigma^a/2$ for the $SU(2)_L$ group, where $a = 1, 2, 3$ and σ^a are the Pauli matrices
- 8 fields G_μ^c with generators $\mathbb{T}^a = \frac{\lambda^a}{2}$ for the $SU(3)$ group, where $a = 1, ..., 8$ and λ^a are the Gell-Mann matrices
- the generators Y, T^a and $\mathbb{T}^a$ follow the Lie algebra:

$$\begin{aligned}
&[\mathbb{T}^a, \mathbb{T}^b] = if^{abc}\mathbb{T}^c \\
&[T^a, T^b] = i\epsilon^{abc}T^c \\
&[\mathbb{T}^a, T^b] = [\mathbb{T}^a, Y] = [T^a, Y] = 0
\end{aligned}$$

The gauge invariant fermionic matter Lagrangian is then given by

$$\mathcal{L}_{\text{fermions}} = i\bar{q}_L \not{\mathcal{D}} q_L + i\bar{l}_L \not{\mathcal{D}}^{(l_L)} l_L + i\bar{q}_{uR} \not{\mathcal{D}}^{(q_R)} q_{uR} + i\bar{q}_{dR} \not{\mathcal{D}}^{(q_R)} q_{dR} + i\bar{l}_R \not{\mathcal{D}}^{(l_R)} l_R$$

where

- q_L and l_L are the left-handed quark and lepton doublets
- q_{uR}, q_{dR} and l_R are the right-handed (up/down) quark and lepton singlets
- $\mathcal{D}_\mu = \partial_\mu - ig'YB_\mu - ig\frac{1}{2}\sigma^a W_\mu^a - ig_s\frac{1}{2}\lambda^a G_\mu^a$
- $\mathcal{D}_\mu^{(l_L)} = \partial_\mu - ig'YB_\mu - ig\frac{1}{2}\sigma^a W_\mu^a$
- $\mathcal{D}_\mu^{(q_R)} = \partial_\mu - ig'YB_\mu - ig_s\frac{1}{2}\lambda^a G_\mu^a$

- $\mathcal{D}_\mu^{(l_R)} = \partial_\mu - ig'YB_\mu$
- g' , g and g_s are the coupling constants associated with $U(1)_Y$, $SU(2)_L$ and $SU(3)_C$ respectively

As with the QED case, the Lagrangian for the kinematics of the gauge fields is:

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4g'^2}B^{\mu\nu}B_{\mu\nu} - \frac{1}{4g^2}W^{a,\mu\nu}W^a_{\mu\nu} - \frac{1}{4g_s^2}G^{a,\mu\nu}G^a_{\mu\nu}$$

and the field strength tensor are:

- $G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + f^{abc}G^b_\mu G^c_\nu$
- $W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu + \epsilon^{abc}W^b_\mu W^c_\nu$
- $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$

2.1.2 The Higgs Field

We note that the gauge fields have no mass term in this formulation since such a term would violate the required gauge symmetry. Fermion masses cannot be added either as this would involve mixing of left and right chiral states that also break the gauge symmetry. The Brout-Englert-Higgs-Kibble mechanism addresses the problem by introducing a scalar doublet under $SU(2)$:

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix}$$

The Lagrangian for the Higgs field preserving gauge invariance is given by:

$$\mathcal{L}_{\text{Higgs}} = (\mathcal{D}^{l_L}\Phi)^\dagger(\mathcal{D}^{l_L}\Phi) - V(\Phi)$$

where:

$$V(\Phi) = -\mu^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2$$

The Higgs potential $V(\Phi)$ is minimized when

$$\langle\Phi^\dagger\Phi\rangle = \frac{\mu^2}{2\lambda} = \frac{v^2}{2}$$

where $v = \mu/\sqrt{\lambda}$ is the vacuum expectation value (VEV). By performing a rotation, the field can be written as an excitation h around the minimum of the field:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix}$$

The excitation h is a physical degree of freedom, which appears in the Lagrangian as the Higgs boson with mass:

$$m_H = \sqrt{2\lambda} v$$

At the minimum, the $SU(2)_L \otimes U(1)_Y$ symmetry is spontaneously broken and the Lagrangian respects only $U(1)_{EM}$ symmetry with generator $Q = T_3 + Y$. In addition, the kinetic term $(\mathcal{D}^{l_L}\Phi)^\dagger(\mathcal{D}^{l_L}\Phi)$ of the Higgs field gives the mass terms of the physical gauge bosons $W^\pm$ and Z:

$$\mathcal{L}_{\text{mass}} = \frac{1}{4} g^2 v^2 W^{\pm\mu} W^{\mp}_\mu + \frac{1}{8}(g^2 + g'^2) v^2 Z^\mu Z_\mu + 0 \times (A^\mu A_\mu)$$

where the gauge fields of the $SU(2)_L \otimes U(1)_Y$ symmetry have mixed to form physical boson fields W^+, W^-, Z and A:

$$W^\pm = \frac{1}{\sqrt{2}}(W_1 \mp iW_2)$$
$$Z = W_3 \cos\theta_W - B \sin\theta_W$$
$$A = W_3 \sin\theta_W + B \cos\theta_W$$
$$\tan\theta_W = \frac{g'}{g}$$

The parameter θ_W, called the *weak mixing angle*, is a rotation angle from the *interaction basis*, where the fields have the transformation properties under the full gauge symmetry, into the *mass basis* of the physically observable vector bosons. The masses acquired are:

$$m_W = \frac{1}{2} v g$$
$$m_Z = \frac{1}{2} v \sqrt{g^2 + g'^2}$$
$$m_A = 0$$

The masses are related to the mixing angle by:

$$\cos^2 \theta_W = \frac{m_W^2}{m_Z^2}$$

Lastly, the leptons and quarks are given masses by adding Yukawa terms to the Lagrangian:

$$\mathcal{L}_{\text{Yukawa}} = -Y_{ij}^u \bar{q}_{Li} q_{uRj} \tilde{\Phi} - Y_{ij}^d \bar{q}_{Li} q_{dRj} \Phi - Y_{ij}^l \bar{l}_{Li} l_{Rj} \Phi + h.c.$$

where $\tilde{\Phi} = i\sigma_2 \Phi^*$ and Y_{ij} are 3×3 Yukawa matrices, which can be chosen in diagonal basis:

$$Y^u = \text{diag}(y_u, y_c, y_t)$$
$$Y^d = \text{diag}(y_d, y_s, y_b)$$
$$Y^l = \text{diag}(y_e, y_\mu, y_\tau)$$

After spontaneous breaking of the gauge symmetry, the Yukawa Lagrangian generates the mass terms for the fermions:

$$\mathcal{L}_{\text{Yukawa}} = -\frac{y_q v}{\sqrt{2}} \bar{q}_L q_R - \frac{y_l v}{\sqrt{2}} \bar{l}_L l_R$$

which gives the fermions their masses:

$$m_q = \frac{y_q v}{\sqrt{2}}, \quad m_l = \frac{y_l v}{\sqrt{2}}$$

Combining the Lagrangians for the fermions, gauge bosons and the Higgs field, the full Lagrangian of the SM is:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{fermions}} + \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}} \tag{2.1}$$

2.2 Deficiencies of the Standard Model

The SM neatly incorporates the electroweak and the strong interactions under the $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ gauge group. Its theoretical predictions for a wide range of processes have been tested at collider experiments with very high accuracies. Some of these predictions are the weak neutral current; the existence and masses of the W and Z bosons; the existence of the fermions τ, c, b, and t; and the last milestone, the Higgs boson, whose discovery was announced at the LHC in 2012. Despite the success of the theory, the SM cannot be a complete description of nature as it is unable to explain several observations, some of which are:

- **Matter-antimatter asymmetry**: The universe appears to be dominated by matter and, to date, astronomical observations have found no large traces of antimatter. While the weak interaction violates CP symmetry, it is insufficient to explain this large asymmetry.

- **Neutrino mass** The SM assumes that neutrinos are massless, but observations of flavor oscillations indicate that neutrinos must have a small non-zero mass through mixing of electroweak eigenstates. One solution, known as the seesaw mechanism, introduces mass to neutrinos by incorporating heavy Majorana neutrinos, whose masses are inversely related to the light neutrino masses. However, we have no experimental evidence for the existence of Majorana neutrinos.

- **Dark matter and dark energy**: Astronomical evidence indicates that only 5% of the universe is ordinary matter. Observations of galactic rotation profiles show a significant amount of weakly interacting matter called dark matter, comprising 24% of the universe. The remaining 71%, responsible for the universe's accelerating expansion, is known as dark energy. Neither of these phenomena is explained by the SM.

2.3 Effective Field Theories

Several models have been suggested to tackle the issues with the SM. These models add new particles and interactions, commonly referred to as "new physics" or NP, to the SM's particle spectrum. However, experimental findings have not yet detected any new resonances beyond the SM. One possible reason for the null result is that our current experiments do not have the energy necessary to generate NP particles. Nevertheless, there may still be indications of NP at present experiment energies. If NP particles did exist, they would modify the strengths of interactions among SM particles and/or introduce anomalous interactions not allowed by the SM. Measuring the interactions between SM particles with high precision might reveal signs of new physics. A deviation from the predictions of the SM can be interpreted under an *effective field theory* (EFT) that generalizes the effects of NP at low energies in a model-independent way. EFTs describe interactions at energies below a chosen scale while disregarding the dynamics of NP that have a significant impact only at a higher energy scale. There are two approaches to constructing an EFT: top-down and bottom-up. In the top-down approach, the operators in the Lagrangian of a known theory are approximated to effective low-energy operators by integrating out the heavy scale fields, whose dynamics have negligible effect. The operators of the effective interactions and the associated coupling constants encode the effects of NP at low energies. In contrast, the bottom-up approach assumes that the correct high energy theory is unknown. Therefore, known particle fields are used to create higher-dimensional operators respecting a set of symmetries to model the effects of new physics in a model-independent manner.

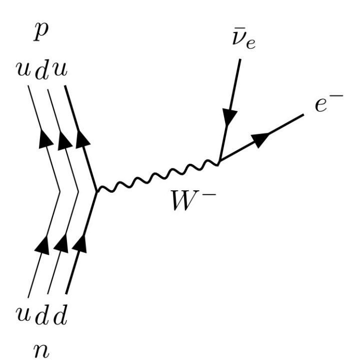


Figure 2.2: Feynman diagram of beta decay

Fermi's theory of weak interactions provides an example where the SM theory of the W boson's interactions with quarks can be approximated by a 4-point contact interaction. Fermi's theory was formulated to describe beta decay, $n(udd) \to p(uud) + e^- + \bar{\nu}_e$, at a time when the existence of the weak gauge bosons was not known. We can show that this theory is a low energy approximation of an interaction mediated by an off-shell W^- boson (Figure 2.2). Consider the SM tree-level amplitude for the interaction [5]:

$$\mathcal{M} = -i\frac{g^2}{2}V_{ud}(\bar{u}\gamma^\mu P_L d)\left(\frac{g_{\mu\nu} - \frac{p_\mu p_\nu}{m_W^2}}{p^2 - m_W^2}\right)(\bar{e}\gamma^\nu P_L \nu_e)$$

Here g is the weak coupling, V_{ud} is the ud CKM matrix element, $P_L = (1 - \gamma^5)/2$, and p^μ is the 4-momentum mediated by the off-shell W^- boson. Since the process is at low energy, i.e. $p^2 \ll M_W^2$, the propagator factor can be approximated as:

$$\frac{g_{\mu\nu} - \frac{p_\mu p_\nu}{m_W^2}}{p^2 - m_W^2} \approx -\frac{g_{\mu\nu}}{m_W^2}$$

The interaction amplitude is then

$$\mathcal{M} = -\frac{ig^2}{2m_W^2}V_{ud}(\bar{u}\gamma^\mu P_L d)(\bar{e}\gamma_\mu P_L \nu_e)$$

which has the same form as the Fermi 4-point contact interaction amplitude given by

$$\mathcal{M}_{\text{Fermi}} = -\frac{4iG_F}{\sqrt{2}}V_{ud}(\bar{u}\gamma^\mu P_L d)(\bar{e}\gamma_\mu P_L \nu_e)$$

Hence, the electroweak theory is encoded as a low energy limit in the Fermi coupling constant G_F, which is related to the weak coupling g by:

$$\frac{G_F}{\sqrt{2}} = \frac{g^2}{8m_W^2}$$

This is an example of the top-down approach to building an effective field theory, i.e. the Fermi 4-point interaction, from the full electroweak theory in the SM.

In the bottom-up approach [6], where the correct theory at high energies is unknown, we first note that the operators in the SM Lagrangian all have mass dimensionality of four or less. These terms are required to respect the gauge symmetry $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$. However, there are higher dimensional operators that still respect the gauge symmetry. These higher dimensional effective operators are constructed with the SM fields and are suppressed by an energy cut-off, so that the SM is recovered at low energies. Here EFT acts as a bridge between heavy scale NP and low energy phenomena (i.e. the SM) without considering the internal structure of the NP. In general, the EFT Lagrangian is taken as:

$$\mathcal{L}_{\mathrm{EFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_{d \geq 4} \sum_{i} \frac{c_i^{(d)}}{\Lambda^{(d-4)}} \mathcal{O}_i^{(d)}$$

where the operators $\mathcal{O}_i^{(d)}$ are built from the SM field and have mass dimensionality d. The coefficients $c_i^{(d)}$, known as the Wilson coeffcients, are the coupling strengths that encode the effects of heavy scale NP and Λ is the energy cut-off scale of the NP. As we go higher in the dimensionality of the operators, their effect becomes smaller due to the suppression factor $1/\Lambda^{(d-4)}$.

2.4 Anomalous Neutral Triple Gauge Couplings (aNTGCs)

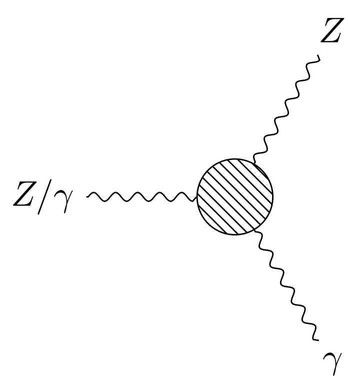


Figure 2.3: The anomalous vertices $Z\gamma\gamma$ and $ZZ\gamma$

In this thesis, we use EFT to study the possible interactions between the neutral gauge bosons of the electroweak sector, focusing specifically on the vertices $Z\gamma\gamma$ and $ZZ\gamma$ (Figure 2.3). These couplings are called *anomalous neutral triple gauge couplings* (aNTGCs).

In the Lagrangian of the SM, there are no terms to allow such interactions at the tree level. However, NP could induce such an effective vertex through a heavy fermionic loop as shown in Figure 2.4.

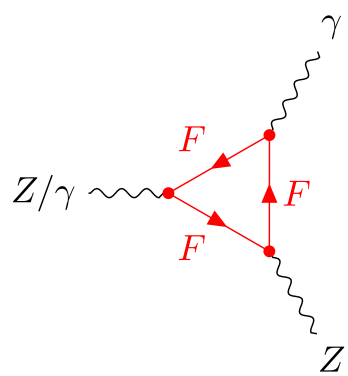


Figure 2.4: Anomalous vertices induced by a heavy fermion F

An EFT has been proposed in [7] to describe such anomalous couplings. In this approach, the symmetry requirement $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ is relaxed to capture more generalized NP effects. Requiring only $U(1)_{\mathrm{EM}}$ symmetry, the EFT Lagrangian with dim-6 and dim-8 operators describing $Z\gamma\gamma$ and $ZZ\gamma$ couplings is given by:

$$\begin{aligned} \mathscr{L}_{ZV\gamma} = -\frac{e}{m_Z^2} \Bigg[\!\!\Bigg[& \Big[h_1^\gamma(\partial^\sigma F_{\sigma\mu}) + h_1^Z(\partial^\sigma Z_{\sigma\mu})\Big] Z_\beta F^{\mu\beta} \\ & + \Big[h_3^\gamma(\partial_\sigma F^{\sigma\rho}) + h_3^Z(\partial_\sigma Z^{\sigma\rho})\Big] Z^\alpha \widetilde{F}_{\rho\alpha} \\ + \Big\{ \frac{h_2^\gamma}{m_Z^2}\Big[\partial_\alpha\partial_\beta\partial^\rho F_{\rho\mu}\Big] & + \frac{h_2^Z}{m_Z^2}\Big[\partial_\alpha\partial_\beta(\Box + m_Z^2)Z_\mu\Big]\Big\} Z^\alpha F^{\mu\beta} \\ - \Big\{ \frac{h_4^\gamma}{2m_Z^2}\Big[\Box\partial^\sigma F^{\rho\alpha}\Big] & + \frac{h_4^Z}{2m_Z^2}\Big[(\Box + m_Z^2)\partial^\sigma Z^{\rho\alpha}\Big]\Big\} Z_\sigma \widetilde{F}_{\rho\alpha} \Bigg]\!\!\Bigg] \end{aligned} \tag{2.2}$$

where F and Z are the bosonic fields (γ and Z), $V = Z/\gamma$, and h_i^V are the anomalous couplings for the $ZV\gamma$ vertex. The interactions associated with the couplings h_1^V and h_2^V are CP-odd while those for h_3^V and h_4^V are CP-even. By constructing a Lagrangian for interactions between the Z and γ bosons respecting only $U(1)_{\mathrm{EM}}$ symmetry, we can capture the effects induced by heavy NP as well as potential violation of the electroweak symmetry, $SU(2)_L \otimes U(1)_Y$.

2.5 Probing aNTGCs in *pp* collisions

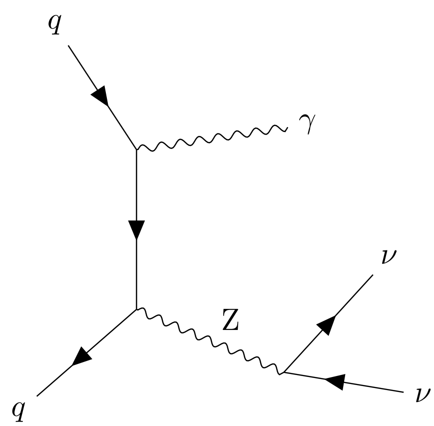


Figure 2.5: $Z(\nu\nu) + \gamma$ production in p-p collisions in the t/u-channels

If the anomalous couplings $ZZ\gamma$ and $Z\gamma\gamma$ existed, their presence may be seen in the measurements of the cross-sections of certain processes in p-p collisions at the LHC. In particular, the process $pp \to Z + \gamma$ is sensitive to these couplings. Under the SM, $Z + \gamma$ is produced at leading order only by the t and u channels (Figure 2.5). However, if anomalous couplings are were present, the production of $Z + \gamma$ can also occur by the s channel (Figure 2.6).

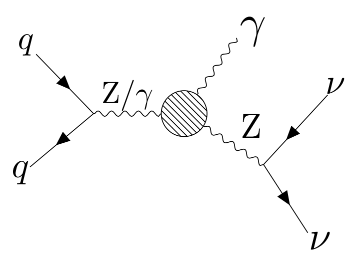


Figure 2.6: $Z(\nu\nu) + \gamma$ production in p-p collisions in the s-channel

As a result of the additional Feynman diagram, the amplitude $\mathcal{M}_{\text{aNTGC}}$ due to aNTGC would add to the SM amplitude,$\mathcal{M}_{\text{SM}}$, giving a total amplitude of:

$$\mathcal{M}_{\text{total}} = \mathcal{M}_{\text{SM}} + \mathcal{M}_{\text{aNTGC}}$$

The aNTGC amplitude $\mathcal{M}_{\text{aNTGC}}$ is linear in the coupling parameters h_i^V and can be written as:

$$\mathcal{M}_{\text{aNTGC}} = \sum_{V=Z,\gamma} \sum_i h_i^V \mathcal{M}_{\text{i}}^V$$

This would modify the cross-section of the process as:

$$\sigma_{\text{total}} = \sigma_{\text{SM}} + \sigma_{\text{int}} + \sigma_{\text{quad}}$$

Here σ_{SM} is the SM cross-section, σ_{int} represents the interference between SM and aNTGC amplitudes, and σ_{quad} is the square term due to aNTGC. Note that σ_{int} is linear in the coupling parameters while σ_{quad} is quadratic in these parameters.

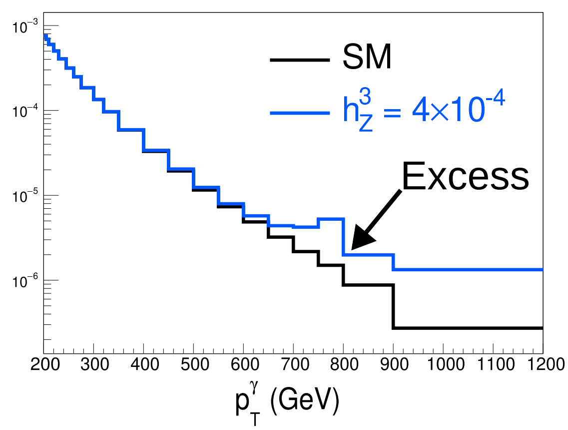
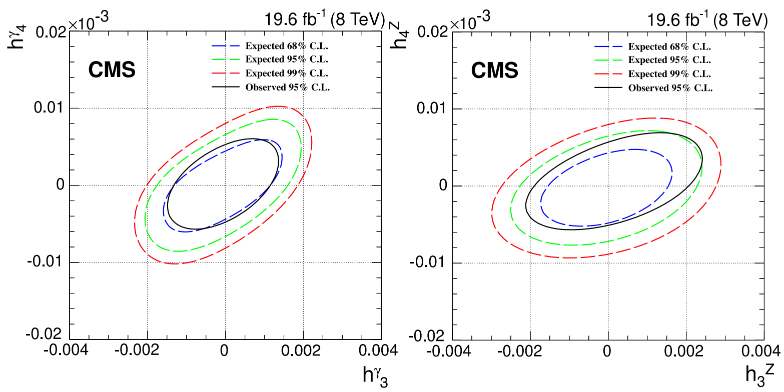


Figure 2.7: Excess in the number of $Z + \gamma$ events due to aNTGC for arbitrary integrated luminosity

The additional contributions from aNTGCs would lead to a deviation in the production cross-section with respect to the prediction of the SM. In particular, an excess is expected at large values of the transverse momentum (p_T^γ) of the photon, as shown in Figure 2.7. A search for these couplings can thus be performed by studying the distribution of the photon transverse momentum in $pp \to Z + \gamma$ events in collider experiments.

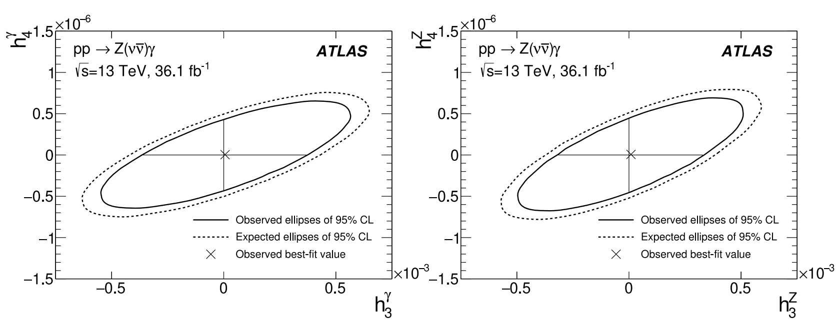
Searches for the anomalous couplings $Z\gamma\gamma$ and $ZZ\gamma$ in the final state of $Z(\to \nu\nu)+\gamma$ in hadron collisions were first conducted by the DØ collaboration [8], and subsequently by the CDF [9, 10], CMS [11, 1] and ATLAS collaborations [12, 2]. In addition, searches have also been peformed in the leptonic final state, $Z(\to ll) + \gamma$ [8, 13, 9, 10, 14, 1].

Figure 2.8 shows the limits measured by the CMS collaboration using LHC Run I *p*-*p* collision data at $\sqrt{s} = 8$ TeV. The tighest limits have been measured by the ATLAS collobaration using LHC Run II *p*-*p* collision data at $\sqrt{s} = 13$ TeV (shown in Figure 2.9).



Coupling	$\sqrt{s} = 8$ TeV	$\sqrt{s} = 7$ TeV
h_3^Z	[-1.5, 1.6]$\times 10^{-3}$	[-2.7, 2.7]$\times 10^{-3}$
h_4^Z	[-3.9, 4.5]$\times 10^{-6}$	[-1.3, 1.3]$\times 10^{-5}$
h_3^γ	[-1.1, 0.9]$\times 10^{-3}$	[-2.9, 2.9]$\times 10^{-3}$
h_4^γ	[-3.8, 4.3]$\times 10^{-6}$	[-1.5, 1.5]$\times 10^{-5}$

Figure 2.8: Limits on the anomalous coupling parameters measured by the CMS collaboration using the final state of $Z(\to \nu\nu) + \gamma$ [1]



Parameter	Limit 95% CL	
	Measured	Expected
h_3^γ	$(-3.7\times10^{-4}, 3.7\times10^{-4})$	$(-4.2\times10^{-4}, 4.3\times10^{-4})$
h_3^Z	$(-3.2\times10^{-4}, 3.3\times10^{-4})$	$(-3.8\times10^{-4}, 3.8\times10^{-4})$
h_4^γ	$(-4.4\times10^{-7}, 4.3\times10^{-7})$	$(-5.1\times10^{-7}, 5.0\times10^{-7})$
h_4^Z	$(-4.5\times10^{-7}, 4.4\times10^{-7})$	$(-5.3\times10^{-7}, 5.1\times10^{-7})$

Figure 2.9: Limits on the anomalous coupling parameters measured by the ATLAS collaboration using the final state of $Z(\to \nu\nu) + \gamma$ [2]

In this thesis, we use the invisible decay channel of the Z boson to search for these anomalous couplings with the parameters (h_3^V, h_4^V) using LHC Run II data at $\sqrt{s} = 13$ TeV collected by the CMS detector with a luminosity of 138 fb^{-1}. The invisible decay of the Z (branching fraction = 20%) is the optimal channel to detect these aNTGCs as the leptonic channel has a lower branching fraction (10.1%) while the hadronic channel (branching fraction =69.9%) has a poor sensitivity to aNTGC due to large background events that arise from the $pp \rightarrow \gamma +$ jets process. In addition, we consider only the CP-even couplings (h_3^V, h_4^V) as it has been shown in [15] that the sensitivities and cross-sections are nearly identical for the same values of $h_{3,4}^V$ and $h_{1,2}^V$. For this search, we study p-p collision events producing a photon with a large transverse momentum (> 225 GeV) and a large missing transverse momentum (> 200 GeV).

Chapter 3

The experimental apparatus

The measurements presented in this thesis are based on data collected in proton-proton collisions at a center-of-mass energy of 13 TeV. The experimental setup consists of the Large Hadron Collider (LHC), which produces the proton beams for collision, and the Compact Muon Solenoid (CMS) detector, which measures the outgoing products to study the properties of the fundamental interactions within the collisions. This chapter provides an overview of the LHC and the key aspects of the CMS detector's subsystems that enable the reconstructing of particles.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [16] is a 27 km two-ring superconducting proton-proton collider located at CERN, spanning the border between France and Switzerland. The LHC was built to study the nature of electroweak symmetry breaking and to search for beyond-the-standard-model phenomena at the TeV scale, including supersymmetry, extra dimensions, and dark matter. Figure 3.1 shows the layout of the CERN accelerator complex.

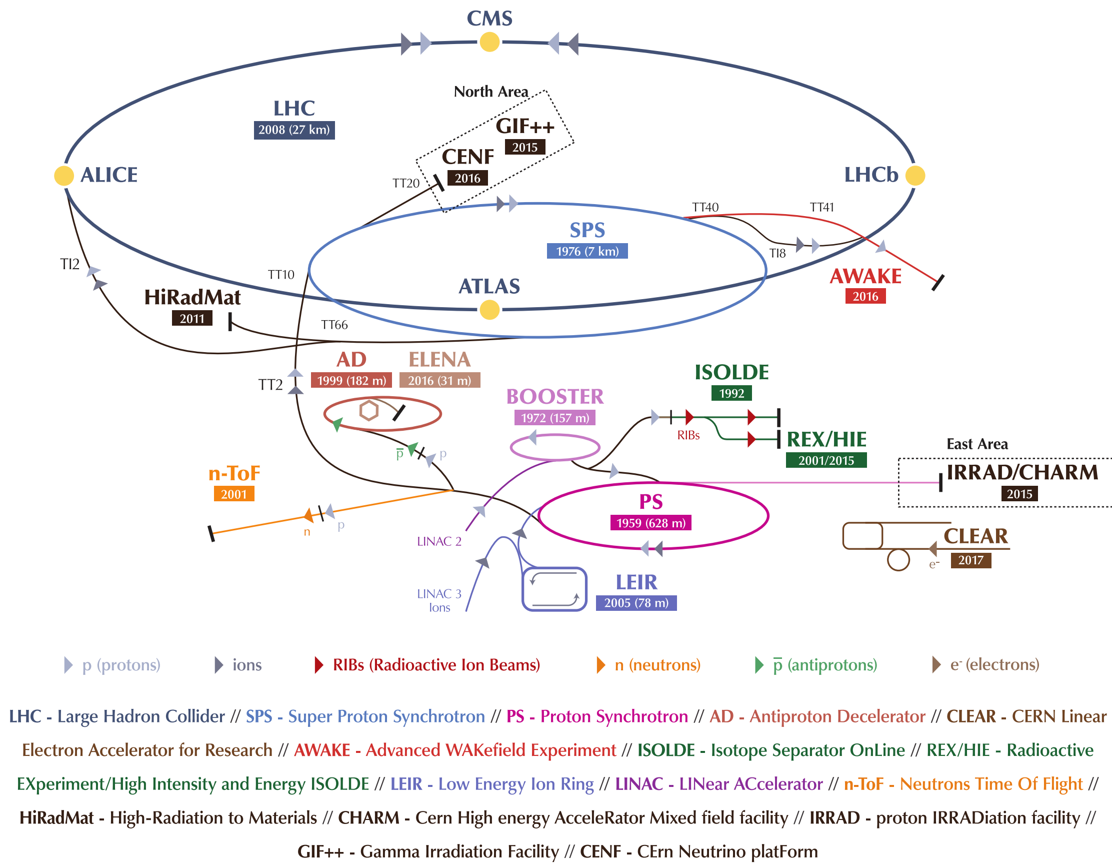


Figure 3.1: An overview of the CERN accelerator complex. The protons are accelerated progressively in the order 50 MeV (LINAC 2) → 1.4 GeV (Booster) → 25 GeV (PS) → 450 GeV (SPS) → 6.5 TeV (LHC).

The protons are accelerated to a center-of-mass energy of 13 TeV through a sequence of stages. The sequence begins at LINAC 2, where hydrogen gas is ionized in an electric field and the resulting protons are linearly accelerated to an energy of 50 MeV. Subsequently, these protons are injected into the Proton Synchrotron Booster (PSB), a 25 m radius ring that uses radio frequency pulses to provide a *kick* to the protons during each revolution, boosting their energy to 1.4 GeV. Once the protons have achieved their maximum energy in the PSB, they are transferred to the Proton Synchrotron (PS), a larger 100 m radius ring. In the PS, the protons are accelerated further to an energy of 25 GeV. After reaching maximum energy in the PS, the protons enter the Super Proton Synchrotron (SPS), a 1 km radius ring that raises their energy to 450 GeV. Finally, the protons are injected into the main LHC ring, where they undergo additional acceleration to reach their final energy of 6.5 TeV. The beams are kept in a circular trajectory by 1232 superconductive dipole magnets and focused by 392 quadrupole magnets distributed along the collider. The radio frequency pulses applied to the protons are tuned to arrange the protons in discrete bunches with a temporal spacing of 25 ns between the bunches, providing a bunch crossing rate of 40 MHz. At full design performance, there are about 2800 bunches, with approximately 1.5×10^{11} protons per bunch. Collision of the beams take place at four interaction points along the ring, where four main detectors, including the CMS, are installed. The number of events for a given process, $pp \to X$, is determined from the cross-section of the process, σ_X, and the instantaneous luminosity, $\mathcal{L}(t)$, of the collisions as:

$$N(pp \to X) = \sigma_X \cdot \int \mathcal{L}(t)\ \mathrm{d}t$$

The instantaneous luminosity is measured by multiple *luminometers* placed within the CMS detector [17]. From 2016 to 2018, the CMS detector recorded a total integrated luminosity of 138 fb^{-1} (Figure 3.2) [17, 18, 19].

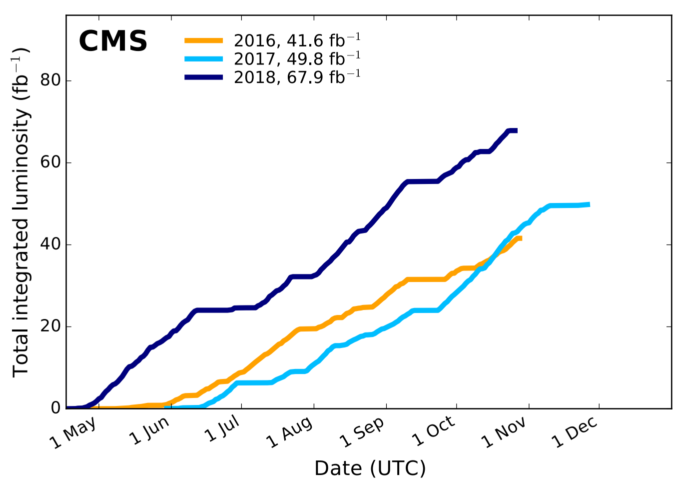


Figure 3.2: Integrated luminosity of p-p collisions versus time recorded by the CMS between 2016 and 2018

3.2 The Compact Muon Solenoid Experiment

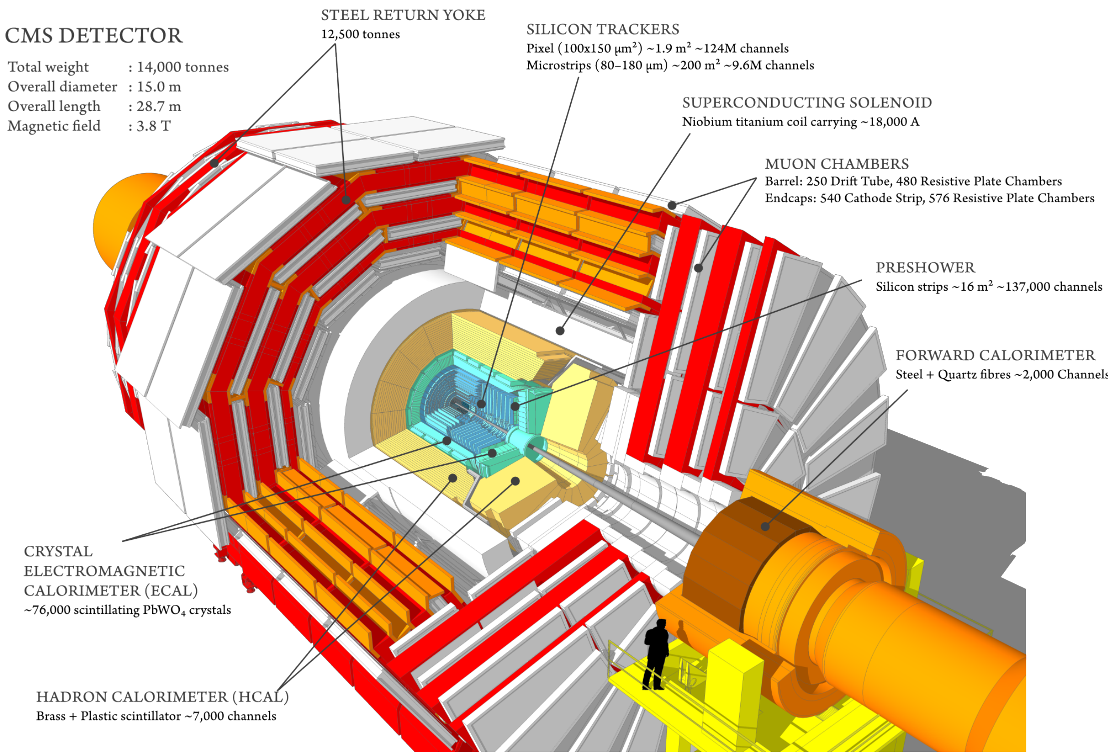


Figure 3.3: Overview of the CMS Detector

The Compact Muon Solenoid (CMS) [20] is one of the two multi-purpose detectors at the LHC. It is located 100 m underground in Cessy, France. Its primary purpose is to explore fundamental interactions at the TeV scale. The detector is designed to achieve high spatial resolution, precise energy measurements, and efficient particle identification. The detector has a cylindrical structure consisting of multiple layers of subdetectors (Figure 3.3). It has a large superconducting solenoid magnet (12.5 m long and 6 m in diameter) that generates a magnetic field of 3.8 T parallel to the beams. The innermost layer comprises a silicon-based tracker for trajectory reconstruction of charged particles, followed by a scintillating crystal electromagnetic calorimeter that measures the energy of electrons and photons. Surrounding these components is a sampling calorimeter dedicated to the measurement of hadronic energy. The tracker and the calorimeters are located inside the solenoid magnet. Outside of the magnet and within the return yoke are the muon chambers. The full detector is 29 m in length, 15 m in diameter, and weights 14,000 tonnes.

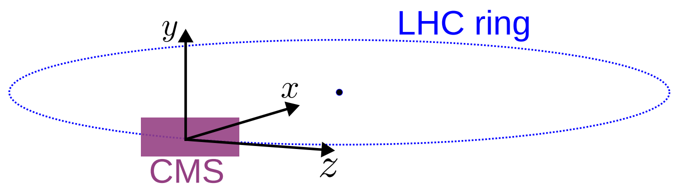


Figure 3.4: Coordinate system of the CMS detector

3.2.1 Coordinate system

Particle directions are measured with a right-handed coordinate system, where the positive x-axis points towards the center of the LHC ring, the positive y-axis points vertically upwards, and the positive z-axis aligns with the direction of the LHC beam (Figure 3.4). The directions are specified in terms of the azimuthal angle, ϕ, relative to the x-axis, and pseudorapidity, η. The pseudorapidity is related to the polar angle, θ, relative to the z-axis by:

$$\eta = -\ln[\tan(\theta/2)]$$

which is approximately equal to the rapidity for relativistic particles. Momentum is measured in the x-y plane, referred to as *transverse momentum* and denoted by $\vec{p}_T$. Correspondingly, particle energy is described by their transverse energy, which is related to their p_T and mass, m, by:

$$E_T = \sqrt{p_T^2 + m^2}$$

For $m \ll p_T$, E_T and p_T are approximately equal. Spatial separation between particles is measured in the (η, ϕ) space and denoted by ΔR:

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

3.2.2 The Tracker

Particles produced in the collisions first encounter the tracker [21, 22, 23, 24]. Its primary function is to precisely measure the trajectory of charged particles, which allows the determination of particle momentum, charge sign, and point of origin (referred to as *vertex*). Vertex reconstruction is especially important for identifying multiple *p*-*p* collisions (*pileup* interactions) that occur simultaneously within a single bunch-crossing. During the 2016 data-taking period, the average number of pileup interactions was 23, which increased to 32 in 2017 and 2018.

The tracker is immersed in a uniform magnetic field of 3.8 T parallel to the beam. It cylindrically covers the interaction point and consists of multiple layers of silicon sensors, with overall dimensions of approximately 1.2 m in radius, 5.6 m in length, and pseudorapidity coverage of $|\eta| < 2.5$. The silicon material is particularly suitable for withstanding the high-radiation environment in close proximity to the interaction point. Each silicon sensor is a *p*-*n* semiconductor junction diode with a bias voltage applied. When a charged particle traverses the tracker, electron-hole pairs are liberated through the waver in the depletion regions of the sensors, generating electronic signals, or *hits*. The track is reconstructed through a fit of the positions of these hits, thereby allowing the momentum to be measured from the curvature of the track in the magnetic field.

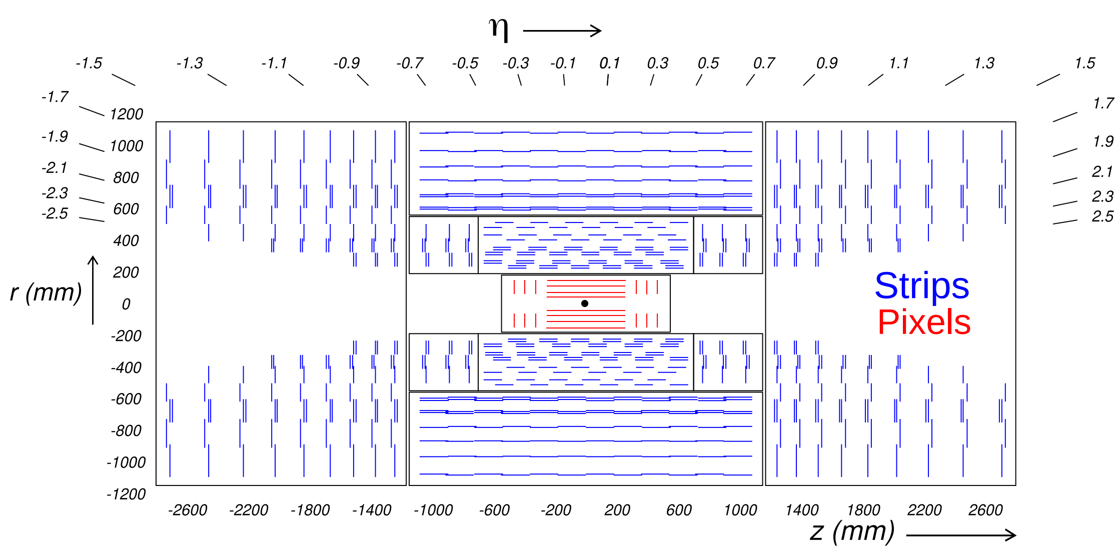


Figure 3.5: Schematic of the tracker shown as a vertical slice along the beam

The tracker is designed to provide high spatial granularity for precise momentum reconstruction while minimizing material thickness to reduce scattering and energy loss. Figure 3.5 shows a schematic diagram of the tracker. It consists of two sections:

- **Pixel detector**: For precise measurement of 3D space coordinates of points on a charged particle's trajectory, layers of high-granularity silicon pixels are placed close to the beamline. In the barrel region, four layers are arranged cylindrically around the beam, while in the endcap region, three layers are arranged as disks transverse to the beam. The pixels have dimensions $100 \times 150\,\mu\text{m}^2$. Due to the strong magnetic field, charge carriers in the silicon drift into adjacent pixels, which allows for an interpolation of the analog electronic signals between adjacent pixels to achieve a spatial resolution in the range of 10-20 μm.

- **Strip detector** Outside of the pixel detector are several layers of elongated rectangular strips consisting of silicon microstrip sensors. The microstrips have a coarser granularity than the pixels, with a typical cell size of 10 cm $\times$ 80 μm. In the barrel region, there are ten concentric strip layers parallel to the beam that provide measurements of the (r, ϕ) coordinates. In the endcap region, the strips point radially towards the beam in twelve layers of disks to provide a measurement in the (z, ϕ) coordinates. In some layers, each strip is supplemented by a second strip mounted back-to-back to the first but rotated through a stereo angle of 100 mrad. This provides a measurement of the third coordinate (z in the barrel and r on the disks).

3.2.3 The Electromagnetic Calorimeter (ECAL)

The electromagnetic calorimeter (ECAL) [25] provides precise measurements of the energies of photons and electrons originating from the interaction point as well as the electromagnetic energy depositions from hadrons. It is homogeneous, hermetic, and composed of scintillating crystals made of lead tungstate ($PbWO_4$) crystals. Due to its high density of 8.28 g/cm^3, short radiation length of 0.89 cm, and small Molière radius of 2.2 cm, the crystal material enables the ECAL to be highly segmented and to effectively capture and localize electromagnetic showers. The crystals also have a fast response, with about 80% of the scintillation emitted within the LHC bunch crossing time of 25 ns.

The ECAL is divided into two main sections: the barrel (EB) and the endcaps (EE). The EB consists of 61,200 crystals and covers the pseudorapidity range $|\eta| < 1.479$. The

EE is made up of two disks, each with 7324 crystals, covering $1.479 < |\eta| < 3.0$. The crystals in the EB are arranged so that they are off-pointing by 3° with respect to the direction towards the interaction point in order to minimize energy being lost in the gaps between the crystals. The EE crystals are arranged in an x-y grid to form an approximately circular shape. The crystals are 23 cm and 22 cm long in the EB and EE. The length is approximately equivalent to one nuclear interaction length, which causes about two thirds of the hadrons to begin showering in the ECAL. The crystal front face dimensions are about 22×22 mm^2, corresponding to approximately one Molière radius of the material, allowing the crystal to contain, on average, 90% of the lateral spread of an electromagnetic shower when the impact is centered on the crystal.

In the endcap region, the angular separation between a photon pair from the decay of a high-energy π^0 meson tends to be small, which produces a collimated deposit in the ECAL. In front of each EE, a preshower detector (ES) with a much finer granularity than the ECAL is installed to help resolve diphoton pairs [26]. The ES covers the pseudorapidity range of $1.65 < |\eta| < 2.6$ and consists of two layers of lead plates interleaved with two layers of strips of silicon sensors. Each strip is 1.9 mm $\times$ 63 mm, and provides position and energy measurements of the early developments of the showers.

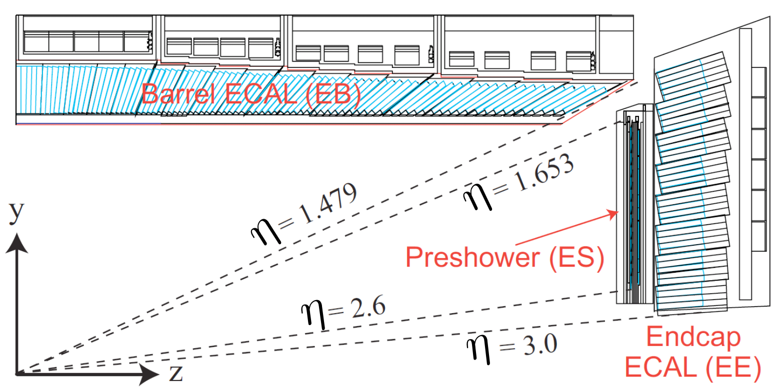
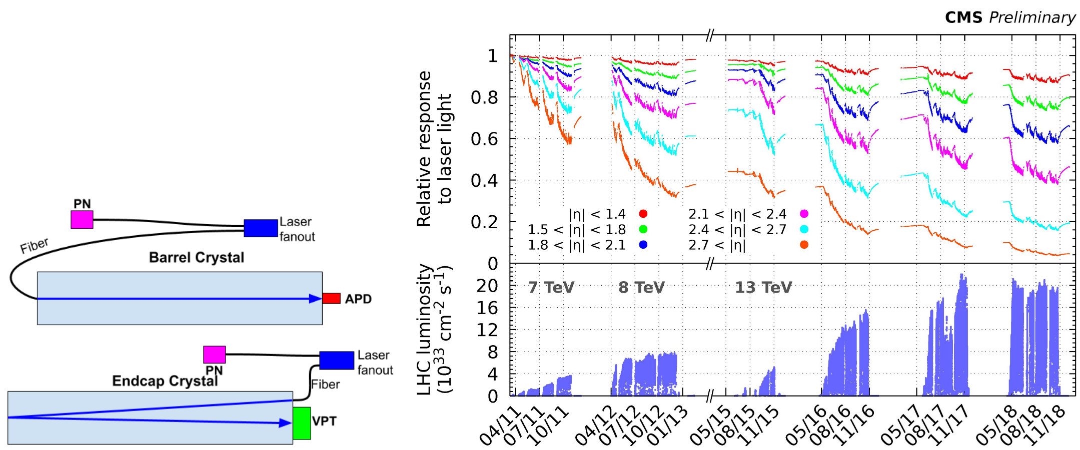


Figure 3.6: Schematic of a quadrant of the Electromagnetic Calorimeter (ECAL) shown as a vertical slice along the beam

The ECAL crystals emit blue-green scintillation light with the peak wavelength at 420–430 nm. The scintillation light is collected and amplified by two types of fast, radiation-tolerant photodetectors attached to the outer faces of the crystals. In the barrel, the light is collected by avalanche photodiodes (APDs) [27]. The endcap crystals are equipped with specialized photodetectors known as a vacuum phototriodes (VPTs) [28] that can withstand the higher radiation flux and the strong magnetic field in this region.



(a) Laser injection method

(b) Evolution of the ECAL response to injected laser in different regions of pseudorapidity

Figure 3.7: Monitoring of crystal transparency with laser injection

During data collection, the crystals lose optical transmission due to the formation of radiation-induced color centers, which absorb a fraction of the transmitted light. This damage partially recovers during the interfill periods of the LHC, but can degrade the performance of the calorimeter if not corrected for during data-taking. To measure the evolution of the transparency loss, the ECAL is equipped with a laser monitoring system that injects laser pulses into the crystals via optical fibers every 40 minutes during LHC operation [29]. The intensity of the pulse is measured separately by silicon photodiodes to provide a reference for normalizing the response of the crystals to the laser. Figure 3.7 shows the evolution of the laser response of the crystals over the course of data-taking between 2011 and 2018 [30]. Because of the differences in optical paths and in the spectra of the injected laser pulses and the scintillation light, the laser response is not linearly related to scintillation response. The relationship between the laser response,

$R(t)$, and the scintillation response, $S(t)$, is approximated by a power law [31]:

$$\frac{S(t)}{S(0)} = \left[\frac{R(t)}{R(0)}\right]^{\alpha}$$

The value of α is depends on the crystal manufacturing method and accumulated radiation damage, and has been calibrated through a study of the evolution of the invariant mass in $Z \rightarrow ee$ events. The correction derived for each crystal is applied to the response signals in the reconstruction process.

3.2.4 The Hadron Calorimeter (HCAL)

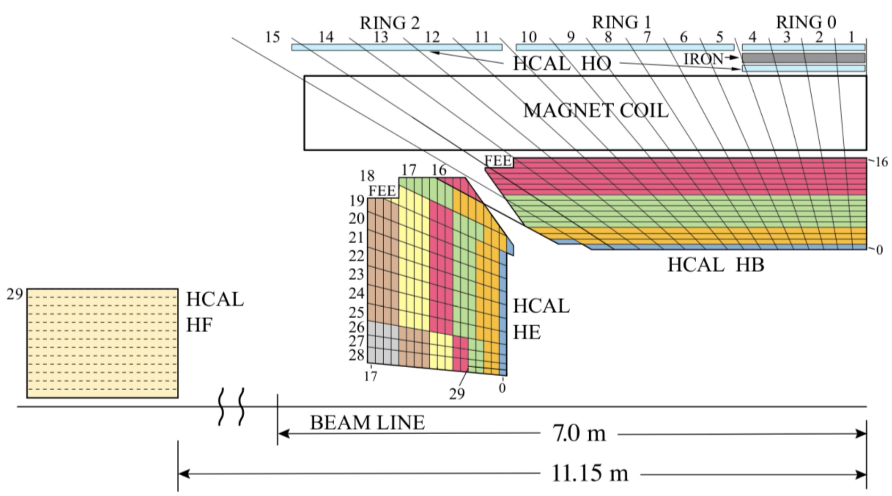


Figure 3.8: Schematic of the Hadron Calorimeter

The HCAL [32, 33] is a sampling calorimeter that measures the hadronic energy of jets with a pseudorapidity coverage of $|\eta| < 5.2$. It also plays a crucial role in the measurement of missing transverse momentum. It consists of several interleaved layers of brass and steel absorber plates instrumented with plastic scintillator tiles. Hadrons from the collision point undergo strong nuclear interactions with the dense material of the absorber plates and produce hadronic showers. Charged hadrons from these showers create scintillation light in the plastic scintillators, which is collected with wavelength shifting fibers and transported to photomultipliers (hybrid photodiodes [34] or silicon photomultipliers) to provide a measurement of the deposited energy.

The HCAL is divided into four sections (Figure 3.8):

- The **hadron barrel (HB)** calorimeter covers pseudorapidities up to $|\eta| < 1.39$. It has 14 layers of brass plates enclosed by two steel layers and 17 layers of plastic scintillators. The depth of the HB varies between $5.8\lambda_l$ at $\eta \sim 0$ to $10.6\lambda_l$ at $\eta \sim 1.3$, where λ_l is the nuclear interaction length (mean distance travelled by a hadron before undergoing an inelastic nuclear interaction).

- The **hadron endcap (HE)** calorimeter extends the pseudorapidity coverage to $|\eta| < 3.0$. It has 16 absorber and 17 scintillator layers.

- The **hadron outer (HO)** calorimeter covers $|\eta| < 1.26$ and consists of one or two layers of scintillator outside the magnet coil. It improves the energy resolution by measuring the energy deposited in the tail of a hadronic shower, which typically contains low-energy secondary particles.

- The **hadron forward (HF)** calorimeter provides coverage of the forward region $2.85 < |\eta| < 5.19$, where particle flux is the highest. It is located 11m away from the interaction point but very close to the beam. The absorbers are made of steel while the scintillation material consists of radiation-hard quartz fibers, where charged particles emit Cherenkov light.

3.2.5 The Muon System

The outermost part of the detector is the muon spectrometer [35, 36], made up of gas-ionization chambers. It is of central importance to the detector since muons are the only visible particles able to penetrate into the muon system, which provides low background rates to studies of processes with muons in the final state (e.g., Higgs studies with $H \to ZZ \to 4\mu$ signature). The muon chambers are interleaved between the steel flux-return yoke, which returns the magnetic field of the solenoid. The trajectory of a muon is reconstructed through a fit of the positions of the signals generated in the cells of the muon system, which allows the momentum of the muon to be measured from the curvature of the track.

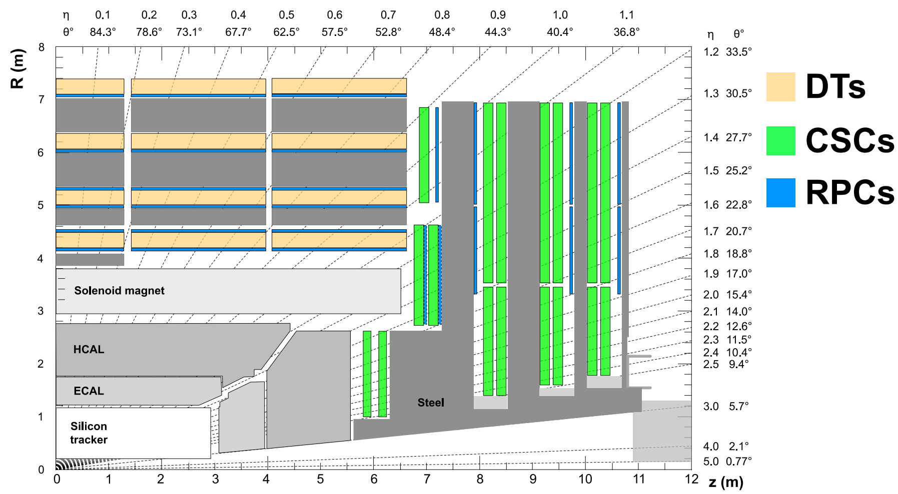


Figure 3.9: Components of the muon system in a vertical slice of a quadrant of the detector

Three different types of muon chambers are used in the detector:

- **Drift tubes (DTs)** : The DT chambers are located in the barrel region ($|\eta| < 1.2$), where the muon and background rates are both low, and the magnetic field is uniform. A DT chamber measures about 42 mm wide, 13 mm high, and 2 to 4 m long. The chambers are filled with an 85/15 mixture of Ar and CO_2 gas, and contain a uniform electric field. Muons traversing the tubes ionize the gas and liberate electrons, which drift through the tube towards a 3.6 kV anode wire and produce an electron avalanche near the wire, creating a current pulse.

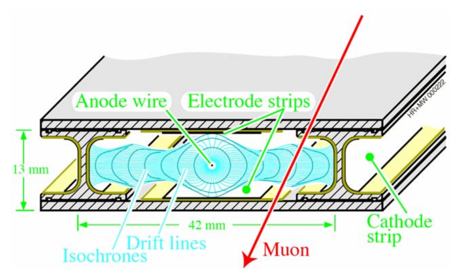


Figure 3.10: Layout of a DT cell, showing the electric field lines in the gas volume

- **Cathode strip chambers (CSCs)** : The CSCs measure muons in the endcaps ($0.9 < |\eta| < 2.4$), where the muon rates are higher, and the magnetic field is both large and uneven. They are radiation hard, have a fast response time, and are finely segmented for precise track reconstruction. The chambers are filled with a mixture of Ar, CO_2 , and CF_4 gases, and are segmented into six layers. Each layer consists of cathode strips aligned radially and anode wires perpendicular to the strips. A passing muon ionizes the gas, and charges are registered on the anodes and cathodes. The grid structure of the anode wires and cathode strips provides a location in (r, ϕ, z) space.

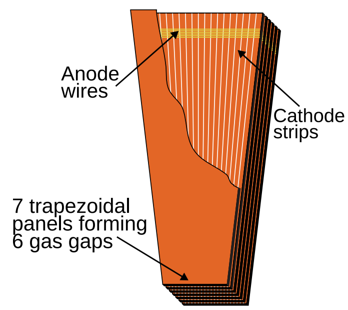


Figure 3.11: Structure of the CSC chamber showing the anode and cathode grid

- **Resistive plate chambers (RPCs)** : RPCs are gaseous parallel-plate muon detectors that provide fast response for trigger purposes. They have two layered gas chambers held at a high potential difference by the positively charged anode plates on the top and bottom, and a pair of negatively charged cathode strips sandwiched between the chambers. The chambers are 2 mm thick and filled with a gas mixture of C_2H_2, F_4 , C_4H_{10}, and SF_6. A muon passing through the RPC ionizes the gas, produces an electron avalanche due to the electric field, and induces a charge on the cathode readout strips.

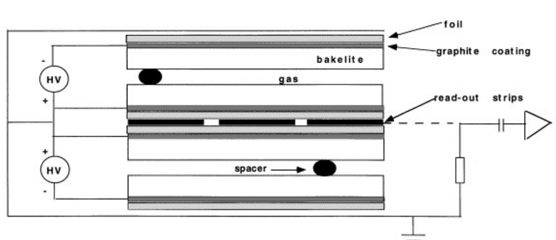


Figure 3.12: Schematic of an RPC

3.2.6 The Triggering System

The collision rate of 40 MHz at the LHC, combined with the large number of readout channels in the CMS detector, results in a large volume of data that cannot be collected and stored by the detector's computational resources. The detector therefore relies on a trigger system to significantly reduce the event recording rate and to select events with signatures that are important for performing fundamental physics measurements.

The trigger system consists of two stages:

- **Level 1 (L1) trigger**: The first triggering stage operates at the hardware level with custom-designed, highly programmable electronics, including FPGAs, ASICs, and programmable memory lookup tables, to quickly identify interesting events at a rate of up to 100 kHz. The decision time between a bunch crossing and delivering an accept signal, the *latency*, is about 4 μs. The trigger receives signals from the front-end electronics of the ECAL, HCAL, and the muon system to perform a coarse reconstruction of the event, and makes trigger decisions based on simple criteria on the presence of energy clusters in the calorimeters, or ionization signals in the muon system [37, 38].

- **High-level trigger (HLT)**: Events passing the L1 trigger are sent to the HLT, which operates at the software level on a large computing farm with thousands of processors to perform a detailed reconstruction of the particles using the full readout of the detector and apply more complex criteria for specific event signatures, reducing the event rate from 100 kHz to about 1 kHz [39] with an average time latency of 260 ms. Events accepted by the HLT is transferred to the CMS computing grid for storage and further refinement through a full reconstruction process described in the next chapter.

Chapter 4

Event reconstruction

As particles produced by the *p*-*p* collisions traverse the detector, they generate signals in different sub-detectors based on their characteristics (Figure 4.1). They first pass through the tracker, where signals from the sensitive layers are used to reconstruct the trajectories (tracks) and origins (vertices) of charged particles. The tracker operates within a magnetic field that causes the paths of charged particles to bend, enabling the measurement of their electric charges and momenta. Subsequently, electrons and photons are absorbed by the electromagnetic calorimeter (ECAL). The ECAL detects the resulting electromagnetic showers as energy clusters in neighboring cells, allowing for the determination of the particles' energy and direction. Charged and neutral hadrons can also initiate hadronic showers in the ECAL, which are then fully absorbed by the hadron calorimeter (HCAL). The clusters observed in the HCAL are used to estimate the energies and directions of these particles. On the other hand, muons and neutrinos traverse the calorimeters with minimal or no interactions. Neutrinos pass through undetected, while muons leave a trace in the outermost layer known as the muon system. Specialized algorithms are implemented to use the readout signals from the sub-detectors to reconstruct the energies and momenta of the particles emitted in the collisions. The reconstructed particles form the basis of the measurement of the cross-section section and the search for aNTGCs in this thesis. The specific criteria for selecting the events will be discussed in Chapter 6. This chapter outlines the principles of the algorithms used in reconstructing photons, electrons, muons, jets, and missing transverse momentum.

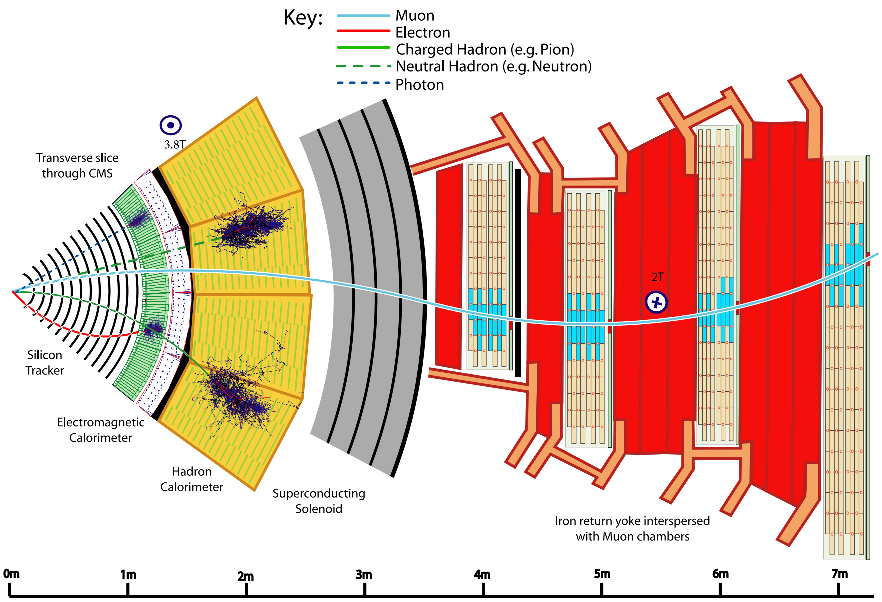


Figure 4.1: Transverse slice of the CMS detector showing interactions of different particles with the sub-detectors

4.1 The particle flow algorithm

The CMS detector employs a comprehensive strategy called the *particle flow* algorithm for particle reconstruction [40]. This approach aims to achieve a holistic understanding of collision events by integrating information from all subdetectors and establishing correlations between measurements to enhance particle reconstruction. It takes advantage of the fine angular granularity of the calorimeters, and the high resolution of the tracker and muon chambers to include low-momentum particles that would otherwise be ignored. This method particularly improves the accuracy of jet energy measurement, since charged hadrons contribute around 65% of the jet energy, photons account for approximately 25%, and neutral hadrons for about 10%. The inputs to the particle flow algorithm are the tracks and the clusters in an event. The tracks are reconstructed from either the inner tracker system or the muon detector, while clusters in the ECAL and HCAL are reconstructed by dedicated clustering algorithms. Based on the reconstructed tracks, vertices associated with these tracks are constructed to identify the interaction points of proton-proton collisions. These tracks and clusters are then gathered to form the inputs to the particle flow algorithm. Subsequently, the algorithm proceeds with a dedicated linking algorithm specific to the target stable particles. This linking algorithm connects vertices, tracks, and showers spatially to associate them with five categories of particles: muons, electrons, photons, charged hadrons, and neutral hadrons.

We briefly describe below the algorithms used in the reconstruction of the tracks and calorimeter clusters.

4.1.1 Tracks and vertices

Track reconstruction is a computationally challenging task due to the high multiplicity of charged tracks produced in the collisions. A dedicated software package, known as the Combinatorial Track Finder (CTF), is used to produce a collection of reconstructed tracks by multiple iterations of the reconstruction sequence. The initial iterations search for tracks that are easiest to find (e.g., those with large transverse momenta and near the beam interaction region). After each iteration, hits associated with the reconstructed tracks are removed to reduce the combinatorial complexity for subsequent iterations that reconstruct more difficult classes of tracks (e.g., low transverse momentum or significantly displaced from the interaction region). A detailed description of the track and vertex reconstruction process can be found in [23]. We summarize the four steps in each tracking iteration:

- **Seed generation**: During this step, *track seeds* are generated using hits in the innermost layers of the tracker. These hits are combined to form track seed candidates, which act as initial approximations for potential tracks. In the initial iterations, three hits in the pixel layers are required to form a track seed. As the iterations progress, the criteria for track seed formation become less stringent. The final iterations specifically target increased muon tracking efficiency by including information from the muon chambers.

- **Track finding**: The track finding step spatially extends the track seeds using an algorithm known as *Kalman filter* [41], which searches for additional hits in neighboring tracker layers that are consistent with the seed trajectory. The hits are assigned to the corresponding tracks based on their proximity to the predicted track position. This process is performed iteratively, gradually extending the track segments and refining the track parameters.

- **Track fitting**: Once the track segments have been extended, a track fitting algorithm is employed to determine the best-fit track parameters. The fitting process minimizes the difference between the predicted positions of the track and the measured hits, taking into account the uncertainties associated with the hits and the material effects (e.g., multiple Coulomb scattering and energy loss) in the tracker.
- **Track Selection**: After the track fitting, a set of criteria is applied to filter out spurious or poorly reconstructed tracks. Various quality criteria, such as the number of hits, track fit quality (χ^2/d.o.f.), and compatibility with the interaction vertex, are used to assess the track quality.

After reconstruction of the tracks, a vertex reconstruction algorithm clusters the tracks that appear to originate from the same interaction vertex, and determines the position of each vertex using its associated tracks. The vertices are ordered by the quadratic sum of the transverse momenta of the tracks associated with each vertex, i.e. $\sum p_T^2$. The *primary vertex* is defined as the one with the largest sum. The other vertices are considered to be *pileup interactions*, which result from multiple *p*-*p* collisions occurring within the same bunch crossing.

4.1.2 Calorimeter clusters

Energy deposits in the ECAL and HCAL are grouped using dedicated clustering algorithms. These algorithms start with seed finding, where local energy maxima are identified as starting points. These seeds are typically chosen as the cells with largest energy deposits or cells exceeding a certain threshold. Starting from each seed, neighboring cells that meet certain criteria, such as energy thresholds and geometric constraints, are added to form clusters. The process of adding neighboring cells continues iteratively until no more cells can be included. This forms clusters representing electromagnetic showers by electrons and photons, or hadronic showers produced by charged and neutral hadrons. Further details on ECAL and HCAL clustering can be found in [40].

4.1.3 The link algorithm

Since a single particle can produce multiple *PF elements* (i.e. tracks or clusters), a *link algorithm* is used to connect the PF elements from different subdetectors. The link algorithm systematically examines pairs of elements and assesses if they satisfy specific linking criteria based on the particle type. Subsequently, the link algorithm forms *PF blocks* consisting of elements that are directly or indirectly linked through shared elements. The main types of links formed are:

- **Track-to-cluster**: The last measured hit of a track is extrapolated to the calorimeters to search for clusters that align with the track. The cluster is *linked* to the track if the extrapolated position is within a cluster. When different clusters are linked to the same track, or when several tracks are linked to the same cluster, the quality of the links are evaluated by considering the separations in the (η, ϕ)-plane, and only the link with the smallest separation is kept.

- **ECAL cluster-to-track**: Electrons bend in the magnetic field and radiate photons, which result in multiple clusters in the ECAL that are separated azimuthally. In order to link the clusters to the track, tangent lines are extrapolated to the ECAL from the intersection points between the track and each of the tracker layers. In addition, photons can also convert to a pair of e^+e^- in the tracker material, producing spatially separated tracks. Two tracks compatible with a single photon origin are linked by a dedicated conversion finder [42].

- **Cluster-to-cluster**: Hadronic showers from jets can produce deposits in both the ECAL and the HCAL. Links between these calorimeter clusters are formed when the cluster position in the more granular calorimeter (preshower or ECAL) is within the cluster envelope in the less granular calorimeter (ECAL or HCAL). When multiple HCAL clusters are linked to the same ECAL cluster, or when multiple ECAL clusters are linked to the same preshower clusters, only the link with the smallest separation is kept.

Once the elements are linked, association criteria are applied to further refine the formation of physics objects. These criteria consider factors such as track-cluster matching, energy sharing between different elements, and consistency with specific particle signatures. The final step involves grouping the linked and associated elements into *PF objects*. These PF objects are classified as electrons, photons, muons, charged hadrons, and neutral hadrons.

4.2 Muons

Muons are reconstructed with high purity and efficiency since they penetrate the whole detector, while other particles are absorbed by the calorimeters and the steel layers of magnet. Their ionization signals in the gas chambers are used to reconstruct tracks independently of the hits in the inner tracker. The muons used in this analysis are reconstructed by extrapolating the tracks in the muon system to the inner tracker and searching for a track with matching parameters. A second reconstruction of the full muon track is performed by combining the hits of the muon system and the inner tracker to improve the momentum measurement. Muons originating from hadrons are rejected using two sets of identification criteria, known as *loose* and *tight* criteria [36]. These criteria are defined by requirements on the quality of the two tracks, including number of hits in the inner tracker, number of segments in the muon system, and impact parameter of the track relative to the primary vertex, as well as requirements on hadronic activity in the regions of the calorimeter near the muon. Further details on muon reconstruction can be found in [36].

4.3 Photons and electrons

Electrons and photons deposit their energy primarily in the ECAL. As they pass through the detector material before reaching the ECAL, interactions can occur, leading to the emission of Bremsstrahlung photons by electrons and the conversion of photons into electron-positron pairs. Consequently, the electron or photon arriving at the ECAL may manifest as a shower of multiple electrons and photons rather than a single particle. To recover the energy of the original electron or photon, a dedicated algorithm combines the clusters from these individual particles into a *supercluster*. Additionally, the trajectory of an electron emitting Bremsstrahlung photons undergoes changes in curvature in the tracker. For electrons, a dedicated tracking algorithm, based on a technique known as

Gaussian sum filter (GSF) that incorporates a sum of Gaussian functions to model the energy loss, is employed to estimate the track parameters [43].

The energy measured in the ECAL supercluster can be affected by electromagnetic shower losses, such as lateral and longitudinal leakage, intermodule gaps, dead crystals, and energy lost in the tracker. These losses introduce systematic variations and degrade the energy resolution. A multivariate technique based on a regression derived from simulation, where the true values of the energies are known, is used to correct the energy estimation. For electrons, where energy and momentum are measured independently in the ECAL and tracker, respectively, the two measurements are combined to improve the energy estimate. This is particularly beneficial at low transverse momentum, where the momentum measurement from the tracker has a better resolution than the corresponding ECAL measurement. After these corrections are applied, small differences remain between real data and simulation in both the electron and photon energy scales and resolutions. In particular, the resolution in simulation is better than that in data. A set of corrections, known as *energy scale and spreading corrections* [44], is derived from the di-electron invariant mass distribution in $Z \to ee$ events in real data and in simulation to minimize these residual differences. A scale offset is applied to the energy in real data to align the energy response with simulation. In simulation, on the other hand, a Gaussian spreading is applied to the energy response to match the resolution observed in real data. As electromagnetic showering in the ECAL is similar for electrons and photons, they are corrected by the same set of scale and spreading corrections. The uncertainty in energy scale after the corrections is estimated to be below 0.8% in the barrel and below 1.1% in the endcaps.

For this analysis, only electrons with a transverse momentum above 10 GeV and within the tracker acceptance ($|\eta| < 2.5$) are considered. To distinguish electrons from hadronic deposits in the ECAL, two sets of identification criteria, referred to as *loose* and *tight* criteria [44], are used. The criteria are based on the shower shape patterns in the ECAL, the quality of associated tracks, matching of the tracks to the ECAL clusters, and requirements on the surrounding activity in the HCAL. The loose criteria are tuned using simulation samples to have a signal efficiency of 90% and under 0.5% background acceptance. The tight criteria are tuned for electrons with transverse momentum above 30 GeV, with a signal efficiency of 70% and under 0.1% background acceptance.

For the identification of photons, a dedicated identification algorithm for photons with transverse momentum above 200 GeV has been developed for this analysis. This is discussed in Chapter 5.

4.4 Jets

Quarks and gluons produced in *p*-*p* collisions create a collimated spray of hadrons due to color confinement and appear as a cluster of energy deposited in a localised area of the detector, which is referred to as a *jet*. On average, 65% of the jet energy is carried by charged hadrons, 25% by photons, and 10% by neutral hadrons. In the particle flow algorithm, charged and neutral hadrons are formed from HCAL clusters and any linked tracks or ECAL clusters. If an HCAL cluster within the tracker acceptance is linked with a track, it is classified as a charged hadron; otherwise, it is considered a neutral hadron. Outside the tracker acceptance, an HCAL cluster along with any linked ECAL cluster becomes a hadron with no charge distinction. After the formation of the objects in the particle flow algorithm, jets are reconstructed in two main steps: mitigating of pileup effects and clustering.

Before the clustering step, an algorithm known as *charged hadron subtraction* (CHS) [45, 46] is applied to remove the contributions from extra *p*-*p* interactions that occur simultaneously in a single bunch crossing, which are known as *in-time pileup*. The algorithm orders the reconstructed vertices by the quadratic sum of the square of the transverse momenta ($\sum p_T^2$) of their associated tracks. The vertex with highest $\sum p_T^2$ is taken as the primary vertex and the rest are treated as pileup vertices. The charged hadrons with tracks originating from pileup vertices are then discarded. Next, the clustering step is performed using the *anti-kT* algorithm [47]. The algorithm assigns a distance metric between pairs of particles based on their transverse momentum values and spatial separation, and iteratively identifies pairs of particles with the smallest distance. The distance between two particles i and j is defined as:

$$d_{ij} = \min\left\{\frac{1}{k_T^2(i)}, \frac{1}{k_T^2(j)}\right\} \cdot \frac{\Delta R_{ij}^2}{R^2}$$

where $k_T(i)$ and $k_T(j)$ are the transverse momenta, ΔR_{ij} is the separation between the two particles in the η-ϕ space, defined as:

$$\Delta R_{ij}^2 = [\Delta\eta(i,j)]^2 + [\Delta\phi(i,j)]^2$$

and R is a characteristic scale known as the *jet radius*. The characteristic scale is chosen to be $R = 0.4$ in this analysis. If the distance, d_{ij}, is smaller than the characteristic scale, R, the particles are combined into a single jet. This process continues until all particles have been assigned to jets. The algorithm clusters particles with high transverse momenta first, creating seed jets, and then gradually adds softer particles, forming a conical jet structure. The 4-momentum of the jet is determined as the vectorial sum of the 4-momenta of the particles in its cluster.

The energy of a jet reconstructed by clustering is subject to instrumental effects of the calorimeters, including electronic noise, non-uniform and non-linear response, as well as pileup contributions from photons, neutral hadrons and particles reconstructed outside the tracker acceptance. Additionally, the response observed in simulation differs from real data. A set of corrections, known as *jet energy scale* corrections, are applied so that the energy of the jet as measured by the detector aligns, on average, with the energy of true jets formed by applying the same clustering algorithm to the *truth-level* particles generated before any interaction with the detector in simulation. These correction are applied sequentially as multiplicative factors to each component of the jet's 4-momentum vector. A detailed description of these corrections can be found in [48]. Below is a summary outlining the key aspects of these corrections:

- **Pileup offset corrections**: While the CHS algorithm removes charged hadron contributions from pileup interactions within the same bunch crossing, particles within the jet cluster can still originate from previous bunch crossings. Additional contributions also arise from neutral particles and from charged particles that lie outside of the tracker acceptance. To remove these contributions, *pileup offset corrections* are derived from simulated samples of events with two jets. The corrections are designed to account for variation of the pileup contribution with the jet transverse momentum and pseudorapidity, as well as the overall amount of pileup activity in the event.

- **Response corrections from simulation**: To align the reconstructed jet energy with true jet energies, corrections are derived from a simulation sample of events with multiple jets. The simulation uses a detailed model of the detector's geometry, alignment, and calibration, as well as a high-precision model of particle interactions, and therefore reliably reproduces jet reconstruction in real data. These corrections are derived as a function of the jet transverse momentum and pseudorapidity in order to make the response uniform in these two variables.

- **Residual corrections for real data**: The response of jets is measured in real data using a number of samples containing a jet recoiling against a reference object. These samples are enriched with the processes $Z(\rightarrow \mu\mu)$+jet, $Z(\rightarrow ee)$+jet, γ+jet, and di-jet events containing one jet with a transverse momentum exceeding 250 GeV within the central region of the detector ($|\eta| < 1.3$). The momentum of the reference object in these samples is measured with high precision. The corrections are derived using the principle that the transverse momentum of the reconstructed jet must balance out the transverse momentum of the reference object.

After these corrections are applied, the resolution of jet energy in simulation is found to be better than in real data. By comparing the spread of jet energy in reference samples from real data and simulation, corrections are derived to align the resolution in simulation with real data. These corrections are applied to the jets in simulation by smearing their energies with a Gaussian distribution. The energy of the jets after all corrections is found to lie within 5% to 20% of the true particle energies in simulation.

4.5 Missing transverse momentum

Neutrinos and hypothetical neutral weakly interacting particles produced the p-p collisions do not interact with the detector. However, their presence can be indirectly inferred from the momentum imbalance in the plane perpendicular to the beam direction. Since the protons in the beam travel longitudinally, the total vectorial transverse momentum of the products of the collision must be zero. Assuming perfect momentum measurement, if the vectorial sum of the transverse momenta of the reconstructed particles is non-zero, the unbalanced transverse momentum must be carried by neutrinos or other invisible particles. The transverse momentum carried by invisible particles is referred to as *missing transverse momentum* ($\not\vec{p}_T$) or *missing transverse energy* ($\not\vec{E}_T$). It is defined as the negative vectorial sum of the transverse momenta of the reconstructed particles:

$$\not\vec{E}_T = -\sum_{i \in \text{visible}} \vec{p}_{i,T}$$

where the index i runs over all reconstructed particles in the event. The corrections to the transverse momenta of the reconstructed particles, including the corrections to the jet energy scale and resolution, are propagated in this sum.

Spurious signals in the detector can also lead to the reconstruction of an artificial missing transverse momentum. These include *beam halo* deposits, where interaction of the beam with materials in the beampipe produce secondary particles that traverse the detector, and anomalous signals in the ECAL barrel produced by direct interaction of hadrons with the active layer of the photodetectors in the crystals. These are discussed in Chapter 6. Further details on the reconstruction of missing transverse momentum can be found in [49].

Chapter 5

Photon identification

In our analysis of $pp \rightarrow Z(\rightarrow \nu\nu) + \gamma$ production, we select events that have a specific signature: a high-energy photon with a transverse momentum greater than 225 GeV and a large missing transverse momentum above 200 GeV. The full selection criteria are discussed in Chapter 6. Here we describe a machine-learning based identification method for selecting the photons. The method has been developed specifically for this study in order to enhance the selection efficiency of the high-energy photons while minimizing the rate of a certain background source.

The events collected by the CMS detector are dominated by processes that produce several jets in the final state. When a high-energy jet deposits energy in the ECAL, it can be misidentified as a photon. This happens particularly when the jet contains a high-energy π^0 or η meson that decays into a pair of photons. The photon pair has low angular separation due to the large boost and appears as a collimated deposit in the ECAL. As a result, the deposit is reconstructed as a single photon.

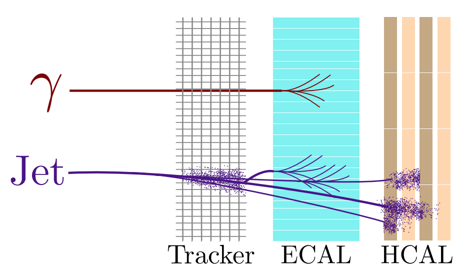
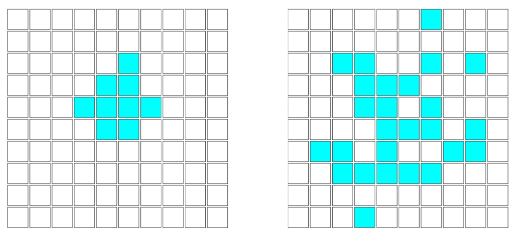


Figure 5.1: Deposits produced in the subdetectors by a jet and a genuine photon

To minimize background arising from jets misidentified as photons, identification criteria are applied on two kinds of discriminating variables. The first kind, termed *isolation variables*, are sums of the energy deposits produced in the subdetectors within a cone surrounding the photon. Photons arising from jets are expected to have significant activity in the subdetectors surrounding the photon cluster, which is referred to as having a large isolation. On the other hand, true photons deposit most of their energy only in the ECAL (Figure 5.1) and have low isolation.



(a) Genuine photon (b) Jet

Figure 5.2: Typical deposits in the ECAL produced by a genuine photon and a jet

The second kind of variables, known as *shower shape* variables, describe the geometric shape of the electromagnetic shower produced in the ECAL. Deposits from the decays of hadrons typically have a broader shower due to additional deposits generated by multiple hadrons produced during the fragmentation of the jet. A photon, on the other hand, leaves a localized deposit that is much narrower than a jet's shower.

In previous CMS analyses of our signature, identification criteria were defined by thresholds on isolation variables and just one shower-shape variable. While this method significantly suppresses the rate of background, the criterion applied to the shower shape variable also reduces the efficiency of selecting true photons. We improve on this method by exploiting further shower shapes variables that also contain discrimination power. To achieve this, we use a machine learning algorithm, XGBoost 1.2.0 [50], to train a *boosted decision tree* (BDT). The BDT allows us to collect the discrimination power contained in several shower shape variables and calculate a single variable called the *BDT score*. The BDT score lies in the range from 0 to 1, where 0 indicates jet-like and 1 indicates photon-like. This score has higher discrimination power than any single shower shape variable alone.

The development scheme of the identification criteria is shown in Figure 5.3. Using simulation samples containing true photons and jets, a training sample is first collected. The BDT is then trained with the shower shape variables. The identification criteria defined by thresholds on the BDT score and three isolation variables are then optimized to minimize the background rate.

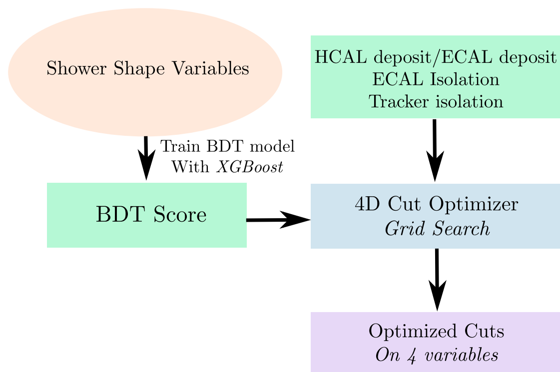


Figure 5.3: Photon ID scheme

5.1 Shower Shape BDT

Showers from jet-induced photons are typically broader than those from real photons. The geometric shape of a shower can be more specifically characterized by a number of variables constructed from the energy deposits in the crystals of the photon cluster (known as a supercluster). To build the BDT model, we identified 14 discriminating shower shape variables for the barrel ECAL and 15 for endcap ECAL. These are:

- $\boldsymbol{E_{\mathbf{max}}/E_{\mathbf{SC}}}$: Ratio of the energy of the most energetic crystal to the super-cluster energy.
- $\boldsymbol{E_{\mathbf{2}}/E_{\mathbf{SC}}}$: Ratio of the energy of the second most energetic crystal to the super-cluster energy.
- $\boldsymbol{E_{\mathbf{1\times3}}/E_{\mathbf{SC}}}$: Ratio of the energies in the 1×3 strip around the central crystal to the super-cluster energy.

- $\boldsymbol{E_{2\times 2}/E_{3\times 3}}$: Ratio of the energies in the 2×2 matrix around the central crystal to the 3×3 matrix.

- $\boldsymbol{E_{2\times 2}/E_{\mathrm{SC}}}$: Ratio of the energies in the 2×2 matrix around the central crystal to the super-cluster energy.

- $\boldsymbol{E_{2\times 5}/E_{\mathrm{SC}}}$: Ratio of the energy of the second most energetic crystal to the super-cluster energy.

- $\boldsymbol{E_{3\times 3}/E_{\mathrm{SC}}}$: Ratio of the energies in the 3×3 matrix around the central crystal to the super-cluster energy.

- $\boldsymbol{\sigma_{\eta}}$: Standard deviation of the η coordinates of the crystals in the supercluster.

- $\boldsymbol{\sigma_{\phi}}$: Standard deviation of the ϕ coordinates of the crystals in the supercluster.

- $\boldsymbol{\sigma_{\eta}/\sigma_{\phi}}$: Ratio of σ_{η} to σ_{ϕ}.

- $\boldsymbol{\sigma_{i\eta i\eta}}$: Standard deviation of the η coordinates (rescaled to units of crystal size) of the crystals in the supercluster.

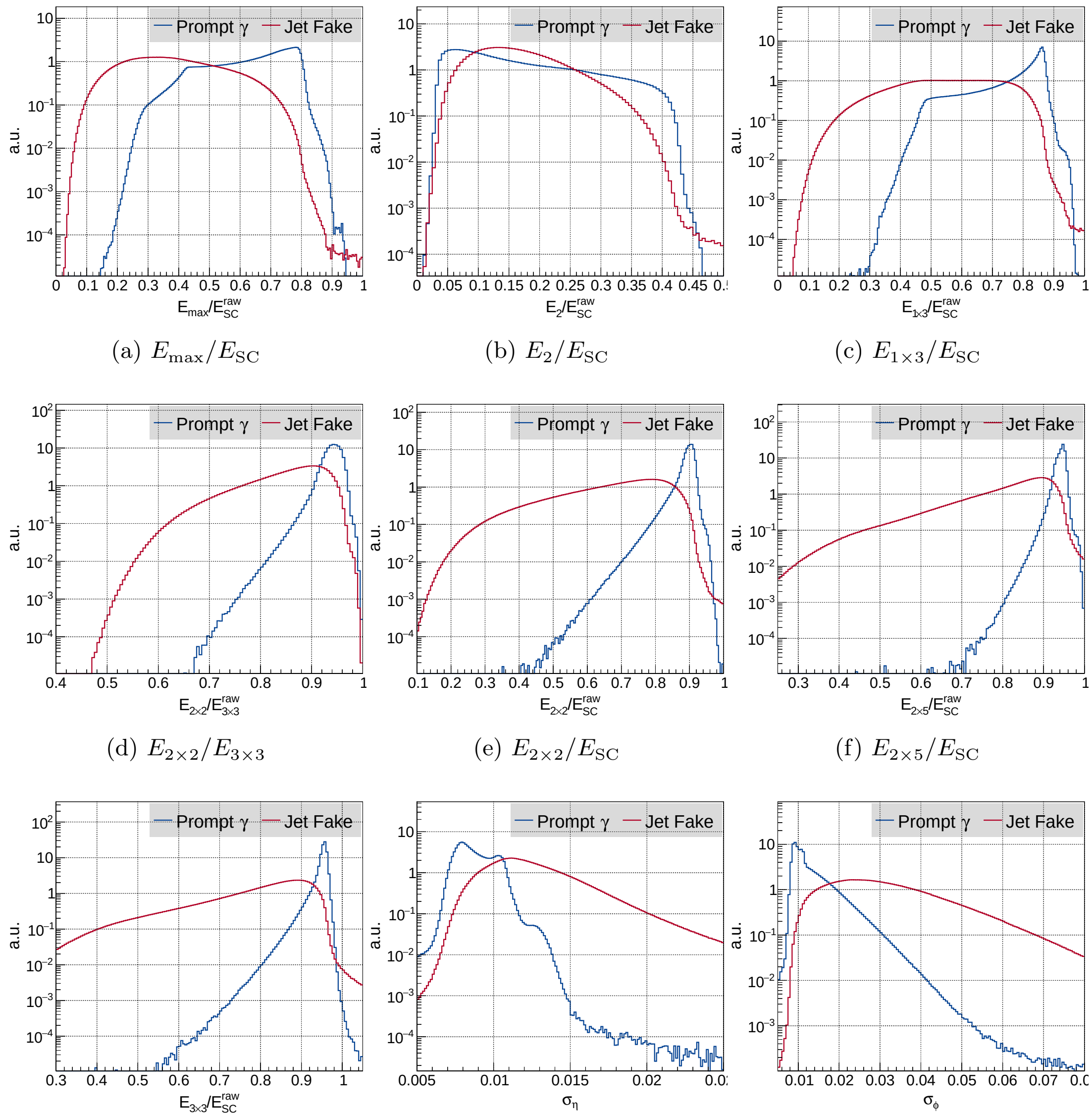
- $\boldsymbol{\sigma_{i\phi i\phi}}$: Standard deviation of the ϕ coordinates (rescaled to units of crystal size) of the crystals in the supercluster.

- $\boldsymbol{\sigma_{i\eta i\phi}}$: Covariance between the η and ϕ coordinates (rescaled to units of crystal size) of the crystals in the supercluster.

- $\boldsymbol{\sigma_{i\eta i\eta}/\sigma_{i\phi i\phi}}$: Ratio of $\sigma_{i\eta i\eta}$ to$\sigma_{i\phi i\phi}$.

- $\boldsymbol{\sigma_{RR}}$: This variable is calculated for the endcap only using the hits in the preshower detector. $\sigma_{RR} = \sqrt{\sigma_{xx} + \sigma_{yy}}$, where σ_{xx} and σ_{yy} measure the lateral spread in the two orthogonal directions of the sensor planes of the preshower detector.

The shapes of the distributions of these shower shape variables for true (prompt) and jet induced (fake) photons in simulated data for photons with $p_T > 200$ GeV are shown in Figures 5.4 and 5.5. It can been seen that the showers produced by the two sources have distinct shapes in the distributions of these variables.

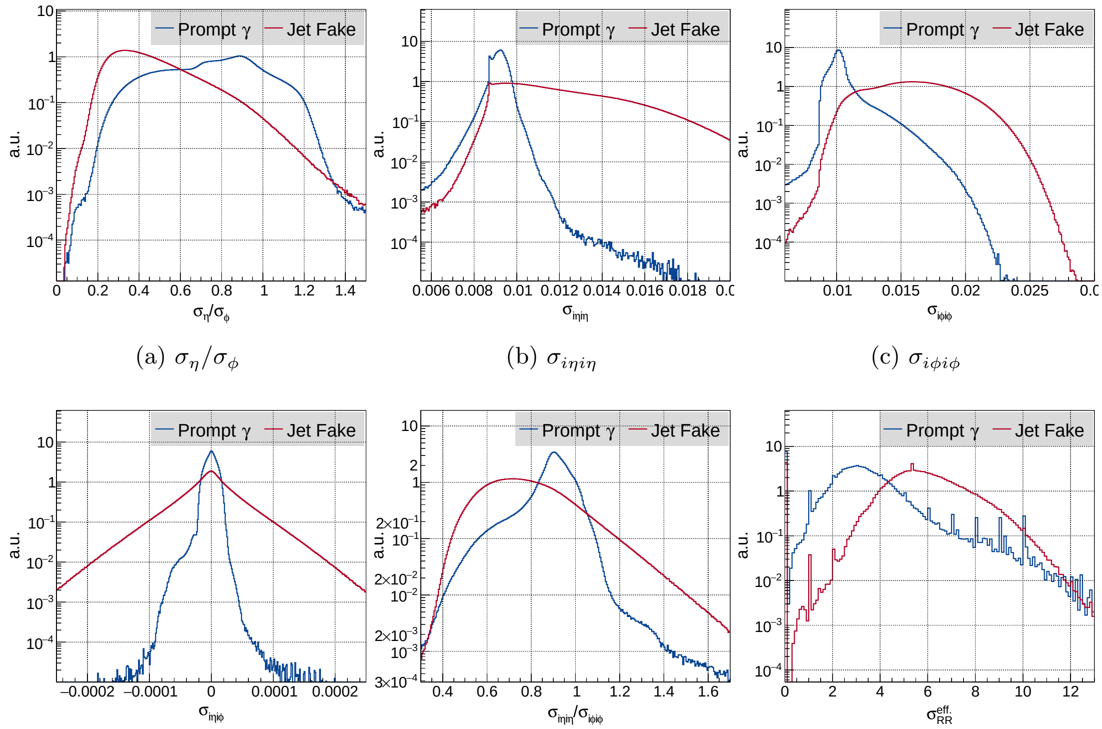


(a) $E_{\mathrm{max}}/E_{\mathrm{SC}}$ (b) E_2/E_{SC} (c) $E_{1\times3}/E_{\mathrm{SC}}$

(d) $E_{2\times2}/E_{3\times3}$ (e) $E_{2\times2}/E_{\mathrm{SC}}$ (f) $E_{2\times5}/E_{\mathrm{SC}}$

(g) $E_{3\times3}/E_{\mathrm{SCs}}$ (h) σ_η (i) σ_ϕ

Figure 5.4: Distributions of shower shape variables for genuine (prompt) and jet-induced (jet fake) photons



(a) σ_η/σ_ϕ (b) $\sigma_{i\eta i\eta}$ (c) $\sigma_{i\phi i\phi}$

(d) $\sigma_{i\eta i\phi}$ (e) $\sigma_{i\eta i\eta}/\sigma_{i\phi i\phi}$ (f) σ_{RR}

Figure 5.5: Distributions of shower shape variables for genuine (prompt) and jet-induced (jet fake) photons

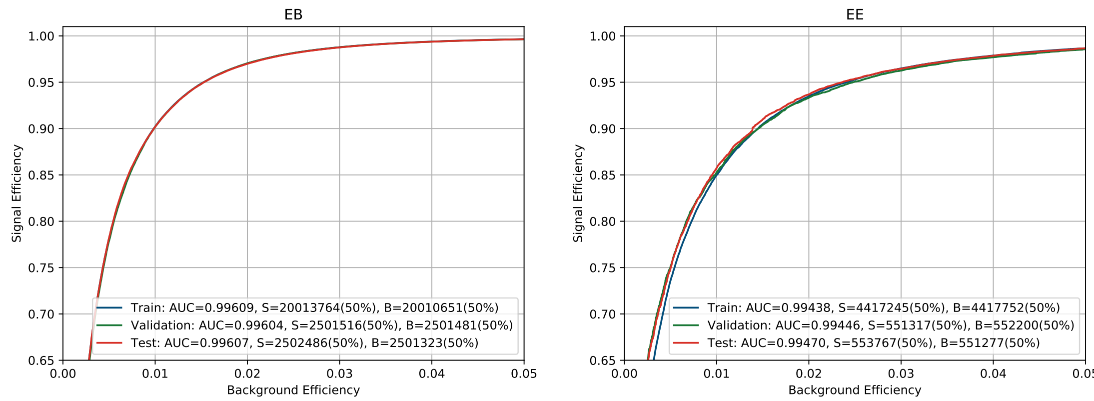
The XGBoost algorithm takes a training sample containing prompt and fake photons from simulation, and builds a large number of decision trees. Each decision tree is a function that takes the input features of a photon and calculates a score. The score from each tree is summed and scaled to determine the BDT score. For true photons, the BDT score tends to be near 1 whereas for jet-induced photons, it is near 0.

Four BDTs were trained: 2× for EB and EE, and 2× for 2016 and 2017/2018. The sizes of the samples used for the training are listed in Table 5.1. The signal and background sizes have approximately a 1:1 ratio. In addition, the complexity of the BDTs has been tuned to ensure that the predictive power of the BDT model is reproducible in independent samples. For this purpose, *validation samples* were set aside for both prompt and fake photons. The package *Hyperopt* [51] has been used to tune the model complexity parameters (*hyperparameters*) by minimizing the difference in performance (measured by the accuracy of the predictions) between training and validation samples. An additional independent sample (*testing sample*) is used to test the performance of the trained model and ensure that the tuning of the hyperparameters does not introduce a bias towards the validation sample.

	Barrel		**endcaps**	
	2016	2017/2018	2016	2017/2018
Prompt γ	19.7M	20M	4.6M	4.4M
Jet Fakes	19.7M	20M	4.6M	4.4M

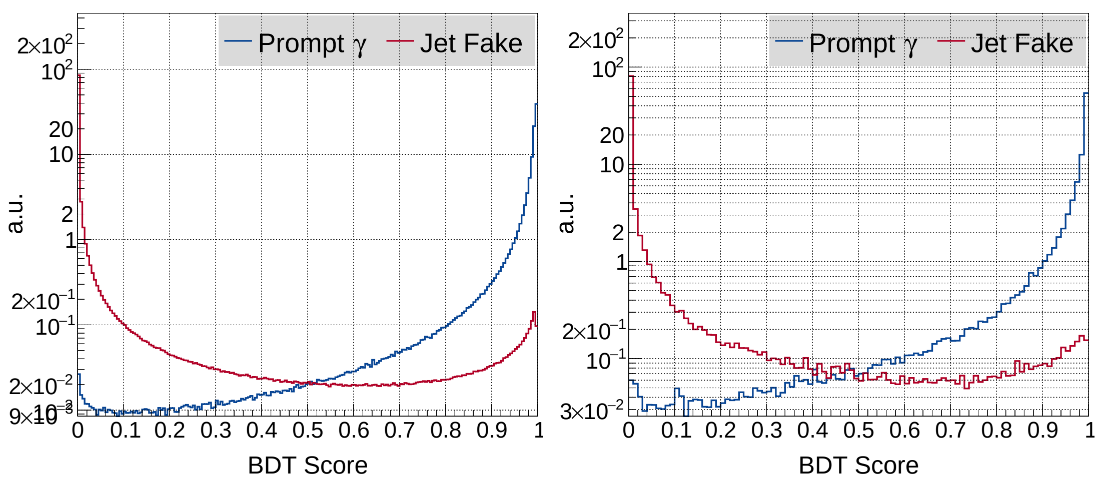
Table 5.1: Number of events in the training samples used for photon BDT training

The performance of the BDT models is shown in Figure 5.6, where the signal efficiency is plotted against the background efficiency as the threshold on the BDT score is varied between 0 and 1. The performance is measured in three sets of samples: the training sample, the validation sample used for tuning the complexity of the BDT model, and an independent testing sample. The area under the curve (AUC) is representative of the predective power of the model (area of 1 for perfect model, 0.5 for no predictive power). The AUCs are over 0.99 in all three samples, indicating that the model has excellent predictive power. In addition, the close overlap between the three curves corresponding to independent samples shows that the model's predictive power is reproducible. The shapes of the distributions of the BDT score for simulated true and fake photons in the barrel and endcap ECAL are shown in Figure 5.7. For the sample of true photons, a sharp peak is seen at 1 whereas the fake photon sample peaks at 0. The BDT distributions for the signal and background samples are well-separated, thus providing a strong discrimination power.



(a) Barrel ECAL (b) Endcap ECAL

Figure 5.6: ROC curves for the BDT model trained for 2017 data. The test samples are independent of the training samples, and they show similar performance.



(a) Barrel ECAL (b) Endcap ECAL

Figure 5.7: BDT distributions for true (prompt) and jet-induced (fake) photons using the model trained for 2017 data. The overlap between signal and background distributions is lower here than that in any single shower shape variable.

5.2 Isolation variables

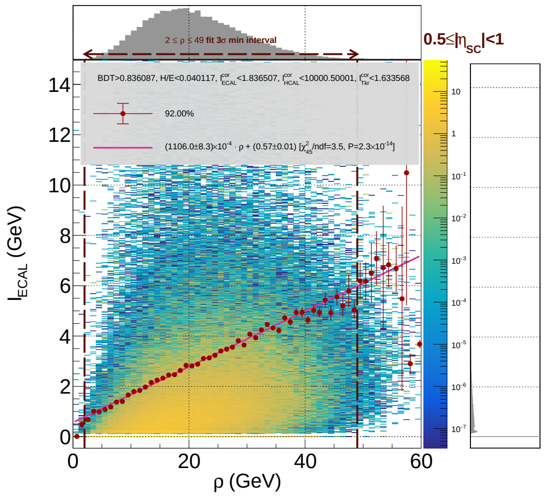
The isolation variables are constructed by summing the deposits in the subdetectors around the photon. The region within which these sums are taken is defined by an upper limit on $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, where η is the pseudorapidity and ϕ is the azimuthal angle. Three isolation variables are employed:

- I_{ECAL}: Sum of transverse energies of ECAL hits not associated with the photon cluster within a cone of $\Delta R < 0.3$ around the photon

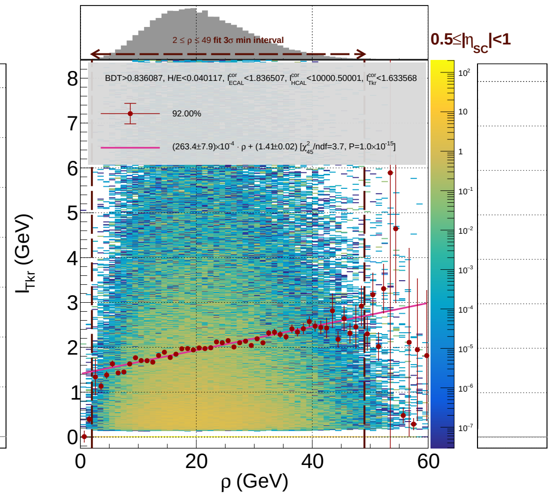
- I_{Tkr}: Sum of transverse energies of tracks within a cone of $\Delta R < 0.3$ around the photon

- H/E: The ratio between the energy deposited in the HCAL in a cone of radius $\Delta R < 0.15$ around the photon and the energy of the photon

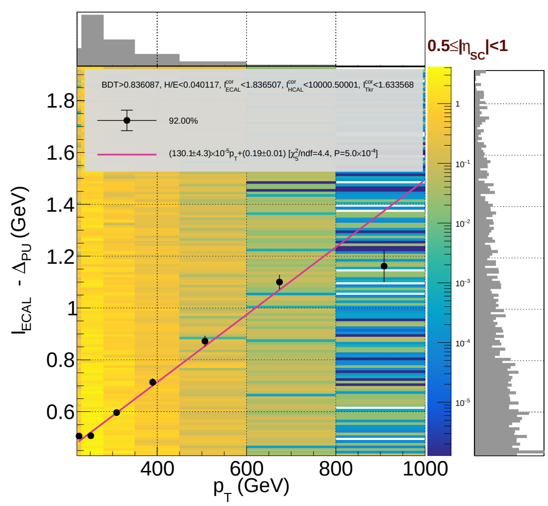
In their raw form, I_{ECAL} and I_{Tkr} contain significant contributions from the products of pileup interactions that are not associated with the hard collision. In addition, due to the imperfect efficiency of the ECAL clustering algorithm, some hits are mis-assigned to I_{ECAL}. These effects in simulated true photons are shown in Figure 5.8 for the pseudorapidity range $0.5 < |\eta| < 1$, where 2D distributions of the isolation variables are plotted against ρ and the photon p_T. Here ρ, called the *pileup energy density*, is proportional to the median jet energy in the event and is representative of the pileup activity in the event [52]. The black points on the graphs represent the 92nd percentile along the y-direction. These points show an approximately linear relationship between the isolation variables, and ρ and p_T. We derive a correction for each of these dependencies by fitting a linear function to the points. This linear dependence is subtracted from the isolation variables. Deriving the corrections at the 92nd percentile ensures approximately that the signal efficiency remains flat at 92% against both ρ and p_T. The corrections are derived in 6 bins of pseudorapidity ($|\eta_{SC}|$) separately for each year of data-taking (see Tables 5.2, 5.3 and 5.4).



(a) Pileup dependence of I_{ECAL}



(b) Pileup dependence of I_{Tkr}



(c) p_T dependence of pileup-corrected I_{ECAL}

Figure 5.8: Pileup and p_T corrections for photon isolations are estimated from 2D distributions in simulation of true photons. The points represent the 92nd percentile of the 1D isolation distributions in the x-bins

ρ corrections to $I_{\mathbf{ECAL}}$

$\lvert\eta_{SC}\rvert$	*2016*	*2017*	*2018*
0 - 0.5	0.125259ρ	0.107304ρ	0.108502ρ
0.5 - 1	0.126289ρ	0.110602ρ	0.108592ρ
1 - 1.4442	0.116536ρ	0.103056ρ	0.096517ρ
1.566 - 1.7	0.08152ρ	0.072804ρ	0.065384ρ
1.7 - 2	0.087155ρ	0.080676ρ	0.076876ρ
2 - 2.5	0.114965ρ	0.101608ρ	0.084448ρ

Table 5.2: Pileup corrections to ECAL isolation

ρ corrections to $I_{\mathbf{Tkr}}$

$\lvert\eta_{SC}\rvert$	*2016*	*2017*	*2018*
0 - 0.5	$0.1230\rho + 1.66$	$0.0242\rho + 1.43$	$0.0276\rho + 1.35$
0.5 - 1	$0.1223\rho + 1.67$	$0.0263\rho + 1.41$	$0.0347\rho + 1.25$
1 - 1.4442	$0.1234\rho + 1.54$	$0.0265\rho + 1.38$	$0.0270\rho + 1.39$
1.566 - 1.7	$0.1120\rho + 1.53$	$0.0159\rho + 1.55$	$0.0257\rho + 1.42$
1.7 - 2	$0.1080\rho + 1.56$	$0.0320\rho + 1.30$	$0.0298\rho + 1.35$
2 - 2.5	$0.1022\rho + 1.31$	$0.0263\rho + 1.29$	$0.0337\rho + 1.15$

Table 5.3: Pileup corrections to tracker isolation

p_T corrections to $I_{\mathbf{ECAL}}$

$\lvert\eta_{SC}\rvert$	*2016*	*2017*	*2018*
0-0.5	$0.000723p_T + 0.247$	$0.000912p_T + 0.309$	$0.000634p_T + 0.318$
0.5-1	$0.001068p_T + 0.196$	$0.001300p_T + 0.193$	$0.001108p_T + 0.181$
1-1.4442	$0.001690p_T + 0.0451$	$0.002382p_T - 0.073$	$0.001327p_T + 0.136$
1.566-1.7	$0.000291p_T + 0.155$	$0.000364p_T + 0.164$	$0.000808p_T + 0.124$
1.7-2	$0.000784p_T + 0.024$	$0.001235p_T - 0.071$	$0.000482p_T + 0.143$
2-2.5	$0.001644p_T - 0.217$	$0.001637p_T - 0.246$	$0.001213p_T - 0.157$

Table 5.4: p_T corrections to ECAL isolation

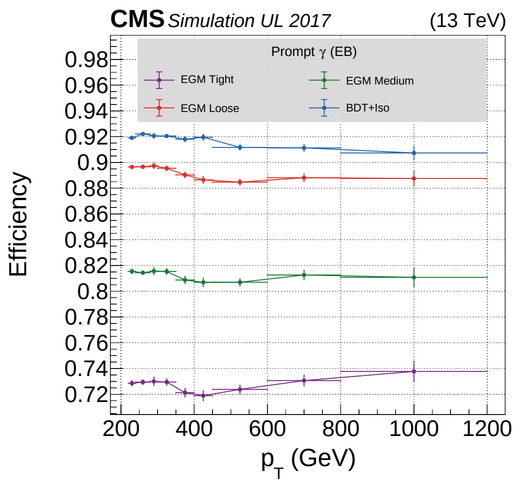
5.3 Cut Optimization

The three isolation variables (I_{ECAL}, I_{Tkr} and H/E) and the BDT score serve as discriminating variables for photon identification. In order to maximize the rejection of jet fakes while maintaining high signal selection efficiency, we impose thresholds (cuts) on each of these variables. To tune the cuts, we developed a 4D optimizer, which takes as inputs simulated signal and background samples, the target signal efficiency, and an initial search grid. Starting with the initial search grid, the optimizer recursively narrows the search grid and descends towards cuts that minimize fake background efficiency while maintaining signal efficiency above the target. It takes 10-15 iterations ($\sim$ 10 minutes) for the cut boundaries to stabilize within $\sim 0.1\%$, when we terminate optimization. We perform this optimzation separately for the barrel ECAL and the endcap ECAL for each of the three years of data taking. Table 5.5 lists the cuts obtained.

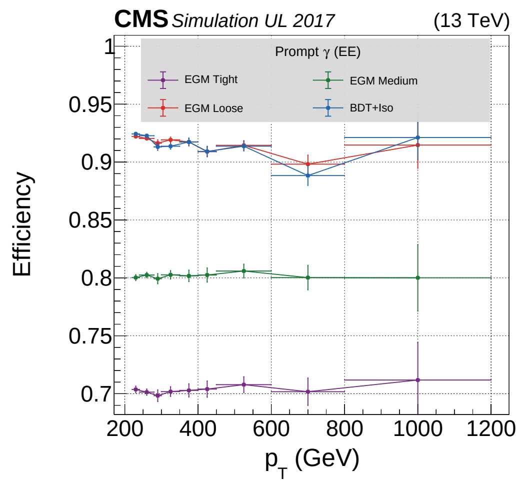
		2016	**2017**	**2018**
Barrel	BDT Score >	0.849	0.836	0.847
	$H/E <$	0.022	0.040	0.040
	$I_{\text{ECAL}}/\text{GeV} <$	2.16	1.84	1.84
	$I_{\text{Tkr}}/\text{GeV} <$	2.19	1.63	1.58
Endcaps	BDT Score >	0.571	0.604	0.695
	$H/E <$	0.032	0.017	0.018
	$I_{\text{ECAL}}/\text{GeV} <$	2.69	2.54	2.94
	$I_{\text{Tkr}}/\text{GeV} <$	3.10	1.82	1.90

Table 5.5: Cuts on the discriminating variables for photon identification

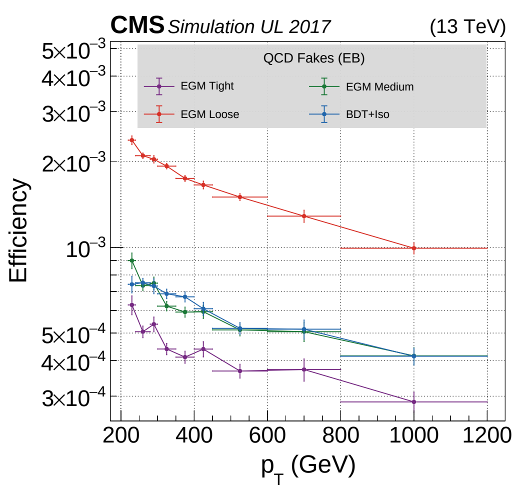
In Figure 5.9, we show the efficiency of selecting true and fake photons in simulated data (under 2017 data-taking conditions) as a function of photon p_T. In addition, we show for comparison the performance of three standard identification methods based on cuts applied to isolation variables and one shower-shape variable ($\sigma_{i\eta i\eta}$). These are indicated in the figure by *EGM Tight*, *EGM Medium* and *EGM Loose*. The efficiency plots show that the use of BDTs trained with several shower shape variables leads to significant improvement in discrimination. Our selection criteria (indicated by *BDT+Iso*) achieves higher signal efficiency than *EGM Loose* while maintaining similar or better background rejection than *EGM Medium*.



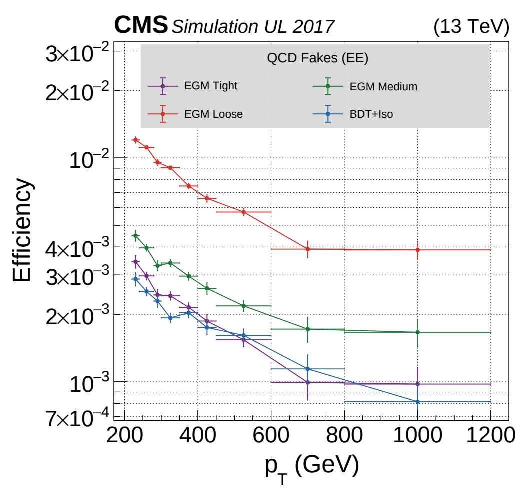
(a) True photons in the barrel ECAL



(b) True photons in the endcap ECAL



(c) Fake photons in the barrel ECAL

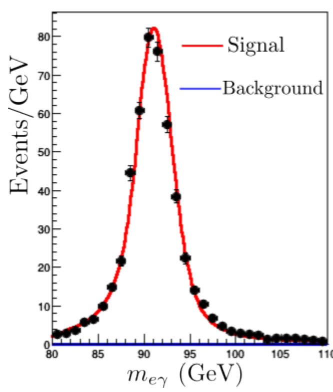


(d) Fake photons in the endcap ECAL

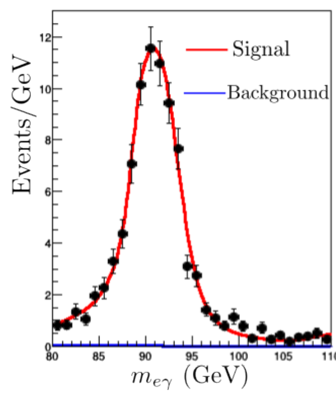
Figure 5.9: True and fake photon selection efficiency in simulated data as a function of photon p_T under our identification criteria (BDT+Iso). For comparison with purely cut-based methods, three other identification criteria are also shown (EGM Tight, EGM Medium and EGM Loose).

5.4 Efficiency Scale Factors

The simulation of photon showering and reconstruction is imperfect since radiation damage over the course of data taking impacts the response of the subdetectors. The detector is periodically re-calibrated using standard candle physics events (e.g. $Z \to ee$ peak) and the simulation conditions are updated; however, granular *in situ* calibration of the detector is not possible and imperfections remain in the simulation. These lead to differences in the photon identification efficiency betwen real data and simulated data. Typically, the efficiency is higher in simulation than in real data.



(a) Probes passing the ID



(b) Probes failing the ID

Figure 5.10: Fits of the invariant mass distributions of $Z \to ee$ pairs where the probe electron passes and fails the photon selection criteria. The probe electron is required have $0 < \eta < 0.6$ and 250 GeV $< p_T <$ 300 GeV.

We correct for the discrepancies by estimating *scale factors*, which are applied as weights to the simulated events in the analysis. These weights are determined using the well-understood physics process $pp \to Z \to ee$. In this procedure, called the *tag and probe* method, a sample of $Z \to ee$ events is selected in real data by applying high-purity identification criteria on the sub-leading (the *tag*) electron and loose criteria on the leading electron (the *probe*). The photon identification criteria for the analysis are applied to the ECAL component of the probe electron (which acts as proxy for photons) in order to split the sample into two parts: N_{pass} events with a probe electron that passes the photon identification criteria and N_{fail} events whose probe electron fails the criteria. In the determination of the event counts, N_{pass} and N_{fail}, the invariant mass distribution is fitted with an analytic function for the Breit-Wigner lineshape of the Z boson, determined from simulation along with a Gaussian convolution to account for the finite resolution of the ECAL, and the background is modeled as the product of an exponential and the Gaussian error function. The efficiency in real data is calculated as:

$$\epsilon_{\text{data}} = \frac{N_{\text{pass}}}{N_{\text{pass}} + N_{\text{fail}}}$$

The efficiency, ϵ_{MC}, in simulated data is measured separately and the scale factor is calculated as the ratio:

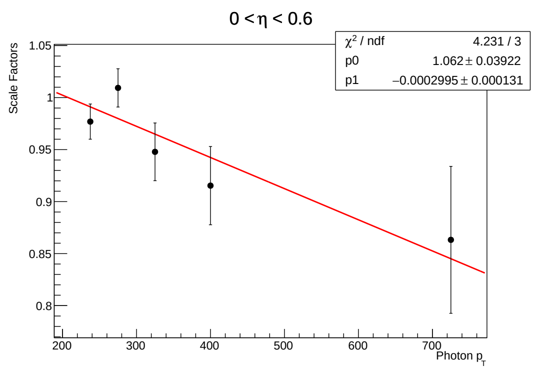
$$\text{Scale factor} = \frac{\epsilon_{\text{data}}}{\epsilon_{\text{simulation}}}$$

The scale factors are calculated with granularity in both pseudorapidity η and photon p_T in order account for variations in these variables. An example is shown in Figure 5.10, where the probe electron is required have pseudorapidity range $0 < \eta < 0.6$ and p_T in the range 250 GeV $< p_T <$ 300 GeV.

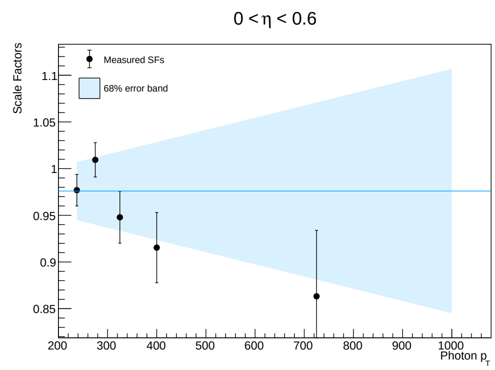
Systematic uncertainties are assigned to the scale factors due to three sources:

- *Choice of Z peak model:* The scale factors are re-computed with the Double Crystal Ball as alternative signal fitting function and the difference in scale factors is taken as a systematic uncertainty.

- *Choice of background model:* The background model is changed to an exponential function and the difference in scale factors is taken as a systematic uncertainty.



(a) Linear fit of scale factors

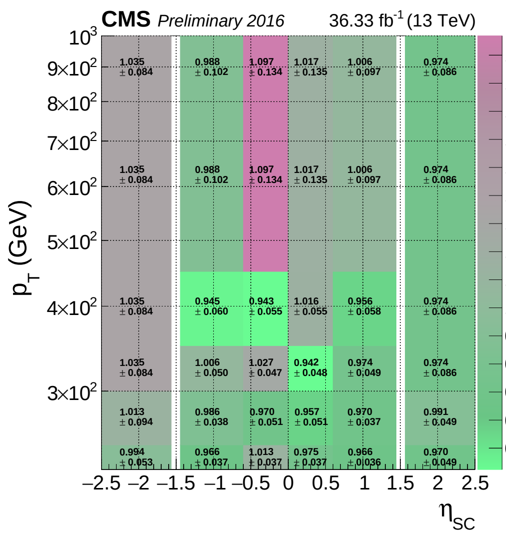


(b) Uncertainty band

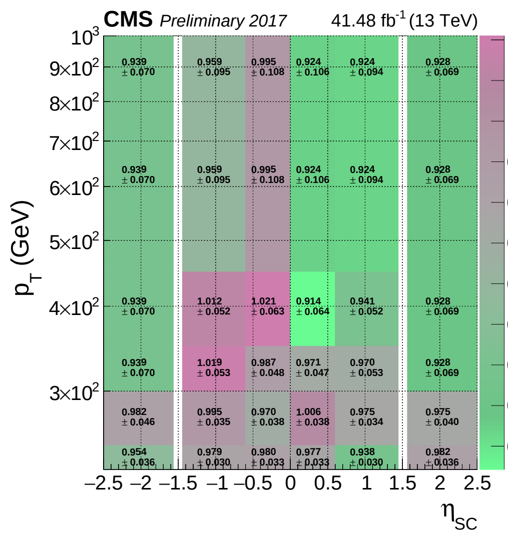
Figure 5.11: Extrapolation uncertainties assigned to scale factors for photons simulated in $0 < \eta < 0.6$ and 250 GeV $< p_T <$ 300 GeV.

- *Extrapolation uncertainty:* The sample size of $Z \rightarrow ee$ event in real data is very limited at high photon p_T. We assume that the scale factors remain flat with p_T beyond the range of available statistics. An uncertainty is assigned for this assumption by fitting the scale factors with a linear function of p_T and taking as a systematic the uncertainty band given by varying the slope by one standard deviation (as obtained from the fit) in positive and negative directions. Figure 5.11 shows the uncertainty estimation for simulated photons with $0 < \eta < 0.6$ and 250 GeV $< p_T <$ 300 GeV for the 2017 data-taking period.

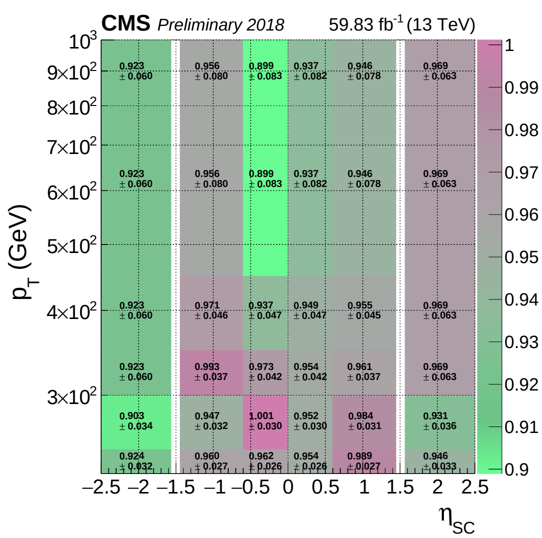
The measured scale factors and their uncertainties for the three years of data-taking are shown in Figure 5.12.



(a) 2016



(b) 2017



(c) 2018

Figure 5.12: Scale factors to correct for mismodeling of photon selection efficiency in simulated data. The corrections are derived separately for different ranges of pseudorapidity (η_{SC}) and photon transverse momentum (p_T)

Chapter 6

Analysis of monophoton events

In this chapter, we present the methodologies employed to carry out the two main measurements of this thesis. The first measurement pertains to the cross-section of $pp \to Z(\nu\nu) + \gamma$ in p-p collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV. The second measurement involves determining the limits on the strength parameters of the anomalous neutral triple gauge couplings (aNTGCs), specifically (h_3^Z, h_4^Z) and (h_3^γ, h_4^γ). To conduct this study, we analyzed data from pp collisions, with an integrated luminosity of 138 fb^{-1}, which were recorded by the CMS detector during Run II spanning the years 2016 to 2018.

The events of interest in these measurements are characterized by the presence of a high-transverse-momentum photon and significant missing transverse momentum. However, this particular signature is not exclusive to $Z\gamma$ production and can be imitated by various other *background* processes. Therefore, we elaborate on the selection criteria implemented to isolate events that enhance the signal while minimizing the presence of background events. The primary backgrounds are assessed using data-driven techniques, whereas the secondary backgrounds are estimated through simulation. By subtracting these background estimations from the observed data, we are able to determine the number of signal events originating from the production of $Z(\nu\nu) + \gamma$.

The transverse momentum of the photon in these events serves as a critical observable for the measurement of aNTGCs. If aNTGCs exist, it is predicted that an excess will manifest in the high-energy tail of this distribution, deviating from the predictions of the Standard Model (SM). Additionally, the same distribution is used to differentially measure the cross-section based on the transverse momentum of the photon. The techniques employed to perform these measurements are based on the *maximum likelihood* formalism, and we discuss their implementation in detail.

6.1 Analysis Strategy

The final state of the process $pp \to Z(\nu\nu)+\gamma$ consists of a photon and missing transverse momentum. At low photon transverse momentum (p_T), the rate of events is dominated by background processes arising from QCD multijet production and the production of $\gamma +$ jets. Therefore, a high threshold is applied on the photon p_T to select the events of interest. In addition, a large missing transverse momentum (denoted by $\not{E}_T$) is also required. In this region of phase space, known as the *signal region*, a number of background processes mimic the signature of our signal. Additional criteria are applied to suppress these contributions. To estimate the sizes of the remaining backgrounds, we select specific control samples, or *control regions*, and extrapolate their contributions in the signal region. An overview of the background processes is shown in Figure 6.1. These processes are:

- **$e \to \gamma$ (electrons misidentified as photons)**
  The largest background arises from the production of W bosons that decay into an electron and a neutrino, where the electron is misidentified as a photon. During the reconstruction process, this can happen if no track is reconstructed due to the finite efficiency of the tracker or if the algorithm for associating tracks with the ECAL hits fails. Particularly, when the electron from the decay of a single W boson decay is misidentified as a photon, the event has the signature of a photon accompanied by missing transverse momentum. To estimate the rate of misidentified electrons, we employ a *fake ratio* technique on an electron-enriched control sample. This involves scaling the event yield of the control sample by the fake ratio, which represents the rate of *fake photons* in the region of interest for each *bona fide* electron in the control sample.

- **$W + \gamma$ background**
  The production of a W boson and a photon is the second largest source of background to our signal. In these events, the W boson decays into a charged lepton

and a neutrino. Occasionally, the lepton escapes the detector's geometric acceptance, particularly when it has a low momentum, resulting in a tightly curved trajectory within the magnetic field. As a result, the lepton remains undetected, mimicking the signature of our signal. To estimate this background, we select two control regions enriched with $W(\rightarrow e\nu) + \gamma$ and $W(\rightarrow \mu\nu) + \gamma$ events. We use a simulation of $W + \gamma$ to determine the efficiency and acceptance in the signal region as well as the control regions. By constructing a joint likelihood function, we perform a simultaneous fit of all three regions, incorporating free parameters that scale the simulated efficiency and acceptance. These parameters, determined in the fit, constrain the background size by extrapolating the measurements from the control regions.

- **Jet$\rightarrow$ γ (jets misidentified as photons)**
  Proton-proton collisions predominantly produce hadrons in the form jets. When a jet contains a highly energetic π^0 or η meson, the $\gamma\gamma$ pair from the decay can appear as an isolated and collimated deposit in the ECAL like a single isolated photon. This contributes as a background when missing transverse momentum is simultaneously produced in the event, particularly from the production of an invisibly decaying Z in association with jets. We estimate the size of this background by the *fake ratio* method, which uses a control sample enriched with events with a non-isolated jet and missing transverse momentum to extrapolate to the signal region via the fake ratio.

- **Minor SM background**
  Smaller contributions to the background arise from a number of processes. Events with a photon and jets in the final state can be identified as signal when the jet energy is significantly mismeasured, resulting in a large missing transverse momentum being reconstructed. These events arise primarily from the production of $\gamma +$ jets, with smaller contributions from top quark processes like $t + \gamma$ and $t\bar{t} + \gamma$. Another background arises from processes producing a W boson or the di-bosons WZ, ZZ or WW. In these events, a boson decays leptonically and the lepton emits Bremsstrahlung radiation in the ECAL, resulting in a photon being reconstructed. Missing transverse momentum is also present in these events as the

decays also produce neutrinos. The size of these background sources is estimated using simulation samples of the processes.

- **Beam halo background**
  When proton beams interact with residual gas or beampipe material in the accelerator, charged pions ($\pi^{\pm}$) are produced. These pions generate highly energetic muons that travel almost parallel to the beamline. As these muons pass through the detector, they can emit photons via Bremsstrahlung radiation, which are detected in the ECAL. In the barrel ECAL, the photon deposits form a straight line parallel to the beamline, enabling the rejection of this background in the barrel region. However, this distinctive feature is not observed in the endcap ECAL. To address this, a BDT-based algorithm has been developed to reject beam halo background in the endcap region, extending the pseudorapidity coverage of photons in this study. The remaining background in the endcaps is estimated by statistically fitting the distribution of the photon's azimuthal angle to the characteristic shapes of real photons (which have a uniform azimuthal distribution) and beam halo photons (which forms concentrated peaks at 0 and π).

- **Barrel ECAL spikes**
  Hadrons produced in proton collisions sometimes interact directly with the photosensors (avalanche photodiodes, or APDs) attached to the scintillation crystals of the barrel ECAL. They produce large anomalous signals through direct ionization of the silicon and are known as a *spikes.* Spikes have a characteristically narrow shower shape and are misidentified as highly energetic photons. Because of this mis-reconstructed energy, a large missing transverse momentum is also falsely reconstructed, which gives the event the same signature as $Z(\rightarrow \nu\nu) + \gamma$ events. To estimate the size of this background, a statistical fit is performed by comparing the timing distribution of observed photons to shape templates (probability density functions) representing real photons and spikes. This method takes advantage of the fact that spike pulses have a shorter rise time and faster decay compared to the response of scintillation light from genuine photons in the ECAL.

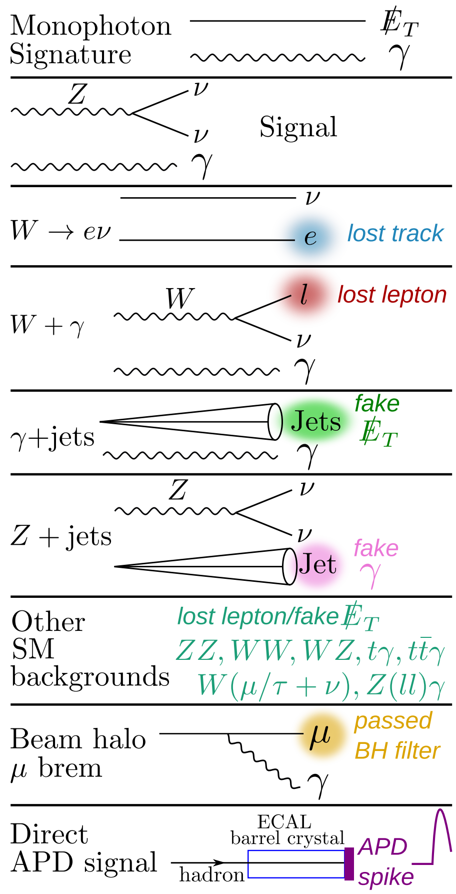


Figure 6.1: Overview of background sources

The events selected in the signal region are divided into distinct intervals, or *bins*, of photon transverse momentum, p_T. This allows us to test for the presence of aNTGCs, which is predicted to manifest as an excess, with respect to the predictions of the SM, in the number of $Z(\nu\nu) + \gamma$ events observed at the high-energy tail of the p_T distribution. An optimization algorithm, known as *Bayesian Blocks binning* [53], is used to choose the bin boundaries. The background processes are individually estimated for each bin. To determine the number of events originating from the signal process, we perform a *binned maximum likelihood fit* of the data to a PDF (probability distribution function) that models the signal and background processes as well the systematic uncertainties. The fit subtracts the background from the data observed in the signal region and provides the best estimate for the number of events originating from $Z(\nu\nu) + \gamma$. These signal events are interpreted separately for the measurement of the cross-section and the measurement of aNTGCs through the use of free parameters.

6.2 Datasets

The dataset employed in this analysis corresponds to the data collected by the CMS detector during LHC Run 2 spanning the period from 2016 to 2018, amounting to a total integrated luminosity of approximately 138 fb^{-1}. The CMS data is organized into distinct *primary datasets* [54] based on the outcomes of the High-Level Trigger selection process. Specifically, we use the primary datasets *SinglePhoton* for the years 2016 and 2017, and the dataset *EGamma* for 2018. These datasets contain comprehensive records of the events that satisfy the trigger requirements specified for this analysis. Each year's dataset is analyzed individually, including the assessment of background contributions. The results obtained from each year are combined to obtain the measurements of the cross-section and aNTGCs.

6.3 Simulations

A number of simulation samples are employed in this study to build the models of the signals and background contributions. These simulations are performed in two steps: simulation of the collisions and simulation of the detector response. The collision simulation is performed using generator packages based on the Monte Carlo technique, which employ theoretical models and known physics processes to generate a range of possible particle interactions with a probability proportional to the phase space and the matrix element for the interactions. In addition, the generators model the parton showering, a process through which high energy partons (quarks and gluons) split into lower-energy partons, and the hadronization process, which transforms the partons into colorless hadrons. The detector simulation step propagates the outgoing particles from the collision step, and simulates the energy deposition and the response of the subdetectors. It uses a detailed model of the magnetic field, the geometry of the detector, the materials within the subdetectors, and the readout electronics. In addition, energy deposition from multiple proton interactions occurring within the same bunch crossing (i.e. pileup)

	Sample	QCD order	Generator packages
Signal	$Z(\nu\nu)+\gamma$	LO	Madgraph5_aMC@NLO Universal FeynRules Output Pythia8
Background	γ+jets $W(l\nu)+\gamma$	LO	Madgraph5_aMC@NLO Pythia8
	$t\bar{t}+\gamma$ $t+\gamma$ $\gamma\gamma$ $Z(ll)+\gamma$	NLO	
	WW WZ ZZ $W(\rightarrow\mu\nu)$ $W(\rightarrow\tau\nu)$	LO	Pythia8

Table 6.1: List of simulation samples used in this analysis

is simulated by overlaying the generated events with randomly selected events from a sample of inelastic proton collisions representative of the pileup interactions [55].

The samples for the SM and aNTGC signal ($pp \rightarrow Z(\nu\nu)+\gamma$) are generated using the generator package Madgraph5_aMC@NLO [56, 57]. The process is generated at QCD and electroweak leading order (LO) with up to two jets in the final state. The aNTGC model is incorporated in the simulation with a Universal FeynRules Output (UFO) model [58], which produces event weights for 440 different sets of non-zero coupling parameter values for each coupling (h_3^γ, h_4^γ) and (h_3^Z, h_4^Z). The generator is interfaced with Pythia 8.2 [59] for simulating both parton showering and hadronization, and the detector response simulation is performed with Geant 4 [60]. For the estimation of $W+\gamma$ background and minor background sources, a number of simulation samples are used. These have been produced within the CMS collaboration for common use in physics analyses. The samples used are summarized in Table 6.1.

6.4 Event Selection Criteria

In order to be sensitive to the signal process, we apply a number of criteria to select the events from the collected data. These criteria, summarized in Table 6.2, are designed to mitigate the presence of background events that mimic the signal signature, while still ensuring a substantial number of signal events for robust statistical analysis.

The first criterion is that the event must pass the *trigger path* for a single photon. This trigger path is configured in the triggering system to store the full detector readout

Trigger	2016: HLT_Photon175 2017, 2018: HLT_Photon200
Photon	Exactly 1 γ in the event passing: • $p_T^{\gamma} > 225$ GeV • $\lvert\eta_{SC}^{\gamma}\rvert < 1.4442$ (barrel) or $1.566 < \lvert\eta_{SC}^{\gamma}\rvert < 2.5$ (endcaps) • 1 or fewer matching hits in the pixel detector • BDT-based identification criteria • Barrel photons satisfy $E^{\text{halo}} < 4.9$ GeV, $\sigma_{i\eta i\eta} > 0.001$, $\sigma_{i\phi i\phi} > 0.001$, and $\eta_W > 0.01$ • Endcap photons pass BDT-based beam halo tagging algorithm
Missing transverse momentum	$\not{E}_T > 200$ GeV $p_T^{\gamma}/\not{E}_T < 1.4$ $\Delta\phi(\not{E}_T, \gamma) > 2$
Jets	No jet in the event passing: • $\Delta\phi(\text{jet}, \not{E}_T) < 0.5$ • $p_T^{\text{jet}} > 30$ GeV • $\lvert\eta^{\text{jet}}\rvert < 5$.
Muons	No μ in the event passing: • $p_T^{\mu} > 10$ GeV • $\lvert\eta^{\mu}\rvert < 2.4$ • Loose identification criteria
Electrons	No e in the event passing: • $p_T^{e} > 10$ GeV • $\lvert\eta_{SC}^{e}\rvert < 1.4442$ or $1.566 < \lvert\eta_{SC}^{e}\rvert < 2.5$ • Loose identification criteria

Table 6.2: List of criteria applied to select the events in the signal region

whenever there is a deposit in the ECAL resembling an isolated photon with a sufficiently large transverse momentum. During the 2016 data-taking period, the minimum transverse momentum threshold for the trigger path was set at 175 GeV, whereas for the subsequent years of 2017 and 2018, this threshold was raised to 200 GeV. In addition, the trigger path includes an upper limit on the amount of deposit in the region of the HCAL that lies within an *isolation cone* relative to the ECAL deposit. The pseudorapidity and azimuthal coordinates, $(\eta_{\mathrm{H}}, \phi_{\mathrm{H}})$, of the cone are defined with respect to the photon's coordinates, $(\eta_\gamma, \phi_\gamma)$, by $\sqrt{(\eta_\gamma - \eta_{\mathrm{H}})^2 + (\phi_\gamma - \phi_{\mathrm{H}})^2} < 0.14$. The total deposit in this region of the HCAL is required to be below 15% (10% for the endcaps) of the size of the ECAL deposit when the photon is in the barrel region of the ECAL. This condition enhances the selection of isolated photons over electromagnetic deposits originating from jets.

Events that pass the trigger path are then required have exactly one isolated photon that passes the BDT-based identification criteria defined in Chapter 5. The photon must have a transverse momentum of at least 225 GeV. In order to avoid selecting electrons, the pixel detector must have fewer than than two matching hits that point towards the ECAL deposit. Additionally, two sets of criteria are employed to reject non-collision background events, namely beam halo and spikes:

- A specialized algorithm is used during photon reconstruction to detect deposits in the ECAL barrel that align to form a line intersecting the photon's cluster, consistent with the trajectory of a beam halo muon radiating via Bremsstrahlung. These deposits are summed to calculate the variable E_{halo}. To filter out such events in the barrel region, a maximum threshold of $E_{\mathrm{halo}} < 4.9$ GeV is applied. To reject this background in the endcaps of the ECAL, a separate algorithm has been developed for this analysis that searches for deposits in the other subdetectors adjacent to the photon's cluster, including the HCAL and pre-shower detector (described in 6.8).

- Spikes in the ECAL barrel exhibit narrow clusters, distinct from the broader showers produced by real photons. In addition, the timing of spikes tend tends to be 10 to 15 ns earlier than that of electromagnetic deposits. To reject spikes, three variables describing the shower shape of the selected photon are subjected to thresholds. The first two variables, $\sigma_{i\eta i\eta}$ and $\sigma_{i\phi i\phi}$, represent the widths of the cluster in the η and ϕ directions, respectively. The third variable, W_η, quantifies the energy sharing between adjacent crystals in the longitudinal direction, calculated as the ratio E_1/E_0, where E_0 is the energy of the crystal with the highest deposit and E_1 is the higher energy of the two neighboring crystals. Based on the localized nature of spike deposits within a single crystal and their minimal energy spread to neighboring crystals, we apply the following criteria to suppress spikes: $\sigma_{i\eta i\eta} > 0.001$, $\sigma_{i\phi i\phi} > 0.001$, and $W_\eta > 0.01$. Furthermore, we require the photon's timing to lie within ± 3 ns of the bunch crossing.

After selecting events containing a single photon, the next set of criteria focuses on the missing transverse momentum ($\vec{\not{E}}_T$) and the jets present in the event. Events with a single isolated high-energy photon are produced primarily by the process $pp \to \gamma + jets$, which has very little intrinsic missing transverse momentum. We remove the majority of these events by applying a lower limit of 200 GeV on the missing transverse momentum. However, the reconstruction of jet energy is limited by the finite resolution of the HCAL, which can result in an imbalance in the total transverse momentum in the event. Specifically, in cases where jet energy is significantly underestimated, a large $\vec{\not{E}}_T$ exceeding the 200 GeV minimum is reconstructed. To address this, we compare the direction of the leading jets with the direction of the missing transverse momentum and require that jets with a transverse momentum greater than 30 GeV be oriented away from the $\vec{\not{E}}_T$ by an azimuthal separation of $\Delta\phi(\text{jet}, \vec{\not{E}}_T) > 0.5$. The remaining background events exhibit a significantly smaller $\not{E}_T$ compared to the photon's p_T, so we impose the condition $p_T/\not{E}_T < 1.4$. However, these criteria alone do not suppress another prominent background originating from the production of W bosons in association with jets, where the boson decays into a neutrino and an electron that is misidentified as a photon. In these events, the W boson recoils against the jets with a large boost, resulting in a smaller azimuthal separation between the electron and the neutrino compared to the separation between the γ and the $Z(\to \nu\nu)$ system in our signal events. To mitigate this background, we select events with an azimuthal separation of $\Delta\phi(\gamma, \vec{\not{E}}_T) > 2$ between the photon and the missing transverse momentum..

The last set of criteria are designed to suppress $W(\to l\nu) + \gamma$ background. A significant portion of these events can be identified in the detector through the presence of an electron or a muon originating from the decay of the W boson. Hence, we discard any event that includes a reconstructed electron or muon with a transverse momentum exceeding 10 GeV and meeting loose identification requirements.

6.5 Estimation of $e \to \gamma$ background

Electrons and photons leave similar shower deposits in the ECAL. In order for an ECAL cluster to be reconstructed as an electron, there must be at least 2 hits in the pixel detector of the tracker that match directionally with the ECAL cluster. When an electron fails to register hits in the pixel layer, the electron is misidentified as a photon. If missing transverse momentum is simultaneously present, the event mimics the monophoton signature and passes the signal selection criteria. These events arise primarily from the production of W bosons that decay into e and ν. The electron in these events are misidentified as photons while the neutrino escapes the detector, producing missing transverse momentum. Smaller contributions to this background also arises from the production of diboson production events, such as $W(\to e\nu) + Z(\to \nu\nu)$.

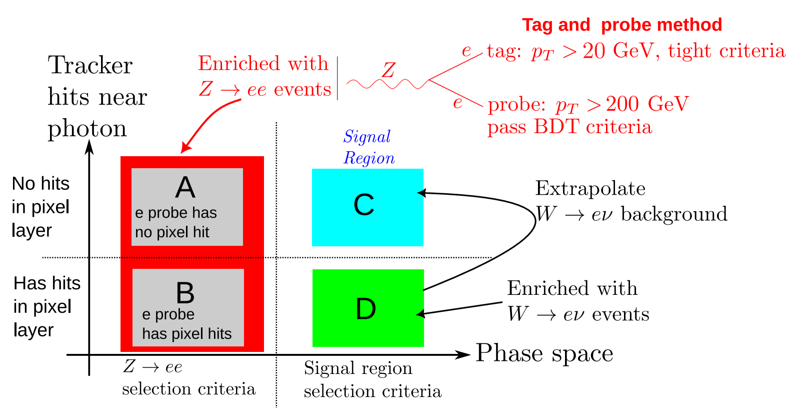


Figure 6.2: Fake ratio method for estimating $e \to \gamma$ background

In order to estimate the size of this background, we use a data-driven method called the *fake ratio* method, alternatively known as the *ABCD* method. The method is illustrated in Figure 6.2, where four distinct phase spaces, labeled A, B, C and D, are defined. Our objective is to calculate the number of electrons misidentified as photons in the signal region, C.

These background events consist of an electron that satisfies the BDT-based identification criteria but is mis-identified as a photon because it lacks any hits in the pixel detector corresponding to its ECAL cluster. The selection criteria for region D are identical to C in all respects except for the condition for the pixel detector: the photons in D must have matching hits in the pixel detector. This requirement ensures that region D is enriched with electrons. We can then assume that the magnitude of the background from $e \rightarrow \gamma$ misidentification in region C is directly proportional to the number of events in region D. This proportionality constant is referred to as the *fake ratio*.

To derive the fake ratio, we introduce two additional regions, A and B. The criteria applied to select the events in regions A and B are specifically designed to isolate events corresponding to the decay of Z bosons into electron-positron pairs. The sole distinction between regions A and B is that the leading electron (the electron with the larger p_T) in region A lacks any matching hits in the pixel detector, resulting in its reconstruction as a photon; whereas the leading electron in region B possesses corresponding hits in the pixel detector. To select events for regions A and B, a *tag and probe* method is used to isolate a sample of $Z \rightarrow ee$ events. The criteria applied are:

- The *tag* electron (sub-leading) passes tight identification criteria and has a p_T larger than 20 GeV
- The *probe* photon (leading) passes the BDT-based identification criteria and has a p_T larger than 200 GeV
- The invariant mass of the electron and photon system lies between 60 GeV and 120 GeV to match the Z mass window

These events are divided among the regions A and B based on whether the probe photon has matching pixel detector hits. Probes that pass the matching criteria are assigned to B while the failing probes are assigned to A.

In order to remove contributions to regions A and B that do not arise from $Z \rightarrow ee$ events, a fit of the invariant mass distribution is performed. The distribution is assumed to be composed of a Z-lineshape and an exponentially falling background. The Z-lineshape is obtained from simulation while the parameters of the background function is determined from the fit. The normalizations of the Z lineshapes of the regions A and B determine the *pixel efficiency*, which is defined as:

$$\epsilon = \frac{N_{\mathrm{B}}}{N_{\mathrm{tot}}}$$

where N_{B} is the number of events corresponding to the Z-lineshape in region B while N_{tot} is the total Z decay yield of regions A and B. Thus, ϵ represents the efficiency of fully reconstructing an electron while $(1-\epsilon)$ is the fraction of electrons that do not have pixel hits in the tracker. ϵ is determined from a simultaneous fit of the invariant mass distributions of regions A and B. The invariant mass distributions for these regions are taken as:

$$\begin{aligned} A: &\quad \hat{N}_{\mathrm{tot}} \cdot (1-\hat{\epsilon}) \cdot Z(m) + b_A \cdot B_A(m) \\ B: &\quad \hat{N}_{\mathrm{tot}} \cdot \hat{\epsilon} \cdot Z(m) + b_b \cdot B_B(m) \end{aligned}$$

where $Z(m)$ is the Z-lineshape PDF (probability density function), $B_A(m)$ and $B_B(m)$ are the background PDFs (exponential decay), and b_A and b_B are the background normalizations. The value of ϵ is determined from a simultaneous fit of the invariant mass distributions of the two regions. The fake ratio is thus obtained as:

$$\text{Fake ratio } = \frac{\text{Number of } Z \text{ events in A}}{\text{Number of } Z \text{ events in B}} = \frac{1-\epsilon}{\epsilon}$$

An example of the fit is shown in Figure 6.3.

The uncertainty in the fake ratios arises mainly from the limited sample sizes of regions A and B. The cross-section of $Z \to ee$ events is extremely low in the phase space of probe $p_T > 200$ GeV. An additional systematic uncertainty is assigned for the choice of the background model in the fits of the invariant mass distribution. The fits are performed with a cubic function as an alternative background model. The difference between the alternative fits and the nominal fits is taken as a systematic uncertainty in the fake ratios.

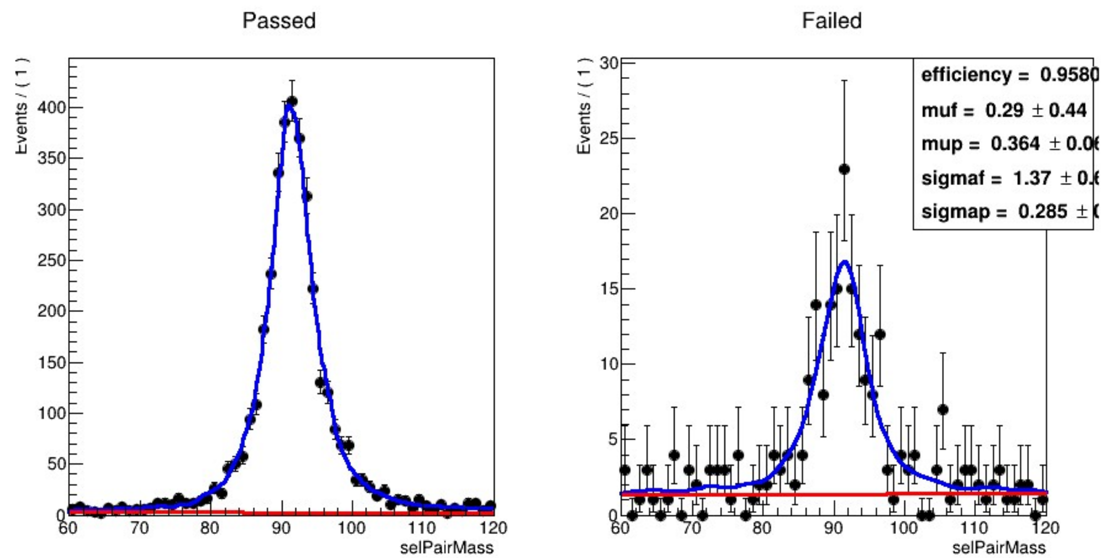


Figure 6.3: Fit of the invariant mass distributions of regions A and B for the pseudorapidity range $|\eta| < 0.55$ corresponding data collected in 2018

The pixel efficiency, ϵ, depends significantly on the pseudorapidity of the electron since, at higher pseudorapidities, an electron shares charge with more pixels and hence has a higher probability of being detected in the pixel layers than at lower pseudorapidities. Therefore, the estimation of the fake ratio is performed in bins of pseudorapidity to account for this variation. Figure 6.4 shows the variation of the fake ratio as a function of pseudorapidity for the data collected in 2017 for the ECAL barrel region. The fake ratios have been fitted to a qudratic function of pseudorapidity.

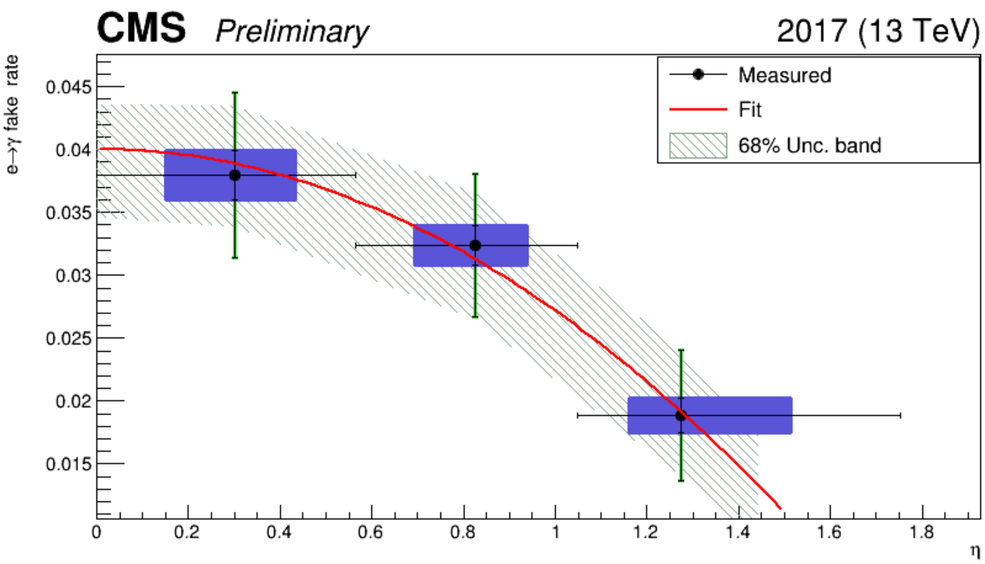


Figure 6.4: Fake ratio for the 2017 data-taking period in bins of pseudorapidity. Systematic uncertainties are shown in blue rectangles. The error bars represent the full uncertainties.

Over the three years of data-taking, some sensors in the pixel layers of the tracker malfunctioned and had a lower efficiency. Regions covered by these sensors had a higher rate of misidentifying electrons as photons. To estimate the background accurately, dedicated fake ratios were estimated for these regions (see Table 6.3).

Year	Region	Fake ratio
2016	$2.4 < \phi < 2.7$ and $\eta < 0$	0.085 ± 0.014
2017	• $\phi > 2.5$and $\eta > 0$ • $1.5 < \phi < 2$ and $\eta > -1$	0.123 ± 0.018
2018	$0.3 < \phi < 0.8$ and $\|\eta\| < 1.4442$	0.29 ± 0.038

Table 6.3: Dedicated fake ratios for dead regions of the pixel detector

The fake ratios for photons in the ECAL endcaps are estimated without any binning in pseudorapidity since the number of $Z \to ee$ events is too low permit a granular estimation. The fake ratios obtained are:

- 2016: 0.0165 ± 0.0046
- 2017: 0.0103 ± 0.0041
- 2018: 0.0058 ± 0.0026

6.6 Estimation of $W + \gamma$ background

The process $pp \to W+\gamma$ appears as a background in our signal region when the W boson decays leptonically, i.e. $W \to l + \nu$, and the charged lepton l fails to be reconstructed in the detector. This primarily occurs when the lepton momentum is low and, due to the strong magnetic field in the detector, the lepton is bent tightly and falls out of the detector acceptance. As a result, these events have the signature of a single photon and missing transverse momentum.

To model the efficiency and acceptance of this background in the detector, we rely on simulation. However, the cross-section of this process has a large theoretical uncertainty due to the absence of higher-order QCD and electroweak effects in present current software. To mitigate the impact of these uncertainties, we introduce two control regions that are enriched with $W + \gamma$ events. Specifically, we select events with a reconstructed electron in the first control region (CR_e) and a reconstructed muon in the second control region (CR_μ).

Initially, the contributions of $W + \gamma$ events to both the signal region and control regions are estimated using simulation. During the likelihood fit stage, the yields of $W + \gamma$ events in all three regions are scaled by a parameter that is allowed to vary freely. In essence, we treat the $W + \gamma$ background yield as an unknown, but the ratios of the yields in the signal region to those in the control regions are constrained by the simulation. Therefore, the simulation provides information about the detector's efficiency and acceptance rates for all three regions, while the fit to the observed data determines the normalization of the $W + \gamma$ background yield through the use of the floating parameter.

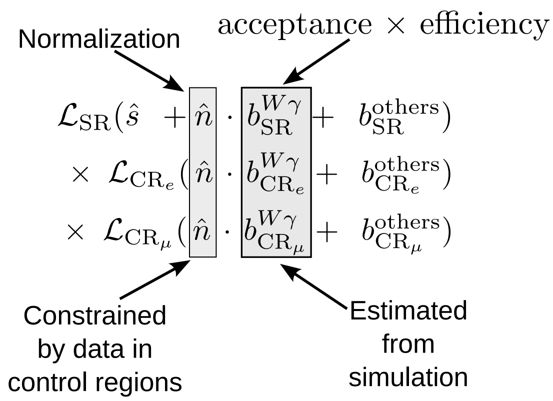


Figure 6.5: Construction of the joint likelihood function for the estimation of $W+\gamma$ background in a single bin of photon p_T. Uncertainty terms are not shown for simplicity.

Figure 6.5 shows a simplified representation of the construction of the joint likelihood function for a given photon p_T bin. The joint likelihood consists of three components: $\mathcal{L}_{\mathrm{SR}}$, $\mathcal{L}_{\mathrm{CR}_e}$ and $\mathcal{L}_{\mathrm{CR}_\mu}$, corresponding to the signal and control regions, respectively. It quantifies the probability of observing events in these regions and involves unknown parameters that are obtained through a fit to the observed data. The signal yield of $Z+\gamma$, denoted as $\hat{s}$, is an unknown free parameter determined by the fit to the data in the signal region. The initial estimates of the $W+\gamma$ background in the three regions are denoted as $b_{\mathrm{SR}}^{W+\gamma}$, $b_{\mathrm{CR}_e}^{W+\gamma}$, and $b_{\mathrm{CR}_\mu}^{W+\gamma}$, and are obtained from simulation. The normalization of this background in all three regions is scaled by a single floating parameter, $\hat{n}$, which is also an unknown. The remaining background estimates in the three regions are represented as $b_{\mathrm{SR}}^{\mathrm{others}}$, $b_{\mathrm{CR}_e}^{\mathrm{others}}$, and $b_{\mathrm{CR}_\mu}^{\mathrm{others}}$ (as determined by the methods described in other sections of this chapter). The normalization parameter $\hat{n}$ is determined through a simultaneous fit to the data in all three regions. This approach effectively mitigates theoretical uncertainties associated with the cross-section of the $W+\gamma$ process by using the measurements from the control regions.

	$\mathbf{CR}_\mu$	$\mathbf{CR}_e$
Photon	Exactly 1 γ passing BDT-based identification criteria	
Lepton	Exactly 1 μ passing tight identification criteria	Exactly 1 e passing tigh identification criteria
Recoil	$\not{E}_T > 50$ GeV $m_T(l, \not{E}_T) < 160$ GeV $R_T > 200$ GeV $p_T^\gamma / R_T < 1.4$ $\Delta\phi(R_T, \gamma) > 2$ where $\vec{R}_T = \vec{\not{E}}_T + \vec{p}_T^{\,l}$ (recoil) and $l = \mu/e$	
Jet veto	No jet passing: • $\Delta\phi(\text{jet}, R_T) < 0.5$ • $p_T^{\text{jet}} > 30$ GeV • $\lvert\eta^{\text{jet}}\rvert < 5.$ • $R(\gamma, jet) > 0.5$ and $R(l, jet) > 0.5$	
Muon veto	No μ (except selected μ in CR_μ) with $p_T^\mu > 10$ GeV and passing loose identification criteria	
Electron veto	No e (except selected e in CR_e) with $p_T^e > 10$ GeV and passing loose identification criteria	

Table 6.4: List of criteria applied to select the events in the $W(\rightarrow \mu\nu) + \gamma$ control region (CR_μ) and $W(\rightarrow e\nu) + \gamma$ control region (CR_e)

The two control regions enriched with $W+\gamma$ events are defined in a similar manner as the signal region with the exception of the requirement of the presence of a reconstructed electron or muon. Both of these regions must satisfy:

- The same high-level trigger requirements as the signal region
- Pass event cleaning filters
- Have exactly one photon in the pseudorapidity range $|\eta| < 1.442$ or $1.566 < |\eta| < 2.5$ with $p_T > 225$ GeV and pass the photon ID requirements

In the control region CR_e, we require the presence of an electron passing *tight* identification requirements [44] defined by criteria on the electron's isolation, ECAL shower shape, and track matching. The electron must have a p_T of at least 30 GeV and lie in the pseudorapidity range $|\eta| < 1.4442$ or $1.566 < |\eta| < 2.5$. In addition, the electron's direction must be separated from the photon by $\Delta R > 0.5$ in order to avoid selecting *converted photons* that produce electron-positron pairs in the detector material.

The control region CR_μ is defined in a similar manner. The events must contain a muon passing tight identification requirements [36] defined by criteria on the muon's track quality and isolation. The muon is also required to have p_T of at least 30 GeV and pseudorapidity within $|\eta| < 2.4$.

We apply further selection requirements common to both CR_μ and CR_e. The missing transverse momentum $\not{E}_T$ must be at least 50 GeV. While the $\not{E}_T$ threshold in the signal region is 200 GeV, we choose the $\nu + l$ (where $l = e/\mu$) system as a proxy for the signal region's $\nu\nu$ system. We define the transverse recoil, $\vec{R}_T$, as the vectorial sum of the charged lepton's transverse momentum $\vec{p}_T^{\,l}$ and the missing transverse momentum $\vec{\not{E}}_T$:

$$\vec{R}_T = \vec{p}_T^{\,l} + \vec{\not{E}}_T$$

The recoil, R_T, must be at least 200 GeV. In addition, we require the ratio of photon p_T^γ to R_T to satisfy $p_T^\gamma / R_T < 1.4$ in order to maintain similarity with the signal region's selection criteria. For the same reason, the jets in these events with a p_T exceeding 30 GeV are required to be separated from the azimuthal direction of R_T by $\Delta\phi(\text{jet}, R_T) > 0.5$. Furthermore, the photon must be separated from recoil in the azimuthal plane by $\Delta\phi(\gamma, R_T) > 2$. In order to reduce background in the control regions, we apply an upper limit of 160 GeV on the transverse mass of the $l + \nu$ system. Lastly, the control regions must not have any extra electron or muon passing *loose* selection criteria [44, 36].

Table 6.4 summarizes the selection criteria for the two control regions.

6.7 Estimation of Jet $\rightarrow \gamma$ background

Collimated γ pairs from the decays of boosted isolated π^0 or η mesons in jets producing low deposits in the HCAL can be misidentified as photons. These *jet fakes* contribute as a background when large $\not{E}_T$ is simultaneously reconstructed in the events. The major sources are $Z(\to \nu\bar{\nu})+$jets production, and processes producing several jets in the final state. In these events, the missing transverse momentum is overestimated due to mismeasurement of the jet energies. We estimate this background by extrapolating the number of events observed in a control region defined by the same selection criteria as for the signal region, but with the photon identification criteria modified to ensure that the photon originates from the electromagnetic deposit of a jet.

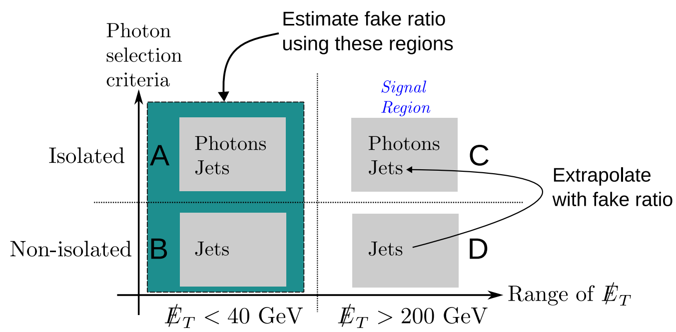


Figure 6.6: Regions used to estimate the background from jets misidentified as photons

In a similar way to the ABCD method for estimating the background from misidentified electrons (Section 6.5), we estimate the number of jets misidentified as photons in the signal region, C, by applying an extrapolation factor, i.e. the fake ratio, to the number of events observed in a control region, D (Figure 6.6). The selection criteria for events in region D are nearly identical to those for region C, with the exception of the photon identification criteria. These criteria are adjusted to ensure that the photon originates from the electromagnetic deposits of a jet. The photon must be surrounded by deposits from other particles, causing it to fail a set of isolation criteria. Specifically, these deposits must be either reconstructed tracks within the photon's isolation cone or electromagnetic deposits around the cluster of the photon. These requirements ensure that the photon is more likely a constituent of a jet, rather than a genuine photon originating from a hard-scattering process in the p-p collision.

The fake ratios used for the extrapolation are estimated using two additional control regions, A and B. Both of these regions are required to have a missing transverse momentum less than 40 GeV in order to be disjointed from regions C and D. Region A consists of two contributions: events producing a genuine photon and jets, and events producing multiple jets. The photon in this region is selected by applying identical identification criteria as in the signal region. For region B, the photon is selected with the same criteria as D. This ensures that the events contain only jets.

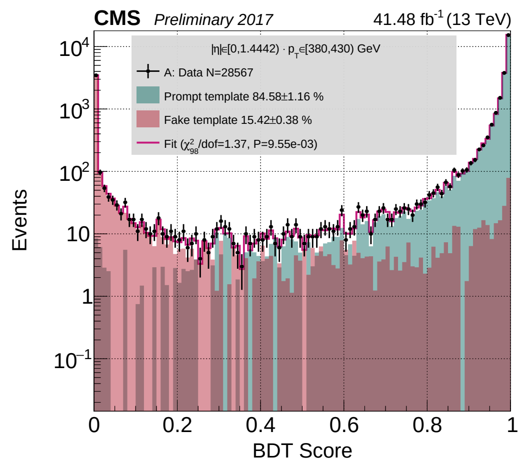


Figure 6.7: Example of a fit of the BDT score distribution of the photons observed in Region A. These photons are in the barrel of the ECAL in the 2017 data-taking period and have a p_T between 380 GeV and 430 GeV.

We estimate the number of events in region A originating from jets by analyzing the BDT score distribution of the photons. We take advantage of the characteristic shapes of the BDT score distributions for genuine photons and photons originating by jets. In Figure 6.7, the observed distribution in region A is shown with black points. The criteria on the BDT score has been dropped to study the full shape of the distribution. We fit it to a combination of two shape templates that represent genuine photons and fake photons:

$$\underbrace{\mathbf{H}(x)}_{\substack{\text{Observed spectrum}\\ \text{in A}}} = N_{\text{genuine}}\cdot\underbrace{\mathbf{h}_{\text{prompt}}(x)}_{\substack{\text{Normalized template}\\ \text{for genuine photons}}} + N_{\text{fake}}\cdot\underbrace{\mathbf{h}_{\text{fake}}(x)}_{\substack{\text{Normalized template}\\ \text{for fake photons}}}$$

The template for genuine photons, $\mathbf{h}_{\text{prompt}}(x)$, is obtained from a simulation sample of the process $pp\to\gamma$+jets. It has a characteristic peak at the value of 1. On the other hand, the template for fake photons, $\mathbf{h}_{\text{fake}}(x)$, peaks at 0. This template has been taken directly from a data sideband chosen by requiring the photon to have a significant deposit in the tracker within the photon's isolation cone. In the specific case shown in the figure, the sideband photons are required to have a tracker isolation (defined as the scalar sum of transverse momenta of tracks within the photon's isolation cone) between

4 GeV and 8 GeV. On the other hand, the criteria applied for selecting genuine photons requires the tracker isolation be less than 1.63 GeV. This requirement on the tracker deposit enriches the template with jet deposits in the ECAL. The sideband otherwise has the same photon identification criteria as region A. To remove contamination from genuine photons in the sideband, we estimate the contribution from a simulation sample of γ+jets and subtract it from the template. This is shown in Figure 6.8.

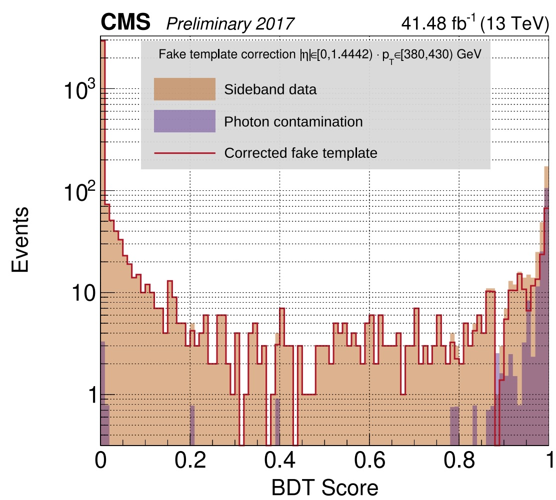
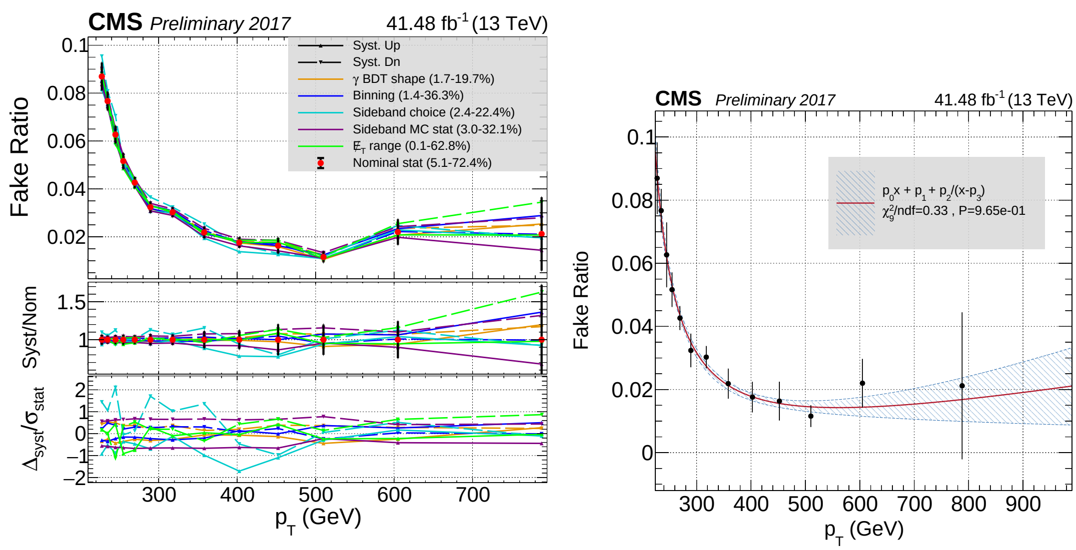


Figure 6.8: Subtraction of contamination from genuine photons in the fake photon template.

By performing this fit, we decompose into the contributions from genuine and from fake photons in region A . Since the BDT criteria was dropped to perform the fit, we integrate the fake template from the minimum threshold on the BDT score to a value of 1. This gives an estimate of the number of fake photons that pass the full identification criteria in the region A. The fake ratio is then obtained as:

$$\text{Fake ratio} = \frac{\text{Number of fake photons in A}}{\text{Number of photons in B}}$$

The size of the background in the signal region, C, is then extrapolated from region D by multiplying its size by this ratio.



(a) Systematic uncertainties in the fake ratio

(b) Parametric fit of the fake ratio as a function of photon transverse momentum

Figure 6.9: Variation of the fake ratios with photon transverse momentum for the barrel region of the 2017 data-taking period

The fake ratios are estimated in bins of photon transverse momentum to account for variation with energy. The sample size drops significantly at high transverse momentum in both simulation and data, and introduces large statistical uncertainties of up to 72% in the fake ratios. In addition, we consider a number of systematic uncertainties in the template fits of region A for the fake ratio estimation:

- **Fake template sideband choice** A correlation is observed between the BDT score of fake photons and their tracker isolation. Since the sideband for the fake photon template is chosen by applying a specific range of allowed values for the tracker isolation, an uncertainty is assigned to account for the bias due to the correlation by varying the upper and lower limits on the tracker isolation by 1 GeV, and re-evaluating the fake ratios. The estimated uncertainty varies between 2.4% and 36% across photon p_T.

- **Photon shower mismodeling**: Since the shape template for the genuine photons is obtained from simulation, we study the mismodeling of the photon shower in the simulation using a sample of $Z \rightarrow ee$ events, where ECAL deposits by electrons serve as a proxy for photons. The mismatch in the simulated and observed BDT distribution is used to derive a correction, known as *quantile morphing technique* [61]. The fake ratios are evaluated with and without these corrections, and the difference is taken as an uncertainty. The estimated uncertainty varies between 8% and 50% across photon p_T.

- **Choice of BDT bin width** The template template fit is performed using binned shapes with 100 bins for the BDT distribution. To account for any biases in this choice, we repeat the fits using 50 bins and 200 bins, and take the variations in the fake ratios as the uncertainty. The estimated uncertainty ranges between 1% and 40% across photon p_T.
- **Choice of the control regions** The control regions A and B used for estimating the fake ratios are chosen to have a missing transverse momentum below 40 GeV. To account for a potential due to this choice, the minimum threshold is varied by 5 GeV in both directions and the fits are repeated. The estimated uncertainty is between 1% and 48% across photon p_T.
- **Finite simulation sample size** The simulation sample of γ+jets has limited size in the sideband used to build the fake photon template. Since this contribution is subtracted from the sideband data, an uncertainty is assigned for the limited sample size. The subtracted contribution is scaled up and down by one standard deviation of this statistical uncertainty, and the fits are repeated. The result uncertainty in the fake ratios is estimated to be between 2% and 32% across photon p_T.

The fake ratios and their systematic uncertainties are derived in bins of photon transverse momentum. The fake ratios are then fitted to a parametric function. Figure 6.9 shows the fake ratios obtained for the data collected in the barrel region of 2017.

6.8 Beam halo background

The phenomenon of *beam halo* [62] occurs at the LHC when protons in the accelerator's beams deviate from the main trajectory and form a diffuse halo around it. This halo contains secondary particles produced when the protons interact with the accelerator's collimators, residual gas or other materials present in the beam pipe. The halo particles, which predominantly travel parallel to the beam, can penetrate the CMS detector and traverse multiple detector cells longitudinally. Among these particles, muons originating from the decay of charged pions within the beam halo radiate in the ECAL and contribute as a background to our analysis. The muons have energies of the order of TeV and their radiation deposits in the ECAL are reconstructed as photons, whose transverse momenta can exceed the 225 GeV threshold applied to select the events in the signal region. As these photons do not originate from the primary collision point, a large missing transverse momentum is also reconstructed, giving the event the same signature as our signal. Dedicated algorithms are implemented in the CMS trigger system and at the offline reconstruction level to filter out events containing beam halo particles [63]. These algorithms look for hit patterns in the cathode strip chambers (CSC) of the muon system consistent with a trajectory parallel to the beam line. In addition, an algorithm at the offline reconstruction level identifies photons induced by beam halo by looking for a trail of deposits in the ECAL barrel (see Figure 6.10) that form a line parallel to the beamline. By imposing an upper limit of 4.9 GeV on the total energy of these deposits, this background is effectively eliminated from the barrel region.

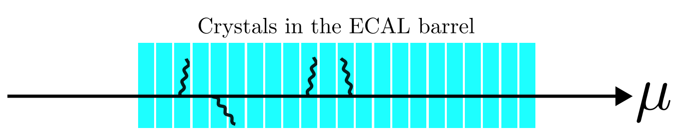


Figure 6.10: Bremsstrahlung radiation by beam halo muons in the ECAL barrel

However, these filtering algorithms are not effective in the endcap region of the ECAL. The coverage of the Cathode Strip Chambers (CSCs) is limited in this region, so they cannot detect muons that travel parallel to the beam axis in this area (see Figure 6.11). Additionally, the muon only passes through one or two crystals of the ECAL due to their projective geometry, so no trail of deposits are formed. To address this issue, we have developed a new tagger based on a BDT that uses information on deposits in the preshower detector and the hadronic calorimeter situated on either side of the ECAL endcaps.

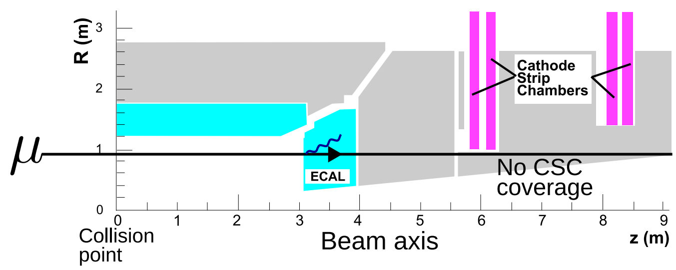


Figure 6.11: Lack of coverage of the cathode strip chambers (CSCs) limiting the identification of beam halo muons

Figure 6.12 shows the concept of the new tagger. Muons from beam halo and photons from the primary collision point have different trajectories with respect to the beam axis. Particles in the beam halo (e.g. hadrons) that accompany the muon leave deposits in the preshower detector and the HCAL. A line connecting the center of any of these energy deposits to the ECAL cluster of the photon tends to be parallel to the beam axis. On the other hand, photons from the primary collision point deposit most of their energy in ECAL, and a line connecting an associated cluster in the preshower or the HCAL forms a non-zero angle with the beam axis.

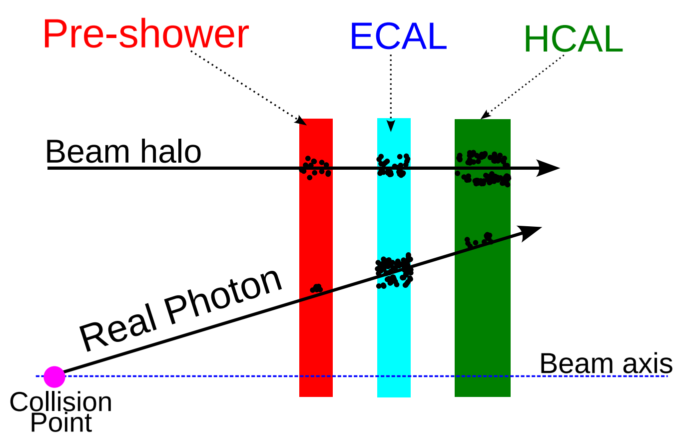


Figure 6.12: Deposits produced by beam halo muons and photons from proton collisions

To form the input variables to the BDT, we construct two sets of preshower/HCAL cluster energies and angles. For a given photon, we form a cluster in the preshower and another in the HCAL that are horizontally adjacent to the ECAL cluster of the photon. The hits in these clusters are required to be within an azimuthal separation of $\Delta\phi < 0.15$

of the photon. The centers of these clusters are connected to the ECAL cluster with a line and the angle of this line with the beam axis is evaluated. This angle corresponds to the *beam halo hypothesis*. Separately, corresponding clusters and an angle are also formed for the *real photon hypothesis* by collecting deposits near the line connecting the collision point to the ECAL cluster. The energies and angles under the two hypothesis are evaluated from a simulation sample of beam halo and a simulation of our signal process to train a BDT. In addition, variables describing the ECAL shower shape of the photon are included in the training. The full set of variables used in the BDT are:

- Four angles formed by clusters under the two hypotheses: two angles corresponding to the preshower and the other two for the HCAL
- Four variables for energy deposited in the clusters
- Five variables describing the shape of the ECAL shower : $E_{3\times3}/E_{\mathrm{SC}}$, $\sigma_{i\eta i\eta}$, $\sigma_{i\phi i\phi}$, σ_η, and σ_ϕ. These are defined in Section 5.1.
- Scalar sum of the transverse momenta of the tracks reconstructed in the event. As beam halo deposits do not arise from the primary collision point, we expect fewer tracks originating from the collision point in these events.

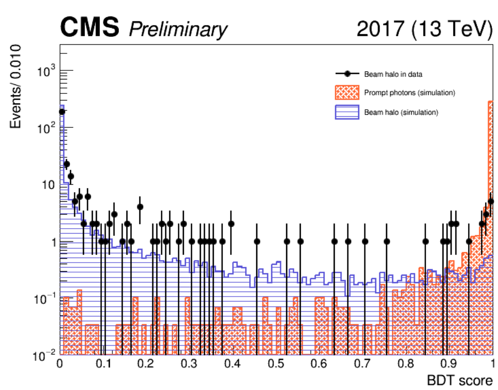
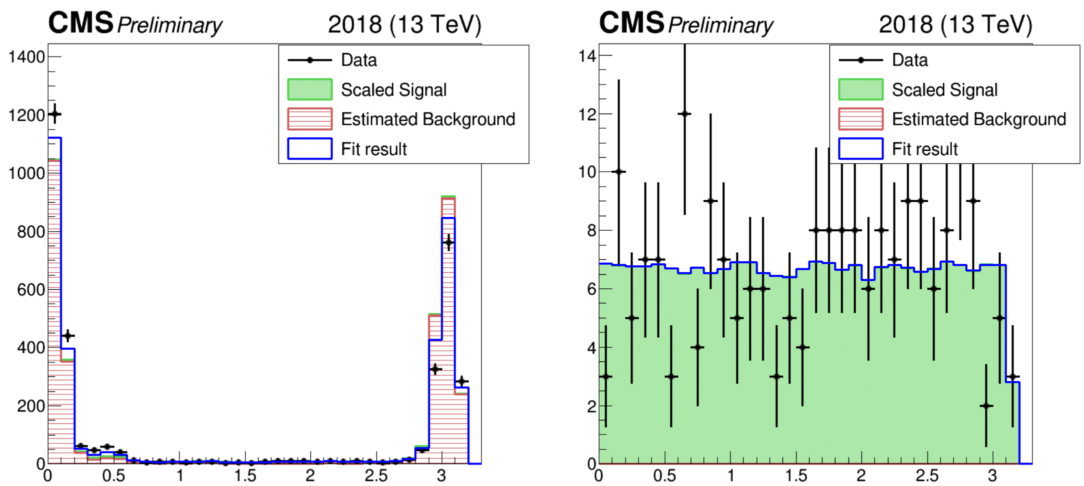


Figure 6.13: Distribution of the BDT score of simulated real photons (red) and photons from beam halo (blue). The BDT score of beam halo photons in real data in a control sample is also shown.

Figure 6.13 shows the distributions of the BDT scores in simulated photons from the signal process and simulated photons in beam halo. In addition, the BDT score distribution in a control sample enriched with beam halo is also shown. The photons from beam halo peak sharply at 0 while real photons peak at 1. To filter out the beam halo events in our signal region, we apply a threshold of 0.99 on the BDT score. This rejects the majority of the beam halo events while reducing the efficiency of the signal events by about 4%.

To estimate the size of the beam halo background that remains after applying the BDT-based filter, we analyze the distribution of the azimuthal angle of the photons observed in the signal region. The LHC's magnets are designed to steer protons along a circular trajectory in the horizontal plane. When charged particles deviate from the beam as part of the beam halo, the LHC magnets direct them towards $\phi = 0$ and $\phi = \pi$ in the CMS coordinate system. Photons originating from the beam halo exhibit distinct peaks at these angles, while genuine photons originating from the primary collision point have a uniform distribution of the azimuthal angle. From a control sample enriched with beam halo photons, we derive the probability density function (PDF), denoted as $p_{\mathrm{BH}}(\phi)$, which characterizes the beam halo photons. The PDF representing real photons is directly obtained from a simulation sample of the signal process. We fit the observed angular distribution using a combination of these PDFs, incorporating fractional contributions f_{BH} and f_{real} respectively.:

$$f_{\mathrm{BH}} \cdot p_{\mathrm{BH}}(\phi) + f_{\mathrm{real}} \cdot p_{\mathrm{real}}(\phi)$$



(a) Without tagger

(b) With tagger

Figure 6.14: Beam halo background in the endcap signal region of data collected in 2017

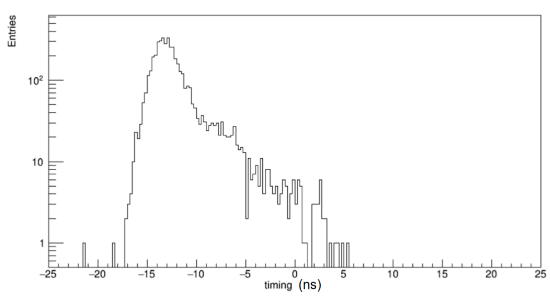
An example of this estimation is shown in Figure 6.14, corresponding to the data collected in 2018. It can be seen that the size of the background overwhelms the signal when the tagger is not applied. After these events are filtered with the tagger, the background disappears. Table 6.5 shows the results of the background estimation for the three years of data-taking.

	Beam halo background events		
	2016	2017	2018
Without tagger	5862.12±106	2079.41±62.4	3256.19±79.9
With BDT-based tagger	9.6±6.2	0±1.9	0±1.6

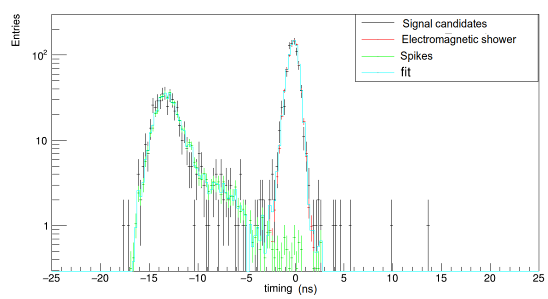
Table 6.5: Size of the beam halo background with and without the use of the BDT-based tagger

6.9 Estimation of spike background in the ECAL barrel

The photodetectors (avalanche photodiodes, or APDs) located at the back of the crystals in the ECAL barrel can generate large anomalous signals when hadrons produced in proton collisions interact directly with them without creating an electromagnetic shower in the crystals. These hadrons lose energy in the adhesive layer between the APD and the crystal surface, causing the ejection of protons directly into the active region of the photodetector. As a consequence, electrons are released into the APD, leading to the production of a spurious signal, known as a *spike*, that can be mistakenly reconstructed as a high-energy photon [64]. As spikes tend to occur in a single crystal, the majority of them are filtered out at the trigger level by identifying clusters in the ECAL barrel with extremely narrow widths. The shower energy distribution in these clusters tends to sharply peak around a single crystal. In addition, the APD does not receive a signal of light from crystal scintillation in spike events, which results in the pulse shape to be constructed, in effect, from only one input signal as opposed to multiple signal samples generated by a real photon shower [65]. These pulses have a shorter rise time and faster decay compared to the response to scintillation light. The shorter rise time of the pulse is reconstructed as an earlier signal. This tends to shift the reconstructed time of a spike signal by 10 to 15 ns before the time of bunch crossing, whereas the timing distribution of real photons peaks at 0 ns. Therefore, we require the timing of the photon to lie within a window -3 ns $< t <$ 3 ns to further reduce the spike background.



(a) Spike timing distribution



(b) Timing distribution of signal events

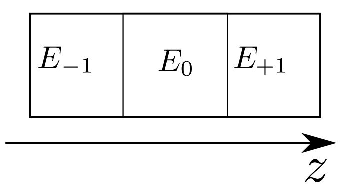
Figure 6.15: Fit of the timing distribution of signal photons in the ECAL barrel

To estimate the remaining spike background, we analyze the timing distribution of the photons selected in the signal region. Figure 6.15 shows the timing distributions of signal photons, spikes and electromagnetic showers. The requirement on the timing of the photons (i.e. -3 ns$< t <$ 3 ns) has been dropped from the selection criteria applied to these events. The contribution of spikes peaks near -15 ns, whereas the electromagnetic shower contribution peaks at 0 ns as expected. We extract the fractional contributions,

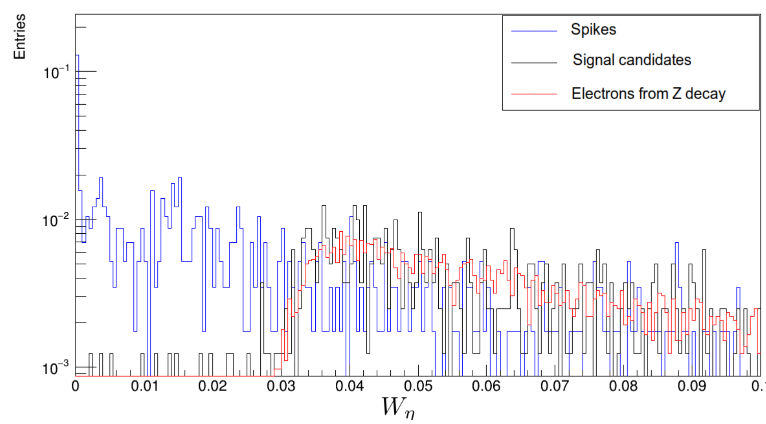
f_{EM} and f_{spikes}, of spikes and electromagnetic showers, respectively, from this distribution by performing a fit to the sum of a PDF describing the timing electromagnetic deposits, $p_{\text{EM}}(t)$, and a PDF describing spikes, $p_{\text{spikes}}(t)$:

$$f_{\text{EM}} \cdot p_{\text{EM}}(t) + f_{\text{spikes}} \cdot p_{\text{spikes}}(t)$$

The PDF $p_{\text{EM}}(t)$ is obtained from a control sample enriched with $Z \to ee$ events where the leading electron passes the photon selection criteria (except for the requirements on the tracker pixel layers).



(a) Crystals defining W_η



(b) Distribution of W_η for spikes, electromagnetic deposits ($Z \to ee$), and signal candidate photons

Figure 6.16: Energy sharing between crystals in the ECAL described by eta wing

In order to select a spike-enriched sample for the construction of $p_{\text{spikes}}(t)$, we use a shower-shape variable called *eta wing*, symbolized W_η. It is defined as the ratio of the energy of the highest-energy crystal (referred to as the *seed* crystal) to the greater energy of the two crystals adjacent to the seed crystal in the z-direction:

$$W_\eta = \frac{\max\{E_{-1}, E_{+1}\}}{E_0}$$

where E_0 corresponds to the energy of the seed crystal, and E_{-1} and E_{+1} are the two adjacent crystals (see Figure 6.16a). This variable describes the sharing of energy between crystals in a cluster. A comparison of the distribution of W_η in spikes, electromagnetic showers (from $Z \to ee$), and from the signal photons is shown in Figure 6.16b. Spikes share very little energy between crystals and therefore the distribution of their W_η peaks at 0. Electromagnetic deposits, on the other hand, typically have $W_\eta > 0.03$. By selecting photons with $W_\eta < 0.01$ (which also pass the signal selection criteria defined in Section 6.4), we obtain a pure sample of spikes to construct the timing PDF, $p_{\text{spikes}}(t)$. After fitting the two PDFs to the observed timing distribution of the signal photons, the spike background is obtained from the area of $p_{\text{spikes}}(t)$ within -3 ns$< t <$ 3 ns and the total spike fraction. The results are shown in Table 6.6.

	2016	2017	2018
Spike events within -3 ns $< t <$ 3 ns	9.4 ± 1.6	10.3 ± 1.5	20.8 ± 2.2

Table 6.6: Estimates of spike background from timing fit for the three years of data-taking

6.10 Statistical Analysis

In order to statistically infer the cross-section and the presence of aNTGCS from the observed data, the first step is to build a likelihood function that subtracts the background contributions from the events selected in the signal region, accounts for the systematic uncertainties, and incorporates a model for the desired interpretation of the signal events. Generally, the likelihood function, $\mathcal{L}(d|\theta)$ represents the probability of obtaining a set of observed data $\{d\}$ under the assumption of a model with unknown parameters θ, where d and θ may be multidimensional. We have a set of observations of event counts, $\{d_i^{\text{SR}}, d_i^{\text{CR}_e}, d_i^{\text{CR}_\mu}\}$, for the photon p_T bins, $\{i\}$, in the signal region, SR, and the two $W+\gamma$ control regions, CR$_e$ and CR$_\mu$, respectively. An event in the SR can arise either from the signal process, the $W+\gamma$ process, or one of the other background processes while the events in CR$_e$ or CR$_\mu$ arise from non-signal processes. We define the expected observations in the three regions in the i'th bin of photon p_T as sums of these processes:

$$\begin{aligned}
\text{SR} &: s_i + n_i \cdot b_i^{W\gamma,\text{SR}} + b_i^{\text{SR}} \\
\text{CR}_e &: \quad n_i \cdot b_i^{W\gamma,\text{CR}_e} + b_i^{\text{CR}_e} \\
\text{CR}_\mu &: \quad n_i \cdot b_i^{W\gamma,\text{CR}_\mu} + b_i^{\text{CR}_\mu}
\end{aligned} \tag{6.1}$$

where s_i is an unknown free parameter representing the signal process, $\{b_i^{W\gamma,x}\}$ are the estimates for $W+\gamma$ from simulation, and $\{b_i^x\}$ are the estimates of the remaining background processes. The free parameter n_i extrapolates adjustments to the $W+\gamma$ background with the observations from the control regions and is common to all three regions. The background estimates are subject to systematic uncertainties that are modeled with a set of *nuisance parameters*, $\{\theta\}$, following Gaussian distributions. Note that the signal terms, s_i, will be re-written a function of the cross-sections and aNTGC coupling parameters in the later sections. To be general, we assume that s_i is a function of a set of *parameters of interest*, $\{\mu\}$, as well as nuisance parameters, $\{\theta\}$. With these definitions, we build a *likelihood function* to quantify the probability of obtaining the observed event counts in the three regions for a given set of parameter values, $\{\mu, n, \theta\}$. It is written as a product of four components, three of which correspond to the probabilities of obtaining the event counts in the three regions, while the fourth term represents the assumed Gaussian distributions of the nuisance parameters representing the uncertainties in the model:

$$\begin{aligned}\mathcal{L}(d^{\mathrm{SR}}, d^{\mathrm{CR}_e}, d^{\mathrm{CR}_\mu}|\mu, n, \theta) =& \mathcal{L}^{\mathrm{SR}}(d^{\mathrm{SR}}|\mu, n, \theta) \times \mathcal{L}^{\mathrm{CR}_e}(d^{\mathrm{CR}_e}|n, \theta) \\ &\times \mathcal{L}^{\mathrm{CR}_\mu}(d^{\mathrm{CR}_\mu}|n, \theta) \times \mathcal{L}^{\theta}(\theta)\end{aligned}$$

The probability of obtaining an event count of x in a given bin is obtained from the Poissonian distribution with mean λ:

$$\mathcal{P}(d|\lambda) = \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$

The mean λ corresponds to sum of signal and background processes on the right hand side in Equation 6.1. The likelihood terms, $\mathcal{L}^{\mathrm{SR}}(d^{\mathrm{SR}}|\mu, n, \theta)$, $\mathcal{L}^{\mathrm{CR}_e}(d^{\mathrm{CR}_e}|n, \theta)$ and $\mathcal{L}^{\mathrm{CR}_\mu}(d^{\mathrm{CR}_\mu}|n, \theta)$, are written as the product of these Poissonian probabilities over the bins of p_T. Thus, the full form of the total likelihood function is:

$$\begin{aligned}\mathcal{L}(d^{\text{SR}}, d^{\text{CR}_e}, d^{\text{CR}_\mu}|\mu, n, \theta) = \Bigg[& \prod_{i \in \text{bins}} \mathcal{P}\Big(d_i^{\text{SR}}|s_i(\mu,\theta) + n_i \cdot b_i^{W\gamma,\text{SR}}(\theta) + b_i^{\text{SR}}(\theta)\Big) \\ & \times \mathcal{P}\Big(d_i^{\text{CR}_e}|n_i \cdot b_i^{W\gamma,\text{CR}_e}(\theta) + b_i^{\text{CR}_e}(\theta)\Big) \\ & \times \mathcal{P}\Big(d_i^{\text{CR}_\mu}|n_i \cdot b_i^{W\gamma,\text{CR}_\mu}(\theta) + b_i^{\text{CR}_\mu}(\theta)\Big)\Bigg] \\ & \times \prod_{j \in \text{nuisances}} \mathcal{N}(\theta_j|\bar{\theta}_j, \sigma_{\theta_j})\end{aligned} \tag{6.2}$$

where $\mathcal{L}^\theta$ has been re-written as a product of Gaussian distributions, $N(x|\bar{x}, \sigma_x)$, over the nuisance parameters. The likelihood function allows us to determine the most likely values of the parameters based on the assumption that, for these parameter values, the probability of obtaining the observed event counts is at its maximum. Another important use of the likelihood functions is in the determination of uncertainties of the parameters of interest, $\{\mu\}$. Using all of the parameters, (μ, n, θ), to find the uncertainties in the parameters of interest μ is equivalent to using the *profile likelihood*, which depends only on μ. It is defined as:

$$\mathcal{L}_P(\mu) = \mathcal{L}(\mu, \hat{\hat{n}}(\mu), \hat{\hat{\theta}}(\mu)) \tag{6.3}$$

where the double-hat notation indicates the *profiled values* of the parameters n and θ that maximize $\mathcal{L}$ for the specified values of μ. We construct the following test statistic with $\mathcal{L}_P$ for obtaining confidence intervals of μ:

$$\lambda(\mu) = \frac{\mathcal{L}_P(\mu)}{\mathcal{L}_P(\hat{\mu})} \tag{6.4}$$

where $\hat{\mu}$ corresponds to the values that maximizes the likelihood. The distribution of $-2\ln\lambda(\mu)$ approaches a χ^2 distribution independently of the values of the nuisance parameters, θ, in the limit where the data sample is large. Here the number of degrees of freedom, is equal to the number of components of μ. By performing a *scan* of $-2\ln\lambda(\mu)$ over μ with the observed data, we can obtain the confidence intervals of μ that correspond to a chosen confidence level for the χ^2 distribution. Specifically, the maximum value $\chi^2_{\max}$ of χ^2 for a given confidence level can be obtained from the inverse cumulative distribution function of the χ^2-distribution. The set of values of μ that correspond to $-2\ln\lambda(\mu) \leq \chi^2_{\max}$ represent the confidence interval.

6.10.1 Choice of bin boundaries

The boundaries for the photon transverse momentum bins are chosen using the *Bayesian-Blocks* algorithm [53]. This technique was originally designed to show the most statistically significant shape variations in time series data. To identify the optimum binning, we start with uniformly binned histograms with width 5 GeV for the total predicted background and SM signal in the barrel and endcap regions. The algorithm identifies, recursively, the binning choice that maximizes the total *Cash statistic* [66], $\sum_k N_k \cdot \log(N_k/T_k)$, where N_k is the number of events in bin k of width T_k. The chosen bin boundaries, in GeV, are:

- Barrel region: 225, 235, 245, 265, 285, 315, 355, 425, 515, 615, 765, 1500 (11 bins)
- Endcap region: 225, 230, 235, 270, 310, 360, 440, 1500 (7 bins)

6.10.2 Signal modeling for cross-section measurement

The first set of measurements involves extracting the cross-section and differential cross-section of the production of $Z(\rightarrow \nu\nu)+\gamma$. The *cross-section* corresponds to the total rate of this process while the *differential cross-section* as a function of photon p_T provides a detailed description of the variation of the cross-section. In a given phase space, if the total number of signal events produced is N, the number of signal events in a specific bin i of photon p_T with width $(\Delta p_T)_i$ is N_i, and the integrated luminosity of the collisions is $\mathcal{L}$, the cross-section and differential cross-section are defined as:

$$\begin{aligned} \sigma &= \frac{N_s}{\mathcal{L}} \\ \left(\frac{\mathrm{d}\sigma}{\mathrm{d}p_T}\right)_i &= \frac{N_i}{\mathcal{L}\cdot(\Delta p_T)_i} \end{aligned} \tag{6.5}$$

respectively. Many efforts have been made to calculate the cross-section and differential cross-sections of $pp \rightarrow Z(\rightarrow \nu\nu) + \gamma$ production with different degrees of precision and calculation schemes, at different orders of perturbation theory, with some incorporating QCD and electroweak corrections [67, 68]. Performing these measurements, therefore, allows direct comparison between theoretical predictions of the SM and observation from experimental data.

While the number of events remaining after subtracting the background through a maximum likelihood fit provides the best estimate of the observed signal, it does not represent the full size, N or N_i, of its production in the proton collisions. Several effects reduce the number of signal events collected:

- **Geometric acceptance** Photons within only the pseudorapidity range of $|\eta| < 1.4442$ or $1.566 < |\eta| < 2.5$ are selected. Signal events with photons outside of this range are not considered.
- **Efficiency of selection** The event selection criteria, as outlined in Section 6.4, impose various thresholds on the photon, missing transverse momentum, and jets to suppress background contributions. These thresholds also decrease the signal size. Additionally, the photon identification efficiency is limited to around 92%.
- **Detector resolution** The resolution of jet energy directly affects the direction and magnitude of the missing transverse momentum. Consequently, the minimum threshold of 200 GeV on $\not{E}_T$, and the criteria applied to $\Delta\phi(\gamma, \vec{\not{E}}_T)$, $\Delta\phi(\text{jet}, \vec{\not{E}}_T)$, and $p_T^\gamma/\not{E}_T$ introduce significant systematic uncertainties in the number of collected signal events. The resolution of the photon energy measurement also has a minor impact due to the 225 GeV transverse momentum threshold.

Object	Criteria at truth-level
Photon	$\lvert\eta\rvert < 1.4442$ or $1.566 < \lvert\eta\rvert < 2.5$ $p_T > 225$ GeV
Missing transverse momentum	$\not{E}_T > 200$ GeV $\frac{p_T}{\not{E}_T} < 1.4$ $\Delta\phi(\not{E}_T, \gamma) > 2$

Table 6.7: Truth-level criteria defining the fiducial region for the cross-section measurement

We follow a technique called *unfolding* to account for the impact of detector acceptance, efficiency, and resolution using a simulated sample of the signal process. The first step in this procedure is to define a *fiducial region* in the *truth level*, which represents the idealized kinematic properties of the particles before any interaction with the detector. The cross-section is measured specifically for this phase space. The criteria defining the fiducial region closely match those applied to the observed data, ensuring alignment of the measured cross-section with the detector-level phase space and minimal extrapolation to truth-level phase spaces not contributing to the observed data. The region's definition, outlined in Table 6.7, relies on thresholds for photon transverse momentum (p_T), pseudorapidity (η), and event missing transverse momentum ($\vec{\not{E}}_T$) at the truth level of the simulation, before detector interaction. We show in Table 6.8 the number of events produced in the fiducial region in simulations, and the events actually selected in the barrel and endcap regions. The table also shows the detected events that are produced outside of the fiducial region.

		Events from the fiducial region		Events from outside the fiducial region
	Year	Detector-level	Truth-level	Detector-level
Barrel region $\lvert\eta\rvert < 1.4442$	2016	405.8	626.7	28.3
	2017	420.4	712.6	31.7
	2018	538.2	1031.5	40.1
Endcap region $1.566 < \lvert\eta\rvert < 2.5$	2016	98.3	211.8	10.7
	2017	82.0	241.4	9.0
	2018	100.6	347.6	10.1

Table 6.8: Simulated fiducial and non-fiducial events in the barrel and endcap regions

Next, we contruct a *response matrix*, R_{ij}, which models the detector's response to the events produced in the fiducial region. It quantifies the probability with which an event produced with a truth-level photon p_T in bin j will be observed with detector-level photon p_T in bin j, and accounts for bin migration as well as efficiency and acceptance. From the simulation sample of the signal, we calculate the number of events, N_{ij}^{detector}, in the detector-level bin i that are actually produced in the truth-level bin j. The sample also provides the total number of truth-level fiducial events, N_j^{truth}, produced in the truth-level bin j. The response matrix is then defined as:

$$R_{ij} = \frac{N_{ij}^{\text{detector}}}{N_j^{\text{truth}}} \tag{6.6}$$

This response is used to model the signal, s_i, within the likelihood function in Equation 6.2. The number of observed signal events in detector bin i is expressed as a summation of contributions from the truth-level bins j, each associated with unknown cross-sections σ_j. Additionally, there is a minor contribution, $s_i^{\text{non-fiducial}}$, originating from signal proces occurring outside of the fiducial region, specified in Table 6.8. These non-fiducial events are treated as background in this measurement. The full signal model is:

$$s_i = s_i^{\text{non-fiducial}} + \sum_j \mathcal{L} \cdot \sigma_j \cdot R_{ij} \tag{6.7}$$

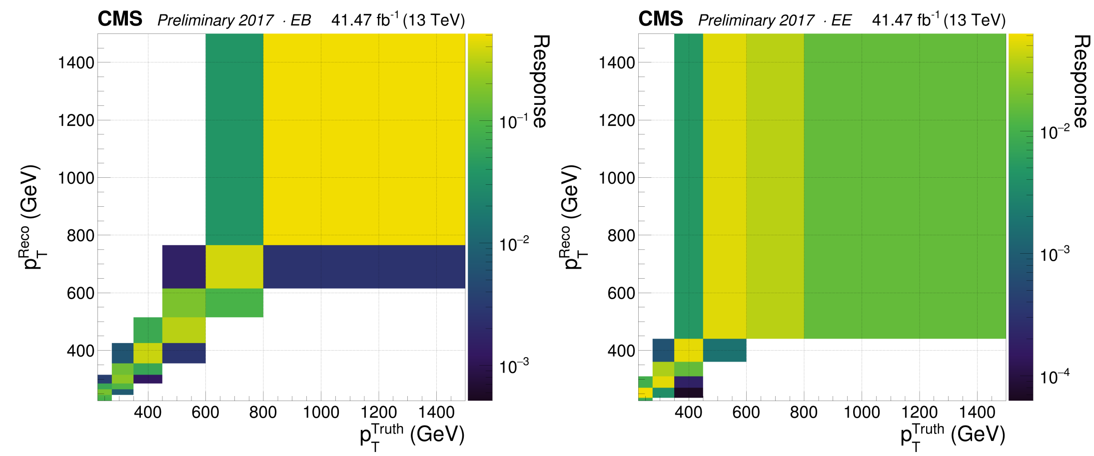
where $\mathcal{L}$ is the integrated luminosity. Notice that $\mathcal{L} \cdot \sigma_j$ represents the true number of events produced in truth-level bin j, which is reduced by the detector response, R_{ij}, in the sum. The terms $\{\sigma_j\}$ are treated as free parameters and obtained from a maximum likelihood fit to derive the differential cross-sections.

In addition to the full fiducial region, we measure the cross-section and differential-cross section separately for the barrel and endcaps, where only the boundaries of η are modified and the other criteria given in Table 6.7 are unchanged. The barrel region corresponds to $|\eta| < 1.4442$ and the endcaps region is given by $1.566 < |\eta| < 2.5$.

For the measurement of the differential cross-sections, the boundaries (in GeV) for the bin of the phton's truth-level p_T are chosen as follows:

- Barrel : 225, 275, 350, 450, 600, 800, 1500 (6 bins)
- Endcaps: 225, 275, 350, 450, 1500 (4 bins)
- Full fiducial region (barrel+endcaps): 225, 275, 350, 450, 600, 800, 1500 (6 bins)

However, when measuring the cross-section, we do not divide the events into truth-level bins. The response matrix contains only one column, and the number of rows corresponds to the detector-level p_T binning. Figure 6.17 shows the response matrix for the full fiducial region for the 2017 data-taking period.



(a) Barrel region (b) Endcap region

Figure 6.17: Response matrix for the measurement of the differential cross-section of the fiducial region for the 2017 data-taking period. The x-axis represents the truth-level p_T and the y-axis is the detector-level p_T

6.10.3 aNTGC signal modeling

The presence of aNTGCs is expected to manifest as an excess in the observed event count from $Z(\rightarrow \nu\nu) + \gamma$ with respect to the predictions of the SM. This excess is significant specifically in the tail of the photon p_T distribution. We search for such an excess by performing a binned maximum likelihood fit with the likelihood function given in Equation 6.2. We perform separate fits for the two different vertices: $ZZ\gamma$ with coupling parameters (h_3^Z, h_4^Z), and $Z\gamma\gamma$ with coupling parameters (h_3^γ, h_4^γ). In these fits, the background is subtracted from the observed events and the signal is interpreted independently for the two vertices.

To include the signal term, s_i, in the likelihood function considering the dependence on the coupling parameters, we construct a parametric function for each bin of photon p_T. This function provides an estimation of the expected event count within the bin for any values of the coupling parameters. This allows us to treat the coupling parameters as free unknown parameters in the maximum likelihood fit. The prediction for bin i

takes the form of a quadratic function of the couplings (h_3^V, h_4^V), where V can be either Z or γ, and is expressed as:

$$s_i(h_3^V, h_4^V) = \underbrace{s_i^{\mathrm{SM}}}_{\mathrm{SM}} + \underbrace{l_i^{3V} \cdot h_3^V + l_i^{4V} \cdot h_4^V}_{\mathrm{SM+aNTGC\ interference}} + \underbrace{q_i^{34V} \cdot h_3^V \cdot h_4^V + q_i^{3V} \cdot (h_3^V)^2 + q_i^{4V} \cdot (h_4^V)^2}_{\mathrm{Pure\ aNTGC}} \tag{6.8}$$

The terms that are linear in the coupling parameters account for the interference between the amplitudes of the SM and aNTGCs, whereas the quadratic terms arise solely from contributions of aNTGCs. The constant term, s_i^{SM}, represents the number of events predicted by the SM in the absence of aNTGCs. To determine the values of the coefficients in the parametrization (l_i^{3V}, l_i^{4V}, q_i^{34V}, q_i^{3V}, and q_i^{4V}), we generated simulation samples for the $ZZ\gamma$ and $Z\gamma\gamma$ vertices. For each vertex, we considered 440 non-zero couplings within the ranges of $-4 \times 10^{-4} < h_3^V < 4 \times 10^{-4}$ and $-5 \times 10^{-7} < h_4^V < 5 \times 10^{-7}$. These boundaries were chosen to align with the current limits on the couplings as reported in [2]. The simulation process includes a comprehensive modeling of $Z(\nu\nu) + \gamma$ event production in p-p collisions, as well as the subsequent particle propagation within the detector, the detector's response to the outgoing particles, and the complete reconstruction process. By applying the selection criteria for the signal region (refer to Section 6.4) to the generated events, we obtained direct predictions of the expected event counts for each simulated coupling point within each bin of photon p_T. To obtain the values of the coefficients for each p_T bin, we performed a fit of the parametric function in Equation 6.8 using the expected event counts at the 441 points. An example of the fit is shown in Figure 6.18, corresponding to the barrel region in the 2017 data-taking period, where the event counts for the simulated coupling points are represented with white markers and the parametrization is drawn as a yellow surface.

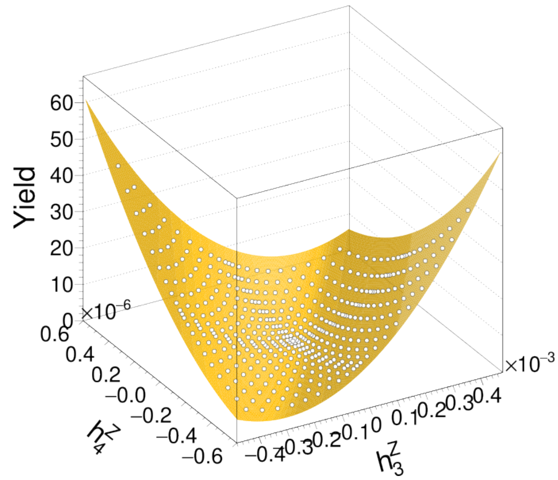


Figure 6.18: Predicted number of events as a function of the coupling parameters (h_3^Z, h_4^Z) in the bin 765 GeV $< p_T <$ 1500 GeV for the barrel region in the 2017 data-taking period

6.11 Systematic Uncertainties

The modeling of the signal is affected by a number of systematic uncertainties. In the measurements of the cross-section and the differential cross-section, the construction of the response matrix relies on simulations. Uncertainties in the simulation, particularly those that impact the efficiency of selecting signal events, contribute to the overall uncertainty of the response matrix. In the case of the measurement of aNTGCs, the search is based on detecting an excess in the number of $Z(\nu\nu)+\gamma$ events beyond the predictions of the SM. Consequently, it is necessary to consider uncertainties in both the SM predictions and the additional contribution from the aNTGC signal in the simulations. These uncertainties can be classified into two categories: theoretical uncertainties associated with simulating the p-p collisions, and experimental uncertainties related to the simulation of the detector response. To account for these uncertainties, we assess the impact of each source individually. The signal model is evaluated by varying the corresponding source of systematic uncertainty by one standard deviation in both positive and negative directions. This involves recomputing the response matrix and re-evaluating the coefficients of the aNTGC parametrization based on the modified predictions. The resulting variations in the signal model are then incorporated into the likelihood function (Equation 6.2), where each uncertainty is represented by a Gaussian-distributed nuisance parameter.

The experimental uncertainties in the signal models arise from differences in the detector performance between simulation and the actual data. In particular, differences in the precision of the energies of photons, jets, and missing transverse momentum, as well as the efficiency of photon identification can introduce biases in the number of simulated events that satisfy the selection criteria defined in Section 6.4. The uncertainties we consider are:

- **Photon identification efficiency** The identification criteria for photons rely on isolation variables of the subdetectors, and ECAL shower shape variables, which are used as inputs to a BDT. Mismodeling of these variables leads to differences in identification efficiency between real data and simulation. A correction weight is applied to each simulated event to align with the efficiency observed in real data (refer to Section 5.4). The uncertainty in this correction is propagated to the signal model by varying the weights by one standard deviation in positive and negative directions, and taking the variation in the signal count as the uncertainty. The uncertainty in the signal efficiency is estimated to be between 3.5% and 4.7% across the three years of data-taking, making it the largest systematic uncertainty in this analysis.

- **Pixel rejection efficiency** In order to avoid selecting electrons, the photon in an event is required to have fewer than two hits in the pixel detector that point towards the ECAL cluster of the photon. Besides rejecting background events, this criterion also reduces the efficiency of selecting true photons when a low-momentum charged particle curves tightly and leaves hits in the pixel layer consistent with the direction of the photon, or when the photon converts to a e^+e^- pair within the pixel detector. However, imperfect modeling of the tracker response leads to a mismatch in the photon selection efficiency between real data and simulation. Corrections have been derived to compensate for this effect by comparing a sample of $pp \to Z \to \mu\mu\gamma$ (where the γ is radiated by a μ) in real data with simulation [44]. These corrections are applied as weights to the simulated events for our signal models. The uncertainty in the correction is propagated to the simulation by varying the weights by one standard deviation and re-evaluating the event count. The corresponding uncertainty in the signal efficiency is estimated to be under 0.2% in the barrel region and under 0.9% in the endcaps.

- **Jet energy scale and resolution** A number of techniques are used to correct the energies of jets, which involve balancing the total transverse momentum in events where jets recoil against a reference object whose transverse momentum is precisely measured [48, 69]. The uncertainties associated with the residual mismatch in jet energy calibration between real data and simulation have an impact on our signal models due to the thresholds applied to the missing transverse momentum (the negative vectorial sum of the $\vec{p}_T$ of reconstructed particles, including jets) and jets in the event selection criteria. To account for this uncertainty, we vary the energy corrections of the jets in the simulated events by one standard deviation, recalculate the missing transverse momentum for each simulated event, and determine the number of events that satisfy the selection criteria. The resulting uncertainty in the signal efficiency ranges between 0.5% and 2.8%. After these corrections, the resolution of jet energy in simulation is found to be better than in real data. To account for the discrepancy, another correction is applied by randomly smearing the jet energies in the simulation using a Gaussian distribution. The uncertainty associated with this correction is taken into account by varying the smearing resolution for each jet in the simulated events by one standard deviation, recalculating the missing transverse momentum, and applying the event selection criteria to the updated events. The uncertainty in the signal efficiency due to the jet energy resolution bias is estimated to be between 0.2% and 2.3%.

- **Photon energy scale and resolution** The energies of reconstructed photon clusters undergo corrections to address instrumental effects, such as energy loss caused by shower leakage in the gaps between crystals and radiation damage affecting the scintillation response [44]. An additional set of corrections is derived using a reference sample of $Z \rightarrow ee$ events to minimize the residual discrepancy between real data and simulation. To account for the associated uncertainty, the corrections for each photon in the simulation are varied by one standard deviation, and the number of events satisfying the selection criteria is re-evaluated. The corresponding uncertainty in the signal efficiency between 0.2% and 1.1%. Furthermore, in order to align the resolution in simulation with that observed in the data, corrections are applied by randomly smearing the photon energies with a Gaussian distribution. The uncertainty due to the resolution correction is evaluated by re-smearing the photon energies in the simulation with the resolution correction varied by one standard deviation and re-counting the simulated events passing the selection criteria. The uncertainty in the signal efficiency due to the resolution bias is estimated to be under 0.2%.

- **Luminosity** The integrated luminosity determines the number of signal events produced in the *p*-*p* collisions. It is derived by integrating the instantaneous luminosity over time. The instantaneous luminosity is a function of the beam parameters, one of which is the beam overlap width, which cannot be measured with sufficient precision during physics runs. Instead, the instantaneous luminosity is measured indirectly using seven subsystems (luminometers) of the detector, which read out the rates of the specific quantities observed in the detector, such as hits, tracks and clusters. These rates are proportional to the instantaneous luminosity, where the constant of proportionality is the visible cross-section. A calibration procedure, known as *van der Meer scan*, is performed to determine the visible cross-section, where the two beams are separated and then moved across each other, and the resulting luminometer rates and beam overlap width are used to determine the visible cross-section. During physics runs, the instantaneous luminosity is calculated using the measured visible cross-section and luminometer rates. The uncertainties in the measured integrated luminosity are 1.2%, 2.3%, and 2.5% for the data-taking periods of 2016 [17], 2017 [18], and 2018 [19] respectively.

- **Pileup modeling** Pileup refers to the simultaneous occurrence of multiple *p*-*p* collisions within a single bunch crossing. The number of additional collisions was, on average, 23 in 2016, and subsequently increased to 32 in 2017 and 2018. The extra interactions lead to additional energy deposits and tracks in the detector, and dedicated algorithms are employed to mitigate the effects of pileup [45]. Mismatch in the number of pileup interactions in simulation and real data biases the energy measurement and identification efficiency of particles in our signal models. To address pileup mismodeling, the simulated events are adjusted by applying weights to match the expected distribution of pileup interactions in real data. The expected pileup distribution is obtained using instantaneous luminosity measurements and an estimated total inelastic cross-section of 69.2 ($\pm$4.6%) mb. To assign an uncertainty for this correction, the expected pileup distribution is re-derived with the inelastic cross-section varied by one standard-deviation, and new weights are applied to the simulated events. This leads to an uncertainty in the signal efficiency between 0.4%-1.5% in the barrel region and between 0.8%-2.2% in the endcaps.

- **Unclustered energy scale** Small deposits are reconstructed in the detector that are not associated with the clusters of any particles or jets. They can arise from instrumental noise, inefficiency of the clustering algorithms, or from low energy particles from pileup interactions. A correction is applied in the reconstruction of missing transverse momentum to mitigate the effect of these unclustered deposits [70]. The uncertainty in this correction is propagated to the signal models by varying the missing transverse momentum by one standard deviation and recounting the number of simulated events passing the selection criteria. The effect of this uncertainty on the signal efficiency is estimated to be under 0.3%.

- **Trigger prefiring in 2016 and 2017** Over the course of data-taking, the timing calibration of the pulses in the endcap crystals of the ECAL was affected by radiation-induced transparency loss. This resulted in an increasing offset in the timing calibration, especially for the endcap crystals at high pseudorapidity. Although the timing calibration was compensated for in the offline reconstruction, it was unfortunately not corrected in the online reconstruction algorithms. As a result, in the data collected in 2016 and 2017, a significant fraction of ECAL deposits that fired the trigger were incorrectly associated with an earlier bunch crossing. The CMS trigger rules forbid the acceptance of two consecutive bunch-crossings, which meant that the detector readout of the correct bunch-crossing was not stored, leading to a reduced trigger efficiency. This effect, known as *trigger prefiring*, is not replicated in the simulation samples. To address this, a correction has been derived [38] to estimate the *non-prefiring probability* of electromagnetic deposits by comparing a simulation sample of $Z \to ee$ with real data. We apply these corrections as event weights to the simulated samples and assess the uncertainty in the signal models by varying the correction weights by one standard deviation. The resulting uncertainty in the signal efficiency is estimated to be about 1.3% in 2016 and 1.9% in 2017 for the endcap region. The uncertainty for the barrel region is under 0.2%. In early 2018, the endcap timing delays in the ECAL front-end electronics were corrected and therefore no correction is necessary for this year.

Source of uncertainty	Relative impact on signal efficiency (%)
Photon selection efficiency	3.5 - 4.7
Jet energy scale	0.5 - 2.8
Jet energy resolution	0.2 - 2.3
Pileup modeling	0.4 - 2.2
Photon energy scale	0.2 - 1.1
Photon energy resolution	< 0.2
Unclustered energy scale	< 0.3
Pixel detector efficiency	< 0.9
Prefiring (2016 and 2017 data only)	0.1 - 1.9

Table 6.9: Relative impact of the experimental uncertainties on the signal efficiency. The uncertainties are given in percent.

Theoretical uncertainties in the signal simulation have an impact on both the number of signal events generated before detector interaction and the number of events selected. In the case of cross-section measurements, these uncertainties have a limited effect because they are correlated between the selected and produced events, resulting in a cancellation when calculating the response matrix ratio. However, in the search for aNTGCs, theoretical uncertainties in the SM prediction can diminish the statistical significance of the excess caused by an aNTGC signal. The signal samples are simulated at leading order in both quantum chromodynamics (QCD) and electroweak theory. This means that higher-order effects are not taken into account in the cross-section of the SM prediction. To address this, corrections are applied as weights to the events, incorporating next-to-next-to-leading-order (NNLO) QCD effects and next-to-leading-order (NLO) electroweak effects. The dominant theoretical uncertainties in the SM prediction arise from the uncertainties in these corrections. Furthermore, an additional uncertainty is assigned to the parton density functions (PDFs) used in the simulations. The theoretical uncertainties are:

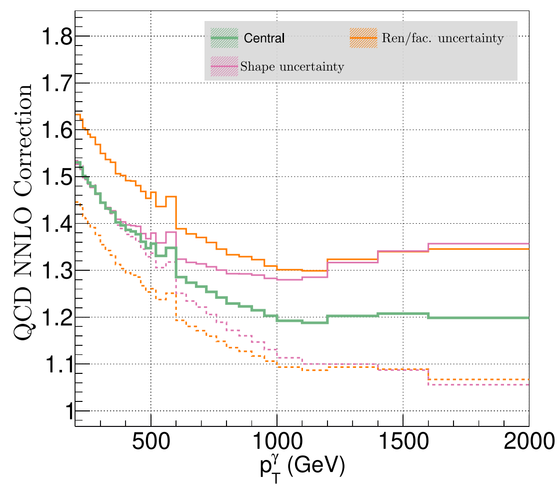


Figure 6.19: Corrections to the SM signal to account for QCD NNLO effects. Uncertainties due to the choice of renormalization and factorization scale, and for the variation in the shape are also shown.

- **Renormalization and factorization scale** The NNLO QCD corrections applied to the SM prediction are derived using the program MATRIX [71, 67]. Two sources of uncertainty are assessed for the cross-section calculated by MATRIX. The first is the choice of renormalization scale, which is a cut-off energy used to remove infinities from divergent integrals in the perturbative calculations. The second is the factorization scale, which separates different energy scales of the scattering process and divides the calculation into two components: the hard scattering cross section and the parton distribution functions (PDFs). The factorization theorem states that the cross section can be factorized into these two terms, with the factorization scale serving as the boundary between them. There is a degree of arbitrariness in the choice of the renormalization and factorization scales since the precise behavior of perturbative calculations at all energy scales is not fully known [72, 73]. We choose the renormalization and factorization scales to be $\sqrt{m_T^Z + p_T^\gamma}$, where m_T^Z is the transverse mass of the outgoing Z boson and p_T^γ is the photon transverse momentum. The uncertainty due to this choice is assessed by varying the two scales independently by factors of 0.5 and 2, and then taking the envelope of the variations [74]. This uncertainty mainly affects the overall normalization of the cross-section. An additonal uncertainty is assigned for possible variation of in the shape of the photon p_T distributions using a prescription given in [75]. These corrections are shown in Figure 6.19. The uncertainty in the predicted number of SM events due to the choice of renormalization and factorization scales is estimated to be between 6% and 7.2% while the shape uncertainty is under 0.1%.

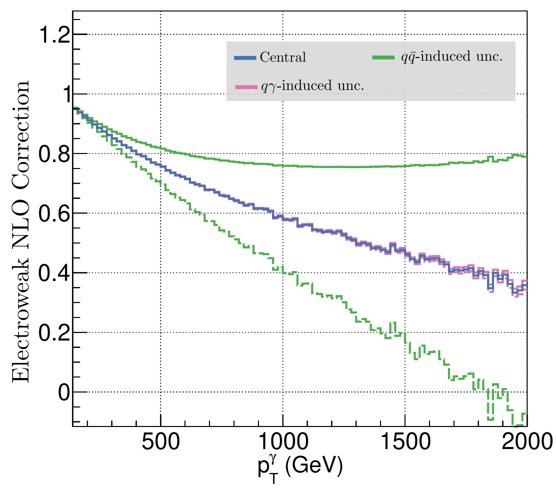


Figure 6.20: Corrections to the SM signal to account for electroweak NLO effects. Uncertainties shown are divided into Feynman diagrams initiated by quark-antiquark and by quark-photon.

- **Electroweak corrections** Higher order electroweak effects significantly modify the cross-section at high energies above the electroweak scale due to contributions that scale logarithmically with the energy [76]. These corrections have been calculated at next-to-leading-order (NLO) in [68] for $\sqrt{s} = 14$ TeV. The authors have kindly re-calculated them with $\sqrt{s} = 13$ TeV for this analysis. The corrections are applied to the simulation as weights to the generated events. The uncertainties in these corrections are divided into two components corresponding to Feynman diagrams initiated by quark-antiquark and by quark-photon. The corrections and the uncertainties are shown in Figure 6.20. The uncertainty in the predicted number of SM events due to these corrections is estimated to be about 2.7%.

- **Parton density functions (PDFs)** PDFs describe the probability distributions of momentum fractions of partons (quarks and gluons) inside the colliding protons [77]. They are determined through a combination of experimental measurements and theoretical calculations. However, due to limited experimental data and the complexity of the proton's internal structure, PDFs have inherent uncertainties associated with them. The signal simulation is performed with the PDF set `NNPDF31_nnlo_hessian_pdfas` calculated by the NNPDF Collaboration [78]. This PDF set is provided in the Hessian format, which characterizes the uncertainties by defining a central set of PDFs and a set of 102 variations around the central values. To assess the uncertainty in the signal model, we evaluate the predicted event counts, $\{N^0, N^1, ..., N^{102}\}$, by applying the weights to the simulated events for each PDF variation. The PDF uncertainty in event count is then given by [79]:

$$\sigma^{\text{PDF}} = \sqrt{\sum_{k=1}^{100} (N^k - N^0)^2}$$

  An additional uncertainty is assigned for the uncertainty in the strong coupling, $\alpha_S = 0.018 \pm 0.002$, using the last two variations of the PDFs:

$$\sigma^{\alpha_S} = \frac{N^{102} - N^{101}}{2}$$

  The uncertainties are then added in quadrature. The uncertainty in the predicted number of signal events is estimated to be 1.7% in the barrel region and 1.5% in the endcaps.

The experimental and theoretical uncertainties in the signal efficiency and the predicted number of signal events are summarized in Tables 6.9 and 6.10 respectively.

	Relative impact on predicted SM signal yield (%)
Renormalization and factorization scale	6 - 7.2
Electroweak corrections	2.7
Parton density functions	1.5 - 1.7

Table 6.10: Relative impact of the theoretical uncertainties on the predicted number of SM signal events. The uncertainties are given in percent.

In addition to the uncertainties in the signal modeling, the estimation of the background processes is also affected by a number of uncertainties which can potentially introduce discrepancies with the actual background in the data. The uncertainties in each source of background are:

- **Electrons misidentified as photons** The uncertainty in the size of this background ranges between 13% and 15% in the barrel region, and between 27% and 45% in the endcaps across the three years. The uncertainty arises from the statistical uncertainty due to the limited sample size of the control region where the fake ratio is applied and from the systematic uncertainty in the fake ratio.

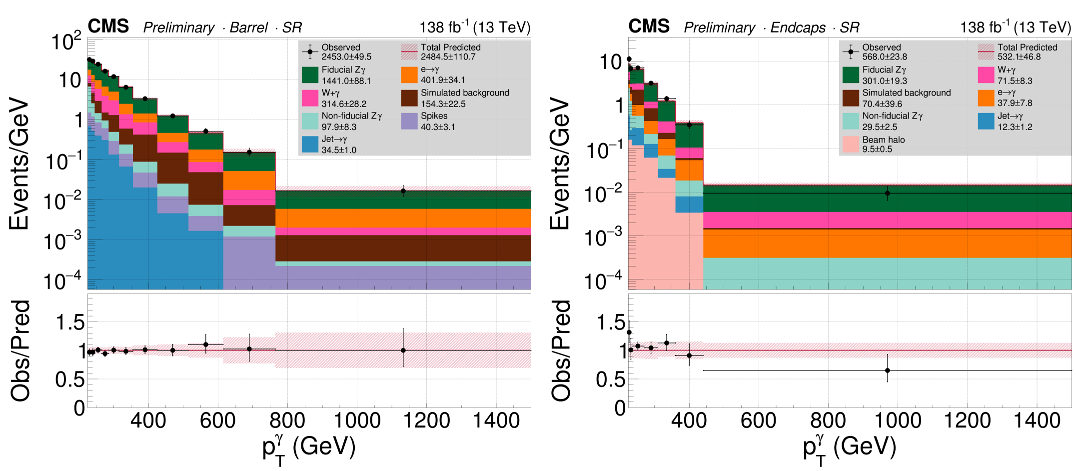
- **$W + \gamma$ background** The uncertainty in this background after the simultaneous fit of the control and signal regions is estimated to be 8.9% in the barrel region and 11.6% in the endcaps for the full data. The simulated estimates of $W + \gamma$ in the signal and control regions are affected by the same sources of experimental uncertainties as those for the signal model, except for the uncertainty in the luminosity since the simulated estimate in the signal region is scaled by the measurements of this process in control regions in real data. An additional systematic uncertainty is assigned for the identification efficiency of electrons and muons for the simulations in the control regions. The uncertainties related to background contributions in the control regions are also taken into account.

- **Jets misidentified as photons** As with the background due to electrons misidentified as photons, systematic uncertainties in the fake ratio and the finite sample size of the control region where the fake ratio is applied both affect the estimation of this background. The uncertainty in the size of this background ranges between 2% and 5.6% in the barrel region, and between 8.2% and 20% in the endcaps.

- **Barrel ECAL spikes** The uncertainty in number of barrel ECAL spikes is estimated to be between 11% and 17% across the three years. These uncertainties arise from the limited sample sizes of the shape templates used in the fit of the timing distribution.
- **Endcap beam halo background** The BDT-based tagger is expected to remove most of the beam halo background. A small background is expected in the 2016 data, with an uncertainty of 64%.
- **Simulation-driven background estimates** The simulation-driven background estimates are affected by the same sources of experimental and theoretical uncertainties in the simulations as those for the signal simulation. A significant systematic uncertainty also arises due to the limited number of events generated in these samples, particularly in the endcap region. The total uncertainty in this background is between 18% and 30% in the barrel region, and between 30%-54% in the endcaps.

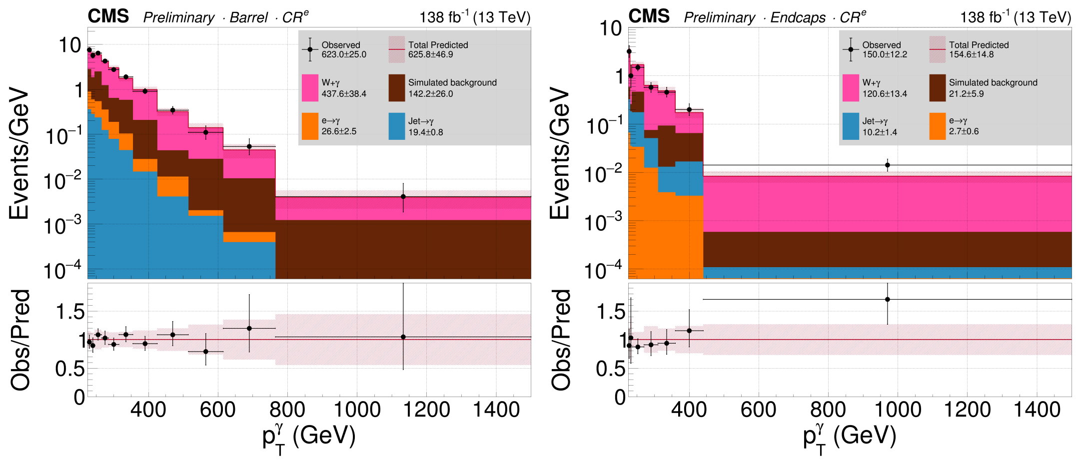
6.12 Summary of estimated contributions

Figures 6.21, 6.22, and 6.23 display the observed and predicted distributions of photon transverse momentum in the signal and $W+\gamma$ control regions separately for the photons in the barrel and in the endcaps. These distributions were obtained by a simultaneous maximum likelihood fit of the three regions for a measurement of the differential cross-section of the process $pp \to Z(\to \nu\nu)\gamma$. The signal component is separated into fiducial and non-fiducial parts as defined in Equation 6.7. During the fit, the differential cross-section parameters, the parameters for constraining the $W+\gamma$ background from the control region measurements, and the nuisance parameters for the systematic uncertainties were adjusted to best match the observed data, resulting in the "predicted" distributions. The top panel of these figures provides a breakdown of the contributions from individual processes, while the lower panel presents the ratio of observed to predicted event counts. The statistical uncertainties of the observed points are indicated by error bars on black points and the systematic uncertainties in the predictions are represented by shaded red bands. Additionally, Tables 6.11, 6.12, 6.13, 6.14, 6.15, and 6.16 provide a detailed breakdown of the observed and predicted event counts for each bin.



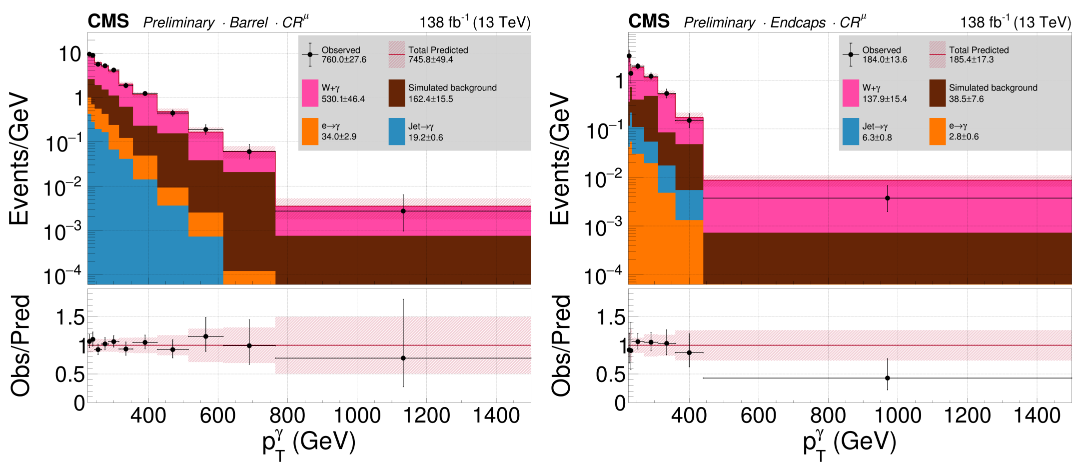
(a) Barrel (b) Endcaps

Figure 6.21: Observed and predicted photon p_T distributions in the signal region in the barrel and endcaps, with breakdown of contributions from individual processes



(a) Barrel (b) Endcaps

Figure 6.22: Observed and predicted photon p_T distributions in the $W(\to e\nu)+\gamma$ control region in the barrel and endcaps, with breakdown of contributions from individual processes



(a) Barrel (b) Endcaps

Figure 6.23: Observed and predicted photon p_T distributions in the $W(\to \mu\nu)+\gamma$ control region in the barrel and endcaps, with breakdown of contributions from individual processes

	Photon p_T range				
	225 - 235 GeV	235 - 245 GeV	245 - 265 GeV	265 - 285 GeV	285 - 315 GeV
Fiducial signal	153.1 ± 12.7	170.0 ± 13.8	290.3 ± 23.2	209.8 ± 13.8	206.0 ± 18.5
Non-fiducial signal	51.4 ± 4.4	13.5 ± 1.1	11.9 ± 1.0	6.8 ± 0.6	5.9 ± 0.6
Misidentified electrons	49.3 ± 4.4	45.0 ± 3.9	75.4 ± 6.5	55.4 ± 4.7	58.4 ± 5.0
$W + \gamma$	43.3 ± 6.9	38.4 ± 6.2	58.0 ± 9.0	47.6 ± 7.1	51.3 ± 7.5
ECAL spikes	8.1 ± 0.6	6.6 ± 0.5	9.0 ± 0.7	5.9 ± 0.5	4.7 ± 0.4
Simulated background	15.1 ± 5.9	21.1 ± 7.3	28.9 ± 11.7	11.1 ± 5.2	24.1 ± 6.8
Misidentified jets	6.7 ± 0.2	5.2 ± 0.4	7.9 ± 0.1	6.1 ± 0.2	4.0 ± 0.3
Total predicted	327.1 ± 17.9	299.7 ± 18.2	481.4 ± 29.6	342.6 ± 18.2	354.4 ± 22.6
Observed	316.0 ± 17.8	290.0 ± 17.0	485.0 ± 22.0	323.0 ± 18.0	356.0 ± 18.9

	Photon p_T range				
	315 - 355 GeV	355 - 425 GeV	425 - 515 GeV	515 - 615 GeV	615 - 765 GeV
Fiducial signal	160.8 ± 13.5	131.1 ± 17.2	68.8 ± 9.7	28.6 ± 5.1	14.7 ± 5.0
Non-fiducial signal	3.9 ± 0.4	2.9 ± 0.3	1.2 ± 0.1	0.4 ± 0.1	0.1 ± 0.0
Misidentified electrons	44.8 ± 4.0	39.7 ± 3.7	17.0 ± 1.7	9.1 ± 1.1	5.1 ± 0.8
$W + \gamma$	29.8 ± 5.2	29.6 ± 4.8	11.0 ± 2.4	3.8 ± 1.2	1.5 ± 0.6
ECAL spikes	2.9 ± 0.2	1.9 ± 0.2	0.7 ± 0.1	0.2 ± 0.0	0.2 ± 0.0
Simulated background	14.3 ± 4.8	23.1 ± 6.2	11.1 ± 4.1	4.1 ± 2.0	0.7 ± 0.4
Misidentified jets	2.6 ± 0.2	1.4 ± 0.2	0.4 ± 0.1	0.2 ± 0.1	0.0 ± 0.0
Total predicted	259.1 ± 16.4	229.6 ± 19.7	110.0 ± 11.2	46.2 ± 5.8	22.4 ± 5.1
Observed	255.0 ± 16.0	232.0 ± 15.2	110.0 ± 10.5	51.0 ± 7.1	23.0 ± 4.8

	Photon p_T range
	765 - 1500 GeV
Fiducial signal	7.7 ± 3.6
Non-fiducial signal	0.0 ± 0.0
Misidentified electrons	2.8 ± 0.5
$W + \gamma$	0.5 ± 0.3
ECAL spikes	0.2 ± 0.0
Simulated background	0.7 ± 0.4
Misidentified jets	0.0 ± 0.0
Total predicted	12.0 ± 3.7
Observed	12.0 ± 3.5

Table 6.11: Breakdown of event counts in the signal region in the barrel in ranges of photon p_T

	Photon p_T range				
	225 - 230 GeV	230 - 235 GeV	235 - 270 GeV	270 - 310 GeV	310 - 360 GeV
Fiducial signal	14.2 ± 1.3	18.5 ± 1.6	121.3 ± 9.7	72.5 ± 5.7	40.3 ± 3.6
Non-fiducial signal	9.0 ± 0.8	4.8 ± 0.5	9.4 ± 0.8	3.4 ± 0.3	1.7 ± 0.2
Misidentified electrons	3.1 ± 0.7	2.7 ± 0.6	14.9 ± 3.1	8.4 ± 1.8	4.8 ± 1.0
$W + \gamma$	10.2 ± 2.4	2.5 ± 1.2	29.8 ± 4.8	14.1 ± 3.3	9.4 ± 2.0
Simulated background	3.2 ± 3.2	3.9 ± 4.7	43.3 ± 29.6	16.2 ± 16.6	3.2 ± 5.4
Misidentified jets	2.1 ± 0.7	0.5 ± 0.2	6.1 ± 0.2	2.6 ± 0.8	0.6 ± 0.3
Beam halo	0.8 ± 0.1	0.8 ± 0.0	4.1 ± 0.2	2.5 ± 0.1	1.1 ± 0.1
Total predicted	42.6 ± 4.4	33.8 ± 5.2	228.8 ± 31.9	119.7 ± 18.1	61.1 ± 7.0
Observed	56.0 ± 7.5	34.0 ± 5.8	246.0 ± 15.7	125.0 ± 11.2	69.0 ± 8.3

	Photon p_T range	
	360 - 440 GeV	440 - 1500 GeV
Fiducial signal	22.5 ± 3.0	11.6 ± 1.8
Non-fiducial signal	0.8 ± 0.1	0.3 ± 0.0
Misidentified electrons	2.8 ± 0.6	1.1 ± 0.3
$W + \gamma$	3.5 ± 1.1	2.1 ± 0.6
Simulated background	0.5 ± 3.4	0.1 ± 0.2
Misidentified jets	0.4 ± 0.3	0.0 ± 0.0
Beam halo	0.3 ± 0.0	0.0 ± 0.0
Total predicted	30.8 ± 4.8	15.4 ± 2.0
Observed	28.0 ± 5.3	10.0 ± 3.2

Table 6.12: Breakdown of event counts in the signal region in the endcaps in ranges of photon p_T

	Photon p_T range				
	225 - 235 GeV	235 - 245 GeV	245 - 265 GeV	265 - 285 GeV	285 - 315 GeV
$W + \gamma$	64.0 ± 10.2	55.3 ± 8.9	84.1 ± 13.0	71.1 ± 10.6	85.0 ± 12.3
Simulated background	15.9 ± 4.0	19.0 ± 4.0	26.9 ± 4.6	21.8 ± 3.3	26.1 ± 4.7
Misidentified jets	4.2 ± 0.4	2.8 ± 0.3	3.8 ± 0.1	3.3 ± 0.1	2.0 ± 0.0
Misidentified electrons	5.8 ± 0.6	4.3 ± 0.5	7.3 ± 0.6	5.2 ± 0.5	5.1 ± 0.4
Total predicted	89.9 ± 11.1	81.4 ± 9.8	122.1 ± 13.9	101.4 ± 11.2	118.3 ± 13.3
Observed	96.0 ± 9.8	90.0 ± 9.5	113.0 ± 10.6	104.0 ± 10.2	126.0 ± 11.2

	Photon p_T range				
	315 - 355 GeV	355 - 425 GeV	425 - 515 GeV	515 - 615 GeV	615 - 765 GeV
$W + \gamma$	55.3 ± 9.5	65.6 ± 10.6	29.2 ± 6.4	12.6 ± 3.8	6.0 ± 2.5
Simulated background	19.8 ± 4.4	12.7 ± 3.5	13.1 ± 4.1	3.6 ± 2.9	3.1 ± 1.3
Misidentified jets	1.6 ± 0.1	1.0 ± 0.0	0.3 ± 0.1	0.1 ± 0.0	0.0 ± 0.0
Misidentified electrons	3.2 ± 0.3	2.4 ± 0.2	0.5 ± 0.1	0.2 ± 0.0	0.0 ± 0.0
Total predicted	79.9 ± 10.6	81.7 ± 11.2	43.1 ± 7.6	16.4 ± 4.8	9.1 ± 2.8
Observed	75.0 ± 8.7	86.0 ± 9.3	40.0 ± 6.3	19.0 ± 4.4	9.0 ± 3.0

	Photon p_T range
	765 - 1500 GeV
$W + \gamma$	2.0 ± 1.2
Simulated background	0.5 ± 0.3
Misidentified jets	0.0 ± 0.0
Misidentified electrons	0.0 ± 0.0
Total predicted	2.6 ± 1.3
Observed	2.0 ± 1.4

Table 6.13: Breakdown of event counts in the $W(\rightarrow \mu\nu)+\gamma$ control region in the barrel in ranges of photon p_T

	Photon p_T range				
	225 - 230 GeV	230 - 235 GeV	235 - 270 GeV	270 - 310 GeV	310 - 360 GeV
$W + \gamma$	15.6 ± 3.6	4.1 ± 1.9	50.7 ± 8.2	27.3 ± 6.3	21.9 ± 4.6
Simulated background	1.2 ± 0.6	2.6 ± 0.9	10.2 ± 2.5	17.0 ± 6.4	3.4 ± 1.1
Misidentified jets	0.4 ± 0.3	0.8 ± 0.4	2.7 ± 0.4	1.4 ± 0.3	0.6 ± 0.2
Misidentified electrons	0.2 ± 0.1	0.3 ± 0.1	1.1 ± 0.2	0.8 ± 0.2	0.2 ± 0.0
Total predicted	17.4 ± 3.7	7.7 ± 2.2	64.7 ± 8.6	46.4 ± 9.0	26.1 ± 4.8
Observed	16.0 ± 4.0	7.0 ± 2.6	69.0 ± 8.3	49.0 ± 7.0	27.0 ± 5.2

	Photon p_T range	
	360 - 440 GeV	440 - 1500 GeV
$W + \gamma$	9.9 ± 3.1	8.6 ± 2.5
Simulated background	3.4 ± 1.7	0.7 ± 0.3
Misidentified jets	0.3 ± 0.2	0.0 ± 0.0
Misidentified electrons	0.1 ± 0.0	0.1 ± 0.0
Total predicted	13.7 ± 3.6	9.3 ± 2.5
Observed	12.0 ± 3.5	4.0 ± 2.0

Table 6.14: Breakdown of event counts in the $W(\rightarrow \mu\nu) + \gamma$ control region in the endcaps in ranges of photon p_T

	Photon p_T range				
	225 - 235 GeV	235 - 245 GeV	245 - 265 GeV	265 - 285 GeV	285 - 315 GeV
$W + \gamma$	51.0 ± 8.1	45.0 ± 7.2	67.7 ± 10.5	58.9 ± 8.8	71.4 ± 10.3
Simulated background	19.4 ± 8.3	12.6 ± 2.9	39.2 ± 17.4	17.7 ± 7.6	13.6 ± 4.9
Misidentified jets	3.6 ± 0.4	2.8 ± 0.3	4.7 ± 0.4	2.5 ± 0.3	2.3 ± 0.1
Misidentified electrons	5.3 ± 0.6	3.2 ± 0.3	6.1 ± 0.8	4.6 ± 0.5	3.3 ± 0.3
Total predicted	79.4 ± 11.7	63.7 ± 7.9	117.7 ± 20.4	83.7 ± 11.7	90.7 ± 11.6
Observed	76.0 ± 8.7	57.0 ± 7.5	128.0 ± 11.3	86.0 ± 9.3	83.0 ± 9.1

	Photon p_T range				
	315 - 355 GeV	355 - 425 GeV	425 - 515 GeV	515 - 615 GeV	615 - 765 GeV
$W + \gamma$	46.5 ± 8.0	54.3 ± 8.8	24.6 ± 5.4	11.1 ± 3.4	5.1 ± 2.2
Simulated background	19.1 ± 4.9	12.7 ± 6.8	2.9 ± 1.7	2.6 ± 1.4	1.5 ± 0.9
Misidentified jets	1.8 ± 0.0	1.0 ± 0.1	0.4 ± 0.0	0.2 ± 0.1	0.1 ± 0.0
Misidentified electrons	2.4 ± 0.2	0.9 ± 0.2	0.7 ± 0.1	0.0 ± 0.1	0.0 ± 0.0
Total predicted	69.8 ± 9.5	68.9 ± 11.1	28.5 ± 5.7	13.9 ± 3.7	6.7 ± 2.3
Observed	76.0 ± 8.7	64.0 ± 8.0	31.0 ± 5.6	11.0 ± 3.3	8.0 ± 2.8

	Photon p_T range
	765 - 1500 GeV
$W + \gamma$	2.0 ± 1.2
Simulated background	0.9 ± 0.4
Misidentified jets	0.0 ± 0.0
Misidentified electrons	0.0 ± 0.0
Total predicted	2.9 ± 1.3
Observed	3.0 ± 1.7

Table 6.15: Breakdown of event counts in the $W(\rightarrow e\nu)+\gamma$ control region in the barrel in ranges of photon p_T

	Photon p_T range				
	225 - 230 GeV	230 - 235 GeV	235 - 270 GeV	270 - 310 GeV	310 - 360 GeV
$W + \gamma$	14.9 ± 3.4	3.6 ± 1.7	43.1 ± 7.0	22.3 ± 5.2	19.9 ± 4.2
Simulated background	1.3 ± 0.6	0.4 ± 0.5	10.2 ± 3.6	1.0 ± 2.8	4.0 ± 2.2
Misidentified jets	1.3 ± 0.5	0.7 ± 0.4	5.0 ± 1.0	1.5 ± 0.5	0.5 ± 0.1
Misidentified electrons	0.3 ± 0.1	0.2 ± 0.1	1.2 ± 0.3	0.5 ± 0.1	0.2 ± 0.0
Total predicted	17.9 ± 3.5	4.8 ± 1.8	59.5 ± 7.9	25.3 ± 5.9	24.5 ± 4.8
Observed	16.0 ± 4.0	5.0 ± 2.2	52.0 ± 7.2	23.0 ± 4.8	23.0 ± 4.8

	Photon p_T range	
	360 - 440 GeV	440 - 1500 GeV
$W + \gamma$	8.6 ± 2.8	8.2 ± 2.3
Simulated background	3.8 ± 1.8	0.5 ± 0.2
Misidentified jets	1.1 ± 0.4	0.0 ± 0.0
Misidentified electrons	0.3 ± 0.1	0.1 ± 0.0
Total predicted	13.8 ± 3.3	8.8 ± 2.4
Observed	16.0 ± 4.0	15.0 ± 3.9

Table 6.16: Breakdown of event counts in the $W(\rightarrow e\nu) + \gamma$ control region in the endcaps in ranges of photon p_T

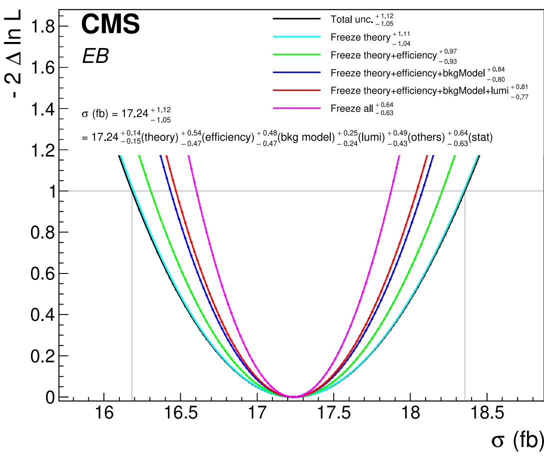
Chapter 7

Results

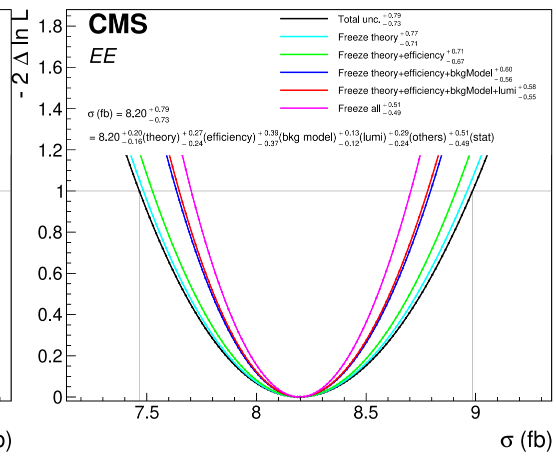
We present the results using the methods described in previous chapters. These include the estimates of the background processes, the systematic uncertainties, and the signal models for the cross-section and differential cross-section (as defined in Equation 6.7) and the signal models for aNTGCs (Equation 6.8).

7.1 Cross-section measurements

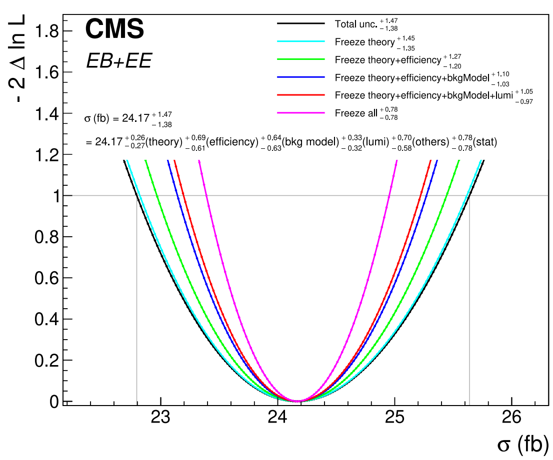
The cross-section of the process $pp \to Z(\to \nu\nu) + \gamma$ in the fiducial region defined in Table 6.7 is measured separately for the barrel and endcap phase spaces, and then combined for the full fiducial region. The cross-section is determined by a maximum likelihood fit to the observed data using a free parameter for the signal model as defined in Section 6.10.2. In order to determine the uncertainty, a scan of the profile likelihood ratio, $\lambda(\sigma)$, defined in Equation 6.4, is performed over the cross-section, σ. The variable $-2\ln\lambda(\sigma)$ follows a χ^2 distribution with one degree of freedom. The uncertainty in σ is determined by the range of σ for which $-2\ln\lambda(\sigma)$ lies below 1, which corresponds to a confidence level of 68% for the χ^2 distribution. The scan is performed separately for each fiducial phase space. The results of these scans are shown in Figure 7.1. The impact of the uncertainties is also assessed by repeating these scans with the associated nuisance parameters fixed to the best-fit values and comparing the change in the uncertainty. A breakdown of the uncertainties is shown in these figures. The statistical uncertainty is between 3% and 6%, and is the largest contribution to the overall uncertainty. The second largest uncertainty arises from the modeling of the identification efficiency of the photons and leptons in the control regions in the simulations, with the photon efficiency having the larger uncertainty. These contribute about 3% to the overall uncertainty. The theoretical uncertainties in the simulated background and the signals contribute between 1-2%. The estimates of the backgrounds derived from measurements in real data contribute between 3-5%. In addition, there is an uncertainty in the measurement of the integrated luminosity, which contributes about 1.5%.



(a) Barrel ($|\eta| < 1.4442$)



(b) Endcaps ($1.566 < |\eta| < 2.5$)



(c) Combination of barrel and endcaps

Figure 7.1: The measured cross-section in separate photon pseudorapidity ranges.

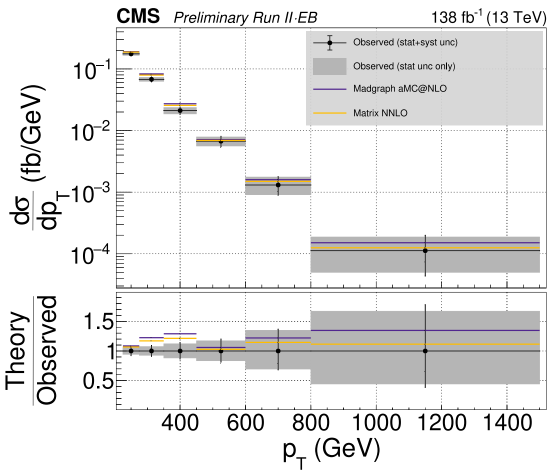
The measured cross-sections are compared with theoretical predictions calculated for the same fiducial phase spaces using the generators Madgraph5 [56, 57] and MATRIX [71, 67], which perform the QCD calculations at next-to-leading order (NLO) and next-to-next-to-leading order (NNLO) respectively. The measurements and predictions are summarized in Table 7.1. The measured cross-sections are found to favor the NNLO predictions more than the NLO predictions.

	Cross-section (fb)		
	Measured	NLO prediction (Madgraph5)	NNLO Prediction (MATRIX)
Barrel $\lvert\eta\rvert < 1.4442$	$17.24^{+1.12}_{-1.05}$	19.98 ± 0.73	$19.33 \pm +0.33$
Endcaps $1.4442 < \lvert\eta\rvert < 2.5$	$8.20^{+0.79}_{-0.73}$	6.56 ± 0.30	6.21 ± 0.09
Combination of barrel and endcaps	$24.17^{+1.47}_{-1.38}$	26.53 ± 0.97	25.45 ± 0.41

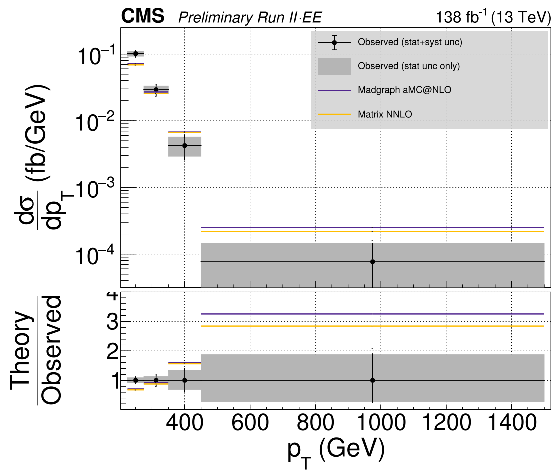
Table 7.1: Measured and predicted cross-sections in the barrel, the endcaps and the combined fiducial phase spaces. The predicted cross-sections are calculated at NLO and NNLO in QCD.

7.2 Differential cross-section measurements

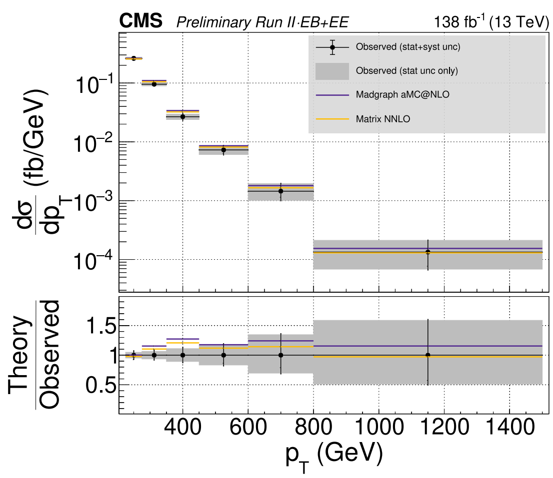
The differential cross-section of this process as a function of photon p_T is determined by a maximum likelihood fit using multiple free parameters that correspond to the cross-sections of the bins of photon p_T at the truth level as defined in 6.10.2. The uncertainties are assessed by performing separate scans of the profile likelihood ratio for each cross-section parameter. Figure 7.2 presents the results for the three phase spaces, including a comparison with theoretical predictions at NLO and NNLO in QCD. In the top panel, the y-axis displays the differential cross-section obtained by dividing the cross-section for each bin by the width of the bin. The lower panel shows the ratio between the predicted and measured cross-sections in each bin. The statistical uncertainties are represented by gray bands, while the full uncertainties are indicated by black error bars. The measurements in the barrel region show no significant disagreement with the predictions, particularly aligning better with the NNLO predictions. However, in the endcaps, there is a discrepancy that increases systematically with the photon p_T. The measurements for the combined phase space exhibit a similar level of agreement as the barrel region, since most of the selected events fall within this region.



(a) Barrel ($|\eta| < 1.4442$)



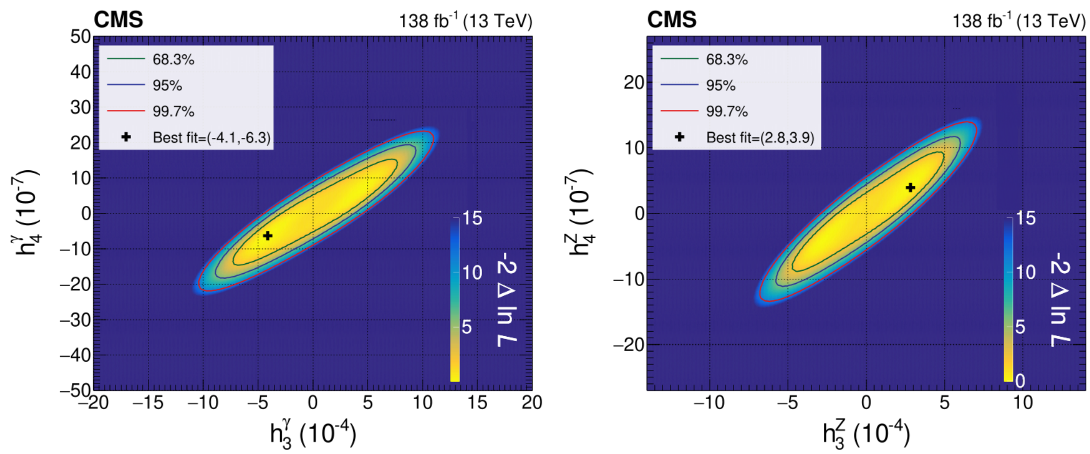
(b) Endcaps ($1.566 < |\eta| < 2.5$)



(c) Combination of barrel and endcaps

Figure 7.2: The measured differential cross-section in separate photon pseudorapidity ranges. Comparison is shown with theoretical predictions calculated at NNLO and NLO in QCD.

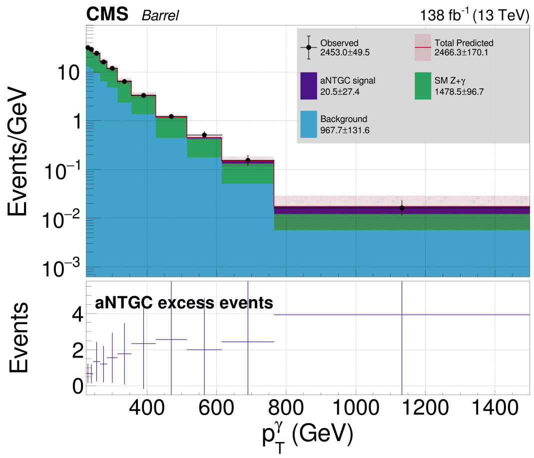
7.3 Search for aNTGCs



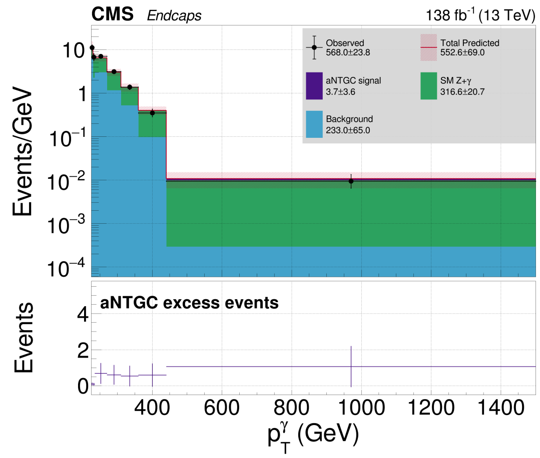
(a) (h_3^γ, h_4^γ) (b) (h_3^Z, h_4^Z)

Figure 7.3: Scan of $-2\ln\lambda(h_3^V, h_4^V)$ over the aNTGC parameters. Best fit values and contours corresponding to 68%, 95% and 99% confidence levels are shown.

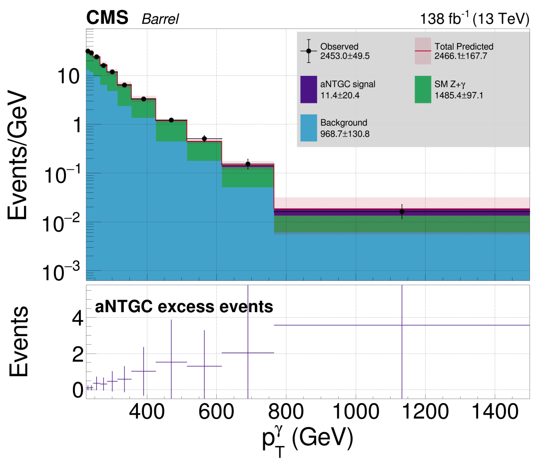
To test for the presence of anomalous neutral triple gauge couplings (aNTGC), the predicted event count for each bin of observed photon transverse momentum is parametrized by a quadratic function of (h_3^V, h_4^V), defined in Equation 6.8. Then a scan of the profile likelihood ratio, $\lambda(h_3^V, h_4^V)$, is performed over each pair of coupling parameters, (h_3^γ, h_4^γ) and (h_3^Z, h_4^Z) using the data observed in both the barrel and endcaps. The results of the scans are shown in Figure 7.3. The variable $-2\ln\lambda(h_3^V, h_4^V)$ follows a χ^2 distribution with two degrees of freedom. The confidence contours for the coupling parameters can be determined by examining the regions where $-2\ln\lambda(h_3^V, h_4^V)$ falls below specific values: 2.30, 5.99, and 9.21. These values correspond to confidence levels of 68%, 95%, and 99% respectively of the χ^2 distribution. In the figures for the scans above, the color scale corresponds to the observed values of $-2\ln\lambda(h_3^V, h_4^V)$ at each point. The best-fit values of the coupling parameters and confidence contours are also shown. The presence of aNTGC would be indicated if the areas within the confidence contours excluded the point $(0, 0)$, which corresponds to the SM prediction. No significant evidence for aNTGCs is found.



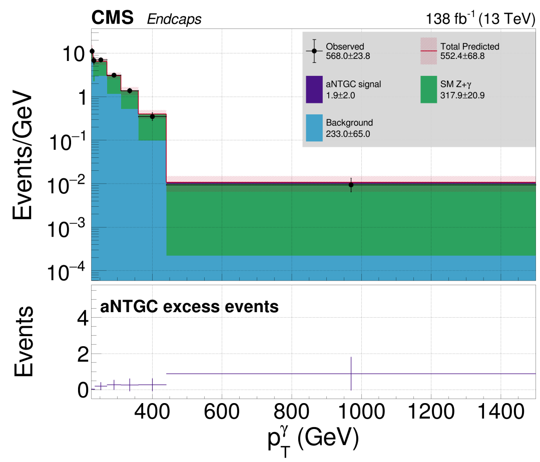
(a) Barrel events with best fit values of (h_3^γ, h_4^γ)



(b) Endcaps events with best fit values of (h_3^γ, h_4^γ)



(c) Barrel events with best fit values of(h_3^Z, h_4^Z)

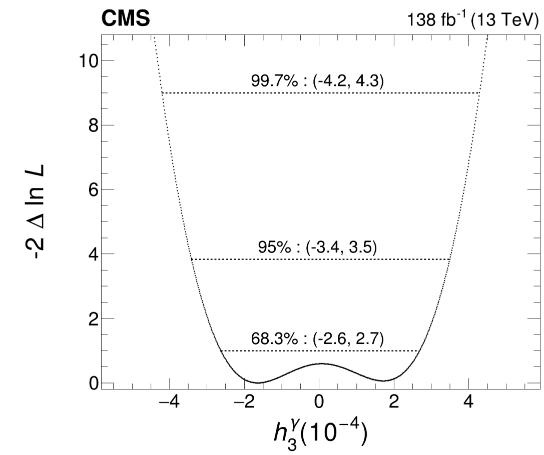


(d) Endcaps events with best fit values of (h_3^Z, h_4^Z)

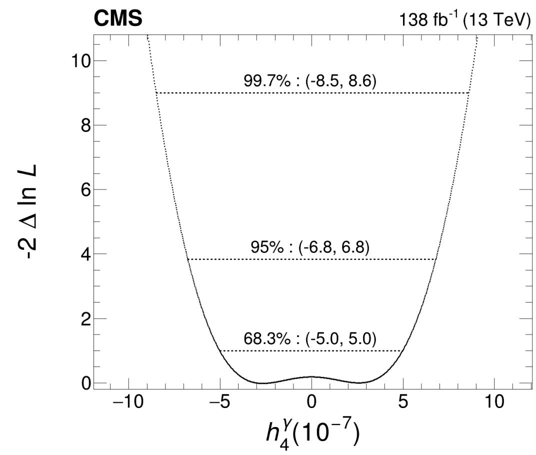
Figure 7.4: Observed and predicted distributions of photon transverse momentum at the best-fit values of (h_3^V, h_4^V)

.

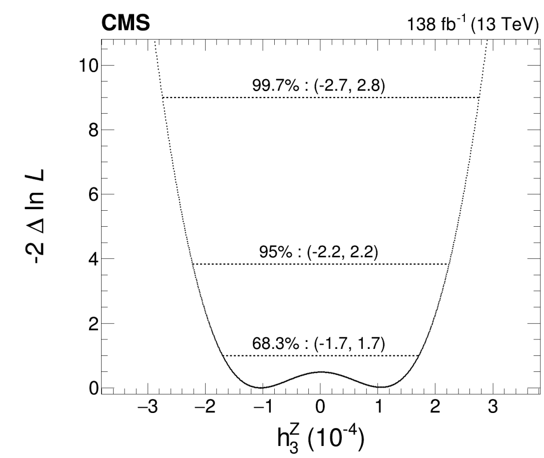
The photon transverse momentum distributions at the best fit values of (h_3^γ, h_4^γ) and (h_3^Z, h_4^Z) are displayed in Figure 7.4 separately for the barrel and endcap regions. The top panels show the density of events while the lower panels show the estimated number of events purely from aNTGC.



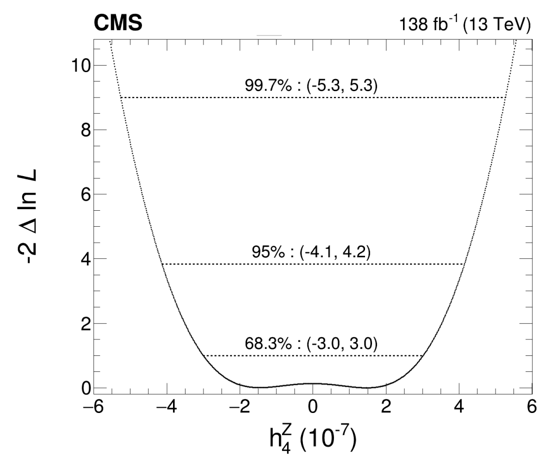
(a) h_3^γ



(b) h_4^γ



(c) h_3^Z



(d) h_4^Z

Figure 7.5: One-dimensional scan of $-2\ln\lambda(h)$ over each aNTGC parameter. Confidence intervals at 68%, 95% and 99% confidence levels are shown.

A 1D scan over each coupling parameter, with the other coupling parameter fixed at zero, is also performed to obtain the confidence intervals over the parameters (Figure 7.5). It should be noted that the confidence intervals obtained in this study have improved the bounds on the aNTGC coupling parameters by about an order of magnitude with respect to the study [1] conducted in the same final state with CMS data collected during LHC Run I at a center-of-mass energy of 8 TeV and integrated luminosity of 19.6 fb^{-1} (shown in Figure 2.8).

Chapter 8

Summary

A search was presented for the anomalous neutral triple gauge couplings $ZZ\gamma$ and $Z\gamma\gamma$ with parameters (h_3^Z, h_4^Z) and (h_3^γ, h_4^γ), respectively. Proton-proton collision data collected by the CMS detector during LHC Run II at a center-of-mass energy of 13 TeV, with an integrated luminosity of 138 fb^{-1}, were used. Measurements of fiducial cross-sections and differential cross-sections as a function of the photon transverse momentum were also performed. The photons were selected using a new machine-learning-based identification method that employs several shower shape variables describing the deposits in the electromagnetic calorimeter. This method, compared to a cut-based approach, increased the selection efficiency by approximately 12% without compromising jet background rejection. Furthermore, the pseudorapidity coverage of the photons was extended from 1.44 to 2.5 using another machine-learning algorithm that suppressed beam halo background in the extended region. The main background sources were estimated using data-driven techniques, while smaller contributions were estimated through simulation. The analysis did not reveal any significant evidence for anomalous couplings. The limits on the coupling parameters were tightened by approximately an order of magnitude compared to CMS analyses conducted with LHC Run I data. The measured cross-sections were compared with theoretical predictions at next-to-leading-order and next-to-next-to-leading-order in QCD, demonstrating closer agreement with the latter. Statistical uncertainties constituted the primary source of uncertainty in these studies due to the low production rate of events containing a high transverse momentum photon. Theoretical uncertainties arose primarily from the precision of the prediction for the $Z(\to \nu\nu) + \gamma$ production cross-section within the framework of the Standard Model, diminishing the significance of any excess associated with aNTGCs. The primary experimental uncertainty resulted from the modeling of the photon selection efficiency in the signal model. Currently, the analysis is undergoing internal review within the CMS collaboration.

Looking ahead, the LHC is currently in its Run III operation and is anticipated to collect data with twice the integrated luminosity of Run II. Subsequently, the LHC and its detectors will undergo upgrades for the High-Luminosity LHC (HL-LHC) phase to enable the collection of data with a significantly higher integrated luminosity of 3 ab^{-1}. In future searches, the statistical uncertainty will be substantially reduced. However, the experimental uncertainties may not necessarily see significant improvements due to the challenges posed by the higher pileup interactions resulting from collisions, which will require further advancements in reconstruction algorithms to mitigate their impact. On the theoretical front, efforts are ongoing to incorporate higher-order perturbative effects due to QCD and electroweak interactions to improve the precision of the predicted SM cross-section.

8.1 Note on contributions

The studies presented in this thesis would not have been possible without the invaluable contributions of numerous individuals within the CMS collaboration. This collaborative effort involves approximately 3,000 physicists, along with an additional 1,500 engineers, technicians, and support staff, who collectively contribute to the operation of the detector, event reconstruction, simulation sample production, and provision of computing resources. In the context of this thesis, several physicists worked closely with the author to achieve the goals of the research.

Dr. Shilpi Jain (UMN and Tata Institute of Fundamental Research) has made invaluable contributions by deriving the fake ratios used to estimate background events arising from electrons misidentified as photons in the barrel region. Dr. Jain's expertise was instrumental in developing a machine-learning-based tagger to mitigate beam halo background, accurately estimating the residual background resulting from beam halo, and deriving corrections for the photon BDT model.

Dr. Giulia Sorrentino (INFN, Trieste and Kansas State University) has played a crucial role by calculating efficiency correction scale factors for the photon identification method, and determining the fake ratios to estimate the background originating from electrons misidentified as photons in the endcap region.

Dr. De-Lin Xiong (Florida State University) conducted a comprehensive study on the background arising from ECAL spikes in the barrel region and estimated its contribution.

The author of this thesis played a central role by developing the photon identification method, generating simulation samples for the signal models, deriving theoretical corrections for these models, processing signal and background samples from the CMS computing servers, applying corrections to reconstructed objects using recipes derived by sub-groups within the CMS collaboration, estimating the background from jets misidentified as photons and from the $W + \gamma$ process, developing the signal models for the measurements of the cross-sections, differential cross-sections and the anomalous couplings, calculating the impact of the individual systematic uncertainties, and performing the statistical analyses for the obtained results.

References

[1] CMS Collaboration. Measurement of the $Z\gamma \to \nu\bar{\nu}\gamma$ production cross section in pp collisions at $\sqrt{s} = 8$ TeV and limits on anomalous $ZZ\gamma$ and $Z\gamma\gamma$ trilinear gauge boson couplings. *Physics Letters B*, 760:448–468, September 2016.

[2] ATLAS Collaboration. Measurement of the $Z\gamma \to \nu\bar{\nu}\gamma$ production cross section in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector and limits on anomalous triple gauge-boson couplings. *Journal of High Energy Physics*, 2018(12):10, December 2018.

[3] Mary K. Gaillard, Paul D. Grannis, and Frank J. Sciulli. The standard model of particle physics. *Rev. Mod. Phys.*, 71:S96–S111, March 1999.

[4] Wikimedia Commons. File:standard model of elementary particles.svg, 2019.

[5] Matthew D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, 1 edition, December 2013.

[6] W. Buchmüller and D. Wyler. Effective lagrangian analysis of new interactions and flavour conservation. *Nuclear Physics B*, 268(3):621–653, 1986.

[7] K. Hagiwara, R.D. Peccei, D. Zeppenfeld, and K. Hikasa. Probing the weak boson sector in $e^+e^- \to W^+W^-$. *Nuclear Physics B*, 282:253–307, 1987.

[8] D0 Collaboration. Study of $Z\gamma$ Events and Limits on Anomalous $ZZ\gamma$ and $Z\gamma\gamma$ Couplings in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV. *Physical Review Letters*, 95(5):051802, July 2005.

[9] CDF Collaboration. Measurement of $Z\gamma$ production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Physical Review D*, 82(3):031103, August 2010.

[10] CDF Collaboration. Limits on Anomalous Trilinear Gauge Couplings in $Z\gamma$ Events from $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV. *Physical Review Letters*, 107:051802, July 2011.

[11] CMS Collaboration. Measurement of the $Z\gamma$ production cross section in pp collisions at $\sqrt{s} = 8$ TeV and search for anomalous triple gauge boson couplings. *Journal of High Energy Physics*, 2015(4):164, April 2015.

[12] ATLAS Collaboration. Measurements of $Z\gamma$ and $Z\gamma\gamma$ production in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Physical Review D*, 93(11):112002, June 2016.

[13] D0 Collaboration. $Z\gamma$ production and limits on anomalous $ZZ\gamma$ and $Z\gamma\gamma$ couplings in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Physics Letters B*, 653(5):378–386, September 2007.

[14] ATLAS collaboration. Measurement of Wγ and Zγ production in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Journal of High Energy Physics*, 2011(9):72, September 2011.

[15] U. Baur and E. L. Berger. Probing the weak-boson sector in $Z\gamma$ production at hadron colliders. *Phys. Rev. D*, 47:4889–4904, June 1993.

[16] Lyndon Evans and Philip Bryant. LHC Machine. *Journal of Instrumentation*, 3(08):S08001, August 2008.

[17] CMS Collaboration. Precision luminosity measurement in proton–proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS. *The European Physical Journal C*, 81(9):800, September 2021.

[18] CMS Collaboration. CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV. CMS Physics Analysis Summary CMS-PAS-LUM-17-004, 2018.

[19] CMS Collaboration. CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV. CMS Physics Analysis Summary CMS-PAS-LUM-18-002, 2019.

[20] CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004, August 2008.

[21] CMS Collaboration. CMS tracking performance results from early LHC operation. *The European Physical Journal C*, 70(4):1165–1192, December 2010.

[22] CMS Collaboration. Alignment of the CMS tracker with LHC and cosmic ray data. *Journal of Instrumentation*, 9(06):P06009, June 2014.

[23] CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *Journal of Instrumentation*, 9(10):P10009, October 2014.

[24] CMS Collaboration. The CMS Phase-1 pixel detector upgrade. *Journal of Instrumentation*, 16(02):P02027, February 2021.

[25] CMS Collaboration. Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV. *Journal of Instrumentation*, 8(09):P09009, September 2013.

[26] Chia Ming and Kuo. The commissioning and first results on the performance of the CMS Preshower Detector. *Journal of Physics: Conference Series*, 293(1):012058, April 2011.

[27] Z. Antunovic, I. Britvitch, K. Deiters, N. Godinovic, Q. Ingram, A. Kuznetsov, Y. Musienko, I. Puljak, D. Renker, S. Reucroft, R. Rusack, T. Sakhelashvili, A. Singovski, I. Soric, and J. Swain. Radiation hard avalanche photodiodes for the CMS detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 537(1):379–382, 2005. Proceedings of the 7th International Conference on Inorganic Scintillators and their Use in Scientific adn Industrial Applications.

[28] K.W. Bell, R.M. Brown, D.J.A. Cockerill, P.S. Flower, P.R. Hobson, B.W. Kennedy, A.L. Lintern, C.S. Selby, O. Sharif, M. Sproston, and J.H. Williams. Vacuum phototriodes for the CMS electromagnetic calorimeter endcap. *IEEE Transactions on Nuclear Science*, 51(5):2284–2287, 2004.

[29] Christopher Rogan. Laser monitoring for the CMS ECAL. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 623(1):267–269, 2010. 1st International Conference on Technology and Instrumentation in Particle Physics.

[30] Marco Cipriani. Performance of the CMS electromagnetic calorimeter during the LHC Run II and its role in precision physics measurements. CMS Note CMS-CR-2018-162, CERN, Geneva, 2018.

[31] P. Bonamy et al. The ECAL Calibration: Use of the Light Monitoring System. Version 2.0. CMS Note CMS-NOTE-1998-013, Feb 1998.

[32] CMS Collaboration. Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data. *Journal of Instrumentation*, 5(03):T03012, March 2010.

[33] CMS Collaboration. Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s}$= 13 TeV. *Journal of Instrumentation*, 15(05):P05002, May 2020.

[34] P Cushman, A Heering, and A Ronzhin. Custom HPD readout for the CMS HCAL. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 442(1):289–294, 2000.

[35] CMS Collaboration. Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV. *Journal of Instrumentation*, 7(10):P10002, October 2012.

[36] CMS Collaboration. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV. *Journal of Instrumentation*, 13(06):P06015, June 2018.

[37] CMS Collaboration. The CMS Level-1 trigger system for LHC Run II. *Journal of Instrumentation*, 12(03):C03021, March 2017.

[38] CMS Collaboration. Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Journal of Instrumentation*, 15(10):P10017, October 2020.

[39] CMS Collaboration. The CMS high level trigger. *The European Physical Journal C*, 46(3):605–667, June 2006.

[40] CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. *Journal of Instrumentation*, 12(10):P10003, October 2017.

[41] Pierre Billoir. Progressive track recognition with a Kalman-like fitting procedure. *Computer Physics Communications*, 57(1):390–394, 1989.

[42] CMS Collaboration. Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. *Journal of Instrumentation*, 10(08):P08010, August 2015.

[43] W Adam, R Frühwirth, A Strandlie, and T Todorov. Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC. *Journal of Physics G: Nuclear and Particle Physics*, 31(9):N9, July 2005.

[44] CMS Collaboration. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *Journal of Instrumentation*, 16(05):P05014, May 2021.

[45] CMS Collaboration. Pileup mitigation at CMS in 13 TeV data. *Journal of Instrumentation*, 15(09):P09018, September 2020.

[46] CMS Collaboration. Jet energy scale and resolution in the cms experiment in pp collisions at 8 tev. *Journal of Instrumentation*, 12(02):P02014, February 2017.

[47] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-kt jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, April 2008.

[48] CMS Collaboration. Determination of jet energy calibration and transverse momentum resolution in CMS. *Journal of Instrumentation*, 6(11):P11002, November 2011.

[49] CMS Collaboration. Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector. *Journal of Instrumentation*, 14(07):P07004, July 2019.

[50] Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, August 2016.

[51] J. Bergstra, D. Yamins, and D. D. Cox. Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures. In *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ICML'13, pages I–115–I–123, Atlanta, GA, USA, June 2013. JMLR.org.

[52] Grégory Soyez. Pileup mitigation at the LHC: A theorist's view. *Physics Reports*, 803:1–158, 2019.

[53] Jeffrey D. Scargle, Jay P. Norris, Brad Jackson, and James Chiang. STUDIES IN ASTRONOMICAL TIME SERIES ANALYSIS. VI. BAYESIAN BLOCK REPRESENTATIONS. *The Astrophysical Journal*, 764(2):167, February 2013.

[54] Giovanni Franzoni. Dataset definition for CMS operations and physics analyses. *Nuclear and Particle Physics Proceedings*, 273-275:929–933, 2016. 37th International Conference on High Energy Physics (ICHEP).

[55] D J Lange, M Hildreth, V N Ivantchenko, and I Osborne. Upgrades for the CMS simulation. *Journal of Physics: Conference Series*, 608(1):012056, April 2015.

[56] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7):79, July 2014.

[57] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5: going beyond. *Journal of High Energy Physics*, 2011(6):128, June 2011.

[58] Céline Degrande, Claude Duhr, Benjamin Fuks, David Grellscheid, Olivier Mattelaer, and Thomas Reiter. UFO – The Universal FeynRules Output. *Computer Physics Communications*, 183(6):1201–1214, 2012.

[59] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Computer Physics Communications*, 191:159–177, 2015.

[60] S. Agostinelli et. al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.

[61] CMS Collaboration. Measurement of the Higgs boson inclusive and differential fiducial production cross sections in the diphoton decay channel with pp collisions at $\sqrt{s} = 13$ TeV. Technical Report arXiv:2208.12279, 2022.

[62] R. Bruce, M. Huhtinen, A. Manousos, F. Cerutti, L. Esposito, R. Kwee-Hinzmann, A. Lechner, A. Mereghetti, D. Mirarchi, S. Redaelli, and B. Salvachua. Collimation-induced experimental background studies at the CERN Large Hadron Collider. *Phys. Rev. Accel. Beams*, 22:021004, February 2019.

[63] CMS Collaboration. The CMS trigger system. *Journal of Instrumentation*, 12(01):P01020, January 2017.

[64] D A Petyt. Mitigation of Anomalous APD signals in the CMS Electromagnetic Calorimeter. *Journal of Physics: Conference Series*, 404:012043, December 2012.

[65] CMS Collabration. Reconstruction of signal amplitudes in the CMS electromagnetic calorimeter in the presence of overlapping proton-proton interactions. *Journal of Instrumentation*, 15(10):P10002, oct 2020.

[66] W. Cash. Parameter estimation in astronomy through application of the likelihood ratio. *The Astrophysical Journal*, 228:939–947, March 1979. ADS Bibcode: 1979ApJ...228..939C.

[67] Massimiliano Grazzini, Stefan Kallweit, and Dirk Rathlev. Wγ and Zγ production at the LHC in NNLO QCD. *Journal of High Energy Physics*, 2015(7):85, July 2015.

[68] Ansgar Denner, Stefan Dittmaier, Markus Hecht, and Christian Pasold. NLO QCD and electroweak corrections to Z + γ production with leptonic Z-boson decays. *Journal of High Energy Physics*, 2016(2):57, February 2016.

[69] CMS Collaboration. Jet energy scale and resolution measurement with Run 2 Legacy Data Collected by CMS at 13 TeV. CMS Detector Performance Summary CMS-DP-2021-033, 2021.

[70] CMS Collaboration. Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV. *Journal of Instrumentation*, 10(02):P02006, February 2015.

[71] Massimiliano Grazzini, Stefan Kallweit, and Marius Wiesemann. Fully differential nnlo computations with matrix. *The European Physical Journal C*, 78(7):537, June 2018.

[72] J Collins. Choosing the renormalization/factorization scale (QCD). *Journal of Physics G: Nuclear and Particle Physics*, 17(10):1547, October 1991.

[73] Matteo Cacciari, Stefano Frixione, Giovanni Ridolfi, Michelangelo L. Mangano, and Paolo Nason. The $t\bar{t}$ cross-section at 1.8 and 1.96 TeV: a study of the systematics due to parton densities and scale dependence. *Journal of High Energy Physics*, 2004(04):068, May 2004.

[74] The NNPDF Collaboration. Parton distributions with theory uncertainties: General formalism and first phenomenological studies. (arXiv:1906.10698), 2019.

[75] J. M. Lindert, S. Pozzorini, R. Boughezal, J. M. Campbell, A. Denner, S. Dittmaier, A. Gehrmann-De Ridder, T. Gehrmann, N. Glover, A. Huss, S. Kallweit, P. Maierhöfer, M. L. Mangano, T. A. Morgan, A. Mück, F. Petriello, G. P. Salam, M. Schönherr, and C. Williams. Precise predictions for V+jets dark matter backgrounds. *The European Physical Journal C*, 77(12):829, December 2017.

[76] W. Beenakker, A. Denner, S. Dittmaier, R. Mertig, and T. Sack. High-energy approximation for on-shell W-pair production. *Nuclear Physics B*, 410(2):245–279, 1993.

[77] J F Owens and W Tung. Parton distribution functions of hadrons. *Annual Review of Nuclear and Particle Science*, 42(1):291–332, 1992.

[78] Richard D. Ball, Valerio Bertone, Stefano Carrazza, Luigi Del Debbio, Stefano Forte, Patrick Groth-Merrild, Alberto Guffanti, Nathan P. Hartland, Zahari Kassabov, José I. Latorre, Emanuele R. Nocera, Juan Rojo, Luca Rottoli, Emma Slade, and Maria Ubiali. Parton distributions from high-precision collider data: NNPDF Collaboration. *The European Physical Journal C*, 77(10):663, October 2017.

[79] Jon Butterworth, Stefano Carrazza, Amanda Cooper-Sarkar, Albert De Roeck, Joël Feltesse, Stefano Forte, Jun Gao, Sasha Glazov, Joey Huston, Zahari Kassabov, Ronan McNulty, Andreas Morsch, Pavel Nadolsky, Voica Radescu, Juan Rojo, and Robert Thorne. PDF4LHC recommendations for LHC Run II. *Journal of Physics G: Nuclear and Particle Physics*, 43(2):023001, February 2016.