UNIVERSITÀ
DEGLI STUDI
DI TRIESTE

UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXXV CICLO DEL DOTTORATO DI RICERCA IN FISICA

Measurement of the $Z\gamma$ production cross section in the $Z \to \nu\bar{\nu}$ channel and limits on anomalous triple neutral gauge couplings in proton-proton collisions at $\sqrt{s} = 13$ TeV with the CMS experiment at the LHC

Settore scientifico-disciplinare: FIS/01 FISICA SPERIMENTALE

DOTTORANDO/A:
Giulia Sorrentino

COORDINATORE:
Prof. Francesco Longo

SUPERVISORE DI TESI:
Dr. Vieri Candelise

CO-SUPERVISORE DI TESI:
Prof. Giuseppe Della Ricca

ANNO ACCADEMICO 2021/2022

Contents

Introduction **iv**

1 The Standard Model of elementary particles **1**
1.1 Historical chronicle . . . 1
1.2 The local gauge invariance principle . . . 4
1.3 Electroweak interactions . . . 5
1.3.1 The neutral current and the electroweak unification . . . 7
1.3.2 Triple and quartic gauge couplings . . . 9
1.4 Strong interactions . . . 9
1.4.1 Colour confinement and hadronization . . . 10
1.4.2 Asymptotic freedom and the running of α_S . . . 10
1.4.3 Electroweak precision physics . . . 12
1.5 The Higgs boson . . . 13
1.5.1 The Spontaneous Symmetry Breaking . . . 13
1.5.2 The Brout-Englert-Higgs mechanism . . . 15
1.5.3 Search for the Higgs boson . . . 17
1.6 Physics beyond the Standard Model . . . 19
1.6.1 Monophoton signatures for dark matter searches at the Large Hadron Collider . . . 21

2 Anomalous gauge couplings in the Effective Field Theory framework **23**
2.1 The Effective Field Theory . . . 23
2.2 Triple gauge couplings . . . 24
2.2.1 Neutral triple gauge couplings . . . 26
2.3 $Z\gamma$ production . . . 28

3 The Compact Muon Solenoid experiment at the Large Hadron Collider **29**
3.1 The Large Hadron Collider . . . 29
3.2 The Compact Muon Solenoid Experiment . . . 31
3.2.1 Coordinate system . . . 31
3.2.2 The superconducting magnet . . . 33
3.2.3 The silicon tracker . . . 33
3.2.4 The electromagnetic calorimeter . . . 34
3.2.5 The hadronic calorimeter . . . 35
3.2.6 The muon system . . . 36
3.2.7 Triggering and data acquisition . . . 38
3.3 The Particle Flow algorithm for particle reconstruction . . . 39
3.3.1 Tracking and vertexing . . . 39
3.3.2 Calorimeter clustering . . . 41
3.3.3 The Link algorithm . . . 42
3.3.4 Particle reconstruction . . . 43

4 The MIP timing detector for the CMS Phase 2 upgrade **57**
4.1 High-luminosity LHC . . . 57
4.2 The CMS Phase 2 upgrade . . . 60
4.2.1 Tracker . . . 60
4.2.2 Endcap calorimeter . . . 60
4.2.3 Endcap muons . . . 61
4.2.4 Trigger and DAQ system . . . 62
4.3 The MIP timing detector project . . . 62
4.4 The MIP timing detector . . . 63
4.4.1 MTD performances in simulation . . . 65
4.4.2 Impact of MTD on physics analysis . . . 67

5 Analysis of $Z(\rightarrow \nu\bar{\nu}) + \gamma$ events **73**
5.1 Datasets and Monte Carlo samples . . . 73
5.1.1 Datasets . . . 73
5.1.2 Monte Carlo samples . . . 74
5.2 Event selection . . . 75
5.3 Photon hybrid identification method . . . 77
5.4 Background estimation . . . 85
5.4.1 Wγ background contribution . . . 86
5.4.2 Electron-photon misidentification background contribution . . . 86
5.4.3 Jet-photon misidentification background contribution . . . 89
5.4.4 Non collision background events in the ECAL barrel . . . 93
5.4.5 Non collision background events in the ECAL endcaps . . . 95

6 Measurement of the Z($\rightarrow \nu\bar{\nu}$) cross section and limits on the anomalous neutral triple gauge couplings **101**
6.1 Statistical analysis . . . 101
6.1.1 Cross section measurement . . . 101
6.1.2 Extraction of the limits on the aNTGCs . . . 102
6.2 Systematic uncertainties . . . 104
6.3 Results . . . 107
6.3.1 The blinded analysis . . . 107
6.3.2 The expected cross section . . . 108
6.3.3 The expected limits on anomalous neutral triple gauge couplings . . . 110
6.3.4 The observed cross section and the observed limits on anomalous neutral triple gauge couplings . . . 110

Conclusions **114**

Appendices **115**

A Systematics impacts **117**

B Photon hybrid ID **121**

C BDT shape corrections **127**

Introduction

The study of multiboson production is a powerful and effective way to probe the electroweak sector of the Standard Model (SM). Multiboson processes are indeed sensitive to possible contributions beyond the SM (BSM) that would manifest themselves as deviations from the SM predictions, for example enhancing the event yield in specific kinematic regions. These new physics effects can be parameterized in terms of anomalous gauge couplings using the framework of Effective Field Theories (EFT), that has the great advantage of being model-independent, thus providing more exploratory power with respect to a specific BSM theory. Among all the multiboson processes that can be explored, the associated production of a photon and a Z boson decaying into two neutrinos is of particular interest. Given that the neutrinos cannot be directly detected in a general-purpose detector like CMS, the signature of the $\mathrm{Z}(\rightarrow \nu\bar{\nu})\gamma$ production consists in a single, isolated photon. The same *monophoton* signature is shared by several physics analysis that search for dark matter at hadron colliders, where pairs of dark matter candidates can mimic the presence of neutrinos, and where the SM $\mathrm{Z}(\rightarrow \nu\bar{\nu})\gamma$ processes therefore represent the most important background contribution. A careful and precise measurement of the $\mathrm{Z}(\rightarrow \nu\bar{\nu})\gamma$ production becomes then compelling, for both its strong capability of testing the SM and for the contribution that it would provide to a wide range of fundamental physics searches. The $\mathrm{Z}(\rightarrow \nu\bar{\nu})\gamma$ cross section has been measured and searches for anomalous gauge couplings have been performed first by experiments operating at the LEP and Tevatron colliders, in electron-positron and proton-antiproton collisions respectively. The most recent results are those obtained at the LHC, by the CMS and ATLAS collaborations. The CMS analysis has been carried out by at a center of mass energy of 8 TeV, while the latest results from ATLAS exploit the data collected during 2016 (with a luminosity of 36.1 fb^{-1}), at a center of mass energy of 13 TeV. One of the most significant limiting factors for such measurements is the low statistics, since the possible presence of BSM effects would be only visibile at very high values (> 200 GeV) of the transverse momentum of the photon. For this reason, the measurement of the $\mathrm{Z}(\rightarrow \nu\bar{\nu})\gamma$ cross section and the evaluation of the limits on the anomalous couplings that are presented in this work are performed using all the data collected by the CMS experiment during the whole Run 2 of the LHC, at a center of mass energy of 13 TeV and for a total luminosity of 137.6 fb^{-1}. Specific strategies and algorithms have been implemented for identifying the signal photon and, in addition, the photons detected in the forward regions of the CMS detector are for the very first time taken into account when defining the phace space. The inclusion of forward photons is possible thanks to a new algorithm specifically developed for this analysis that allows to reject beam-halo events, the otherwise dominant and ineliminable background events when looking at the monophoton signature in the forward region. A beam halo event occurs when the protons interact with residual gas molecules in the vacuum chamber or with the walls of the beam pipe, giving rise to a flow of secondary muons traveling along the beam line that can lead to electromagnetic showers similar to those produced by photons.

Chapter 1 opens this work with an "historical" overview of the SM, underlining the theoretical and experimental milestones that led to its current formulation. The electroweak and the strong interactions are then described, with a special attention paid to

the triple and quartic couplings arising from the interactions between electroweak gauge bosons. A section is dedicated to the Brout-Englert-Higgs mechanism and the discovery of the Higgs boson. The first chapter closes with a summary of the still open questions in particle (but not only) physics, like the so-called hierarchy problem, or the dark matter. Within this context, the most studied channels for the production of dark matter candidates at the LHC using a monophoton signature are also presented. Chapter 2 focuses on anomalous gauge couplings and their parameterization through EFT. Anomalous triple gauge couplings (aTGCs) are first introduced using both the Lagrangian approach and the vertex approach, while a dedicated section describes in more detail the neutral TGCs (NTGCs) arising in the associated production of two neutral gauge bosons. The state of art of the $Z(\rightarrow \nu\bar{\nu})\gamma$ cross section measurements and of the limits on the aNTGCs is presented in the last section of the chapter. The Large Hadron Collider and the Compact Muon Solenoid experiment are introduced in Chapter 3, specifying the design and the main features of the different CMS subdetectors. The Particle Flow algorithm used for particle reconstruction is then presented, together with a description of the reconstruction flow for the physics objects (electrons, photons, muons, hadrons, jets and missing transverse energy) and of the mitigation techniques when dealing with pileup events. The post-reconstruction corrections to be applied on jets and missing transverse energy are outlined.

Chapter 4 is dedicated to the High-Luminosity upgrade of LHC, and to the consequent upgrade of the CMS subdetectors. Among them, a special attention is paid to the new timing detector for minimum ionizing particles (MTD) that will be installed in 2025. In parallel with the study of the $Z(\rightarrow \nu\bar{\nu})\gamma$ production at 13 TeV, studies on the timing performances have been carried out and hence they are included in this thesis work: the main contributions include the development of new monitoring tools for the MTD timing performances and for the evaluation of the MTD impact on the reconstructed tracks and vertices in terms of the track-vertex association. A concise description of the MTD design is also provided, and the expected performances evaluated through simulation are shown. The last section of the chapter contains an overview of the main physics analysis which will benefit from the inclusion of the timing information provided by the MTD.

Chapters 5 and 6 contain the core part of this work. In Chapter 5, all the datasets and simulated samples used for the analysis of $Z(\rightarrow \nu\bar{\nu})\gamma$ events at 13 TeV are described. The event selection is presented, and a section is dedicated to the development of the customized ID used for the identification of the signal photon. The estimation of the background contributions is then described, defining two categories of background processes: those which have a real photon in the final state, and those where another physics object is wrongly reconstructed as one. A description of the dedicated algorithm developed for the rejection of beam-halo events is provided. At the beginning of Chapter 6, the statistical procedure for the extraction of the cross section and of the limits on the anomalous couplings is defined. The systematic uncertainties affecting the measurement are then listed, and the results for the expected values for the cross section and the limits on the anomalous couplings are presented.

Chapter 1

The Standard Model of elementary particles

The Standard Model (SM) of particle physics is the theoretical model that describes the elementary particles and their reciprocal interactions. It includes three of the four fundamental forces existing in the Universe: the electromagnetic, the electroweak and the strong force, while the gravitational force is too weak to have a significant role at quantum level. This chapter opens with a historical overview of particle physics, with the aim of conveying the milestones that have contributed to the modern formulation of the SM. The particle content and the theoretical features of the SM are then presented, starting with the description of the local gauge invariance, the fundamental principle on which the whole theory relies. The theory of the strong interactions and the Higgs-Englert-Brout mechanism are described, together with a brief overview of the renowed discovery of the Higgs boson in 2012. A special attention is paid to the theory of the electroweak interactions of gauge bosons and to the triple gauge couplings arising from the electroweak Lagrangian. Eventually, the current issues faced by the SM are briefly listed, and some of the most popular Beyond the Standard Model theories are presented.

1.1 Historical chronicle

The idea that all matter is composed of elementary particles dates back to at least the 6th century BC [1]. In the 19th century, Dalton formally introduced the atomic theory into chemistry, stating that all the elements are made of tiny particles called atoms. Atoms were believed to be fundamental and indivisible objects until 1897, when Thompson discovered the electron [2] and proposed the *plum pudding* atomic model, where electrons were supposed to be immersed in a uniform "sea" of positive charge providing the atom's neutrality. Rutherford's famous scattering experiment [3] disproved Thomas' model in 1909, demonstrating that most of the mass and positive charge of an atom is instead concentrated in a tiny core, or *nucleus*, at the center of the atom. In 1932 Chadwick [4] discovered the neutral counterpart of the proton, the neutron, explaining why atoms heavier than hydrogen, and hence consisting of more than one proton, carried a bigger weight compared to the simple sum of the protons' masses. Around the same time of Thompson's electron discovery in 1900, Planck conceived the idea of the *quantization* of the electromagnetic radiation [5] while trying to solve the problem of the ultraviolet catastrophe in the black-body spectrum for the electromagnetic radiation emitted by a hot object. The concept of a *light quantum* was introduced by Einstein in 1905 to explain the photoelectric effect [6]. In spite of the heavy criticism received about the idea of light behaving as a particle on the subatomic scale, studies conducted by Millikan in 1916 on the photoelectric effect [7], and Compton's scattering experiment in 1923 [8] eventually confirmed the

Einstein's thesis, paving the way to quantum mechanics and quantum field theory. The term *photon* associated to this new "light particle" was suggested by the chemist Lewis in 1927 [9].

In 1931, while studying the composition of cosmic rays, Anderson discovered the *positron* [10], the positive-charged twin of the electron and first antiparticle to be detected in the history of particle physics. The existence of electron counterparts was originally postulated by Dirac in 1928, to explain the anomalous negative-energy quantum states predicted by its equations for relativistic electrons [11]. Going back to the classical model of the atom, as defined after the discovery of neutron in 1932, a significant problem can be (and it was) immediately observed: being positively charged, the protons packed in an atomic nucleus should violent repel each other. Since this does not clearly happen, there should be a kind of force binding protons and neutrons together. The first significant theory of the *strong force* was proposed in 1934 by Yukawa [12], assuming the existence of a field keeping protons and neutrons together, similar to the electric field that attracts an electron to a nucleus, and to the gravitational field that keeps the Moon rotating around the Earth. According to the quantum theory, such field should have been quantized with a proper particle acting as a mediator for the strong force, analogously to the photon when dealing with the electromagnetic force. Yukawa called this particle *meson* (from the greek term for "middle-weight"), and in the same spirit the electron took the name of *lepton* ("light-weight") while the protons and neutrons that of *baryons* ("heavy-weight"). In 1947, Powell and his colleagues discovered two middle-weight particles in cosmic rays [13], which they called π (or *pion*) and μ (or *muon*), and at the end of the same year the neutral *kaon* K^0 was detected in Rochester and Butler's cloud chamber [14]. In 1949 and 1950, the charged kaons and the *lambda* Λ were discovered, by Brown and Anderson's group respectively [15, 16], followed by many others mesons and baryons in the very next years. These new particles were called *strange* particles, for several reasons: first of all, they were completely unexpected, as the Yukawa model did not imply the existence of any other meson. Furthermore, they were produced on a time scale of about 10^{-23} seconds, but decayed much more slowly (in about 10^{-10} seconds). Pais and others hence suggested that the mechanism responsible for their production was different from the one regulating their decay [17].

In 1953, Gell-Mann and Nishijima [18, 19] implemented Pais' idea assigning a new property (the so-called *strangeness*) to each particle, conserved in any strong interaction (and hence in the production of strange particles) but not conserved in a *weak interaction* (responsible instead for the decay of strange particles). The first theory of the weak interaction was proposed by Fermi in 1933 [20], addressing a problem observed in the nuclear beta decay. In the beta decay, as known at the time, a radioactive nucleus decays into a slighly lighter one, with the emission of an electron. In a two-body decay, the energies of all the partecipating particles are kinematically determined in the center of mass frame, but what was observed experimentally was a continuous spectrum for the electron's energy. Instead of a two-body decay then, Fermi postulated the existence of a third, light neutral particles (which he called *neutrino*) emitted together with the electron and the lighter nucleus. He expressed the relativistic Hamiltonian of the beta decay and its corresponding matrix element in terms of a contact interaction, hence involving the presence of a new contact force, precisely the weak force, with no range. In 1954, Yang and Mills [21] extended the concept of gauge theory for abelian groups (quantum electrodynamics) to nonabelian groups, to provide an explanation for the strong interactions. In 1961, after Wu demonstrated the parity violation in the weak interactions [22], Glashow unified the description of the electromagnetic and weak forces [23], laying the basis for the present definition of the Standard Model. In the same year, all the known particles were arranged by Gell-Mann in a scheme called the *Eightfold Way* [24], that ordered baryons and mesons

into specific geometrical patterns according to their charge and strangeness. For example, the eight lightest baryons were fit into a hexagonal array with two particles at the center, where the horizontal lines in the hexagon connected particles with same strangeness, and the diagonal lines those with same charge. Starting from the Eightfold Way, Gell-Mann and Zweig [25, 26] independently proposed a model in which all *hadrons* (from the greek word for "massive") were formed by even more elementary constituents, named *quarks*[1] by Gell-Mann. According to this model, all the baryons are composed of three quarks (or three *anti*quarks in the case of *anti*baryons), and all the mesons of one quark and one antiquark. All the multiplets of the Eightfold Way easily emerged by postulating the existence of three quarks (u, d and s, or *up*, *down* and *strange*), and the quark model also provided an explanation for the deep inelastic scattering experiments involving protons that were carried out in the 1960s: the results from all those experiments indeed pointed toward a protonic substructure made of three massive "lumps". In 1974, the model was definitely affirmed by the simultaneous discovery of the J/ψ [27, 28] by the Ting's and Burton's groups, a particle characterized by an exceptional long lifetime (of the order of 10^{-20} seconds) compared to the typical lifetimes of hadrons (about 10^{-23} seconds). The quark model provided a striking and natural explanation for the J/ψ, implying the presence of a fourth quark (named *charm*, or c) with the J/ψ being a bound state of a c quark and antiquark. The idea of a fourth quark was already proposed by Glashow many years earlier, based on the simple idea of a parallel between the number of quarks and leptons. After the discovery of neutrinos by Cowan and Reines in 1956 [29], the known leptons were indeed four (the electron, the muon, and their corresponding neutrinos). In 1975 a new lepton, the τ (*tau*), was discovered [30]. Together with its corresponding neutrino, it brought the number of leptons up to six. Two years later, the discovery of the Υ meson [31] proved the existence of a fifth quark (named *beauty*, or b), and eventually in 1998 also the sixth and last quark predicted (named *top*, or t) was found [32, 33].

The current formulation of the SM was developed around 1968, when Weinberg and Salam incorporated the Higgs mechanism into Glashow's electroweak interaction. In the current view, the SM predicts three families of elementary particles: leptons, quarks, and mediators (Figure 1.1). The constituents of matter are leptons and quarks, 1/2-spin particles called *fermions*, whose dynamics is described by the Dirac equations of relativistic quantum mechanics. For each of the twelve fermions, an antiparticle with exactly the same mass and opposite charge exists. Antiparticles are conventionally labelled with a bar on top (i.e. an antiquark is labelled as $\bar{q}$). There are three negatively charged leptons, the electron (e), the muon (μ) and the tau (τ), and three neutrinos (ν_e, ν_μ, ν_τ). Leptons are grouped in three *generations*, each one constituted of a charged lepton and its corresponding neutrino; the charged lepton mass increases from the first ($m_e \simeq 0.511$ MeV) to the third ($m_\tau \simeq 1777$ MeV) family, while neutrino masses are very small ($\lesssim$ 1 eV) and their hierarchy has not been determined yet. There are six *flavors* of quarks: up (u), down (d), charm (c), strange (s), top (t) and bottom (b). Quarks are classified by the charge and the isospin 3-component: there are three up-type quarks (the isospin 3-component is 1/2 and the charge is 2/3 of the electron charge), and three down-type quarks (the isospin 3-component is -1/2 and the charge is -1/3 of the electron charge). As leptons, quarks are grouped in three generations, each of them constituted of an up-type quark and a down-type quark. Due to the nature of the strong interaction potential, quarks cannot be observed as free particles but they are always confined in bound states (hadrons). Each of the three forces of relevance to particle physics is described by a Quantum Field Theory (QFT) corresponding to the exchange of a mediator, a spin-1 force-carrying particle known as a gauge boson. The photon is the gauge boson of Quantum Electrodynamics (QED), the QFT of electromagnetic interaction; the gluon is the gauge boson of Quantum

[1] From the word "quark" in James Joyce's 1939 book "Finnegans Wake".

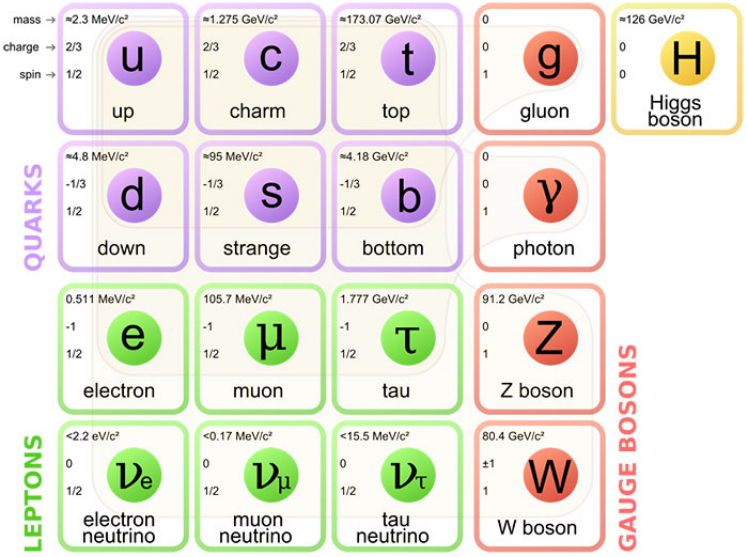

Figure 1.1: The Standard Model of particle physics: leptons, quarks, gauge and Higgs boson, together with their charge, mass and spin.

Chromodynamics (QCD), the QFT of strong interaction. The gluon and the photon are massless. The weak interaction is mediated by the charged W^+ and W^- bosons, and by the neutral Z boson. All gauge bosons of weak interactions are massive. The final element of the SM is the Higgs boson, which was discovered by the CMS [34] and ATLAS [35] experiments at the Large Hadron Collider (LHC) in 2012. It is a massive spin-0 scalar particle, and it provides the mechanism by which all other particles acquire mass. The process of interaction by boson exchange can be described by means of Feynman diagrams. The left-hand side of the diagram represents the initial state, while the right-hand side represents the final state, with time flowing from the left to the right. The central part of the diagram shows the particles exchanged and the SM vertices involved in the interaction: each vertex is characterized by a coupling constant which determines the interaction strength, and energy and momentum are conserved. The Feynman diagram for the scattering process $a + b \to c + d$ through the exchange of another particle X is shown in Fig. 1.2.

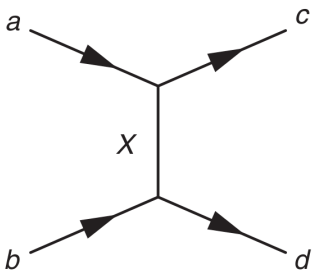

Figure 1.2: The Feynman diagram for the $a + b \to c + d$ scattering.

1.2 The local gauge invariance principle

The current formulation of the SM is based on a $\mathrm{SU(3)_C} \times \mathrm{SU(2)_L} \times \mathrm{U(1)_Y}$ gauge structure, meaning that the SM Lagrangian is invariant under local transformations of the gauge group. The concept of local gauge invariance can be introduced starting from electromagnetism [36], where the physical electric and magnetic fields do not change under

gauge transformation of the scalar and vector potential ϕ and $\vec{A}$ from where they are derived,

$$\phi \to \phi' = \phi - \frac{\partial \chi}{\partial t} \text{ and } \vec{A} \to \vec{A'} = \vec{A} + \nabla\chi.$$

Consider a free fermion field ψ with mass m, which has a global U(1) symmetry such that its Lagrangian is invariant under the U(1) global phase transformation $\psi \to \psi'(x) = \psi(x)e^{i\alpha}$, where α is an arbitrary real number:

$$\mathcal{L}_0 = i\overline{\psi}(\gamma^\mu \partial_\mu - m)\psi.$$

Suppose now to require an U(1) local phase transformation such that α depends on the space-time coordinate, $\alpha \to \alpha(x)$. In order to preserve the invariance of $\mathcal{L}_0$, the derivative ∂_μ should be replaced by the covariant derivative D_μ:

$$\partial_\mu \to D_\mu = \partial_\mu + iqA_\mu,$$

provided that the new field A_μ, corresponding to a massless gauge boson, transforms as:

$$A_\mu \to A'_\mu = A_\mu - \partial_\mu \alpha.$$

Therefore, the gauge-invariant Lagrangian of QED can be written as:

$$\mathcal{L}_0 = i\overline{\psi}(\gamma^\mu \partial_\mu - m_e)\psi + e\overline{\psi}\partial^\mu A_\mu \psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu},$$

where the kinetic term $F_{\mu\nu}F^{\mu\nu} = (\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\nu A^\mu - \partial^\mu A^\nu)$ is already invariant under U(1) local gauge transformation, and the term containing A_μ describes the interaction of the fermions with a new gauge field that can be identified as the photon. The QED four-vector current $j^\mu = \overline{\psi}\gamma^\mu\psi$ hence becomes

$$j^\mu = \overline{u}(p')\gamma^\mu u(p),$$

where u and $\overline{u}$ are the Dirac spinor and its adjoint (depending on the fermions four-momentum p and p') and γ^μ refers to the four Dirac matrices. By requiring the invariance of the Lagrangian under U(1) transformation all the QED can be derived, including Maxwell's equations.

1.3 Electroweak interactions

The same gauge formalism can be applied to the SU(2) group, requiring Lagrangian invariance under SU(2) local phase transformation [36]:

$$\psi \to \psi' = \exp\left(ig_W \vec{\alpha}(x) \cdot \vec{T}\right).$$

Here, the $T_i = \frac{1}{2}\,\sigma_i$ with $i \in [1, 2, 3]$ are the three Pauli matrices corresponding to the three generators of SU(2), g_W is the weak coupling constant and $\vec{\alpha}(x)$ are three functions that specify the local phase at each point in the space-time. Since Pauli matrices do not commute, meaning $[\sigma_i, \sigma_j] = 2i\epsilon^{ijk}\sigma_k$, the SU(2) group is termed non-Abelian. The covariant derivative then becomes:

$$\partial_\mu \to D_\mu = \partial_\mu + ig_W \vec{T} \cdot \vec{W}_\mu,$$

where $\vec{W}(x)$ are the three new gauge fields associated to three gauge bosons $W^{(1)}$, $W^{(2)}$ and $W^{(3)}$. Due to the experimental observation of parity violation in weak interactions,

the four-vector current formulation instead differs from QED, as it requires to be written as a linear combination of vector and axial vector currents:

$$j^{\mu} = \overline{u}(p')\gamma^{\mu}u(p) + \overline{u}(p')\gamma^{\mu}\gamma^{5}u(p) = g_V j_V^{\mu} + g_A j_A^{\mu},$$

where g_V and g_A are the vector and the axial vector coupling constants. The relative strength of the parity violation is then given by:

$$\frac{g_V g_A}{g_V^2 + g_A^2}.$$

If either g_V or g_A is zero, parity is conserved in the interaction, while parity violation is maximal for pure vector minus/plus axial vector (V-A/V+A) interactions, meaning $|g_V| = |g_A|$. From experiments [36], it was observed that the weak charged current due to the exchange of a $\mathrm{W}^{\pm}$ bosons takes the form of a V-A interaction. Through the left- and right-handed projection operators:

$$P_R = \frac{1}{2}(1+\gamma^5), \qquad P_L = \frac{1}{2}(1-\gamma^5),$$

any Dirac spinor can be decomposed into left-handed and right-handed components and, as a consequence of the V-A interaction, only left-handed (LH) particle states or right-handed (RH) antiparticle states are allowed to participate in the charged-current weak interaction. The current term resulting for RH particle states or LH antiparticle states is indeed zero: for this reason, the symmetry group of the weak interaction is referred to as $\mathrm{SU(2)_L}$. Given that the generators of the SU(2) gauge transformation are the 2×2 Pauli spin-matrices, ψ is written in terms of two components and it is termed a weak isospin doublet, in analogy with the definition of isospin. In order to mantain the chirality structure of the weak interaction, a weak isospin doublet $\psi_L(x)$ must only contain LH particles and RH antiparticles, differing by one unit of electric charge, like for example:

$$\psi_L(x) = \begin{pmatrix} \nu_{\mathrm{e}} \\ \mathrm{e}^- \end{pmatrix}_L.$$

The total weak isospin I_W is 1/2, while its third component is 1/2 for the neutrino ν_{e} and -1/2 for the electron e^-. RH particle and LH antiparticle states are placed in weak isospin singlets with $I_W = 0$, and they are therefore unaffected by the $\mathrm{SU(2)_L}$ local gauge transformation. The new interaction term arising from the $\mathrm{SU(2)_L}$ local gauge invariance is the following:

$$i g_W T_k \gamma^{\mu} W_{\mu}^k \psi_L = \frac{i}{2} g_W \sigma_k \gamma^{\mu} W_{\mu}^k \psi_L,$$

originating three weak currents, one for each of the Pauli spin-matrices:

$$j_{\mu}^1 = \frac{g_W}{2}\overline{\psi}_L \gamma^{\mu}\sigma_1 \psi_L, \qquad j_{\mu}^2 = \frac{g_W}{2}\overline{\psi}_L \gamma^{\mu}\sigma_2 \psi_L, \qquad j_{\mu}^3 = \frac{g_W}{2}\overline{\psi}_L \gamma^{\mu}\sigma_3 \psi_L.$$

The four-vector currents corresponding to the exchange of the physical $\mathrm{W}^{\pm}$ bosons can be hence written as:

$$j_{\mu}^{\pm} = \frac{1}{\sqrt{2}}(j_{\mu}^1 \pm i j_{\mu}^2) = \frac{g_W}{2\sqrt{2}}\overline{\psi}_L \gamma^{\mu}(\sigma_1 \pm i\sigma_2)\psi_L = \frac{g_W}{\sqrt{2}}\overline{\psi}_L \gamma^{\mu}\sigma_{\pm}\psi_L,$$

and the physical W bosons are defined as:

$$\mathrm{W}_{\mu}^{\pm} = \frac{1}{\sqrt{2}}\left(\mathrm{W}_{\mu}^1 + i\mathrm{W}_{\mu}^2\right).$$

Therefore, the Lagrangian of weak charged interaction becomes:

$$\mathcal{L}_{CC} = -\frac{g_W}{2\sqrt{2}} \sum_i \overline{\psi}_i \gamma^\mu \left(1 - \gamma^5\right) \left(T^+ W_\mu^+ + T^- W_\mu^-\right) \psi_i$$

where the sum is over the three fermion generations, ψ_i is a $\mathrm{SU(2)_L}$ doublet and $T^\pm = \frac{1}{2}\left(T^1 \pm T^2\right)$ are the projector operators. It was mentioned in the previous section that the first description of the weak interaction was provided by Fermi, in terms of a contact interaction. Comparing the matrix element derived from the Fermi Hamiltonian to that obtained from the formulation of the electroweak interaction developed after the discovery of parity violation, it is possible to relate g_W with the Fermi constant G_F, that first expressed the strength of the weak coupling:

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

From the observed decay rates of muons and tau leptons it is found that the strength of the weak interaction is the same for all lepton flavors, namely $G_F^{(\mathrm{e})} = G_F^{(\mu)} = G_F^{(\tau)}$. The leptons' property of coupling to the gauge boson independently from their flavor is called *lepton universality*. On the other hand, the strength of the weak interaction for quarks, which can be determined from the study of nuclear β-decay, is not the same for all quark flavors. Such differences arise considering that the weak eigenstates of quarks do not coincide with the mass eigenstates, but are related to them by the unitary Cabibbo-Kobayashi-Maskawa (CKM) matrix V_{CKM} [37, 38]:

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

The relative strength of the interaction is therefore defined by the relevant element of the CKM matrix. The CKM matrix has four independent parameters, three rotation angles and a complex phase:

$$V_{CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix},$$

where c_{ij} stands for $\cos\theta_{ij}$, and s_{ij} for $\sin\theta_{ij}$. The values of these parameters are not predicted by SM, and they have to be evaluated from experiments [39].

1.3.1 The neutral current and the electroweak unification

The $\mathrm{SU(2)_L}$ symmetry of the weak interaction implies the existence of a third current j_μ^3, corresponding to the neutral vertices:

$$j_\mu^3 = \frac{g_W}{2} \overline{\psi}_L \gamma^\mu \sigma_3 \psi_L. \tag{1.1}$$

It would be hence possible to identify the boson for the neutral current as W^3. If that was the case, such boson would only couple to LH particles (and RH antiparticles), but it has been observed from experiments that also RH particles (and LH antiparticles) can interact via neutral current. In the 1960s, Glashow [40], Salam [41] and Weinberg [42] (GSW) proposed a model, based on the Yang-Mills theory, capable of unifying the electromagnetism and the weak interaction into a single electroweak interaction. In the GSW

model, the U(1) gauge symmetry of electromagnetism is replaced with a new $\mathrm{U(1)_Y}$ local gauge symmetry, such that:

$$\psi(x) \to \psi'(x) = \exp\left(ig'\frac{Y}{2}\zeta(x)\right)\psi(x).$$

In order to preserve local gauge invariance, a new interaction term has to be introduced, together with a new gauge field B_μ that couples to a new kind of charge, termed the weak hypercharge Y:

$$g'\frac{Y}{2}B_\mu\gamma^\mu\psi.$$

With this formalism, the photon and the boson associated to the weak neutral current (called Z boson) can be written as a linear combinations of B_μ and W^3:

$$A_\mu = B_\mu\cos\theta_W + W^3_\mu\sin\theta_W, \tag{1.2}$$

$$Z_\mu = -B_\mu\sin\theta_W + W^3_\mu\cos\theta_W, \tag{1.3}$$

where θ_W is termed the weak mixing angle. Therefore, the physical currents of QED and the weak neutral current get the following form:

$$j^{em}_\mu = j^Y_\mu\cos\theta_W + j^3_\mu\sin\theta_W,$$

$$j^Z_\mu = -j^Y_\mu\sin\theta_W + j^3_\mu\cos\theta_W.$$

The Lagrangian of weak, neutral current hence becomes:

$$\mathcal{L}_{NC} = -\frac{g_W}{2\cos\theta_W}\sum_i \overline{\psi_i}\gamma^\mu(g^i_V - g^i_A\gamma^5)\psi_i Z^\mu,$$

where ψ_i can be RH or LH fermionic singlets, and g^i_V and g^i_A are expressed as

$$g^i_V = T^i_3 - 2Q^i\sin^2\theta_W \qquad \text{and} \qquad g^i_A = T^i_3.$$

The GSW model of electroweak unification implies that the couplings of the weak and the electromagnetic interactions are related through the mixing angle θ_W, as

$$e = g_W\sin\theta_W = g'\cos\theta_W,$$

while the definition of hypercharge can be obtained by imposing the invariance under $\mathrm{U(1)_Y}$ and $\mathrm{SU(2)_L}$ local gauge transformations:

$$Y = 2(Q - I^3_W).$$

The electroweak Lagrangian can be then written in a compact way as:

$$\mathcal{L}_{EW} = \overline{\chi}_L\gamma^\mu\left(i\partial_\mu - g_W\vec{T}\cdot\vec{W} - g'\frac{Y}{2}B_\mu\right)\chi_L + \overline{\psi}_R\left(i\partial_\mu - g'\frac{Y}{2}B_\mu\right)\psi_R + \mathcal{L}_{KIN},$$

where χ_L and ψ_R are the fermionic left-handed and right-handed states, respectively [43], and $\mathcal{L}_{KIN}$ describes the gauge field kinetics:

$$\mathcal{L}_{KIN} = -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}\vec{W}_{\mu\nu}\vec{W}^{\mu\nu},$$

where

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \qquad \text{and} \qquad \vec{W}_{\mu\nu} = \partial_\mu\vec{W}_\nu - \partial_\nu\vec{W}_\mu + g_W\vec{W}_\mu\times\vec{W}_\nu.$$

1.3.2 Triple and quartic gauge couplings

The last term in $\vec{W}_{\mu\nu}$ is due to the non-Abelian structure of $\mathrm{SU(2)_L}$. As a consequence of the presence of such extra-term, the kinetic term of the electroweak Lagrangian $\mathcal{L}_{KIN}$ gives rise to triple and quartic self-interactions among the gauge fields, whose strength is given by the same $\mathrm{SU(2)_L}$ constant coupling g_W. After transforming the gauge fields into the physical fields (see Eq. 1.2 and 1.3), the explicit Lagrangian for the triple gauge couplings becomes [43]:

$$\begin{aligned}\mathcal{L}_3 = &- ie\cot\theta_W\{(\partial^\mu W^\nu - \partial^\mu W^\nu)W^\dagger_\mu Z_\nu - (\partial^\mu W^{\nu\dagger} - \partial^\mu W^{\nu\dagger})W_\mu W_\nu + W_\mu W^\dagger_\nu(\partial^\mu Z^\nu - \partial^\mu Z^\nu)\}\\ &- ie\{(\partial^\mu W^\nu - \partial^\mu W^\nu)W^\dagger_\mu A_\nu - (\partial^\mu W^{\nu\dagger} - \partial^\mu W^{\nu^\dagger})W_\mu A_\nu + W_\mu W^\dagger_\nu(\partial^\mu A^\nu - \partial^\mu A^\nu)\},\end{aligned}$$

while for the quartic gauge couplings [43]:

$$\begin{aligned}\mathcal{L}_4 = &- \frac{e^2}{2\sin^2\theta_W}\{(W^\dagger_\mu W^\mu)^2 - W^\dagger_\mu W^{\mu\dagger} W_\nu W^\nu\} - e^2\cot^2\theta_W\{W^\dagger_\mu W^\mu Z_\nu Z^\nu - W^\dagger_\mu Z^\mu W_\nu Z^\nu\}\\ &- e^2\cot\theta_W\{2W^\dagger_\mu W^\mu Z_\nu A^\nu - W_\mu Z^\mu W_\nu A^\nu - W_\mu A^\mu W_\nu Z^\nu\}\\ &- e^2\{W^\dagger_\mu W^\mu A_\nu A^\nu - W^\dagger_\mu A^\mu W_\nu A^\nu\}.\end{aligned}$$

In each term of the Lagrangian there are always at least a pair of charged W bosons. In order to preserve the gauge invariance, the $\mathrm{SU(2)_L}$ algebra does indeed not generate any neutral couplings involving only photons or Z bosons. However, they can arise in the context of the theory of anomalous gauge couplings, as it will be discussed in the next chapter.

1.4 Strong interactions

Quantum Chromodynamics (QCD) is the quantum field theory of strong interaction. Similarly to QED, whose interaction is mediated by a massless photon, the generator of U(1) symmetry, QCD is mediated by eight massless gluons that correspond to the eight generators of the SU(3) group. As for QED and weak interactions, Lagrangian must be invariant under SU(3) local phase transformation, meaning that:

$$\psi(x) \to \psi'(x) = \exp\left[ig_S\vec{\beta}(x)\cdot\vec{\hat{T}}\right]\psi(x),$$

where $\vec{\hat{T}} = \{T^a\}$ represents the eight generators which are related to the Gellmann matrices λ^a as $T^a = \frac{1}{2}\lambda^a$, while g_S is the coupling constant. In addition, to preserve the SU(3) local gauge symmetry, the derivatives have to be replaced with the covariant derivative:

$$D^{ij}_\mu = \partial_\mu\delta^{ij} + ig_S(T^a)^{ij}G^a_\mu,$$

where the G^a_μ are the new fields corresponding to the eight gluons with color index $a \in [1, ..., 8]$. The SU(3) group is non-Abelian as its generators λ do not commute (meaning that $[\lambda_i, \lambda_j] = 2if_{ijk}\lambda_k$). This feature gives rise to gluon-self interactions and *asymptotic freedom*, which will be discussed below. Being the SU(3) generators 3×3 matrices, the wavefunction ψ must be written including three additional degrees of freedom, that can be represented by a three component vector. This new degree of freedom is called *color*, with *red*, *blue* and *green* labelling the three color wavefunctions

$$r = \begin{pmatrix}1\\0\\0\end{pmatrix}, \qquad g = \begin{pmatrix}0\\1\\0\end{pmatrix}, \qquad b = \begin{pmatrix}0\\0\\1\end{pmatrix}.$$

The single charge of QED is then replaced by three color charges, r, g and b. Color must be conserved at the interaction vertex, and so gluons must carry both color and anticolor charge. Only particles that have non-zero color charge can couple to gluons: leptons, for example, do not feel the strong force because they are color neutral. The current term for QCD is written in analogy with the QED:

$$j^{\mu} = \overline{\psi}_j \left(\frac{1}{2} i g_S \lambda^a \gamma^{\mu} \right) \psi_i,$$

where ψ_i represents a SU(3) triplet and j is the color index. Therefore, the Lagrangian for QCD becomes:

$$\mathcal{L}_{QCD} = \psi_i \left((i\gamma^{\mu} D_{\mu})_{ij} - m\delta_{ij} \right) \psi_j - \frac{1}{4} \widetilde{F}^a_{\mu\nu} \widetilde{F}^{\mu\nu}_a,$$

where the last term is the kinetic term, with $\widetilde{F}^a_{\mu\nu} = \partial_{\mu} G_{\nu} - \partial_{\nu} G_{\mu} + g_S f^{abc} G^b_{\mu} G^c_{\nu}$.

1.4.1 Colour confinement and hadronization

One of the issues that soon arose after the quark model was first proposed by Gell-Mann, in the 1960s, was the lack of a direct observation of such particles. Quarks, differently from leptons, indeed cannot be seen directly. Their non-observation is now explained by the hypothesis of *color confinement*, which states that objects having color charge cannot propagate as free particles. In spite of its importance in the understanding of QCD processes, the mechanism of color confinement has not been analytically proved yet, but it is believed to originate from gluon-gluon self-interactions. A qualitative understanding of the process can be obtained by considering two free quarks that being apart. The mutual exchange of virtual gluons between them squeezes the color field into a tube (see Fig. 1.3) and, as a consequence, the energy stored in the field becomes proportional to the distance between the quarks. It would require an infinity amount of energy to separate the quarks to infinity, meaning that quarks can only exist in colorless hadrons. Furthermore, since also gluons carry color charge they can not propagate over macroscopic distances, and they are confined to colorless object. In particle collisions, as it will be described later, quarks and gluons are observed as *jet* of colorless particles. A qualitative description of the hadronization process is shown in Fig. 1.4. The quark and the antiquark produced in an interaction separate at high velocities, restricting the color field in a tube, consequently to color confinement; the energy stored in the field increases as the quarks separate further, until it becomes sufficient to produce a new $q\bar{q}$ pair, breaking the color field in smaller tubes. The process continues until the energy of quarks is enough to combine and form colorless hadrons. Two jets result from the initial $q\bar{q}$ pair, following the initial directions of the quark and the antiquark respectively.

Figure 1.3: Qualitative picture of the color field between a $q\bar{q}$ pair [36].

1.4.2 Asymptotic freedom and the running of α_S

The strength of an interaction is usually espressed in terms of a *coupling constant* α that varies as a function of the energy scale or transfered momentum Q. Such variation is called the *running* of α, and it formally originates from the *renormalization* procedure. In the case of QED, the magnitude of the coupling decreases at larger distances while

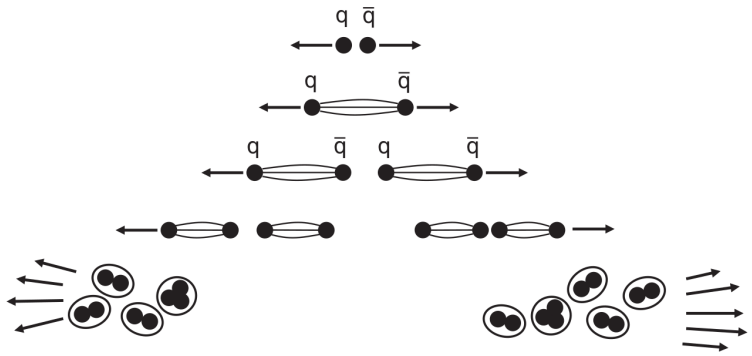

Figure 1.4: A qualitative description of the hadronization process [36].

it gets stronger at small distances, tending to its macroscopic value of $\alpha(0) \simeq 1/137$. Such behavior of the coupling constant is to be researched in the expression of the electromagnetic potential created by a test charge q_c. Due to the vacuum polarization, pairs of particle-antiparticle are created around q_c, with particles of charge opposite to the test charge tending to be closer to it, analogously to the electric charge screening in a dielectric medium. By the Gauss' law, the total charge inside a sphere of radius r will be smaller than q_c, and the larger r the more the total charge will tend to the test charge value. The potential determined by the screening is hence:

$$V(r) \propto q_C \frac{e^{-r}}{r}.$$

For the QCD the situation is different because gluons carry color charge. Since the $SU(3)_C$ group is non-Abelian, the Gauss' law has to include an additional term containing f_{abc} (with a, b and c being the color indices). This new term determines the anti-screening, that enhances the field at large distances and hence leads the QCD coupling constant to decrease with small distance and to increase at large distance. The strong coupling constant is usually rewritten as:

$$\alpha_S = \frac{g_S}{4\pi}.$$

The energy dependence of the strong coupling is controlled by the so-called β-function [44], defined as follows:

$$Q^2 \frac{\partial \alpha_S}{\partial Q^2} = \beta(\alpha_S) = -\left(\frac{\alpha_s}{4\pi}\right)^2 \sum_{n=0} \left(\frac{\alpha_s}{4\pi}\right)^n \beta^n,$$

where the β_n depends on the quark and gluon loops and on the renormalization scheme used. The exact analytical solution for α_S is known only to β_0 order, while at higher orders numerical methods need to be used. At β_0 order:

$$\alpha_S(Q^2) = \frac{\alpha_S(\mu_R^2)}{1 + \beta_0 \ln(Q^2/\mu_R^2)\alpha_S(\mu_R^2)}.$$

This equation relates the strength of the coupling at a scale Q to the one at scale μ_R (introduced to remove the divergences that arise in QCD calculations), assuming both scales to be in the perturbative regime. The value of α_S is currently extrapolated at the value of the Z mass, at which it is sufficiently small for the perturbation theory to be applied, and its value is $\alpha_S(m_Z^2) = 0.1181 \pm 0.0011$ [45]. When the coupling approaches unity, the perturbation theory is not valid anymore and the QCD scale parameter Λ_{QCD} is introduced as the scale at which $\alpha_S(Q^2)$ diverges [44]:

$$\alpha_S(Q^2) = \frac{4\pi}{\beta_0 \ln(Q^2/\Lambda_{QCD})}.$$

At value of $\Lambda_{QCD} \sim 200$ MeV (the *Landau pole*), the strong coupling becomes infinite, perturbation theory can not be applied anymore and quarks and gluons combine into hadrons.

1.4.3 Electroweak precision physics

The weak mixing angle θ_W is one of the key parameters in the electroweak sector of the SM, with its measurement providing one of the most stringent test of the consistency of the SM itself. Up to now, the most precise value of θ_W has been obtained by combining the results from the electron-positron Stanford Linear Collider (SLC) and the Large Electron-Positron (LEP) collider, while the value which is currently used includes also the measurements performed at the LHC, resulting in [46]:

$$\sin^2 \theta_W = 0.23150 \pm 0.00006$$

The conceptually simplest and most used method to measure the weak mixing angle exploits the backward-forward asymmetry in the $e^+e^- \to Z \to \ell\bar{\ell}$ production. The differential cross section for the process can be written as:

$$\frac{d\sigma}{d\cos\theta^*} \propto a(1 + \cos^2\theta^* + A_4 \cos\theta^*), \tag{1.4}$$

where θ^* is the angle of the negative lepton in the Collins-Soper frame [47]. The $((1 + \cos^2\theta^*)$ term arises from the spin-1 of the exchanged boson, while the $\cos\theta^*$ term originates from interference between vector and axial-vector contributions. The forward-backward asymmetry A_{FB} is related to Eq. 1.4 as follows:

$$A_{FB} = \frac{\sigma_F - \sigma_B}{\sigma_F + \sigma_B} = \frac{3}{8} A_4,$$

where σ_F and σ_B are the cross section for the negative lepton produced in the forward ($\theta^* < \pi/2$) and backward ($\theta^* > \pi/2$) direction. The $\cos\theta^*$ can be calculated using the following variables in the laboratory frame:

$$\cos\theta^* = \frac{2(P_1^+ P_2^- - P_1^- P_2^+)}{\sqrt{m_{\ell\ell}^2 (m_{\ell\ell}^2 + p_{T,\ell\ell}^2)}} \cdot \frac{p_{z,\ell\ell}}{|p_{z,\ell\ell}|},$$

where $m_{\ell\ell}$, $p_{T,\ell\ell}$, and $p_{z,\ell\ell}$ are the mass, transverse momentum and longitudinal momentum, respectively, of the dilepton system, while the $P_i^\pm$ are defined in terms of the energies (E_i) and longitudinal momenta ($p_{z,i}$) of the negatively and positively charged leptons as $P_i^\pm = (E_i \pm p_{z,i})/\sqrt{2}$. The weak mixing angle is finally obtained by fitting the mass and rapidity distribution of A_{FB} as a function of θ_W^2 [48]. Besides the measurement of θ_W, the study of the Z boson provides a number of significative tests of the SM, such as the precise measurements of the Z mass and width. Among them, a remarkable results has been recently achieved by the CMS collaboration, that has obtained the first precise measurement of the invisible width of the Z boson at a hadron collider, using proton-proton collisions at $\sqrt{s} = 13$ TeV (with an integrated luminosity of 36.3 fb^1):

$$\Gamma_{\text{inv}} = 523 \pm 3 \text{ (stat)} \pm 16 \text{ (syst) MeV}$$

This is also the single most precise direct measurement up to date, competitive with the combined result of the direct measurements from the LEP experiments [49]. The indirect method, which uses the total Z boson width extracted from the Z boson lineshape, is however the most precise, with a combined indirect measurement of 499.0 ± 1.5 MeV [45]

obtained from the LEP experiments. Further tests of the electroweak sector can be obtained from the study of the W boson: the mass, width and branching ratios have been also measured at LEP with excellent precision, and the same measurements are also currently ongoing at the LHC. As it is shown in Fig. 1.5, the global fit to the latest electroweak data impressively demonstrates the predictive power of the SM in the electroweak sector. Of particular importance is the measurement of the W mass. The presence of a neutrino

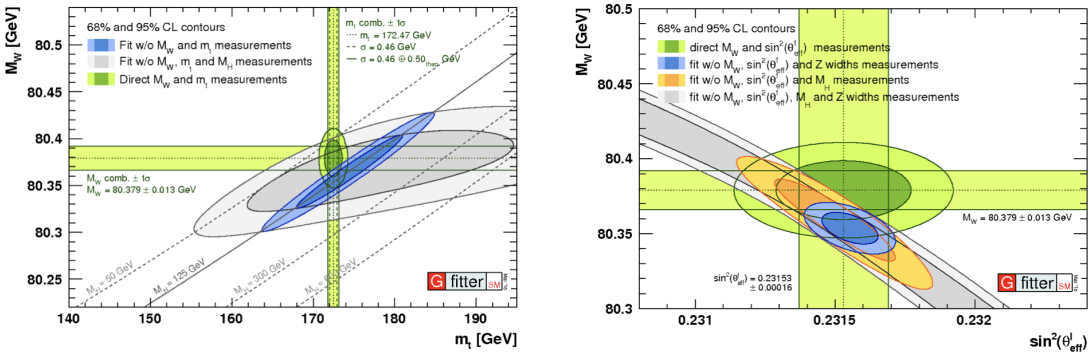

Figure 1.5: (Left) Contours of 68% and 95% confidence level obtained from scans of fits with fixed variable pairs of W mass M_{W} and top mass m_{t}. The narrower blue and larger grey allowed regions are the results of the fit including and excluding the M_{H} measurement, respectively. The horizontal bands indicate the 1σ regions of the M_{W} and m_{t} measurements. (Right) Contours of 68% and 95% confidence level obtained from scans of fits with fixed variable pairs of M_{W} and the effective weak mixing angle $\sin^2\theta_{eff}$. The narrower blue and larger grey allowed regions are the results of the fit including and excluding the M_{H} measurements, respectively. The horizontal bands indicate the 1σ regions of the M_{W} and $\sin^2\theta_{eff}$ measurements (world averages) [50].

in the decay products of the boson means that the W mass cannot be reconstructed by exploting the shape of its resonance, as it is the case of the Z boson. Therefore, different techniques have to be adopted, such as the extraction of the mass boson from the shape of the $\mathrm{e^-e^+} \to \mathrm{WW}$ cross section close to the threshold $\sqrt{s} = 2m_{\mathrm{W}}$. The results obtained by the LEP experiments and ATLAS (using LHC Run 1 data) are compatible with the SM prediction for the W mass. However, a surprising result has instead been obtained in 2022 by the Collider Detector at Fermilab (CDF) Collaboration at Tevatron [51]. The CDF measurement of m_{W} has indeed provided a value which is 7 standard deviation away from the SM prediction (Fig. 1.6), strongly suggesting the possibility of improvements of the SM precision calculation, or SM extensions with new sectors.

1.5 The Higgs boson

1.5.1 The Spontaneous Symmetry Breaking

The required local gauge invariance of SM is broken by the terms in the Lagrangian that arise from the boson masses. Consider, as an example, the simple case of a massive photon mediating the electromagnetic interaction: in order to account for its mass, the QED Lagrangian should contain the additional term: $\frac{1}{2}m_\gamma^2 A_\mu A^\mu$. However, the photon field under a U(1) local gauge transformation would therefore transform as:

$$A_\mu \to A'_\mu = A_\mu - \partial_\mu \alpha,$$

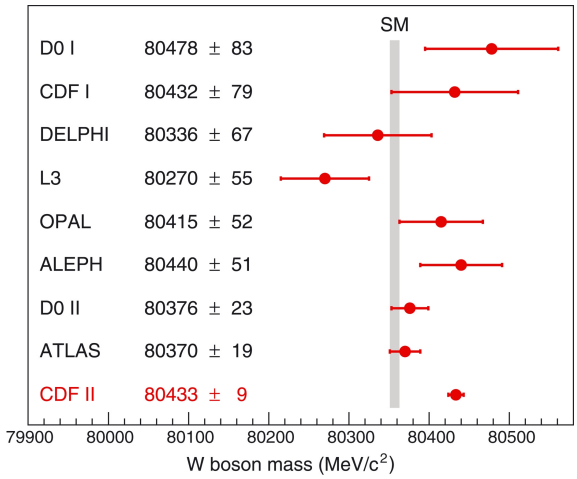

Figure 1.6: Comparison of the CDF II measurement and past m_{W} measurements with the SM expectation [51].

with the mass term becoming:

$$\frac{1}{2}m_\gamma^2 A_\mu A^\mu \rightarrow \frac{1}{2}m_\gamma^2(A_\mu - \partial_\mu\alpha)(A^\mu - \partial^\mu\alpha) \neq \frac{1}{2}m_\gamma^2 A_\mu A^\mu,$$

and breaking the local gauge U(1) invariance. This is not a problem for QED and QCD, where the gauge bosons are massless, but it is conflicting with the observation of the large masses of the W and Z bosons [52, 53]. The same problem has to be faced when considering the fermion masses: in the gauge transformation of the weak interaction, the LH particles transform as weak isospin doublets and the RH particles as singlets, implying that the fermion mass term in the SM Lagrangian should be written as:

$$m_f(\overline{\psi}_L\psi_R + \overline{\psi}_R\psi_L),$$

breaking the required $\mathrm{SU(2)_L}$ gauge invariance. In order to preserve the correct properties of the SM Lagrangian, in the 1960s Higgs, Englert and Brout [54, 55] introduced the idea of the *spontaneous symmetry breaking* (SSB), a mechanism that brings in the Lagrangian the mass term for a scalar field as a consequence of a broken symmetry. As an example, consider a complex scalar field:

$$\phi = \frac{1}{2}(\phi_1 + i\phi_2).$$

The associated Lagrangian is invariant under a U(1) global symmetry transformation:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)^*(\partial^\mu\phi) - V(\phi) \quad \text{with} \quad V(\phi) = \mu^2(\phi^*\phi) + \lambda(\phi^*\phi)^2. \tag{1.5}$$

The potential has a finite minimum if $\lambda > 0$, and its shape depends on the sign of μ^2, as shown in Figure 1.7. If $\mu^2 > 0$, the potential minimum occurs when both ϕ_1 and ϕ_2 are zero; if $\mu^2 < 0$, the potential has an infinite set of minima defined by:

$$\phi^\dagger\phi = \frac{1}{2}(\phi_1^2 + \phi_2^2) = \frac{-\mu^2}{\lambda} = v^2. \tag{1.6}$$

Without loss of generality, the physical vacuum state can be chosen to be any minimum of such set, hence breaking the global symmetry of the Lagrangian. Choosing the vacuum state in the real direction, $(\phi_1, \phi_2) = (v, 0)$ and expanding the field around it by writing $\phi_1(x) = \eta(x) + v$ and $\phi_2(x) = \xi(x)$, one obtains:

$$\phi = \frac{1}{2}(\eta + v + i\xi).$$

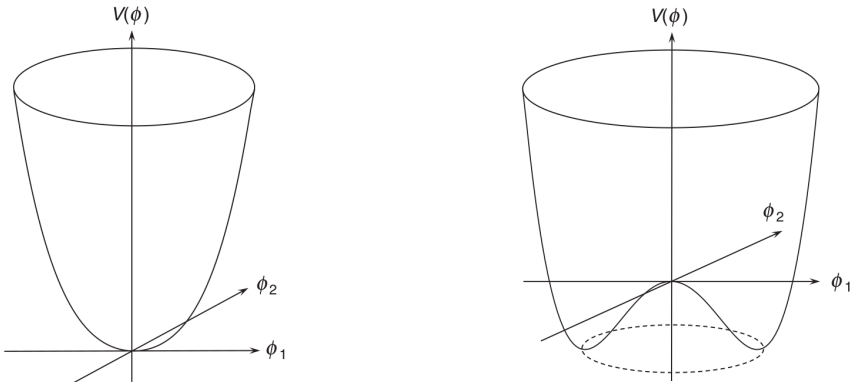

Figure 1.7: The $V(\phi)$ potential for (left) $\mu^2 > 0$ and (right) $\mu^2 < 0$.

The Lagrangian can be therefore written in term of the fields η and ξ, exploiting the relation $\mu^2 = -\lambda v^2$ (Eq. 1.7). It represents a scalar field η with mass $m_\eta = \sqrt{2\lambda v^2}$ and a massless scalar field, whose associated massless scalar particle is known as a Goldstone boson [56]. The term which is quadratic in the field η can be identified as the mass term, while the terms with three or four powers of the fields are the interaction terms V_{int}:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \eta)(\partial^\mu \eta) - \frac{1}{2} m_\eta^2 \eta^2 + \frac{1}{2}(\partial_\mu \xi)(\partial^\mu \xi) - V_{int}(\eta, \xi). \tag{1.7}$$

1.5.2 The Brout-Englert-Higgs mechanism

In order to include the W^+, W^- and Z masses in the SM Lagrangian, the SSB mechanism is integrated with the local $SU(2)_L \times U(1)_Y$ gauge symmetry, giving rise to the so-called Brout-Englert-Higgs (BEH) mechanism. To formalize such mechanism, it is possible to starting again with the simpler U(1) case. Consider first the Lagrangian for a complex scalar field: the Lagrangian is not invariant under U(1) local gauge transformation $\phi \to \phi'(x) = \exp(ig\alpha(x)) \cdot \phi(x)$, and the derivatives have hence to be replaced by the covariant derivatives $\partial_\mu \to D_\mu = \partial_\mu + igB_\mu$, where the new gauge field B_μ transforms as $B_\mu \to B'_\mu = B_\mu - \partial_\mu \alpha(x)$. In this way, the combined Lagrangian for the complex scalar field ϕ, the gauge field B and the potential $V(\phi) = \mu^2\phi^2 - \lambda\phi^4$ can be written as:

$$\mathcal{L} = -\frac{1}{4} F^{\mu\nu} F^{\mu\nu} + (D_\mu \phi)^* (D_\mu \phi) - \mu^2 \phi^2 - \lambda \phi^4.$$

As before, the choice of the physical vacuum state spontaneously breaks the symmetry: choosing $\phi_1 + i\phi_2 = v$, and expanding around such vacuum state, the following Lagrangian is obtained:

$$\begin{aligned}\mathcal{L} = \frac{1}{2}(\partial_\mu \eta)(\partial^\mu \eta) - \lambda v^2 \eta^2 + \frac{1}{2}(\partial_\mu \xi)(\partial^\mu \xi) + \\ - \frac{1}{4} F^{\mu\nu} F^{\mu\nu} + \frac{1}{2} g^2 v^2 B_\mu B^\mu - V_{int}(\eta, \xi) + gvB_\mu(\partial^\mu \xi).\end{aligned}$$

The SSB produces again a massive scalar field η and a massless Goldstone boson ξ but, in addition, the requirement of local gauge invariance makes the gauge field B acquiring a mass term equal to $\frac{1}{2} g^2 v^2 B_\mu B^\mu$. The last term in the Lagrangian, corresponding to a direct coupling between the Goldstone field ξ and the gauge field B, can be eliminated by making the following gauge transformation:

$$B_{\mu(x)} \to B'_\mu(x) = B_\mu(x) + \frac{1}{gv} \partial_\mu \xi(x),$$

which corresponds to choose $\alpha(x) = -\xi(x)/gv$, such that $\phi(x) \to \phi'(x) = \exp(-i\frac{\xi(x)}{v}) \cdot \phi(x)$. As a consequence, the complex scalar field $\phi(x)$ after the expansion about the vacuum state appears to be entirely real:

$$\phi(x) = \frac{1}{\sqrt{2}}(v + \eta(x)) \equiv \frac{1}{\sqrt{2}}(v + h(x)). \tag{1.8}$$

Here, the field $\eta(x)$ has been written as the "Higgs field" *h(x)* to emphasize that this is the physical field in this gauge (called the *Unitary gauge*). After cancelling the Goldstone field, the Lagrangian can be rewritten (ignoring constant terms) as:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu h)(\partial^\mu h) - \lambda v^2 h^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}g^2v^2B_\mu B^\mu + \\ + g^2 v B_\mu B^\mu h + \frac{1}{2}g^2 B_\mu B^\mu h^2 - \lambda v h^3 - \frac{1}{4}\lambda h^4,$$

describing a massive scalar Higgs field h and a massive gauge boson B associated with the U(1) local gauge symmetry. It contains interaction terms between the Higgs boson and the gauge boson, and Higgs boson self-interaction terms. The mass of the gauge boson is $m_B = g \cdot v$, while the mass of the Higgs boson is $m_{\mathrm{H}} = \sqrt{2\lambda}v$.

In the case of a $\mathrm{SU(2)_L \times U(1)_Y}$ gauge symmetry, the simplest Higgs model consists of two complex scalar fields which are placed in a weak isospin doublet. In order to generate the masses of the electroweak bosons, the scalar fields ϕ^0 must be neutral while the others must be charged such that ϕ^+ and $(\phi^+)^* = \phi^-$ give the longitudinal degrees of freedom of the $\mathrm{W^+}$ and $\mathrm{W^-}$:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}.$$

The Lagrangian for the Higgs field and the associated Higgs potential have the same form of Eq. 1.5. Expanding the fields around the chosen minimum and writing the doublet in a unitary gauge, one obtains:

$$\phi(x) = \frac{1}{2} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}.$$

In order to respect the $\mathrm{SU(2)_L \times U(1)_Y}$ local gauge symmetry of the electroweak model, the derivatives have to be replaced with the appropriate covariant derivatives:

$$\partial_\mu \to D_\mu = \partial\mu + ig_W \vec{T} \cdot \vec{W}_\mu + ig' \frac{Y}{2} B_\mu,$$

and the Higgs doublet ϕ then becomes:

$$D_\mu \phi = \frac{1}{2}\left[2\partial_\mu + \left(ig_W \vec{\sigma} \cdot \vec{W}_\mu + ig' B_\mu\right)\right]\phi.$$

The gauge bosons masses are determined by the terms in $(D_\mu\phi)^\dagger(D^\mu\phi)$ that are quadratic in the gauge boson fields (D_μ is a 2×2 matrix, and the identity matrix multiplying ∂_μ and B_μ terms is implicit in the expression):

$$(D_\mu\phi)^\dagger(D^\mu\phi) = \frac{1}{2}(\partial_\mu h)(\partial^\mu h) + \frac{1}{8}g_W^2(W_{1\mu} + iW_{2\mu})(W^{1\mu} + iW^{2\mu})(v+h)^2 + \\ + \frac{1}{8}(g_W W_\mu^3 - g' B_\mu)(g_W W^{3\mu} - g' B^\mu)(v+h)^2.$$

The mass terms for the W^1 and W^2 fields are $\frac{1}{2}m_W^2 W_\mu^1 W^{1\mu}$ and $\frac{1}{2}m_W^2 W_\mu^2 W^{2\mu}$, and hence the mass of the W boson is $m_W = \frac{1}{2}g_W v$. The mass terms for the W^3 and B fields instead can be written as:

$$\frac{v^2}{8}\begin{pmatrix} W_\mu^3 & B_\mu \end{pmatrix}\begin{pmatrix} g_W^2 & -g_W g' \\ -g_w g' & g'^2 \end{pmatrix}\begin{pmatrix} W^{3\mu} \\ B^\mu \end{pmatrix} = \frac{v^2}{8}\begin{pmatrix} W_\mu^3 & B_\mu \end{pmatrix}\mathbf{M}\begin{pmatrix} W^{3\mu} \\ B^\mu \end{pmatrix}$$

where $\mathbf{M}$ is the non-diagonal mass matrix that allows the mixing of the two fields. The physical boson fields (A_μ and Z_μ) correspond to the basis in which the mass matrix is diagonal, while their masses correspond to the eigenvalues of the mass matrix:

$$\frac{v^2}{8}\begin{pmatrix}A_\mu & Z_\mu\end{pmatrix}\begin{pmatrix}0 & 0\\ 0 & g_W^2+g'^2\end{pmatrix}\begin{pmatrix}A^\mu\\ Z^\mu\end{pmatrix}=\frac{1}{2}\begin{pmatrix}A_\mu & Z_\mu\end{pmatrix}\begin{pmatrix}m_A^2 & 0\\ 0 & m_Z^2\end{pmatrix}\begin{pmatrix}A^\mu\\ Z^\mu\end{pmatrix}$$

Therefore, the masses of the photon and the Z boson become, respectively:

$$\begin{aligned} m_A &= 0,\\ m_Z &= \frac{1}{2}v\sqrt{g_w^2+g'^2}. \end{aligned}$$

The mass of the Higgs boson, m_{H}, is obtained from the term which is quadratic in h, hence $m_{\mathrm{H}} = \sqrt{2\lambda v^2}$. The vacuum expectation value of the Higgs field is $v = 246$ GeV, while λ is a free parameter. This means that the value of the Higgs mass is not predicted by theory. The fermion masses are obtained from the fermions' coupling to the Higgs field, through the so-called Yukawa interaction. A LH doublet L combined with a scalar field ϕ and a RH singlet R, $\overline{L}\phi R$, is invariant under $SU(2)_L$ and $U(1)_Y$ gauge transformations, as it is its Hermitian conjugate $(\overline{L}\phi R)^\dagger = \overline{R}\phi^\dagger L$. The mass term for all the Dirac fermions can be then obtained from the Yukawa Lagrangian:

$$L_Y = -\sum_{i,j=1}^{3}\left[y_{ij}^u u_{Ri}\phi_C Q_{Lj} + y_{ij}^d \overline{d}_{Ri}\phi^\dagger Q_{Lj} + y_{ij}^{\mathrm{e}}\overline{\mathrm{e}}_{Ri}\phi^\dagger Q_{Lj}^{\mathrm{e}}\right] + \text{hermitian conjugate.} \quad (1.9)$$

The sum is performed over the three generations of quarks and leptons, ϕ_C is the hermitian conjugate of the Higgs doublet ϕ and Q_{Lj} and Q_{Lj}^{e} are the LH doublets for quarks and leptons, respectively. The values $y^{u,d,\mathrm{e}}$ correspond to the Yukawa coupling constants which are related with the fermion masses through the vacuum expectation value v.

1.5.3 Search for the Higgs boson

The Higgs boson was theorized for the first time in 1964, and the first attempts to verify its existence started in the early 1970s [57]. However, the LEP experiments have been the first ones to approach the energy necessary to the Higgs boson detection and to place a significant limit on its mass, which was contrained to be greater than 114 GeV [58]. The value of the Higgs boson mass is indeed not predicted by the SM theory. With the beginning of the LHC operations, the mass range was further restricted, and the discovery of the Higgs boson was finally announced by ATLAS and CMS on July 4th 2012. The next year, on October 8th 2013, the Nobel prize in physics was awarded jointly to Englert and Higgs "for the theoretical discovery of a mechanism that contributes to our understanding of the origin of mass of subatomic particles, and which recently was confirmed through the discovery of the predicted fundamental particle, by the ATLAS and CMS experiments at CERN's Large Hadron Collider" [59].

Higgs production modes

The Higgs boson couples preferentially to heavy particles (W, Z, top and bottom quarks and τ-leptons) and it can be produced in proton-proton collisions through a number of different processes [45]. The *gluon-gluon fusion* processes ($\mathrm{gg} \to \mathrm{H} + \mathrm{X}$) are the dominant processes at the LHC; the gluon coupling to the Higgs boson in the SM can be mediated by triangular top- and bottom-quark loops. The second important Higgs production channel is the *vector-boson fusion* ($qq \to qq\mathrm{H}$), whose cross section is about one order of magnitude smaller than the gluon-fusion one. The Higgs boson is here produced together with two

forward jets coming from the break-up of the colliding protons, without the QCD radiation that instead accompany the gluon-fusion process, producing more easily identifiable final states. Other relevant production modes are the *Higgs-strahlung* ($q\bar{q} \rightarrow$ WH or $q\bar{q} \rightarrow$ ZH), which was the dominant channel at the LEP, where an electron and a positron collided to form a virtual Z boson, and the $t\bar{t}$ associated production ($q\bar{q} \rightarrow t\bar{t}H$). These last two processes have a cross section from one to two orders of magnitude smaller than the gluon-fusion cross section. At the LHC, at a center of mass of about 8 TeV, the total production cross section for a Higgs boson with a mass of 125 GeV was approximately 20 pb^{-1} [45].

Higgs decay modes

The largest Higgs boson branching ratios are listed in Tab. 1.1. The fermionic decay ($H \rightarrow b\bar{b}$ and $H \rightarrow \tau^-\tau^+$) is the most promising channel for probing the coupling of the Higgs field to the quarks and leptons, but the presence of very large backgrounds makes the isolation of a Higgs boson signal in these channels quite challenging [45]. In the $H \rightarrow W^+W^-$ channel, the W boson can subsequently decay either into a quark and an antiquark or into a charged lepton and a corresponding neutrino. The decays into quarks are difficult to distinguish from the background, and the decays into leptons can not be fully reconstructed due to the presence of neutrinos. A cleaner signal is given by the $H \rightarrow \gamma\gamma$ and the $H \rightarrow ZZ$ decay, with each Z boson decaying into a lepton-antilepton pair.

Decay channel	Branching ratio
$H \rightarrow \gamma\gamma$	2.27×10^{-3}
$H \rightarrow ZZ$	2.62×10^{-2}
$H \rightarrow W^+W^-$	2.14×10^{-1}
$H \rightarrow \tau^+\tau^-$	6.27×10^{-2}
$H \rightarrow b\bar{b}$	5.84×10^{-1}
$H \rightarrow Z\gamma$	1.53×10^{-3}
$H \rightarrow \mu^+\mu^-$	2.18×10^{-4}

Table 1.1: The predicted branching ratios of the Higgs boson for m_H=125 GeV [45].

The Higgs discovery

For a low-mass Higgs (between 110 and 150 GeV) the five final states that play an important role at the LHC are $\gamma\gamma$, ZZ^*, WW^*, $\tau^-\tau^+$ and $b\bar{b}$. In particular, despite the low branching ratios, the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4\ell$ channels are those providing the best mass resolution (1-2 GeV) [34]. For this reason, such channels have been called the *golden channels* and both ATLAS and CMS experiments relied on these to measure the Higgs boson mass. In the first case, the search was performed for a narrow peak over a smoothly falling background in the invariant mass distribution of two photons with high p_T . An excess of events was found, with significance above 5σ, resulting in an estimate for the mass of the observed particle of $126.0 \pm 0.4(\text{stat.}) \pm 0.4(\text{syst.})$ GeV [34]. In the second case, the search was performed for a narrow mass peak over a small continuous background dominated by non-resonant ZZ^* production, deriving from quark-antiquark annihilation and gluon-gluon fusion processes [45]. An excess of events was found also here, with a significance above 5σ, resulting in an estimate for the mass of the observed particle of $125.3 \pm 0.4(\text{stat.}) \pm 0.5(\text{syst.})$ GeV [35]. In Fig. 1.8 the distribution of the reconstructed invariant mass of the two photons in $H \rightarrow \gamma\gamma$ events in the ATLAS detector is shown on the left, while on the right the distribution of the invariant masses of the four charged leptons in the CMS $H \rightarrow ZZ^* \rightarrow 4\ell$ search is shown. The combined mass measurements gave the following result [60]:

$$m_H = 125.09 \pm 0.21(\text{stat.}) \pm 0.11(\text{syst.}) \text{ GeV}.$$

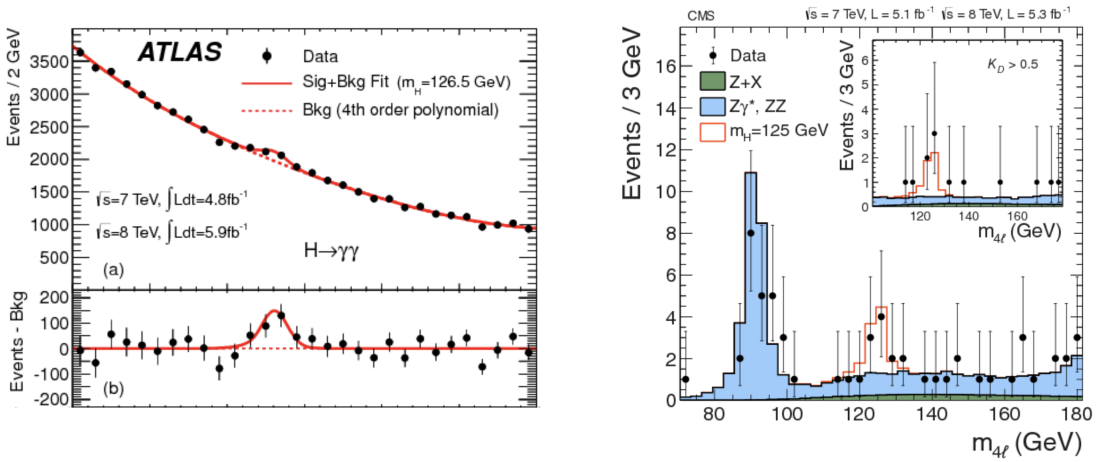

Figure 1.8: (Left) The reconstructed invariant mass distribution of the diphotons candidates in the ATLAS experiment [35]. The points represent the data, the dotted line represents the background, and the solid line the fit to the data of the sum of a signal component fixed to $m_H = 126.5$ GeV and of a background component. (Right) The distribution of the reconstructed invariant masses of the four leptons candidates in the CMS experiment [34]. The points represent the data, the filled histograms represent the background, and the open histogram shows the signal expectation for a Higgs boson of mass 125 GeV.

1.6 Physics beyond the Standard Model

Thanks to the discovery of the Higgs boson in 2012, the SM is considered to be complete as it was formulated in the 1980s. The validity of its predictions has been widely verified (see Sec. 1.4.3), and the precision evaluation of its parameters have been performed at LEP, SLC, and LHC [61]. However, despite being the most successful theory of particle physics to date, the SM is not a complete theory of Nature:

- **The hierarchy problem**. The mass of the Higgs boson, as measured by the CMS and ATLAS experiments, is of the same order of magnitude of the other gauge bosons ($M_{W,Z,H} \sim 100$ GeV). However, SM calculations show that it receives large quantum contributions from all energy scales, all the way up to the highest energy scale at which the SM is valid [62], which is the Plank mass ($M_P \sim 10^{19}$ GeV). Consequently, the value for the Higgs mass should be instead of the order of magnitude of M_P, which is 10^{17} bigger compared to the measurements obtained at the LHC. Such vaste difference in scale, besides not being fully understood yet, also arises *naturalness* issues within the SM. In order to retrieve in the calculations the value of the measured mass, indeed, the bare mass parameter of the Higgs in the SM must be fine-tuned in a way that almost completely cancels these quantum contributions.

- **Matter-antimatter asymmetry**. In the early universe, the Big Bang should have created equal amounts of baryonic matter (or *matter*) and antibaryonic matter (or *antimatter*). However, as it can be experienced in everyday life, the current observable universe is dominated by matter, and the SM does not provide an explanation of what happened to the antimatter or about what determined such huge disproportion between the two. In 1967, Sakharov [63] elaborated a set of three conditions to be satisfied during baryogenesis in order to originate a matter-antimatter imbalance: the violation of the baryonic number, the violation of the CP-symmetry, and the presence of interactions out of thermal equilibrium. Currently, there is no exper-

imental evidence for the baryon number violation, while CP violation has already been observed in the decay of strange [64] and beauty particles [65, 66]. In 2019, the LHCb (Large Hadron Collider beauty) Collaboration at the LHC announced the first observation of CP violation also in charm particle decays [67], opening the way for the study of CP-violating effects in the sector of up-type quarks and for new physics searches using charm CP asymmetry measurements. Besides the Sakharov conditions, other explanations have been provided to the matter-antimatter asymmetry, including possible widely distant regions of the universe where the antimatter could dominate, or the theory of the *mirror anti-universe* [68], but currently a consistent model for the description of such phenomenon does not yet exist.

- **Dark matter and dark energy**. According to the current cosmological model and observations, only the 5% of the energy and matter present in the Universe can be explained by the SM. Out of this undisclosed, remaining 95%, around 26% should be a new, invisible kind of matter, not interacting with SM and that is called *dark matter* (DM). The remaining 69% is instead referred to as *dark energy*, an unknown form of energy with exotic physical properties and affecting the universe on the largest scale, that was also hypothesized to explain the acceleration rate of the Universe expansion [69]. On the other hand, although dark matter has never been directly observed yet, there are many evidences supporting its existence. One of most important comes from the analysis of the rotation curves of the disc galaxies, obtained by studying the orbital speeds of the luminous materials (stars and gas) in a galaxy as a function of their radial distance r from the galaxy's center. Assuming that most of the galaxy mass is concentrated in the central bulge, the tangential velocities of the stars is supposed to decrease as $M(r)/r$. However, the rotation curve remains flat as the distance from the center increases, with the stars' velocities slowly decreasing with r. One of the most straightforward solution proposed to explain this unexpected behavior is to hypothesize the existence of DM particles affecting the mass distribution of the galaxy. The *weakly interacting massive particles* (WIMPs) are currently the most popular search candidates for DM, which are supposed to interact via gravity or any other force potentially not part of the SM itself [70]. The WIMPs are predicted by several Beyond Standard Model (BSM) theories, such as the theory of Supersymmetry (SUSY), that predicts a *super-symmetric partner* for each particle in the SM. Such partners are characterized by the same mass, but different spin by half of a unit, and they may decay into pairs of WIMPs. Similar searches can be also performed underground, through the possible observation of the scattering between WIMPs and nuclei, as it is the case of the XENON [71] and PandaX [72] collaborations. Other experiments, such as the Axion Dark Matter Experiment [73], look instead for *axions* as possible DM candidates, and searches for DM particles are also performed at the LHC, for example through the invisible decay of the Higgs boson [74] or looking for DM signatures in monophoton and monojet events.

- **Neutrino masses**. The experimental observation of neutrino oscillations [75, 76] has demonstrated that, differently from the original SM predictions, neutrinos have mass. Such mass terms can be added to the SM but this leads to new theoretical inconsistencies. In fact, considering neutrinos as normal Dirac particles, the introduction of their masses "by hand" would imply an extraordinarily, and hence unnaturally, small Yukawa coupling to the Higgs field. One of the hypothetical model developed to provide an alternative explanation for the exceptional smallness of these masses is the so-called *seesaw mechanism* [77]. According to this mechanism, neutrinos are Majorana particles, meaning they are their own antiparticles as well ($\nu \equiv \bar{\nu} \equiv \nu_M$). A variety of experiments (for example CUORE [78] and

KamLAND-Zen [79]) is currently trying to probe the intrinsic nature of the neutrino by searching for the neutrinoless double beta decay. Such process is a theoretical radioactive decay process that, if observed, would imply the violation of total lepton number conservation, consequently confirming the Majorana nature of neutrinos.

- **Gravity**. While the SM describes three of the four fundamental forces, gravity is interpreted in the theoretical framework of general relativity which appears to be incompatible with the underlying SM QFT. Quantum gravity is the field of theoretical physics that aims to reconcile gravity and quantum mechanics under a unique framework, and in many QG theories the gravitational field itself is quantized. Such hypothetical quantum of gravity is called *graviton*, and it is supposed to mediate the force of gravitational interaction as the SM gauge bosons mediate the electroweak and the strong interactions. Tests for quantum gravity are currently ongoing at the LHC, with the aim of searching for possible emission of gravitons in a variety of theoretical models.

In addition to these unexplained phenomena, there are some experimental results that disagree with the SM predictions, with a significance that exceeds or approaches 5σ, the widely accepted threshold for a discovery in particle physics. The anomalous mass of the W boson from the CDF Collaboration reported in Sec.1.4.3, for example, is one of them, while the Muon g-2 experiment at Fermilab has measured the magnetic dipole moment of the muon to differ from the SM prediction by 4.2σ [80]. Scientists have developed many BSM theories, in order to provide a more exhaustive formulation of the SM that would solve the aforementioned issues, and consequently explain the observed discrepancies with the predictions. With this goal in mind, in the last decades a common effort between the theoretical and experimental communities led to powerful and different strategies for BSM searches (like model-independent approaches) and to the development of new, sophisticated tools for particle reconstruction and data analysis (for example by using cutting-edge machine learning techniques) with the aim of improving the accuracy and the precision of the measurements.

1.6.1 Monophoton signatures for dark matter searches at the Large Hadron Collider

As mentioned in the previous section, DM searches at the LHC can be performed by looking at events with a large amount of missing energy and a single photon. In proton-proton collisions, DM candidates χ can be produced through the $q\bar{q} \rightarrow \gamma\chi\bar{\chi}$ process if the new particles interact with SM quarks or gluons via new couplings. The monophoton final state is identifiable with high efficiency and purity and, although DM cannot be directly detected at the LHC, its production can be inferred from an observation of events with large transverse momentum imbalance. The monophoton final state has also the advantage of being a model-independent signature, meaning that it has sensitivity to several DM theories and models of extra spatial dimensions, such as the case of the Arkani-Hamed, Dimopoulos and Dvali model (ADD), that also provides a possible solution to the hierarchy problem. The ADD model postulates the existence of a 4+n dimensional space-time where gravitons can propagate freely, implying that true mass scale of the gravitational interaction in this multidimensional space-time has the same order of magnitude of the electroweak scale. According to the ADD model, the graviton can couple directly to the photon or to a quark (Fig. 1.9 right). On the other hand, in DM production through a vector or axial vector mediator, a photon can be radiated from incoming quarks (Fig. 1.9 left), while other processes under investigation involve couplings of DM with electroweak particles via effective interaction (Fig. 1.9 center) [81]. Searches for new physics in the monophoton final state have been carried out in the past years by the CMS and ATLAS

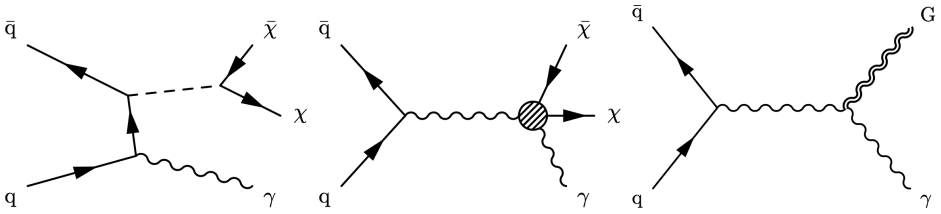

Figure 1.9: Leading-order diagrams of the simplified DM model (left), electroweak-DM effective interaction (center), and graviton (G) production in the ADD model (right), with a monophoton final state [81].

experiments [81–84], providing exclusion limits on DM production cross sections and on the parameters in models containing extra spatial dimensions. The primary background to such DM searches is the irreducible contribution from SM $Z(\rightarrow \nu\bar{\nu})\gamma$ production. Therefore, a careful characterization of this final state and an accurate measurement of its SM production cross section are essential to reduce the associated uncertainty in DM analysis, and to provide more stringent limits.

Chapter 2

Anomalous gauge couplings in the Effective Field Theory framework

A powerful way to search for physics beyond the SM is to use a model-independent approach to study the production of multiple electroweak bosons at the LHC. Multiboson processes are indeed sensitive to triple and quartic gauge couplings, that can be interpreted in the framework of an Effective Field Theory (EFT) in order to probe possible new physics effects and to test the electroweak sector of the SM at the TeV scale. The first part of this chapter gives a concise description of the theoretical context where an EFT can be built, and it outlines its principal features. Then, the formulation of the charged and neutral gauge couplings is presented, connecting the original anomalous couplings approach with the modern EFT approach. The last section of the chapter illustrates the latest limits obtained at the LHC for the $Z\gamma$ production.

2.1 The Effective Field Theory

The simplest strategy to search for physics beyond the SM is to directly look for direct signals of on-mass shell production of new particles. However, despite the many BSM theories that have been developed in the years, the Higgs boson discovered in 2012 was the last particle to be observed directly and no one of the hypothetical new particles has been observed yet. A second, powerful method to hunt for new physics is to study the well-known interactions between the SM particles, looking for possible deviations between the measurements and the SM predictions. New physics can indeed manifest also in the form of novel interactions of known particles, originated by the virtual presence of new particles in quantum loops and in new amplitudes generated by their exchange at tree-level, which would lead to a change in the value of the measured cross sections. This approach does not refer to any particular SM extension but it is completely model-independent, allowing the method to be potentially sensitive to any BSM effect that could manifest and, as for the rest of the searches, in the case that no new physics appears it is able to quantify the accuracy with which the new physics is excluded. When building a model-independent framework for studying non-standard interactions, the following features have to be incorporated [85]:

- The axioms of the S-matrix (unitary, analyticity...) should be respected. The S-matrix connects the initial and the final state of a physical system undergoing a scattering process, and therefore it is closely related to the transition probability amplitude in QFT and to the cross sections of the interactions.

- The Lorentz invariance and the $SU(2)_L \times SU(3)_C \times U(1)_Y$ gauge symmetry of the SM should be respected.

- The SM should be recovered in the appropriate limit(s).

- In the extended theory, it should be possible to evaluate the radiative corrections at any order in the SM interactions, and in the new interactions as well.

Using an EFT approach is a powerful way to include all the above features. An EFT is constructed by extending the SM Lagrangian $\mathcal{L}_{SM}$, whose all operators are restricted to be of mass dimension four. By dimensional analysis, the new operators are then required to have coefficients with dimensions of inverse powers of mass, and hence they are suppressed if such mass is large compared with the experimentally-accessible energies. The extended Lagrangian $\mathcal{L}_{EFT}$ can be written as:

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

where Λ is termed the mass scale, and it represents the scale at which the new physics should appear. The dimensionless coefficients c_i of the higher-dimension operators $\mathcal{O}_i$ parameterize the strength with which the new physics couples to the SM particles. The dimension of the operators is labeled with d. From this formulation of the Lagrangian, it can be clearly seen that the dominant extended operators will be those of the lowest dimensionality and that the SM is recovered in the limit $\Lambda \to \infty$. Operators with an odd number of dimensions do not conserve baryon and lepton numbers, and they are usually neglected in SM EFT. In addition, at dimension five there is only one operator, that is responsible for generating Majorana masses for neutrinos [86]. Consequently, the largest contribution which can affect the triple gauge couplings (TGCs) and the quartic gauge couplings (QGCs) comes from dimension-six operators. This chapter does not provide a description of the QGCs, given that only TGCs can play a role in the Zγ processes.

2.2 Triple gauge couplings

The anomalous couplings of electroweak bosons were originally introduced when it was not certain yet that the electroweak interaction was a spontaneously broken gauge theory [87]. For this reason, the triple gauge couplings have been first studied within the *anomalous couplings* framework, using either Lagrangian or a vertex function approach. Considering the case of a triple charged verted, the Lagrangian approach is based on the following Lagrangian [85]:

$$\begin{aligned}\mathcal{L}_{WWV} = ig_{WWV} \Bigg(& g_1^V \left(W_{\mu\nu}^+ W^{-\mu} - W^{+\mu} W_{\mu\nu}^- \right) V^\nu + \kappa_V W_\mu^+ W_\nu^- V^{\mu\nu} + \frac{\lambda_V}{M_W^2} V^{\mu\nu} W_\nu^{+\rho} W_{\rho\mu}^- + \\ & + ig_4^V W_\mu^+ W_\nu^- (\partial^\mu V^\nu + \partial^\nu V^\mu) - ig_5 \epsilon^{\mu\nu\rho\sigma} (W_\mu^+ \partial_\rho W_\nu^- - \partial_\rho W_\mu^+ W_\nu^-) V_\sigma + \\ & + \tilde{\kappa} W_\mu^+ W_\nu^- \tilde{V}^{\mu\nu} + \frac{\tilde{\lambda}_V}{m_W^2} W_\mu^{\nu+} W_\nu^{-\rho} \tilde{V}_\rho^\mu \Bigg),\end{aligned} \tag{2.1}$$

where $V = \gamma, \mathrm{Z}$, $W_{\mu\nu}^\pm = \partial_\mu W_\nu^\pm - \partial_\nu W_\mu^\pm$, and $V_{\mu\nu} = \partial_\mu V_\nu - \partial_\nu V_\mu$. The first three terms conserve CP, while the last four terms are CP-violating. Due to the electromagnetic gauge invariance $g_1^\gamma = 1$ and $g_4^\gamma = g_5^\gamma = 0$. Taking into account only SM contributions, the CP-violating terms are not considered and the charged TGCs are given by $g_1^Z = \kappa_Z = \kappa_\gamma = 1$ and $\lambda_Z = \lambda_\gamma = 0$, retrieving the formulation of the Lagrangian in Sec. 1.3.2, where $g_{WWZ} = e \cot\theta_W$ and $g_{WW\gamma} = e$. The momentum-space analogue of the Lagrangian approach is referred to as the vertex function approach, which parameterizes the trilinear

boson vertex function as [85]:

$$\begin{aligned}\Gamma_V^{\alpha\beta\mu} = f_1^V(q-\overline{q})^\mu g^{\alpha\beta} - \frac{f_2^V}{m_W^2}(q-\overline{q})^\mu P^\alpha P^\beta + f_3^V(P^\alpha g^{\mu\beta} - P^\beta g^{\mu\alpha}) + \\ + i f_4^V(P^\alpha g^{\mu\beta} - P^\beta g^{\mu\alpha}) + i f_5^V \epsilon\mu\alpha\beta\rho(q-\overline{q})_\rho + \\ - i f_6^V \epsilon\mu\alpha\beta\rho P_\rho - f_7^V(q-\overline{q})^\mu \epsilon\alpha\beta\rho\sigma P_\rho (q-\overline{q})_\sigma,\end{aligned}$$

where P, q and $\overline{q}$ are the four momenta of V, W^-, and W^+, respectively. The coefficients f_i^V are form factors depending on P^2: the first three conserve CP, while the remaining four violate CP. One of the main problems of the anomalous couplings framework is that it leads to a violation of the unitarity bounds. Considering the Lagrangian approach, indeed, the anomalous couplings turn out to be energy-independent, yielding amplitudes growing like s/m_W^2, and hence violating the unitarity bound at high energy. This issue could be solved by switching to the vertex function approach and applying form factors falling off at large s, which is however considered overly restrictive [85]. On the other hand, reinterpreting the anomalous couplings as coefficients of higher (> 4) dimension operators would allow the unitarity bounds to be automatically respected. The reason for that relies on the fact that EFT is, by construction, a low-energy theory used to fit the actual data: since the data necessarily respect the unitarity bound, an effective field theory fitting the data will respect the bound as well. The five dimension-six operators that contribute to the triple charged vertex are the following [85]:

$$\begin{aligned}\mathcal{O}_{WWW} &= Tr[W_{\mu\nu}W^{\nu\rho}W_\rho^\mu], \\ \mathcal{O}_W &= (D_\mu\phi)^\dagger W^{\mu\nu}(D_\nu\phi), \\ \mathcal{O}_B &= (D_\mu\phi)^\dagger B^{\mu\nu}(D_\nu\phi), \\ \mathcal{O}_{\tilde{W}WW} &= Tr[\tilde{W}_{\mu\nu}W^{\nu\rho}W_\rho^\mu], \\ \mathcal{O}_{\tilde{W}} &= (D_\mu\phi)^\dagger \tilde{W}^{\mu\nu}(D_\nu\phi).\end{aligned}$$

Only the dimension-six operators are considered here, since they are expected to be the dominant operators, and they are constructed from the Higgs doublet field ϕ and the usual SM fields $W_{\mu\nu}$ and $B_{\mu\nu}$. The covariant derivative for ϕ is given by $D_\mu \equiv \partial_\mu + \frac{i}{2}g'B_\mu + ig_W\vec{W}_\mu\frac{\vec{T}^\mu}{2}$, while $\tilde{W}_{\mu\nu} = \sum_j W_{\mu\nu}^j\frac{\sigma_j}{2}$. The first three operators are CP-conserving, while the remaining two are CP-violating. The parameters of the Lagrangian in Eq. 2.1 can be hence expressed in terms of coefficients (c_{WWW}, c_W, c_B, $c_{\tilde{W}WW}$ *and* $c_{\tilde{W}}$) of these operators [85]:

$$\begin{aligned}g_1^Z &= 1 + c_W\frac{m_Z^2}{2\lambda^2}, \\ \kappa_\gamma &= 1 + (c_W + c_B)\frac{m_W^2}{2\lambda^2}, \\ \kappa_Z &= 1 + (c_W - c_B\tan^2\theta_W)\frac{m_W^2}{2\lambda^2}, \\ \lambda_\gamma &= \lambda_Z = c_{WWW}\frac{3g^2m_W^2}{2\lambda^2}, \\ g_4^V &= g_5^V = 0, \\ \tilde{\kappa}_\gamma &= c_{\tilde{W}}\frac{m_W^2}{2\lambda^2},\end{aligned}$$

$$\tilde{\kappa}_{\rm Z} = -c_{\tilde{W}} \tan^2\theta_W \frac{m_W^2}{2\lambda^2},$$
$$\tilde{\lambda}_\gamma = \tilde{\lambda}_{\rm Z} = c_{\tilde{W}WW} \frac{3g^2 m_W^2}{2\lambda^2}.$$

Similarly, the form factors used in the vertex function approach become [85]:

$$\begin{aligned}
f_1^\gamma &= 1 + c_{WWW}\frac{3g^2P^2}{4\Lambda^2}\\
f_1^{\rm Z} &= 1 + c_W\frac{m_Z^2}{\lambda^2} - c_{WWW}\frac{3g^2P^2}{4\Lambda^2},\\
f_2^\gamma &= f_2^{\rm Z} = c_{WWW}\frac{3g^2m_W^2}{2\Lambda^2},\\
f_3^\gamma &= 2 + (c_W(1+\cos^2\theta_W) - c_B\sin^2\theta_W)\frac{m_Z^2}{2\lambda^2} + c_{WWW}\frac{3g^2m_W^2}{2\lambda^2},\\
f_4^V &= f_5^V = 0,\\
f_6^\gamma &= c_{\tilde{W}}\frac{m_W^2}{2\lambda^2} - c_{\tilde{W}WW}\frac{3g^2m_W^2}{2\Lambda^2},\\
f_6^{\rm Z} &= c_{\tilde{W}}\tan^2\theta_W\frac{m_W^2}{2\lambda^2} - c_{\tilde{W}WW}\frac{3g^2m_W^2}{2\Lambda^2},\\
f_7^\gamma &= f_7^{\rm Z} = -c_{\tilde{W}WW}\frac{3g^2m_W^2}{4\Lambda^2}.
\end{aligned}$$

2.2.1 Neutral triple gauge couplings

Differently from the charged TGCs, the SM forbids vertices of three neutral bosons (ZZZ, ZZγ and Z$\gamma\gamma$) when all particles are on-shell (Fig. 2.1 left). This means that neutral TGCs could only be realized as *anomalous* neutral triple gauge couplings (aNTGCs) in the EFT approach (Fig. 2.1 right). According to the vertex function formalism, the anomalous vertices arising from the interactions of three neutral gauge bosons can be parameterized as [87–91]:

$$\begin{aligned}
\Gamma^{\alpha\beta\mu}_{\rm ZZV}(q_1,\ q_2,\ p) =& \frac{-e(p^2-m_V^2)}{m_Z^2}\left[f_4^V(p^\alpha g^{\mu\beta} + p_3^\beta g^{\mu\alpha}) - f_5^V\epsilon^{\mu\alpha\beta\rho}(q_1-q_2)_\rho\right]\\
\Gamma^{\alpha\beta\mu}_{Z\gamma V}(q_1,\ q_2,\ p) =& \frac{-e(p^2-m_V^2)}{m_Z^2}\left\{h_1^V\left(q_2^\mu g^{\alpha\beta} - q_2^\alpha g^{\mu\beta}\right) + \frac{h_2^V}{m_Z^2}p^\alpha\left[(pq_2)g^{\mu\beta} - q_2^\mu p^\beta\right] + \right.\\
& \left. -h_3^V\epsilon^{\mu\alpha\beta\rho}q_{2\rho} - \frac{h_4^V}{m_Z^2}p^\alpha\epsilon^{\mu\beta\rho\sigma}p_\rho q_{2\sigma}\right\},
\end{aligned} \tag{2.2}$$

where p, q_1 and q_2 are the four-momenta of the incoming off-shell Z or γ (as in Fig 2.1 right) and of the two outcoming on-shell bosons. In Eq. 2.2, V can be a photon or a Z boson. Six-dimensions operators are the lowest dimension operators expected to provide a contribution to these vertices, but by construction none of them can induce nTGCs [92, 93]. Therefore, the dominant operators that contribute are those of dimension-eight, namely:

$$\begin{aligned}
\mathcal{O}_{BW} &= i\phi^\dagger B_{\mu\nu}W^{\mu\rho}\left\{D_\rho, D^\nu\right\}\phi,\\
\mathcal{O}_{WW} &= i\phi^\dagger W_{\mu\nu}W^{\mu\rho}\left\{D_\rho, D^\nu\right\}\phi,\\
\mathcal{O}_{BB} &= i\phi^\dagger B_{\mu\nu}B^{\mu\rho}\left\{D_\rho, D^\nu\right\}\phi,\\
\mathcal{O}_{\tilde{B}W} &= i\phi^\dagger \tilde{B}_{\mu\nu}W^{\mu\rho}\left\{D_\rho, D^\nu\right\}\phi.
\end{aligned}$$

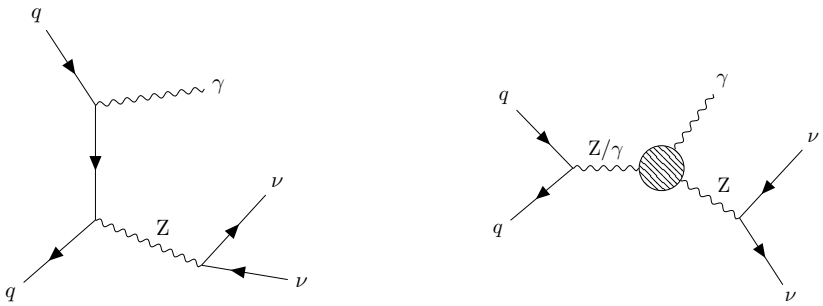

Figure 2.1: Production channels for the Zγ process: SM t-channel (left) and s-channel including the presence of the aNTGC (right).

Therefore, the anomalous couplings involved in the ZZV production and expressed in terms of operators' coefficients are given by [91]:

$$
\begin{aligned}
f_5^Z &= 0, \\
f_5^\gamma &= \frac{v^2 m_z^2 c_{\tilde{B}W}}{4 c_w s_w \lambda^4}, \\
f_4^Z &= \frac{m_Z^2 v^2 \left(c_w^2 \frac{c_{WW}}{\lambda^4} + 2 c_w s_w \frac{c_{BW}}{\lambda^4} + 4 s_w^2 \frac{2 c_{BB}}{\lambda^4} \right)}{2 c_w s_w}, \\
f_4^\gamma &= -\frac{m_Z^2 v^2 \left(-c_w s_w \frac{c_{WW}}{\lambda^4} + \frac{c_{BW}}{\lambda^4} (c_w^2 - s_w^2) + 4 c_w s_w \frac{c_{BB}}{\lambda^4} \right)}{4 c_w s_w}.
\end{aligned}
$$

where f_5^V are CP-conserving and f_4^V are CP-violating. On the other hand, for the ZγV vertex the couplings are the following [91]:

$$
\begin{aligned}
h_3^Z &= \frac{v^2 m_Z^2 c_{B\tilde{W}}}{4 c_w s_w \lambda^4}, \\
h_4^Z &= h_3^\gamma = h_4^Z = 0, \\
h_1^Z &= \frac{m_Z^2 v^2 \left(-c_w s_w \frac{c_{WW}}{\lambda^4} + \frac{c_{BW}}{\lambda^4} (c_w^2 - s_w^2) + 4 c_w s_w \frac{c_{BB}}{\lambda^4} \right)}{4 c_w s_w}, \\
h_2^Z &= 0, \\
h_1^\gamma &= -\frac{m_Z^2 v^2 \left(s_w^2 \frac{c_{WW}}{\lambda^4} + 2 c_w s_w \frac{c_{BW}}{\lambda^4} + 4 c_w^2 \frac{c_{BB}}{\lambda^4} \right)}{4 c_w s_w}, \\
h_2^\gamma &= 0.
\end{aligned}
$$

The simplest way to generate anomalous couplings is through a fermionic triangle loop, mediated by SM or new fermions. For example, in MSSM such new fermions are neutralinos or charginos, with the magnitude of the contribution depending on the MSSM parameters. However, considering the contribution of fermionic triangles loop only leads many CP-conserving couplings to vanish, as a direct consequence of the symmetries of the fermionic trace of the triangular diagram and of Shoutens relation [89]. Such couplings can though happen perturbatively at a higher-loop level, or new strong interactions can generate non

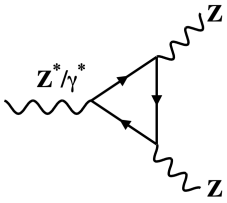

Figure 2.2: Fermionic triangle contributing to the γZZ and γZZ couplings at 1-loop level.

perturbative couplings among the vector bosons, like for example those expected in the Vector Dominance Model [94].

2.3 Zγ production

Possible effects originated by anomalous couplings in Zγ processes were first investigated at LEP and Tevatron [95, 96], considering several final states ($\ell^{\pm}\ell^{\mp}$, $q\bar{q}$, $\nu\bar{\nu}$) for the decay of the Z boson. The CP-conserving couplings h_3^V and h_4^V always lead to real amplitudes interfering with the SM ones, implying that the various observables are linearly sensitive to these terms. On the contrary, the CP-violating couplings produce purely imaginary amplitudes that have negligible interference with the SM ones [97]. The latest measurements performed at the LHC only placed the limits on the magnitude of h_3^V and h_4^V. The analysis of Zγ processes has already been carried out by the CMS and the ATLAS experiments using events produced in proton-proton collision at $\sqrt{s} = 7$ and 8 TeV, considering both the leptonic and the invisible channels for the Z-decay. ATLAS has estimated the limits in the $Z \to \nu\bar{\nu}$ channel also at $\sqrt{s} = 13$ TeV [98], however only using the data collected during 2015 and 2016 (36.1 fb^{-1}). From the overview of the limits on the anomalous NTGCs (Fig. 2.3) it can be observed that the current best limits are those evaluated by ATLAS using part of the Run 2 data. This work aims to improve such results by providing both the first CMS measurement of the limits at a center of mass energy of 13 TeV as well as the first limits' estimation using the full Run 2 data.

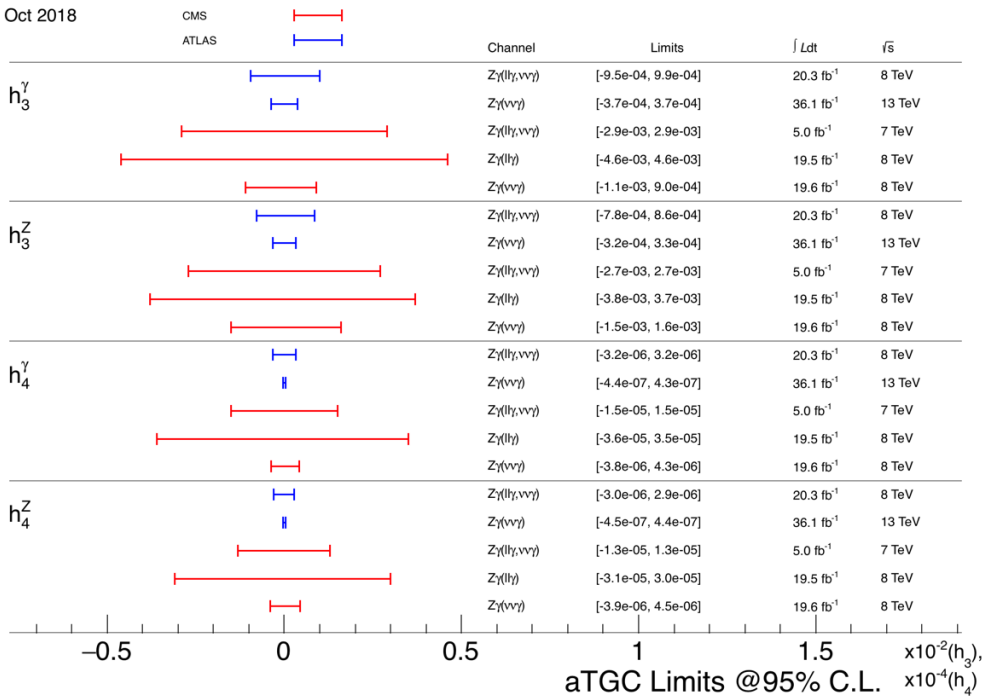

Figure 2.3: Limits on neutral ZZγ and ZZZ aTGC couplings. The limits have been placed on the anomalous parameters h_3 and h_4 [99].

Chapter 3

The Compact Muon Solenoid experiment at the Large Hadron Collider

Particle accelerators are sophisticated machines able to, as their name implies, accelerate charged particles at speeds close to the speed of light and to contain them in well-defined beams. The particles can then be directed and smashed onto a fixed target, or they can collide with other particles traveling in the accelerator in the opposite direction. By studying the outcome of the collisions, physicists can probe a variety of physics phenomena making precise measurements of known processes and interactions, and searching for new, unpredicted ones. In this chapter, the biggest and more powerful accelerator in the world, the LHC, is introduced, and the four main experiments installed alongside the accelerator ring are presented. Among them, the Compact Muon Solenoid (CMS) experiment is described in detail, providing a brief description of its structure and of the different components of its sub-detectors. The second part of the chapter is focused on the particles reconstruction in CMS, performed by using the information coming from the particles interaction in the different sub-detectors and dedicated reconstruction algorithms.

3.1 The Large Hadron Collider

The Large Hadron Collider [100] is a hadron and ion accelerator and collider that operates at CERN (Conseil Europeen pour la Recherche Nucleaire), an organization for particle physics research established in 1954 and located near Geneva, on the border between Switzerland and France. The LHC has been built in a 26.7 km circular tunnel, constructed between 1984 and 1989 for the LEP collider, that lies between 45 and 170 meters underground. The LHC started its operations in 2008, and after the so-called Run 1 (2010-2013) and the Run 2 (2015-2018) [101], the Run 3 has officially started in July 2022 reaching the world's record energy of 13.6 TeV in the proton-proton center of mass [102]. In the LHC, protons are grouped in around 2800 bunches, with about $1.5{\cdot}10^{11}$ protons per bunch, and they are supplied by a complex injection chain, as shown in Fig. 3.1. First, protons are accelerated up to 50 MeV in the LINAC (LINear ACcelerator). Then, they enter the Proton Synchrotron Booster (PSB) to reach 1.4 GeV and finally the Super Proton Synchrotron (SPS) increases their energy up to 450 GeV before they are injected into the main ring, where they are accelerated in two different beam pipes kept in ultra-high vacuum regime. The beams are focused by 392 magnetic quadrupoles distributed along the collider, where the circular trajectory is maintained by means of a 8.4 T magnetic field produced by 1232 superconductive-14.3 m long dipoles. The magnets are kept at an operating temperature of 1.9 K, provided by 96 tonnes of superfluid helium-

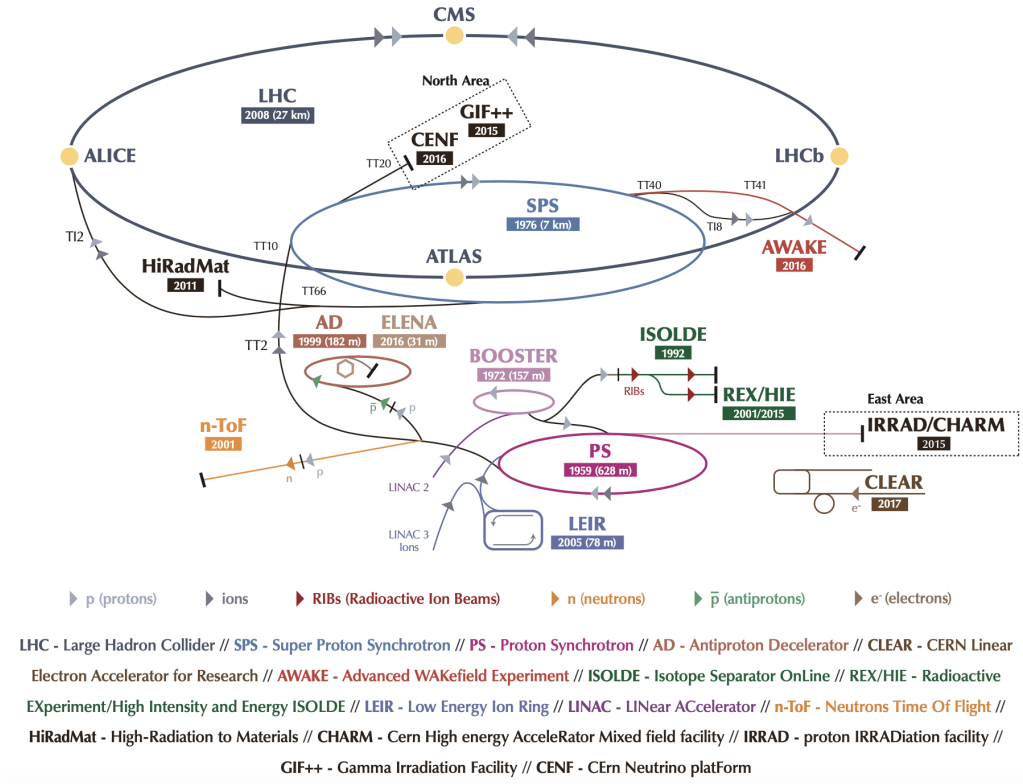

Figure 3.1: The injection chain for the Large Hadron Collider [103].

4. Once the protons in each beam reach the energy of 6.8 TeV, the collisions take place in four interaction points along the ring, where the four main experiments are installed:

- **ATLAS** [104] (A Toroidal LHC ApparatuS): a general-purpose detector whose aim is to study high-energy proton-proton collisions to perform precise tests of the SM, and to investigate new physics phenomena.

- **CMS** [105] (Compact Muon Solenoid): it is the other large general-purpose detector at LHC, and it is described in detail in Sec. 3.2. Together with ATLAS [35], CMS discovered the Higgs boson in July 2012 [34].

- **LHCb** [106] (Large Hadron Collider beauty): a dedicated b-physics experiment, designed primarily to measure the parameters of CP violation in the decay of charm and beauty hadrons. The LHCb detector is a single-arm forward spectrometer, because the beauty hadrons at high energies are mostly produced in the same forward or backward cone.

- **ALICE** [107] (A Large Ion Collider Experiment): this experiment is optimized to study heavy-ion (Pb-Pb nuclei) collisions and it is focused on physics of strong interactions at very high energy and density. One of its goals is to verify the existence of the quark-gluon plasma, whose properties would be crucial for a better understanding of the quantum chromodynamics.

The collision rate at LHC is 40 MHz, meaning that the interactions take place every 25 ns. For a process of cross section σ, the instantaneous luminosity is defined as:

$$L = \frac{dN}{dt} \cdot \frac{1}{\sigma},$$

where dN is the number of events detected in time dt. In order to express the size of the collected data, the integrated luminosity $\int L dt$ is used, measured in inverse of barns (b^{-1}), where a barn is a metric unit of area equal to 10^{-28} m^2. The luminosity depends on the beam parameters and can be written as:

$$L = \frac{N_b^2 \; n_b \; f_{rev} \gamma_{rev}}{4\pi \; \epsilon_n \; \beta^*} F,$$

where N_b is the number of protons in each bunch and n_b the number of bunches per beam. f_{rev} is the revolution frequency, γ_r is the Lorentz factor, ϵ_n the normalized transverse beam emittance and β^* is the Twiss beta function at the collision point. F is the geometric luminosity reduction factor due to the crossing angle θ_C at the interaction point:

$$F = \left(1 + \left(\frac{\theta_C \sigma^z}{2\sigma^*}\right)^2\right)^{1/2},$$

where σ^z and σ^* are the root mean square of the bunch length and the transverse root mean square of the beam size at the interaction point, respectively. The LHC was designed to collide protons at a center of mass energy of 14 TeV with a nominal peak instantaneous luminosity L = $10^{34} cm^{-2} s^{-1}$.

3.2 The Compact Muon Solenoid Experiment

The CMS experiment is one of the two general purpose experiments installed at LHC. It is located near Cessy, in France, at the interaction point 5 along the LHC ring about 100 meters underground. It was conceived to study the high-energy, high-luminosity proton-proton collisions and it has a wide range of research goals, from precision measurements of SM parameters and study of the Higgs boson properties, to investigation of new BSM physics, extra dimensions and dark matter. The CMS detector has a cylindrical structure, symmetrical around the interaction point, and can be divided into two regions: the *barrel*, the central part that is coaxial with the beamline, and two *endcaps* that close the central part on both sides. It consists of a series of concentric sub-detectors, each one with a specific function: the silicon tracker, the electromagnetic calorimeter, the hadronic calorimeter, and the muon chambers. A superconducting solenoid magnet surrounds the tracker and the calorimeter providing a magnetic field of 4 T, while the muon chambers are placed inside the magnet steel return yoke. The whole structure has a length of 21.6 m and a diameter of 14.6 m, and weights 12500 t. An overview of the CMS detector is shown in Fig. 3.2.

3.2.1 Coordinate system

The coordinate system adopted at the LHC is a cylindrical right-handed system (Fig. 3.3) with the origin placed in the collision point. The z-axis follows the beam direction and the x-axis points to the center of the LHC ring, while the y-axis points upward and it is perpendicular to the beam direction. The following coordinates are hence defined:

- $r = \sqrt{x^2 + y^2}$: the radial coordinate in the x-y plane, corresponding to the distance from the z-axis.
- $\theta = \tan^{-1}(x/y)$: the polar angle measured from the positive z-axis.
- $\phi = \arctan(y/z)$: the polar angle measured from the positive x-axis in the x-y plane.

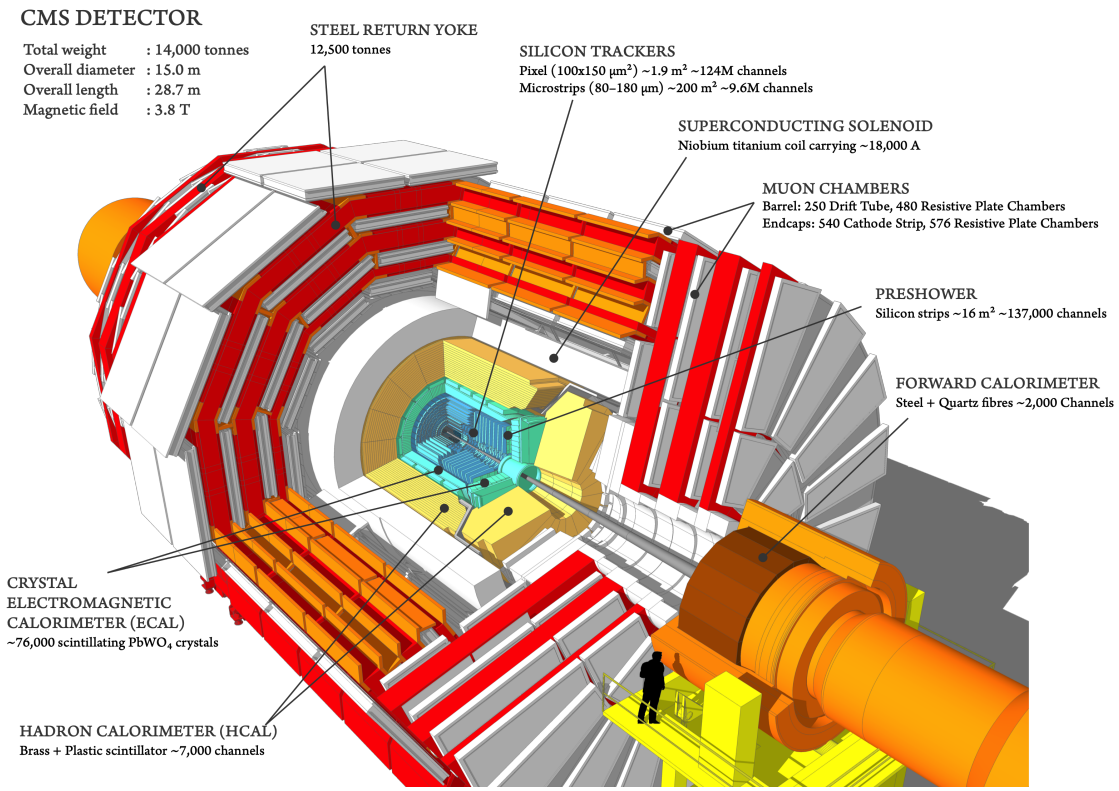

Figure 3.2: The CMS detector and its subsystems. From the innermost to the outermost layer: the silicon tracker, the preshower, the electromagnetic calorimeter, the hadronic calorimeter, the superconducting solenoid and the muon chambers, placed inside the magnet return yoke [108].

However, rather than using θ to describe the angle of a particle with respect to the beam axis, the pseudorapidity η is preferred since it is invariant under boosts along the longitudinal axis, and particle production is also constant over η. It is defined as:

$$\eta = -\ln[\tan(\theta/2)] = \frac{1}{2}\ln\frac{|\vec{p}| + p_z}{|\vec{p}| - p_z},$$

where $\vec{p}$ is the particle momentum and p_z is its component of along the beam axis (i.e. the longitudinal momentum). In the limit where the particle is traveling close to the speed of light, the pseudorapidity converges to the rapidity y:

$$y = \frac{1}{2}\ln\frac{E + p_z}{E - p_z}.$$

In proton-proton events, the actual colliding particles are the quarks contained in the protons. Each quark carries an unknown fraction of the proton energy, and hence the total momentum of the collision remains unidentified. However, since the protons are accelerated only along the z-axis, in the transverse plane (x-y) the total momentum is determined, as it must be zero before and after the collision for the law of conservation of momentum. For a particle with mass m, the transverse energy E_T and the transverse momentum p_T are hence preferred over the corresponding total variables. They are defined as:

$$p_T = \sqrt{p_x^2 + p_y^2},\ E_T = \sqrt{m^2 + p_T^2}.$$

The conservation of transverse quantities has a great importance in the evaluation of the fraction of energy and momentum carried by the neutrinos, which cannot be detected by CMS and whose presence can only be indirectly measured.

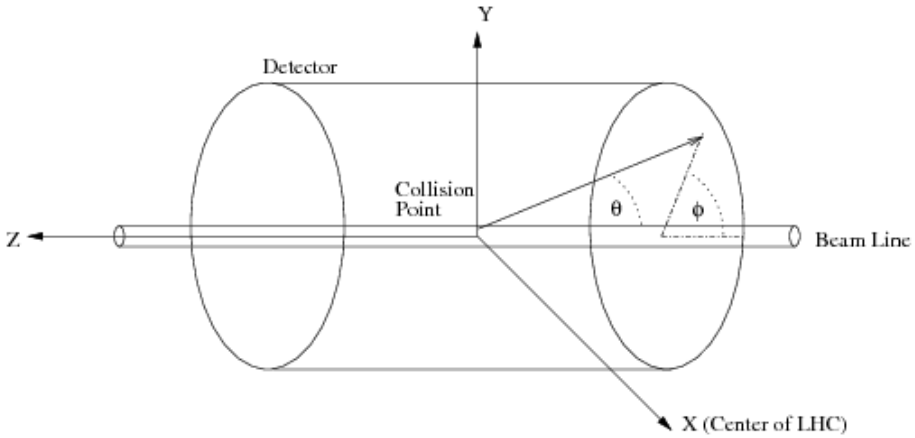

Figure 3.3: Schematic illustration of the coordinate system used by the LHC experiments.

3.2.2 The superconducting magnet

The superconducting solenoid magnet is one of the distinctive features of CMS as well as the backbone of the experiment, with all sub-detectors being supported from it [109]. It is 12.5 m long, with a diameter of 6 m, and it provides a uniform magnetic field of 4 T obtained with a four-layers superconducting solenoid coil. As superconducting material the niobium-titanium (NbTi) was chosen, arranged in Rutherford-type cables co-extruded with pure aluminium for the field stability. The coil is cooled by saturated helium at 4.5 K. The magnetic flux is returned using an iron yoke that allows through only muons and weakly interacting particles, and it is split into five barrel rings, each one having a mass of 1200 t, and three endcap disks (600, 600, and 250 mm thick) [110]. The features of the CMS magnetic field, with the superconducting magnet in the inner part of the detector and the magnetic field lines closed with steel yokes on the outside, makes the muons traveling in the characteristic "s" shaped trajectory, recognizable also on the well known CMS logo.

3.2.3 The silicon tracker

A robust tracking and a good and detailed vertex reconstruction are two fundamental requirements for an experiment as CMS, which is designed to address the full range of physics accessible at the LHC [111]. Two different substructures within the tracker system, the silicon strips tracker and the silicon pixels detector, provide high precision measurement of the charged particles' trajectory up to pseudorapidities of $|\eta| < 2.5$. The pixel detector is the innermost and closest to the interaction point (IP), and consists of four concentric barrel layers, and three disks on each end, with a total silicon area of 1.9 m^2. In the barrel pixel detector (BPIX) 1184 silicon sensor modules are used, while 672 modules are used for the forward disks (FPIX). Each module consists of a sensor with 160 $\times$ 416 pixels connected to 16 readout chips (ROCs), for a total of 124 million readout channels. The sensors are built following the n-in-n approach [112] with strongly n-doped (n+) pixelated implants on an n-doped silicon bulk and a p-doped back side. In a reverse-bias configuration the n+ implants collect electrons, that have higher mobility compared to holes and therefore are less affected by charge trapping caused by radiation damage in the silicon after high irradiation [113, 114], leading to a high signal charge even after a high fluence of charged particles [115]. The outermost part of the tracker consists of silicon strip detectors, arranged as in Fig. 3.4. Four regions are defined: the inner barrel part (TIB) and the outer barrel part (TOB), composed respectively of four and six concentric layer barrel shell structures; the inner disks (TID) made of three disk structures on each

side, each divided in three concentric rings, and the outer end-caps (TEC), made of nine disk structures on each side, each made of four to seven rings. The tracker is composed of 15148 detector modules distributed among the four regions described above, and each module has one or two silicon sensors, for a total of 24244 sensors. The sensors are made of n-type phosphorus doped silicon with p+ single sided strips, with variable resistivity and thickness. The inner parts of the tracker (TIB, TID and the four innermost TEC rings) have been instrumented with sensors of lower resistivity and thickness, while the outer parts of the tracker have higher resistivity and thicker sensor wafers. These differences in the resistivity and sensors' thickness are needed to cope with the effects of irradiation, to avoid a detector current breakdown induced by the higher bias voltage needed for depletion with the increase of the radiation damage [116].

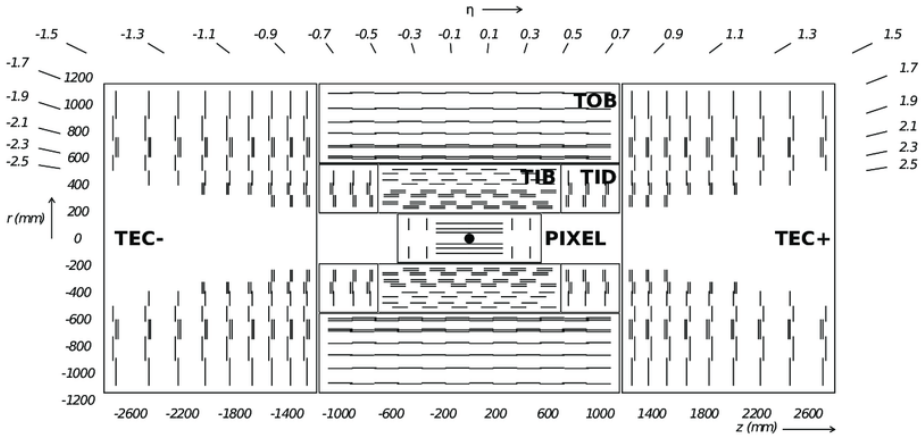

Figure 3.4: Projected view of the CMS tracker layout in the r-z plane [117].

3.2.4 The electromagnetic calorimeter

The Electromagnetic Calorimeter (ECAL) measures the energy of electrons and photons that are produced in electromagnetic eletron/photon-induced showers. It is a hermetic homogeneous calorimeter made of 75848 lead tungstate ($PbWO_4$) scintillating crystals and, as for the tracker, it has a cylindrical central part called barrel, and two endcaps that close the calorimeter at both ends. The lead tungstate has a high density ($\rho = 8.28$ g/cm^3, a short radiation length ($X_0 = 0.89$ cm) and a small Moliere radius ($R_M = 2.2$ cm), making this material ideal to built a compact detector with high granularity. The geometrical crystal coverage extends to $|\eta| = 3.0$, with the barrel section of the detector covering the regions up to $|\eta| = 1.479$ and the endcaps covering the region $1.479 < |\eta| < 3.0$. However, precision energy measurement involving photons and electrons can be carried out only up to $|\eta| = 2.5$. This limit has been determined by considerations on the radiation dose and amount of pileup energy and matches the geometric acceptance of the inner tracking system [118]. The barrel crystals has a 26×26 mm^2 front face area and a length of 23 cm, which corresponds to nearly 26 radiation lengths, while the endcap crystals have an area of 30×30 mm^2 and a length of 22 cm corresponding to nearly 25 radiation lengths. The presence of a silicon preshower detector in the endcap regions, mounted to provide π^0/γ separation over the pseudorapidity range 1.65 to 2.6, allows the use of shorter crystals. A sketch of the section of the ECAL calorimeter is shown in Fig. 3.5. Two types of photodetectors are used: avalanche photodiodes in the barrel, as they can operate in strong transverse magnetic fields, and vacuum phototriodes in the endcaps, to cope with the higher level of radiation. The energy resolution in ECAL can been parameterized as

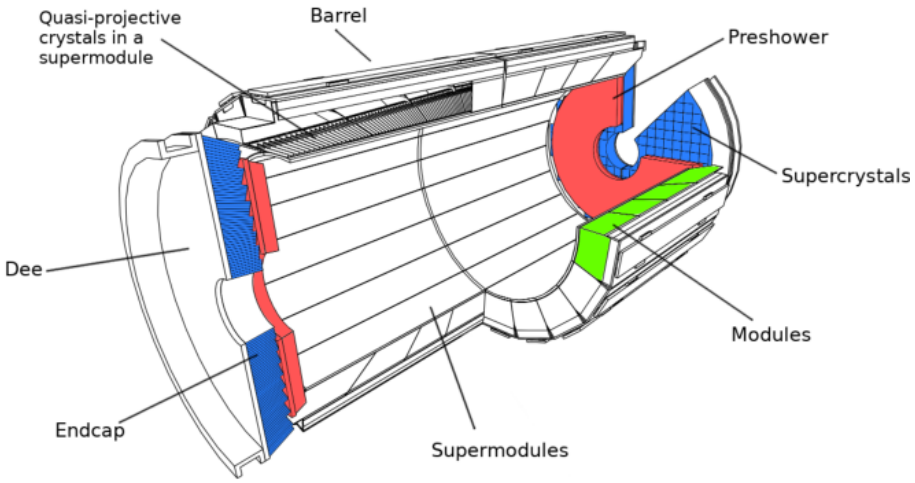

Figure 3.5: A schematic view of the ECAL calorimeter: besides barrel, endcaps and preshower, in the figure are also highlighted the modules, supermodules and supercrystals structures where the ECAL crystals are arranged [118].

[118]:

$$\frac{\sigma_E}{E} = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C = \frac{2.8\%\ \text{GeV}^{1/2}}{\sqrt{E}} \oplus \frac{0.12\ \text{GeV}}{E} \oplus 0.3\%.$$

In the above formula, S is the stochastic term including fluctuations in the shower containment and the contribution from photostatistics, N is the noise term including electronic noise and pileup energy, and C is the constant term. The constant term accounts for non-uniformity of the longitudinal light collection, energy leakage from the back of the calorimeter and single-channel response uniformity and stability, and it dominates the energy resolution for high-energy electron and photon showers. The values of S, B and C values have been extracted from beam test measurements [119]. A correction on the energy deposited in the crystals must be applied to address the radiation-induced effects. The production of color centers induced by irradiation indeed reduces the transparency of the crystals, with this loss increasing with $|\eta|$. To compensate for this effect, a laser monitoring system is used [120]. A reference laser light at 440 nm is injected into the crystals during one orbit gap (calibration gap) of the LHC orbit, and the crystals' response is then used to compute calibration constants, to be then used in the offline reconstruction [121]. In addition, a relative calibration is performed between the responses of the ECAL crystals. Some *pre-calibration* factors are obtained from laboratory measurements, beam tests, and from exposure to cosmic rays, while the final intercalibration is obtained from collision data. Different intercalibration methods can be used. The ϕ-symmetry method is based on the expectation that, for a large sample of minimum bias events, the deposited energy should not depend on the azimuthal angle of the crystal, and hence it provides a fast intercalibration of crystals located within the same η-ring. The intercalibration can be also performed using the invariant mass of photon pairs from π^0 and η mesons, and with isolated electrons from W and Z bosons decays, comparing the energy measured in ECAL to the track momentum measured in the silicon tracker [121].

3.2.5 The hadronic calorimeter

Like electrons and photons, hadrons can also initiate particle showers when they interact with atomic nuclei. By collecting the energy of hadronic showers, the Hadron Calorimeter (HCAL) can measure the energy of hadrons, and also provide measurements of missing transverse energy due to the production of undetected particles like neutrinos.

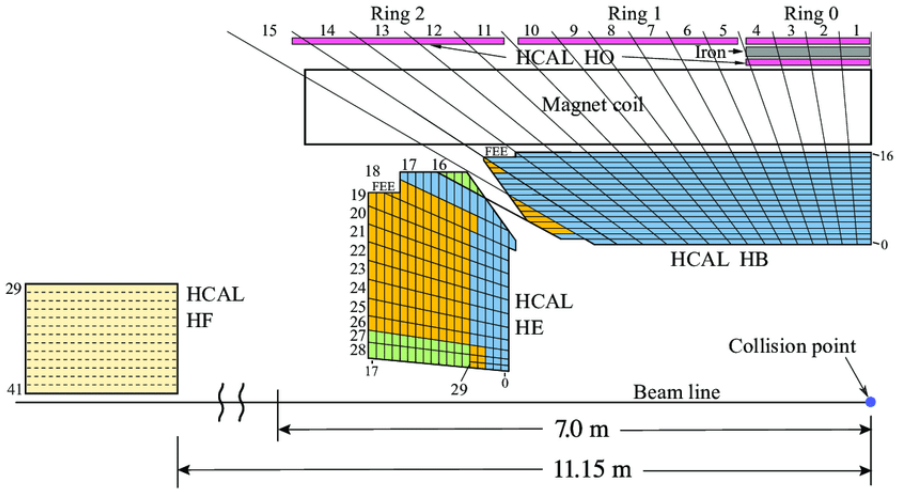

Figure 3.6: A schematic view of one quarter of the CMS HCAL during 2016 LHC operation, showing the positions of its four major components: the hadron barrel (HB), the hadron endcap (HE), the hadron outer (HO), and the hadron forward (HF) calorimeters [126].

It is a hermetic sampling calorimeter divided into four sections: barrel (HB), endcaps (HE), outer hadron calorimeter (HO) and forward hadron calorimeter (HF), covering different space regions for a total of 9072 readout channels [122]. The HB and HE consist of alternating layers of brass and plastic scintillator plates, where HB covers the pseudo-rapidity range $|\eta| < 1.3$, and HE the range $1.3 < |\eta| < 3$. Cartridge brass (70% Cu and 30% Zn) is the absorber material, being non-magnetic. A plastic scintillator is chosen as the active material, giving its long-term stability and radiation hardness, a density of 8.83 g/cm^3 and a nuclear interaction length λ_I of 16.42 cm [122]. The HF is a Cherenkov calorimeter, composed of steel absorber with embedded fused-silica-core optical fibers as active medium, and it covers the region range $3 < |\eta| < 5$. The steel absorber structure is composed of 5 mm thick grooved plates, with fibers inserted in the grooves. The choice of optical fibers was constrained by the exceptional harshness of the radiation conditions faced by the HE, which has to endure a ∼760 GeV energy deposit per proton-proton interaction, compared to only 100 GeV for the remaining parts of the detector [123]. The HF is also used for online luminosity measurements, with two methods for extracting a real-time relative instantaneous luminosity implemented in the HF firmware [124]. The last part of the detector has been built to retrieve the energy which cannot be measured by HB due to its geometrical constraints. The HB is indeed placed between the ECAL barrel and the magnet cryostat and coil, providing a total absorber thickness that corresponds to nearly 6 interaction lengths, increasing with the polar angle up to 10.6 λ_I at $|\eta| = 1.3$. The electromagnetic calorimeter provides about one additional interaction length (independently from η). In order to ensure and adequate sampling depth (at least 11 λ_I) for all the particles detected in HB, and hence to avoid large energy leakage, the HO was placed outside the cryostat. The HO indeed utilises the cryostat and the solenoid coil as an additional absorber, and thick scintillators tiles as active material. The tiles are organized in ring layers and they are placed in front of the first layer of the muon barrel detector [125]. A schematic view of the CMS HCAL, with all its subsystems, is shown in Fig. 3.6.

3.2.6 The muon system

The outermost CMS sub-detectors are the muon chambers, implemented with different technologies in order to correctly detect, identify and measure with highest precision the

momentum of muons produced in a high-energy collision. Similarly to the other subsystems, a barrel region and two endcap regions can be identified within the muon system. Drift tubes (DT) are used in the barrel region, cathode strip chambers (CSC) in the endcap regions, and resistive plate chambers (RPC) in both the barrel and endcap regions [127]. The DT chambers cover the pseudorapidity region $|\eta| < 1.2$ and they are organized into four stations alternated between the plates of the magnet flux return yoke. The basic elements of the DT system are the drift cells, equipped with a gold-plated stainless-steel anode wire at the center and filled with a gas mixture of 85% Ar and 15% CO_2. The drift field is obtained using 4 electrodes: 2 on the side walls of the tube and 2 above and below the wires on the ground planes between the layers. A DT chamber is composed by two superlayers (formed by four staggered layers of parallel cells), measuring the $r - \phi$ coordinates, and an orthogonal superlayer that measures the r-z coordinate, where r is the nominal distance from the beam collision point. The choice of DT was possible due to the expected low muons rate and to the low, uniform magnetic field (below 0.4 T) in the barrel region [128]. At higher pseudorapidities, the higher rates and the non-uniformity of the magnetic field require the employment of more suitable detectors, and hence for covering $0.9 < |\eta| < 2.4$ the CSCs are used. They operate as standard multi-wire proportional counters instrumented with cathode strips, and they use a gas mixture of 50% CO_2, 40% Ar, and 10% CF_4. The CSCs have various physical dimensions but they all share the same trapezoidal shape, and they are organized in rings along the r direction. Every chamber contains six detecting layers, each of them equipped with a plane of radial cathode strips and a plane of anode wires almost perpendicular to the strips, to provide the position measurement. For achieving optimal trigger performances the muon system is also instrumented with RPCs, each one made by two parallel bakelite plates filled with a a gas mixture of 96% $C_2H_2F_4$, 3.5% $i - C_4H_{10}$ and 0.5% of SF_6. The RPCs are able to provide fast response and good time resolution even in a high pileup environment or for high background rates. The layout of the muon system is shown in Fig. 3.7.

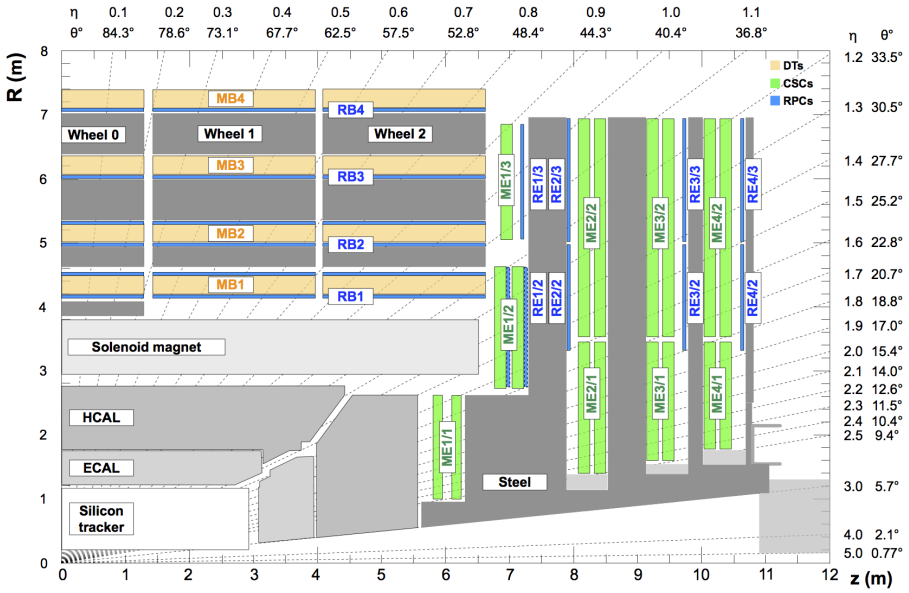

Figure 3.7: An R-z cross section of a quadrant of the CMS detector with the axis parallel to the beam (z) running horizontally and the detector's radius (R) increasing upward. The drift tube stations (DTs) are labeled MB ("Muon Barrel") and the cathode strip chambers (CSCs) are labeled ME ("Muon Endcap"). Resistive plate chambers (RPCs) are mounted in both the barrel and endcaps of CMS, where they are labeled RB and RE, respectively [128].

3.2.7 Triggering and data acquisition

The proton-proton interaction rate provided by the LHC, with proton bunches colliding 40 million times per second, is not manageable by the current computer facilities and only a fraction of all the events produced in the collisions can be stored for subsequent analysis. In addition, an effective selection of the events originated from hard processes is required, because at the LHC the dominant processes are those coming from soft interactions, that could easily saturate the available resources at the expense of the hard interactions. For this reason, CMS has developed a two-levels trigger system. The first level (L1) is implemented in custom hardware and operates online, restricting the output rate to 100 kHz. To decide which events have to be kept or discarded, the L1 trigger exploits the so-called *trigger primitives* (TP). The TPs are built starting from the energy deposits in ECAL and HCAL, and from the hits collected in the muon system. They are then processed in several steps, with the information coming from the calorimeters and the muon system encoded in a final Global Calo Trigger (GCT) and a Global Muon Trigger (GMT), respectively. In the final step, the information from GCT and GMT are combined into a Global Trigger (GT) which makes the ultimate decision of whether keep or reject the event [129]. A schematic overview of the L1 trigger system is shown in Fig. 3.8. The second level (high level trigger, HLT) operates offline, and it is implemented in software. The HLT is structured around the concept of a *HLT path*, that consists in a set of ordered algorithms whose aim is to further refine the purity of the L1 objects, reducing the rate down to 1 kHz [130] for the final storage. The data acquisition system (DAQ) is responsible

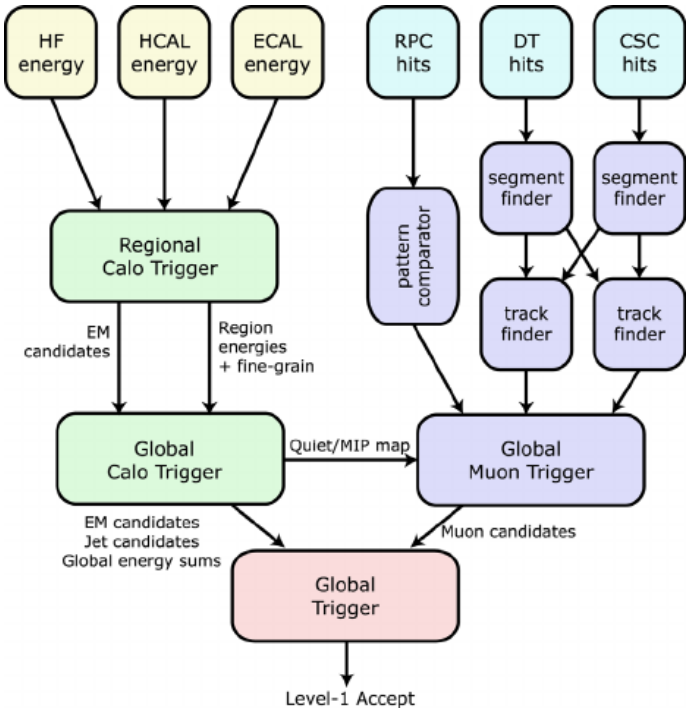

Figure 3.8: Overview of the CMS L1 trigger system. Data from the hadronic and from the electromagnetic calorimeter are processed first regionally (RCT) and then globally (GCT). The hits collected in the muon system are processed and sent onwards to a global muon trigger (GMT). The information from the GCT and GMT is combined in a global trigger (GT), which makes the final trigger decision [131].

for transporting event data to the HLT farm, where the HLT event filtering operates, and then to the Tier-0 computing center at CERN [132]. Its main components are:

- Custom electronics modules, the Frontend Readout (Optical) Link boards, receiving

data from the detector back-end boards that in CMS are called Frontend Driver (FED) boards.

- Readout Unit (RU) servers that aggregate data from about 10 FEDs into the so-called *super-fragments*.
- Builder Unit (BU) servers, whose aim is to build the full events from the *super-fragments*.
- Filter Unit (FU) servers connected to BUs, that execute the HLT algorithm.
- A Storage and Transfer System (STS), that merges events that have passed the HLT into larger files and transfers them to the Tier-0 at CERN.

In addition, a Trigger Throttling System (TTS), consisting of custom electronics modules, is able to collect fast status information from all the FEDs in order to throttle the trigger and avoid buffer overflows.

3.3 The Particle Flow algorithm for particle reconstruction

The proton-proton collisions taking place at the LHC produce a shower of particles that travel across the subsystems of the CMS detector, interacting with the different sub-detectors according to their charge, energy and type. Starting from the beam interaction region, particles first enter the CMS tracker. There, the trajectories (tracks) of charged ones and their origins (vertices) are reconstructed starting from the hits collected in the tracker layers, and the magnetic field allows the measurement of particles' charge and momentum. Right after, electrons and photons are fully absorbed in the electromagnetic calorimeter, where the measurement of their energy and direction is obtained from the energy deposits in the calorimeter cells. In a similar way, the energy and the direction of charged and neutral hadrons is estimated from the energy deposits in the hadron calorimeter, where the hadronic showers are fully absorbed. Muons are then detected in the outermost part of the detectors, where the muon chambers are located, while neutrinos escape undetected, and their energy can be only inferred using indirect methods. In Fig. 3.9, a sketch of a transverse slice of the CMS detector summarizes the above description of the particles interactions in the different subsystems. Starting from this simplified view, a complete description of a physics event can be obtained by combining the information from all the sub-detectors, in order to identify and reconstruct the properties of the detected particles. This *holistic* approach is called particle-flow (PF) reconstruction [133], and it was first developed in the 1990s for the ALEPH experiment at LEP [134]. Given the fine spatial granularity of the CMS detector, the high magnetic field which allows good separation between neutral and charged hadrons, and the excellent performances in the identification of the muon tracks, the PF algorithm is able to deliver a global description of the event with high resolution and efficiency, as well as reduced misidentification rate. In the following sections, the reconstruction of the main object needed for the $Z\gamma$ analysis are discussed, starting with the description of the particle-flow elements as extracted from the different subsystems.

3.3.1 Tracking and vertexing

Track reconstruction refers to the process of using the hits collected in the tracker layers to obtain estimates for the momentum and position parameters of the charged particles traversing the detector. The procedure adopted by CMS is called *iterative tracking* as each track is reconstructed by multiples steps of the Combinatorial Track Finder (CTF) algorithm [135]. The process starts from well recognised tracks, for example tracks that have large p_T and are produced near the interaction region. At each iteration, the number

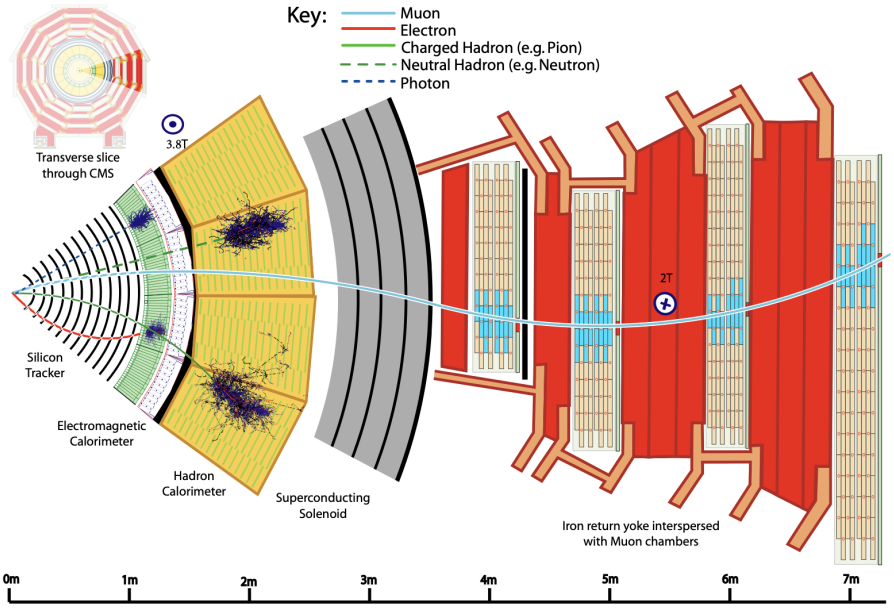

Figure 3.9: A sketch of the specific particle interactions in a transverse slice of the CMS detector. The muon and the charged pion are positively charged while the electron is negatively charged [133].

of hits associated with the tracks is reduced, thus relaxing the selection criteria. In this way, the iterative procedure is able to reduce the combinatorial complexity that could lead to misrecontructed objects when looking at more difficult classes of tracks (with low-p_T, or displaced tracks). Each iteration consists of four steps [135]:

- **Seed generation**: it provides an initial set of track candidates and trajectory parameters using only a few hits. In particular, the seed generation relies on the hits collected on the four layers of the pixel detectors and it exploits an algorithm called Triplet Propagation [136]. The algorithm starts from one combination of four pixel layers, and then it looks for pairs of hits compatible with the region configuration created between the first two layers in the combination. Every created pair is propagated to the third layer where a third hit is searched, in order to build a triplet. Similarly, every triplet is then propagated to the fourth layer in order to build quadruplets. The Cellular Automaton (CA) technique [137] is used to look for hit pairs, forming hit triplets and quadruplets after checking the seed compatibility with the beam spot. A scheme of the CA track seed automaton is shown in Fig. 3.10.
- **Track finding**: it is based on the Kalman filter algorithm [138]. It starts with a first estimation of the track parameters provided by the seed, and then it builds candidates by adding hits from successive detectors layers, updating the parameters at each layer. It continues until there are no more layers, or there is more than one missing hit. Besides the location and the uncertainty of the detected hits, the algorithm takes also into account the amount of material crossed, which is used to estimate the effects of multiple Coulomb scattering and energy loss.
- **Track fitting**: once the full trajectory information is available and all its hits are known, the best possible parameters of each track are estimated by a refitting procedure.
- A set of *quality selections* is then applied, to reject the possible fake tracks (tracks reconstructed from a combination of unrelated hits, or genuine particle trajectory whose reconstruction is flawed due to the inclusion of spurious hits). The main

selection criteria involve the number of layers that have hits, the χ^2/dof of the track fit and the compatibility with the primary vertex.

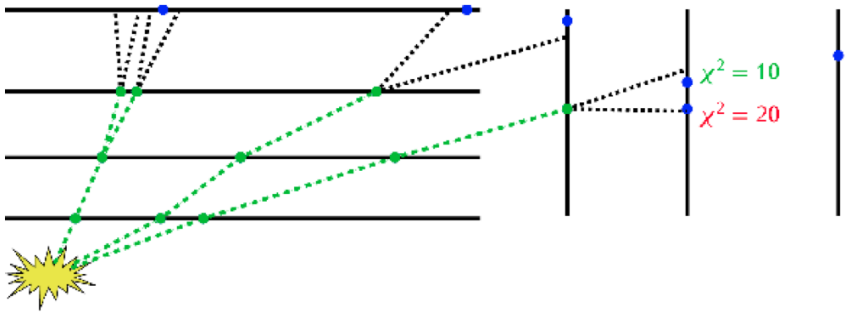

Figure 3.10: The scheme of the Cellular Automaton track seed algorithm when propagating a triplet to the fourth pixel layer [136].

At the end of the iterative procedure, all the selected tracks are clustered using a deterministic annealing algorithm [139] that takes into account the z-coordinates of the points of closest approach of the tracks to the center of the beam spot and their associated uncertainties. This clustering allows the reconstruction of all proton-proton interaction vertices of each event, including the signal vertex and any vertices from pileup collisions [135]. Once the vertex candidates have been identified, those containing at least two tracks are fitted via an adaptive vertex fitter [140]. The adaptive vertex fit associates a weight to each track according to the probability they really belong to the vertex, and computes the best estimate of the vertex parameters. The tracking efficiency is measured in data applying the *Tag and Probe* technique to events with a Z boson decaying into a pair of muons. The tag muon is reconstructed using both the information from the muon chambers and the tracker, and it has to satisfy stringent selection criteria. The probe muon is instead reconstructed using only the information from the muon sub-detectors, and it passes looser identification requirements. If a probe can be matched to at least one track in a cone ($\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.3$) around it, it is classified as a passing probe. Otherwise, it is classified as a failing probe. The efficiency is then evaluated as the ratio between the number of passing probes and the total probes. As can be observed in Fig. 3.11, an excellent tracking efficiency of about 99.9% for tracks associated with muons is guaranteed over the whole muon pseudorapidity acceptance for all the Run 2 data.

3.3.2 Calorimeter clustering

The main purpose of the clustering algorithms is to estimate the energy of particles crossing the calorimeters by collecting the energy deposits released inside ECAL, HCAL and ES. In addition, such algorithms are also able to evaluate the direction of neutral particles (photons and neutral hadrons) that do not leave hits in the tracker, to discriminate between deposits originated by neutral and charged particle, and to correctly identify the electromagnetic showers produced by electrons, collecting the accompanying bremsstrahlung photons. The energy measurement of charged hadrons is also used to improve the accuracy of the parameters of low quality tracks. The clustering algorithms run separately in the barrel and endcap regions of ECAL and HCAL, and in the ES. The clustering procedure starts by identifying a *cluster seed*, which is a calorimeter cell whose energy exceeds a given threshold and it is larger than the energy of the neighboring cells (the four closest cells sharing a side with the seed candidate, or the eight closest cells sharing a corner with the seed candidate). Then, a *topological clustering* is performed to

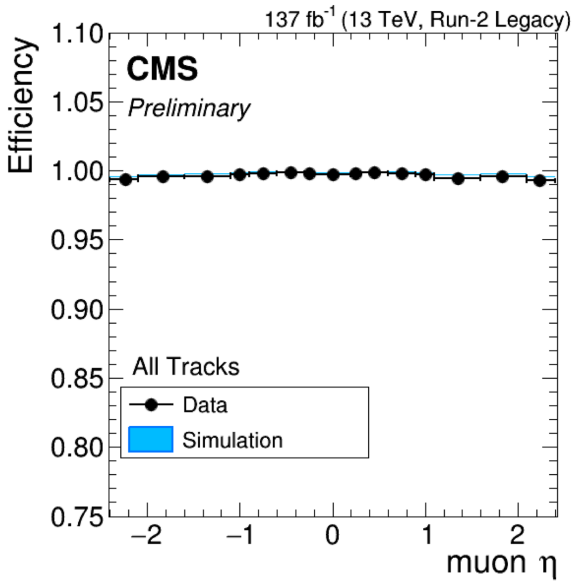

Figure 3.11: The tracking efficiency as a function of the muon pseudorapidity. The efficiency in data (black dots) is compared to that obtained using a simulated $Z \to \mu^-\mu^+$ sample (blue boxes) [141].

aggregate cells that share at least a corner in common with the cells already in the cluster, and whose energy exceeds a second threshold, set to twice the noise level. The ultimate energy and position of the clusters are defined by using a dedicated algorithm that relies on the assumption that the energy deposits in the cells of the topological cluster arise from Gaussian energy deposits in the cluster seeds. An example of ECAL and HCAL clustering are shown in Fig. 3.12.

3.3.3 The Link algorithm

According to its charge and type, a particle crossing the different CMS sub-detectors can produce different objects (tracks and clusters) for each layer it passes through. Then, in order to obtain the final reconstruction and the ultimate identification of the particle, the information carried by these objects has to be combined. This is achieved by using a dedicated *link algorithm* that connects the reconstructed tracks and clusters originated by the same particle, producing the so-called *PF blocks*. The specific conditions required to link two elements are described in the list below [133]:

- **Track-cluster link**. The link between a track and a calorimeter cluster starts by extrapolating the track into the calorimeters using its last measured hit. For ECAL, its position is evaluated at a depth corresponding to the maximum of a typical electron shower, and at a depth corresponding to one interaction length for HCAL. The extrapolated position is then enlarged by up to the size of a cell in each direction, to account for the presence of gaps or cracks in the calorimeters and for the uncertainty in the position of the shower maximum. If the extrapolated position is found to be within a cluster, the track is matched to it. If different clusters are linked to the same track, or if several tracks are linked to the same cluster, the distance between the extrapolated track position and the cluster position, evaluated in the (η, ϕ) plane, is used to quantify the quality of the links, keeping only the link with the smallest distance.
- **ECAL cluster-track link**. In order to link the clusters produced by bremsstrahlung photons to their corresponding electron track, the tangent to the electron trajectory

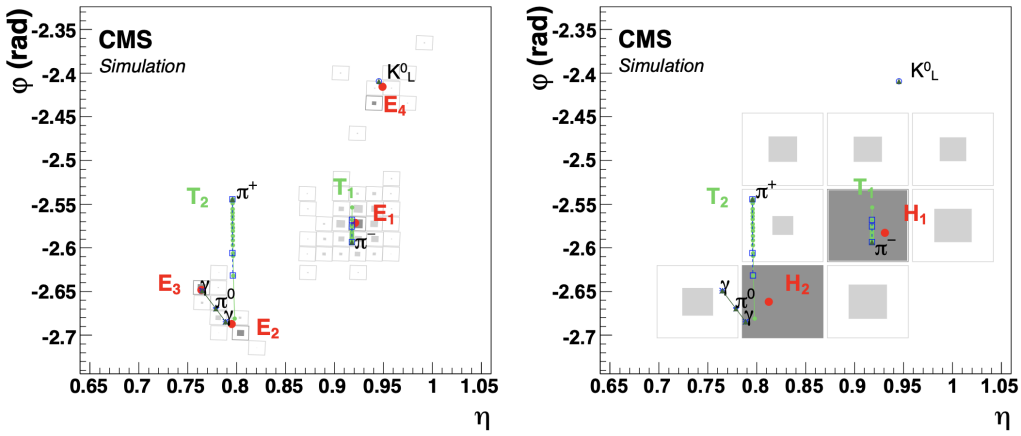

Figure 3.12: ECAL and HCAL clusters originated by a jet made of five particles (K^0_L, $\pi^+, \pi^-, \pi^0(\rightarrow \gamma\gamma)$). The ECAL and HCAL cells are represented as squares, with an inner area proportional to the logarithm of the cell energy. Cells with an energy larger than those of the neighboring cells are shown in dark grey, the cluster positions are represented by red dots, the simulated particles by dashed lines, and the positions of their impacts on the calorimeter surfaces by open markers. (Left) Four separated ECAL clusters originated by the K^0_L, the μ, and the two photons from the π^0, denoted as E_1, E_2, E_3, E_4. The π^+ does not create a cluster in the ECAL. The two charged π are also reconstructed as charged-particle tracks T_1 and T_2. (Right) Two HCAL clusters, originated by the two charged π and denoted as H_1 and H_2 [133].

is extrapolated to the ECAL for each layer of the tracker. If the extrapolated tangent position falls within the boundaries of the cluster, the cluster is associated to the original track as a potential bremsstrahlung photon.

- **Cluster-cluster link**. A link can be also established between HCAL and ECAL clusters, or between ECAL and preshower clusters. The connection is performed when the cluster position in the more granular calorimeter (preshower or ECAL), is within the cluster boundary in the less granular calorimeter (ECAL or HCAL). As for the track-cluster linking, in case multiple clusters are matched to the same one, the one with the shortest $\eta - \phi$ distance (for HCAL-ECAL link) or $x - y$ distance (for an ECAL-preshower link) is kept.
- **Track-track link**. For nuclear-interaction reconstruction, further links are built for charged-particle tracks that could share a secondary vertex. In this case, at least three tracks must be associated together through a common displaced vertex, where at most one of the tracks is the incoming track.
- **Muons**. A link between a track in the tracker and the information in the muon chambers is also established, in order to define the different muon classes.

3.3.4 Particle reconstruction

Electrons and photons

The algorithm for the reconstruction of electron and photons starts from ECAL clusters, formed as described in Sec. 3.3.2. The clusters are formed by grouping neighboring crystals whose energies exceed a predefined threshold, typically around 80 MeV for EB and 300 MeV for EE, chosen to be 2 or 3 times bigger than the expected electronic noise. Among all the clusters, the one containing most of the energy (with a minimum transverse energy above 1 GeV) is taken as the *seed cluster*. When propagating through the material

in front of ECAL, electrons and photons have a significant probability of showering, producing bremsstrahlung photons and e^-e^+ pairs that also could start new electromagnetic showers. The original object could thus consist of several others electrons and photons, and hence the multiple ECAL clusters they originate need to be combined into a single *supercluster* (SC) that captures the energy of the original electron/photon. The superclustering procedure employs two algorithms. The first one is called the *mustache* algorithm, since it aggregates clusters that fall into a shape similar to a mustache around the seed cluster, as shown in Fig. 3.13. The second superclustering algorithm is known as the *re-*

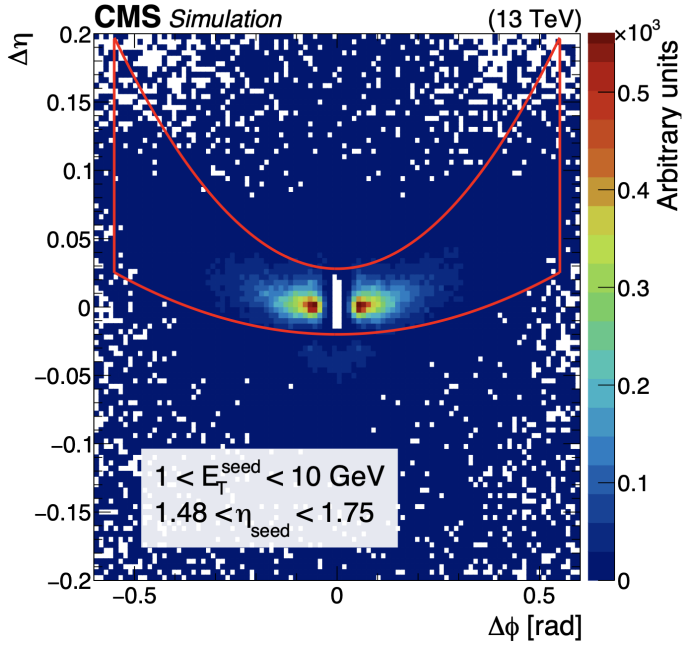

Figure 3.13: Distribution of $\Delta\eta = \eta_{seed\ cluster} - \eta_{cluster}$ versus $\Delta\phi = \phi_{seed\ cluster} - \phi_{cluster}$ for simulated electrons. The energy of the seed cluster is between 1 and 10 GeV and $1.48 < \eta_{seed\ cluster} < 1.75$. The red line contains the clusters selected by the mustache algorithm to form the SC and the white region at the center of the plot represents the seed cluster [142].

fined algorithm, and it uses the tracking information to further refine the SC, minimizing the risk of including spurious clusters.

The energy deposited by electrons and photons as collected by the superclustering algorithm is still subject to several losses, for lateral and longitudinal shower leakage, intermodule gaps and dead crystals, or possible energy deposits in the tracker material. This leads to systematic variations of the energy measured in ECAL that could significantly degrade the energy resolution for the reconstructed particles. A multivariate technique is used to correct the energy estimation, based on a set of regression fits targeting the ratio y between the reconstructed and true energy of an electron or a photon. The regression employs the following log-likelihood function:

$$\mathcal{L} = -\sum_{e/\gamma} \ln p(y|\vec{x})$$

where $p(y|\vec{x})$ gives the estimated probability for a reconstructed electron or photon to have the observed value y, given a vector of input electron and photon variables $\vec{x}$. The regression is performed in simulated samples, where the true values of the energies are known, and the probability density function used is a double-sided Crystal Ball (DSCB) [143], which has a Gaussian core with power law tails on both sides. The most probable value of

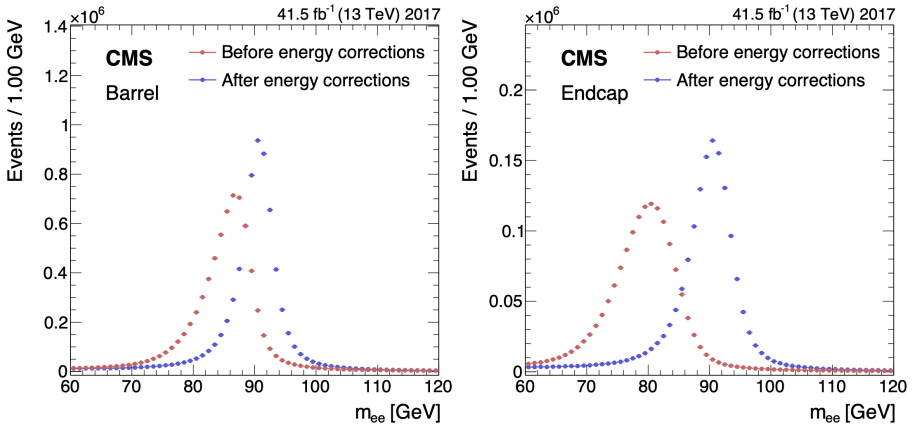

Figure 3.14: Dielectron invariant mass distribution before and after the application of the energy and scale corrections for barrel (left) and endcap (right) electrons for $Z \to e^- e^+$ events [142].

the DSCB estimated by the algorithm provides the energy corrections, while the energy resolution is extracted from the width of the Gaussian core of the distribution. An additional correction factor is applied to electron and photon energy scales, to take into account possible residual differences between data and simulation. Two dedicated methods, the *fit method* and the *spreading method*, have been developed to estimate this correction, using $Z \to e^- e^+$ events. The fit method performs a fit to the invariant mass distribution of the Z boson, obtained from data and from simulated events, and extracts a scale offset by comparing the results. On the other hand, the spreading method takes the simulated Z boson invariant mass distribution as a probability density function in a maximum likelihood fit to the data, and evaluates a correction function to be applied to the simulation. A multistep procedure based on both methods is implemented to derive the ultimate corrections on the energy scales [142]. The importance of the energy resolution and scale correction can be clearly observed in Fig. 3.14, where the distributions of the reconstructed $Z \to e^- e^+$ mass is shown, before and after applying the corrections. The SC information is also exploited in the iterative tracking procedure at the fitting step. As explained above, the selected tracks are refitted once all the hits in the tracker layers are known, in order to achieve the best track parameters. A Gaussian-sum filter (GSF) [144] is used instead of the Kalman filter, as it allows for sudden and substantial energy losses along the trajectory and hence it better suits the reconstruction of electrons' tracks. However, this is a very CPU consuming process and therefore it cannot run over all the reconstructed hits in the particles. The reconstruction of electron tracks therefore begins with the identification of a *seed*, a hit pattern selected in order to lie on an electron trajectory, and that can be either ECAL-driven or tracker-driven. The ECAL-driven seeding first selects SCs with transverse energy greater than 4 GeV and H/E smaller than 0.15, where E and H are, respectively, the SC energy and the sum of the energy deposits in the HCAL towers within a cone of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.15$ centered on the SC position. The ϕ and z position of the SC are then used to extrapolate the trajectory of its corresponding electron, assuming it to be helical. If the hits of a tracker seed are matched to the predicted trajectory, this is selected for seeding a GSF track. On the other hand, the tracker-driven approach iterates over all generic tracks, as extracted by the Kalman filter, and then it looks for compatibility with an ECAL SC. The compatibility is established applying selections on the track quality and track-cluster matching variables. The ECAL-driven approach performs better for high-p_T isolated electrons, while the tracker-driven approach is designed to improve efficiency for low-p_T or non isolated electrons [142]. The improvement in the

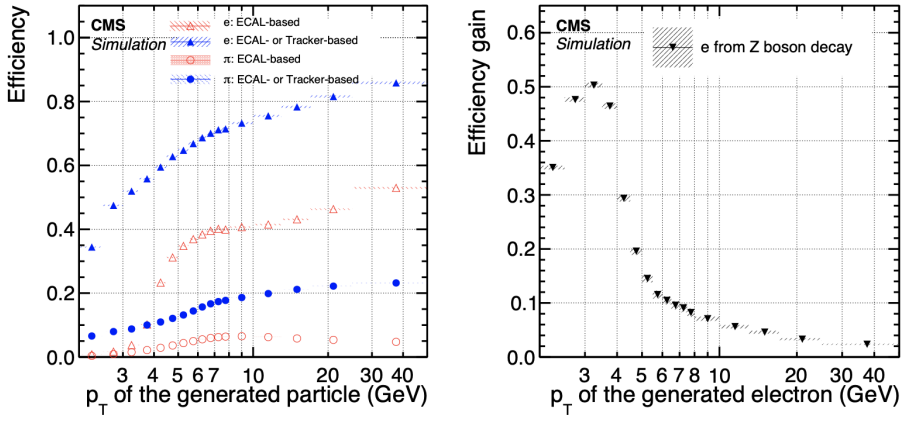

Figure 3.15: (Left) Electron seeding efficiency for electrons (triangles) and pions (circles) as a function of their transverse momentum, from a simulated sample enriched in b quark jets with at least one semileptonic b hadron decay. Both the efficiencies for ECAL-based seeding only (hollow symbols) and with the tracker-based seeding added (solid symbols) are shown. (Right) Efficiency gain from the tracker-based seeding for electrons from Z boson decays as a function of p_T [133].

seeding efficiency for electrons in b quark jets (non isolated electrons) is shown on the left in Fig. 3.15. The addition of the tracker-based seeding increases the probability that a charge hadron gives rise to an electron seed, but the PF algorithm is able to take care of the possible misidentification at later stages of the reconstruction. For isolated electrons, the tracker-based seeding also improves the overall efficiency by several per cent (Fig. 3.15 right) allowing to reconstruct electrons with p_T below 4 GeV.

Hadrons and jets

Hadrons deposit their energy in both the ECAL and the HCAL calorimeters. The HCAL clustering is performed as described in Sec. 3.3.2, where the energy threshold of the HCAL cell for being considered by the clustering algorithm is 800 MeV. Both the ECAL and HCAL cluster energies need to be substantially recalibrated in order to get a good estimate of the true hadron energy, given that ECAL is calibrated for photons, and the HCAL response is not linear and depends on the energy deposited in ECAL. At the LHC, hadrons can originate from jet fragmentation and hadronization, and they can be detected as charged hadrons, neutral hadrons, and non isolated photons (deriving from π^0 decays). If the ECAL and HCAL clusters are not matched to any track, the particle would be usually reconstructed as a photon or a neutral hadrons. However, beyond the tracker acceptance or if the track association fails, a charged hadron could be identified with a neutral one and its energy wrongly associated to a non isolated photon. Since hadrons leave the 25% of the jet energy in the ECAL, this incorrect identification could heavily affect the jet reconstruction. For this reason, the algorithm has been refined so that ECAL clusters without an associated track are reconstructed as photons only if they are not linked to HCAL clusters [133]. Otherwise, the algorithm identifies the particle as a neutral hadron. Starting from PF candidates, jets are then clustered using the anti-k_T algorithm [145]. First, the algorithm calculate the distance between two particles i and j:

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

where R is a radius parameter that characterizes the method, ΔR is the distance between the two particles in the $(\eta - \phi)$ plane and k_T denotes the magnitude of the particles' transverse momentum. The distance between the i^{th} particle and the beamline B is then

evaluated as:

$$d_{iB} = 1/k_T^2(i)$$

and the minimum between d_{iB} and d_{ij} is calculated. If this minimum is d_{iB}, the algorithm removes the particle i from the list of candidates and defines the object as a "jet". If the minimum is d_{ij}, the momenta k of the particle i and j are summed to construct a new particle, and the algorithm is iterated again until only jets are left. Since no particle with $d > R$ can be counted, the jet is reconstructed as cone around the original particle direction, with R being the radius of this cone.

Pileup measurement and mitigation tecniques

At the luminosities delivered by the LHC, multiple proton-proton interactions can occur at the same time when two proton bunches collide. During the Run 2 the average number of these additional collisions was 40, reaching the average value of 50 for significant fraction of events during the 2017 data taking. The separation of the proton-proton interaction of interest from these overlapping collisions, that go under the name of pileup events, becomes then a crucial factor to associate the reconstructed particles to the correct vertex, and therefore to the correct event. With this purpose, several techniques for mitigating the effect of pileup on the particle reconstruction have been developed at CMS. The most widely used is the Charged-Hadron Subtraction (CHS), which is the standard technique employed to suppress the contribution of the so-called in-time pileup events, the additional collisions taking place in the same bunch crossing of the primary vertex of interest. The CHS algorithm first orders the primary vertices according to the quadratic sum of the p_T of their associated tracks. The one with highest $\sum p_T^2$ is taken as the hard-scatter vertex and all the others as pileup vertices. With this classification of the vertices, all the charged hadrons whose track is associated to one pileup vertex are excluded from the list of physics objects participating in the particle reconstruction. Charged hadrons accounts for around two thirds of the total pileup contribution, while the remaining fraction consists of photons, neutral hadrons and particles reconstructed outside the tracker acceptance. The impact of these particles is addressed exploiting the uniformity of the p_T-density of pileup in the (η, ϕ) plane, and then subtracting this contribution when reconstructing the physics objects [133]. Another approach is the PileUp per Particle Identification (PUPPI), that sums over all the particles j in the events and defines a *shape* α for each particle i as [146]:

$$\alpha = log \sum_{j, \Delta R_{ij} < R_0} \left(\frac{p_T^j}{\Delta R_{ij}} \right)^2 ,$$

where ΔR_{ij} is the distance between the particles i and j, and $R_0 = 0.4$. The α value is compared to the median of the charged pileup distribution in order to extract a weight for each particle between 0 (the particle comes from the hard-scatter vertex) and 1 (the particle comes from a pileup event). The four-momentum of the particles is then rescaled according to the weight, discarding particles with very small weights or rescaled p_T, and a final set of pileup-corrected particles is obtained [147].

Jet energy resolution and corrections

In CMS the jets are reconstructed by applying the anti-k_T algorithm mentioned in the previous section, starting from the PF candidates within the chosen cone of clusterization. However, like all the other experimentally-reconstructed physics objects, the non-linearity of the detector response and several other physics aspects (pileup, discrepancies in data and simulation) need to be considered to obtain the correct jet energy value. A set of Jet Energy Corrections (JEC) has been developed, factorizing the different contributions that could affect the jet energy estimation. The fixed sequence of the JEC is illustrated in Fig. 3.16, where the output of each step (level) is used as input to the next one, and

the different levels of correction are described in the list below. Each one provides a scaling factor to the jet four momentum, depending on various jet related quantities (p_T, η, flavor).

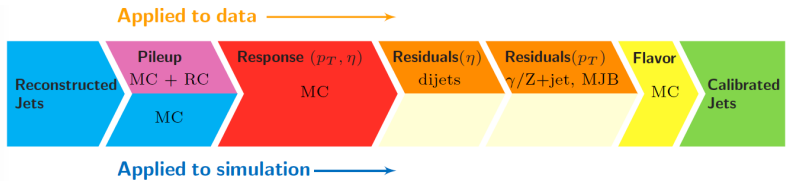

Figure 3.16: Levels of JEC for data and MC simulation, to be applied to jets as reconstructed by the anti-k_T algorithm in order to obtained the final calibrated jets. The corrections marked with MC are derived from simulation studies, RC stands for random cone, and MJB refers to the analysis of multijet events [148].

- **Level 1 (L1: pileup offset)**. This correction takes into account the contribution of the pileup events, differentiating the aformentioned in-time pileup from the out-of-time pileup, the latter arising from proton-proton collisions that occur in the previous or in the subsequent bunch crossing with respect to the reference one. The in-time pileup is corrected using the CHS algorithm described in the previous section. For the out-of-time contribution, the signal time-integration window is shortened. Corrections for residual differences between data and detector simulation are determined using the random cone method [149].

- **Level2-Level3 (L2L3: MC-truth corrections)**. These corrections are derived as a function of the jet p_T and η, to make the jet energy response uniform over these two variables. First, an accurate and detailed description of the jet response is provided from simulation; then, the reconstructed jets are matched to particle-level jets and the JEC are derived by comparing the corresponding η and p_T. The definition of particle level jets refers to those jets that are built from simulation by applying the jet clustering algorithms to all stable particles, excluding neutrinos.

- **Residuals**. Residuals corrections are meant to correct for remaining small differences (of the order of %) within jet response in data and simulation, and are obtained as a function of η and p_T using the simulated samples mentioned in Fig. 3.16.

- **Level 5 (L5: flavor corrections)**. The jet flavor is defined in terms of the flavor of the parton that originated the jet, while jets resulting from hard gluon radiation have an undefined flavor. Possible differences in the energy response for different jet flavors arise mainly from variations in the fragmentation energy and variations in its particle composition. L5 corrections are derived similarly to the L2L3 ones, but using simulated samples that account for the various flavors.

The jet response, defined as the average value of the ratio of the transverse momentum of the reconstructed jet over the transverse momentum of the particle-level jet, is shown in Fig. 3.17 before and after the use of the JEC. For detailed measurements of the SM jet production, as well as for new physics searches and analysis that rely on well-reconstructed missing momentum, the jet p_T resolution is also of fundamental importance. However, the resolution has been measured to be worse in data than in simulation and hence the use of some scaling factors on MC jets is needed, in order to match the resolution observed in data. After the JEC, the jet p_T resolution in data is evaluated with two techniques: the dijet asymmetry methods, in events containing two jets, and the $\gamma+$jets balance method in $\gamma+$jets events. In the first case, the resolution is extrapolated from the scalar balance

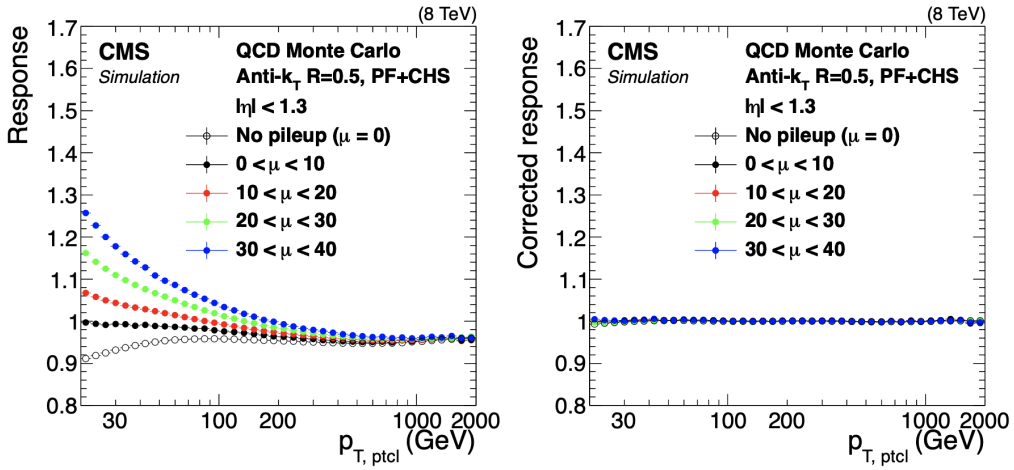

Figure 3.17: Average ratio of the reconstructed jet p_T over the particle-level jet p_T in QCD simulation, as a function of the particle-level jet p_T. The jets are reconstructed with the anti-k_T clustering algorithm (with R=0.5) and the CHS method for pileup mitigation. The ratio is shown before any corrections (left) and after (right) all the stages of the JEC, while μ denotes the average number of pileup interactions per bunch crossing [148].

between the transverse momenta of the two jets, while in the second case the measurement is performed by comparing the transverse momentum of the photon with that of the jet. The scaling factors obtained with the two methods at 8 TeV are shown in Fig .3.18. As can be observed, the discrepancies between data and simulation can increase up to 40% for very forward ($|\eta| \sim 3$) jets.

Muons

The CMS detector is able to reconstruct muons with high efficiency over the full detector acceptance. A high purity is guaranteed by the inner calorimeters and the steel layers of the return yoke that absorb all the other particles (except neutrinos), while the tracker provides a very precise measurement of the muons' momentum. In addition, the hits measured in the muon chambers can be used together with the information coming from the tracker in order to extract the muon trajectories, leading to three different classes of particles:

- **Standalone muons.** They rely entirely on the information of muon chambers. The hits collected in the DT or in the CSC detectors are used to form segments that act as track seeds for the pattern recognition in the other detectors. The final trajectory is then built using DT, CSC and RPC information only, by means of the Kalman filter.
- **Global muons.** They are built using an *outside-in* approach. The standalone track obtained exclusively from the muon chambers is matched to a track as built in the tracker (hereafter referred to as an inner tracks), and then a combined fit is performed using the Kalman filter.
- **Tracker muons.** They are built using an opposite *inside-out* approach with respect to the global muons tracks. The inner tracks are propagated to the muon system by matching them to CSC or DT segments, and then the full track is obtained by a combined fit, as for the global muons tracks.

Reconstruction of global muons requires the particle to cross more than one detectors plane in order to form a segment in the muon chambers, and then it provides higher efficiency for muons whose transverse momentum exceeds about 200 GeV. For momenta below about 10

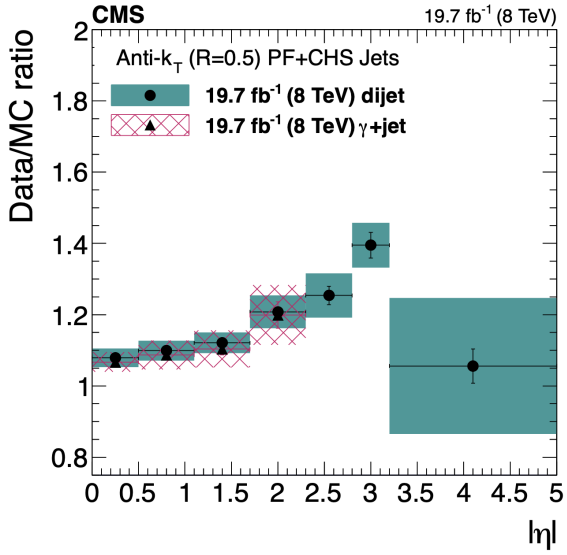

Figure 3.18: Distribution of the scaling factors for the jet p_T resolution as a function of the absolute jet pseudorapidity. The jets are reconstructed with the anti-k_T clustering algorithm (with R=0.5) and the CHS method for pileup mitigation. Results determined from γ + jets (hatched boxes) and dijet data (solid boxes) are shown [148].

GeV, the multiple scattering in the steel of the return yoke leads for the majority of muons not to cross more than one plane, and therefore the tracker muon reconstruction is more efficient. Starting from the three muons classes, various methods are employed to fit the tracks [150], and the ultimate momentum of the muon is determined by applying the *Tune-P* algorithm [151] which selects the best fit according to the goodness-of-fit information and the momentum resolution. Further selections are applied in order to optimize the muon identification (ID) accordingly to the needs of the different CMS physics analysis. The main variables taken into account are the χ^2 of the track fit, the number of hits per track and the quality of the matching between the inner tracks and the muon tracks. To discriminate the muons produced in the collision from those coming from weak decays in jets (punch-through), some requirements on the muon isolation are also applied, where the isolation is defined by summing up the energy around the muon direction, using either the reconstructed tracks or the neutral particles and charged hadrons reconstructed by the PF algorithm [150]. The efficiency for the reconstruction and identification of muons is shown in Fig. 3.19, for two different IDs.

Missing transverse energy

Given their exceptionally small interaction cross section (see Fig. 3.20), neutrinos produced in a proton-proton collision can cross and escape all the detector's layers without interacting. However, their presence can be indirectly estimated by exploiting the imbalance of the total transverse momentum, where the transverse momentum p_T for a single particle is defined as in Sec. 3.2.1. Once all the particles in the events have been reconstructed, for the law of conservation of momentum the obtained total transverse momentum must be zero, as it is zero for the system of two quarks colliding. If this does not happen it means that in the collision one or more neutrinos have been produced, with a transverse momentum corresponding to the quantity that would be necessary to restore the zero-value balance. This transverse momentum fraction associated to the presence of neutrinos is called missing transverse energy (MET), and it is defined as:

$$E_T^{miss} = -\sum_i |\vec{p_{T,i}}| \equiv -\sum_i p_{T,i}, \tag{3.1}$$

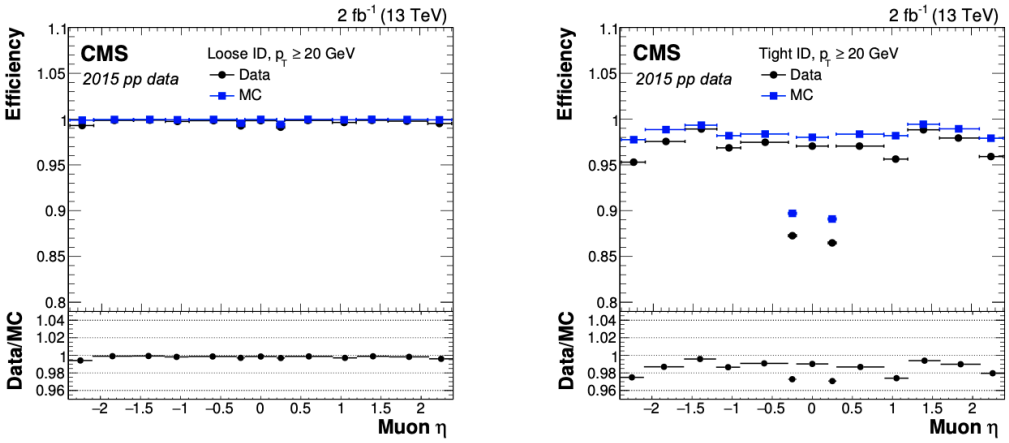

Figure 3.19: Efficiency for muon reconstruction and identification in 2015 data (circles) and simulation (squares), evaluated by applying the *Tag and Probe* method to muons coming from J/ψ and Z decays. The efficiencies for probe muons with $p_T > 20$ GeV passing the Loose ID (left) and the Tight ID (right) are shown in the upper panels, while the ratio between efficiencies in data and simulation is shown in the bottom pnales. The statistical uncertainties are smaller than the symbols used to display the measurements [150].

where the sum is performed over all the reconstructed particles. In CMS, three different algorithms are used to measure the MET, leading to three different MET types:

- **PF $\boldsymbol{E_T^{miss}}$**: defined as the negative vectorial sum over the transverse momenta of all reconstructed PF particles. It is used in the majority of CMS analyses.
- **Calo $\boldsymbol{E_T^{miss}}$**: evaluated using the information of the energies contained in calorimeter towers, excluding deposits below the noise threshold. Since a muon deposits only a fraction of their energy in the calorimeter, the muon p_T is also included in the calculation, while the small calorimetric energy deposit associated with the muon track is excluded.
- **TC $\boldsymbol{E_T^{miss}}$**: starts from the Calo E_T^{miss} but also includes in the calculation the transverse momentum of the inner tracks, removing the expected calorimetric energy deposit of each track.

Large values of MET can be caused not only by the production of invisible particles in the collisions. More often, the missing energy is overestimated due to the contribution of anomalous events that lead to the reconstruction of fake high energy particles, increasing in this way the total measured transverse energy. The presence of these anomalous events can be due to several causes [152]:

- Anomalous signal in the calorimeters. In ECAL they are usually associated with particles directly interacting with the EB sensors, or hitting the transducers, or to noisy sensors in the photodetectors. In HCAL, they can arise from noise in the photodiodes and electronics, as well as from particles interacting in the light guides and photomultiplier tubes of the HF. The filter employs information from energy deposits, timing and pulse shape in order to exclude the anomalous contributions from the MET calculation.
- Beam halo contribution. The interaction of protons with the residual gas molecules in the vacuum chambers, or with the wall of the beam pipe, produces a flow of secondary muons that travel parallel to the collision axis and that leave energy deposits in ECAL. In this case, the information from CSCs is exploited: the hits

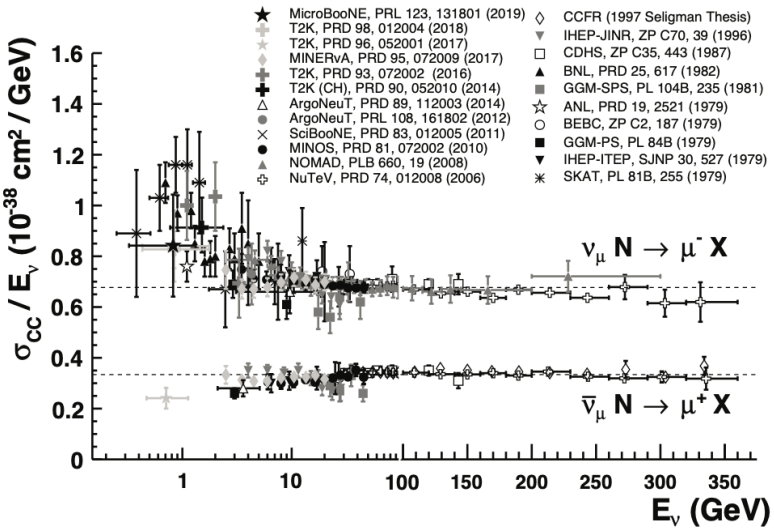

Figure 3.20: Total inclusive charged current cross section for a neutrino (ν_μ N $\rightarrow$ μ^- X) scattering off nucleons, as a function of its energy. Neutrino cross sections are typically twice as large as their corresponding antineutrino counterparts. Neutral current cross sections (not shown here) are generally smaller compared to the charged current case [45].

collected for a beam halo event will indeed be in line with the calorimeter deposits, allowing for an effective identification and removal of such events.

- Non-instrumented or bad functioning detector regions. The filter operates by rejecting events where the track of a muon or a charged hadron has been reconstructed with low quality, according to the *Tune-P* algorithm.

Dedicated filtering algorithms are used to suppress the contribution from these anomalous events. The filters are designed to identify more than 85-90% events contributing to fake E_T^{miss}, with a misidentification rate smaller than 0.1% [153]. In Fig. 3.21, the E_T^{miss} and the ϕ distribution of the jet are shown, before and after the application of the filters, for the dijet and monojet samples respectively. It can be seen that, for both cases, the agreement between the simulated distributions and the data is restored only after the application of all the filters. The value of the MET obtained at this stage still needs to be corrected before it can be safely used in a physics analysis. The non-linearity of the response of the calorimeter for neutral and charged hadrons, neutrinos from semileptonic decays of particles, and possible inefficiencies in the tracker can indeed lead to an underestimation of the MET. For the PF E_T^{miss} three types of corrections are recommended to be applied:

- **Type-0 correction**. It takes into account the degradation of the MET reconstruction due to invisible particles, such as neutrinos from K decays produced in pileup interactions. This correction is based on the CHS algorithm applied to jets but, besides removing charged hadrons originating from the pileup vertices, the Type-0 correction also removes an estimate of neutral pileup contributions. The estimate is based on the assumptions that, at the true level, the contribution to the total transverse momentum provided by charged particles and neutral particles originated in pileup interactions sums to zero (Eq. 3.2), and that the contribution from charged particles is measured with enough precision that Eq. 3.3 holds. Here, $\vec{E}^{miss}_{\mathrm{T,uncorr}}$ refers to the uncorrected vectorial MET, while $\vec{p}_T^{\,true}$ and $\vec{p}_T$ to the true and reconstructed transverse momentum of the particle i, respectively:

$$\sum_{i \in neuPU} \vec{p}_{T,i}^{\,true} + \sum_{i \in chPU} \vec{p}_{T,i}^{\,true} = 0, \tag{3.2}$$

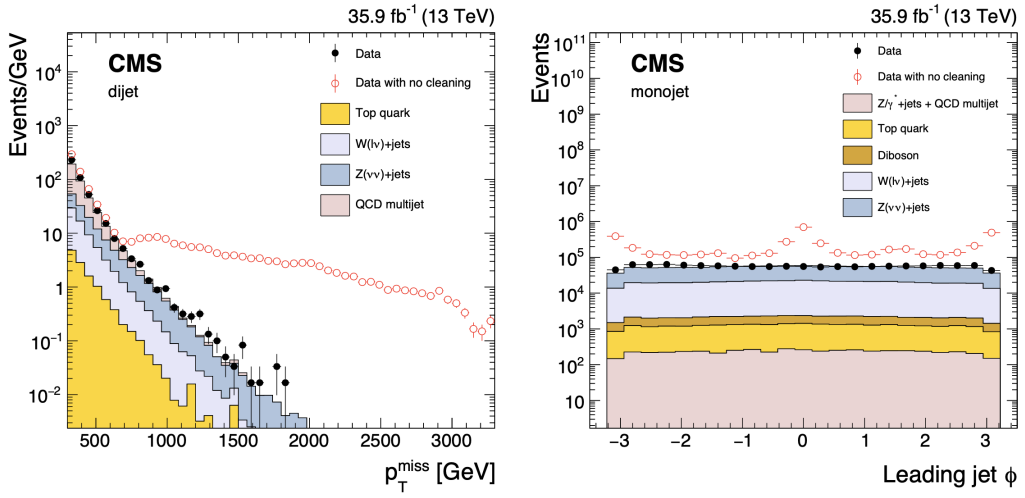

Figure 3.21: The E_T^{miss} (left) and ϕ (right) distributions obtained in data for events passing the dijet (left) and monojet (right) selection with (filled dots) and without (open dots) the event filtering algorithms applied, compared with simulation (solid histograms) [153].

$$\sum_{i \in chPU} \vec{p}_{T,i}^{\,true} = \sum_{i \in chPU} \vec{p}_{T,i}. \tag{3.3}$$

The contribution from charged particles is evaluated by the CHS algorithm, and hence the estimation of the neutral contribution can be written as:

$$\sum_{i \in neuPU} \vec{p}_{T,i} = (1 - R_0) \sum_{i \in chPU} \vec{p}_{T,i}$$

where R_0 represents a systematic shift that needs to be introduced due to the calibration of calorimeters, which is optimized for high-p_T particles. As a consequence, the energy measurement of low-p_T particles, such as those coming from pileup vertices, results to be systematically off. The Type-0 correction factor $\vec{C}^{Type-0}$ and the corrected MET $\vec{E}^{miss}_{T,Type-0}$ are therefore defined as:

$$\vec{C}^{Type-0} = (1 - R_0) \sum_{i \in chPU} \vec{p}_{T,i},$$

$$\vec{E}^{miss}_{T,Type-0} = \vec{E}^{miss}_{T,uncorr.} + \vec{C}^{Type-0} = \vec{E}^{miss}_{T,uncorr.} + (1 - R_0) \sum_{i \in chPU} \vec{p}_{T,i}.$$

- **Type-1 correction**. This correction is a propagation of the JEC to MET. It replaces the vector sum of transverse momenta of particles which can be clustered as jets with the vector sum of the transverse momenta of the jets to which JEC is applied. First, the sum on the PF particles is split into two disjoint sets, either including the particles that can or cannot be clustered as jets:

$$\vec{E}^{miss}_{T,uncorr.} = -\sum_{i \in jets} \vec{p}_{T,i} - \sum_{i \in unclus.} \vec{p}_{T,i}. \tag{3.4}$$

Then, assuming that the first vector sum is the same as the vector sum of the transverse momentum of all (uncorrected) jets (Eq. 3.5), the Type-I correction can be written as the difference between two vector sums (Eq. 3.7):

$$\sum_{i \in jets} \vec{p}_{T,i} = \sum_{jets} \vec{p}_T^{\,uncorr.} \tag{3.5}$$

$$\vec{C}^{Type-1} = \sum_{jets} \vec{p_T}^{uncorr.} - \sum_{jets} \vec{p}_T^{JEC}. \tag{3.6}$$

As a result, the Type-1 corrected MET can be written as:

$$\vec{E}^{miss}_{T,Type-1} = \vec{E}^{miss}_{T,uncorr.} + \vec{C}^{Type-1} = -\sum_{jets} \vec{p}_T^{JEC} - \sum_{i \in unclus.} \vec{p}_{T,i}. \tag{3.7}$$

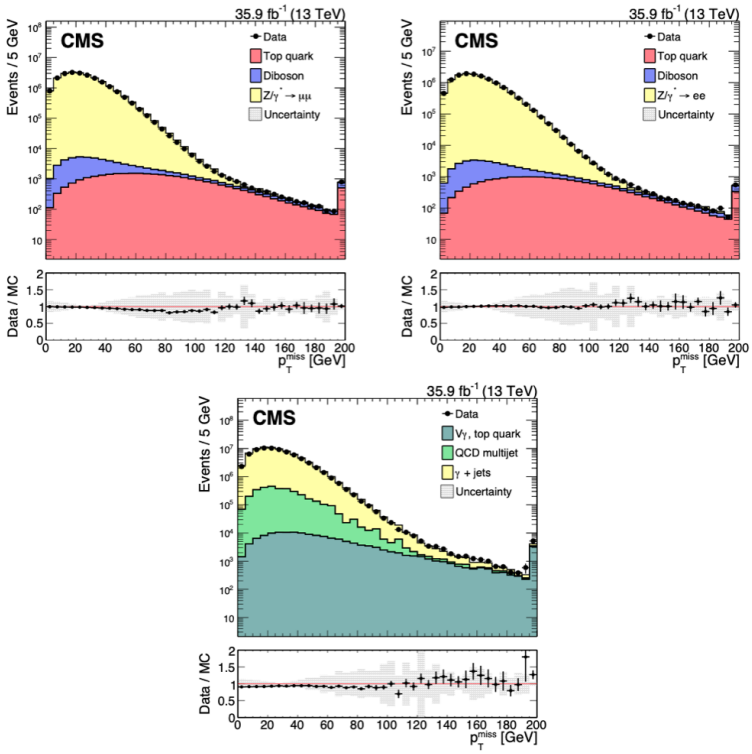

Figure 3.22: Distributions of MET in $Z \to \mu^- \mu^+$ (top left), $Z \to e^- e^+$ (top right), and γ + jets events (bottom) in data and simulation. The last bin includes all events with MET > 195 GeV. Lower panel: Data to simulation ratio. The shaded band includes the systematic uncertainties due to the energy scale and resolution of all reconstructed objects entering the MET computation [153].

- **xy-shift correction**. It is applied to restore the ϕ-independence of the MET. The distribution of true MET indeed must be independent of the polar angle because of the rotational symmetry of the collisions around the beam axis. However, the reconstructed MET shows a modulation with ϕ that could be due to anisotropic detector responses, inactive calorimeter cells, detector misalignment, and displacement of the beam spot. It has been observed that the amplitude of the modulation increases roughly linearly with the number of the pileup interactions. Such modulation can be reduced by shifting the origin of the coordinate in the transverse momentum plane by a constant factor $\vec{c}$ (Eq. 3.8) that depends on the number of vertices. With this shift, the MET can be rewritten as in Eq. 3.9, where n is the total number of particles:

$$\vec{p}_{T,i} \to \vec{p}_{T,i} - \vec{c}, \tag{3.8}$$

$$\vec{E}^{miss}_{T,xy} = -\sum_{i \in all} (\vec{p}_{T,i} - \vec{c}) = \vec{E}^{miss}_{T,uncorr.} + n\vec{c}. \tag{3.9}$$

The xy-shift correction C^{xy} and the xy-corrected MET are then defined as:

$$\vec{C}^{xy} = n\vec{c},$$

$$\vec{E}^{miss}_{T,xy} = \vec{E}^{miss}_{T,uncorr.} + \vec{C}^{xy}.$$

- **Jet smearing**. In simulated samples, the jet momenta are smeared in order to account for possible jet resolution differences between data and simulation [154]. The smearing function is obtained from a parameterizations of the full jet p_T resolution distribution, and the simulated MET is then recomputed taking into account the new jet momenta.

In Fig. 3.22 the MET distribution is shown after the application of all the above corrections, for $Z \to \mu^-\mu^+$, $Z \to e^-e^+$ and γ+jets events, and very good agreement is observed between data and simulation. The performances of the PF E_T^{miss} reconstruction algorithm are studied in term of the E_T^{miss} response and resolution, in samples with an identified Z boson decaying to a pair of electrons or muons, or with an isolated photon. Such events have typically little or no true MET. The hadronic recoil system $\vec{u}_T$ is defined as the vector p_T-sum of all PF candidates, except for the vector boson (or its decay products in the case of the Z boson decay), while the vector boson momentum in the transverse plane is denoted with $\vec{q}_T$. The momentum conservation in the transverse plane imposes $\vec{q}_T + \vec{u}_T + \vec{E}_T^{miss} = 0$. In Fig. 3.23 all the kinematic definitions used are summarized. The

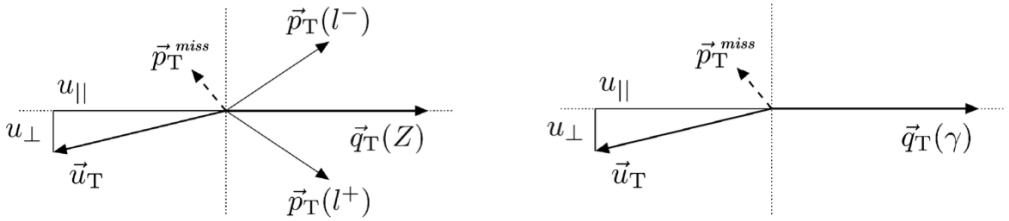

Figure 3.23: Scheme of the Z boson (left) and photon (right) event kinematics in the transverse plane. $u_{||}$ and $u_\perp$ are the magnitude of the parallel and perpendicular components of the hadronic recoil $\vec{u}_T$, respectively. The vectorial missing transverse energy is labeled here as $\vec{p}_T^{\,miss}$ [154].

scalar quantity $-\langle u_{||}\rangle/q_T$ is referred to as the MET response, and it is shown in Fig. 3.24 as a function of the magnitude of the boson transverse momentum q_T. The MET resolution is instead extracted from the $u_\perp$ and the $u_{||} + q_T$ distributions, and it is expressed in terms of the resolution of $u_\perp$ and $u_{||}$ (Fig. 3.25). For processes without true E_T^{miss}, the distribution of the vectorial sum of $u_{||}$ and q_T is expected to be symmetric around zero, with any deviations from this behavior implying imperfect MET reconstruction. The $u_\perp$ distribution is expected to be also symmetric and with a mean value of zero, due to the assumed isotropic nature of the energy fluctuations of the detector noise and of the underlying events.

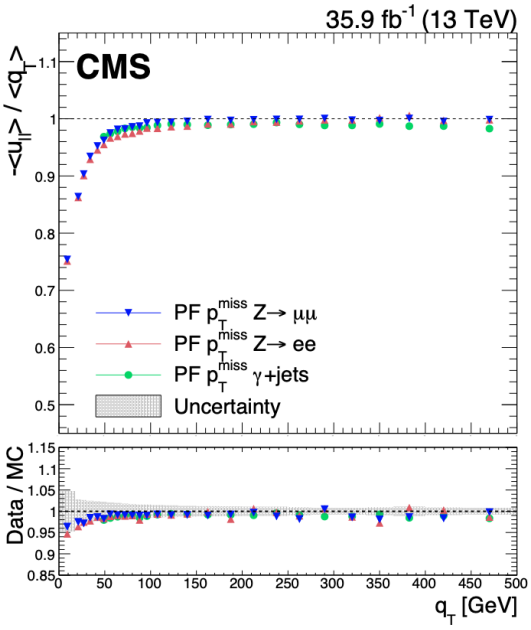

Figure 3.24: The MET response in data, in $Z \to \mu^-\mu^+$ (blue), $Z \to e^-e^+$ (red), and γ+jets (green) events is shown in the upper panel, while the ratio of the MET response in data and simulation is shown in the lower one. The grey band corresponds to the systematic uncertainties due to the energy scale and resolution of all reconstructed objects entering the MET computations, and it is estimated from the $Z \to e^-e^+$ sample [153].

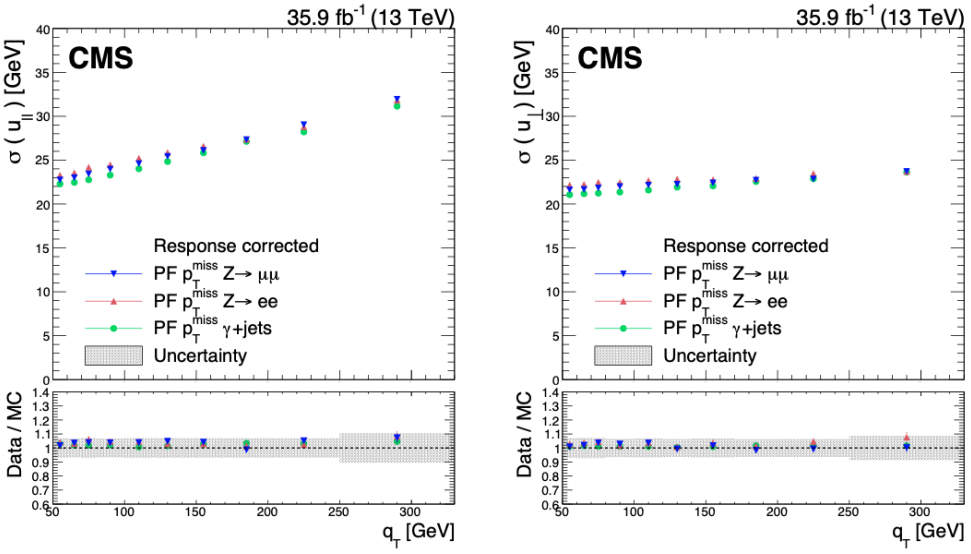

Figure 3.25: Resolution of the parallel (left) and perpendicular (right) component of the hadronic recoil as a function of the magnitude of the boson transverse momentum, in $Z \to \mu^-\mu^+$ (blue), $Z \to e^-e^+$ (red), and γ +jets (red) events. In each plot, the upper panel shows the resolution in data, whereas the lower panel shows the ratio of data to simulation. The grey band corresponds to the systematic uncertainties due to the energy scale and resolution of all reconstructed objects entering the MET computations, and it is estimated from the $Z \to e^-e^+$ sample [153].

Chapter 4

The MIP timing detector for the CMS Phase 2 upgrade

The performances of particle colliders such as the LHC are typically quantified in terms of beam energy and delivered luminosity. The design energy of 14 TeV is foreseen to be reached by the LHC during the Run 3, while the instantaneous luminosity achieved during Run 2 has already exceeded the design value of $1 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ by a factor of two. After the Run 3, the LHC is undergoing a significant series of upgrades that will allow the instantaneous luminosity to reach five times the design value, opening a new era for discoveries in particle physics. In parallel, to manage such high luminosity, CMS is expected to replace the majority of its sub-detectors and to install new ones. In particular, this chapter is focused on the new timing detector for minimum ionizing particle, and on the timing performance studies carried out along with the analysis work described in this thesis. The chapter opens with a brief overview of the LHC upgrades foreseen for this new high luminosity phase, together with the physics program expected to be delivered by the four main experiments operating along the accelerator. The CMS upgrades are introduced and the timing detector is described in detail. Then, the timing performances studied in simulation are presented, and in the last section the impact of the timing detector on the CMS physics program is quantified.

4.1 High-luminosity LHC

At the end of the ongoing Run 3, expected in 2025, the LHC will enter the third long period of technical stop since the beginning of its operations in 2008. During the so-called Long Shutdown 3 (LS3), both the accelerator and the detectors will undergo an important series of upgrades in order to start the new High-Luminosity (HL) LHC phase in 2029. The Run 3 is already expected to double the collected luminosity with respect to the previous Run 2, reaching a total of 300 fb^{-1}, but for many of the most interesting physics searches foreseen for the next future, especially those involving multiple Higgs bosons production and possible new resonances, a much higher statistics will be needed. In order to improve the potential for discoveries, the HL-LHC upgrade is hence focused on a substantial increase of the delivered luminosity, up to twenty times the expected total for Run 3 (to give an example of the expected performances, the LHC will produce at least 15 million Higgs bosons per year during the HL phase, compared to around three million in 2017). Physicists will have the chance to observe extremely rare phenomena that could not be revealed otherwise, and the magnified statistics will allow to study SM mechanisms with unprecedented precision. The main topics of the physics program carried out by the LHC experiments during the HL phase are [155]:

- **Precision measurements in the Higgs boson sector**. In Fig. 4.1, the ex-

pected improvements in the measurements of the Higgs boson couplings to elementary fermions and bosons with 3000 fb^{-1} of data are illustrated. The accurate estimate of such couplings has great importance in the understanding of the quantum instability related to the Higgs potential. Moreover, the increased luminosity will allow to probe for the very first time important rare processes, such as $gg \rightarrow HH$. The di-Higgs production is one of the main goal of the LHC physics program, given that such process is the only one where the trilinear Higgs-self coupling can be directly measured, providing direct access to the shape of the Higgs scalar potential.

- **Beyond Standard Model searches**. In the last decades, many BSM theories have been developed. As described in Sec. 1.6, an extended and more exhaustive formulation of the SM is indeed needed, providing a theory that would include all the four fundamental forces. One of the most popular is the theory of Supersymmetry (SUSY), predicting a so-called "supersymmetric partner" for each particle in the SM. Among the new particles predicted, the class benefiting more from the large luminosity provided by the HL-LHC is that of the electroweak SUSY particles (as neutralinos and charginos), whose cross sections is two to three orders of magnitude smaller than generic QCD cross-sections. Of course, besides SUSY, the HL phase will allow searches for possible new exotic particles and resonances in many different BSM scenarios.
- **Flavor studies**. The LHCb experiment (designed to study flavor physics), but also ATLAS and CMS will benefit from the increased luminosity, with the latter reaching sensitivity comparable to those of dedicated experiments for several processes (for example $B_{s,d} \rightarrow \mu\mu, B_s^0 \rightarrow J/\psi, B_0 \rightarrow K^{*0}\mu\mu$). In flavor physics there is plenty of discovery potential, as it is evinced by the latest measurements at the LHCb experiment [156]: their results indeed provide hints of a violation of lepton flavor universality, and they are consistent with a pattern of flavor anomalies already measured by LHCb [157, 158] and other experiments worldwide over the past decade [159, 160]. Flavor probes are especially useful since they are able to test the virtual effects of new particles above the TeV scale, whose direct production is not accessible at the current energy of the LHC. In particular, in processes involving Flavor Changing Neutral Currents (FCNCs) there is still a 20% possible new physics contribution not yet constrained by the precision of current measurements.

In order to achieve at least the total luminosity goal of 3000 fb^{-1}, an annual integrated luminosity of around 250 fb^{-1} is expected. This will lead to a peak luminosity of $5 \times 10^{34} cm^{-2} s^{-1}$, which is already five times the original designed luminosity of the LHC, and in its ultimate configuration the HL-LHC is expected to reach up to $7.5 \times 10^{34} cm^{-2} s^{-1}$. In such luminosity configurations, the number of events per bunch crossing, the pileup events, is expected to vary from 140 to 200. The key parameters that have been targeted, and which need to be carefully tuned to deliver such impressive and certainly unprecedented performances, are the number of protons per bunch and the transverse beam size at the interaction points (IPs) [161]. Another relevant parameter could be the number of bunches per beam, but the bunch spacing at the LHC is already at its minimum value of 25 ns, with the aforementioned maximum number of bunches of approximately 2800. The increase in the bunch population (up to 2×10^{11} protons per bunch in the baseline scenario [162]) will be provided by a renovated injector accelerator chain, with major upgrades required in the linear accelerator, the Proton Synchrotron Booster, the Proton Synchrotron and the Super Proton Synchrotron. The bunch population is constrained by several factors, such as the impact of one bunch on another when they collide at the IP, possible beam-beam effects and the robustness of the machine protection. This protection is currently ensured by Carbon-Fibre-Carbon composite collimators, designed to absorbe particles that might stray from the beam trajectory and

damage the machine. For the HL-LHC, the majority of the collimators will be substituted with new ones made by a novel Molybdenum-Graphite composite which guarantees higher performances and less electromagnetic interference on the beam. The second key factor in maximizing the luminosity performance will be the reduction of the β^* parameter below 30 cm (the nominal value of the LHC is 55 cm). This will be achieved by new quadrupoles made of niobium-tin (Nb_3Sn) replacing the current NbTi ones. This novel material is able to generate a magnetic field up to 12 T compared to the present 9 T maximum, and together with the beam optics scheme will be able to provide a more focused and squeezed proton beam. With the reduction of β^* the crossing angle between the bunches needs to be adjusted in order to ensure that they are enough separated when reaching the IP, and to avoid potential unwanted bunch-bunch interactions. This necessary adjustment of the crossing angle, however, can lead to significant luminosity reduction up to 70% of the β^* design values. To recover this loss, transversely deflecting RF devices, the so-called *crab cavities* [163] will be installed near the ATLAS and CMS IPs. Their goal is to tilt bunches, providing an effective overlap, thus higher luminosity, even when the beam trajectories have a crossing angle. Besides the aspects discussed so far, a relevant number of refurbishments are still foreseen for the LHC to perform at its best during the HL phase, including upgrades of the power lines, instrumentation, devices for tackling with the high level of collision debris and high radiation levels, and a huge amount of civil-engineering work will be needed as well to build the new HL-LHC underground areas.

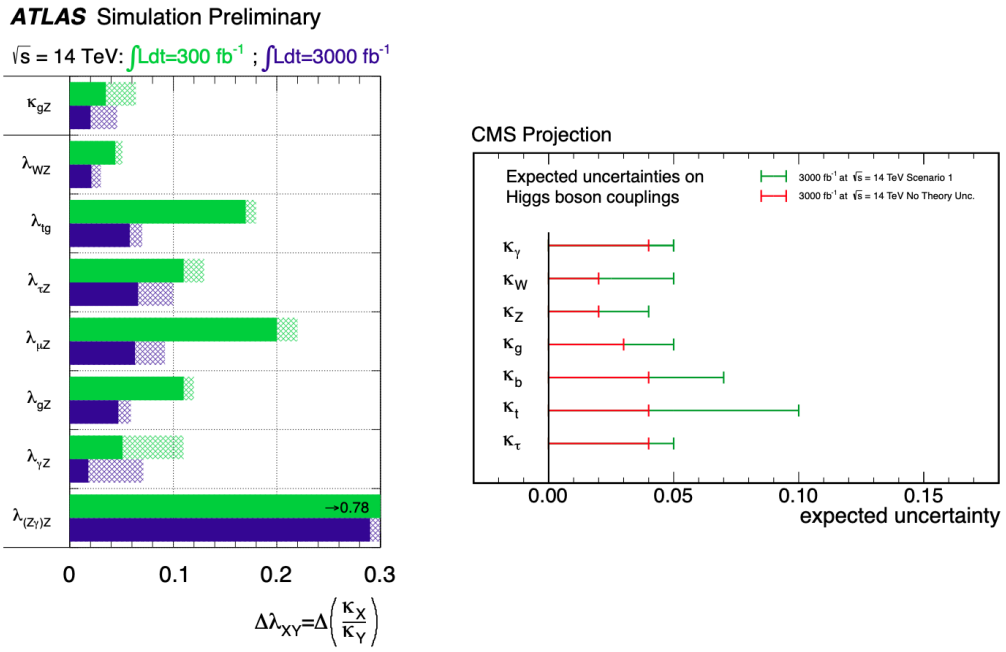

Figure 4.1: Estimated precision of the Higgs boson coupling modifiers with an integrated dataset of 3000 fb^{-1} and assuming a center of mass energy of 14 TeV. (Left) ATLAS relative uncertainty on the expected precision for the determination of coupling scale factor ratios λ_{XY}. The hashed areas represent the increase of the estimated error due to current theory systematic uncertainties. (Right) CMS projection for the expected uncertainties in the couplings. Scenario 1 assumes the current systematic uncertainties while Scenario 2 considers the experimental uncertainties scaling as $1/\sqrt{L}$ and reduces the theoretical uncertainties by half [155].

4.2 The CMS Phase 2 upgrade

In preparation for the HL-LHC operations, the CMS experiment is going to replace or refurbish the majority of its detectors under a program named "Phase 2 Upgrade". The goal of this program is to maintain excellent detectors performances, and optimal resolution and particle reconstruction under the demanding conditions of a high luminosity environment. In more detail, the three main challenges that CMS has to face with are related to the substantial increase of the data throughput, the degradation of the sensing elements and electronics induced by the radiation, and the high number of pileup events which worsen the detectors' resolution. All the CMS subsystems will undergo specific renovations and new detectors are going to be installed in order to cope with the new HL conditions. A part of the Phase 2 operations has already started during the 2018-2021 technical stop (the Long Shutdown 2, LS2) preceding the Run 3, while the majority of the upgrades will take place during the LS3, expected to start in 2026 and last for three years. In the following, a short overview of the main Phase 2 upgrades is provided, except for the timing detector which is described in a more exhaustive way in Sec. 4.3.

4.2.1 Tracker

The CMS tracking system will suffer from significant radiation damage by LS3, and will be completely substituted for Phase 2. It will consist of a new inner tracker (IT) implemented with new silicon pixel modules, and an outer tracker (OT) implemented with new silicon modules with strip and macro-pixel sensors [164]. The IT will be composed of a barrel part with four cylindrical layers and twelve disc-like structures in each forward direction, providing a coverage up to about $|\eta| = 4$. The OT instead will be instrumented with dedicated "p_T-modules", which are capable of rejecting signals from particles below a certain p_T threshold. Two versions of the p_T-modules will be realized, one with two strip sensors and one with a strip and a macro-pixel sensor, in order to provide the z (r) coordinate measurements in the barrel (endcaps), thus helping to identify particles coming from different vertices. The new tracker design will ensure improved η coverage, granularity, and high radiation tolerance, with no maintenance intervention for the OT and the possibility to access the IT to replace radiation-damaged elements without removing the CMS beam pipe. The smaller pixels and thinner sensors implemented in the IT will provide better impact parameter resolution and two-track separation, while the information from the p_T modules will contribute to the L1 online event selection, anticipating part of the reconstruction currently performed by the HLT. In addition, the designed reduced amount of material in the tracker volume will be beneficial for both the tracker system and the calorimeters, reducing the energy loss of the particles before they enter the calorimeters.

4.2.2 Endcap calorimeter

Given the high radiation levels in the endcap regions, the electromagnetic and hadronic endcap calorimeters must be replaced by the end of the Run 3, and a novel high-granularity sampling calorimeter (HGCAL) will be installed, covering the region $1.5 < |\eta| < 3$ [165]. The HGCAL will consist of an electromagnetic compartment (CE-E) followed by a hadronic compartment (CE-H), featuring an unprecedented transverse and longitudinal segmentation for both sectors, which will be able to provide three dimensional images of showers. The CE-E will be composed of 28 sampling layers with a total thickness of 34 cm and a depth of approximately 26 radiation lengths X_0 and 1.7 electromagnetic interaction lengths λ. The active detector elements will be hexagonal silicon sensors of different thickness, according to the particles fluence of the different regions, while tungsten and copper plates will be used as absorber. The CE-H will employ the same silicon sensors as active elements,

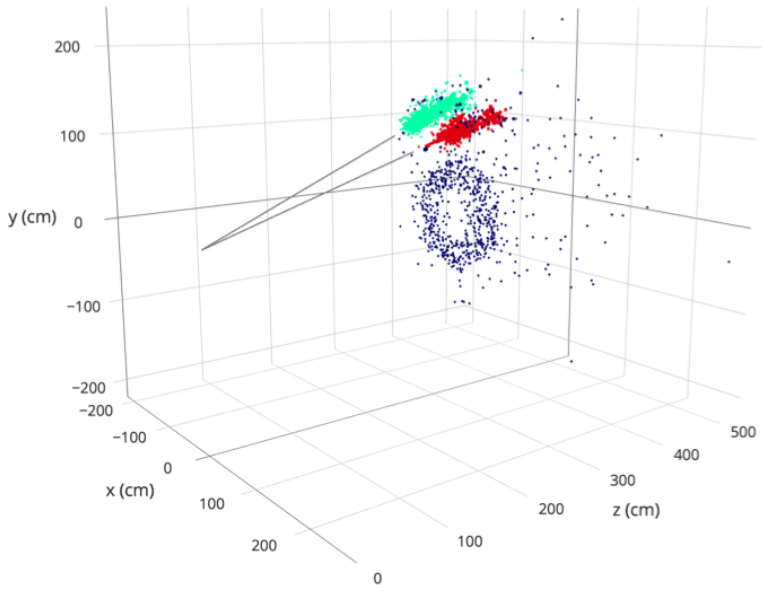

Figure 4.2: Two pions, separated by 30 cm, clustered by the (first exploratory version of) direct 3D clustering algorithm in HGCAL [165].

featuring 12 planes of stainless steel as absorber, corresponding to a depth of 3.5 λ. The hadronic compartment will be followed by a stainless-steel-scintillator sampling calorimeter, providing a total calorimeter depth to about 10 λ. The clustering algorithm for the 3D reconstruction of the shower is going to run on each of the layers of the calorimeter, identifying the local energy deposits in the calorimeter's cells. The (x, y, z) position of the cells will be then projected onto the other planes by using the $\eta - \phi$ coordinates in order to form the 3D cluster associated to the shower (see Fig. 4.2). The materials and design of the HGCAL will provide high radiation tolerance, fine lateral and longitudinal granularity and precision measurement of the time of high energy showers, helping the rejection of energy from pileup and the identification of the primary vertex.

4.2.3 Endcap muons

The main reason for a HL upgrade of the CMS muon system is the current limited acceptance ($|\eta| < 2.4$) provided by the muon chambers. The Phase 2 tracker will be able to cover a pseudorapidity region up to $|\eta| = 4$, that has to be matched by a muon detector since the silicon tracker alone cannot identify muons with enough accuracy. The installation of four additional muon stations is foreseen during the LS3, in order to extend the muons acceptance. Two of them will use improved Resistive Plate Chambers (iRPCs) with smaller electrodes and reduced gas gap thickness, to satisfy the space requirements of the experiment [166], while the other two chambers will implement the Gas Electron Multiplier (GEM) technology. The GEM chambers will be filled with a mixture of argon and CO_2, and they will consist of two printed circuit boards with a stack of three GEM foils in between. The GEM foils are polyimide foils of 50 μm thickness perforated by conical holes and surrounded on the top and bottom with copper conductors. The upgrade of the muon system has already started, with the insertion of "demonstrator" GEM during the winter shutdown between 2016 and 2017 [167].

4.2.4 Trigger and DAQ system

The Phase 2 upgrade of the trigger and DAQ system will use the same two-level strategy of the Phase 1 described in Sec. 3.2.7, but with increased acceptance rate and latency in order to maintain similar performances in terms of event selection and to exploit the information coming from all the sub-detectors [168]. The maximum trigger frequencies for the L1 and the HLT, which currently are 100 kHz and 1 kHz, respectively, will be increased to 750 kHz and 7.5 kHz. The allowed latency, defined as the time needed to form and distribute the trigger decision, will also increase from the current value of 4 μs to 12.5 μs, enabling for the first time the inclusion of the tracker and high-granularity calorimeter information. Another key feature of the new trigger system will be the introduction of a Correlator Trigger (CT), working in two steps called correlator layers. The first layer will create particle-flow candidates which can be filtered based on the information of the primary vertex, while the second step will reconstruct the physics objects. This new trigger system will allow to achieve trigger performances already at the L1 stage that approach those of the HLT. In parallel, the HLT algorithms will be improved and refined, deploying more specialized algorithms for dedicated event selections.

4.3 The MIP timing detector project

The average amount of pileup (PU) interactions that CMS has faced during the Run 2 is about 40 per event, and several PU mitigation methods (as the CHS algorithm described in Sec. 3.3.4) are currently employed to cope with this amount of overlapping interactions and to identify the hard vertex. However, as previously mentioned, the PU is going to increase during the HL-LHC growing from 140 up to 200. With this huge number of secondary vertices, the growth in the probability of spatial overlaps could make the PU mitigation algorithms failing at a substantial rate. The vertices overlap can be quantified in

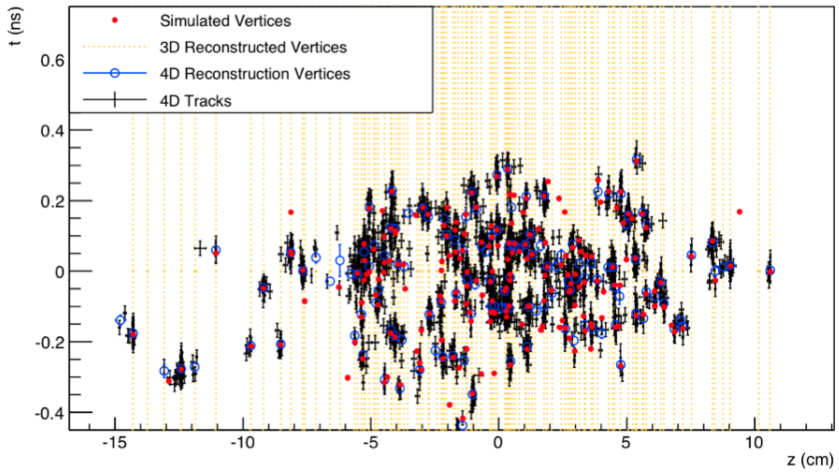

Figure 4.3: Simulated and reconstructed vertices in a bunch crossing with 200 PU interactions assuming a MIP timing detector with 30 ps time resolution covering the barrel and endcap regions. The simulated vertices are the red dots. The vertical yellow lines represent the vertices reconstructed without the timing information (3D vertices), while the black crosses and the blue open circles represent tracks and vertices reconstructed including the time information, and hence they are labeled as 4D. The horizontal axis refers to the z position along the beam line, while the vertical axis to the vertex time. As can be seen, many of the 3D vertices that would not be resolved solely on the basis of the spatial coordinates are clearly separated when using the time information [169].

terms of the *line density*, defined as the density of vertices along the z-axis, dN_V/dz. The *line density* measured at the LHC with around 30 PU interactions is 0.3 mm^{-1} and it is going to increase up to 1.9 mm^{-1} in the 200 PU scenario, contaminating the reconstruction of the hard vertex with tracks from pileup and hence worsening resolution and efficiency of the physics objects. The CMS timing detector for minimum ionizing particles (MTD) that will be installed during the LS3 aims to reduce the PU overlapping by measuring the time of arrival of minimum ionizing particles, and by using this information to reconstruct the time of the vertices. The inclusion of the timing in the vertex reconstruction algorithms will allow to identify spatially coincident PU vertices that otherwise would be reconstructed as the same vertex (see Fig. 4.3), providing a reduction of the PU levels down to 40-60 events within the time uncertainty (as during Phase 1). The time resolution provided by the MTD is expected to be 30-40 ps (degrading to 50-60 ps at the end of operations, due to the accumulated radiation), compared to a resolution of 180-200 ps resulting from a level of 200 PU events in a no-MTD scenario. The number of PU tracks incorrectly associated with the hard interaction vertex is shown in Fig. 4.4 as a function of the collision line density, for different resolutions. It can be observed that even for a degraded time resolution of 60 ps, the MTD is already able to reduce the number of the wrong association by around a half.

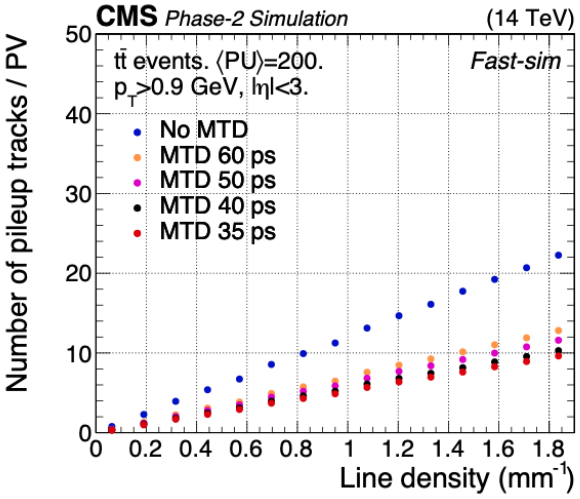

Figure 4.4: Number of pileup tracks incorrectly associated to the primary vertex as a function of the line density, for different time resolutions [169].

4.4 The MIP timing detector

The design of the MTD is similar to those of the CMS tracker and calorimeters, consisting in a Barrel Timing Layer (BTL), up to $|\eta| < 1.5$, and an Endcap Timing Layer (ETL) covering the pseudorapidity region up to $|\eta| = 3$. Given the strict geometrical requirements provided by the compactness of the CMS detector, there are only two locations where the timing detectors can be installed: the BTL is going to be placed between the last layer of the OT and the ECAL barrel, while the ETL will be integrated between the tracker bulkhead and the CE-E. In order to cope with the different environment conditions, different technologies are implemented in the two pseudorapidity regions. A schematic view of the geometry of the MTD is shown in Fig. 4.5, while a more detailed description of its two sub-systems is provided below.

The Barrel Timing Layer

The Barrel Timing Layer has a cylindrical shape, with an inner radius of 1148 mm and an outer radius of 1188 mm. Its sensitive elements are Lutetium Yttrium Orthosilicate crystal bars doped with Cerium (abbreviated in L(Y)SO:Ce crystal bars), of 57 mm length and 3.12 mm width. The crystals are instrumented with Silicon Photo Multipliers (SiPM) at each side, with the aim of doubling the readout information and ensuring a uniform time response along the crystal. The crystals' thickness along the z-axis is optimized to have the smallest impact possible on the ECAL energy resolution while maintaining the maximum light extraction from the crystal. Three crystal thickness are used, decreasing from 3.75 mm to 3.0 mm and 2.4 mm according to the η region where they will be placed, to account for the non-uniformity in the radiation level expected within the BTL pseudo-rapidity range. In a similar way, the SiPM active area is also reduced from 3.75 mm^2 to 9 mm^2 and 7.2 mm^2. The BTL will consist of 165888 crystals, resulting in a total channel count of 331776. The energy deposit of a MIP particle in a L(Y)SO:Ce crystal follows a Landau distribution whose most probable value is located around 0.86 MeV/mm. The average energy deposit in the BTL is 4.2 MeV which, considering the performances of the photomultipliers and the properties of the scintillating crystals, would lead to a total number of 5100 photoelectrons expected at each SiPM for a MIP. The measurement of the time at which the MIP crosses the detector is referred to as "time stamp", and it is averaged over the two sides of the crystal, $t_{avg} = (t_{right} + t_{left})/2$, in order to provide an uniform time stamp across the single bar, minimizing the dependance of the time resolution on the position of the hit in the crystal. A time correction for amplitude variations (amplitude walk) is also applied.

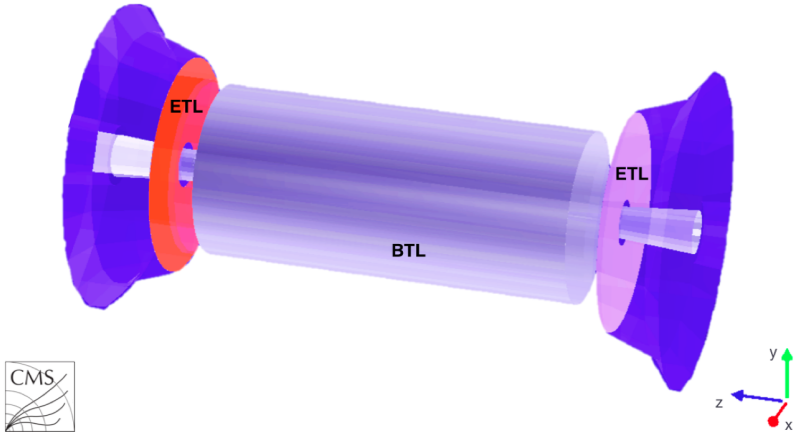

Figure 4.5: A schematic view of the MTD geometry, as it is used by CMS for simulation studies. The grey cylinder represents the BTL, while the red and light violet discs represent the ETL, closing the barrel region at both sides [169].

The Endcap Timing Layer

Each side of the ETL is equipped with two discs instrumented with Ultra-Fast Silicon Detectors (UFSDs) based on the Low-Gain Avalanche Detector (LGAD) technology. They are n-on-p silicon sensors with an extra thin p-layer below the junction that allows a lower and more controlled gain in the signal formation mechanism, reducing the occurence of breakdown. The time measurement is performed by considering the sensor as a capacitor, with a current source in parallel which is read out by a pre-amplifier: in order to evaluate the time of arrival, the output from the pre-amplifier is compared to a fixed threshold. The arrangement of the UFSDs on the two faces of the discs is such that the area allocated for the service devices (readout, power, cables) is covered by sensors on the opposite face. This

approach ensures a good coverage of the endcap regions, minimizing the inactive areas. According to the Technical Design Report proposal, the sensors will have a dimension of 21×42 mm^2, with square pixel pads of 1.3×1.3 mm^2. The small size provides a good granularity, as well as a more efficient use of the silicon wafer where the pixels will be placed. The total amount of space required along the beam axis for the ETL to be installed is 45 mm, and 20 more millimeters would be needed for isolating the timing layer from the CE by the insertion of a thermal screen.

4.4.1 MTD performances in simulation

The PF algorithm used by CMS for the event reconstruction is expected to significantly benefit from the timing information provided by the MTD. In order to quantify the improvements, optimizing the inclusion of the timing in the algorithms and studying its impact on the physics objects, several event reconstruction studies have been performed and are currently ongoing, based on a complete GEANT4 [170] simulation of the MTD detector and on a modeling of the readout electronic response. The time and position of particles crossing MTD are reconstructed starting from the energy deposits in the active detector elements. A simple topological clustering is performed to associate adjacent MTD hits above the readout threshold. The cluster's barycenter, weighted by the energy of the single hits, provides an estimate of the cluster position. In a similar way, the corresponding time is obtained from the average of single-hits time measurements. Once the time and space coordinates of the clusters are defined, the reconstructed track is propagated to MTD and, if a spatially compatible cluster is found, the track is matched to it. The efficiency of the cluster-track association in simulation is shown in Fig. 4.6, and it is observed to be robust against the PU level and independent of the track p_T. Then, the track parameters are refitted with the additional time and space information provided by MTD and the new track is propagated backwards in order to compute the total pathlength L_{track}. At this stage, the track time t_{track} at the beam line is hence defined as:

$$t_{track} = t_{MTD} - \frac{L_{track}}{p_{track}/m_\pi c},$$

where t_{MTD} is the time of the MTD cluster, p_{track} is the track momentum, L_{track} is the pathlength and m_π the pion mass. The uncertainty assigned to this measurement is inflated by adding in quadrature the difference in time of flight (TOF) under the pion and proton mass hypothesis. As mentioned in Sec. 3.3.1, the reconstruction of the primary vertex in CMS is performed via a deterministic annealing algorithm that can be easily extended to more than three dimensions to include the additional timing information. The choice of the mass hypothesis between pions, kaons and protons becomes here a crucial step in the evaluation of the vertex time, because for momenta of the order of a few GeV their difference in TOF becomes significant with respect to the time resolution of the detector. A first 4D vertex reconstruction is performed using the pion hypothesis to compute the time of the tracks entering the algorithm, and a first estimate of the vertex time is obtained. Then, the compatibility of the tracks associated to the reconstructed 4D vertex is tested under the three mass hypotheses. The compatibility is expressed in terms of:

$$\chi^2_{hyp} = \frac{(z_{PCA} - z_{PV})^2}{\sigma^2_{z_{PCA}}} + \frac{(t_{PCA,hyp} - t_{PV})^2}{\sigma^2_{t_{MTD}}},$$

where z_{PV} and t_{PV} are the z-position and time of the primary vertex, as evaluated by the 4D deterministic annealing algorithm, while z_{PCA} and $t_{PCA,hyp}$ are the z-position and time of the track extrapolated at the point of closest approach, the latter evaluated in the mass hypothesis (pion, kaon or proton) *hyp*. The corresponding estimated uncertainty on

the track time measured in MTD is labeled with σ_{MTD}. Times and uncertainties are then reassigned to tracks, using the mass hypothesis providing the best χ^2_{hyp}, and with these recomputed information the 4D vertex reconstruction runs a second and final time.

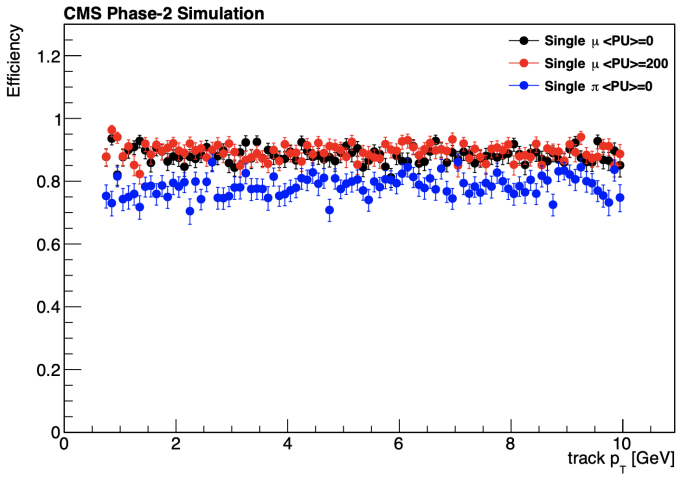

Figure 4.6: Efficiency to find an MTD cluster associated with a reconstructed track as a function of transverse momentum. The efficiency obtained using a sample of simulated muons without pileup is shown in black. The results are compared to those obtained with a simulated sample of muons with an average of 200 pileup events (red) and a simulated sample of pions without pileup (blue). The efficiency is about 90% for muons, and about 80% for pions, lowered by the effects of nuclear interactions in the tracker volume [169].

Quality of the track-vertex association

The inclusion of the MTD information in the CMS algorithms is still in progress, and different studies are currently ongoing aiming for a comprehensively and optimized integration of the timing in the event reconstruction. The first assessment of the 4D vertexing performances and of the goodness of the particle identification can be found in the Technical Design Report, taken as reference for the present studies. In Fig. 4.7, for example, the reconstructed vertex time before and after the particle identification step is compared to the simulated time of the hard interaction for tracks reconstructed within $|\Delta z| < 1$ mm of the simulated primary vertex, both for events with no PU (Fig. 4.7 left) and with 200 PU events (Fig. 4.7 right). It can be observed that before particle identification the distribution has an asymmetric tail, due to kaons and protons misinterpreted as late pions, which is removed by the second step of the reconstruction. My contribution to the study of the MTD performances includes the monitoring of the MTD impact on the reconstructed tracks and 4D vertices, in terms of association between the reconstructed track and the simulated vertices, and between the reconstructed vertices and the simulated ones. Studying the track-vertex association is of crucial importance for the rejection of tracks coming from PU vertices. PU tracks are indeed discarded on the basis of a selection on the distance between the reconstructed track and the primary vertex. The track-vertex association relies on an algorithm that first associates a simulated (SIM) track with a reconstructed (RECO) track according to the number of hits they share. Then, if one of the SIM tracks linked to the SIM vertex is associated to a RECO track, the RECO track is matched with the SIM vertex as well. The time resolution of the matched tracks is then obtained, defined as the difference between the time of the RECO track, t_{REC}, and of the SIM track, t_{SIM}. An "estimated time" is also evaluated, which is defined as:

$$t_{est} = t_{MTD} - \frac{L_{track}}{p_{track}/m_{true}c},$$

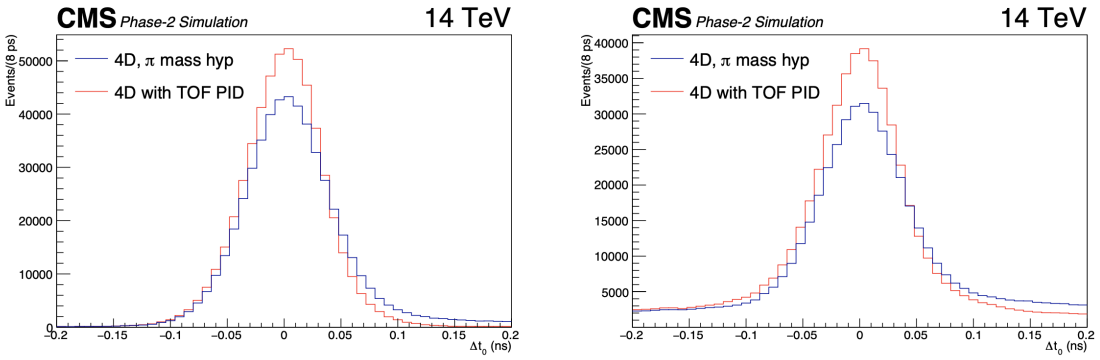

Figure 4.7: Difference between the reconstructed and the simulated track time at the vertex in $t\bar{t}$ events, for particles reconstructed within $|\Delta z| < 1$ mm of the generated hard interaction. Events with no pileup (left) and 200 PU (right) are shown.

where m_{true} is the true value of the particle's mass. Comparing the different quantities $t_{REC} - t_{est}$ and $t_{est} - t_{SIM}$ allows to separate the effects due to an inaccurate PID from other contributions (for example hits wrongly assigned to the track) that could produce incorrect track-vertex association. The study of the association between the RECO and the SIM vertices is also fundamental in the evaluation of the MTD perfomances, since it allows to quantify the precision of the 4D vertex reconstruction by comparisons to the true vertex time obtained from simulation. The association starts by matching RECO tracks to SIM vertices, as described in the previous section. A global weight is then evaluated for each SIM vertex using the weight with which each associated track has contributed to the fit of a RECO vertex. A RECO vertex is hence associated to the SIM vertex that has the highest global weight, and it is classified according to the number of possible "associable" SIM vertex candidates. Having association quality 1 means that for each RECO vertex there is only one SIM candidate, association quality 2 means that there are two SIM candidates and so on, up to quality 7. The vertex timing performances can be then evaluated in terms of the vertex time resolution, defined as the difference between the time of the RECO vertex and the time of its associated SIM vertex (see Fig. 4.8).

Conclusion

The studies performed in this thesis work aims to the final and optimized integration of the timing information in the full event reconstruction. A new tool for the monitoring the MTD cluster-track efficiency of association has been implemented, together with the monitoring of other fundamental variables like cluster and hit energy. In addition, the quality of the association between tracks and vertices, and between reconstructed and simulated vertices has been studied. The development of such new tools is fundamental to monitor the MTD performances while the reconstruction algorithms are being modified and updated toward their ultimate version for Phase 2. Moreover, implementing a method for the track-vertex association lays the groundwork for more sophisticated studies, such as the evaluation of the PID performances, which is of great importance for a wide range of physics analysis, as it is discussed in the following section.

4.4.2 Impact of MTD on physics analysis

The inclusion of timing information in the event reconstruction is expected to provide great benefits in sensitivity for many different physics measurements and searches of the HL-LHC program. The impact of MTD has been studied for a few benchmark cases, representative of three different ways of exploiting the detector, using simulated samples

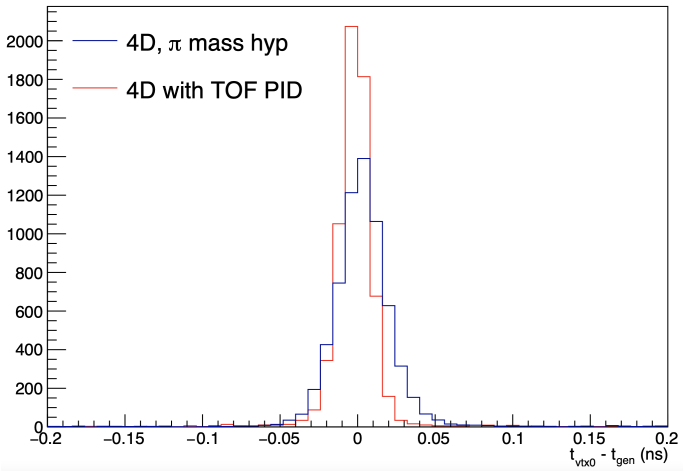

Figure 4.8: Distribution of the difference of the reconstructed (t_{vtx0}) and simulated (t_{gen}) time of the primary vertex in simulated $t\bar{t}$ events with 200 PU as reported in the MTD Technical Design Report [169]. The resolution resulting after the first step of the 4D vertex reconstruction (using the π mass hypothesis) is shown in blue, while the ultimate resolution (after the particle identification step) is shown in red. The root mean square of the distribution obtained after the second step is about 10 ps.

and considering an estimated average MTD time resolution of 30-40 ps:

- Mitigation of the PU effects to improve reconstruction and selection of the physics objects. In this case, as a benchmark analysis, the search for Higgs boson pair (HH) production in several final states is used.

- Introduction of new time-based variables that improve discrimination power between prompt and non-prompt particles, improving the discrimination between signal and background in searches for Long Lived Particles (LLP) in BSM models.

- PID capabilities provided by time-of-flight measurements, especially useful for precise measurements of heavy flavor hadron production in Heavy Ion (HI) collisions, and for b-physics analysis.

Higgs boson pair production

The study of the HH production at the LHC is of great importance, as it constitutes the only way to directly measure the value of the trilinear self-coupling of the Higgs boson, which is fundamental in the determination of the shape of the Brout-Englert-Higgs potential. The process is originated via gluon-gluon fusion and it has a very small cross section ($36.69^{+2.1\%}_{-4.9\%}$ fb) at $\sqrt{s}$= 14 TeV, meaning that a highly integrated luminosity is needed to achieve enough sensitivity to the SM signal and to measure the trilinear coupling. Five decay channels are explored and combined: bbbb, bb$\tau\tau$, bbWW (WW $\rightarrow \ell\nu\ell'\nu'$ with $\ell, \ell' = \mathrm{e}, \mu$), bb$\gamma\gamma$, and bbZZ (ZZ $\rightarrow \ell\ell\ell'\ell'$ with $\ell, \ell' = \mathrm{e}, \mu$). The analysis strategy assumes an MTD resolution of 35 ps with the BTL performances slowly drifting to 50 ps, as a result of the radiation damage, and an integrated luminosity of 3000 fb^{-1}. It also assumes that a HH signal exists, with the strength and properties as predicted by the SM. The improved physics objects, as a result of the integration of the MTD in the event reconstruction, would bring an improvement in the sensitivity in the combined measurement of about 20%. The expected significances for the SM HH signal for the five decay channels (single and combined) are listed in Tab. 4.1.

	Expected significance	
HH decay channel	No MTD	With MTD
bbbb	0.88	0.94
bb$\tau\tau$	1.3	1.48
bb$\gamma\gamma$	1.7	1.83
bbWW	0.53	0.58
bbZZ	0.38	0.42
Combined	2.4	2.63

Table 4.1: Expected significance in units of standard deviation for the SM HH signal for the five decay channels considered and their combination [169].

Long lived particles

The LLPs are particles that travel inside the experiment before decaying into SM objects (leptons, photons and jets), and that are thus detected with a delay compared to analogous objects produced in SM processes [171]. The existence of the LLPs is predicted by several BSM theories and their searches is of high importance in the context of searches for new physics at the LHC. The precise vertex timing ensured by the MTD, together with the information from the Phase 2 CMS calorimeters, is expected to allow the measurement of the TOF of the LLPs between primary and secondary vertices, and provides new powerful variables to identify the possible presence of a LLP signature. The performances of the MTD for this physics case are explored in a 30 ps time resolution scenario, using different topologies. The first class of explored topologies considers a gauge-mediated SUSY breaking (GMSB) scenario, and the production of a $\widetilde{\chi}_1^0$ (neutralino) in top-squarks pairs $\widetilde{\mathrm{t}}\bar{\widetilde{\mathrm{t}}}$ processes is first studied (where $\widetilde{\mathrm{t}} \to \mathrm{t} + \widetilde{\chi}_1^0$, $\widetilde{\chi}_1^0 \to \mathrm{Z} + \tilde{G}$ and $\mathrm{Z} \to \mathrm{e}^-\mathrm{e}^+$). The velocity of the neutralino can be measured from the time difference between the production and the decay vertex, and the combination with the kinematic properties of visible decay products allows the reconstruction of the neutralino mass under the assumption of a massless gravitino (Fig. 4.9). In other processes, like the production of two gluinos that decay into a quark-squark pair with the squark decaying into another quark and a neutralino, the two gravitinos in the final state (from the decay $\widetilde{\chi}_1^0 \to \tilde{G} + \gamma$) are not detected, meaning that the mass of the neutralino cannot be directly reconstructed as in the previous case. However, the neutralino lifetime can be indirectly measured from the TOF of the photon. In this case, the MTD will serve to measure the time of the primary vertex and, for converted photons, the time of arrival of the photon. In Fig. 4.9, the analysis sensitivity is shown as a function of the Λ scale (and therefore of the neutralino mass) and of the neutralino lifetime, for three different values on the timing resolution. Measuring the time of arrival to the MTD can provide significant discrimination of signal events with respect to background events also in searches of LLPs resulting in charged final states. For example, in exotic models where a Higgs boson mediates the production of two LLPs (X) decaying into quarks, the new X particles generated at the PV travel some distance before decaying into pairs of jets. The time of arrival of the charged constituents of the jets would hence be significantly higher than the time taken by a SM particle traveling from the PV to the MTD at the speed of light. Furthermore, MTD has a direct impact on the searches for Heavy Stable Charged Particles (HSCP), through the measurement of the particle velocity (β), obtained using the pathlength and the time difference between the primary vertex and the particle hits in the MTD. The velocity is indeed used to discriminate between signal ($\beta > 1$) and background SM processes ($\beta = 1$), and therefore the time resolution becomes the main factor distorting the measurement.

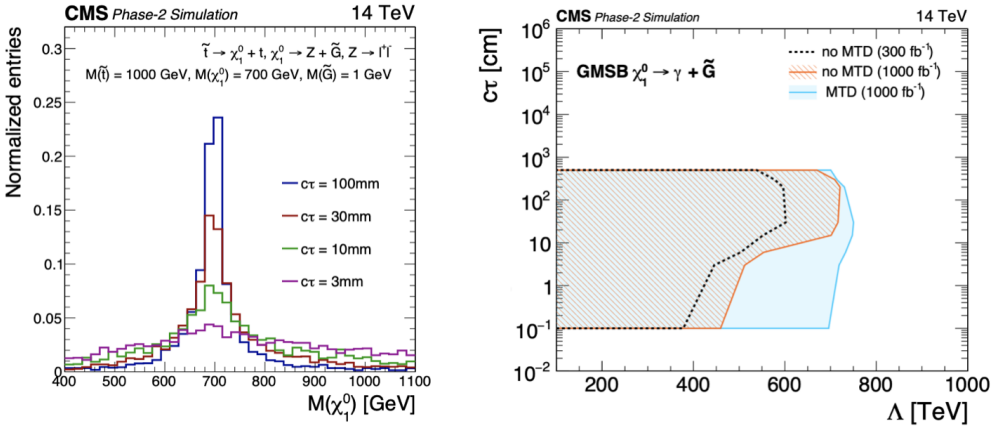

Figure 4.9: (Left) Distribution of the reconstructed mass of the neutralino, assuming a top squark mass of 1000 GeV and a neutralino mass of 700 GeV. The mass distributions are shown for different values of the $c\tau$ of the neutralino, where τ is its lifetime. (Right) Sensitivity to the $\tilde{\chi}_1^0 \rightarrow \tilde{G} + \gamma$ signals in terms of the neutralino lifetimes and the Λ scale, for 300 (red), 180 (green) and 30 (blue) ps of the resolution. The three scenarios correspond to the time resolution of the current CMS detector, and the Phase 2 detector with photon timing performed without MTD and with MTD [169].

Heavy Ion and b-physics analysis

The integration of the MTD within the CMS detector is expected to have a huge impact on the heavy-flavor analysis. Given that heavy-flavor quarks (charm and bottom) are primarily produced via initial hard scattering, and hence thay are decoupled from the bulk production of soft gluons and light-flavor quarks in heavy-ion collisions, they are used to probe the properties and dynamics of the quark-gluon plasma. The particle identification (PID) capabilities enabled by the MTD will extend the heavy flavor studies, which are currently limited to not very low-p_T regions ($p_T > 2$ GeV for D^0 mesons and $p_T > 7$ GeV for B mesons) to a larger range, from 0 up to several hundred GeV. As a case study, the decay of the $\mathrm{D}^0 \rightarrow \pi^+\mathrm{K}^-$ and of the $\Lambda_\mathrm{c}^+ \rightarrow \pi^+\mathrm{K}^- p$ are chosen. A simple track-by-track PID is performed, and the proper tracks are then combined in order to define the D^0 or the Λ_c^+ vertex. In addition, the candidates need to pass some topological selection according to the χ^2 of their decay vertex, their distance between the primary vertex, and the value of their pointing angle (defined as the angle between the line connecting the primary and decay vertices and the momentum of the particle candidates, in the plane transverse to the direction of the beam). The same analysis is performed both on the background and signal simulated sample. In Fig. 4.10, the ratios of background candidates with and without the MTD is shown, observing an overall background reduction when using the MTD information. Besides HI studies, PID plays a significant role in b-physics, where the study of CP violation in the $\mathrm{B}_\mathrm{s}^0 \rightarrow \mathrm{J}/\psi\phi(1020)$ decay is used as benchmark analysis. The performances of the flavor-tagging of the B-meson affect terms sensitive to CP violating phase, meaning that a very accurate tagging is needed to obtain a precise measurement. The figure of merit used to estimate the tagging performances is the *tagging power* P_{tag}:

$$P_{tag} = \epsilon_{tag}(1 - 2\omega_{tag})^2, \tag{4.1}$$

where $\epsilon_{tag} = N_{tag}/N_{tot}$ and $\omega_{tag} = N_{mistag}/N_{tag}$ are the tagging efficiency and the mistag fraction, respectively. The tagging power can be interpreted as an effective efficiency that takes into account the performance degradation due to events wrongly tagged: this means that a gain in the tagging power is equivalent to a similar improvement to the statistics. Currently, CMS uses a Deep Neural Network to implement a *same-side tagger*

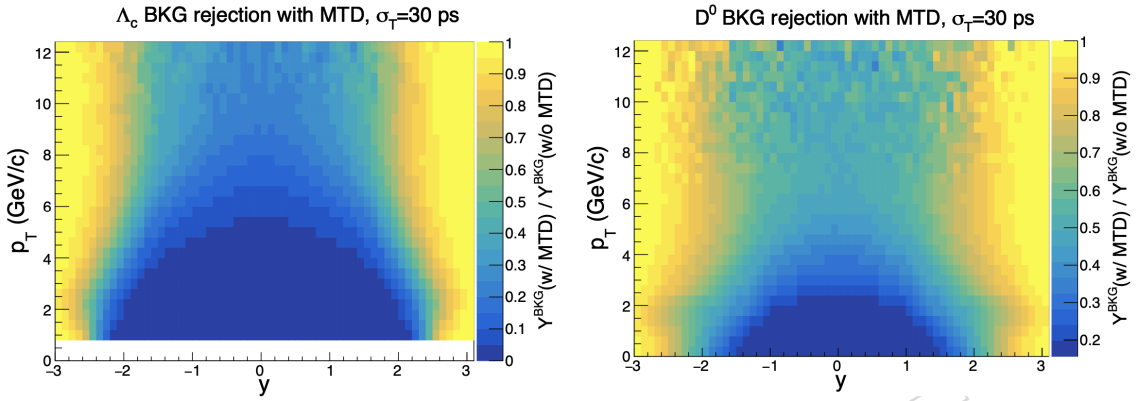

Figure 4.10: The ratios of background Λ_c^+ (left) and D^0 (right) candidates with and without the MTD, considering a 30 ps time resolution, as a function of the candidate p_T and rapidity y [169].

that exploits the charge correlation between the s-quark sign and the charge of a soft kaon coming from the same PV of the B meson. The inclusion of the PID information in the DNN decreases the effective mistag probability by more than 15% with respect to a no-PID tagger, resulting in a 24% gain in terms of tagging power for a MTD resolution of 40 ps. Such gain corresponds to an increase in luminosity of 35%.

Chapter 5

Analysis of $Z(\rightarrow \nu\bar{\nu}) + \gamma$ events

The purpose of this work is to study the diboson Zγ production where the Z boson decays in a neutrino-antineutrino pair, in order to measure the cross section and to place upper limits on the strength of the anomalous ZZγ couplings. The power of the full LHC Run 2 statistics (137.6 fb^{-1}) allows to explore regions of the phase space that are more sensitive to the possible presence of new physics effects, and that were not accessible by previous analysis, due to the lower statistics. Another remarkable aspect of this work consists in exploring forward events for the first time, including photons detected in the endcaps of ECAL. In this chapter, after listing the datasets and samples used, the requirements applied to identify the signal events are described. A special attention is paid to the criteria implemented to identify the signal photon, specifically built for this analysis using a boosted decision tree algorithm. Then, the different background contributions are estimated, defining two background categories: the first includes events that have a real photon in their final state, while the second includes events in which another physics object (like a jet or an electron) is erroneously reconstructed as a photon. Two sections are dedicated to the study of the background contribution coming from non-collision events in the ECAL barrel and endcaps, respectively. The extraction of the cross section and the limits on the anomalous triple neutral gauge couplings will be presented in Chapter 6.

5.1 Datasets and Monte Carlo samples

5.1.1 Datasets

The data used in this analysis were collected by the CMS experiment from 2016 to 2018, during the so-called LHC Run 2 data-taking period, for a total luminosity of about 137.6 fb^{-1}. During 2016, a data acquisition issue was observed, related to high-energy deposition in the bulk of silicon strip sensors: the inelastic interactions between hadrons and the nuclei of silicon sensors produced high-ionized deposits that saturated the APV25 front-end chip [172], introducing a significant dead-time (∼700 ns) into the detector readout system. To recover hits efficiency, a specific Preamplifier Feedback Voltage Bias (VFP) was set during 2016, and hence the data collected this year are further divided into pre-VFP and post-VFP datasets. For this analysis, the *SinglePhoton* (2016-2017) and the *EGamma* (2018) datasets are used, with a single photon being the signature of a $Z(\rightarrow \nu\bar{\nu})\gamma$ event. In addition, events that have been filtered online by the trigger system are used, requiring them to be accepted by the High Level Trigger paths *HLTPhoton200* (2017-2018) and *HLTPhoton175* (2016). This means that only events that contained at least one high-energetic photon whose transverse momentum exceeds the 200 (175) GeV threshold were selected. In this way, the number of the processed collision can be reduced during data taking to a rate manageable by the computing facilities. The higher threshold adopted

for 2017 and 2018 is due to the higher pileup rate recorded during these years, as a consequence of the increased luminosity. All datasets and associated luminosities are listed in Tab. 5.1 to 5.3. The analysis is performed separately for each year, and then the results are combined together.

Run	Number of Events	Luminosity (fb^{-1})
2016B-preVFP	56878553	5.83
2016C-preVFP	23147235	2.6
2016D-preVFP	29801360	4.29
2016E-preVFP	22322869	4.07
2016F-preVFP	12806145	2.72
2016F-postVFP	1860761	0.42
2016G-postVFP	33288854	7.65
2016H-postVFP	35035661	8.74

Table 5.1: Name, number of events, and luminosity of the 2016 datasets used for the analysis.

Run	Number of Events	Luminosity (fb^{-1})
2017B	15950935	4.80
2017C	42182948	9.57
2017D	9753462	4.25
2017E	19011446	9.31
2017F	29783015	13.54

Table 5.2: Name, number of events, and luminosity of the 2017 datasets used for the analysis.

Run	Number of Events	Luminosity (fb^{-1})
2018A	339013231	14.03
2018B	153822427	7.06
2018C	147827904	6.89
2018D	752528074	31.83

Table 5.3: Name, number of events, and luminosity of the 2018 datasets used for the analysis.

5.1.2 Monte Carlo samples

Monte Carlo (MC) event generators are employed to produce simulations of the final states of high-energy collisions: they have manifold applications in all physics analyses. With simulated samples one can correctly compare the observed data with the predictions, and handle the irreducible background contributions arising from physics processes that share exactly the same signature of the final state of interest, but that cannot be removed trough event selection only. In addition, the simulated signal samples including the contributions arising from the presence of the anomalous couplings are needed for the estimation of the limits. In Tab. 5.4, the list of the signal and background samples used in this analysis are shown, together with the precision order at which they are generated. All the samples are centrally produced by CMS, with the only exception of the signal samples

Sample	Cross section (fb^{-1})	Generator order
$Z(\to \bar{\nu}\nu)\gamma$ $(200 < p_T^{\gamma} < 600)$	49.49	LO
$Z(\to \bar{\nu}\nu)\gamma$ $(600 < p_T^{\gamma} < Inf)$	0.774	LO
$\gamma\gamma$	127255	NLO
$\gamma +$ jets $(200 < H_T < 400)$	2191211	LO
$\gamma +$ jets $(400 < H_T < 600)$	259944	LO
$\gamma +$ jets $(600 < H_T < \infty)$	86332	LO
$t\gamma$	2997	NLO
$t\bar{t}\gamma$	3783	NLO
$W\gamma$	806.83	LO
$W(\to \mu\nu)$	174162	LO
$W(\to \tau\nu)$	174132	LO
WW	75940	LO
WZ	27564	LO
ZZ	12135	LO
$Z(\to \ell\bar{\ell})\gamma$	205.72	NLO

Table 5.4: List of the Monte Carlo signal and background samples. The value of the cross section and the order at which the samples are generated are also presented.

$Z(\to \nu\bar{\nu})\gamma$, that are generated privately using Madgraph5_aMC@NLO [173, 174], importing in the datacards the Universal FeynRules Output (UFO) model [175] that accounts for the couplings' variations. The $\gamma\gamma$, $t\bar{t}\gamma$ and the $Z(\to \ell\bar{\ell})\gamma$ samples are generated with Madgraph5_aMC@NLO and interfaced with Pythia8 [176] for the parton shower. The FxFx matching scheme [177] is used for the jet matching at NLO. For the $t\gamma$ sample the same generators are used, and the NLO corrections are included using MadSpin [178]. The $\gamma +$ jets samples are generated at LO using Madgraph5_aMC@NLO interfaced with Pythia8, performing the jet matching at LO with the MLM merging scheme [179, 180]. The $W\gamma$ sample is produced at LO using Madgraph5_aMC@NLO interfaced with Pythia. The WW, WZ, ZZ, $W(\to \mu\nu)$ and $W(\to \tau\nu)$ samples are also generated at LO, with Pythia8 only. The latter sample is interfaced with Tauola [181], specific for the generation of the tau-lepton decays. For all the samples, the latest CP5 MC tune [182] and the recommended NNPDF3.1 set [183] of parton distribution functions (PDFs) [184, 185] at the NLO precision is used (with $\alpha_S = 0.118$). The PDFs $f_i(x, Q^2)$ provide the probability of finding in the proton a parton i carrying a fraction x of the total proton momentum, with Q being the energy scale of the hard process. Their functional forms depend on the detailed dynamics of the proton, therefore they are not *a priori* known and have to be determined from experimental measurements, as for example Deep Inelastic Scattering (DIS) experiments like those performed at HERA (Hadron Elektron Ring Anlage), where electrons (or positron) were collided with protons at a center of mass energy between 225 and 318 GeV [186].

5.2 Event selection

The $Z(\to \nu\bar{\nu})\gamma$ final state is characterized by one high-energy photon, and high missing transverse energy (MET) E_T^{miss} originating from the two neutrinos of the Z boson decays. The PF E_T^{miss} is used (see Sec. 3.3.4), defined as follows:

$$E_T^{miss} = -\sum_i |\vec{p}_T(i)| \equiv -\sum_i p_T(i), \tag{5.1}$$

where p_T is the transverse momentum of a visible particle i. In addition to the cross section measurement, this analysis aims to place limits on aNTGCs, whose presence in Zγ processes would manifest as a deviation from the SM prediction at high photon p_T (Fig. 5.1). For this reason, exactly one photon with $p_T^{\gamma} > 225$ GeV, and $E_T^{miss} > 200$

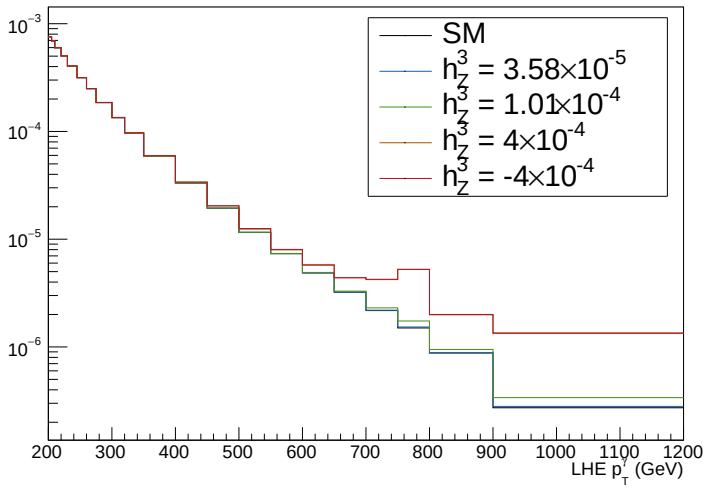

Figure 5.1: Examples of aNTGC excesses over the SM predictions in the photon p_T spectrum.

GeV are required in the event. The recommended MET filters are applied to mitigate the effect of possible bad reconstruction of the missing transverse energy, or that of sporadic noise signals affecting the MET evaluation (as was observed for example in the HB and HE sections of HCAL). The photon is selected using a dedicated identification (ID) method, described in Sec. 5.3. Its associated supercluster (SC) in the electromagnetic calorimeter (ECAL) is required to lie in the barrel/endcaps (EB/EE), i.e. $|\eta_{SC}^{\gamma}| < 1.4442/1.566 < |\eta_{SC}^{\gamma}| < 2.5$. The photon must not have a pixel seed in the pixel detector (which means that a *pixel veto* is applied), and it has to be matched with the photon object at the HLT level considering $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.3$. In order to suppress the beam halo contribution in the ECAL barrel that could produce spurious photons, the requirement $MIP_{tot} < 4.9$ GeV is applied to EB photons, where MIP_{tot} estimates the beam halo deposits in ECAL by summing the energies of hits along a beam parallel line intersecting the SC. The EB photons are also required to have $\sigma_{i\eta i\eta} > 0.001$ and $\sigma_{i\phi i\phi} > 0.001$, in order to suppress the contribution of spike events in the ECAL barrel, and for the same reason an additional selection on the variable "eta wing", $\eta_W > 0.01$, is applied. The description of this new variable is provided in Sec. 5.4.4. One of the major background including one real photon in the final state arises from $W(\rightarrow \ell\nu)\gamma$ processes where the lepton is lost or badly reconstructed: to suppress it, events with an electron whose SC is in EB/EE ($|\eta_{SC}^{e}| < 1.4442/1.566 < |\eta_{SC}^{e}| < 2.5$), or a muon with $|\eta^{\mu}| < 2.4$ are vetoed. Leptons have to pass loose identification criteria and have $p_T > 10$ GeV. In order to remove the contribution coming from events with fake E_T^{miss} due to limited jet energy resolution (as in the case of γ+jets events) events in which there is at least one jet with $p_T > 30$ GeV, passing tight identification criteria, and with $p_T^{jet}/E_T^{miss} < 1.4$ and $\Delta\phi_{\min}(\text{jet}, E_T^{miss}) > 0.5$ are vetoed. A separation of $\Delta\phi(E_T^{miss}, \gamma) > 2$ between the photon and the MET is applied to suppress $W(\rightarrow e\nu)$ events in which the electron does not have the pixel seed. During the 2018 data taking, the power supply of two modules of HCAL stopped working. This impacted the jet energy measurement for -1.57 $< \phi <$ -0.87 and -3 $< \eta <$ -1.3, and hence events that have a jet with $p_T > 15$ GeV in this region are excluded from the phase

space. The selected photon is also required to be detected out of this region. The two modules involved are labelled as Hadronic Endcaps Minus (HEM) 15 and 16, therefore this particular veto in the η-ϕ region is termed the *HEM veto*.

5.3 Photon hybrid identification method

As mentioned beforehand, the signature of Zγ events consists in one energetic photon and high MET. The identification of photon candidates becomes thereby a crucial factor for achieving reliable and competitive results, and to put stringent limits on aNTGCs. The available identification criteria provided by the CMS e/γ (EGM) group are not explicitly tuned for photons with such high transverse momentum. Therefore, a dedicated high-p_T photon ID, as a hybrid of a shower-shape boosted decision tree (BDT) and a cut-based ID, has been developed. The ID is defined by optimized selections on four variables:

- *BDT score*, described in detail in the next section.
- Hadronic over electromagnetic energy ratio (H/E): the H/E ratio is defined as the ratio between the energy H deposited in HCAL in a cone of radius $\Delta R = 0.15$ around the SC direction and the energy E of the photon candidate. The three main sources that contribute to measured hadronic energy of a genuine electromagnetic object are: HCAL noise, pileup and leakage of electrons or photons through the inter-module gaps [142].
- ECAL isolation (I_{ECAL}): the scalar sum of p_T of all ECAL PF candidates in a ΔR = 0.3 cone around the photon direction in the ECAL.
- Tracker isolation (I_{Tkr}): for 2017 and 2018 is used the scalar sum of p_T of all tracks in a $\Delta R = 0.3$ cone around the direction of the photon, but outside $\Delta R = 0.05$, where $\gamma \rightarrow$ ee conversions can create tracks. Due to the loss in tracker efficiency caused by chip saturation in the 2016 data-taking, a less susceptible alternative for this year is preferred, defined as the maximum sum of the energies of all charged HCAL PF candidates matched to a vertex in a $\Delta R = 0.3$ cone around the photon direction.

For the development of the ID, the aim is to achieve the maximum signal efficiency that can be obtained while maintaining a similar background rejection as the official EGM medium ID, whose signal efficiency is $\epsilon_S = 80\%$. For the hybrid ID, this is achieved at $\epsilon_S = 92\%$.

Shower Shape BDT

The electromagnetic shower in a reconstructed photon's ECAL cluster can be described by variables formed from ratios of energies in crystal matrices and by energy-weighted variances and covariances of the crystal coordinates. As showers from jet fakes are broader than those from real photons, the shower shape variables have distinct distributions that can be used to discriminate between real and fake photon candidates. Exploiting the power of machine learning algorithms, a BDT is trained with XGBoost 1.2.0 [187, 188], in order to convert the information contained in several shower shape variables into a single number called *BDT score* in the range 0-1 (see Fig. 5.2), that has higher discrimination power than any single shower shape variable alone. A total of 14 discriminating shower shape variables for EB and 15 for EE are identified (shown in Fig. 5.4 and 5.5) to train separate BDTs for the two sub-detector sections. In addition, p_T and $|\eta_{SC}|$ are also included in the input features, to account for any dependence that the shower shapes may have on them. The shower shape variables entering the BDT algorithm are the following:

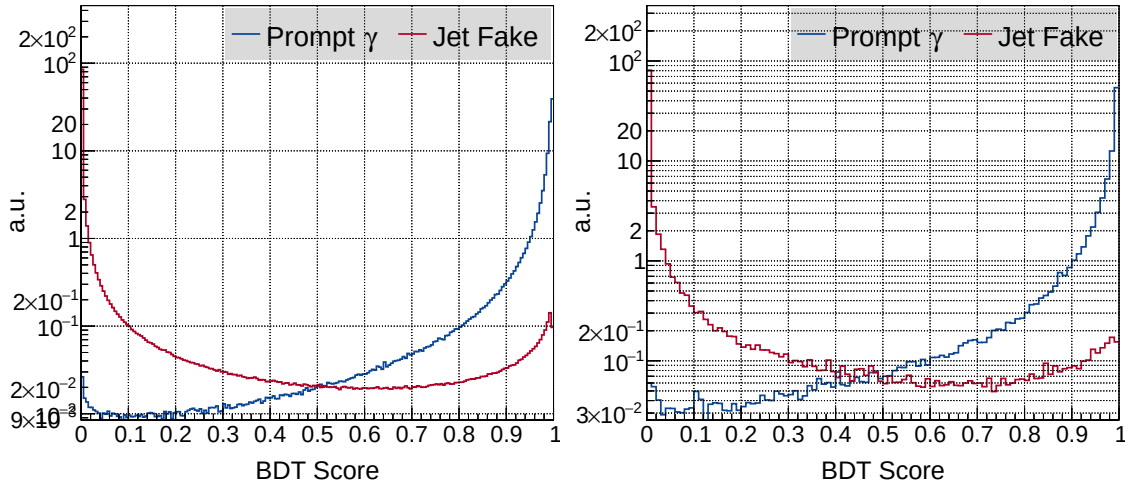

Figure 5.2: BDT distributions for simulated prompt (blue) and fake (red) photons for 2017 EB (left) and EE (right). The overlap between signal and background distributions is lower here than in any single shower shape variable, hence providing a better discrimination power.

- E_{max}/E_{SC}^{raw}: ratio of crystal's highest energy over total SC's energy;
- E_2/E_{SC}^{raw}: ratio of the crystal's second highest energy over total SC's energy;
- $E_{3\times3}/E_{SC}^{raw}$: ratio of highest energy 3×3 block containing seed over total SC's energy;
- $E_{2\times2}/E_{3\times3}$: ratio of highest energy 2×2 block over highest energy 3×3 block, with both blocks containing seed;
- $E_{2\times2}/E_{SC}^{raw}$: ratio of highest energy 2×2 block containing seed over total SC's energy;
- $E_{1\times3}/E_{SC}^{raw}$: ratio of highest energy 1×3 block containing seed over total SC's energy;
- $E_{2\times5}/E_{SC}^{raw}$: ratio of highest energy 2×5 block containing seed over total SC's energy;
- $\sigma_{i\eta i\eta} = \sqrt{\frac{\sum_i^{5\times5} w_i \left(\eta_i - \bar{\eta}_{5\times5}\right)^2}{\sum_i^{5\times5} w_i}}$, where $w_i = 4.2 + \ln(E_i/E_{5\times5})$;
- $\sigma_{i\phi i\phi} = \sqrt{\frac{\sum_i^{5\times5} w_i \left(\phi_i - \bar{\phi}_{5\times5}\right)^2}{\sum_i^{5\times5} w_i}}$;
- $\sigma_{i\eta i\phi} = \sqrt{\frac{\sum_i^{5\times5} w_i \left(\eta_i - \bar{\eta}_{5\times5}\right)\left(\phi_i - \bar{\phi}_{5\times5}\right)}{\sum_i^{5\times5} w_i}}$;
- $\sigma_{i\eta i\eta}/\sigma_{i\phi i\phi}$;
- $\sigma_\eta = \sum_i^{\mathrm{SC}} \sqrt{\frac{\mathrm{E_i}}{\mathrm{E_{SC}}} \left(\eta_\mathrm{i} - \eta_\mathrm{SC}\right)^2}$;
- $\sigma_\phi = \sum_i^{\mathrm{SC}} \sqrt{\frac{\mathrm{E_i}}{\mathrm{E_{SC}}} \left(\phi_\mathrm{i} - \phi_\mathrm{SC}\right)^2}$;
- σ_η/σ_ϕ;
- σ_{RR} (EE) $= \sqrt{\sigma_{xx}^2 + \sigma_{yy}^2}$, where σ_{xx} and σ_{yy} measure the lateral spread in the two orthogonal directions of the sensor planes of the preshower detector.

The XGBoost algorithm builds a large number of decision trees, taking as input γ+jets and $W(\rightarrow \ell\nu)\gamma$ MC samples, that contribute for the prompt photons, and a MC QCD sample that contributes for the description of the fake photons. Each tree is a function that calculates a *score* from the input features of a candidate photon. The *score* for each tree is summed and scaled to determine the total *BDT score*. Four BDTs are trained: 2 for EB and EE, and 2 for 2016 and 2017/2018. The input sample is split into a training sample (40% statistics), a testing one (30%) and a validating one (30%). The statistics of the training samples are listed in Tab. 5.5. The samples are reweighted to produce flat 2D distribution in (p_T, η_{SC}) so that the BDT performed uniformly in these variables, and the hyperparameters (listed in Tab. 5.6) for the BDT are tuned using the *Hyperopt* [189, 190] tool. The parameters *max_depth* and *min_child_weights* optimize complexity, *alpha, lambda* and *gamma* optimize regularization, *subsample* and *colsample_bytree* the robustness, and *eta* the learning rate. The receiver operating characteristic (ROC) curves for the four BDTs are shown in Fig. 5.3.

	EB		**EE**	
	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 5.5: Number of events in the training samples used for photon BDT training.

	max_depth	eta	gamma	lambda	alpha	min_child_weight	subsample	colsample_bytree
EB	4	0.12	0.4	8.0	145	4740	0.51	0.69
EE	5	0.15	3.64	4.6	47.2	3.7	0.41	0.65

Table 5.6: The XGBoost hyperparameters used for training. They were tuned to minimize the difference in performance between the training sample and the validation sample.

Cut Optimization

In order to define the selections on the ID variables (*BDT score*, H/E, I_{ECAL} and I_{Tkr}) a 4D optimizer is developed. However, before entering the optimizer, isolation variables need to be corrected for pileup and p_T dependence. This is because the isolation cones may contain contributions from the products of pileup interactions that are not associated with the hard collision, and that have to be subtracted. In a similar way the energy footprint of the photon, which increases as its transverse momentum increases, has to be removed from the isolation cone. In order to correct for pileup (p_T), the standard procedure starts from the distribution of the isolation as a function of ρ (p_T) on a 2D histogram. The variable ρ is defined as the event-specific average pileup energy density per unit area in the (η, ϕ) plane. The 92nd percentile is located for each ρ (p_T) bin and the corresponding points are then fitted with a linear function:

$$I^{corr}(\rho) = a \cdot \rho + b.$$

The same parameterization is used for $I^{corr}(p_T)$. The final isolations values are then obtained by subtracting $I^{corr}(\rho)$ and $I^{corr}(p_T)$ (the latter is applied on I_{ECAL} only) from the uncorrected isolations. The corrections are evaluated in 6 bins of $|\eta_{SC}|$ (see Tab. 5.7, 5.8 and 5.9).

The isolation variables are then given as input to the optimizer, together with H/E and the *BDT score*. Starting with an initial search grid, the optimizer recursively narrows the search grid and descends towards variables' boundaries that maximize the signal efficiency while maintaining a background rejection similar to the EGM Medium ID. The selections

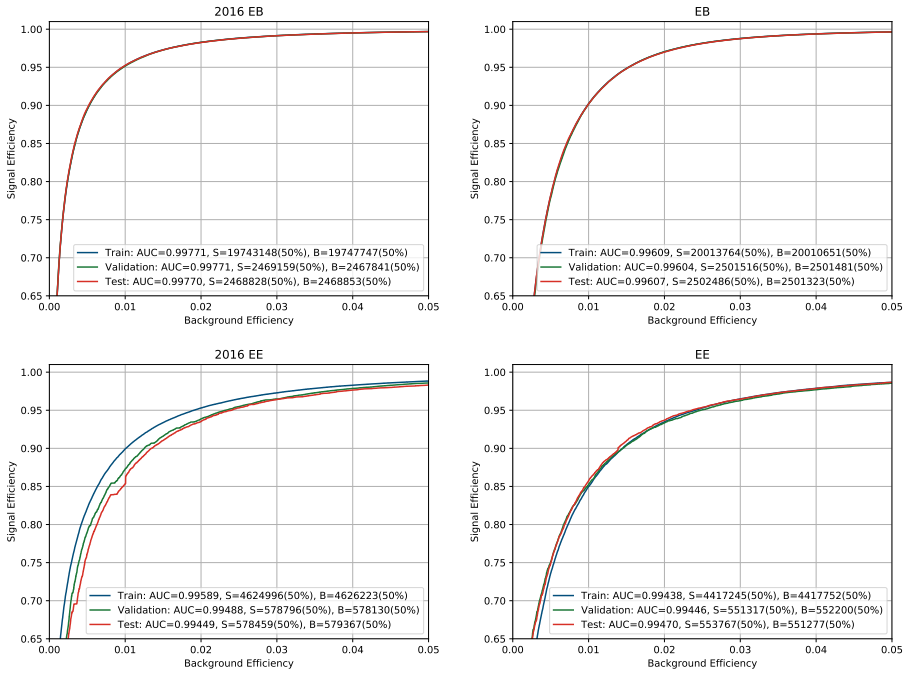

Figure 5.3: Photon BDT ROC curves for 2016 EB (top left) and EE (bottom left), and for 2017/18 EB (top right) and EE (bottom right). The test samples are independent of the training samples, and they show similar performance.

$\lvert\eta_{SC}\rvert$	*2016*	*2017*	*2018*
0 - 0.5	$0.1253\rho + 0.45$	$0.1073\rho + 0.57$	$0.1085\rho + 0.50$
0.5 - 1	$0.1263\rho + 0.51$	$0.1106\rho + 0.57$	$0.1086\rho + 0.50$
1 - 1.4442	$0.1165\rho + 0.53$	$0.1031\rho + 0.60$	$0.0965\rho + 0.51$
1.566 - 1.7	$0.0815\rho + 0.23$	$0.0728\rho + 0.24$	$0.0654\rho + 0.33$
1.7 - 2	$0.0872\rho + 0.24$	$0.0807\rho + 0.25$	$0.0769\rho + 0.27$
2 - 2.5	$0.1150\rho + 0.22$	$0.1016\rho + 0.17$	$0.0845\rho + 0.15$

Table 5.7: Pileup (ρ) corrections to ECAL isolation I_{ECAL}.

obtained for the 3 years are listed in Tab. 5.10. The ID efficiencies for prompt and fake photons are shown in Appendix B, along with the reference performances of the EGM IDs. The use of BDTs trained with several shower shape variables clearly leads to significant improvement in signal-background discrimination, and the signal efficiency achieved is 92%.

Scale factors

Once the new photon IDs is computed for the three years, the respective scale factors (SFs) are estimated accounting for efficiency differences in data and simulation. For their evaluation, the *Tag and Probe* method is used, selecting a tag electron with $p_T > 37$ GeV, HEEP (High-energy Electron Pairs) ID and $|\eta| < 2.5$, and a probe photon in EB (EE) with $p_T > 225$ GeV and a pixel seed, in order to suppress contribution from true photons. The tag and the probe must be separated by $\Delta R(\mathrm{e}, \gamma) > 0.3$, and the reconstructed Z mass must lie between 80 and 110 GeV. The events are required to pass the analysis HLT and to have $E_T^{miss} < 160$ GeV. Once the tag+probe events are selected, the efficiency is computed respectively in data and simulation as

$$\epsilon = \frac{\#passing\ probes}{\#passing\ probes\ +\ \#failing\ probes}.$$

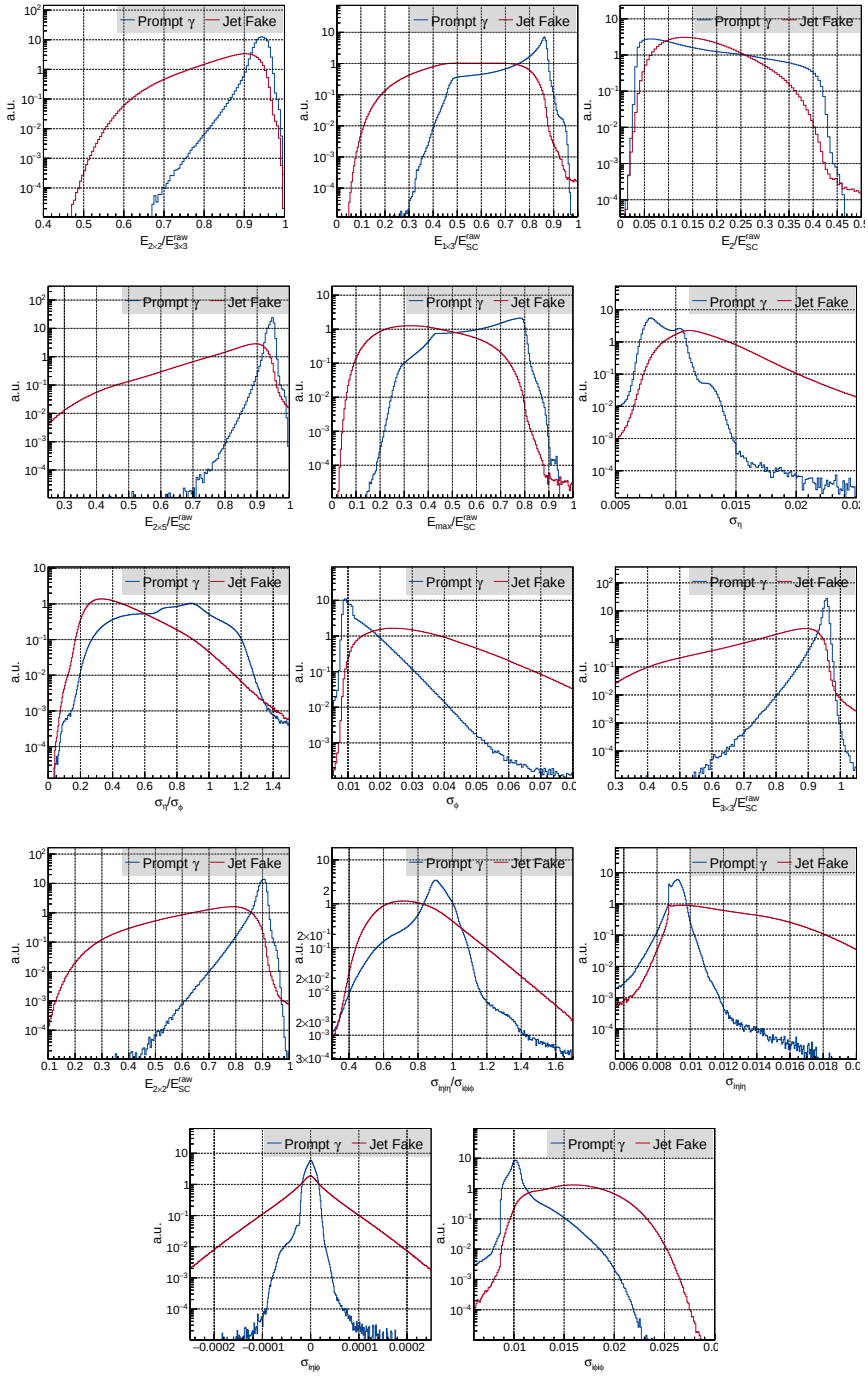

Figure 5.4: The 14 shower shape variables used as input features to the photon BDT in EB. The distributions are obtained using simulated samples.

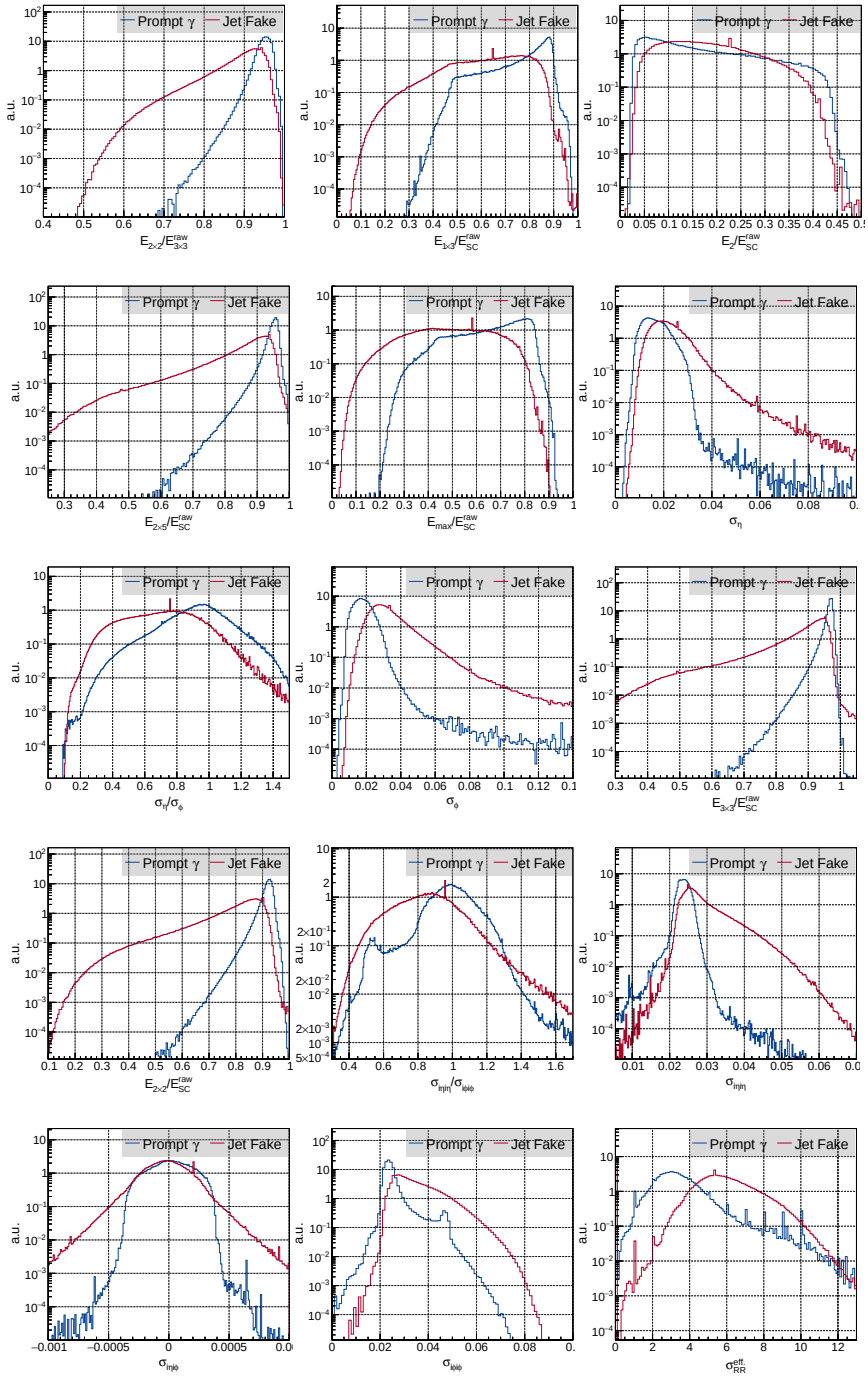

Figure 5.5: The 15 shower shape variables used as input features to the photon BDT in EE. The distributions are obtained using simulated samples.

$\lvert\eta_{SC}\rvert$	*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 5.8: Pileup (ρ) corrections to tracker isolation I_{Tkr}.

$\lvert\eta_{SC}\rvert$	*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 5.9: p_T corrections to ECAL isolation I_{ECAL}.

The passing (failing) probe events are those in which the probe photon has passed (failed) the new photon ID. A SF is hence defined as the ratio between the efficiency obtained in data and in simulation, $\frac{\epsilon_{data}}{\epsilon_{MC}}$. The efficiencies are computed in η-p_T bins, and they are shown in Appendix B, together with the SFs. For data, the number of passing and failing events is estimated by fitting the passing and failing Z distribution and extracting the number of signal events, while in simulation a simple *cut-and-count* method is used. The signal fitting function used in the data is obtained by extracting the line shape from the simulated passing and failing distributions, and then smearing them with a Gaussian. The background is modelled using the product of an exponential decay function and an error function. The following uncertainties have been addressed on the evaluated scale factors:

- Signal fitting choice: a Double Crystal Ball function convoluted with a Gaussian is taken as alternative signal (using the nominal background function).
- Background fitting choice: an exponential decay function is taken as alternative background (using the nominal signal function).
- Scale factor p_T-dependency: for very high values of the transverse momentum, the direct measurement of the scale factors is not performed due to the reduced statistics.

		2016	**2017**	**2018**
EB	*BDT Score* $>$	0.8493	0.8361	0.8466
	$H/E <$	0.0222	0.0401	0.0402
	I_{ECAL}/GeV $<$	2.16	1.84	1.84
	I_{Tkr}/GeV $<$	2.19	1.63	1.58
EE	*BDT Score* $>$	0.5714	0.6035	0.6948
	$H/E <$	0.0315	0.0171	0.0175
	I_{ECAL}/GeV $<$	2.69	2.54	2.94
	I_{Tkr}/GeV $<$	3.10	1.82	1.90

Table 5.10: Selections applied on the *BDT score*, H/E and isolation variables for the definition of the photon ID in the three years.

Hence, for photons in EB (EE) that have $p_T > 1000$ (800) GeV, the last p_T − bin scale factor is considered, relying on the assumption that the scale factor distribution is flat over p_T. For this reason, an additional systematic for this extrapolation is needed. To do so, the SF distribution as a function of p_T is fitted with a linear function, and the error on the slope has been taken as the 68% limit on the gradient, providing the error band that gives the additional systematics (see Appendix B).

Validation with data and shape's quantile-quantile corrections

Since the BDT is trained using simulated samples only, the shape of the *BDT score* has to be validated on data. γ+jets events required to pass the analysis HLT and to have $E_T^{miss} <$ 160 GeV are used. Each event must have at least one jet with with $p_T > 200$ GeV, Tight ID and $|\eta| < 5$, and at least one photon in EB (EE) with $p_T > 225$ GeV. In order to reduce the contribution from the QCD background, the following requirements are applied on the photon [191]:

- Pixel seed veto;
- H/E < 0.02148;
- PF Charged Isolation (ρ-corrected) < 0.65 GeV;
- PF Neutral Isolation (ρ-corrected) $< 0.317 + 0.01512 \cdot p_T + 2.259\ 10^{-5} \cdot p_T^2$ GeV;
- Track isolation (ρ-corrected) < 1 GeV;
- HCAL PF cluster isolation (ρ-corrected) < 0 GeV.

In addition, the photon and the highest p_T jet have to be separated in $\Delta\phi$, i.e. $\Delta\phi(jet, \gamma) > 2.9$, and a maximum of two additional jets with $p_T > 30$ GeV and Loose ID is allowed in the event. The appropriate photon variables are then given as input to the BDT algorithm, and the distributions of the *BDT score* in data and MC are compared (see Appendix C). The *BDT score* distribution is used in the evaluation of the jet-to-photon fake ratio, as it will be described in Section 5.4.3, and hence the discrepancy observed in the BDT shape between data and simulation can give rise to a biased estimate of the fake ratio. Such discrepancy is corrected for using the quantile morphing method [192], thus assessing a systematic uncertainty on the fake ratio. This method was originally developed in the context of $H \to \gamma\gamma$ analysis [193] and it is based on comparisons of Cumulative Distribution Functions (CDF) in the data and simulation taking into account all the backgrounds in the simulation (see Fig. 5.6). A transformation factor *t(x)* is evaluated as a function of

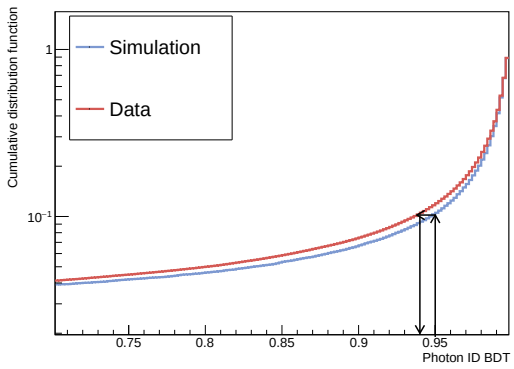

Figure 5.6: Figure illustrating the idea of quantile morphing in simulation to get the desired shape of the BDT photon ID as it is in data.

the *BDT score* x as follows:

$$t(x) = CDF_{DT}^{-1}(CDF_{MC}(x)), \tag{5.2}$$

where CDF_{DT} and CDF_{MC} are the corresponding CDFs of the photon ID BDT distributions in data and simulation. In order to retrieve from the simulated BDT distributions the shape observed in the data, the transformation factor has to be applied to all the simulated samples. An orthogonal $Z \to e^- e^+$ sample is also used to check the agreement between data and simulation, giving us further confidence about the usage of these corrections. Electron-photon pairs are selected as follows:

- Electron with $p_T > 37$ GeV, HEEP ID and $|\eta| < 2.5$;
- Photon in EB (EE) with pixel seed, $p_T > 225$ GeV and H/E < 0.05;
- Photon-corrected track isolation < 18 GeV;
- Photon-corrected PF HCAL isolation < 10 GeV;
- $\Delta R(e, \gamma) > 0.3$;
- 80 GeV $<$ Z($\to e\gamma$)mass $<$ 110 GeV.

Once the eγ pair has been selected, the appropriate photon variables are given in input to the BDT algorithm and the *BDT score* is evaluated. However, the correction factor evaluated in Eq. 5.2 cannot be applied directly to the $Z \to e^- e^+$ events, since the electron and photon have different longitudinal shower shapes and hence their BDT distributions differ. For this reason, a further transformation of the electron BDT in both the data and the simulation is needed. A second transformation factor ($t_{ele}(x)$) is calculated, comparing the CDF of the photon BDT distributions for the photon (γ+jets) and the electron ($Z \to e^- e^+$) samples:

$$t_{pho}(x) = CDF_{pho}^{-1}(CDF_{ele}(x)) \tag{5.3}$$

This factor has to be applied to the simulated samples before *t(x)*. In Appendix C, the comparisons between data and simulation in $Z \to e^- e^+$ and γ+jets events before and after the corrections are shown. The overall agreement is improved in both the orthogonal samples.

5.4 Background estimation

The background processes that mimic the Zγ final state can be divided in two categories, listed below: those who have one real photon in the final state, and those in which another physics object (mainly a jet or an electron) has been mistakenly reconstructed as one.

- Prompt photons backgrounds:
  - $W(\to \ell\nu)\gamma$;
  - γ + jets (estimated using simulation);
  - $t\gamma, t\bar{t}\gamma, Z(\to \ell\bar{\ell})\gamma, WW, WZ, ZZ, \gamma\gamma, W(\to \mu\nu), W(\to \tau\nu)$ (estimated using simulation).
- Fake photons backgrounds:
  - Electrons misidentified as photons;
  - Jets misidentified as photons;

– Spikes (in EB);
– Beam halo.

In the following sections, the methods used to evaluate the contribution of the background sources that cannot be directly estimated from simulation are described.

5.4.1 Wγ background contribution

A relevant contribution to the total background yield is given by $W(\to \ell\nu)\gamma$ processes where the lepton happens to be badly reconstructed, or not detected at all. This background is estimated in dedicated control regions, and *transfer factors* are applied in order to evaluate the $W(\to \ell\nu)\gamma$ yield in the signal region. Each transfer factor is defined as the ratio of the expected yields as given by simulation in a given bin of the signal region, N_{MC}^{SR}, and the corresponding bin of one of the control samples, N_{MC}^{CR}. One then extrapolates the expectation for the event yield in a bin of the signal region, N_{obs}^{CR}, from the event yield measured in the corresponding bin of the control sample, N_{obs}^{CR}, via the expression:

$$N_{pred}^{SR} = \frac{N_{MC}^{SR}}{N_{MC}^{CR}} \cdot N_{obs}^{CR}. \tag{5.4}$$

Two $W(\to \ell\nu)\gamma$ control regions are selected. One, CR_e, is defined via the electronic decay of the W. The other, CR_μ, is defined using the W muonic decay. The HLT requirement, MET filters and the selection of the photon remain the same as described in Sec. 5.2. The selections used for the leptons are shown in Tab. 5.11. The jet veto described in Sec. 5.2 is used, with the additional condition that $\Delta R(\ell, jet) > 0.5$. In addition, $E_T^{miss} > 50$ GeV and $m_T(\ell, E_T^{miss}) < 160$ are required. Some selections on the transverse recoil $\vec{R}_T$ are also applied, namely, $R_T > 200$ GeV, $p_T^\gamma/R_T < 1.4$ and $\Delta\phi(R_T, \gamma) >$, where $\vec{R}_T = \vec{E}_T^{miss} + \vec{p}_T^{\,\ell}$ and ℓ identifies the electron or the muon. The *HEM veto* is applied as in signal region, with the additional condition that neither the lepton should lie in that region. The plots for the p_T yieds in the control regions and additional kinematic variables are shown below, for EB and EE events, in Fig. 5.7 and Fig. 5.8.

CR_e	CR_μ
$p_T^e > 30$ GeV	$p_T^\mu > 30$ GeV
$\lvert\eta_{SC}^e\rvert$ <1.4442 or 1.566 < $\lvert\eta_{SC}^e\rvert$ <2.5	$\lvert\eta^\mu\rvert$ <2.4
Tight ID	Tight ID
$\Delta R(\mathrm{e}, \gamma) > 0.5$	-

Table 5.11: Electron and muon selections applied in CR_e and CR_μ, respectively.

5.4.2 Electron-photon misidentification background contribution

As described in Sec. 3.3.4, the reconstructed electrons in CMS can be tracker driven or ECAL driven, with the ECAL driven ones requiring the presence of at least 2 or 3 hits in the pixel detector of the tracker. Hence, if an ECAL driven electron is not registered in the inner layers of the pixel detector it may be reconstructed as an energy deposit in the ECAL without a track, which is consistent with the signature of a photon. This contribution of electrons faking photons comes mainly from single boson and diboson production such as $\mathrm{W} \to \mathrm{e}\nu$ and $\mathrm{W}(\to \mathrm{e}\nu)\mathrm{Z}(\to \nu\nu)$, that may mimic our final state. To estimate this fake ratio we use $\mathrm{Z} \to \mathrm{e}^-\mathrm{e}^+$ events and the *Tag and Probe* method, looking for electrons that fail to be registered by the pixel detector. The following selection is applied to the data, looking for $\mathrm{Z} \to \mathrm{e}^-\mathrm{e}^+$ events in which one of the two electrons is reconstructed as a photon:

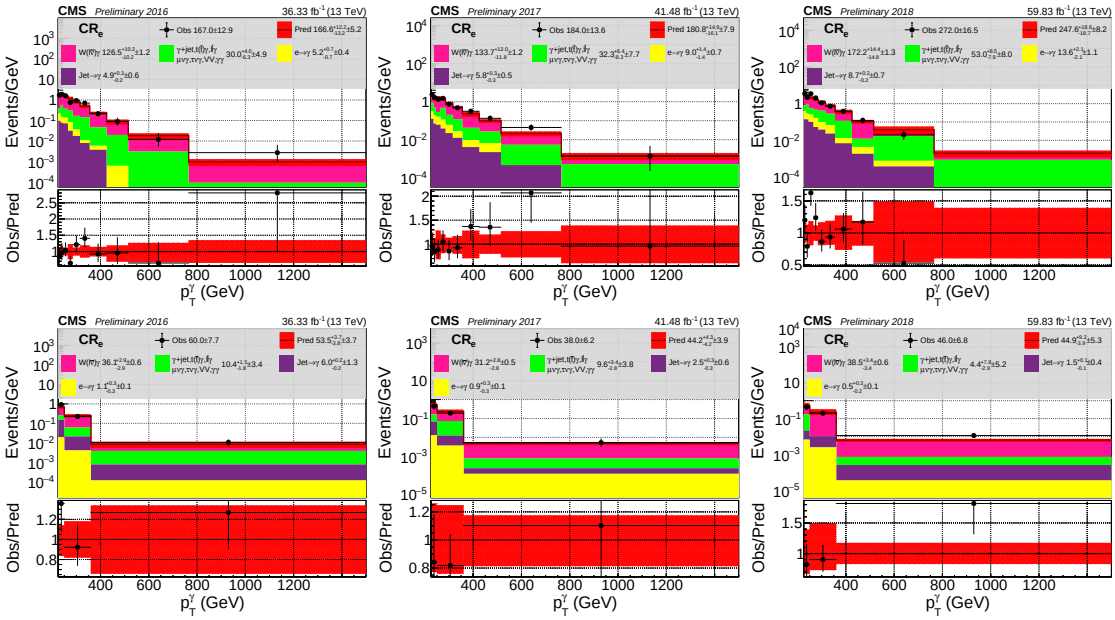

Figure 5.7: Distribution of photon p_T in the W($\rightarrow$ eν)γ control region for 2016 (left), 2017 (center) and 2018 (right) for events in EB (top) and EE (bottom).

- HLT_DoublePhoton70, HLT_DoublePhoton60 (2016);
- $E_T^{miss} < 160$ GeV;
- Electron with $p_T > 20$ GeV, passing Tight ID;
- Photon with $p_T > 200$ GeV, passing the analysis (hybrid) ID;
- 60 GeV $<$ Z($\rightarrow$ eγ) mass $<$ 120 GeV;
- $\Delta R(\mathrm{e}, \gamma) < 0.05$.

This fake ratio is highly dependent on the photon's pseudorapidity since the pixel inefficiency is η-dependent, and hence its estimation is done in bins of $|\eta|$. This dependence arises because, at higher pseudorapidities, an electron will share charge with more pixels, and hence will have a higher probability of being detected. In each bin, a fit to the dielectron-invariant mass (around the Z peak) is done for passing and failing probes simultaneously, in order to evaluate the number of photons passing and failing the pixel veto. The signal fitting function is obtained applying the selection above and extracting the line shape of the Z mass from a Z $\rightarrow$ e$^-$e$^+$ MC sample. The background is modeled using an exponential function. The fake ratio (FR) for an electron to be misidentified as a photon is then defined as:

$$FR = (1 - \epsilon)/\epsilon, \tag{5.5}$$

where ϵ is the efficiency to reconstruct an electron in the $Z \rightarrow$ e$^-$e$^+$ final state and (1-ϵ) is then the fraction of electrons which do not have pixel hits in the tracker. During the three years of data-taking, some sensors in the pixel layers of the tracker experienced bad functioning, resulting in an important decrease of the seeding efficiency in some η-ϕ regions. Since seeding inefficiencies would determine a higher fake ratio, remarkably increasing the probability of an electron to be reconstructed as a photon if the pixel sensors are off, a dedicated FR is estimated in the regions affected by such inefficiencies. For the remaining regions, the fake ratio is instead estimated in 3 bins of $|\eta|$, and then fitted with a 2$^{\mathrm{nd}}$ order polynomial. The dedicated FRs are shown in Tab. 5.12, while the FR distributions are shown in Fig. 5.9. Two sources of systematic uncertainties are assessed:

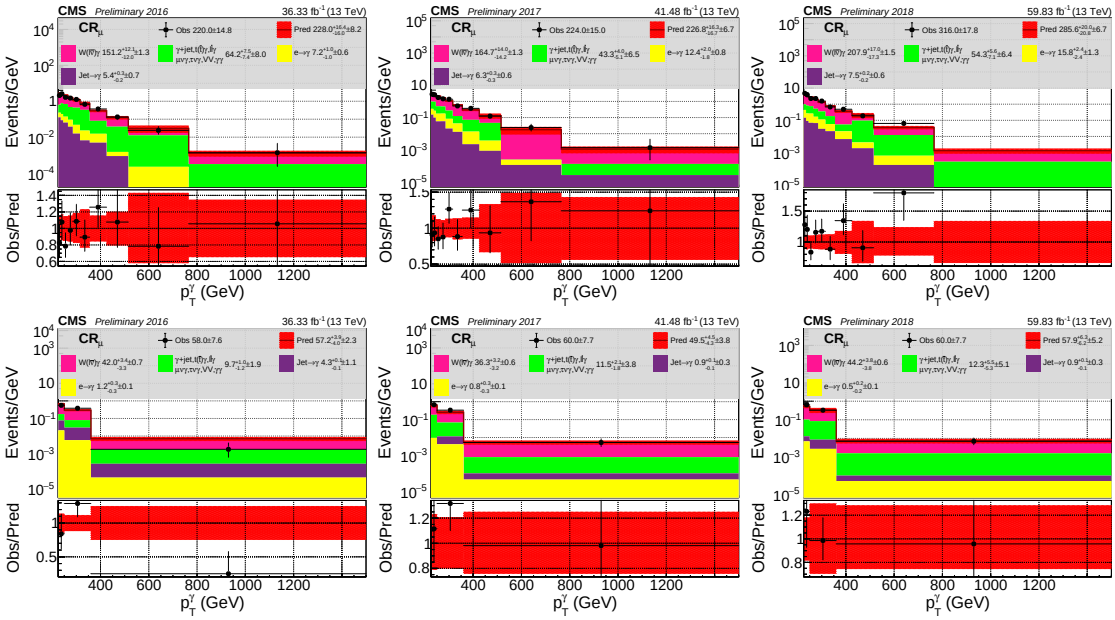

Figure 5.8: Distribution of the photon p_T in the W($\to \mu\nu$)γ control region for 2016 (left), 2017 (center) and 2018 (right) for events in EB (top) and EE (bottom).

- Choice of mass window: evaluated using a mass window of [80-110] GeV instead of [60-120] GeV for the fit;
- Choice of the fit function: a second round of fit is performed using a Breit-Wigner function convoluted with a Gaussian to model the signal (keeping the same background function).

In Fig. 5.9, the violet band shows the systematics and the green error bars are the fit errors. This trend is then fitted with a 2$^{\text{nd}}$ order polynomial. In order to get the error band to this fitted trend, the covariance matrix of the fit is used and a multi-variate Gaussian random numbers (10000 times) are thrown which give 3 numbers corresponding to the fitted parameters. For each random throw, the corresponding fake ratio is estimated for each bin in $|\eta|$, and the 68% interval of the fake ratio distribution forms the low and high error band. The above method is used also for the estimation of the background from

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

Table 5.12: Dedicated electron to photon fake ratio in the pixel inefficient regions.

misidentified electrons in the EE region. However, due to the lack of statistics at high $|\eta|$, the fake ratio is evaluated in one $|\eta|$-bin only, obtaining the following results for the three years:

- UL16 FR = 0.0165$\pm$0.0046 (stat.);
- UL17 FR = 0.0103$\pm$0.0041 (stat.);

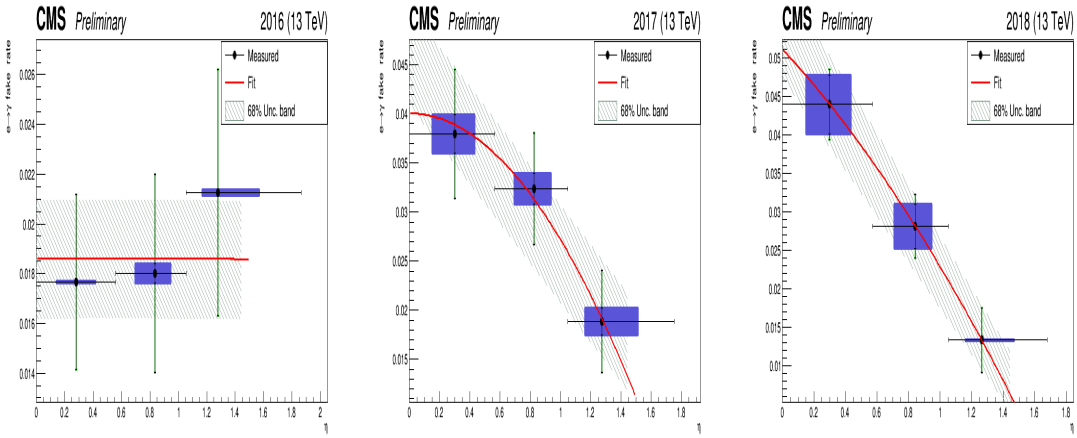

Figure 5.9: Electron to photon fake ratio as a function of the absolute value of pseudo-rapidity η for 2016 (left), 2017 (center) and 2018 (right). Green error bars are statistical error and blue rectangles are the systematic errors. The error band corresponds to the 68% coverage of the fake ratio distribution as a result of the uncertainty on the fitted parameters of a 2$^{\text{nd}}$ order polynomial.

- UL18 FR $= 0.0058 \pm 0.0026$ (stat.).

5.4.3 Jet-photon misidentification background contribution

Photon pairs coming from the decays of isolated and highly energetic π^0/η mesons produced inside a jet can be misidentified as photons, and can contribute as a background when large E_T^{miss} is simultaneously reconstructed in the event. The major sources of these jets-faking-photons objects are QCD multijets processes where E_T^{miss} is mismeasured due to limited jet energy resolution, and $Z(\to \nu\bar{\nu}) + \text{jets}$ production. This background contribution cannot be reliably simulated due to the difficulty in producing a sufficient yield of γ-like jets, and to the large systematic uncertainties associated with parton showering and detector response. A data-driven approach is therefore used, called the *ABCD method*.

ABCD Method

The region of interest C, corresponding to the signal region where the jet$\to \gamma$ background has to be estimated, is characterized by the presence of one photon identified by the ID criteria (selections on BDT *score*, H/E, I_{ECAL} and I_{Tkr}) described in Sec. 5.3 plus all the other phase space selection. Then, a complementary region A is defined by requiring:

- One photon with identical ID requirements with respect to the region A, except for the one on the BDT *score*;
- $E_T^{miss} < 40$ GeV.

By construction, the region A consists primarily of $\gamma + \text{jets}$ and QCD multijets. The number of jet faking photon events in A that satisfy the full ID (including the BDT requirement) can be extracted by fitting the BDT spectrum in A with shape templates for prompt photons and fake photons. The procedure is described in the next section. In addition, a region B is defined, with one photon passing some special ID requirements (listed in Tab. 5.13) tuned to enrich the contribution from jets faking photons and to have negligible prompt photon contamination. The ID used for photons in the B region was determined using the cut optimization method described in Sec. 5.3, but targeting 95% prompt photon efficiency and then inverting the resulting ID selections. The upper

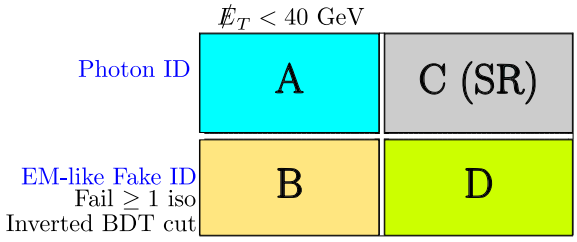

Figure 5.10: The four disjoint regions used to estimate the jet→ γ fake ratios.

boundary on E_T^{miss} is retained as in A. Analogously, a fourth region D is defined with selections identical to the region of interest C except for the photon ID criteria, which are kept the same as in B. It must then hold that:

$$\frac{\text{Number of fakes passing full ID in } A}{\text{Number of events in } B} = \frac{\text{Number of fakes passing full ID in } C}{\text{Number of events in } D}.$$

The fraction on the left-hand side is defined as the fake ratio (FR) which allows to estimate the jet fake component (N_C^{fake}) in C that passes the full ID from the yield (N_D) of D:

$$N_C^{fake} = FR \times N_D.$$

			2016	**2017**	**2018**
EB	Fail at least 1	I_{ECAL}/GeV <	2.04	2.94	2.65
		I_{Tkr}/GeV <	3.12	2.18	2.18
	Pass all	*BDT score* <	0.7292	0.6671	0.6533
		H/E <	0.0481	0.0215	0.0204
		I_{ECAL}/GeV <	4×	4×	4×
		I_{Tkr}/GeV <	4×	4×	4×
EE	Fail at least 1	I_{ECAL}/GeV <	5.23	3.10	3.06
		I_{Tkr}/GeV <	4.95	2.25	2.30
	Pass all	*BDT score* <	0.2702	0.4526	0.4599
		H/E <	0.0806	0.0240	0.0218
		I_{ECAL}/GeV <	2×	4×	4×
		I_{Tkr}/GeV <	2.5×	4×	4×

Table 5.13: ID definitions for regions B and D. The threshold for the "Pass all" categories is defined as n times ($n\times$) the threshold for the corresponding "Fail at least 1" categories.

Shape template fitting

The BDT spectrum of A obtained from data has contributions from prompt photons and jet faking photon events that can be extracted from a template fit, through the following parameterization:

$$H(x) = N_{prompt} \times h_{prompt}(x) + N_{fake} \times h_f(x),$$

where $H(x)$ is the observed spectrum in region A, N_{prompt} and N_{fake} are the number of prompt and fake events, respectively, $h_{prompt}(x)$ is the normalized template extracted from A using γ+ jets events in MC, and $h_f(x)$ is the normalized template extracted from the data sideband of A.

- Prompt γ template : Showering and reconstruction of prompt photons are well simulated in MC samples. Their template shape is obtained from γ+jets MC samples by applying the selections relevant to A. Residual data/MC differences in the shape are treated as a systematic uncertainty.

- Jet faking photon template : Shower simulation of photon-like jets has large systematic uncertainties and does not match well with observed data. The template is obtained directly from a data sideband A_S which has identical event requirements as in A, except for that on I_{Tkr} that has been modified to enrich the sideband with jet faking photons events containing significant tracker energy in their isolation cone. The sideband A_S is defined in Tab. 5.14. A systematic uncertainty on the fake template is assigned for the choice of sideband. In addition, the template is refined by subtracting the contamination contribution coming from prompt photons, estimated using $\gamma+$ jets MC (see example in Fig. 5.11).

	A Fitting Region	A_S Sideband for Fake Template
2016	$I_{Tkr} < 2.19$ GeV	4 GeV $< I_{Tkr} < 8$ GeV
	$H/E < 0.0222$ & $I_{ECAL} < 2.16$ GeV	
2017	$I_{Tkr} < 1.63$ GeV	4 GeV $< I_{Tkr} < 8$ GeV
	$H/E < 0.0401$ & $I_{ECAL} < 1.84$ GeV	
2018	$I_{Tkr} < 1.58$ GeV	4 GeV $< I_{Tkr} < 8$ GeV
	$H/E < 0.0402$ & $I_{ECAL} < 1.84$ GeV	

Table 5.14: Definitions of templates for jet fake ratio estimation in EB. A and A_S differ only in the cut on I_{Tkr}.

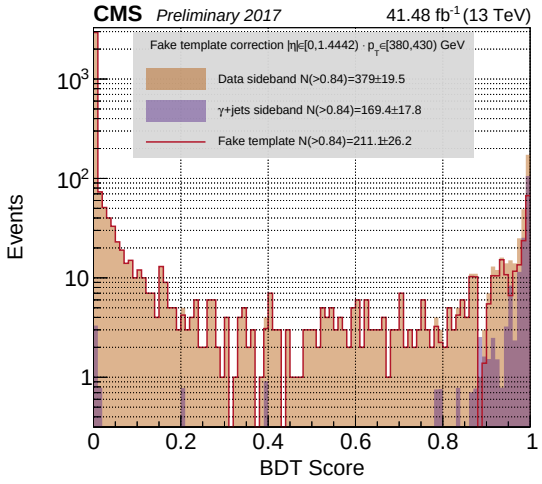

Figure 5.11: The spectrum obtained from the fake sideband A_S (orange) is corrected by subtracting the prompt photon contamination estimated by $\gamma+$ jets MC (violet).

The template fit is performed using a maximum likelihood fit in order to extract the number of prompt photons N_{prompt} and the number of fakes N_{fake} in the observed spectrum of A (see Fig. 5.12 left). The integral of the fake template above the BDT threshold gives the number of fake photons passing the full ID. This is divided by the yield of B to obtain the fake ratio (see Fig. 5.12 right). The fake ratio is evaluated in bins of photon p_T, in order to account for the reduction in faking probability with p_T and to improve the per-bin background estimation in the signal region. The distribution of the fake ratio as a function of p_T is then fitted with a 4-parameter function:

$$f(x) = ax + b + \frac{c}{x - d}.$$

The results are shown in Fig. 5.13 and 5.14. For 2016 EE, a linear fit function is instead used.

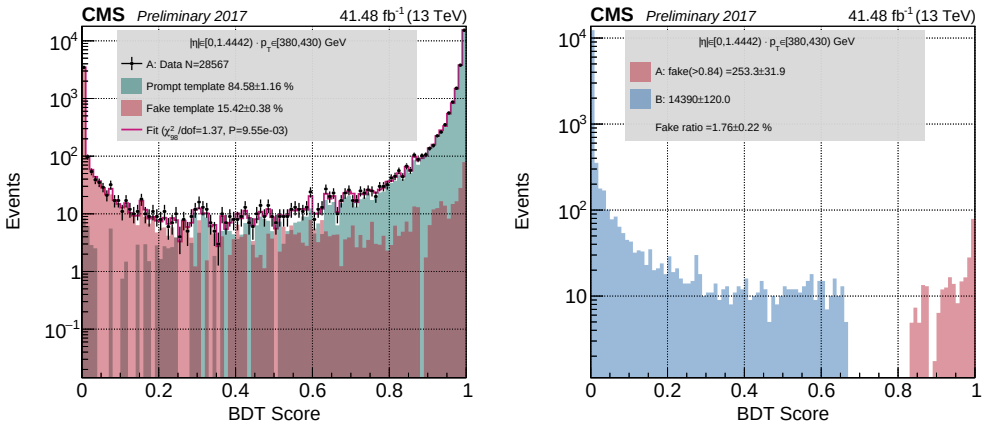

Figure 5.12: (Left) Template to the photon BDT distribution in data for 2017. The prompt template is shown in green, while the fake template in shown in red. (Right) Fake ratio extraction for 2017 EB. The number of events in the region A (red) are estimated selecting those that have the same $BDTscore$ requirement as defined for the photon ID ($BDTscore > 0.8361$). The $BDTscore$ selection for the events falling in region B (blue) is instead defined in Tab. 5.13 ($BDTscore < 0.6671$).

Fake ratio uncertainties

The following uncertainties in the estimation of the jet $\rightarrow \gamma$ background are considered:

- Fake template sideband choice: estimated by simultaneously raising and lowering the tracker isolation boundaries by 1 GeV;
- Photon shower mismodeling: the prompt template and the prompt removal in the fake template is corrected using quantile morphing corrections obtained in Sec. 5.3;
- Binning: estimated by varying the number of bins (200, 100, 50) used in the templates;
- E_T^{miss} cut choice: raise and lower the cut on E_T^{miss} by 5 GeV;
- MC statistics in fake sideband: estimated by raising and lowering the integral of the prompt component subtracted from the fake template by the MC statistical uncertainty.

The uncertainties are treated as uncorrelated and added in quadrature for each p_T bin.

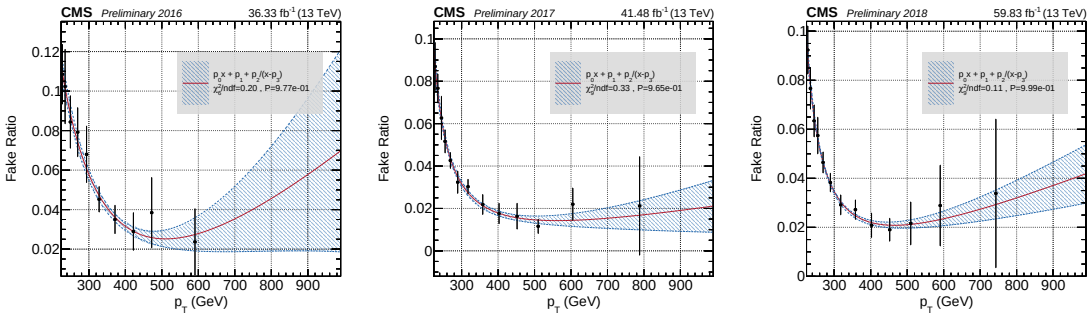

Figure 5.13: Distribution of the jet$\rightarrow \gamma$ fake ratios as a function of the photon p_T in EB for 2016 (left), 2017 (center) and 2018 (right).

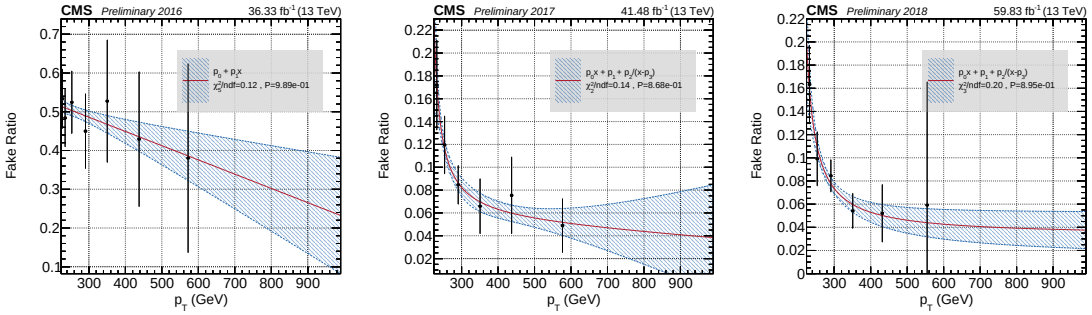

Figure 5.14: Distribution of the jet$\rightarrow \gamma$ fake ratios as a function of the photon p_T in EE for 2016 (left), 2017 (center) and 2018 (right).

5.4.4 Non collision background events in the ECAL barrel

Non-collision backgrounds are defined as those backgrounds that do not come directly from processes originated in the proton-proton collisions. In the ECAL barrel, these contributions come from the so-called *spikes* and beam halo, whose energy deposits in ECAL could be consistent with the signature of a photon. Spikes are events that occur when particles directly ionize within the active silicon layers of the avalanche photodiodes (APDs), giving rise to small clusters associated to high-energy deposits (see Fig. 5.15) with a rate that is proportional to the collision rate of the proton beams [194]. A first rejection of spike events is performed online with a dedicated trigger, being otherwise the spikes the dominant component of the CMS Level-1 trigger rate bandwidth. On the other hand, beam halo is produced when the protons interact with residual gas molecules in the vacuum chamber or with the wall of the beam pipe. These interactions give rise to a flow of secondary muons that travel along the beam line and that can lead to broad electromagnetic showers in ECAL: a Beam Halo Monitor (BHM) system is already implemented in CMS, with the purpose of detecting and correctly identifying beam-halo particles. A dedicated offline study is however still needed, to estimate the remaining contribution of both these background sources to the signal region. To do so, a data-driven approach is used, starting from the distribution of the photons candidates' time in data. The photon candidates are defined as in the signal region, except for the one on W_η (as defined below). All the other event selections described in Sec. 5.2 hold as well. The time distribution is then fitted using three templates: one for spike, one for prompt photon and one for beam halo. The spike template is built applying a selection on the variable W_η, called "eta-wing" and defined as:

$$W_\eta = \frac{E_0}{E_1},$$

where E_0 is the energy of the seed crystal and E_1 is defined as $\max\left\{E_{0-1_\eta}, E_{0+1_\eta}\right\}$, being E_{0-1_η} and E_{0+1_η} the two crystals adjacent to the seed (in the η direction). The seeds originated by spikes are expected to have smaller energy deposits in the neighboring crystals when compared to the seeds originated by electrons or photons' showers. This new variable can be hence fruitfully exploited to extract the spike template, separating spike events (that have small W_η) from other contributions (that have large W_η). In order to construct the beam halo template, the events are required to fail the central beam halo filter and to have $MIP_{tot} > 4.9$ GeV. The prompt template is then built using $\mathrm{Z} \rightarrow \mathrm{e}^-\mathrm{e}^+$ events, by requiring a first photon in the event that passes all the analysis selections, and a second photon with $p_T > 10$ GeV, pixel seed, and same H/E and isolation selections of the first one. The ΔR separation between the two photons must be greater than 0.2, and the reconstructed $\mathrm{Z}(\rightarrow$ "γe") mass must lie in the [60, 120] GeV window. To perform

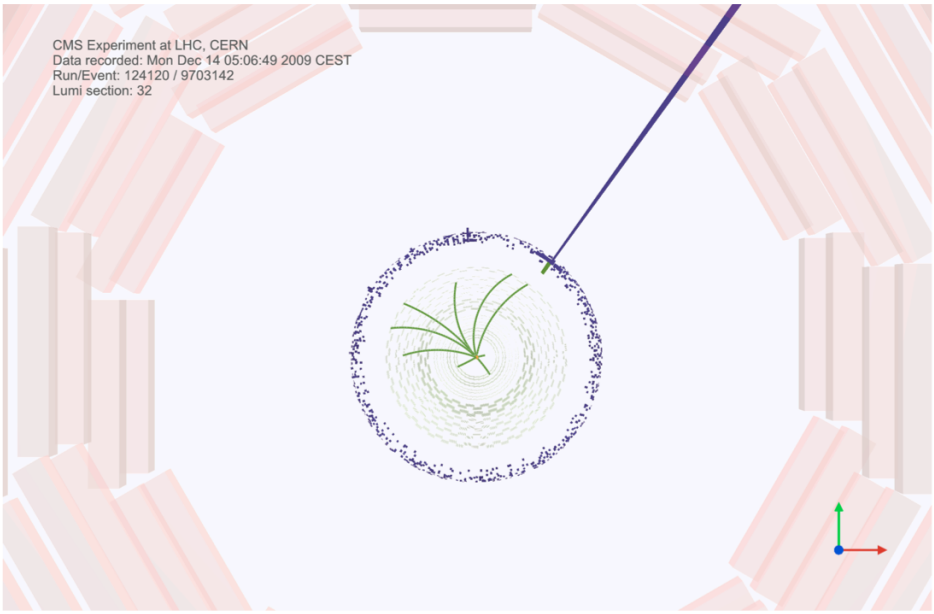

Figure 5.15: CMS event display of a pp collision showing an isolated ECAL spike corresponding to a 690 GeV transverse energy deposit [194].

the fit, the datasets with the information for in-time photons (i.e. photons in the prompt window, with time equal to *seedtime* ± 3 ns) and for the out-of-time photons (photons not in the prompt window) have been merged, as it was observed (Fig. 5.16) that the timing structure of the spikes was distributed mostly outside of the prompt time window. This is due to the fact that the spike pulses have a shorter rise time and faster decay compared to the response of scintillation light in ECAL. This shorter rise time is interpreted in the reconstruction as an earlier signal, originating negative timing. The templates are fit to the candidate events distribution, thus obtaining the number of spike events in the prompt window. From Fig. 5.17, one can see that the beam halo events give a very small contribution to the in-time candidate events. This is confirmed by fitting again with only spike and prompt templates, and verifying that the event yields with and without the beam halo template do not change. The number of spikes estimated in the prompt window is shown in Tab. 5.15 for the three years. The results are then cross-checked

2016	2017	2018
9.4 ± 1.57	10.3 ± 1.53	20.8 ± 2.2

Table 5.15: Number of spike events estimated from the template fit on the timing distribution.

using the matrix method. This ensures that no potential spikes or beam halo contribution are hidden via scaling one or the other templates from the timing fit. The method works by using two uncorrelated variables, the timing information for the photon and W_η, and exploiting the difference of efficiencies under selections on these variables between the two components (spikes and prompt) of the candidate events. The candidate events are selected in the prompt window (*seedtime* ± 3 ns) and the spike events are selected in the early time (*seedtime* < -12.5 ns). The in-time photon candidate can be then split into two other disjoint regions, with $W_\eta > 0.01$ and $W_\eta < 0.01$, respectively. Splitting the total number of candidate (N_{can}) into the collision and beam halo component (N_Z) and

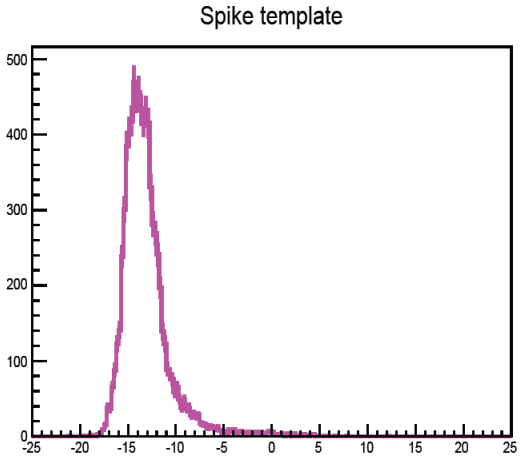

Figure 5.16: Timing distribution of the spike template as observed in [195]. On the x-axis the time associated to spikes is shown, considering as t=0 the *seedtime* of the supercluster. As it can be seen, spike events are mostly distributed at negative time, outside the prompt ($seedtime \pm 3$ ns) window.

the spike component (N_s), one obtains the following set of equations:

$$\begin{cases} N_{can} = N_s + N_Z \\ N_{can}^{W_\eta} = N_s^{W_\eta} + N_Z^{W_\eta} = \epsilon_s N_s + \epsilon_Z N_Z. \end{cases} \tag{5.6}$$

The label W_η identifies events in the region with $W_\eta > 0.01$. The efficiency ϵ_s and ϵ_Z are measured in a spike control sample and in a $Z \rightarrow ee$ sample, respectively. The control samples have been defined with the same requirements used for making the prompt and the spike template, and the efficiency is measured using a cut-and-count method based on the number of events passing $W_\eta < 0.01$ ($W_\eta > 0.01$) in the spike ($Z \rightarrow e^- e^+$) control region. Therefore, the number of spikes can be evaluated by solving the above system for N_s. The results are shown in Tab. 5.16. The spike contributions for all the three years are in good agreement within the uncertainties with those extracted from the template fit.

2016	2017	2018
8.6 ± 7.2	9.8 ± 7.9	18.5 ± 11.3

Table 5.16: Number of spike events estimated using the matrix method.

5.4.5 Non collision background events in the ECAL endcaps

While the beam halo provides a negligible contribution to analysis characterized by a single-photon signature in the ECAL barrel, it turns out to be the most important source of background events when looking for the same signature in the ECAL endcaps. The beam halo filter provided by CMS is indeed not optimised for EE as it relies on signals detected in the CSC chambers, which do not cover large regions of the EE ($|\eta| > 1.8$). Therefore, for the present analysis, a dedicated beam halo tagger has been developed, based on BDT that uses information from preshower (ES), ECAL endcaps and the HCAL endcaps (HE). This tagger exploits the fact that beam halo photons and prompt photons depositing energy in ECAL do not share the same travel path with respect to the beam axis. For a beam halo photon, the line joining the energy deposits in EE and in the HE/ES tends to be parallel to the z-axis, since beam halo mostly travels parallel to the beam lime. On the other

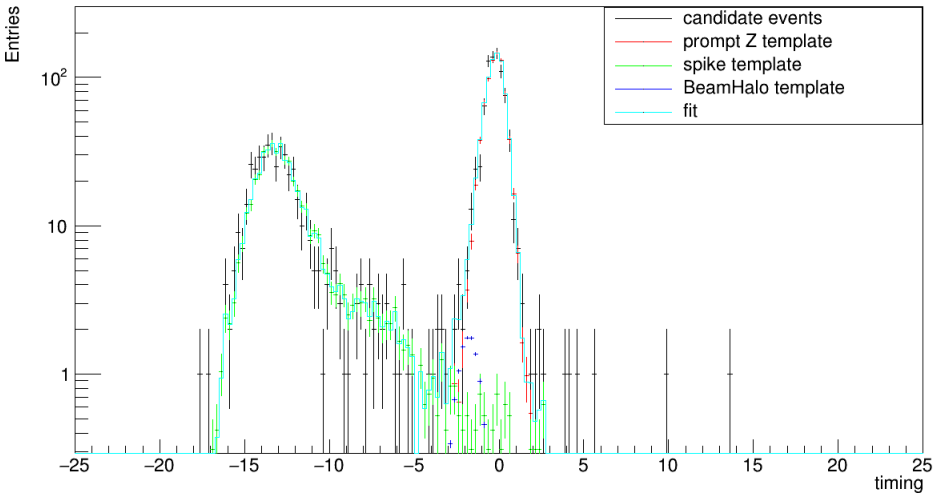

Figure 5.17: Timing fit to candidate events using prompt, spike and beam halo templates for 2017.

hand, for prompt photon coming from the primary vertex, the angle between the beam axis and the line joining the energy deposits in the EE and HE/ES is non-zero. Fig. 5.18 shows the two photon hypothesis between EE and HE, with the candidate photon being a prompt photon (PP) or beam halo (BH). For each hypothesis, the energies deposited in ES/HE are evaluated as follows, starting from the energy-weighted position of the ECAL supercluster (SC):

- BH hypothesis: the energy in HE is evaluated collecting all the HE hits in a cone of $\Delta\phi < 0.15$ along the line joining the SC and the HE. In addition, the search for HE hits is restricted in a 3×3 HCAL towers space by requiring $\sqrt{(\Delta x)^2 + (\Delta y)^2} <$ 26 cm, where $\Delta x = x_{HE} - x_{SC}$ and $\Delta y = y_{HE} - y_{SC}$. A selection that takes into account the EE transverse coverage is also applied, requiring $31 < \sqrt{(x_{HE}^2 + y_{HE}^2)} <$ 172 cm. The energy in ES is instead evaluated looking for the best matched hit in ES for which $\sqrt{(\Delta x)^2 + (\Delta y)^2} < 2.2$ cm, where $\Delta x = x_{ES} - x_{SC}$ and $\Delta y = y_{ES} - y_{SC}$. A maximum of five strips (in the positive and in the negative x and y direction) around the main one with $\Delta x < 1$ cm and $\Delta y < 1$ cm is used, summing all the associated energy deposits.

- PP hypothesis: the energy in HE is evaluated collecting all the hits in the HE falling in a cone of $\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.15$. For ES, the z coordinate in the ES planes is extrapolated from the position of the SC. Then, as for the BH hypothesis, a maximum of five strips (in the positive and in the negative x and y direction) around the main one with $\Delta x < 1$ cm and $\Delta y < 1$ cm are used, summing all the associated energy deposits.

The angles between the beam axis and the line joining the energy deposits in the EE and HE/ES are also evaluated for the two hypothesis. Then, a BDT is trained with XGBoost, using a privately generated beam halo sample and the $\mathrm{Z}(\rightarrow \nu\bar{\nu})\gamma$ sample for the prompt photon. The beam halo sample has been produced with the BeamHaloGenerator [196] provided by CMS. The following variables as used as input to the BDT algorithm:

- Shower shape variables of the photons: $E_{3\times3}/E_{SC}^{raw}$, $\sigma_{i\eta i\eta}$, $\sigma_{i\phi i\phi}$, σ_η, σ_ϕ;

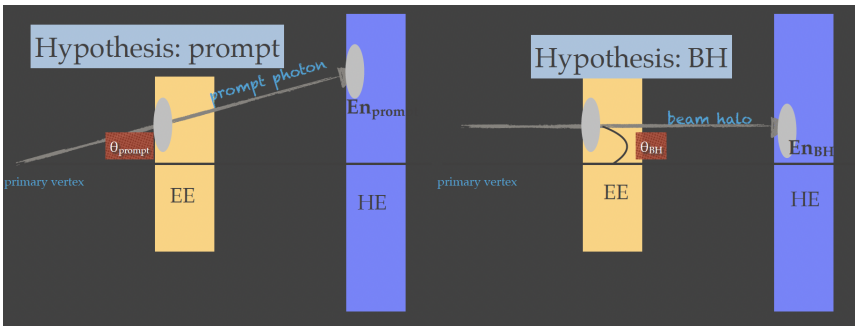

Figure 5.18: PP and BH hypothesis for a photon super-cluster reconstructed in EE.

- Energy deposits in ES and HE, as evaluated above;
- Energy weighted average angle in ES and HE, respectively, where the average is over the two hypothesis:

$$\theta_{avg}^{ES/HE} = \frac{E_{PP} \cdot \theta_{PP} + E_{BH} \cdot \theta_{BH}}{E_{PP} + E_{BH}},$$

  where E_{PP} and θ_{PP} are the energy and the angle in the PP (BH) hypothesis;
- Scalar sum of the transverse momentum p_T of all the tracks at the primary vertex. Since photons from beam halo do not arise from collisions, the scalar p_T-sum of all the tracks in the event is generally smaller.

The distributions of the *BDT score* for signal and beam halo events is shown in Fig. 5.19. The beam halo events are hence filtered requiring the *BDT score* to be greater than 0.99. In Fig. 5.20, the signal photon's ϕ distribution is shown in data.The variable ϕ is used

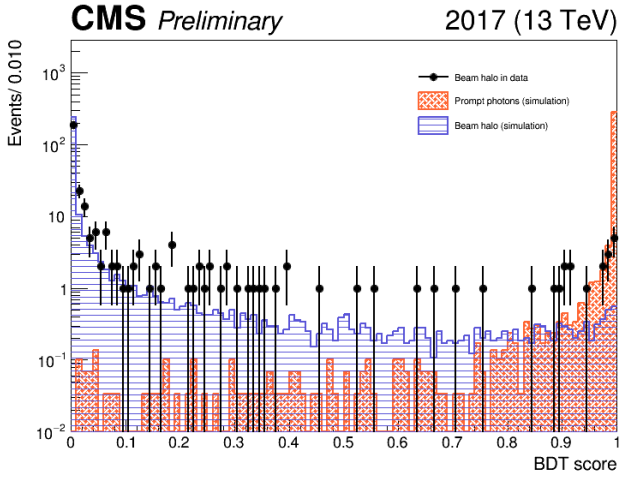

Figure 5.19: Distribution of the *BDT score* of simulated prompt photons originating from collisions (red) and beam halo (blue). The *BDT score* of beam halo photons in data is also shown, where the beam halo dataset is built using events that fail the central beam halo filter.

because one of the peculiar feature of beam halo events is to give higher contributions around $\phi = 0$ and $\phi = \pm\pi$, and hence this variable is an optimal handle to estimate the presence of beam halo in the analysis. The same distribution is shown before (using the already available tagger) and after the application of the new tagger. A template fit is performed to data in order to extract the number of signal and beam halo events, building

the signal template from the MC $(Z \rightarrow \nu\bar{\nu})\gamma$ sample, and the background template from data using a disjoint control region enriched in beam halo ($BDT\ score < 0.05$). The estimates of beam halo and signal efficiency in both cases are shown in Tab. 5.17 and, as it can be clearly seen, the newly developed BH tagger is highly effective in rejecting the beam halo from the analysis. Since the BDT tagger is built from simulated samples, in order to assess for any differences in the performances when it is applied to data, the corresponding scale factors are evaluated. The efficiency of the new tagger is estimated in the simulated $Z(\rightarrow \bar{\nu}\nu)\gamma$ sample and in data, selecting γ+jets events. The photon is required to have $p_T > 150$ GeV and to have Loose ID, while the jets are required to pass the Tight ID and to be have $\Delta\phi(\gamma, jet) > 2.9$. The efficiencies are then evaluated in data and Monte Carlo in bins of the photon's p_T, using a simple cut-and-count method, namely:

$$\epsilon = \frac{\text{photons passing the BH filter}}{\text{all selected events}}.$$

The efficiency and the scale factors are shown in Fig. 5.21. The estimated scale factors are all compatible with 1 within the statistical uncertainties and hence we take their values as 1, without assessing any additional uncertainty.

Table 5.17: The beam halo contributions and the signal efficiencies for the central BH tagger and the BDT tagger

Selections used	Signal efficiency / beam halo contribution		
	2016	2017	2018
Central tagger	99.6 / 5862.12±106	99.4 / 2079.41±62.4	99.4 / 3256.19±79.9
BDT based tagger	96.7 / 9.6±6.2	95.9 / 0±1.9	96.4 / 0±1.6

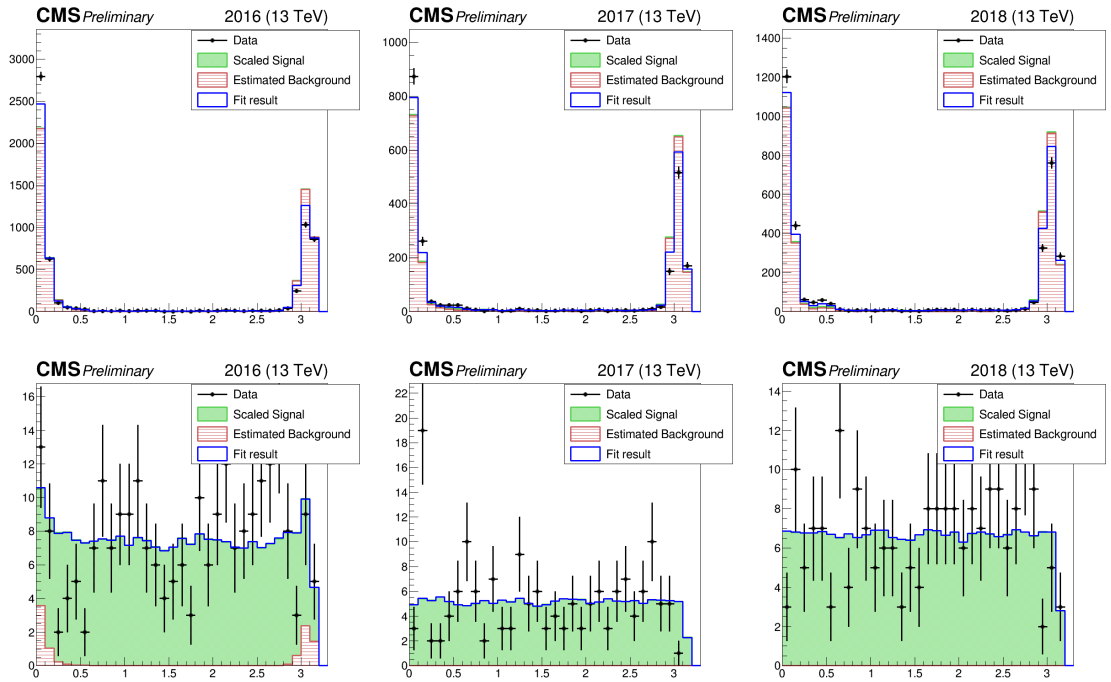

Figure 5.20: Template fits of the ϕ distribution of the signal photon in the EE region, for 2016 (left), 2017 (center) and 2018 (right). The distributions are obtained, respectively, using the CMS central tagger (top) and the new BDT BH tagger (bottom).

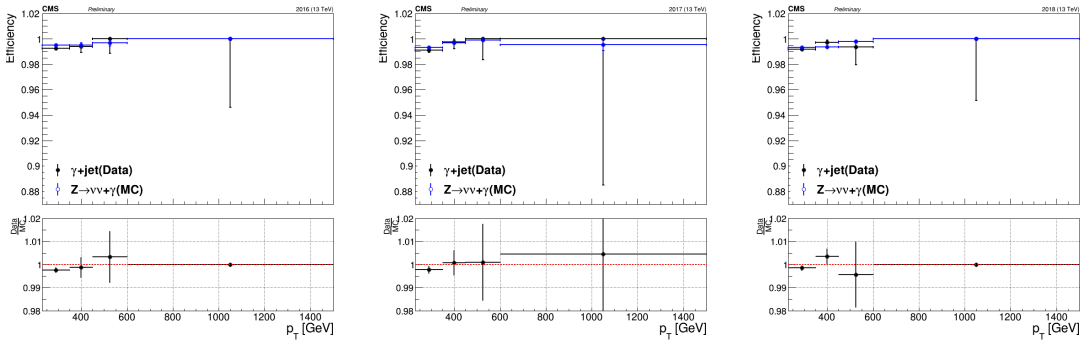

Figure 5.21: Efficiency comparison on γ +jets events in data and $\mathrm{Z}(\rightarrow \nu\bar{\nu})\gamma$ events in the simulation for 2016 (left) 2017 (center) and 2018 (right).

Chapter 6

Measurement of the $\mathrm{Z}(\rightarrow \nu\bar{\nu})$ cross section and limits on the anomalous neutral triple gauge couplings

In this chapter, the statistical procedure adopted to measure the inclusive $\mathrm{Z}(\rightarrow \nu\bar{\nu})\gamma$ production cross section, and to place the upper limits on the $\mathrm{ZZ}\gamma$ aNTGCs (h_3^{Z} and h_4^{Z}) is presented, together with the description of the systematic uncertainties affecting the measurement. The expected results are extracted from an *Asimov* dataset, by fitting a binned maximum likelihood to the photon p_T spectrum. The impact of the systematics on the cross section measurement is evaluated, and the expected limits on h_3^{Z} and h_4^{Z} are extracted. They are then compared to the expected limits obtained by CMS and ATLAS at a center of mass energy of 8 TeV and 13 TeV, respectively.

6.1 Statistical analysis

6.1.1 Cross section measurement

The cross section σ provides a measurement of the probability that an event (in the present case, the associated production of a photon and a Z boson decaying into a neutrino-antineutrino pair) occurs. While the theoretical cross sections are evaluated by computing the matrix element of the process using Feynman rules, and then integrating over its Lorentz-invariant phase space, the experimental cross sections, as measured at the colliders, are defined as follows:

$$\sigma = \frac{N - N_b}{A\epsilon\mathcal{L}}, \tag{6.1}$$

where N is the number of candidate events, N_b is the estimated number of background events and A, ϵ and $\mathcal{L}$ are the geometrical acceptance, detector efficiency and luminosity, respectively. When performing a study of collisions events, the probability to observe a number n of candidate events, given their expected number to be ν, can be expressed in terms of a Poissonian distribution:

$$f(n,\nu) = \frac{e^{-\nu}}{n!}\nu^n,$$

where the number of expected events includes both expected background and expected signal events ($\nu = \nu_{sig}+\nu_{bkg}$), and it depends on the signal strength μ and on the systematic

uncertainties $\vec{\theta}$ associated to the measurement. The signal strength is defined as:

$$\mu = \frac{\sigma_{obs}}{\sigma_{SM}}, \tag{6.2}$$

where σ_{obs} is the experimental observed cross section and σ_{SM} the SM prediction, that in this analyisis corresponds to the SM Zγ cross section. If no new physics appears, the signal strength is measured to be compatible with 1. In this work, the cross section is extracted using a binned likelihood function to fit the photon p_T spectrum. This is the observable most sensitive to the possible presence of new physics, considering that the presence of aNTGCs in Zγ processes would determine a higher yield of events with photon $p_T > 200$ GeV with respect to the SM predictions (as explained in Sec. 5.2). In particular, a *profile* likelihood method is used. Considering x_i variables with probability distribution function $f(x; \psi, \lambda)$, where ψ is the parameter of interest and λ a nuisance parameter, the log-likelihood can be written as:

$$\mathcal{L}(x \mid \psi, \lambda) = \prod_i f(x_i \mid \psi, \lambda). \tag{6.3}$$

Given that the direct simultaneous estimation of ψ and λ via likelihood maximization can be difficult to obtain, the nuisance parameter can be *profiled* out. This means that, for each ψ on a grid of chosen values, $\mathcal{L}_\psi(\lambda)$ is maximized in order to evaluate the estimator $\hat{\lambda}_\psi$ for the nuisance parameter. Then, the best parameter of interest is obtained by maximizing $\mathcal{L}(x \mid \psi, \hat{\lambda}_\psi)$. For the purpose of the present analysis, the profile likelihood is built taking into account the expected number of events in each bin of the photon p_T distribution as a function of μ and of the systematic uncertainties:

$$\mathcal{L}(data \mid \mu, \vec{\theta}) = \prod_i Poisson(d_i \mid s_i(\mu, \vec{\theta}) \ + b_i(\vec{\theta})) \cdot p(\tilde{\vec{\theta}} \mid \vec{\theta}), \tag{6.4}$$

where d_i, s_i and b_i are the number of observed, signal (SM $Z\gamma$) and background events in each p_T-bin i, respectively. Systematic uncertainties are incorporated via nuisance parameters (NP) labeled as $\vec{\theta}$, while $\tilde{\vec{\theta}}$ represents the measured values for each NP. The value of μ is estimated by maximizing $\mathcal{L}$ in data, and the measured cross section is hence obtained as $\mu \times s$. To test a hypothesized value of μ, the profile likelihood ratio can be used:

$$\Lambda(\mu) = \frac{\mathcal{L}(\mu = 0, \hat{\hat{\lambda}})}{\mathcal{L}(\hat{\mu}, \hat{\lambda})},$$

where $\hat{\hat{\lambda}}$ denotes the value of λ that maximizes the likelihood for the specified μ. Then, the following test statistic can be defined:

$$t_\mu = -2 \ln \Lambda(\mu). \tag{6.5}$$

The confidence interval around the best fit value is obtained by applying the Wilk's theorem, such that the 68% confidence interval corresponds to the region for which $t_\mu < 1$ [197].

6.1.2 Extraction of the limits on the aNTGCs

In order to evaluate the strength of the CP-even aNTGCs, the signal s_i in Eq. 6.4 is parameterized as a function of h_3^Z and h_4^Z, rewriting the likelihood as follows:

$$\mathcal{L}(data \mid \mu, h_3^Z, h_4^Z, \vec{\theta}) = \prod_i Poisson(d_i \mid s_i(\mu, h_3^Z, h_4^Z, \vec{\theta}) \ + b_i(\vec{\theta})) \cdot p(\tilde{\vec{\theta}} \mid \vec{\theta}), \tag{6.6}$$

With this formulation, the cross section can be simply obtained by setting h_3^Z and h_4^Z equal to zero, namely using the SM prediction for the signal yield s_i in each p_T-bin, and estimating μ. On the other hand, when estimating h_3^Z and h_4^Z, the value of the signal strength of the SM Zγ prediction is set to 1 and a likelihood fit is performed in the two dimensional h_3^Z-h_4^Z plane. The choice of the parameterization function for s_i is motivated by the corresponding modification that the anomalous couplings produce in the amplitude of the matrix element, therefore in the cross section and consequent number of events N. For a generic anomalous coupling c, associated to a new BSM operator in the EFT Lagrangian, N becomes:

$$N = N_{SM} + c \cdot Lin + c^2 \cdot Quad,$$

where N_{SM} is the SM event yield, Lin identifies the linear interference term between the SM and the new operator effect, and $Quad$ is the quadratic term that arises due the pure BSM contribution. When handling more operators at the same time, the previous formulation becomes:

$$N = N_{SM} + \sum_i \left(c_i \cdot Lin_i + c_i^2 \cdot Quad_i \right) + \sum_{i \neq j} Mixed_{ij},$$

with i being the i-th operator, and $Mixed_{ij}$ the interference term between the two operators. The Zγ event yield s is hence parameterized using the following function of h_3^Z and h_4^Z:

$$s\left(h_3^Z, h_4^Z\right) = \underbrace{s_0}_{\text{SM}} + \underbrace{2 \cdot \left[a_3^Z \cdot h_3^Z + a_4^V \cdot h_4^Z\right]}_{\text{SM+aNTGC interference}} + \underbrace{c^Z \cdot h_3^Z \cdot h_4^Z\Big] + \left[b_3^Z \cdot \left(h_3^Z\right)^2 + b_4^Z \cdot \left(h_4^Z\right)^2\right]}_{\text{Pure aNTGC}} \tag{6.7}$$

The signal is simulated at 440 non-zero values of the couplings for the ZZγ vertex, in the coupling ranges $-4 \times 10^{-4} < h_3^Z < 4 \times 10^{-4}$ and $-5 \times 10^{-7} < h_4^Z < 5 \times 10^{-7}$. The boundaries are chosen to be close to the latest ATLAS limits published in Ref. [98]. A least squares fit to the signal yield in each bin of the p_T^γ distribution is then performed, to extract the ultimate function parameters to be used in the signal parameterization when evaluating the limits on h_3^Z and h_4^Z. The confidence intervals for the aNTGC parameters are determined using a test statistic similar to that in Eq. 6.5:

$$t_{\vec{\alpha}} = -2 \ln \Lambda(\vec{\alpha}),$$

where $\vec{\alpha}$ are the anomalous coupling parameters, and $\Lambda(\vec{\alpha})$ is the likelihood ratio:

$$\Lambda(\vec{\alpha}) = \frac{\mathcal{L}(\vec{\alpha}, \hat{\hat{\lambda}})}{\mathcal{L}(\hat{\vec{\alpha}}, \hat{\lambda})}.$$

By applying the Wilk's theorem, the 68% and 95% confidence intervals for the values of the anomalous couplings can be determined [197]. The choice of the p_T^γ binning is optimized using the recursive Bayesian-Block algorithm, with the goal of separating the statistically significant features from the ever-present random observational errors [198]. The optimization process starts from an uniformly binned histogram with a bin width of 10 GeV, obtained with a simulated Z($\rightarrow \nu\bar{\nu}$)γ sample where the aNTGC h_3^Z and h_4^Z are set to be equal to 4×10^{-4} and 5×10^{-7}. The binning choice is made by recursively maximizing the total *Cash statistic* $\sum_k N_k \cdot \log(N_k/T_k)$, where N_k is the yield in the k-bin of width T_k [199]. The optimized bins are: [225-235], [235-245], [245-265], [265-285], [285-315], [315-355], [355-425], [425-515], [515-615], [615-765], [765,∞].

6.2 Systematic uncertainties

The measurement of the cross section and of the aNTGCs is affected by two different kind of uncertainties: a statistical uncertainty due to the limited amount of data and MC simulated events, and a series of different systematic uncertainties. Systematic uncertainties arise in different parts of the analysis, for example from the particular choices made to estimate background contributions, the theoretical uncertainties associated to the modelling of the process, or from the corrections applied to the physics objects. All the systematic uncertainties evaluated in this analysis are described in the following list.

Luminosity

CMS is able to measure luminosity both online and offline, providing the best luminosity measurements to be used by physics analysis. Online luminosity measurements are jointly performed by several subsystems: the Pixel Luminosity Telescope (PLT), the Hadronic Forward (HF) calorimeter, the Beam Conditions Monitor-Fast (BCM1F), and the Drift Tubes (DT). Offline luminosity measurements employ the Pixel Cluster Counting (PCC) algorithm: the recorded number of hit pixel clusters per bunch crossing is indeed expected to be a linear function of the number of interactions per crossing, and therefore a very good measure of luminosity [200]. In addition, luminosity is calibrated using the so-called "van der Meer scans" [201], that allow the determination of the absolute luminosity as a function of the beam parameters. The uncertainties in the integrated luminosity of the data collected by CMS during Run 2 are 1.2% (2016), 2.3% (2017) and 2.5% (2018). When combining the three years, correlations are taken into account by introducing five nuisance parameters in the likelihood in Eq. 6.4. Three parameters correspond to uncorrelated uncertainties for each year, the fourth handles the correlation across all years, and the last one handles the partial correlation between 2017 and 2018.

Pileup

In Monte Carlo samples, pileup events are added to the nominal hard scattering process using minimum bias events generated according to the value of the measured minimum bias cross section (σ_{MB} = 69.2 mb). The uncertainty on σ_{MB} is taken into account by reweighting the simulated events with ±4.6% variations in the cross section.

Scale factors uncertainties

Simulated events are corrected using a set of scale factors [142] (SFs) that take into account efficiency differences between data and simulation in photons, electrons and muons reconstruction and identification. The evaluation of the photon SFs and of their associated systematics is described in Sec. 5.3. The uncertainties on the SFs applied to electrons and muons in CRs include statistical and systematic uncertainties, the latter assessed for the choice of the signal and background fit functions when applying the Tag and Probe method (similarly to what is done in Sec. 5.3) for the histogram binning, fit window, and isolation of the tag muon candidate. Pixel seed SFs are also applied for dealing with the mismodeling in the simulations of the pixel seed efficiency that may affect the photon yield in SR and CRs.

Photon energy scale

In proton-proton collisions, the electroweak energy scale is measured with an uncertainty smaller than 0.1 (0.3)% in the EB (EE) [142]. The electron and photon energies are varied accordingly, and the uncertainty is then propagated as a nuisance parameter in Eq. 6.4.

Jet energy scale and resolution

Uncertainties on the jet energy correction and resolution are related to the different types of corrections described in Sec. 3.3.4, such as the pseudorapidity dependance of the de-

tector response, the momentum uniformity in the energy reconstruction, and the pile up subtraction. Such uncertainties affect the yields both in SR and in CRs, due to the phase space cut on $\Delta\phi_{min}(jet, E_T^{miss})$ (SR) and $\Delta\phi_{min}(\text{jet}, R_T)$ (CRs). The uncertainty is evaluated by shifting all reconstructed jet transverse momenta by ± 1 standard deviation of the measured jet energy scale. The same procedure is applied to evaluate the systematic uncertainty of the jet energy resolution. Furthermore, the jet corrections propagate to E_T^{miss} as described in Sec. 3.3.4, and they have to be taken into account when evaluating the uncertainties on E_T^{miss}. The jet energy variations and the associated E_T^{miss} variations are hence performed simultaneously.

L1 ECAL prefiring

In 2016 and 2017, the gradual timing shift of ECAL signals was not properly propagated to the L1 trigger primitives (TP), with the result that a significant fraction of TPs at high pseudorapidity was mistakenly associated to the previous bunch crossing (BX). The CMS trigger rules forbid the acceptance at L1 of two consecutive BXs, meaning that if an L1 TP is assigned to an earlier BX, and is accepted by the L1 trigger, the earlier BX is triggered and the BX that originally contained the TP is lost. This effect is known as *trigger prefiring*, and it is not described by the simulation. The probability for a prefiring to happen is therefore estimated in data and parameterized as a function of offline variables (photon and jet p_T and η), using *unprefirable* events. In an unprefirable event, the previous BX is blocked from being accepted by the L1 due to trigger rules, meaning that the event cannot be lost due to prefiring. An event weight ω is then obtained as the product of the non prefiring probability of all objects (measured using unprefirable events), namely:

$$\omega = 1 - P(prefiring) = \prod_{i=\gamma,\mu,jets} \left(1 - \epsilon_i^{pref.}(\eta, p_T)\right) \tag{6.8}$$

Prefiring weights are only applied to the MC events for 2016 and 2017. The associated uncertainty is obtained by shifting all prefiring probabilities plus or minus their uncertainties.

Jet $\rightarrow \gamma$ and e $\rightarrow \gamma$ fake ratio

The sources of uncertainty on the jet $\rightarrow \gamma$ and e $\rightarrow \gamma$ fake ratios are described in Sec. 5.4.3 and Sec. 5.4.2, respectively. The impact on the background yields in SR is assessed by varying each fake ratio within its uncertainty.

Theoretical uncertainties

Theoretical uncertainties affecting the cross section measurement come from the uncertainty on the set of PDFs used and the choice of the strong coupling constant α_S. Starting from the nominal PDFs set, 100 additional sets are computed with the Hessian method, embedding variations in the PDFs parameters. The associated uncertainty on the expected event yield per bin is then evaluated as:

$$\delta^{PDF} N = \sqrt{\sum_{k=1}^{100} \left(N^{(k)} - N^{(0)}\right)^2},$$

where $N^{(0)}$ is the event yield in the bin N obtained with the central PDFs set, while $N^{(k)}$ is the event yield obtained with the k-th set. Two additional PDFs sets are also computed, corresponding to the PDFs central values but with $\alpha_S = 0.116$ and $\alpha_S = 0.120$, respectively. The uncertainty due to the choice of α_S on the event yield per bin is therefore evaluated as:

$$\delta^{\alpha_S} N = \frac{N(\alpha_S = 0.116) - N(\alpha_S = 0.120)}{2}.$$

The PDF+α_S uncertainties are then combined adding them in quadrature.

Electroweak and QCD corrections

The signal Zγ and the background Wγ samples are simulated at leading QCD and electroweak order, meaning that corrections have to be applied to account for QCD next-to-next-to-leading order (NNLO) and electroweak next-to-leading order (EW NLO) effects. The EW NLO corrections are provided by Denner et. al. in [202] and [203] for $\sqrt{s} = 14$ TeV. For the purpose of this analysis, the authors have re-evaluated them (Fig. 6.1) at $\sqrt{s} = 13$ TeV, with kinematic selections and PDFs as they are used in this analysis. The corrections are computed as a function of the transverse momentum of the photon, and they have the following form:

$$k = 1 + \delta_{q\bar{q}} + \delta_{q\gamma},$$

where $\delta_{q\bar{q}}$ and $\delta_{q\gamma}$ refer to the quark-antiquark and photon-induced corrections, respectively. It can be observed that the correction becomes larger at increasing p_T^γ, due to the presence of Sudakov double logarithms: around 1 TeV, the correction for Zγ (W$^\pm\gamma$) is of the order of 40% (30%). The uncertainty of $\delta_{q\bar{q}}$ is $\delta^2_{q\bar{q}}$, while the considered uncertainty of $\delta_{q\gamma}$ is taken as $\delta_{q\gamma}$, due to the high uncertainty on the photon PDF at large values of the Bjorken variable x [202]. QCD NNLO corrections are evaluated with the pro-

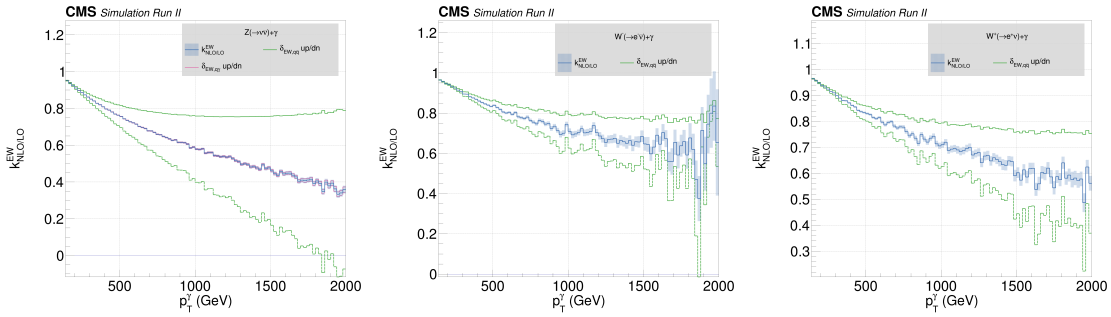

Figure 6.1: Electroweak NLO/LO k-factors as a function of the photon transverse momentum, as they are applied to Zγ (left), W$^-\gamma$ (center) and W$^+\gamma$ (right) simulated events.

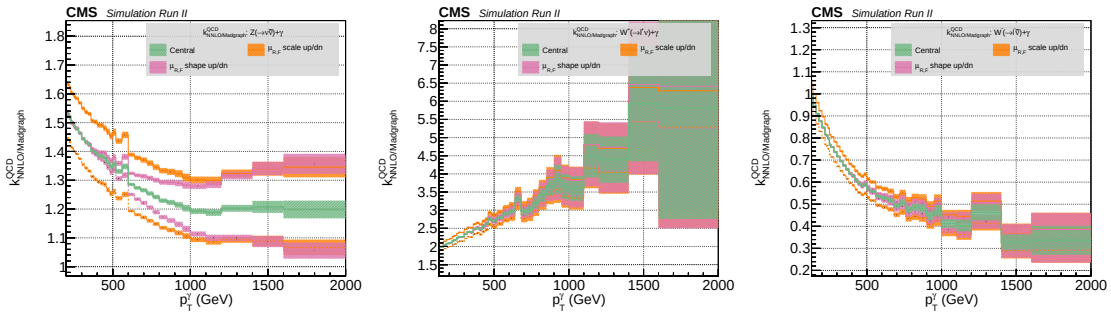

Figure 6.2: QCD NLO/LO k-factors as a function of the photon transverse momentum, as they are applied to Zγ (left), W$^-\gamma$ (center) and W$^+\gamma$ (right) simulated events.

gram MATRIX [204], and two sources of uncertainty are assessed on them. One is due to missing higher orders, and it is provided by MATRIX by means of factorization and renormalization scale (μ_R and μ_F) variations. The renormalization factor μ_R is an energy cut-off that has to be introduced in the QCD Lagrangian in order to deal with ultraviolet (UV) divergences that can arise in the evaluation of perturbation theory integrals. Another energy cut-off that has to be introduced is the factorization scale μ_F, representing the resolution with which the hadrons are probed in a collision. This means that emissions with transverse momenta above μ_F are included in the hard scattering cross section term, while emissions with transverse momenta below μ_F are included in the PDFs term. Like

the renormalization factor, the factorization scale is arbitrary, and the usual choice for both parameters is to take $\mu_R = \mu_F = Q$, where Q is the energy scale of the process. In these V + γ samples, $Q = \sqrt{(m_T^V)^2 + (m_T^\gamma)^2}$. The uncertainty due to the arbitrary choice of μ_R and μ_F mainly affects the overall normalization of the p_T distributions, and it is estimated by independently varying $\mu_{F/R}$ by factors of 0.5 and 2 and then taking the envelope of the 7-point variations [205]. Such uncertainty is taken into account also for all the other MC samples used to estimate the background contribution coming from prompt photons. The second source of uncertainty is instead assigned to account for possible variation in the shape of the distributions, and it is estimated from the scale uncertainty on the NNLO cross-section $\Delta\sigma^{\mu\text{ scale}}$ using [206]:

$$\Delta\sigma^{\mu\text{ shape}} = \frac{p_T^2 - p_{T,0}^2}{p_T^2 + p_{T,0}^2} \cdot \Delta\sigma^{\mu\text{ scale}}$$

where p_T^0 is chosen to be 850 GeV. The QCD NNLO corrections are shown in Fig. 6.2.

6.3 Results

The likelihood function in Eq. 6.6 can be reformulated to explicitely point out the terms referring to the Wγ control regions:

$$\mathcal{L}(data \mid \mu, \vec{\theta}) = \prod_i \left[\mathcal{P}(d_i | s_i(\mu) + \Sigma_k b_i^k) \cdot \prod_{K=e\gamma,\mu\gamma} \mathcal{P}\left(d_{K,i} | R_{K,i}^{W\gamma}(\theta) N_i^{W\gamma} + b_{K,i}(\theta)\right) \right] \cdot \prod_j \mathcal{N}(\theta_j), \tag{6.9}$$

where i identifies the p_T-bin index, the Poisson and Gaussian distribution are respectively represented with $\mathcal{P}$ and $\mathcal{N}$, and d_i and s_i are the number of observed event and the prediction for the SM $Z\gamma$ yield in bin i, respectively. The sum Σb_i^k, includes all the backgrounds k in the i-bin. The observed number of events in the i-bin of a control region K is $d_{K,i}$, while $N_i^{\mathrm{W}\gamma}$ is the $W\gamma$ yield sample in the i-bin of the SR. The transfer factors evaluated in Sec. 5.4.1 are labeled with $R_{K,i}^{W\gamma}$. Each systematic uncertainty j (as listed in Sec. 6.2) enters the Lagrangian as a nuisance parameter θ_j. The per-bin yield $N_i^{\mathrm{W}\gamma}$ associated to the Wγ background contribution is also estimated from the fit.

6.3.1 The blinded analysis

When performing a physics analysis it would be easy to fall into the temptation of exploiting the degrees of arbitrariness of the chosen methods (for example, the choice of a fitting function) or of the applied selections in order to achieve better results, i.e. results which are closer to what we would like to obtain. In order to avoid such bias and to ensure a objective and robust results, the data in the signal region are never used when defining the analysis structure, and the extraction of the cross section and the evaluation of the limits on the aNTGCs is first performed according to the so-called *blinding* procedure. This means that the likelihood fit is first performed on an *Asimov* dataset, which is toy data sample constructed by generating events under the SM hypothesis $\mu = 1$. A blind analysis has no power in measuring the cross sections and signal strength, but it is particularly useful to assess the impact of the related systematics and to deliver a first estimation of the tightness of the limits on the aNTGCs, provided that the real signal is not too different from the MC simulation. A *fiducial* phase space is then defined, choosing the selections

as close as possible to those applied for defining the signal region in data:

$$
\begin{aligned}
|\eta^{\gamma}| < 1.4442 \text{ (EB)},\ 1.566 < |\eta^{\gamma}| < 2.5 \text{ (EE)}, \\
p_T^{\gamma} > 225 \text{ GeV}, \\
E_T^{miss} > 200 \text{ GeV}, \\
\frac{p_T^{\gamma}}{E_T^{miss}} < 1.4, \\
\Delta\phi(\gamma,\, E_T^{miss}) > 2.
\end{aligned}
$$

The above set of fiducial selections is applied to the generated sample at the *truth* level. The finite resolution and limited acceptance of the detectors, however, could reflect in migration effects that would bring an event, passing the selections at the truth level, out of the fiducial region once it is fully reconstructed. Therefore, for the measurement of the fiducial cross sections, the signal expectation s_i in each p_T-bin is constructed as:

$$
s_i = \sum_j \left[\sigma_j \cdot \mathcal{L} \cdot R_{ij}(\vec{\theta}) + \sigma_j \cdot \frac{\sigma_j^{E_T^{miss},MC}}{\sigma_j^{fid,MC}} \cdot \mathcal{L} \cdot R_{ij}^{E_T^{miss}}(\vec{\theta}) \right] + s_i^{\text{OOA},MC}(\vec{\theta}) \tag{6.10}
$$

where σ_j is the measured cross-section for the p_T-bin j (for the purpose of this analysis only one bin is considered), $\mathcal{L}$ is the luminosity, and R_{ij} is the response matrix, estimated from simulation, that represents the probability that an event in the fiducial truth bin j will be observed in the reco bin i. $R_{ij}^{E_T^{miss}}$ is the response matrix obtained for *out of acceptance* (OOA) events that fail the fiducial E_T^{miss} requirement. In this measurement, in fact, the most significant OOA contribution ($> 99\%$) comes from events that pass all the fiducial selections with the sole exception of the requirement on E_T^{miss}. The corresponding OOA cross sections $\sigma_j^{E_T^{miss}}$, however, cannot be measured, and therefore they are related to σ_j using the ratio between $\sigma_j^{E_T^{miss},MC}$ (the OOA cross section estimated in MC for events failing the fiducial E_T^{miss} selection) and $\sigma_j^{fiducial,MC}$ (the fiducial cross section estimated in MC). The remaining OOA contributions $s_i^{OOA,MC}$ are treated as an additional background.

6.3.2 The expected cross section

Fig. 6.3 shows the likelihood scans of the fiducial cross section (with h_3^Z and h_4^Z set to zero in Eq. 6.6), where $-2\Delta \ln \mathrm{L}$ is equal to the test statistic shown in Eq. 6.5. The contribution of the different sources of systematic uncertainty is also highlighted, defining four categories. The global theoretical uncertainty (labeled as *theory*) refers to the contribution from the PDFs, α_S, μ_R, μ_F, and the EWK and QCD corrections. The *efficiency* term refers to the uncertainties associated to the scale factors, prefiring and pileup. The *calibration* term takes into account the uncertainty on luminosity, photon energy scale and jet and MET corrections, while all the remaining uncertainties are included in the *others* category. It can be observed that the theoretical uncertainties are those providing the smallest contribution to the estimate of the cross section, whose expected values is measured to be $6.956^{+0.164}_{-0.154}$ fb in EB, and $1.741^{+0.073}_{-0.060}$ fb in EE. The expected fiducial cross section, when combining the results from EB and EE, is:

$$
\sigma_{fid}(\text{fb}) = 8.689^{+0.004}_{-0.008}(\text{theory})^{+0.141}_{-0.121}(\text{calibration})^{+0.069}_{-0.077}(\text{efficiency})^{+0.103}_{-0.092}(\text{others})^{+0.038}_{-0.038}(\text{stat}).
$$

The *impacts* of the systematic uncertainties are evaluated in order to examine the effects that such uncertainties have on the cross section measurement. They are obtained by

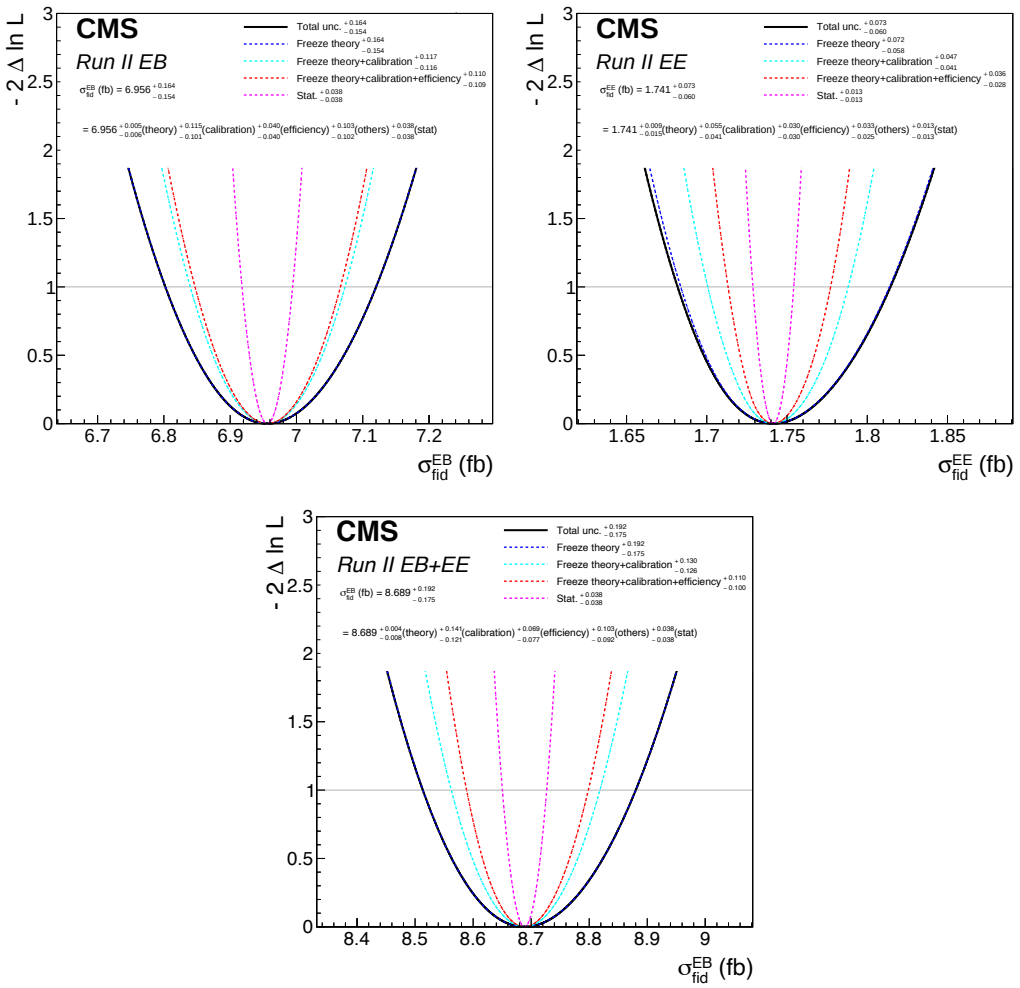

Figure 6.3: Likelihood scans of the fiducial cross section for EB (top left), EE (top right), and combining EB+EE (bottom). The solid black line takes into account all the uncertainties (statistical and systematic), while the dashed lines refer to likelihood scan in which one or more sources of uncertainties are not considered (*freezed*).

shifting each nuisance parameter by its $\pm 1\sigma$ post-fit uncertainty values, and then evaluate the related shift in the signal strength with respect to the best-fit value. The impacts provided by the 21 main sources of systematic uncertainties are quantified in Fig. 6.4. The variation on the cross section produced by a $\pm 1\sigma$ variation of the nominal value for each systematics is shown, together with the associated pulls, defined as:

$$\frac{\hat{\theta} - \theta_0}{\Delta\theta},$$

where $\hat{\theta}$, θ_0 and $\Delta\theta$ are the best-fit value, the input value and the input uncertainty for each source of systematic uncertainty, respectively. The uncertainty on the event yield in each bin of the Wγ control regions, which is also fitted when minimizing the likelihood, as explained in Sec. 6.3, is treated as an additional systematic. Given that the yields are constrained to be non negative, their impact on the cross section is not symmetric. The complete list of the impacts can be found in Appendix A. It can be observed that the greatest impact is provided by the JES uncertainty, which affects the evaluation of E_T^{miss}, followed by the uncertainty associated to the estimation of the e $\rightarrow \gamma$ fake ratio and to the photon ID scale factors.

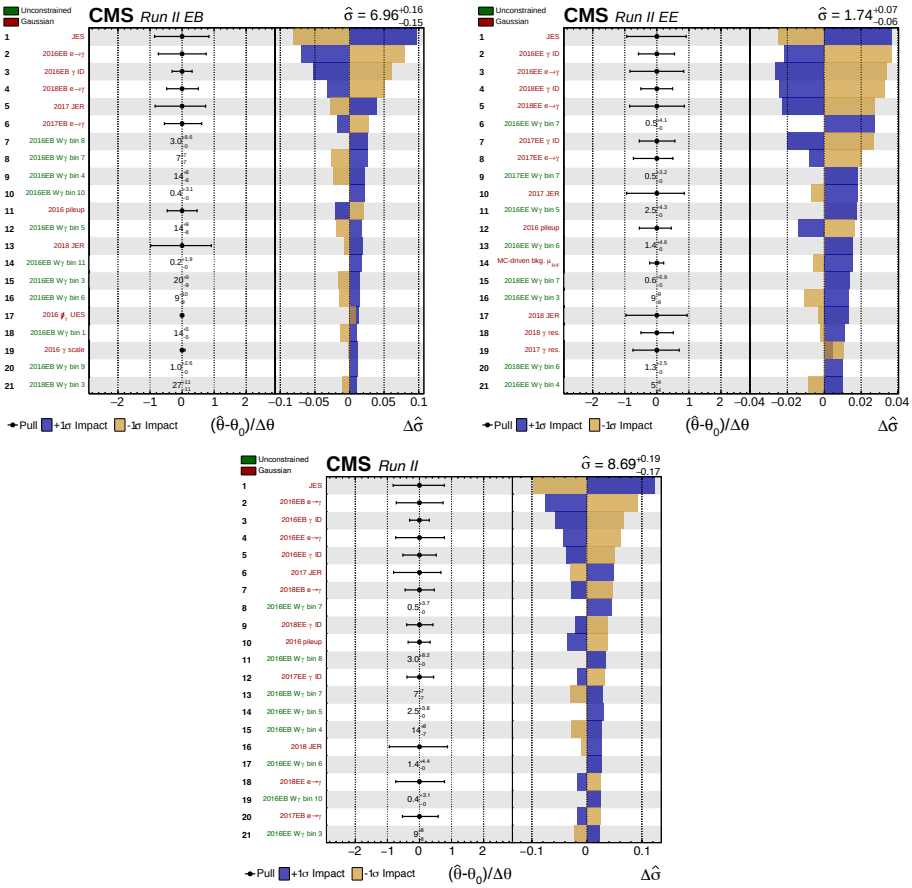

Figure 6.4: Impacts and pulls of the 21 main sources of systematics on the expected cross section measured in EB (top left), EE (top right) and EB+EE (bottom).

6.3.3 The expected limits on anomalous neutral triple gauge couplings

The expected limits on aNTGCs are estimated performing 2D likelihood scans of the anomalous coupling parameters (considering the signal parameterization described in Sec. 6.1.2) with the same Asimov dataset used for the evaluation of the expected cross section. The signal strength parameter is fixed at 1, while the anomalous couplings are taken as free parameters. The 2D likelihood scans in the (h_3^Z, h_4^Z) plane are shown in Fig. 6.5, for events in EB and EE, respectively, and for the EB+EE combination. The expected sensitivity to h_3^Z, under the SM hypothesis, is obtained by setting h_4^Z to zero, and viceversa (see Fig. 6.6 and Fig. 6.7). The expected limits at the 95% confidence level (CL) are shown in Tab. 6.1, where they are compared with the expected results obtained by CMS and ATLAS at 8 TeV and 13 TeV, respectively. The expected limits extracted from this analysis are more stringent by one order of magnitude with respect to the expected limits in the previous CMS analysis, and by a factor that varies between around 1.4 and 2 with respect to the expected limits in the 13 TeV ATLAS analysis. The inclusion of forward events leads to a 1-2% improvement with respect to the results obtained using events in EB only.

6.3.4 The observed cross section and the observed limits on anomalous neutral triple gauge couplings

After the expected cross section and the limits on the anomalous couplings were evaluated, as well as the expected impacts of the systematics on the cross section measurement,

the analysis has started a standard internal review process by the CMS Collaboration. This procedure aims to provide all the cross checks needed in order to get the best possible unbiased, unblinded observed results. This process is currently ongoing. The observed cross section and limits will therefore be available once the review process is finalized by the CMS collaboration.

	CMS 8 TeV	ATLAS 13 TeV	CMS 13 TeV
h_3^Z	$[-1.5, 1.6] \times 10^{-3}$	$[-3.8, 3.8] \times 10^{-4}$	$[-1.818, 1.841] \times 10^{-4}$
h_4^Z	$[-4, 4.5] \times 10^{-6}$	$[-5.3, 5.1] \times 10^{-7}$	$[-3.662, 3.625] \times 10^{-7}$

Table 6.1: Expected limits on the ZZγ aNTGCs h_3^Z and h_4^Z obtained by CMS at 13 TeV and comparisons with the previous expected limits.

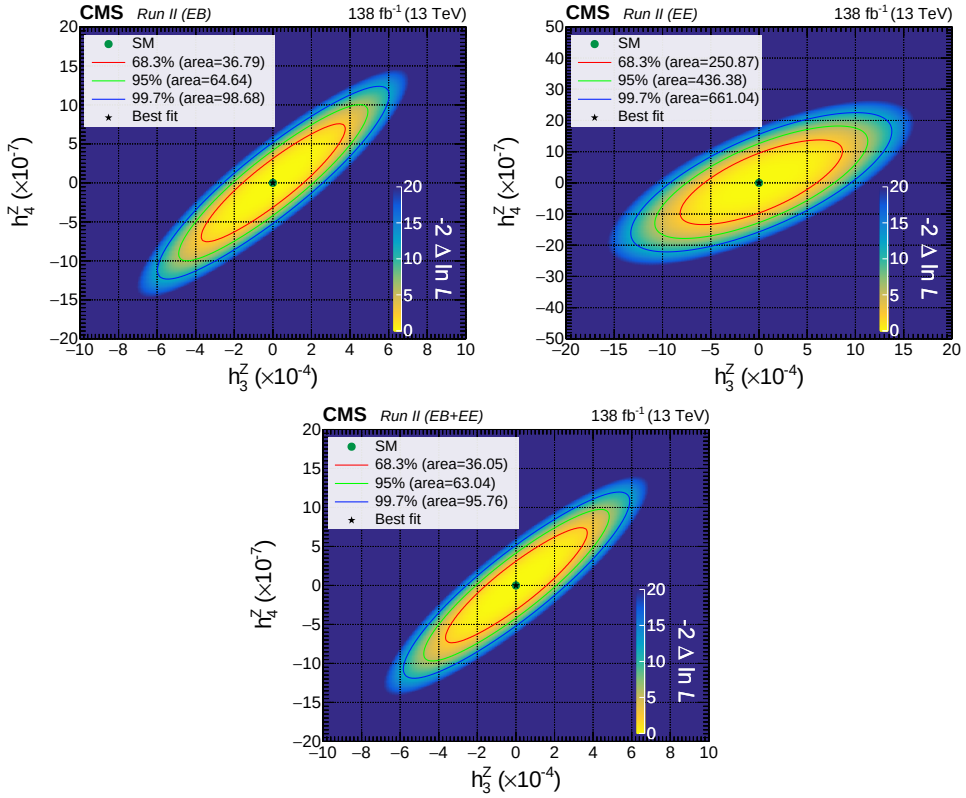

Figure 6.5: 2D likelihood contours in the (h_3^Z, h_4^Z) plane performed for signal events in EB (top left), EE (top right) and for the EB+EE combination (bottom).

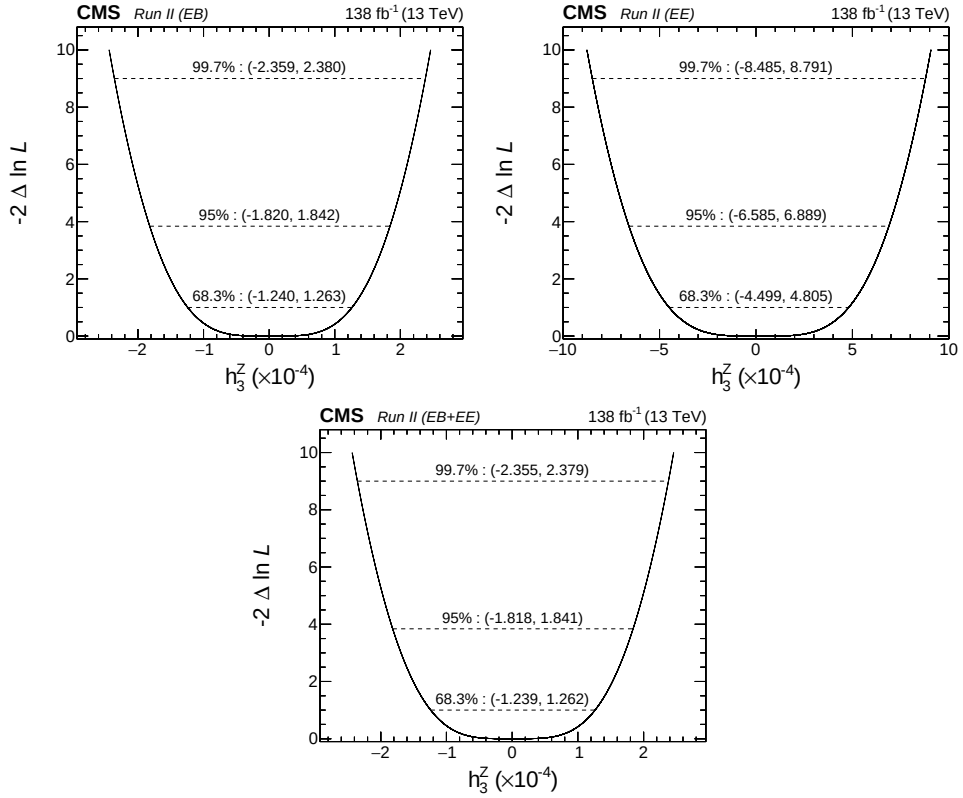

Figure 6.6: Likelihood scans for the estimation of the limits on h_3^Z in EB (top left), EE (top right) and for the EB+EE combination (bottom), with h_4^Z set to zero.

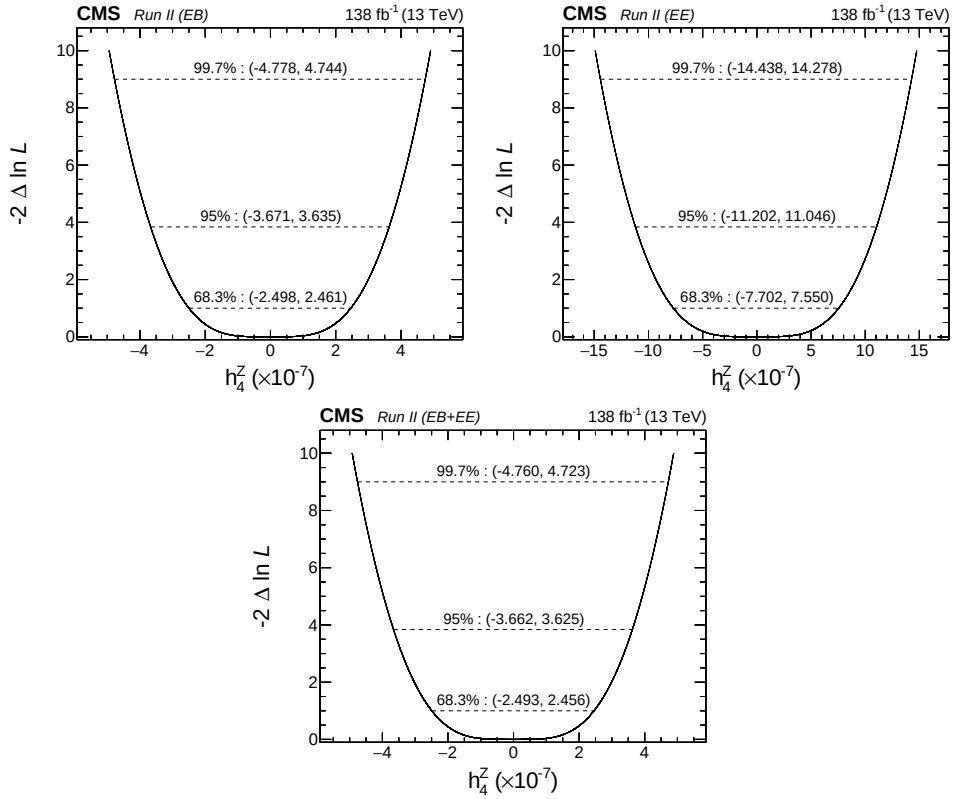

Figure 6.7: Likelihood scans for the estimation of the limits on h_4^Z in EB (top left), EE (top right) and for the EB+EE combination (bottom), with h_3^Z set to zero.

Conclusions

In this thesis work, the associated production of a Z boson decaying into a neutrino-antineutrino pair and a photon is studied. The analysis of the $Z(\rightarrow \nu\bar{\nu})\gamma$ process has crucial importance in testing the electroweak sector of the SM and in probing possible effects of new physics by placing limits on the presence of anomalous triple neutral gauge couplings. Moreover, when performing searches for dark matter candidates in proton-proton collision, the SM production of $Z(\rightarrow \nu\bar{\nu})\gamma$ events represents the most important background, and a precise measurement of its cross section is therefore needed in the view of achieving the best possible measurements when looking for dark matter signatures. The data collected by the CMS experiment during the full Run 2 of the LHC at a center of mass energy of 13 TeV are used, for a total luminosity of 137.6 fb^{-1}. In addition, this is the first time that photons detected in the forward part of the electromagnetic calorimeter are included in a $Z(\rightarrow \nu\bar{\nu})\gamma$ CMS analysis. Such improvement is possible thanks to a dedicated algorithm developed for the rejection of beam-halo events, the dominant contribution to the signal region when looking for a monophoton signature in the forward region. A dedicated set of criteria specifically built and validated for this analysis using a multivariate technique based on a boosted decision tree algorithm is implemented to select the signal photon candidate. A data driven approach is used to estimate the major background contribution which arises from electrons that do not leave a hit in the tracker, and that are hence reconstructed as photons. The associated production of a W boson and a photon provides the second highest background contribution when the W decays leptonically but the lepton is lost or poorly reconstructed. Such contribution is estimated using control regions built from data. The remaining, minor, sources of background are evaluated using data-driven techniques (such as for spikes and jets faking photons), or from simulation. The expected cross section is evaluated performing a maximum likelihood fit to the p_T distribution of the photon built from an Asimov dataset, giving $\sigma = 8.689^{+0.192}_{-0.175}$ fb. The uncertainties that mostly affect the evaluation of the cross section are those coming from the jet energy scale, the photon ID scale factors in 2016 and the estimation of the background due to electrons faking photons in 2016. The uncertainties on the scale factors and on the $e \rightarrow \gamma$ background suffer from the limited statistics collected in the 2016 data-taking, reflecting in a higher impact on the measurement. The expected limits on the anomalous triple neutral gauge couplings h_3^Z and h_4^Z are estimated using the same Asimov dataset, parameterizing the signal as a quadratic function of them. The expected limits at the 95% confidence level are $[-1.818, 1.841] \times 10^{-4}$ for h_3^Z and $[-3.662, 3.625] \times 10^{-7}$ for h_4^Z. The inclusion of forward events provides an additional 1-2% improvement in the limits' estimation. The expected results provide more stringent limits by one order of magnitude with respect to the expected limits in the previous CMS analysis, and by a factor that varies between around 1.4 and 2 with respect to the expected limits in the 13 TeV ATLAS analysis. Such improvements are achieved thanks both to the new multivariate techniques employed in the analysis and to the magnified luminosity delivered by the LHC during the Run 2. The observed cross section and limits will be available once the review of the analysis will be finalized by the CMS collaboration. The ongoing Run 3, and even more the High-Luminosity phase of the LHC (HL-LHC), are foreseen

to further increase the statistics and hence to provide the most stringent limits on the possible presence of anomalous couplings, as well as the most precise measurement of the $Z(\rightarrow \nu\bar{\nu})\gamma$ cross section. For the HL-LHC phase, the CMS experiment will undergo a significant series of major upgrades, refurbishing the majority of its subdetectors and installing new ones. With this perspective in mind, a part of this thesis work is also dedicated to the studies of the perfomances of the new timing detector for minimum ionizing particles. The final goal of these studies is to achieve the best time resolution for vertices and tracks, providing good pileup rejection and improved particle reconstruction even in such a high-luminosity environment, opening a new era for new physics searches as well as for precise measurements of fundamental SM processes, such as the aforementioned $Z(\rightarrow \nu\bar{\nu})\gamma$ production studied in this work.

Appendices

Appendix A

Systematics impacts

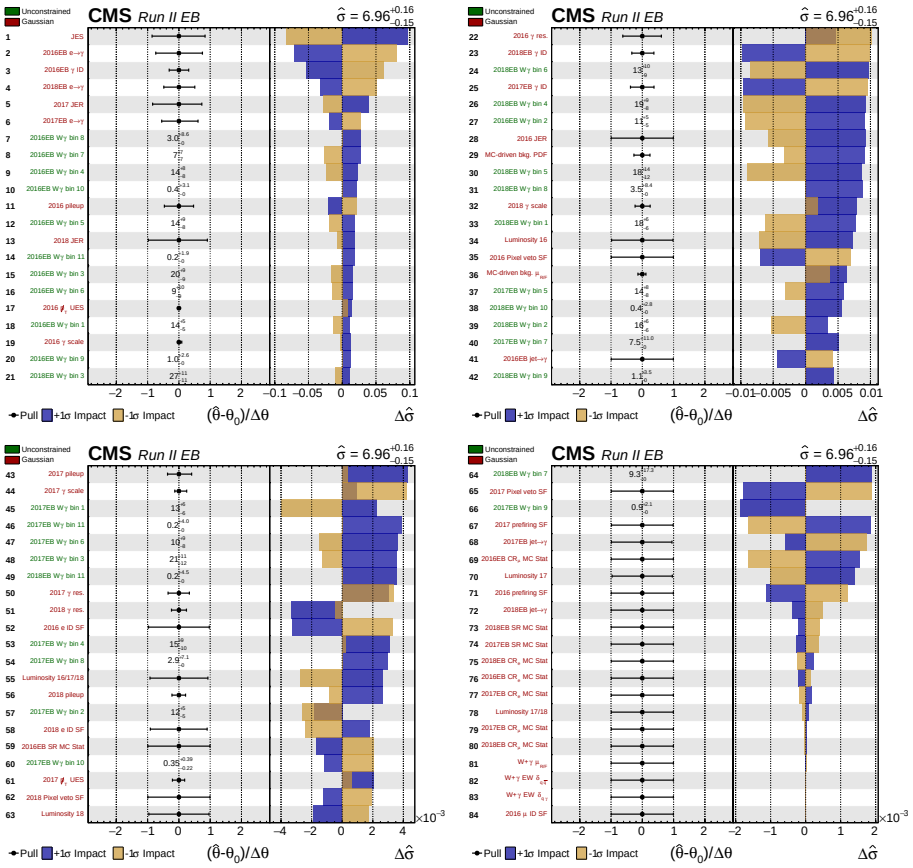

Figure A.1: Impacts and pulls of the systematics on the expected cross section measured in EB.

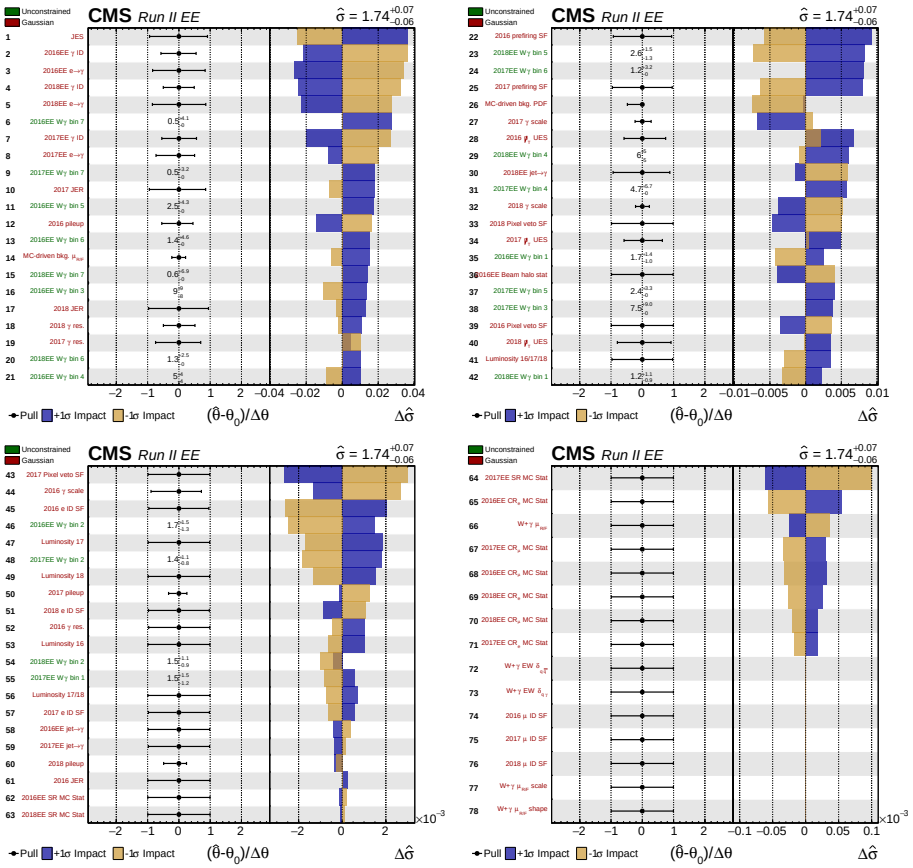

Figure A.2: Impacts and pulls of the systematics on the expected cross section measured in EE.

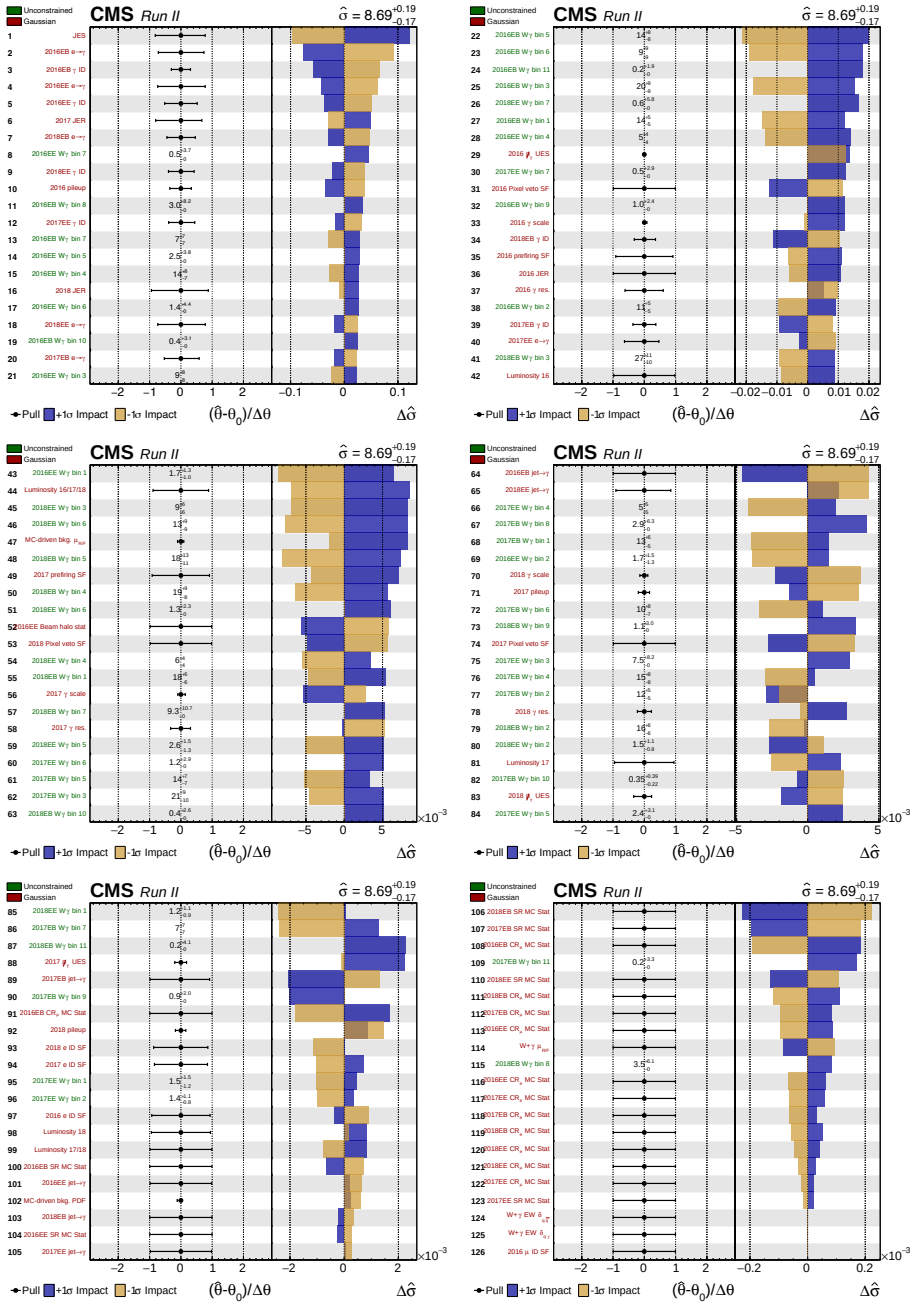

Figure A.3: Impacts and pulls of the systematics on the expected cross section measured combining events in EB and EE.

Appendix B

Photon hybrid ID

The signal and background efficiencies for the photon ID described in Sec. 5.3, the scale factors (SFs) and the SF efficiencies are presented below, for EB and EE photons. The distributions of the SFs as a function of the photon p_T are also shown, together with the error band used for addressing the p_T systematics, as described in Sec. 5.3.

ID efficiencies

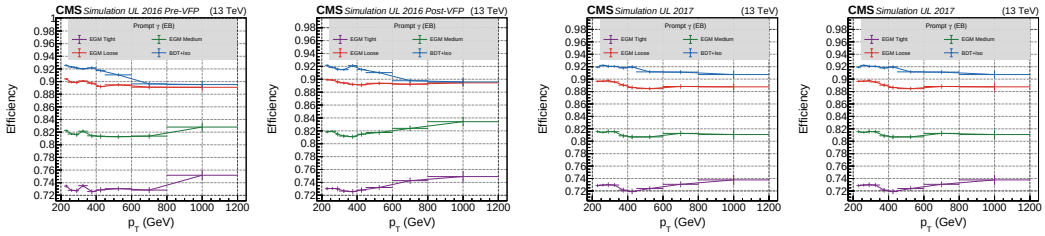

Figure B.1: From left to right: signal efficiency evaluated in EB from MC $\gamma+$jets samples as a function of p_T for years 2016-preVFP, 2016-postVFP, 2017 and 2018.

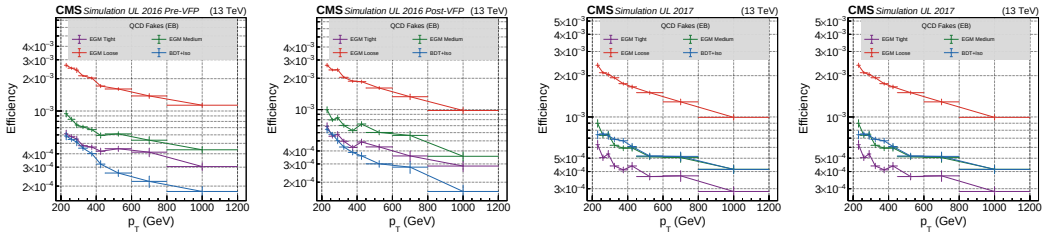

Figure B.2: From left to right: background efficiency evaluated in EB from MC QCD samples as a function of p_T for years 2016-preVFP, 2016-postVFP, 2017 and 2018.

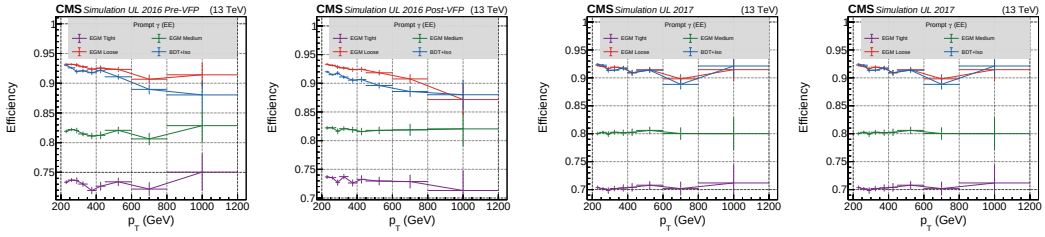

Figure B.3: From left to right: signal efficiency evaluated in EE from MC $\gamma+$jets samples as a function of p_T for years 2016-preVFP, 2016-postVFP, 2017 and 2018.

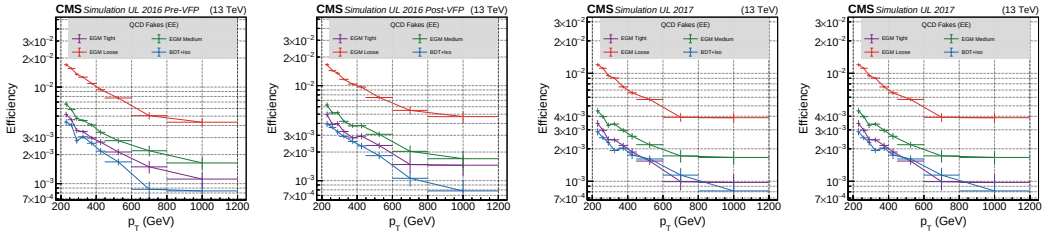

Figure B.4: From left to right: background efficiency evaluated in EE from MC QCD samples as a function of p_T for years 2016-preVFP, 2016-postVFP, 2017 and 2018.

Scale factors and related efficiencies

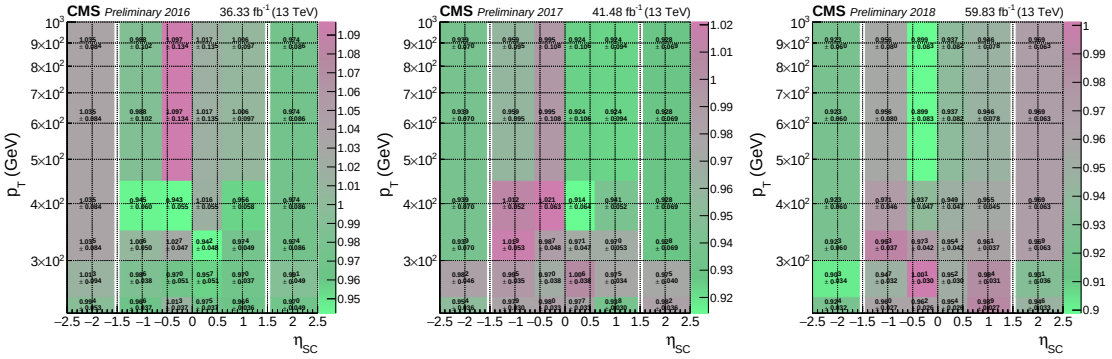

Figure B.5: Photon ID scale factors to correct for differences between data and MC in the estimation of the ID efficiency in 2016 (left), 2017 (center) and 2018 (right).

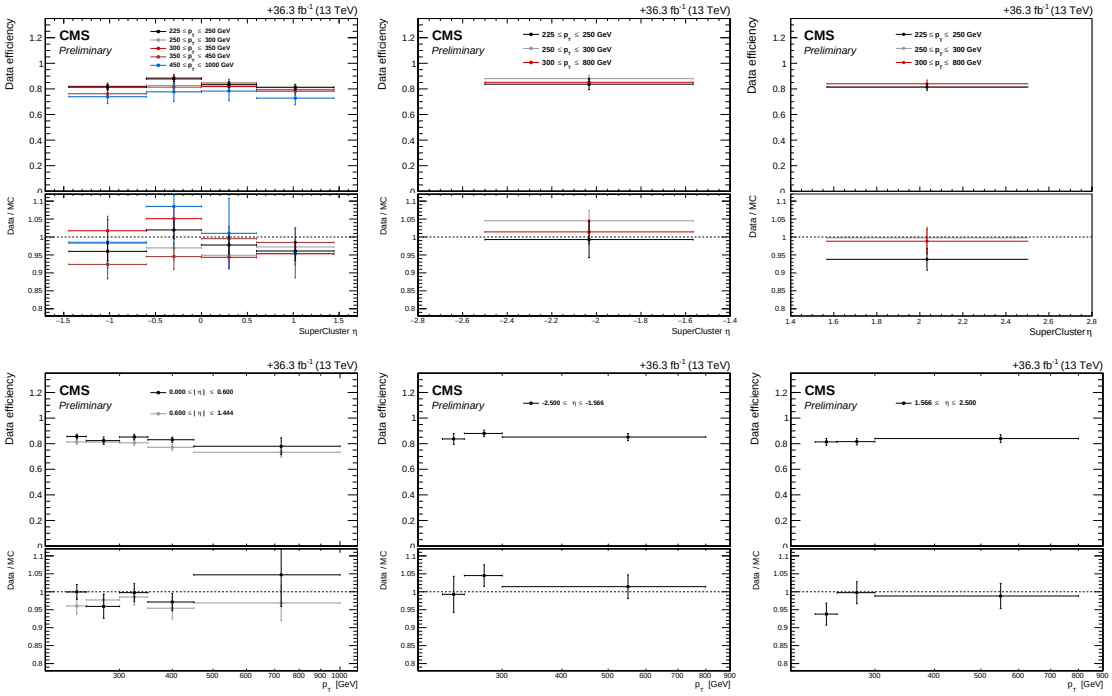

Figure B.6: Photon ID efficiencies in data as a function of η (top) and p_T (bottom), for the 2016 year, in EB (left) and EE (center and right). The ratio between the data and the MC efficiencies is also shown in the bottom panel of each plot.

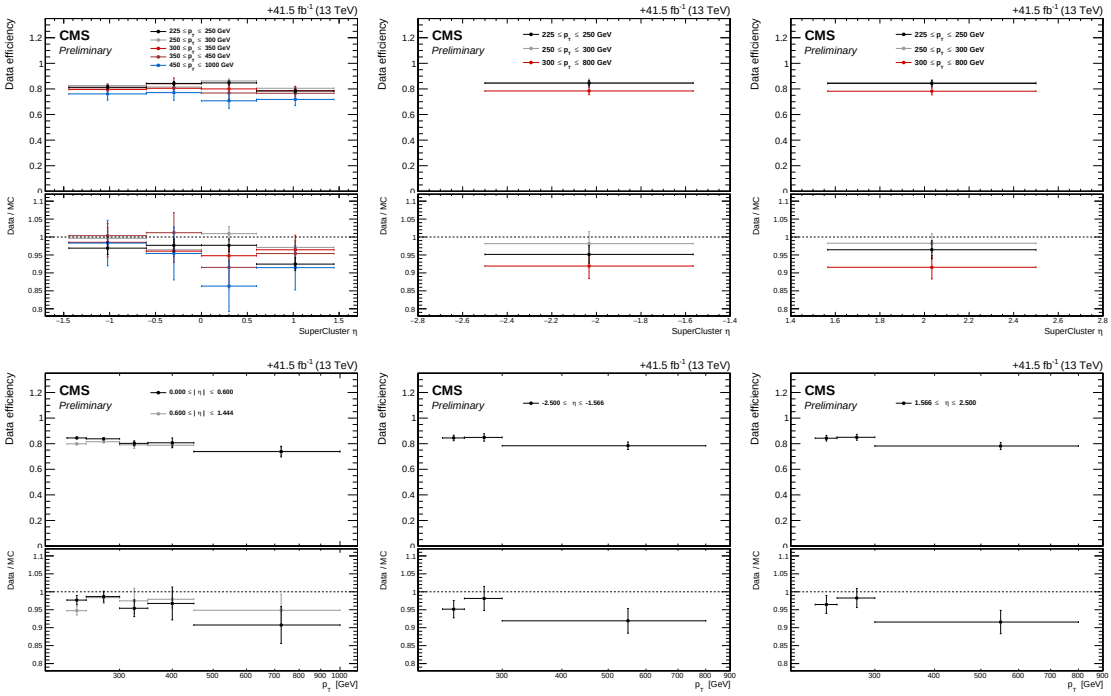

Figure B.7: Photon ID efficiencies in data as a function of η (top) and p_T (bottom), for the 2017 year, in EB (left) and EE (center and right). The ratio between the data and the MC efficiencies is also shown in the bottom panel of each plot.

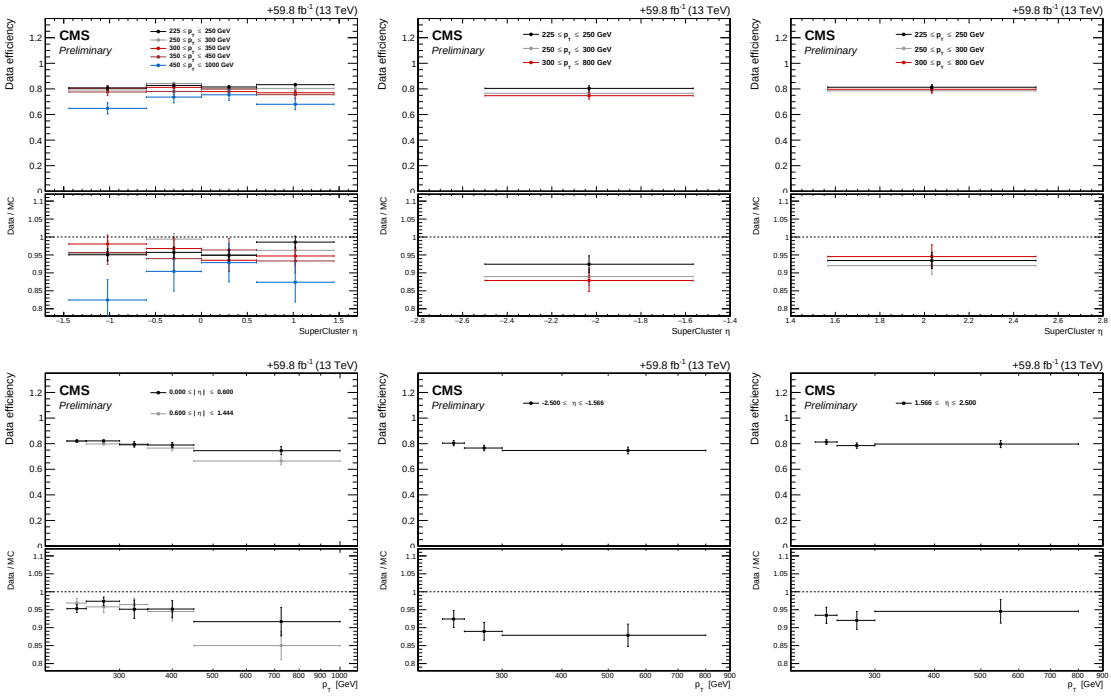

Figure B.8: Photon ID efficiencies in data as a function of η (top) and p_T (bottom), for the 2018 year, in EB (left) and EE (center and right). The ratio between the data and the MC efficiencies is also shown in the bottom panel of each plot.

Scale factors p_T systematics

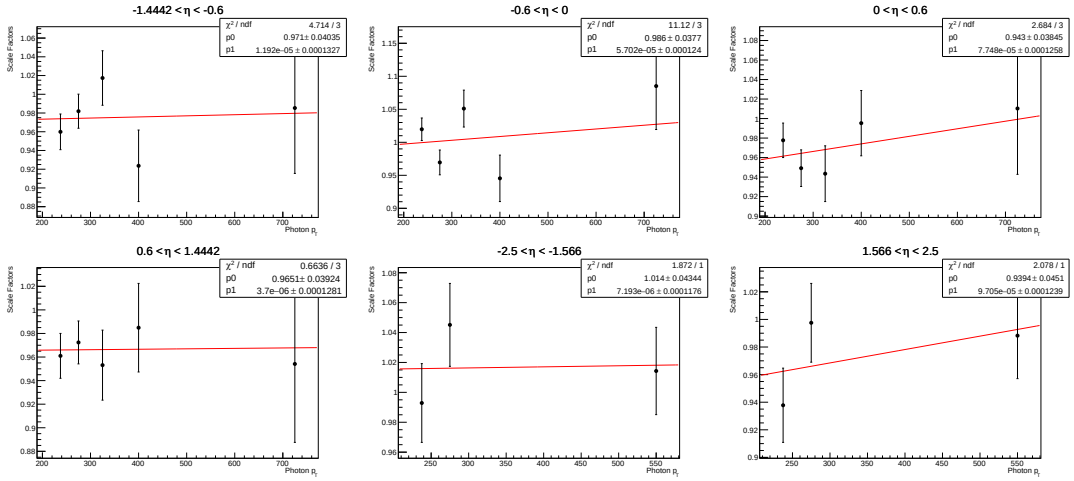

Figure B.9: Fit of the scale factors distribution as a function of p_T for 2016.

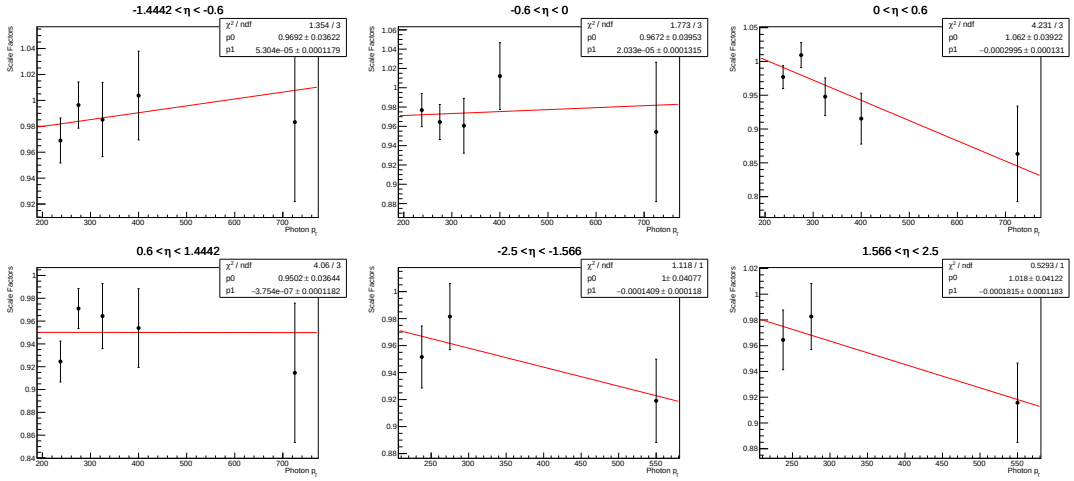

Figure B.10: Fit of the scale factors distribution as a function of p_T for 2017.

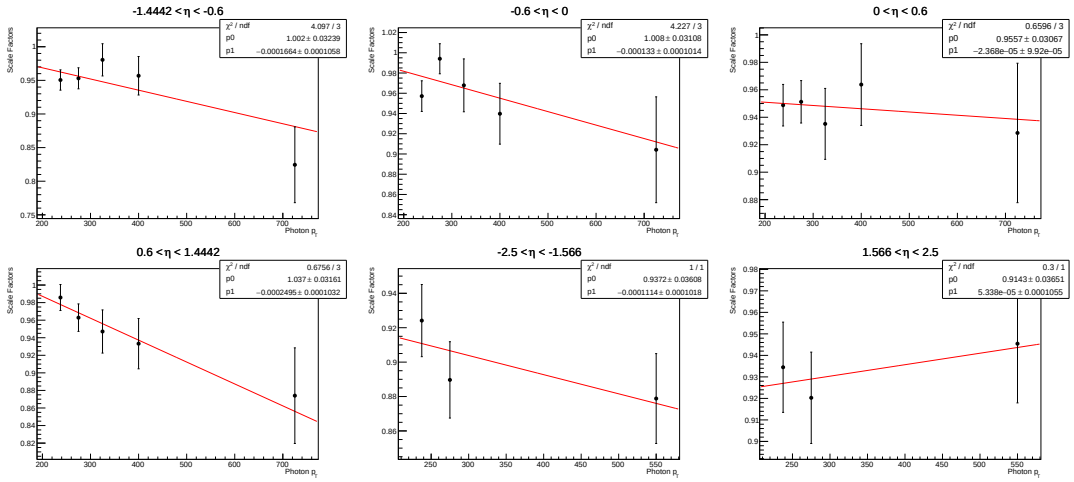

Figure B.11: Fit of the scale factors distribution as a function of p_T for 2018.

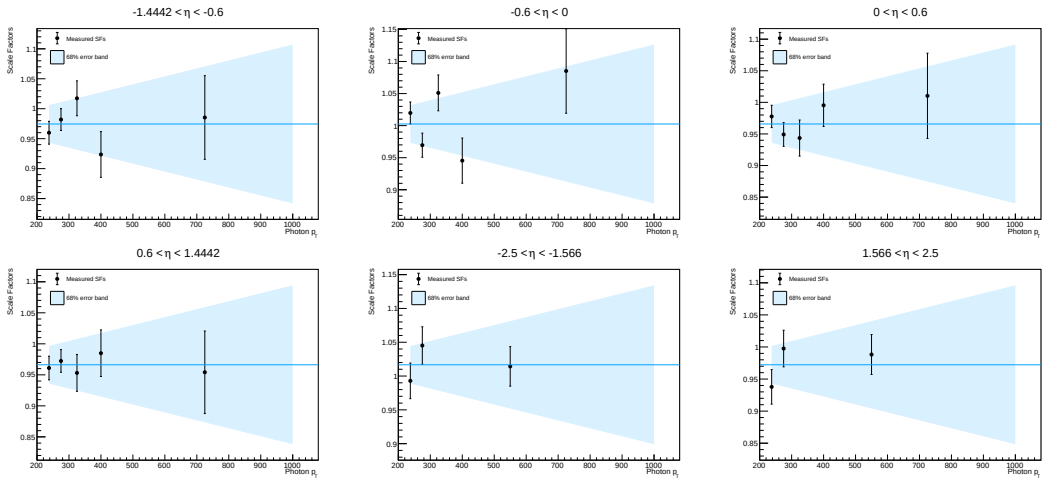

Figure B.12: Additional p_T-systematics estimated for 2016. The error band is obtained from the 68% limit on the gradient of the SFs linear fit.

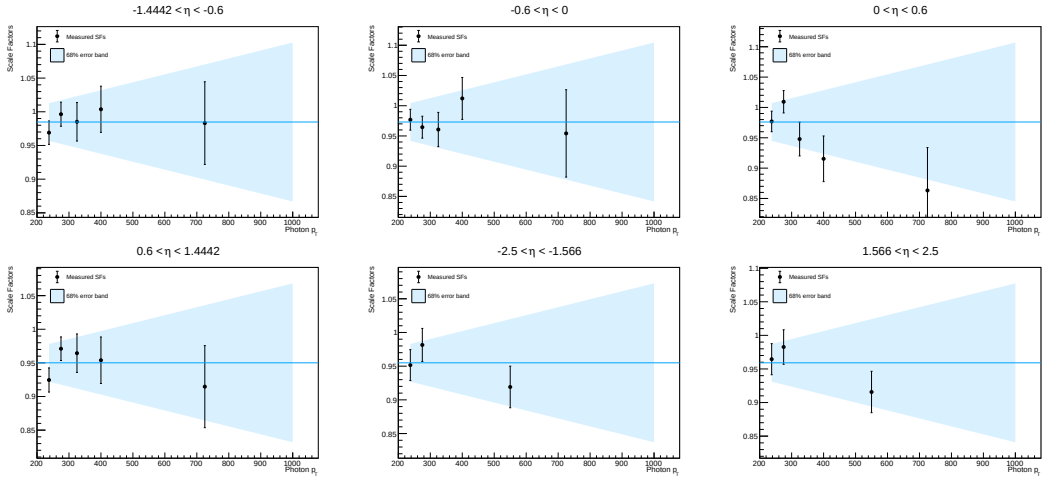

Figure B.13: Additional p_T-systematics estimated for 2017. The error band is obtained from the 68% limit on the gradient of the SFs linear fit.

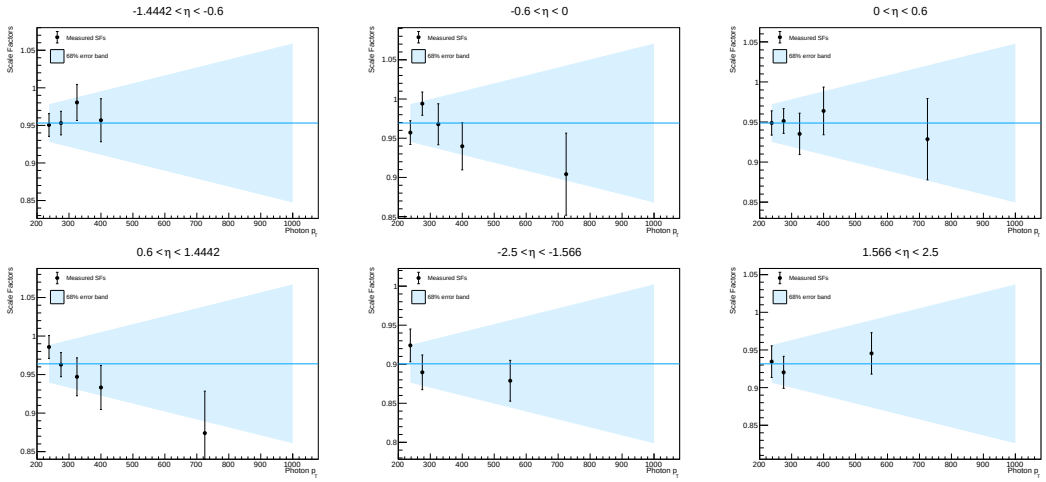

Figure B.14: Additional p_T-systematics estimated for 2018. The error band is obtained from the 68% limit on the gradient of the SFs linear fit.

Appendix C

BDT shape corrections

The data-MC comparisons of the *BDT score* and of the related CDF are shown, for $Z \to e^- e^+$ and γ+jets events, before and after the corrections. In each figure, the *BDT score* is shown in the top distributions, while the CDF is shown in the bottom ones. The first four left plots (2 top and 2 bottom plots) refer to EB events, while the four on the right refer to EE events.

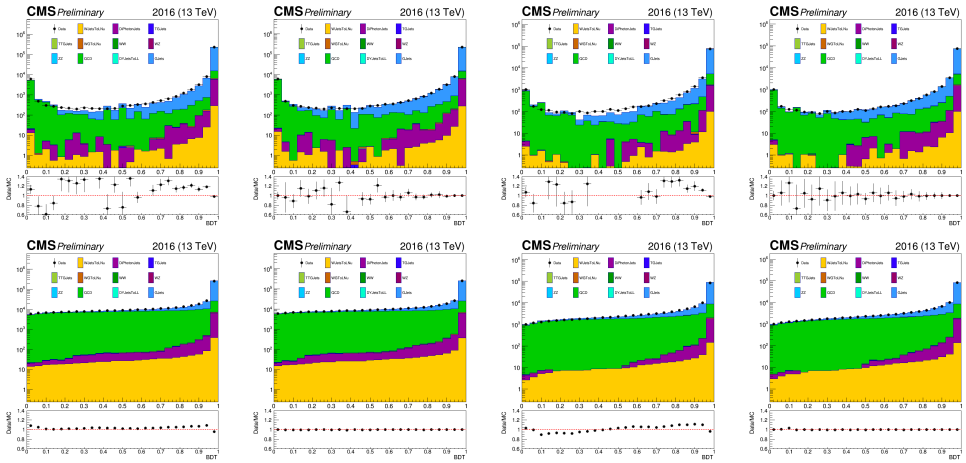

Figure C.1: Data-MC comparison of the *BDT score* and CDF in γ + jets events for 2016, before (left) and after (right) the application of quantile-quantile corrections.

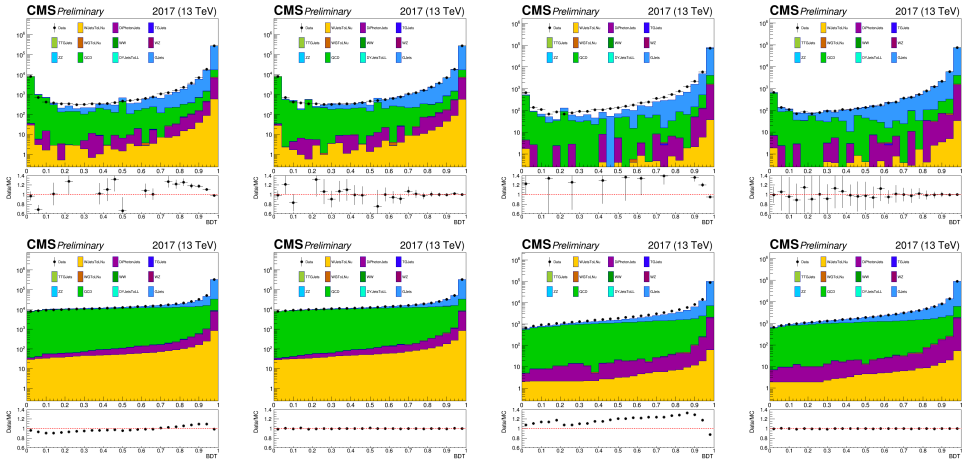

Figure C.2: Data-MC comparison of the *BDT score* and CDF in γ + jets events for 2017, before (left) and after (right) the application of quantile-quantile corrections.

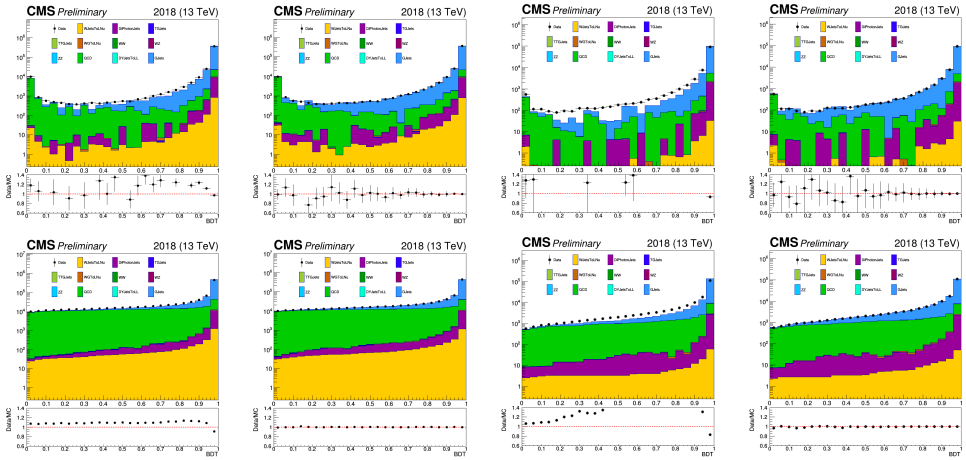

Figure C.3: Data-MC comparison of the *BDT score* and CDF in γ + jets events for 2018, before (left) and after (right) the application of quantile-quantile corrections.

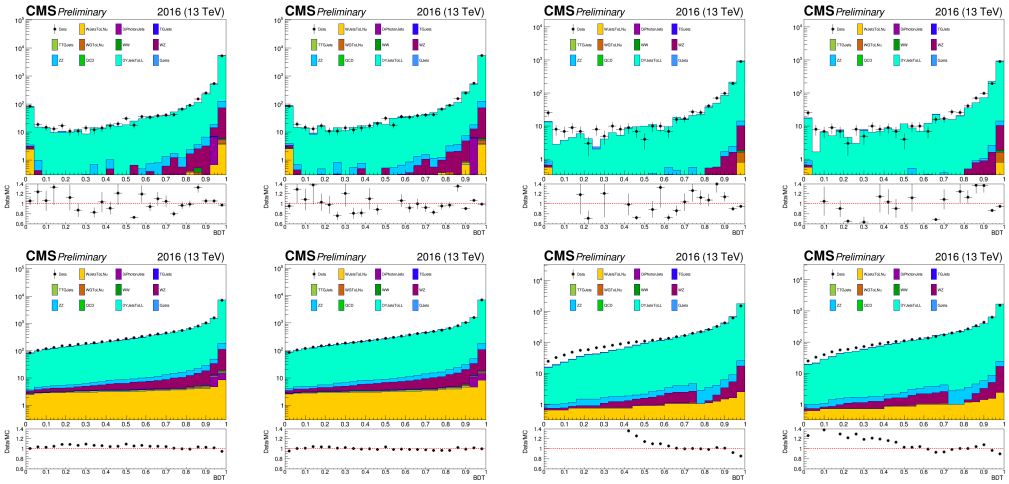

Figure C.4: Data-MC comparison of the *BDT score* and CDF in $Z \to e^- e^+$ + jets events for 2016, before (left) and after (right) the application of quantile-quantile corrections.

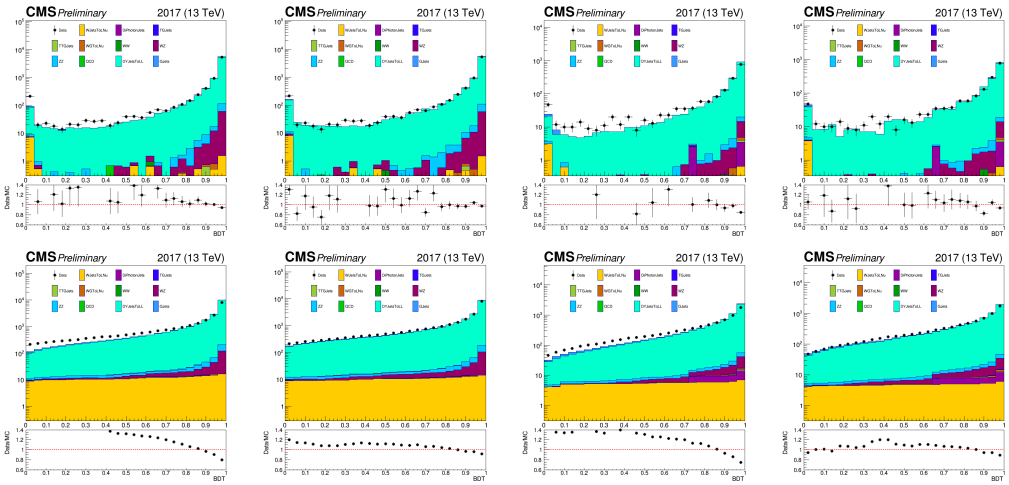

Figure C.5: Data-MC comparison of the *BDT score* and CDF in $Z \to e^- e^+$ + jets events for 2017, before (left) and after (right) the application of quantile-quantile corrections.

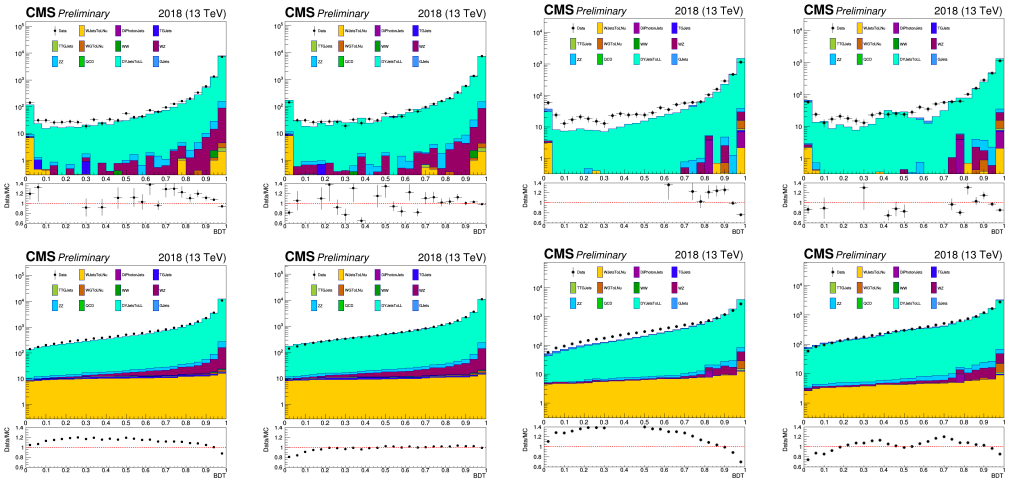

Figure C.6: Data-MC comparison of the *BDT score* and CDF in $Z \to e^- e^+$ + jets events for 2018, before (left) and after (right) the application of quantile-quantile corrections.

Bibliography

1. Salerno, D. in *The Higgs Boson Produced With Top Quarks in Fully Hadronic Signatures* 7–44 (Springer International Publishing, 2019).
2. Thomson, J. J. *On the masses of the ions in gases at low pressures. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **48,** 547–567 (1899).
3. Rutherford, E. *The scattering of alpha and beta particles by matter and the structure of the atom. Phil. Mag. Ser. 6* **21,** 669–688 (1911).
4. Chadwick, J. *The Existence of a Neutron. Proc. Roy. Soc. Lond. A* **136,** 692–708 (1932).
5. Planck, M. in *The Old Quantum Theory* (ed HAAR, D. T.) 82–90 (Pergamon, 1967).
6. Einstein, A. *Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. Annalen der Physik* **322,** 132–148 (1905).
7. Millikan, R. A. *A Direct Photoelectric Determination of Planck's "h". Phys. Rev.* **7,** 355–388 (Mar. 1916).
8. Compton, A. H. *A Quantum Theory of the Scattering of X-rays by Light Elements. Phys. Rev.* **21,** 483–502 (1923).
9. Gilbert, N. L. *The Conservation of Photons. Nature* **118,** 874–875 (1926).
10. Anderson, C. D. *The Apparent Existence of Easily Deflectable Positives. Science* **76,** 238–239 (1932).
11. Dirac, P. A. M. *The quantum theory of the electron. Proc. Roy. Soc. Lond. A* **117,** 610–624 (1928).
12. Yukawa, H. *On the Interaction of Elementary Particles I. Proc. Phys. Math. Soc. Jap.* **17,** 48–57 (1935).
13. Lattes, C. M. G., Muirhead, H., Occhialini, G. P. S. & Powell, C. F. *Processes involving charged mesons. Nature* **159,** 694–697 (1947).
14. Rochester, G. D. & Butler, C. C. *Evidence for the Existence of New Unstable Elementary Particles. Nature* **160,** 855–857 (1947).
15. Brown, R. *et al. Observations With Electron Sensitive Plates Exposed to Cosmic Radiation. Nature* **163,** 82 (1949).
16. Hopper, V. D. & Biswas, S. *Evidence Concerning the Existence of the New Unstable Elementary Neutral Particle. Phys. Rev.* **80,** 1099–1100 (Dec. 1950).
17. Pais, A. *Isotopic spin and mass quantization. Physica* **19,** 869–887 (1953).
18. Gell-Mann, M. *Isotopic Spin and New Unstable Particles. Phys. Rev.* **92,** 833–834 (1953).
19. Nakano, T. & Nishijima, K. *Charge Independence for V-particles. Prog. Theor. Phys.* **10,** 581–582 (1953).

20. Fermi, E. *Tentativo di una teoria dell'emissione dei raggi beta. Ric. Sci.* **4,** 491–495 (1933).

21. Yang, C.-N. & Mills, R. L. *Conservation of Isotopic Spin and Isotopic Gauge Invariance. Phys. Rev.* **96** (eds Hsu, J.-P. & Fine, D.) 191–195 (1954).

22. Wu, C. S., Ambler, E., Hayward, R. W., Hoppes, D. D. & Hudson, R. P. *Experimental Test of Parity Conservation in β Decay. Phys. Rev.* **105,** 1413–1414 (1957).

23. Glashow, S. L. *Partial Symmetries of Weak Interactions. Nucl. Phys.* **22,** 579–588 (1961).

24. Gell-Mann, M. *The Eightfold Way: A Theory of strong interaction symmetry* CTSL-20, TID-12608 (Mar. 1961).

25. Gell-Mann, M. *A Schematic Model of Baryons and Mesons. Phys. Lett.* **8,** 214–215 (1964).

26. Zweig, G. in *Developments in the quark theory of hadrons. Vol. 1. 1964-1978* (eds Lichtenberg, D. B. & Rosen, S. P.) 22–101 (Feb. 1964).

27. Aubert, J. J. *et al. Experimental Observation of a Heavy Particle J. Phys. Rev. Lett.* **33,** 1404–1406 (1974).

28. Augustin, J. E. *et al. Discovery of a Narrow Resonance in e^+e^- Annihilation. Phys. Rev. Lett.* **33,** 1406–1408 (1974).

29. Cowan, C. L., Reines, F., Harrison, F. B., Kruse, H. W. & McGuire, A. D. *Detection of the free neutrino: A Confirmation. Science* **124,** 103–104 (1956).

30. Perl, M. L. *et al. Evidence for Anomalous Lepton Production in e^+e^- Annihilation. Phys. Rev. Lett.* **35,** 1489–1492 (1975).

31. Herb, S. W. *et al. Observation of a Dimuon Resonance at 9.5 GeV in 400 GeV Proton-Nucleus Collisions. Phys. Rev. Lett.* **39,** 252–255 (1977).

32. Abe, F. *et al. Observation of top quark production in $\bar{p}p$ collisions. Phys. Rev. Lett.* **74,** 2626–2631. arXiv: hep-ex/9503002 (1995).

33. Abachi, S. *et al. Observation of the top quark. Phys. Rev. Lett.* **74,** 2632–2637. arXiv: hep-ex/9503003 (1995).

34. Chatrchyan, S. *et al. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. Phys. Lett. B* **716,** 30–61. arXiv: 1207.7235 [hep-ex] (2012).

35. Aad, G. *et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. Phys. Lett. B* **716,** 1–29. arXiv: 1207.7214 [hep-ex] (2012).

36. Thomson, M. *Modern particle physics* (Cambridge University Press, New York, 2013).

37. Kobayashi, M. & Maskawa, T. *CP Violation in the Renormalizable Theory of Weak Interaction. Prog. Theor. Phys.* **49,** 652–657 (1973).

38. Cabibbo, N. *Unitary Symmetry and Leptonic Decays. Phys. Rev. Lett.* **10,** 531–533 (1963).

39. Descotes-Genon, S. & Koppenburg, P. *The CKM Parameters. Ann. Rev. Nucl. Part. Sci.* **67,** 97–127. arXiv: 1702.08834 [hep-ex] (2017).

40. Glashow, S. L. *The renormalizability of vector meson interactions. Nucl. Phys.* **10,** 107–117 (1959).

41. Salam, A. & Ward, J. C. *Weak and electromagnetic interactions. Nuovo Cim.* **11,** 568–577 (1959).

42. Weinberg, S. *A Model of Leptons. Phys. Rev. Lett.* **19,** 1264–1266 (1967).

43. Knippen, G. *Triple and Quartic Gauge Couplings* in. Particle Physics at the LHC - Seminar Report (June 2014).

44. Deur, A., Brodsky, S. J. & de Teramond, G. F. *The QCD Running Coupling. Nucl. Phys.* **90,** 1. arXiv: 1604.08082 [hep-ph] (2016).

45. Workman, R. L. *et al. Review of Particle Physics. PTEP* **2022,** 083C01 (2022).

46. Baak, M. *et al. The Electroweak Fit of the Standard Model after the Discovery of a New Boson at the LHC. Eur. Phys. J. C* **72,** 2205. arXiv: 1209.2716 [hep-ph] (2012).

47. Collins, J. C. & Soper, D. E. *Angular Distribution of Dileptons in High-Energy Hadron Collisions. Phys. Rev. D* **16,** 2219 (1977).

48. Sirunyan, A. M. *et al. Measurement of the weak mixing angle using the forward-backward asymmetry of Drell-Yan events in pp collisions at 8 TeV. Eur. Phys. J. C* **78,** 701. arXiv: 1806.00863 [hep-ex] (2018).

49. CMS Collaboration. *Precision measurement of the Z boson invisible width in pp collisions at* $\sqrt{s}$ *= 13 TeV*. arXiv: 2206.07110 [hep-ex] (June 2022).

50. Haller, J. *et al. Update of the global electroweak fit and constraints on two-Higgs-doublet models. Eur. Phys. J. C* **78,** 675. arXiv: 1803.01853 [hep-ph] (2018).

51. Aaltonen, T. *et al. High-precision measurement of the W boson mass with the CDF II detector. Science* **376,** 170–176 (2022).

52. Arnison, G. *et al. Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at* $\sqrt{s}$ *= 540 GeV. Phys. Lett. B* **122,** 103–116 (1983).

53. Arnison, G. *et al. Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/*c^2 *at the CERN SPS Collider. Phys. Lett. B* **126,** 398–410 (1983).

54. Higgs, P. W. *Broken Symmetries and the Masses of Gauge Bosons. Phys. Rev. Lett.* **13** (ed Taylor, J. C.) 508–509 (1964).

55. Englert, F. & Brout, R. *Broken Symmetry and the Mass of Gauge Vector Mesons. Phys. Rev. Lett.* **13** (ed Taylor, J. C.) 321–323 (1964).

56. Goldstone, J., Salam, A. & Weinberg, S. *Broken Symmetries. Phys. Rev.* **127,** 965–970 (1962).

57. Ellis, J., Gaillard, M. K. & Nanopoulos, D. V. *A Historical Profile of the Higgs Boson*. arXiv: 1201.6045 [hep-ph] (Jan. 2012).

58. Barate, R. *et al. Search for the standard model Higgs boson at LEP. Phys. Lett. B* **565,** 61–75. arXiv: hep-ex/0306033 (2003).

59. *NobelPrize.org, The Nobel Prize in Physics 2013* https://www.nobelprize.org/prizes/physics/2013/summary/.

60. Aad, G. *et al. Combined Measurement of the Higgs Boson Mass in pp Collisions at* $\sqrt{s}$ *= 7 and 8 TeV with the ATLAS and CMS Experiments. Phys. Rev. Lett.* **114,** 191803. arXiv: 1503.07589 [hep-ex] (2015).

61. Erler, J. & Schott, M. *Electroweak Precision Tests of the Standard Model after the Discovery of the Higgs Boson. Prog. Part. Nucl. Phys.* **106,** 68–119. arXiv: 1902.05142 [hep-ph] (2019).

62. Smith, E. *The Hierarchy Problem* in. The University of Chicago, QFT III Final Paper (June 2019).

63. Sakharov, A. D. *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe. Pisma Zh. Eksp. Teor. Fiz.* **5,** 32–35 (1967).

64. Christenson, J. H., Cronin, J. W., Fitch, V. L. & Turlay, R. *Evidence for the* 2π *Decay of the* K_2^0 *Meson. Phys. Rev. Lett.* **13,** 138–140 (1964).

65. Abe, K. *et al. Observation of large CP violation in the neutral B meson system. Phys. Rev. Lett.* **87,** 091802. arXiv: hep-ex/0107061 (2001).

66. Aubert, B. *et al. Observation of CP violation in the* B^0 *meson system. Phys. Rev. Lett.* **87,** 091801. arXiv: hep-ex/0107013 (2001).

67. Aaij, R. *et al. Observation of CP Violation in Charm Decays. Phys. Rev. Lett.* **122,** 211803. arXiv: 1903.08726 [hep-ex] (2019).

68. Boyle, L., Finn, K. & Turok, N. *CPT-Symmetric Universe. Phys. Rev. Lett.* **121,** 251301. arXiv: 1803.08928 [hep-ph] (2018).

69. Weinberg, D. H. *et al. Observational probes of cosmic acceleration. Physics Reports* **530,** 87–225. arXiv: 1201.2434v2 [astro-ph.CO] (2013).

70. Drees, M. *Dark Matter Theory. PoS* **ICHEP2018,** 730. arXiv: 1811.06406 [hep-ph] (2019).

71. *XENONnT - The XENON Experiment* http://www.xenon1t.org.

72. *PandaX* https://pandax.sjtu.edu.cn.

73. Bartram, C. *et al. Search for Invisible Axion Dark Matter in the 3.3-4.2* μeV *Mass Range. Phys. Rev. Lett.* **127,** 261803. arXiv: 2110.06096 [hep-ex] (2021).

74. Felcini, M. *Searches for Dark Matter Particles at the LHC* in *53rd Rencontres de Moriond on Cosmology* (2018), 327–336. arXiv: 1809.06341 [hep-ex].

75. Fukuda, Y. *et al. Evidence for oscillation of atmospheric neutrinos. Phys. Rev. Lett.* **81,** 1562–1567. arXiv: hep-ex/9807003 (1998).

76. Ahmad, Q. R. *et al. Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory. Phys. Rev. Lett.* **89,** 011301. arXiv: nucl-ex/0204008 (2002).

77. Akhmedov, E. K., Branco, G. C. & Rebelo, M. N. *Seesaw mechanism and structure of neutrino mass matrix. Phys. Lett. B* **478,** 215–223. arXiv: hep-ph/9911364 (2000).

78. *CUORE - Cryogenic Underground Observatory for Rare Events* https://cuore.lngs.infn.it/en.

79. Shirai, J. *KamLAND-Zen experiment* in *Proceedings of XIV International Conference on Heavy Quarks and Leptons PoS(HQL2018)* **332** (2018), 050.

80. Abi, B. *et al. Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm. Phys. Rev. Lett.* **126,** 141801. arXiv: 2104.03281 [hep-ex] (2021).

81. Sirunyan, A. M. *et al. Search for new physics in the monophoton final state in proton-proton collisions at* $\sqrt{s}$ *= 13 TeV. JHEP* **10,** 073. arXiv: 1706.03794 [hep-ex] (2017).

82. Khachatryan, V. *et al. Search for new phenomena in monophoton final states in proton-proton collisions at* $\sqrt{s}$ *= 8 TeV. Phys. Lett. B* **755,** 102–124. arXiv: 1410.8812 [hep-ex] (2016).

83. Aad, G. *et al. Search for new phenomena in events with a photon and missing transverse momentum in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector. Phys. Rev. D* **91.** [Erratum: Phys.Rev.D 92, 059903 (2015)], 012008. arXiv: 1411.1559 [hep-ex] (2015).

84. Aaboud, M. *et al. Search for new phenomena in events with a photon and missing transverse momentum in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector. JHEP* **06,** 059. arXiv: 1604.01306 [hep-ex] (2016).

85. Degrande, C. *et al. Effective Field Theory: A Modern Approach to Anomalous Couplings. Annals Phys.* **335,** 21–32. arXiv: 1205.4231 [hep-ph] (2013).

86. Weinberg, S. *Baryon and Lepton Nonconserving Processes. Phys. Rev. Lett.* **43,** 1566–1570 (1979).

87. Gaemers, K. J. F. & Gounaris, G. J. *Polarization amplitudes for $e^+e^- \to W^+W^-$ and $e^+e^- \to ZZ$. Z. Phys. C* **1,** 259 (1979).

88. Hagiwara, K., Peccei, R. D., Zeppenfeld, D. & Hikasa, K. *Probing the weak boson sector in $e^+e^- \to W^+W^-$. Nucl. Phys. B* **282,** 253–307 (1987).

89. Gounaris, G. J., Layssac, J. & Renard, F. M. *New and standard physics contributions to anomalous Z and γ self-couplings. Phys. Rev. D* **62,** 073013. arXiv: hep-ph/0003143 (2000).

90. Gounaris, G. J., Layssac, J. & Renard, F. M. *Off-shell structure of the anomalous Z and γ self-couplings. Phys. Rev. D* **62,** 073012. arXiv: hep-ph/0005269 (2000).

91. Degrande, C. *A basis of dimension-eight operators for anomalous neutral triple gauge boson interactions. JHEP* **02,** 101. arXiv: 1308.6323 [hep-ph] (2014).

92. Buchmuller, W. & Wyler, D. *Effective Lagrangian Analysis of New Interactions and Flavor Conservation. Nucl. Phys. B* **268,** 621–653 (1986).

93. Grzadkowski, B., Iskrzynski, M., Misiak, M. & Rosiek, J. *Dimension-Six Terms in the Standard Model Lagrangian. JHEP* **10,** 085. arXiv: 1008.4884 [hep-ph] (2010).

94. Kopp, G., Walsh, T. & Zerwas, P. *Hadron production in virtual photon-photon annihilation. Nuclear Physics B* **70,** 461–485 (1974).

95. Ellison, J. & Wudka, J. *Study of trilinear gauge boson couplings at the Tevatron collider. Ann. Rev. Nucl. Part. Sci.* **48,** 33–80. arXiv: hep-ph/9804322 (1998).

96. Spagnolo, P. *Boson gauge couplings at LEP* in *40th Rencontres de Moriond on Electroweak Interactions and Unified Theories* (June 2005), 363–370. arXiv: hep-ex/0506048.

97. Gounaris, G. J., Layssac, J. & Renard, F. M. *Signatures of the anomalous Z_γ and ZZ production at the lepton and hadron colliders. Phys. Rev. D* **61,** 073013. arXiv: hep-ph/9910395 (2000).

98. Aaboud, M. *et al. Measurement of the $Z \to \nu\bar{\nu}\gamma$ production cross section in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector and limits on anomalous triple gauge-boson couplings. JHEP* **12,** 010. arXiv: 1810.04995 [hep-ex] (2018).

99. *Limits on anomalous triple and quartic gauge couplings* https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSMPaTGC.

100. *LHC Machine. JINST* **3** (eds Evans, L. & Bryant, P.) S08001 (2008).

101. *High-Luminosity Large Hadron Collider (HL-LHC)*: Technical Design Report V. 0.1. **4/2017** (eds Apollinari, G. *et al.*) (2017).

102. *Launch of the LHC Run 3* https://home.cern/events/launch-lhc-run-3.

103. Mobs, E. *The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018.* https://cds.cern.ch/record/2636343 (2018).

104. Aad, G. *et al. The ATLAS Experiment at the CERN Large Hadron Collider. JINST* **3,** S08003 (2008).

105. Chatrchyan, S. *et al. The CMS Experiment at the CERN LHC. JINST* **3,** S08004 (2008).

106. Alves Jr., A. A. *et al. The LHCb Detector at the LHC. JINST* **3,** S08005 (2008).

107. Aamodt, K. *et al. The ALICE experiment at the CERN LHC. JINST* **3,** S08002 (2008).

108. *About CMS* https://cms.cern/detector.

109. Acquistapace, G. *et al. CMS, the magnet project: Technical design report* CERN-LHCC-97-10, CMS-TDR-1 (May 1997).

110. Hervé, A. *et al. Status of the construction of the CMS magnet. IEEE Trans. Appl. Supercond.* **14** (ed Shintomi, T.) 542–547 (2004).

111. Karimäki, V. *The CMS tracker system project: Technical Design Report* CERN-LHCC-98-006, CMS-TDR-5 (1997).

112. Allkofer, Y. *et al. Design and performance of the silicon sensors for the CMS barrel pixel detector. Nucl. Instrum. Meth. A* **584,** 25–41. arXiv: physics/0702092 (2008).

113. Lindstrom, G. *et al. Radiation hard silicon detectors developments by the RD48 (ROSE) Collaboration. Nucl. Instrum. Meth. A* **466** (eds Ohsugi, T., Unno, Y., Sadrozinski, H. F. W. & Nakano, I.) 308–326 (2001).

114. Rohe, T. *et al. Fluence dependence of charge collection of irradiated pixel sensors. Nucl. Instrum. Meth. A* **552** (eds Borchi, E. & Bruzzi, M.) 232–238. arXiv: physics/0411214 (2005).

115. Adam, W. *et al. The CMS Phase-1 Pixel Detector Upgrade. JINST* **16,** P02027. arXiv: 2012.14304 [physics.ins-det] (2021).

116. Azzurri, P. *The CMS Silicon Strip Tracker. Journal of Physics: Conference Series* **41,** 127–134 (May 2006).

117. Khachatryan, V. *et al. CMS Tracking Performance Results from Early LHC Operation. Eur. Phys. J. C* **70,** 1165–1192. arXiv: 1007.1988 [physics.ins-det] (2010).

118. *The CMS electromagnetic calorimeter project: Technical Design Report* CERN-LHCC-97-033, CMS-TDR-4 (1997).

119. Chatrchyan, S. *et al. Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at* $\sqrt{s}$ *= 7 TeV. JINST* **8,** P09009. arXiv: 1306.2016 [hep-ex] (2013).

120. Orimoto, T. J. *The CMS Ecal Laser Monitoring System* in *10th ICATPP Conference on Astroparticle, Particle, Space Physics, Detectors and Medical Physics Applications* (2008), 481–485.

121. Valsecchi, D. *ECAL trigger performance in Run 2 and improvements for Run 3. JINST* **15,** C06037 (2020).

122. Abdullin, S. *et al. Design, performance, and calibration of CMS hadron-barrel calorimeter wedges. Eur. Phys. J. C* **55,** 159–171 (2008).

123. Bayatian, G. *et al. Design, performance and calibration of the CMS forward calorimeter wedges. Eur. Phys. J. C* **53,** 139–166 (2008).

124. *Measurement of CMS Luminosity* CMS-PAS-EWK-10-004 (2010).

125. Abdullin, S. *et al. Design, performance, and calibration of the CMS Hadron-outer calorimeter. Eur. Phys. J. C* **57,** 653–663 (2008).

126. Sirunyan, A. M. *et al. Calibration of the CMS hadron calorimeters using proton-proton collision data at* $\sqrt{s}$ *= 13 TeV. JINST* **15,** P05002. arXiv: 1910.00079 [physics.ins-det] (2020).

127. *The CMS muon project: Technical Design Report* CERN-LHCC-97-032, CMS-TDR-3 (1997).

128. Chatrchyan, S. *et al. The Performance of the CMS Muon Detector in Proton-Proton Collisions at* $\sqrt{s}$ *= 7 TeV at the LHC. JINST* **8,** P11002. arXiv: 1306.6905 [physics.ins-det] (2013).

129. Khachatryan, V. *et al. The CMS trigger system. JINST* **12,** P01020. arXiv: 1609.02366 [physics.ins-det] (2017).

130. Sert, H. *CMS Run 2 High Level Trigger Performance. PoS* **EPS-HEP2019,** 165 (2020).

131. Brooke, J. J. *et al. Hardware and Firmware for the CMS Global Calorimeter Trigger* in *9th Workshop on Electronics for LHC Experiments* (2003), 226–229.

132. Bawej, T. *et al. The New CMS DAQ System for Run-2 of the LHC* in *19th IEEE-NPSS Real-Time conference 2014* (2014), 1099–1103.

133. Sirunyan, A. M. *et al. Particle-flow reconstruction and global event description with the CMS detector. JINST* **12,** P10003. arXiv: 1706.04965 [physics.ins-det] (2017).

134. Buskulic, D. *et al. Performance of the ALEPH detector at LEP. Nucl. Instrum. Meth. A* **360,** 481–506 (1995).

135. Chatrchyan, S. *et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. JINST* **9,** P10009. arXiv: 1405.6569 [physics.ins-det] (2014).

136. Pantaleo, F. *New Track Seeding Techniques for the CMS Experiment* CERN-THESIS-2017-242 (2017).

137. Schwartz, J. T. *Theory of Self-Reproducing Automata. Mathematics of Computation* **21,** 745–745 (1967).

138. Fruhwirth, R. *Application of Kalman filtering to track and vertex fitting. Nucl. Instrum. Meth. A* **262,** 444–450 (1987).

139. Rose, K. *Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. Proceedings of the IEEE* **86,** 2210–2239 (1998).

140. Fruhwirth, R., Waltenberger, W. & Vanlaer, P. *Adaptive vertex fitting. J. Phys. G* **34,** N343 (2007).

141. Elmetenawee, W. *CMS track reconstruction performance during Run 2 and developments for Run 3. PoS* **ICHEP2020,** 733. arXiv: 2012.07035 [physics.ins-det] (2021).

142. Sirunyan, A. M. *et al. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. JINST* **16,** P05014. arXiv: 2012.06888 [hep-ex] (2021).

143. Oreglia, M. *A Study of the Reactions* $\psi' \to \gamma\gamma\psi$ SLAC-0236 (Dec. 1980).

144. Adam, W., Frühwirth, R., Strandlie, A. & Todor, T. *Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC* CMS-NOTE-2005-001 (Jan. 2005).

145. Cacciari, M., Salam, G. P. & Soyez, G. *The anti-k_t jet clustering algorithm. JHEP* **04,** 063. arXiv: 0802.1189 [hep-ph] (2008).

146. Sirunyan, A. M. *et al. Pileup mitigation at CMS in 13 TeV data. JINST* **15,** P09018. arXiv: 2003.00503 [hep-ex] (2020).

147. Bertolini, D., Harris, P., Low, M. & Tran, N. *Pileup Per Particle Identification. JHEP* **10,** 059. arXiv: 1407.6013 [hep-ph] (2014).

148. Khachatryan, V. *et al. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. JINST* **12,** P02014. arXiv: 1607.03663 [hep-ex] (2017).

149. Chatrchyan, S. *et al. Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS. JINST* **6,** P11002. arXiv: 1107.4277 [physics.ins-det] (2011).

150. Sirunyan, A. M. *et al. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$ = 13 TeV. JINST* **13,** P06015. arXiv: 1804.04528 [physics.ins-det] (2018).

151. Chatrchyan, S. *et al. Performance of CMS Muon Reconstruction in pp Collision Events at $\sqrt{s}$ = 7 TeV. JINST* **7,** P10002. arXiv: 1206.4071 [physics.ins-det] (2012).

152. Chatrchyan, S. *et al. Missing transverse energy performance of the CMS detector. JINST* **6,** P09001. arXiv: 1106.5048 [physics.ins-det] (2011).

153. Sirunyan, A. M. *et al. Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s}$ = 13 TeV using the CMS detector. JINST* **14,** P07004. arXiv: 1903.06078 [hep-ex] (2019).

154. Khachatryan, V. *et al. Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s}$ = 8 TeV. JINST* **10,** P02006. arXiv: 1411.0511 [physics.ins-det] (2015).

155. *ECFA High Luminosity LHC Experiments Workshop: Physics and Technology Challenges. 94th Plenary ECFA meeting* ECFA-13-284 (2013).

156. Aaij, R. *et al. Test of lepton universality in beauty-quark decays. Nature Phys.* **18,** 277–282. arXiv: 2103.11769 [hep-ex] (2022).

157. Archilli, F., Bettler, M. O., Owen, P. & Petridis, K. A. *Flavour-changing neutral currents making and breaking the standard model. Nature* **546,** 221–226 (2017).

158. Ciezarek, G. *et al. A Challenge to Lepton Universality in B Meson Decays. Nature* **546,** 227–233. arXiv: 1703.01766 [hep-ex] (2017).

159. Wehle, S. *et al. Lepton-Flavor-Dependent Angular Analysis of $B \to K^* \ell^+ \ell^-$. Phys. Rev. Lett.* **118,** 111801. arXiv: 1612.05014 [hep-ex] (2017).

160. Crivellin, A. *B-anomalies related to leptons and lepton flavour universality violation. PoS* **BEAUTY2016,** 042. arXiv: 1606.06861 [hep-ph] (2016).

161. Brüning, O. *et al. The scientific potential and technological challenges of the High-Luminosity Large Hadron Collider program. Rept. Prog. Phys.* **85,** 046201 (2022).

162. Brüning, O. & Myers, S. *Challenges And Goals For Accelerators In The XXI Century* (World Scientific Publishing Company, 2016).

163. Calaga, R. *Crab Cavities for the High-luminosity LHC* in *18th International Conference on RF Superconductivity* (2018), THXA03.

164. Tumasyan, A. *et al. The Phase-2 Upgrade of the CMS Tracker* CERN-LHCC-2017-009, CMS-TDR-014 (June 2017).

165. *The Phase-2 Upgrade of the CMS Endcap Calorimeter* CERN-LHCC-2017-023, CMS-TDR-019 (2017).

166. Pedraza-Morales, M. I. *et al. RPC upgrade project for CMS Phase II. JINST* **15,** C05072. arXiv: 1806.11503 [physics.ins-det] (2020).

167. Hebbeker, T. & Korytov, A. *The Phase-2 Upgrade of the CMS Muon Detectors* CERN-LHCC-2017-012, CMS-TDR-016 (Sept. 2017).

168. Zabi, A., Berryhill, J. W., Perez, E. & Tapper, A. D. *The Phase-2 Upgrade of the CMS Level-1 Trigger* CERN-LHCC-2020-004, CMS-TDR-021 (2020).

169. Butler, J. N. & Tabarelli de Fatis, T. *A MIP Timing Detector for the CMS Phase-2 Upgrade* CERN-LHCC-2019-003, CMS-TDR-020 (2019).

170. Agostinelli, S. *et al. GEANT4–a simulation toolkit. Nucl. Instrum. Meth. A* **506,** 250–303 (2003).

171. Coccaro, A., Curtin, D., Lubatti, H. J., Russell, H. & Shelton, J. *Data-driven Model-independent Searches for Long-lived Particles at the LHC. Phys. Rev. D* **94,** 113003. arXiv: 1605.02742 [hep-ph] (2016).

172. Jones, L. L. *et al. The APV25 deep submicron readout chip for CMS detectors. Conf. Proc. C* **9909201,** 162–166 (1999).

173. Alwall, J., Herquet, M., Maltoni, F., Mattelaer, O. & Stelzer, T. *MadGraph 5 : Going Beyond. JHEP* **06,** 128. arXiv: 1106.0522 [hep-ph] (2011).

174. Alwall, J. *et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. JHEP* **07,** 079. arXiv: 1405.0301 [hep-ph] (2014).

175. Degrande, C. *et al. UFO - The Universal FeynRules Output. Comput. Phys. Commun.* **183,** 1201–1214. arXiv: 1108.2040 [hep-ph] (2012).

176. Sjöstrand, T. *et al. An introduction to PYTHIA 8.2. Comput. Phys. Commun.* **191,** 159–177. arXiv: 1410.3012 [hep-ph] (2015).

177. Frederix, R. & Frixione, S. *Merging meets matching in MC@NLO. JHEP* **12,** 061. arXiv: 1209.6215 [hep-ph] (2012).

178. Artoisenet, P., Frederix, R., Mattelaer, O. & Rietkerk, R. *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations. JHEP* **03,** 015. arXiv: 1212.3460 [hep-ph] (2013).

179. Mangano, M. *The so-called MLM prescription for ME/PS matching* in *Fermilab ME/MC Tuning Workshop* (Oct. 2002).

180. Mangano, M. L., Moretti, M., Piccinini, F. & Treccani, M. *Matching matrix elements and shower evolution for top-quark production in hadronic collisions. JHEP* **01,** 013. arXiv: hep-ph/0611129 (2007).

181. Chrzaszcz, M., Przedzinski, T., Was, Z. & Zaremba, J. *TAUOLA of τ lepton decays-framework for hadronic currents, matrix elements and anomalous decays. Comput. Phys. Commun.* **232,** 220–236. arXiv: 1609.04617 [hep-ph] (2018).

182. Sorrentino, G. *Monte Carlo tuning in CMS. Nuovo Cim. C* **44,** 79 (2021).

183. Ball, R. D. *et al. Parton distributions from high-precision collider data. Eur. Phys. J. C* **77,** 663 (2017).

184. Forte, S. *Parton distributions at the dawn of the LHC. Acta Phys. Polon. B* **41,** 2859–2920 (2010).

185. Gao, J., Harland-Lang, L. & Rojo, J. *The Structure of the Proton in the LHC Precision Era. Phys. Rept.* **742,** 1–121 (2018).

186. Plaakyt, R. *Structure Functions and PDF determination at HERA* in *48th Rencontres de Moriond on QCD and High Energy Interactions* (2013), 349–352.

187. Chen, T. & Guestrin, C. *XGBoost: A Scalable Tree Boosting System* in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, California, USA, 2016), 785–794.

188. *XGBoost 1.2.0* https://xgboost.readthedocs.io/en/release_1.2.0/index.html.

189. Bergstra, J., Komer, B., Eliasmith, C., Yamins, D. & Cox, D. D. *Hyperopt: a Python library for model selection and hyperparameter optimization. Computational Science and Discovery* **8,** 014008 (July 2015).

190. *Hyperopt: Distributed Hyperparameter Optimization* https://github.com/hyperopt/hyperopt.

191. *Cut Based Photon ID for Run 2* https://twiki.cern.ch/twiki/bin/viewauth/CMS/CutBasedPhotonIdentificationRun2.

192. Devroye, L. *Non-Uniform Random Variate Generation* (Springer, originally published with Springer-Verlag, New York, 1986).

193. Sirunyan, A. M. *et al. Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at* $\sqrt{s}$ *= 13 TeV. JHEP* **11,** 185. arXiv: 1804.02716 [hep-ex] (2018).

194. Petyt, D. A. *Mitigation of anomalous APD signals in the CMS electromagnetic calorimeter. J. Phys. Conf. Ser.* **404** (ed Akchurin, N.) 012043 (2012).

195. Sirunyan, A. M. *et al. Search for new physics in final states with a single photon and missing transverse momentum in proton-proton collisions at* $\sqrt{s}$ *= 13 TeV. JHEP* **02,** 074. arXiv: 1810.00196 [hep-ex] (2019).

196. Drollinger, V. *Simulation of beam halo and cosmic muons* CMS-NOTE-2005-012 (July 2005).

197. Cowan, G., Cranmer, K., Gross, E. & Vitells, O. *Asymptotic formulae for likelihood-based tests of new physics. Eur. Phys. J. C* **71.** [Erratum: Eur.Phys.J.C 73, 2501 (2013)], 1554 (2011).

198. Scargle, J. D., Norris, J. P., Jackson, B. & Chiang, J. *Studies in Astronomical Time Series Analysis. VI. Bayesian Block Representations. Astrophys. J.* **764,** 167. arXiv: 1207.5578 [astro-ph.IM] (2013).

199. Cash, W. *Parameter estimation in astronomy through application of the likelihood ratio. Astrophys. J.* **228,** 939–947 (1979).

200. *CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update* CMS-PAS-LUM-13-001 (2013).

201. Van der Meer, S. *Calibration of the Effective Beam Height in the ISR* CERN-ISR-PO-68-31 (June 1968).

202. Denner, A., Dittmaier, S., Hecht, M. & Pasold, C. *NLO QCD and electroweak corrections to* $W + \gamma$ *production with leptonic W-boson decays. JHEP* **04,** 018. arXiv: 1412.7421 [hep-ph] (2015).

203. Denner, A., Dittmaier, S., Hecht, M. & Pasold, C. *NLO QCD and electroweak corrections to* $Z + \gamma$ *production with leptonic Z-boson decays. JHEP* **02,** 057. arXiv: 1510.08742 [hep-ph] (2016).

204. Grazzini, M., Kallweit, S. & Wiesemann, M. *Fully differential NNLO computations with MATRIX. Eur. Phys. J. C* **78,** 537. arXiv: 1711.06631 [hep-ph] (2018).

205. Abdul Khalek, R. *et al. Parton Distributions with Theory Uncertainties: General Formalism and First Phenomenological Studies. Eur. Phys. J. C* **79,** 931. arXiv: 1906.10698 [hep-ph] (2019).

206. Lindert, J. M. *et al. Precise predictions for V+jets dark matter backgrounds. Eur. Phys. J. C* **77,** 829. arXiv: 1705.04664 [hep-ph] (2017).