Università degli Studi di Torino
Scuola di Dottorato

Dottorato in Fisica ed Astrofisica

Search for the rare decay of the W boson into a pion and a photon in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the CMS experiment

Riccardo Salvatico

Tutor: Prof. Stefano Argirò

Co-tutor: Dr. Mario Pelliccioni

"*Things are only impossible until they're not!*"

Captain Jean-Luc Picard

Preface

The W boson, one of the two mediators of the weak interaction, has been broadly studied at both e^+e^- and hadron colliders since its discovery in 1983. Despite the increasing accuracy in the measurement of its properties, it still remains relatively poorly known if compared to the other mediator, the Z boson, mainly due to the bigger experimental challenges in the reconstruction of the W boson decays.

The work documented in this thesis regards the search for the rare decay of the W boson into a pion and a photon, using the data collected in proton-proton collisions at the center-of-mass energy of 13 TeV with the CMS experiment. Similarly to the other exclusive hadronic decays of the vector bosons, the decay $W \to \pi\gamma$ is foreseen by the Standard Model of particle physics, but has never been observed. On the one hand, the observation would represent a good probe for the Standard Model and would provide insights into quantum chromodynamics factorization and meson form factors at high energy scales; on the other, it could offer a new way to measure the mass of the W boson that is based solely on visible single-particle decay products.

Performing such a search at CMS presents a number of challenges, from the definition of a suitable trigger strategy, to the individuation of the most effective methods for background suppression and signal enhancement. Differently from the only previous search for $W \to \pi\gamma$, performed by the CDF Collaboration, this novel analysis exploits the W production in top quark-antiquark pair events, thanks to the relatively large cross section for such process at the LHC. The leptonic decay of the W boson from one of the top quarks is used to tag the event, and the b quark jets are exploited to reduce the background from the hadronization of light-flavor quarks and gluons. The W boson originating from the other top quark is then used to search for the $W \to \pi\gamma$ decay. Such events are characterized by an isolated track and an isolated photon of large transverse momentum.

The analysis selection criteria are designed before inspecting the signal region in the collision data, so to avoid any observational bias. Both cut-based and multivariate techniques are employed in the offline event selection, which also makes use of a dedicated pion-isolation variable, conceived to discriminate between pions originating from electroweak processes and pions surrounded by jet activity. The $W \to \pi\gamma$ branching fraction is then determined by fitting the reconstructed pion-photon invariant mass spectrum. No significant excess is observed above the expected background. The first upper limit on the $W \to \pi\gamma$ branching fraction at the LHC is set.

The contents of this thesis are organized as follows: Chapter 1 provides a historical and physical background for this analysis, from the description of the Standard Model and the electroweak interactions, to the measurements of the W boson properties at collider experiments. There I discuss the physics motivations behind the search for the decay $W \to \pi\gamma$, and I outline the analysis strategy adopted.

The LHC and the CMS detector are described in Chapter 2, with particular emphasis on the identification and reconstruction of the particles and physics objects involved in this search.

Chapter 3 offers a detailed description of the data samples and the simulated processes the analysis is based on.

In Chapter 4, I discuss in detail the offline event selection procedures, which aim at suppressing the background sources and isolating signal-like events.

The core of the analysis, that is the method used to extract the branching fraction of $W \to \pi\gamma$ from a fit to the pion-photon invariant mass spectrum, is presented in Chapter 5.

Chapter 6 outlines the sources of systematic uncertainty that affect this search, describing the way they are estimated and how they affect the final results.

The final results are presented in Chapter 7, including the extraction of the branching fraction of $W \to \pi\gamma$ and the calculation of an upper exclusion limit on this parameter.

Contents

Chapter 1

Physics motivations

The Standard Model of particle physics includes what nowadays we know as weak (or electroweak) force, one of the possible ways the particles in the model interact. In fact, the understanding and rigorous formulation of the electroweak force were not achieved until the second half of the last century. The discovery of the mediators of electroweak interactions, the W and Z bosons, followed the theoretical construction and represented a great success, paving the way to an even deeper understanding of Nature.

1.1 Historical perspectives

The idea of the existence of the W boson as mediator of the weak force is strictly bound to the phenomenon of radioactivity. At the end of the nineteenth century, three kinds of radiation had been observed: α, β, and γ rays. As they all indicated different transformations of their mother nuclei, α and β rays were found to change the electric charge of the initial nucleus, resulting in a different element in the final state. On the other hand, γ rays were emitted by excited nuclei to reach their ground energy state. At the time and for a few decades, though, the most puzzling difference among these physical processes was the observation of the continuous energy spectrum of β rays, as opposed to the discrete energy spectrum of the α and γ radiation. A β decay was thought to proceed as $N_i \to N_f + \mathrm{e}^-$, where N_i and N_f are the initial and final nuclei, respectively. The kinematics of such a process, under the hypothesis of energy conservation and negligible recoil of the final nucleus (T_f), is:

$$\begin{aligned} & m_i c^2 = m_f c^2 + T_f + m_\mathrm{e} c^2 + T_\mathrm{e} \\ & \Rightarrow T_\mathrm{e} = m_i c^2 + m_f c^2 - m_\mathrm{e} c^2 - T_f \approx m_i c^2 + m_f c^2 - m_\mathrm{e} c^2, \end{aligned} \tag{1.1}$$

Therefore, it was expected that the kinetic energy of the electron was a fixed quantity, corresponding to $T_\mathrm{e} \approx m_i c^2 + m_f c^2 - m_\mathrm{e} c^2$, while the observations indicated a continuous energy spectrum. Wolfgang Pauli tried to address the problem in such a way that the energy conservation precept was not violated: in 1931, he proposed the existence of a still undetected, electrically neutral particle that now we call *neutron*. The neutron was supposed to be a constituent of the nuclei, and at the same time to be emitted in β decays, thus explaining the continuous β spectrum. A couple of years later, when James Chadwick had discovered the neutron [1] and it was clear that this particle could not fix the β spectrum problem, the search was directed to a much lighter particle, possibly even massless, and with a penetrating power exceeding many times that of photons. Such particle played a crucial role in the famous theory of β decays formulated by Enrico Fermi, and was named *neutrino* by the Italian physicist.

Fermi proposed that, in β decays, both an electron e^- and a neutrino $\bar{\nu}$ are emitted from a neutron n [2]. Drawing inspiration from quantum electrodynamics (QED), his

theory led to the following expression for the Lagrangian density of β decay:

$$\mathcal{L}_\beta = G_\mathrm{F} j_\mu^{\mathrm{n}\to\mathrm{p}} j^\mu_{\nu\to\mathrm{e}} = G_\mathrm{F}(\bar{u}_\mathrm{p}\gamma_\mu u_\mathrm{n})(\bar{u}_\mathrm{e}\gamma^\mu u_\nu), \tag{1.2}$$

where the u terms are the spinors associated to proton, neutron, electron, and neutrino, $\bar{u}$ are their complex conjugates, and $\gamma_\mu^{(\mu)}$ is a Dirac matrix. G_F, known as the Fermi constant, was a completely new coupling constant which, therefore, indicated the existence of a new force of nature.

A strength of Fermi's thesis was that it was not limited to β^- decays ($\mathrm{n} \to \mathrm{p}\,\mathrm{e}^-\bar{\nu}$), but it was actually descriptive of several other processes, an important one being the muon decay ($\mu^- \to \mathrm{e}^-\nu_\mu\bar{\nu}_\mathrm{e}$ or $\mu^+ \to \mathrm{e}^+\nu_\mathrm{e}\bar{\nu}_\mu$), proposed by Bruno Pontecorvo in 1947 [3] and observed two years later by Jack Steinberger [4]. Nevertheless, the astounding discovery that parity was violated in weak interactions [5] highlighted a flaw in Fermi's theory, fully symmetric under spatial inversion. In 1957 that the efforts of the physics community gave birth to the V–A theory [6], according to which the Lagrangian of weak interactions comprises a vector and an axial vector terms that correctly describe parity violation.

Despite the great advancement it introduced, the V–A theory was not able to solve another criticality: the fields representing the four fermions in Eq. (1.2) are evaluated at the same point in space and time, in what is called a "contact" interaction. This aspect depicts a force with no range. While low-energy weak interactions are well described in this framework, problems arise when considering high-energy processes, since the theory predicts that the involved cross sections rise linearly with energy. This yields to a violation of the unitarity principle, which states that the probability for a certain process to happen must be less than or equal to unity.

The attempts to amend Fermi's theory proceeded through the replacement of the four-fermion, point-like interaction with a particle exchange mechanism, and culminated in the late '60s with the unification of electromagnetic and weak forces using the language of group theory at the base of the Standard Model.

1.2 The Standard Model

The Standard Model of particle physics is a quantum field theory that describes all the known elementary particles. Its foundations were laid by Sheldon Glashow [7], Steven Weinberg [8], Abdus Salam [9], and other physicists operating in the '60s, who conjectured that the electromagnetic and the weak interactions were different manifestations of a single, unified *electroweak* force. In its current formulation, the Standard Model also describes the strong interaction between particles, thus failing to incorporate only those phenomena generated by the gravitational interaction. Gravity remains therefore the only one of the four fundamental forces of nature to miss a predictive quantum description. Nevertheless, its contribution at the energy ranges spanned by modern particle colliders can be considered as negligible. Matter is described in terms of elementary spin-1/2 particles, called *fermions*. The model theorizes the existence of twelve elementary fermions, along with their respective anti-fermions, which have all been experimentally observed or inferred with a great level of confidence. Depending on the way they are allowed to interact with other particles, fermions are divided into *quarks* and *leptons*. Quarks interact via both the electroweak and the strong force, and consequently have non-zero *weak hypercharge* and *color charge*, which are the quantum numbers conserved by electroweak and strong interactions, respectively. On the other hand, leptons do not carry any color charge and only have electroweak interactions with other particles. Leptons and quarks can be further categorized into three

flavor families or generations, the reason behind the number three still representing a remarkable open question concerning the Model. Each of these families contains a quark doublet, composed by quarks with electric charge equal to $+2/3$ and $-1/3$ of the proton charge, respectively, and a leptonic doublet accommodating a lepton with electric charge -1 and an electrically neutral neutrino. The three fundamental forces comprised in the Standard Model, acting on elementary particles and on composite particles made of quarks, are mediated by spin-1 particles called *bosons*. In particular, the carriers of the strong force are massless *gluons* (g), while the electroweak force is carried by massless photons (γ), responsible for electromagnetic interactions, and by the massive W^+, W^-, and Z bosons. The two W bosons and the Z boson mediate the *charged* and the *neutral currents* in weak processes, respectively. The Model is completed by the scalar (spin-0) Higgs boson, produced by the quantum excitation of the field that allows elementary particles to acquire mass. A complete picture of the fundamental constituents of the Standard Model is shown in Figure 1.1.

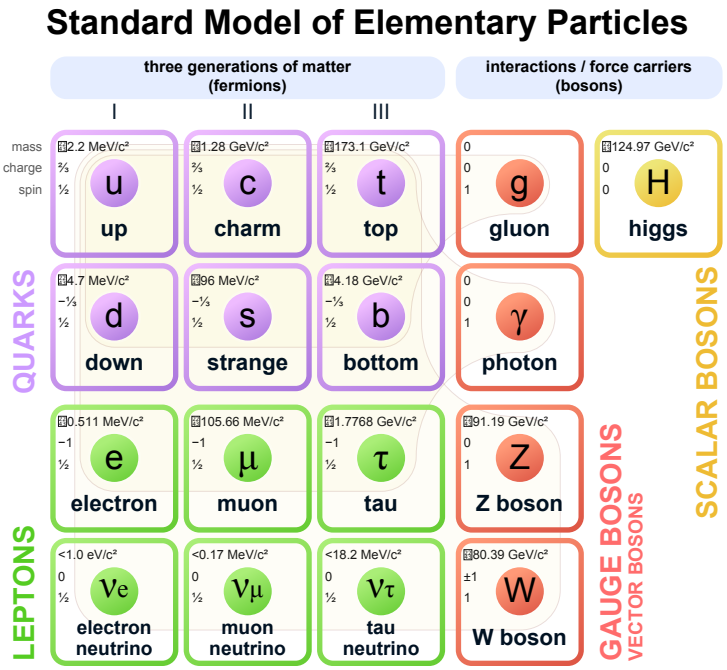

FIGURE 1.1: The fundamental constituents of the Standard Model.

1.2.1 Electroweak interactions

In order to explain the complex phenomenology of elementary particles, the Standard Model adopts a principle of symmetry conservation under local *gauge* transformations, which are geometrical transformations in the four-dimensional spacetime. In particular, it is possible to explain both electromagnetic and weak phenomena by imposing the invariance of the fermionic Lagrangian under local transformations of the group $SU(2)_L \otimes U(1)_Y$, where the subscripts L and Y indicate the coupling with the left-handed fermions and the weak hypercharge, respectively. The generators of $SU(2)_L$ are the three components of the weak isospin operator ($T^i = 1/2\tau^i$, where τ^i are the three Pauli matrices) and the generator of $U(1)_Y$ is the weak hypercharge operator. The weak isospin (T) and hypercharge are related to the electric charge (Q)

as follows:

$$Q = T_3 + \frac{Y}{2}, \tag{1.3}$$

where T_3 is the eigenvalue of the third component of the isospin. This model inherits from the framework used in the V–A theory, with quark and lepton fields organized in left-handed doublets and right-handed singlets:

$$\begin{array}{ccc} \begin{pmatrix} u \\ d \end{pmatrix}_L,\ u_R,\ d_R & \begin{pmatrix} c \\ s \end{pmatrix}_L,\ c_R,\ s_R & \begin{pmatrix} t \\ b \end{pmatrix}_L,\ t_R,\ b_R \\ \begin{pmatrix} \nu_{\mathrm{e}} \\ \mathrm{e} \end{pmatrix}_L,\ \mathrm{e}_R & \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L,\ \mu_R & \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L,\ \tau_R \end{array} \tag{1.4}$$

Since only the weak isospin third component of the fields in the doublets is non-null, the left-handed fermionic fields may undergo weak interactions, while right-handed particles are blind to it. This provides an explanation to the phenomenon of maximal parity violation in weak interactions observed in experiments. In addition, it should be noted that the right handed neutrino ν_R does not carry $SU(2)_L$ or $U(1)_Y$ charges, and thus decouples from the electroweak interaction.

The theory contains four spin-1 gauge fields, or bosons, associated to the generators of the gauge transformation: W^i_μ $(i = 1, 2, 3)$ and B_μ. The interaction among these gauge bosons and the fermionic fields (ψ) can be mathematically described by the Lagrangian:

$$\mathcal{L}_{\mathrm{EW}} = -\frac{1}{4}\overrightarrow{W}_{\mu\nu} \cdot \overrightarrow{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} + \bar{\psi}i\gamma^\mu D_\mu \psi, \tag{1.5}$$

using the field strength tensors

$$W^i_{\mu\nu} \equiv \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g\varepsilon^{ijk}W^j_\mu W^k_\nu \qquad B_{\mu\nu} \equiv \partial_\mu B_\nu - \partial_\nu B_\mu \tag{1.6}$$

and the covariant derivative

$$D_\mu \equiv \partial_\mu - i\frac{g}{2}\vec{\tau} \cdot \overrightarrow{W}_\mu - i\frac{g'}{2}B_\mu Y, \tag{1.7}$$

with gauge couplings g and g'. The mass eigenstates of these bosons can be expressed as a combination of the electroweak eigenstates W^i_μ and B_μ:

$$\begin{aligned} \mathrm{W}^\pm_\mu &= \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu), \\ \mathrm{Z}_\mu &= \frac{gW^3_\mu - g'B_\mu}{\sqrt{g^2 + g'^2}} = W^3_\mu \cos\theta_W - B_\mu \sin\theta_W, \\ \mathrm{A}_\mu &= \frac{g'W^3_\mu + gB_\mu}{\sqrt{g^2 + g'^2}} = W^3_\mu \sin\theta_W + B_\mu \cos\theta_W, \end{aligned} \tag{1.8}$$

with θ_W the weak (or "Weinberg") mixing angle, and $g'/g = \tan\theta_W$. Nevertheless, the gauge and chiral symmetries forbid to write a mass term for bosons and fermions. Thus, the $SU(2)_L \otimes U(1)_Y$ Lagrangian (1.5) contains only massless fields, whereas from experimental observations we certainly know they are not.

1.2.2 Electroweak symmetry breaking

A natural way of allowing the Standard Model elementary particles to acquire mass is through the existence of a complex scalar field ϕ, in the form of an $SU(2)$ doublet:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}. \tag{1.9}$$

Choosing for the scalar hypercharge the value $Y = 1$ in the covariant derivative (1.7), the Lagrangian in Eq. (1.5) can be further implemented with an interaction and a potential term arising from the scalar field ϕ:

$$\mathcal{L} = \mathcal{L}_{\mathrm{EW}} + (D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi), \tag{1.10}$$

with the potential $V(\phi)$ given by:

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \tag{1.11}$$

For $\lambda > 0$ (since the potential has to be bound inferiorly) and $\mu^2 < 0$, the state of minimum energy does not correspond to $\phi = 0$ and is in fact not unique anymore. Choosing the vacuum state so that $\phi_1 = \phi_2 = \phi_4 = 0$ and $\phi_3 = v$, the system is no longer invariant under rotations in the (ϕ^+, ϕ^0) plane, and the void expectation value of ϕ results in:

$$\langle 0|\phi|0\rangle \equiv \langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + \mathrm{H} \end{pmatrix}, \qquad v^2 = -\frac{\mu^2}{\lambda}, \tag{1.12}$$

where the unitary gauge has been adopted to remove "unphysical" fields called Goldstone bosons. This choice of vacuum breaks the $SU(2)_L \otimes U(1)_Y$ symmetry to the electromagnetic subgroup $U(1)_{QED}$, which still remains a true symmetry of the vacuum, since $Q \langle \phi \rangle = (T_3 + Y/2) \langle \phi \rangle = 0$. This represents an example of spontaneous symmetry breaking, and, in the particular case of the electroweak interactions, it is commonly referred to as Brout-Englert-Higgs mechanism [10, 11], from the names of the physicists who first formulated it in 1964. In the kinetic piece $(D^\mu \phi)^\dagger (D_\mu \phi)$ of the Lagrangian (1.10), the scalar multiplet is coupled to the gauge bosons by the covariant derivative. After spontaneous symmetry breaking, such kinetic component takes the form:

$$(D^\mu \phi)^\dagger (D_\mu \phi) = \frac{1}{2} \partial_\mu \mathrm{H}\, \partial^\mu \mathrm{H} + (v + \mathrm{H})^2 \left(\frac{g^2}{4} \mathrm{W}^+_\mu \mathrm{W}^{-\mu} + \frac{g^2}{8 \cos^2 \theta_W} \mathrm{Z}_\mu \mathrm{Z}^\mu \right). \tag{1.13}$$

The vacuum expectation value chosen has generated quadratic terms for $\mathrm{W}^\pm$ and Z, namely those gauge bosons have acquired masses:

$$m_{\mathrm{W}} = m_{\mathrm{Z}} \cos \theta_W = \frac{1}{2} v g. \tag{1.14}$$

Since the $U(1)_{QED}$ symmetry of electromagnetism is conserved, the photon keeps being massless. Nevertheless, after spontaneous symmetry breaking, a physical scalar boson remains in the spectrum: the Higgs field (H), through which, similarly to bosons, fermions (quarks and leptons) acquire mass as well. One of the most remarkable successes of the Standard model was indeed the prediction of the existence of the Higgs boson, then discovered by the ATLAS and the CMS Collaborations at the CERN LHC in 2012 [12, 13].

1.2.3 The Cabibbo-Kobayashi-Maskawa matrix

Similarly to bosons, also quarks and leptons interact with the Higgs field and acquire mass thanks to spontaneous symmetry breaking, but they do so via a specific coupling, the *Yukawa coupling*. Nevertheless, the quark mass eigenstates do not correspond to pure weak interaction eigenstates, but are a mixture of the latter. The unitary transformation connecting the two bases of mass and weak eigenstates is represented by the Cabibbo-Kobayashi-Maskawa (CKM) matrix [14, 15]. By convention, the u, c, and t quarks are chosen to be pure states, and the flavor mixing is described in terms of a 3×3 matrix operating on the d, s and b quark states:

$$\begin{pmatrix} \mathrm{d}' \\ \mathrm{s}' \\ \mathrm{b}' \end{pmatrix} = \begin{bmatrix} V_{\mathrm{ud}} & V_{\mathrm{us}} & V_{\mathrm{ub}} \\ V_{\mathrm{cd}} & V_{\mathrm{cs}} & V_{\mathrm{cb}} \\ V_{\mathrm{td}} & V_{\mathrm{ts}} & V_{\mathrm{tb}} \end{bmatrix} \begin{pmatrix} \mathrm{d} \\ \mathrm{s} \\ \mathrm{b} \end{pmatrix}. \tag{1.15}$$

The weak interaction and the mass eigenstates are reported on the left and on the right side of the equation, respectively. The CKM matrix summarizes all the properties of the weak quark interaction. In the unitarity constraint on its diagonal terms

$$\sum_k |V_{ik}|^2 = \sum_i |V_{ik}|^2 \tag{1.16}$$

is codified the preservation of the weak coupling universality, a consequence of the fact that all $SU(2)$ doublets couple with the same strength to the vector bosons of weak interactions. At the same time, the CKM matrix describes the ranking among the transitions occurring within quarks of the same family and quarks belonging to different families. Furthermore, this mechanism provides an explanation for the non conservation of quark flavor under weak interactions mediated by $\mathrm{W}^{\pm}$ bosons, namely the weak charged currents, and incorporates the Glashow-Iliopulos-Maiani (GIM) mechanism [16], which explains why flavor-changing neutral currents mediated by a Z boson are suppressed. Considering the additional constraint of unitarity

$$\sum_k V_{ik} V_{jk}^* = 0, \tag{1.17}$$

it is possible to parametrize the CKM matrix in terms of three Euler angles (θ_{12}, θ_{23}, θ_{13}) and a complex CP-violating phase (δ_{13}). This imaginary part of the CKM matrix is the source of all the CP-violating phenomena accounted for in the Standard Model. The determination of the magnitude of the CKM matrix elements has been targeted by several experiments in the past 20 years, including the BaBar experiment at SLAC, the BELLE experiment at KEK, and the LHCb experiment at CERN [17].

1.2.4 Strong interactions

Unlike the other fermions in the Standard Model zoology of elementary particles, quarks also interact among each others via the strong force, carried by mediators called gluons. The quantum field theory of strong interactions is called quantum chromodynamics (QCD) and is based on a local gauge symmetry group $SU(3)_C$, with the subscript C indicating the color charge, which is the conserved quantity under local gauge transformations of this group. The non-Abelian structure of $SU(3)_C$, that is the non-commutativity of its transformations, gives rise to some peculiar properties which are specific of the strong interaction. The Lagrangian of QCD can be expressed

as:

$$\mathcal{L}_{\mathrm{QCD}} = \sum_q \bar{\psi}_{q,i} \left(i\gamma^\mu \partial_\mu \delta_{ij} - g_S \gamma^\mu t^C_{ij} A^C_\mu - m_q \delta_{ij} \right) \psi_{q,j} - \frac{1}{4} G^i_{\mu\nu} G^{\mu\nu}_i, \tag{1.18}$$

where $\psi_{q,i}$ are the Dirac spinors of the quark fields, with the indices q, i running over the quark flavor and color; t^C_{ij} (with $C = 1, \ldots, 8$) is a set of 3×3 Gell-Mann matrices, generators of the $SU(3)_C$ group, and A^C_μ are the associated 8 bosonic fields (the gluons); $G^i_{\mu\nu}$ is the color fields tensor; $g_S^2 = 4\pi\alpha_{\mathrm{S}}$, with α_{S} the strong coupling constant, whose nature is one of the peculiar aspects of QCD. α_{S}, which sets the strength of the interactions involving quarks and gluons, is actually a *running constant*, that is a function of the momentum Q transferred in a given process. A first order solution is:

$$\alpha_{\mathrm{S}}(Q^2) = \frac{4\pi}{\beta_0 \ln(Q^2/\Lambda^2_{\mathrm{QCD}})}, \tag{1.19}$$

where β_0 is the first term of a perturbative series in the $\beta(\alpha_{\mathrm{S}})$ function [18, 19], and $\Lambda_{\mathrm{QCD}} \simeq 200\,\mathrm{MeV}$ is a scale parameter that prevents soft divergencies. The value of α_{S} at the electroweak scale ($Q^2 \simeq m_{\mathrm{Z}}^2$) is of the order of 0.1, but increases at lower energy, unlike its electroweak partner. An important consequence of this regards the application of perturbation theory in QCD calculations, which is thus possible only for a high momentum transfer, in the so-called *hard scattering* processes. On the contrary, at low energies the bound between quarks becomes very strong and their dynamics cannot be described with perturbative approximations. In this latter scenario, the most precise results are obtained by applying *lattice QCD* calculations, based on the numerical evaluation of path integrals on a discretized Euclidean space-time. In the low Q^2 region, interactions between quarks are so strong that they inevitably condense into colorless $SU(3)_C$ singlets, or *hadrons*: their most common combinations are scalar $q^i \bar{q}^i$ states, known as *mesons*, or anti-symmetrical $\varepsilon^{ijk} q^i q'^j q''^k$ states, the *baryons*, such as the protons and the neutrons the atomic nuclei are made of. This phenomenon is referred to as *color confinement*, since it affects all the particles carrying color charge, and is responsible for the *hadronization*, the process of creation of hadrons out of quarks and gluons emitted in a hard scattering. All quarks but the top are subject to hadronization. The exceptional behavior of the heaviest quark is due the time scale of its weak decay ($\approx 0.5 \times 10^{-24}\,\mathrm{s}$), shorter than that at which the strong force acts ($1/\Lambda_{\mathrm{QCD}} \simeq 3 \times 10^{-24}\,\mathrm{s}$). Therefore, a top quark decays before it can hadronize. On the other hand, in the high-energy/short-distance regime, the coupling constant α_{S} becomes weaker and a consistent description is possible within the perturbation theory. This condition is known as *asymptotic freedom* [18].

1.3 W boson: from the discovery to recent measurements

Despite the fact that the formulation of the electroweak theory dates back to the late '60s, the discovery of its massive bosons had to wait until particle accelerators were powerful enough to produce them, almost fifteen years later. The first experiments ever to observe the evidence of the existence of the W boson were UA1 and UA2, in the CERN laboratories. UA1 and UA2, which shared the same main purpose with a different detector design, were placed at two interaction points of the 7-kilometers-long collider called Super Proton-Antiproton Synchrotron (Sp$\bar{\mathrm{p}}$S). This accelerator was the result of important modifications of the Super Proton Synchrotron (SPS), now part of the acceleration chain for protons and heavy ions culminating with the LHC. Following a farsighted idea by Carlo Rubbia, also spokesperson of the UA1 experiment,

the SPS was converted from a one-beam synchrotron to a two-beam collider, where a single vacuum chamber could be used to circulate protons and antiprotons in opposite directions, generating collisions at a center-of-mass energy of 540 GeV. The production of intense antiproton beams was achieved through the stochastic cooling technique, invented by Simon van der Meer a few years before [20]. The discovery of the W boson was announced during a seminar at CERN in January 1983, and published independently by the UA1 and UA2 Collaborations soon after [21, 22]. The discovery of the Z boson followed in some months time [23, 24]. Rubbia and van der Meer were awarded the Nobel Prize for Physics in 1984, for the ideas and the technological innovations that made such milestones possible.

1.3.1 The Jacobian peak and the first W-mass measurement

The search for the W boson, as it was performed by the UA1 and UA2 Collaborations, targeted the signature of the decay $\mathrm{W}^{\pm} \to \mathrm{e}^{\pm}\nu_{\mathrm{e}}(\bar{\nu}_{\mathrm{e}})$, characterized by:

- an isolated electron of large transverse momentum (p_{T}), that is the momentum component orthogonal to the beam axis;
- a peak in the electron p_{T} distribution at $m_{\mathrm{W}}/2$ (the *Jacobian peak*), with m_{W} the mass of the W boson;
- large missing transverse momentum ($p_{\mathrm{T}}^{\mathrm{miss}}$), due to the neutrino escaping detection.

A few clear electron-neutrino events were selected out of an integrated luminosity $\mathcal{L}_{\mathrm{int}} \simeq 18\,\mathrm{nb}^{-1}$. The mass of the W boson, whose most accurate Standard Model prediction was $m_{\mathrm{W}} = 82 \pm 2.4\,\mathrm{GeV}$, was reconstructed through the *Jacobian peak* procedure. The differential cross section for the process $\mathrm{p\bar{p}} \to \mathrm{W} \to \mathrm{e}\nu$ is a slow-varying function of the helicity angle θ^* between the charged lepton and the beam direction in the W boson rest frame:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\cos\theta^*} = A\left(\cos\theta^*\right), \tag{1.20}$$

and

$$\frac{\mathrm{d}\sigma}{\mathrm{d}p_{\mathrm{T}}^{\mathrm{e}}} = \frac{\mathrm{d}\sigma}{\mathrm{d}\cos\theta^*}\frac{\mathrm{d}\cos\theta^*}{\mathrm{d}p_{\mathrm{T}}^{\mathrm{e}}}. \tag{1.21}$$

The electron transverse momentum $p_{\mathrm{T}}^{\mathrm{e}}$ can be expressed as:

$$p_{\mathrm{T}}^{\mathrm{e}} = p^{\mathrm{e}}\sin\theta^* = \frac{m_{\mathrm{W}}}{2}\sin\theta^*. \tag{1.22}$$

Given the following expressions for the sinusoidal functions of θ^*:

$$\sin\theta^* = \frac{2p_{\mathrm{T}}}{m_{\mathrm{W}}} \quad \Rightarrow \quad \cos\theta^* = \sqrt{1-\sin^2\theta^*} = \sqrt{1-\left(\frac{2p_{\mathrm{T}}^{\mathrm{e}}}{m_{\mathrm{W}}}\right)^2}, \tag{1.23}$$

it is possible to write:

$$\frac{\mathrm{d}\cos\theta^*}{\mathrm{d}p_{\mathrm{T}}^{\mathrm{e}}} = \frac{4p_{\mathrm{T}}^{\mathrm{e}}/m_{\mathrm{W}}}{2\sqrt{1-\left(\frac{2p_{\mathrm{T}}^{\mathrm{e}}}{m_{\mathrm{W}}}\right)^2}} = \frac{2p_{\mathrm{T}}^{\mathrm{e}}}{m_{\mathrm{W}}\sqrt{1-\left(\frac{2p_{\mathrm{T}}^{\mathrm{e}}}{m_{\mathrm{W}}}\right)^2}}. \tag{1.24}$$

Finally, the expression for the differential cross section

$$\frac{\mathrm{d}\sigma}{\mathrm{d}p_{\mathrm{T}}^{\mathrm{e}}} = A\left(\cos\theta^*\right)\frac{\mathrm{d}\cos\theta^*}{\mathrm{d}p_{\mathrm{T}}^{\mathrm{e}}} \simeq K\frac{2p_{\mathrm{T}}^{\mathrm{e}}}{m_{\mathrm{W}}\sqrt{1-\left(\frac{2p_{\mathrm{T}}^{\mathrm{e}}}{m_{\mathrm{W}}}\right)^2}} \tag{1.25}$$

reveals the presence of the Jacobian peak in the $p_{\mathrm{T}}^{\mathrm{e}}$ distribution, as shown in Figure 1.2. After correcting for a non-null transverse momentum of the W boson itself, which can arise from the presence of other particles originating from the $\mathrm{p\bar{p}}$ collision (e.g., the associated production of a W boson and jets), the $p_{\mathrm{T}}^{\mathrm{e}}$ value that was found to maximize the peak represents an estimation of m_{W}:

$$m_{\mathrm{W}} = 81 \pm 5\,\mathrm{GeV} \tag{1.26}$$

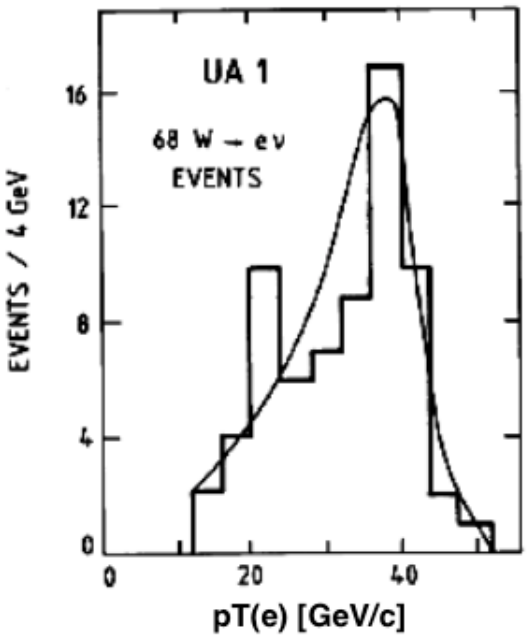

FIGURE 1.2: Distribution of the $p_{\mathrm{T}}^{\mathrm{e}}$ spectrum from candidate $\mathrm{W}^{\pm} \to \mathrm{e}^{\pm}\nu_{\mathrm{e}}(\bar{\nu}_{\mathrm{e}})$ events.

1.3.2 Recent W-boson mass measurements

Following the discovery, the properties of the W boson have been broadly investigated at various facilities, including the Tevatron proton-antiproton collider, the Large Electron-Positron collider (LEP), and, more recently, the LHC. Among these properties, the W-mass obviously holds a privileged position. In fact, a precise determination of m_{W} would positively affect the precision on other interdependent SM parameters, such as the mass of the top quark (m_{t}) and the Higgs boson (m_{H}). Such dependences are observable in Fig. 1.3, which shows a global fit to several SM parameters. Such fit may include or not the experimental constraints derived from direct measurements of m_{W}, m_{t} or m_{H}, thus leading to different allowed regions of the parameters space. In particular, the gray contour represents the phase space allowed by the theory without including any measurement of the three aforementioned parameters; the blue contour is obtained by using $m_{\mathrm{H}} = 125\,\mathrm{GeV}$; finally, the horizontal and vertical green bands arise from the measurement of the three masses and identify a region which is expected

to intersect with the blue contour. Although this happens within the uncertainties, a more precise determination of m_{W} might confirm or overturn this observation.

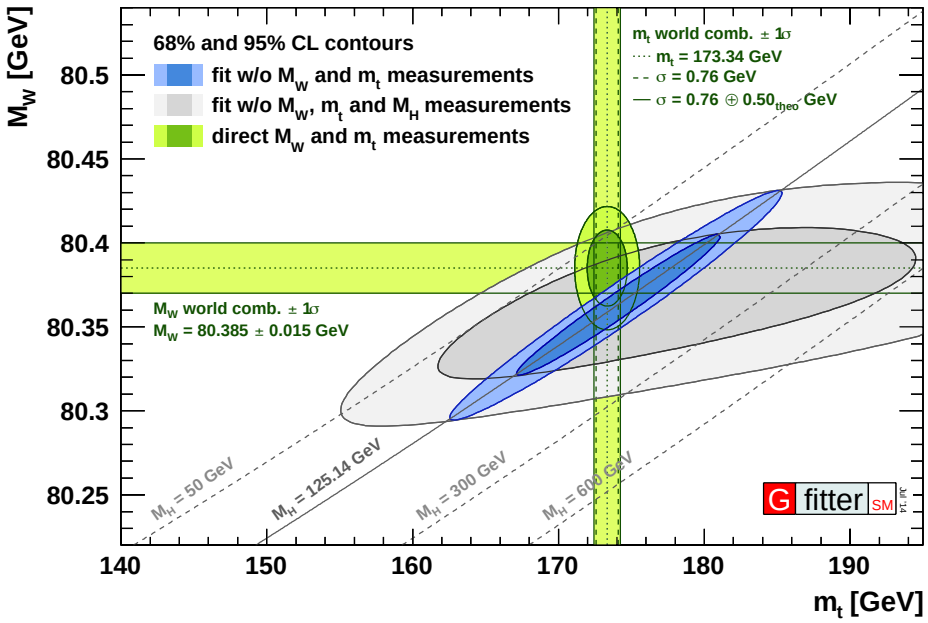

FIGURE 1.3: Contours of 68% and 95% confidence level obtained from scans of fits to electroweak SM parameters for fixed values of m_{W} and m_{t}. The blue and grey regions are the results of the fit including and excluding the m_{H} measurements, respectively. The green bands indicate the $1\,\sigma$ regions of the m_{W} and m_{t} measurements (world averages are used).

In the case of hadron colliders, which will be treated more in depth in the next section, the production of on-shell W bosons is generally tagged by the charged lepton (muon or electron) of large p_{T} arising from their decay. Given the unknown parton-parton effective energy in each hadron collision and the presence of missing energy in the longitudinal direction, the strategy of hadron collider experiments (such as CDF and D0 at Tevatron, and ATLAS at the LHC) consists in reconstructing the transverse mass of the W boson

$$m_{\mathrm{T}} = \sqrt{2\, p_{\mathrm{T}}^{\ell}\, p_{\mathrm{T}}^{\mathrm{miss}} (1 - \cos\Delta\phi)}, \tag{1.27}$$

and derive m_{W} from comparing the transverse mass distribution with MC predictions as a function of m_{W}. In the expression above, p_{T}^{ℓ} denotes the charged lepton (e, μ) transverse momentum, whereas $\Delta\phi$ is the azimuthal opening angle between the charged lepton and the missing transverse momentum. The measurements of the W-mass carried out with such strategy are at present the most forefront. In particular, both the CDF and the ATLAS Collaborations have reached comparable precision [25, 26], as can be observed in Fig. 1.4.

In the case of LEP, the nature of lepton collisions permitted a precise knowledge of the beam energy, which allowed the experiments to determine the $\mathrm{e^+e^-} \to \mathrm{W^+W^-}$ cross section as a function of the center-of-mass energy, as well as to reconstruct the W mass precisely from its decay products. Close to the $\mathrm{W^+W^-}$ production threshold ($\approx$161 GeV), the dependence of the W-pair production cross section on m_{W} is large, and thus this characteristic was exploited to determine m_{W}. At higher energies, this dependence is much weaker. Therefore, WW boson pairs were reconstructed through their fully hadronic, semi-leptonic, and fully leptonic decays, and m_{W} was determined as the invariant mass of the W boson decay products. The results obtained

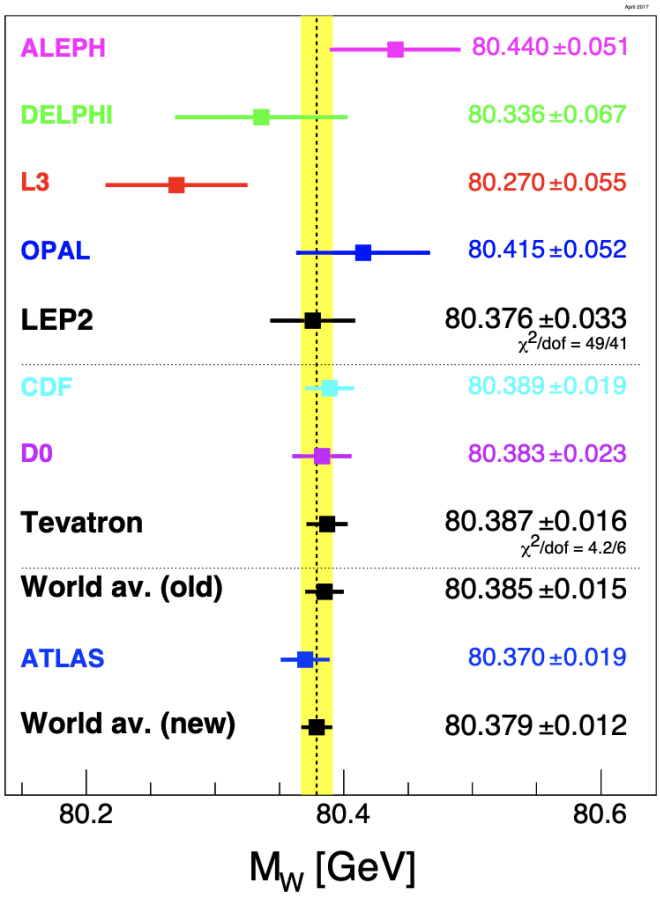

FIGURE 1.4: Measurements of the mass of the W boson (m_{W}) by Tevatron, LEP, and LHC experiments [17]. The world average of these measurements is also shown, before and after the most recent m_{W} measurement performed by the ATLAS Collaboration.

at LEP are comparable and generally in good agreement with those of hadron collider experiments, providing a valuable cross-validation.

1.4 The W boson at the LHC

As anticipated in the previous section, since its discovery at the Sp$\bar{\mathrm{p}}$S proton-antiproton accelerator, the properties of the W boson have been studied at both $\mathrm{e^+e^-}$ (LEP) and hadron (Tevatron, LHC) colliders. The initial state is reflected into the production mechanisms of the W boson, and may also originate different underlying conditions which can make one kind of collider more or less suitable for a specific measurement. The search this thesis reports on exploits collision data collected at the LHC, a machine which accelerates and circulates protons. Therefore, the next section is intended to provide a brief description of the phenomenology of hadron collisions. Afterwards, the main properties of the W boson will be presented.

1.4.1 Hadron collisions

As discussed previously in this chapter, protons are not elementary particles, but composite objects constituted of three valence quarks (one down and two up quarks) surrounded by the gluons they emit and by quark-antiquark pairs from gluon splitting.

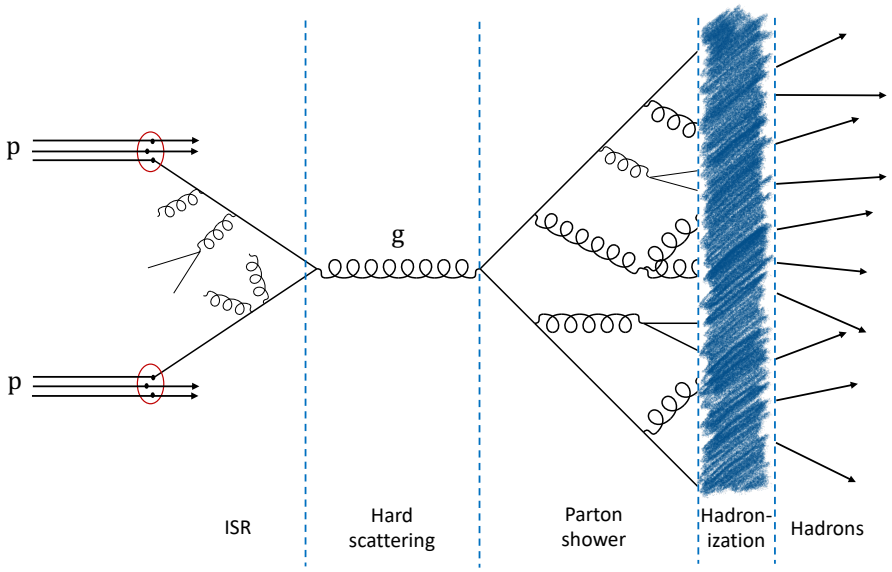

FIGURE 1.5: Scheme of the main phases of parton interactions.

Thus, a collision between two protons is in fact an interaction which involves their fundamental constituents, generically referred to as *partons*. Despite the complexity of the phenomenology of such process, the theory of QCD may be used to describe its main phases. Following the scheme in Fig. 1.5, we can identify:

- possible emission of *initial state radiation* (ISR) from the incoming partons before the actual interaction takes place, thus reducing the beam energy prior to the momentum transfer;
- *hard scattering*: the main parton-parton interaction, which generates other particles;
- *parton shower*: in case colored particles are produced in the ISR phase or the hard scattering, they tend to generate a shower of other partons;
- *hadronization*: as the shower evolves and the momentum transfer gets lower, partons inevitably start to recombine into color-singlet states through gluon radiation, thus forming hadrons.

The cascades of collimated particles generated at hadronization stage, mainly formed by hadrons but also by leptons and photons, are observed as *jets* by particle detectors. The definition of a jet in a physics experiment is not unique, but depends on the clustering algorithm used to group the jet constituents according to their kinematic properties. Similarly, the description of this process in Monte Carlo (MC) generators is achieved through different effective models.

Occasionally, two hard parton-parton interactions may occur within a single proton-proton (pp) collision. These are commonly referred to as double-parton scattering events. Nevertheless, the hard scattering is normally accompanied by additional softer activity originated from the interaction of the other (less energetic) partons in the same protons. The result is the generation of extra parton showers and hadronization processes. This activity is known as *underlying event*.

The compositeness of protons colliding at given center-of-mass energy $\sqrt{s}$ is reflected on the fact that the energy $\sqrt{s'}$ available for the hard scattering is necessarily

a fraction of $\sqrt{s}$:

$$\sqrt{s'} = \sqrt{(x_1 p_1 + x_2 p_2)^2} \approx \sqrt{x_1 x_2 s}, \tag{1.28}$$

where $p_1,\ p_2$ are the four-momenta of the two protons and $x_1,\ x_2$ the fractions of these four-momenta carried by the hard-scattering partons. The aleatoric nature of x_1 and x_2 is expressed in terms of probability density functions known as parton distribution functions (PDFs), which participate in the definition of any particle production cross section at hadron colliders. Therefore, a precise determination of the PDFs is the target of several measurements at the LHC.

Moreover, the protons in the LHC vacuum pipes circulate in bunches containing a large number of particles each (order of 10^{11}). The needle-like shape of the proton cloud composing each bunch, with its transverse (longitudinal) size of the order of a few μm (cm), warrants a certain probability that multiple pp interactions occur within a single bunch crossing. Usually, only one of these interactions generates an interesting physics event, with large momentum particles. In high-energy physics experiments, the corresponding hard scattering point is commonly referred to as the *primary vertex*. The other concurrent interactions are defined *pileup* (PU), and produce low-momentum particles which overlap with those from the primary vertex, degrading the reconstruction performance for the event of interest. It is therefore paramount for the detectors to be able to precisely identify the primary vertex of interaction, and for physicists to implement effective methods for PU rejection or subtraction.

1.4.2 Production of the W boson

At the LHC, W bosons are primarily produced, alone or in association with one or more quarks or gluons, from the interaction of a quark-antiquark pair. Given the quark composition of colliding protons (uud), the most common scenarios are the production of a W^+ from a u valence-quark and a $\bar{d}$, or of a W^- from a d valence-quark and a $\bar{u}$, where the $\bar{q}$ quarks need to be pulled from the proton sea with enough energy to generate a W. This introduces an asymmetry in the total number of W^+ and W^- produced at the LHC, with the positive charge sign being favored. Moreover, the W boson production can be initiated by the quarks of the second family, charm and strange; the fraction of production from these heavy quarks up to 20% at a center-of-mass energy of 13 TeV. On the other hand, the production from across-family quarks is suppressed by the small size of the off-diagonal elements of the CKM matrix, described in Section 1.2.3.

Another relevant W boson creation mechanism exists at the LHC. The Large Hadron Collider is often regarded as a "top factory", because of the large cross section for single top quark and, in particular, for top-antitop pair production. In pp collisions at $\sqrt{s} = 13$ TeV, the $t\bar{t}$ production mainly proceeds through gluon-gluon fusion, which dominates over $q\bar{q}$ annihilation (see Feynman diagrams in Fig. 1.6). The theoretical calculations for its cross section, exceeding 800 pb, are now performed at next-to-next-to-leading order (NNLO) in QCD [27, 28], and are supported by increasingly precise measurements by the LHC experiments [29]. Given the aforementioned cross section and considering an average instantaneous luminosity $\mathcal{L} \simeq 2 \times 10^{34}\,\mathrm{cm^{-2}s^{-1}}$, it can be inferred that roughly 15 $t\bar{t}$ pairs are produced at the LHC every second. Afterwards, top (anti)quarks decay with a probability close to 100% into a W boson and a bottom (anti)quark. Despite the cross section for this production mode being more than two orders of magnitude smaller than the direct W boson production (which amounts to about 105000 pb [30]), this is an interesting process for the analysis reported in this thesis.

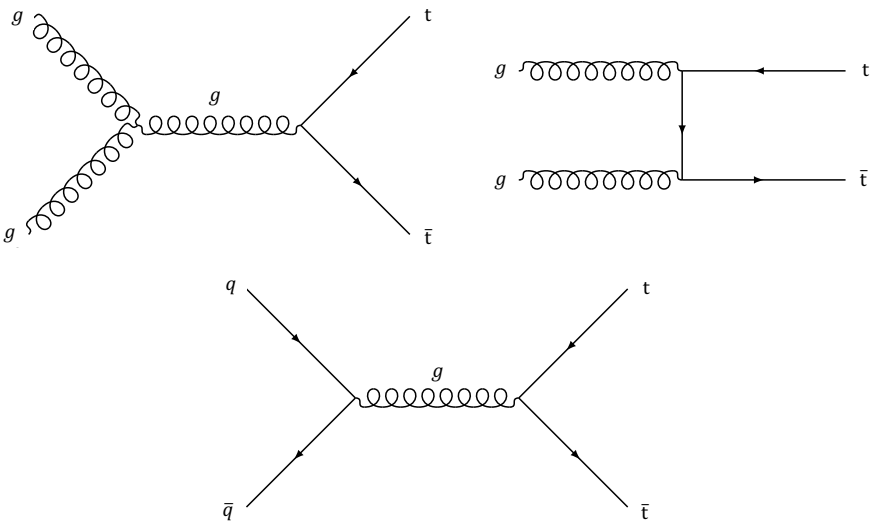

FIGURE 1.6: Tree-level Feynman diagrams for $t\bar{t}$ production at the LHC. The two upper diagrams describe the contribution through the gluon-gluon fusion mechanism, while the lower diagram describes $q\bar{q}$ annihilation.

Further production modes of the W boson are accessible at the LHC, but their cross sections can be considered negligible with respect to the other ones examined in this section.

1.4.3 Decay of the W boson

With a lifetime of the order of 10^{-25} s, the W boson cannot be directly detected with the current technology. Therefore, past and current experiments have exploited either *fully leptonic* or *fully hadronic* decay modes to realize measurements of its properties. Despite the great efforts accomplished so far, our knowledge of the W boson remains poor if compared to that of the mediator of neutral currents in weak interactions, the Z boson. The leptonic and the fully hadronic decays of the Z boson are indeed known with a precision over an oder of magnitude (i.e., at the per mille level) better than those of the W boson, mainly reflecting the fact that a great amount of Z bosons were produced at the LEP e^+e^- collider, in an experimentally clean environment. Similarly, the uncertainty in the measurement of the mass of the charged vector boson (m_W) is currently of the order of 10 MeV, while m_Z is known with an uncertainty of few MeV. The search for rare Z decays, then, is much richer, with over 50 different measurements of semi-exclusive hadronic final states (e.g., $Z \to J/\psi X$ [17]), as well as *upper limits* on exclusive hadronic final states (e.g., $Z \to \pi^0\gamma$ [31]) and on lepton flavor violating decays (e.g., $Z \to e^{\pm}\mu^{\mp}$ [32]). Comparatively, a very small number of rare W decays has been investigated so far. Thus, it is evident why, almost 40 years after its discovery, the W boson and its decays remain an important subject of study.

Fully leptonic decays

The decay amplitude for the process

$$W \to \ell\nu_\ell \qquad \ell = e, \mu, \tau \tag{1.29}$$

constitutes ≈32% of the total decay amplitude of the W boson [17], a concept which is commonly expressed by stating that the *branching ratio* or *branching fraction*

$\mathcal{B}(\mathrm{W} \to \ell\nu_\ell) \simeq 0.32$. Generally, the charged lepton, if electron or muon, can be easily detected, and its momentum can be measured with high accuracy and resolution. The neutrino, though, escapes detection in common experimental apparati. For this reason, if a measurement of the m_W has to be performed, it cannot be based on the lepton-neutrino invariant mass reconstruction; the energy of the escaping neutrino must be inferred instead, exploiting the approximately null initial transverse momentum for a collision in a symmetrical accelerator. It is worth underlining that such approximation is very close to the truth also at hadron colliders, where the interacting particles are composite. Nonetheless, each pp interaction might produce more than one (anti)neutrino, and the detection efficiency is in general < 1 for any particle and detector, thus reducing the precision obtainable with the $p_\mathrm{T}^\mathrm{miss}$ method.

Fully hadronic decays

The process

$$\mathrm{W} \to q\bar{q'} \to hadrons \tag{1.30}$$

represents the most likely decay of the W boson, with a branching ratio of ≈67% [17]. Even though, in principle, the invariant mass of the two jets arising from the $q\bar{q}$ pair can be fully reconstructed, the typical jet momentum resolution at collider experiments is around 10%, much larger than that of charged leptons. In addition, hadronic jets may be an actual source of $p_\mathrm{T}^\mathrm{miss}$, due to the presence of neutrinos from certain hadron decays. Therefore, most of the precision measurements of m_W and other properties of the W boson performed so far have exploited the fully leptonic decay mode.

Rare decays

Once accounted for the fully hadronic and the fully leptonic decays, which add up to a branching fraction of about 99%, the Standard Model foresees a few other options for the W boson to decay, though with a very low probability. These are known as *exclusive hadronic* decays, with the term "exclusive" indicating a final state that contains one specific hadron, and not a cascade generated by the hadronization of free quarks. According to the theory, low-multiplicity hadronic decays of the gauge and the Higgs bosons are subject to strong suppression mechanisms (the so-called "Sudakov effects"), and calculations of their branching fractions, if ever performed, have rather large uncertainties and may span various orders of magnitude. At present, none of these decays has been observed by experiments. Nevertheless, for what concerns the W boson, upper limits at 95% confidence level (CL) were set on the branching fractions $\mathcal{B}(\mathrm{W}^\pm \to \pi^\pm\gamma) < 7.0 \times 10^{-6}$ [33], $\mathcal{B}(\mathrm{W}^\pm \to \mathrm{D_s}^\pm\gamma) < 1.3 \times 10^{-3}$ [34], and $\mathcal{B}(\mathrm{W}^\pm \to \pi^\pm\pi^\pm\pi^\mp) < 1.01 \times 10^{-6}$ [35]. Besides the obvious interest of determining the aforementioned rare decay branching ratios, the observation of these processes could on the one hand represent a further probe of the Standard Model, and on the other offer a new way of measuring essential properties of the W boson, as will be discussed in the next sections.

1.5 The rare decay of the W boson into a pion and a photon

The exclusive radiative decays $\mathrm{W}^\pm \to \mathrm{P}^\pm\gamma$, where P denotes a generic pseudoscalar meson, were studied in a few theoretical works since before the discovery of the W

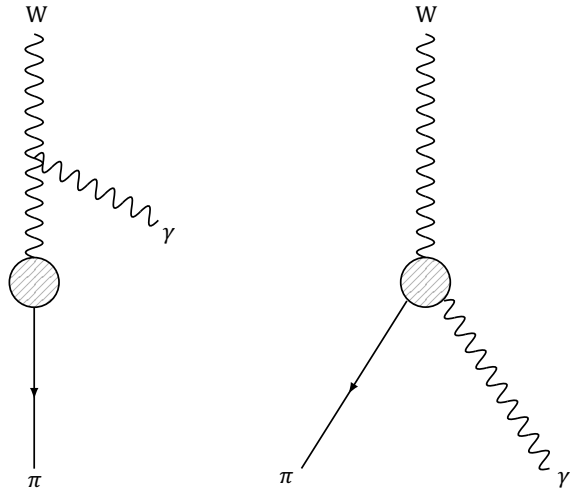

FIGURE 1.7: The two diagrams contributing to the decay $\mathrm{W} \to \pi\gamma$.

boson. Therefore, we know that the process $\mathrm{W}^{\pm} \to \pi^{\pm}\gamma$ receives two main kinds of contribution, as illustrated in Fig. 1.7. In one of them, which corresponds to the left diagram, the photon couples directly to the W boson. According to a relatively recent paper by M. Mangano and T. Melia [36], its rate can be related to the pion decay constant, $f_\pi = 93\,\mathrm{MeV}$, by evaluating the current

$$\left\langle \pi^+(p) \left| J^\rho_{\mathrm{W}}(0) \right| 0 \right\rangle = \frac{f_\pi}{\sqrt{2}}\, p^\rho, \tag{1.31}$$

where p indicates the momentum of the pion state $|\pi^+(p)\rangle$, and $J^\rho_{\mathrm{W}} = \bar{\mathrm{d}}\gamma^\rho P_{\mathrm{L}}\mathrm{u}$ the weak charged current, with $P_{\mathrm{L}} = \frac{1}{2}(1-\gamma_5)$ and $(\bar{\mathrm{d}}, \mathrm{u})$ the anti-down and up quarks. On the other hand, contributions of the type shown in the right diagram involve the calculation of

$$\int d^4x e^{ik\cdot x} \langle \pi^+(p)|T[J^\lambda_W(0) J^\mu_\gamma(x)]|0\rangle, \tag{1.32}$$

where k is the photon momentum and $J^\mu_\gamma = \sum_{i=\mathrm{u,d}} Q_i \bar{q}_i \gamma^\mu q_i$ is the electromagnetic current, with Q_i the charge of the quark q_i. This can be evaluated by adapting existing calculations [37] of the decay width of $\mathrm{Z} \to \mathrm{W}^{\pm}\pi^{\pm}$, and subsequently estimating the decay $\mathrm{Z} \to \pi^0\gamma$ using an operator product expansion at leading order (LO) in α_{S}. The order of magnitude estimate obtained for the branching ratio of $\mathrm{W} \to \pi\gamma$ is 10^{-9}, even though the expansion is not convergent unless important higher order corrections (not yet calculated) are introduced.

Mangano and Melia's estimate is in agreement with a previous result by L. Arnellos, W. Marciano, and Z. Parsa [38], who adopted a different approach. On the contrary, using a one-loop calculation, Y. Kneum and X. Pham [39] hypothesize an enhancement of the rare radiative decays in exam, leading to $10^{-8} < \mathcal{B}(\mathrm{W}^{\pm} \to \pi^{\pm}\gamma) < 10^{-6}$. It has to be noted, though, that in this latter case the upper bound 10^{-6} follows a specific choice for the value of the quark masses and the momentum running in the loop. If such momentum is truly at the scale of m_{W}, then some of the assumptions of the authors do not hold, and the estimate of 10^{-6} becomes too optimistic.

Since no experiment has ever been able to observe the rare decay of $\mathrm{W} \to \pi\gamma$, we are for the moment only aware that its probability lies in between 10^{-9} and 10^{-6}, probably shifted towards the lower bound. The intent of the next section is thus to discuss how a process as rare as that described above could be observed at the LHC.

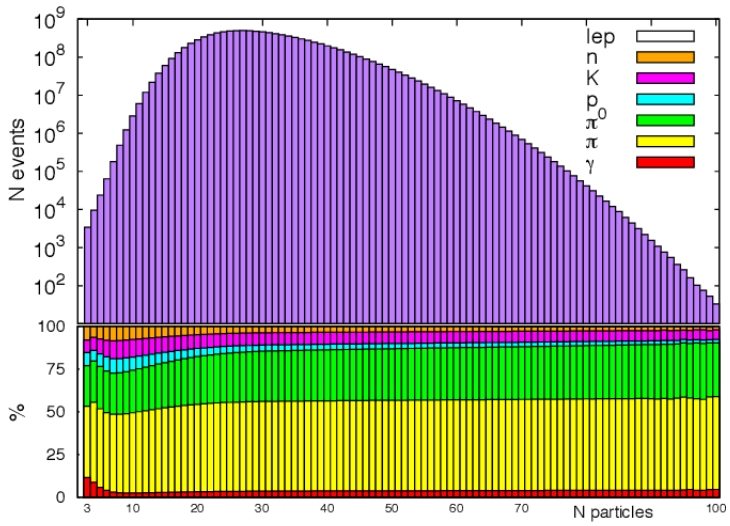

FIGURE 1.8: Event distribution as a function of the number of final state particles in 10^{10} events of $W^+ \to u\bar{d}$ decay produced with PYTHIA 8.1. Top panel: total number of events. Bottom panel: average fraction of particle type, in decreasing order (note that leptons are not visible).

1.5.1 Analysis strategy

The huge number of W bosons produced at the LHC makes this machine suitable to investigate rare W-decay modes. Nevertheless, the search is complicated by the large background of QCD processes, giving rise to a vast multiplicity of particles observed by the detector within every pp collision, including leptons, photons, and charged and neutral hadrons. This is illustrated in Fig. 1.8, which is the result of a study [36] performed by simulating about 10^{10} events of $W^+ \to u\bar{d}$ decay with PYTHIA 8.1 [40], a MC generator that handles the showering and the hadronization steps as well. Even in this rather intricate scenario, the study of a few rare decays leading to low-multiplicity final states seems plausible. Among them, the decays $W^\pm \to \pi^+\pi^+\pi^-$ and $W^\pm \to \pi^\pm\gamma$ appear to be the most promising. In particular, the charged pion and the photon from the latter process show larger momenta than the pions from the three-body decay, which also suffers from a strong combinatorial background in pp collisions. On the other hand, at a center-of-mass energy of 13 TeV, a substantial fraction of the many particles produced in a collision carries a large momentum, forcing the detectors to set high thresholds on the minimum p_T that triggers the data acquisition, in order to not saturate the trigger system. For instance, at the CMS experiments at the LHC and during the period 2016–2018, the lowest trigger threshold for single photons was about 170 GeV, way above the typical p_T expected for a photon from the rare W boson decay, which is on average around 40 GeV.

A work around these issues was proposed in the already mentioned study by M. Mangano and T. Melia [36], which has been a source of inspiration for the analysis presented in this thesis. The authors suggest that the abundance of $t\bar{t}$ events at the LHC, as described in Section 1.4.2, could be exploited to obtain a clear signature of W boson production. As the top (anti)quarks decay, two W bosons are available; the search strategy can then be articulated in three main items, as Fig 1.9 also illustrates:

- isolate the events in which one of the W bosons decays into leptons. As previously discussed, such decay has a branching fraction $\mathcal{B}(W \to \ell\nu_\ell) \simeq 0.1$ for each of the leptons ($\ell = e, \mu, \tau$). From an experimental point of view, the detection of electrons and muons is generally much simpler than that of tau leptons, given

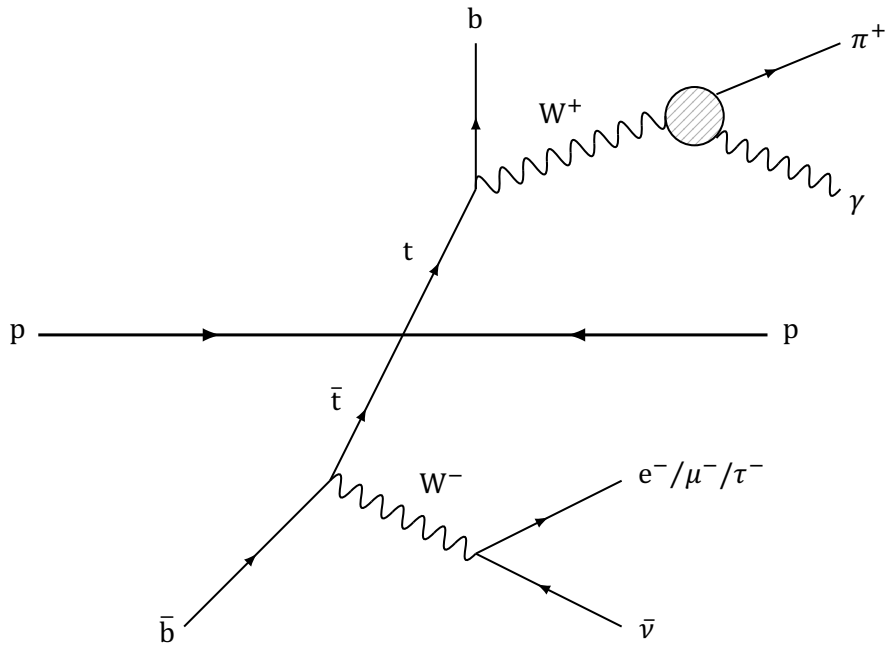

FIGURE 1.9: Pictorial view of the typical topology for the decay of $W \to \pi\gamma$ in $t\bar{t}$ events. The charge signs may of course be flipped.

the short lifetime of the latter. Nevertheless, the process $W \to \tau\nu_\tau$ could in principle be considered too. In the context of this analysis, it is accounted for only in case the τ subsequently decays into leptons, namely $\tau \to e(\mu)\nu_{e(\mu)}\nu_\tau$;

- identify the *b jets*, namely the jets originating from the hadronization of the b quarks from the (anti)top quark decay. Doing so helps suppressing the background arising from the hadronization of light quarks and gluons;
- use the other W boson in the event to search for the topology characteristic of the rare decay $W \to \pi\gamma$.

The latter partially depends on the detectors used to perform the search; at CMS, the topology of the rare decay $W \to \pi\gamma$ is characterized by an isolated *track* and an isolated photon of large transverse momentum.

Since the selection of $t\bar{t}$ events reduces the cross section for W boson production by two orders of magnitude with respect to the direct production from pp interaction, the success of this search is determined first of all by the size of the collision data available, but also by the detector performance and efficiency, and by the ability to suppress the background effectively with thoughtful discrimination procedures. In the period 2016–2018 (commonly referred to as *Run 2* by LHC experiments), the LHC has delivered an unprecedented integrated luminosity $\mathcal{L}_{\text{int}} \simeq 137\,\text{fb}^{-1}$, which allows for precision measurements of Standard Model parameters and searches for rare phenomena as never before. At the same time, the CMS experiment at the LHC is able to identify muons and electrons with an efficiency and a purity close to 100%, and to determine the energy of photons in the typical momentum range for this rare decay with a resolution below 1%. Therefore, it can be regarded as the perfect candidate in the search for the $W \to \pi\gamma$ decay.

1.5.2 The physical interest of $W \to \pi\gamma$

The low-multiplicity exclusive hadronic decay $W \to \pi\gamma$ represents a very interesting probe of strong interaction dynamics, at the boundary between the perturbative and nonperturbative domains of QCD. From a theoretical perspective, the rare decay amplitude may be calculated using the *QCD factorization* approach [41, 42], with the addition of soft form factor contributions to the hard-scattering contributions described in the factorization framework. If at leading order such decay amplitude can be expanded in powers of Λ_{QCD}/m_W and calculated in a relatively easy way, already the first-order power corrections involve ill-defined overlap integrals. This introduces poorly known model parameters and makes phenomenological predictions less precise. In this sense, the observation of the decay $W \to \pi\gamma$ would provide insights into factorization and meson form factors at large energy scales.

In addition, a number of recent theoretical works [43–48] has proposed that exclusive hadronic decays of the Higgs boson, of the type $H \to VM$ (with $V = W, Z$ and M a generic meson), could be used for a precise determination of the on- and off-diagonal couplings of the Higgs to quarks. Thus, the observation of an exclusive hadronic decay of the W boson would represent a proof of principle that these very challenging measurements are pursuable at the LHC.

Finally, the decay $W \to \pi\gamma$ could provide a new way to measure the mass of the W boson that is based solely on visible decay products. Although a precision measurement of this fundamental parameter of the Standard Model that exploits the rare decay must be excluded at the LHC, future hadron colliders, such as the High Luminosity LHC (HL-LHC), might be able to provide the sufficient amount of collision data to do so. Therefore, depending on its actual branching fraction, this decay channel could offer a novel handle in the effort to constrain m_W to a precision of the order of the MeV.

1.5.3 State of the art

At present, the most forefront measurement of the branching fraction of the decay $W^\pm \to \pi^\pm\gamma$ is due to the CDF Collaboration at Tevatron [33]. This section is intended to present an overview of their analysis strategy and the results obtained. The main differences and affinities with respect to the search presented in this thesis, which exploits data collected with the CMS experiment, will be highlighted. For a full description of the CMS detector, the reader should refer to Chapter 2.

Exploiting $p\bar{p}$ collisions at $\sqrt{s} = 1.96\,\text{TeV}$, the CDF Collaboration set the 95% CL upper limit on the ratio of the decay amplitudes (Γ):

$$\Gamma(W^\pm \to \pi^\pm\gamma)/\Gamma(W^\pm \to e^\pm\nu) < 6.4 \times 10^{-5}, \tag{1.33}$$

based on an integrated luminosity $\mathcal{L}_{int} = 4.3\,\text{fb}^{-1}$. Using the world average for $\mathcal{B}(W^\pm \to e^\pm\nu)$, this result translates into an upper limit on the rare decay branching fraction

$$\mathcal{B}(W^\pm \to \pi^\pm\gamma) < 7.0 \times 10^{-6}. \tag{1.34}$$

The strategy adopted by CDF was to analyze events containing a pion and a photon candidates, collected using an inclusive photon trigger with no tracking requirement. Thanks to this latter characteristic, $W^\pm \to e^\pm\nu$ events were also collected with the same trigger and used to renormalize the result of the search, allowing for the cancelation of many common systematic uncertainties in the measurement of the two decay amplitudes.

The offline reconstruction and identification of electron, charged pion, and photon candidates exploited information from the central tracker and the electromagnetic and hadronic calorimeters. Similarly to the CMS case, the CDF experiment did not comprise a detector specifically devoted to particle identification. Therefore, a charged pion candidate was defined as a track consistent with originating from the primary vertex of the $p\bar{p}$ interaction, pointing to a narrow calorimeter cluster, incompatible with that of an electron.

An important characteristic of both the particles emerging from the rare decay of the W boson in exam consists in the reduced jet activity expected in their vicinity, due to their electroweak production process. At detector level, this is observed as the presence of isolated tracks and calorimetric clusters. Thus, specific techniques aiming to quantify the pion and photon *isolation* were developed by the Collaboration, in a similar fashion to what is done in the analysis presented in this thesis (see for instance Sections 2.3.2 and 4.2.6). In particular, these are based on the computation of the energy deposited in the calorimeters in a solid angle cone surroundings each candidate, and of the sum of the transverse momenta of the particles measured in the tracker within the same cone.

The signal from the $W^{\pm} \to \pi^{\pm}\gamma$ decay, if present, was expected to appear as a peak in the pion-photon invariant mass spectrum, centered at the W boson mass and with a resolution of about 2.5 GeV, which includes the full width of the W boson and the experimental resolution of the CDF detector. The invariant mass distribution was therefore reconstructed from the 1398 $\pi^{\pm}\gamma$ candidates surviving the event selection, and a signal region was identified between 75 and 85 GeV (Fig. 1.10).

After determining the background contribution to the signal region through a fit to the sidebands of the same spectrum, a total of 219 ± 10 events were expected in the signal region, and 206 were observed. Since the data were consistent with the expected background, the signal region was further split into four 2.5 GeV bins. Upper limits were calculated for each bin and then combined, obtaining a gain in

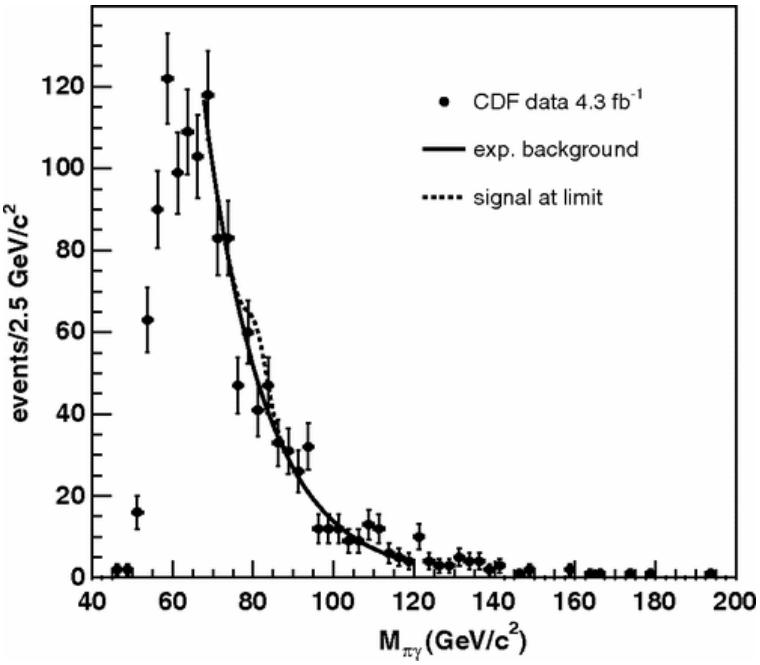

FIGURE 1.10: Event distribution as a function of the pion-photon invariant mass. The background expectation for the 1398 events passing the full selection and the signal expectation at the 95% CL upper limit are also shown. The uncertainties are purely statistical.

sensitivity by using information about the shape of the expected $W^{\pm} \to \pi^{\pm}\gamma$ mass peak. The main uncertainties affecting the final result were of statistical nature, with small systematic contributions mainly arising from the calculation of the product of efficiency and acceptance for the $\pi^{\pm}\gamma$ selection.

Under some aspects, the analysis presented in this thesis shows some similarities with the search for the decay $W \to \pi\gamma$, especially in the offline reconstruction of the particles involved in the decay. Nevertheless, the overall strategy pursued by the CDF Collaboration is unfortunately hardly repeatable at CMS, where the lowest trigger thresholds for single photons are way above the typical p_T expected for the products of the rare decay in exam. Therefore, my work focuses on W bosons produced in $t\bar{t}$ decays, as anticipated in the previous sections of this chapter. As will be discussed in the next chapters, this provides a rather clear signature, but implies a strong reduction of the signal cross section with respect to the direct production of W bosons from pp (or $p\bar{p}$) interactions. The different definition of the signal process results in diverse background contributions as well. The dominant backgrounds in the search by CDF came from photon+jet events, where the jet fragmented into a single charged particle, and from QCD multijet events, where one of the jets fragmented into a single charged particle and another was misidentified as a photon. Drell–Yan pair production and W/Z decays, especially to τ leptons, also contributed to the background at a level of $\approx$10%. If Drell–Yan processes constitute a background source for my search too, the characteristic topology of $t\bar{t}$ decays and the requirement of one charged lepton with large p_T per event determine that the main background contributions arise from any non-signal $t\bar{t}$ production and from the associated production of a vector boson and a photon, as will be discussed in Section 3.2.2 and throughout the following chapters.

Moreover, it has to be noted that W bosons from top quark decay are produced in specific polarization states, whereas the W boson candidates reconstructed at CDF were unpolarized. In my analysis, this characteristic is accounted for and yields to the evaluation of a systematic uncertainty in the signal efficiency (see Sections 6.2.3 and 6.2.3).

In summary, the final result of the search presented in this thesis is the first upper limit on the branching ratio of $W \to \pi\gamma$ measured at the LHC. Even though it is not yet as stringent as the limit obtained at CDF, it demonstrates the possibility of performing a search for such rare hadronic decays of the W boson at the LHC and defines a suitable search strategy.

Chapter 2

The CMS experiment at the LHC

Despite the remarkable accuracy in the description of elementary particle interactions, the Standard Model as we know it today does not provide an answer to some outstanding questions. Among these, the mysterious nature of Dark Matter, the asymmetry between matter and antimatter observed in our universe, the evidence that neutrinos have a finite mass, or the possible existence of additional generations of quarks and leptons. The Large Hadron Collider was conceived to be an instrument for the scientific community to address these and other questions and, at the same time, to open new scenarios and interrogatives. Therefore, the LHC physics program goes far beyond the Higgs discovery. Several areas of high energy physics are being and will be explored, while fundamental Standard Model parameters are measured with increased precision. This unprecedented chance to deepen our understanding of Nature makes the LHC and its experiments among the most fruitful research facilities in the world. A brief overview of the LHC machine is presented in the first section of this chapter, followed by a more detailed description of the CMS experiment and its subsystems.

2.1 The Large Hadron Collider

The LHC [49], which started its operation in 2009, is a machine capable of accelerating and colliding protons and heavy ions. The center-of-mass energy reached in pp collisions increased since its activation, and was kept stable at 13 TeV during Run 2 (2016–2018). The instantaneous luminosity above $2 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ is one of the largest ever achieved, and the largest for a hadron collider. Thanks to the high production rate of gauge bosons and heavy quarks quarks at LHC, the considerable amount of statistics available allows for many precision measurements, in spite of the large hadronic background. The LHC is also capable of accelerating heavy ions (lead, xenon, etc.), whose collisions are used to produce the Quark Gluon Plasma state of matter, recreating the conditions expected in the primordial universe, just after the Big Bang, and looking for hints on the nature of quark confinement.

The LHC is an unprecedented project, not only for its energy and luminosity, but also in terms of cost, complexity of the detectors, and human effort.

2.1.1 The accelerator

The LHC (Fig. 2.1) is a 27 km-long synchrotron installed at an average depth of 100 m underneath the ground level, in the tunnel previously built to host the LEP e^+e^- accelerator. A magnetic field of $\approx$8.3 T is necessary to keep protons and ions at energies of $6.5 - 7$ TeV on the LHC orbit, and is delivered by 1232 super-conducting dipole magnets. From a simple calculation, it is possible to show that the magnetic

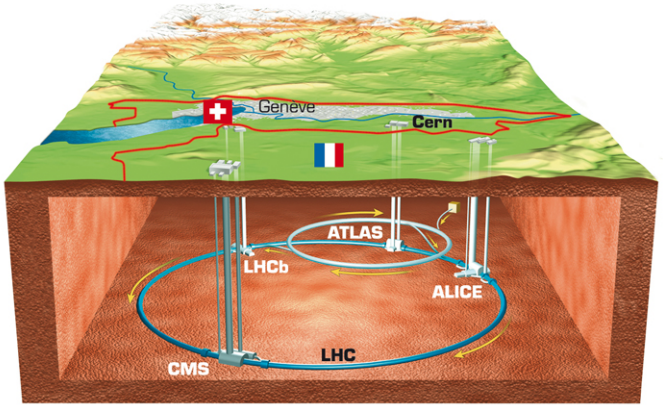

FIGURE 2.1: Geographical position of the LHC and its main experiments.

field required to keep a 6.5 TeV proton on the orbit is:

$$B(\mathrm{T}) = \frac{p(\,\mathrm{GeV})}{0.3\,R(\mathrm{m})} \simeq 5\,\mathrm{T}, \tag{2.1}$$

where p is the proton momentum and R is the LHC radius, namely $\approx$4300 m. Nevertheless, the dipole magnets do not cover the entire LHC circumference, so the magnetic field needs to be greater. The Nb–Ti magnets are kept to a temperature of 1.9 K by means of super-fluid helium. Since collisions occur between particles of the same charge, two separate beam pipes are required, with two opposite magnetic field configurations. 392 additional quadrupole magnets placed along the circumference are employed to focus the beams, which would tend to spread because of the Coulomb interaction among the particles that compose them.

As illustrated in Fig. 2.2, the LHC is the last component of a complex apparatus [50] that starts with the production of the protons and their first acceleration in the LINAC 2. The proton bunches are then injected into the Proton Synchrotron Booster (PSB), which accelerates them to 1.4 GeV before pushing them into the Proton Synchrotron (PS), where they reach an energy of 25 GeV. Afterwards, the Super Proton Synchrotron (SPS) further accelerates them up to 450 GeV and eventually injects them into the LHC. Here, 8 radio-frequency resonant cavities oscillating at 400 MHz accelerate the bunches to their final energy with "kicks" of $\Delta E = 0.5$ MeV per turn. Each orbit can contain up to 2800 bunches of protons, with a nominal number of the order of 10^{11} protons each. Each bunch has a small transverse spread (about 15 μm) and a length around 7.5 cm in the beam direction. The minimal bunch time separation is 25 ns. Bunches collide with a peak rate of 40 MHz.

Four detectors are installed along the LHC tunnel, corresponding to the four interaction points: CMS and ATLAS [51], both high-luminosity, multi-purpose detectors, LHCb [52], whose main field of study are rare decays and CP violation in the "b sector", and ALICE [53], devoted to heavy-ion physics. The analysis presented in this thesis refers to data collected with the CMS experiment, whose description is given in the following sections.

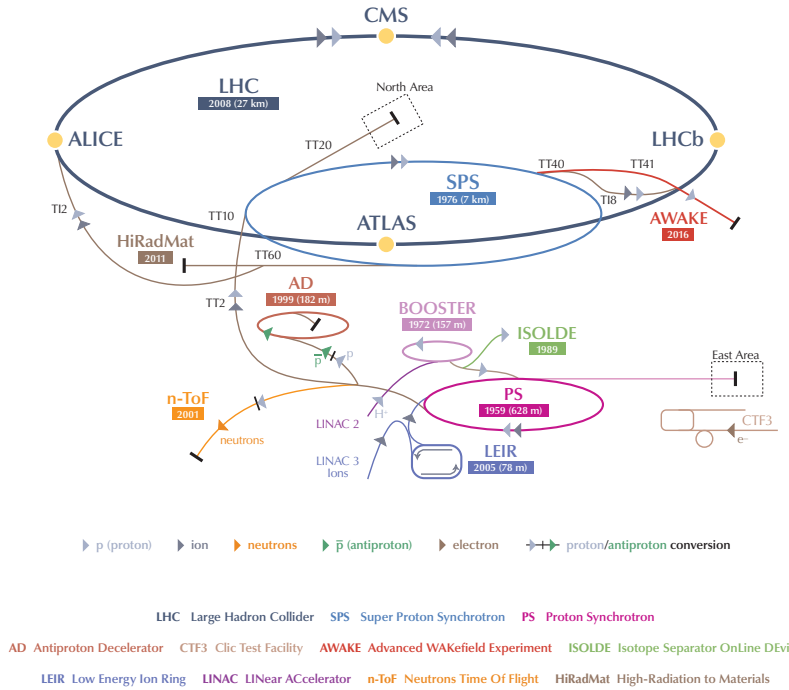

FIGURE 2.2: Schematic view of CERN's accelerator complex, culminating in the LHC.

2.2 The CMS experiment

The Compact Muon Solenoid is a multi-purpose detector [54], installed at interaction point number 5 in the LHC tunnel. Its design philosophy, which was driven by the search for the Higgs boson, is in fact intended to reconstruct a wide range of particles and to explore several phenomena originated in LHC collisions. The highlights in the design of CMS are:

- a high-performance *muon system*, capable to measure accurately the transverse momentum of this kind of particles up to 1 TeV. Connected to this is the presence of a large superconducting solenoidal magnet that produces an intense magnetic field of 3.8 T, allowing a compact muon spectrometer with high performances in track and charge measurements;
- a high-resolution *electromagnetic calorimeter* (ECAL) to detect and measure electrons and photons, with a size compatible with the presence of the magnet;
- an *inner tracking system* for accurate momentum measurements of charged particles;
- a hermetic *hadronic calorimeter* for precise reconstruction of jets and missing transverse energy.

A section of CMS is shown in Fig. 2.3. The cylindrical structure is 28 m long, with a 7.5 m radius, resulting in an overall weight of about 14000 tons. The central section (called *barrel*) is coaxial to the beam axis. Two orthogonal hermetic discs (called *end-caps*) close the structure and extend the acceptance to detect particles traveling close to the beam axis. Figure 2.4 shows a transverse slice of CMS sub-detectors. Moving outwards from the interaction point, emerging particles find the silicon tracker, the electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL), the superconducting solenoid, and the iron return yoke, equipped with muon chambers.

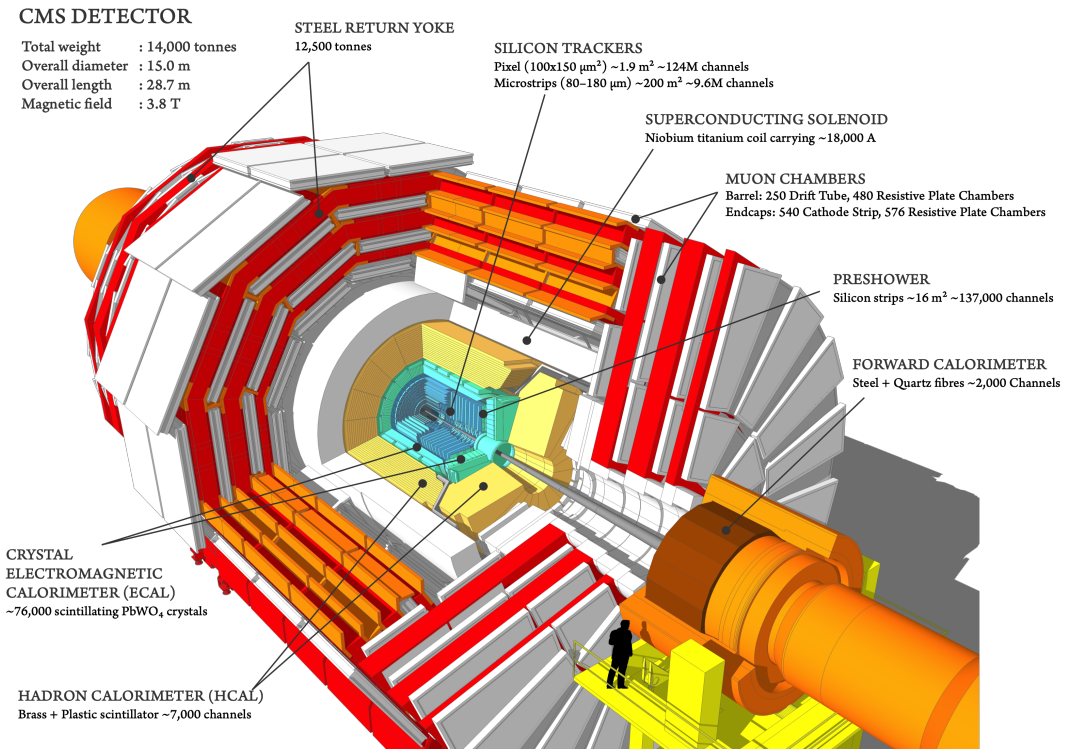

FIGURE 2.3: Schematic view of the CMS detector and its main components.

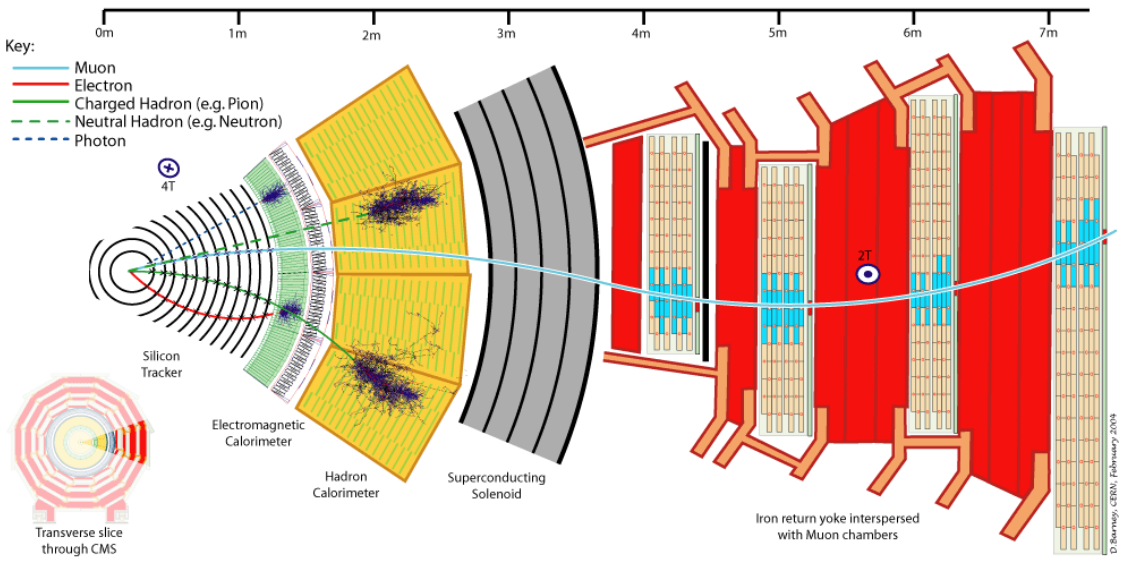

FIGURE 2.4: A sliced view of CMS and its subsystems. The paths of different particles are shown, as well as their interactions with the subdetectors.

CMS uses a right-handed cartesian coordinate system, with the origin at the nominal interaction point at the center of the detector. The x-axis points to the centre of the LHC, the y-axis points upwards, perpendicularly to the LHC plane, and the z-axis towards the anti-clockwise beam direction. Since in hadron colliders the fractions x_1 and x_2 of the parent proton momenta carried by the colliding partons is unknown, it is impossible to determine the particle boost along the beam line in the lab frame. Thus, a pseudo-angular reference system is defined:

- $\theta \longrightarrow$ **polar angle**, measured from the positive z-axis;
- $\phi \longrightarrow$ **azimuthal angle**, in the x–y plane;
- $p_{\mathrm{T}} = p \sin\theta \longrightarrow$ **transverse momentum**, i.e., the particle's momentum in the transverse plane with respect to the beam direction. This quantity is Lorentz-invariant for boosts along the longitudinal (beam) direction;
- $\eta = -\log\left(\tan\frac{\theta}{2}\right) \longrightarrow$ **pseudorapidity**.

The pseudorapidity is in fact the high-energy limit of the *rapidity*, defined as

$$y = \frac{1}{2}\ln\left(\frac{E + p_{\mathrm{z}}}{E - p_{\mathrm{z}}}\right). \tag{2.2}$$

Differences of rapidity (and in the LHC case, of pseudorapidity too) are Lorentz-invariant under boosts along the z-axis. Therefore, the use of pseudorapidity is generally preferred to the use of the polar angle θ. Figure 2.5 shows the pseudorapidity values corresponding to certain angles θ.

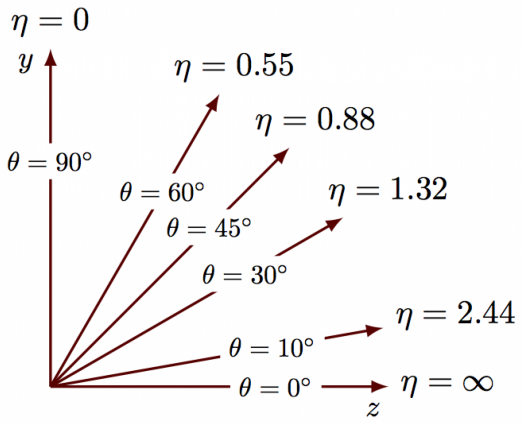

FIGURE 2.5: Pseudorapidity values corresponding to certain values of the polar angle θ.

Distance between particles can be computed in terms of $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, that is the distance in the $\phi - \eta$ plane, also invariant under Lorentz boosts along the z-axis.

2.2.1 The Magnet

The role of the magnetic field is to bend the trajectory of the charged particles, in order to achieve the best transverse momentum resolution possible. This is obtained using a relatively small solenoid that produces an intense field. This innovative feature has driven the design of all the other subdetectors.

The superconducting magnet [55] is a 13 m-long cylinder of niobium-titanium wired coils, with a diameter of 5.9 m. It provides a uniform magnetic field of 3.8 T

at its center, thanks to the flow of an electric current of 18 kA. The magnet flux is returned via a 1.5 m-thick saturated iron yoke, which is also instrumented with four stations of muon chambers.

2.2.2 The Silicon Tracker

The silicon tracker [56] is designed to provide a precise and efficient measurement of the particle trajectories in a high-multiplicity and highly-radioactive environment. Entirely made of silicon, it is capable of measuring few points (10–14) for each particle passing through, with an outstanding spatial resolution of $\approx$10 μm. This results in a momentum resolution below 2% (6%) in the barrel (endcap) for muons with 20 GeV $< p_{\mathrm{T}} <$ 100 GeV, and better than 10% for p_{T} up to 1 TeV. The reconstructed particle trajectory is called a *track*.

The tracker system is composed of three regions, with decreasing granularity moving outwards from the interaction point:

- a *pixel detector* is placed close to the primary vertex, covering the area with the highest particle flux. The size of each pixel is approximately 100 μm $\times$ 150 μm;
- the intermediate region is made of layers of *microstrips* with a minimum strip size of 10 cm $\times$ 80 μm;
- in the outermost region, the particle flux is low enough to adopt larger *strips* with a maximum strip size of 25 cm $\times$ 80 μm.

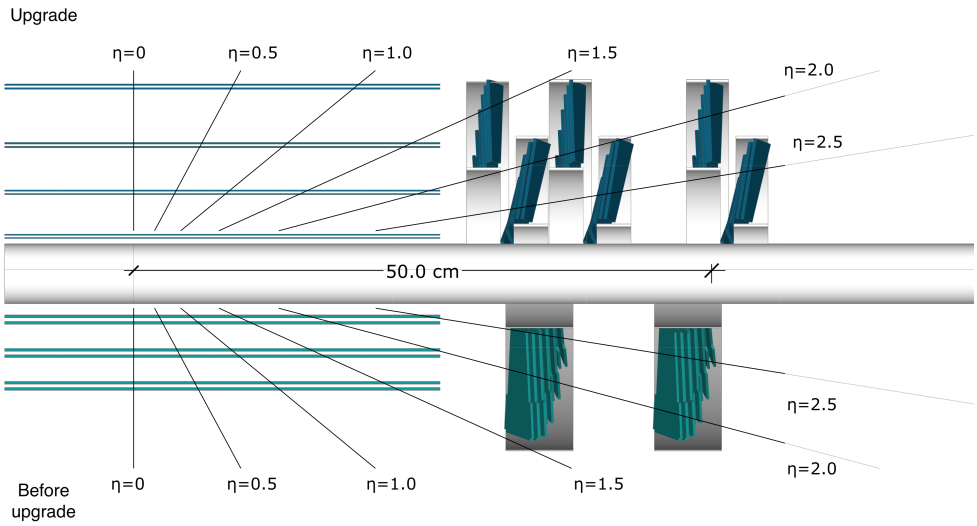

FIGURE 2.6: Longitudinal section of the pixel detector, in its 2016 (below) and post-2016 configurations (above).

About 124 million pixels are accommodated in the 4-layer barrel of the pixel detector, with 3-layer endcaps on each side. This configuration reflects a major upgrade the pixel detector underwent in between the 2016 and 2017 data-taking periods [57], with the addition of one layer in both the barrel and the endcaps with respect to the previous configuration, as illustrated in Fig. 2.6. The strip tracker extends outwards to a radius of 110 cm, up to the calorimetric system. There are approximately 10 million strips arranged in 10 cylindrical layers in the barrel region, and in 9 disks in each of the two endcaps. Figure 2.7 offers a schematic view of the longitudinal section of the whole silicon tracker.

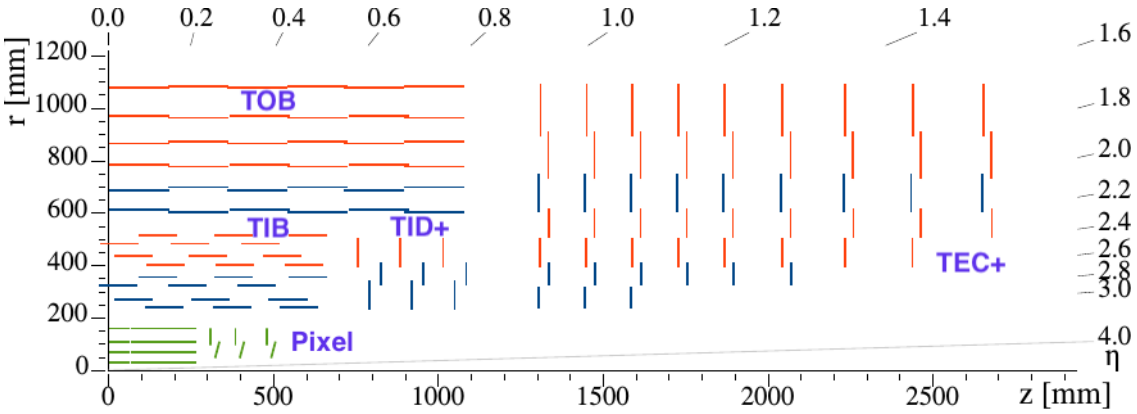

FIGURE 2.7: Schematic of a quarter of the longitudinal section of the CMS silicon tracker. The nominal interaction point is at $z = 0$. Silicon pixels are marked in green, single-sided strip modules are represented as red lines and strip stereo modules are shown as blue lines.

One of the goals of the CMS silicon tracker is to limit the material budget before particles reach the other detectors, since this represents one of the main sources of error in accurate calorimetric measurements of electrons (which emit bremsstrahlung radiation) and photons (which convert into e^+e^- pairs). Furthermore, the multiple scattering of charged particles in the tracker tends to worsen its performances. The material budget of the tracker as a function of pseudorapidity is shown in Fig. 2.8. The tracker adds up to less than half a radiation length (X_0) in the center of the barrel, increasing to a maximum of about $1.8\,X_0$ around $|\eta| = 1.5$.

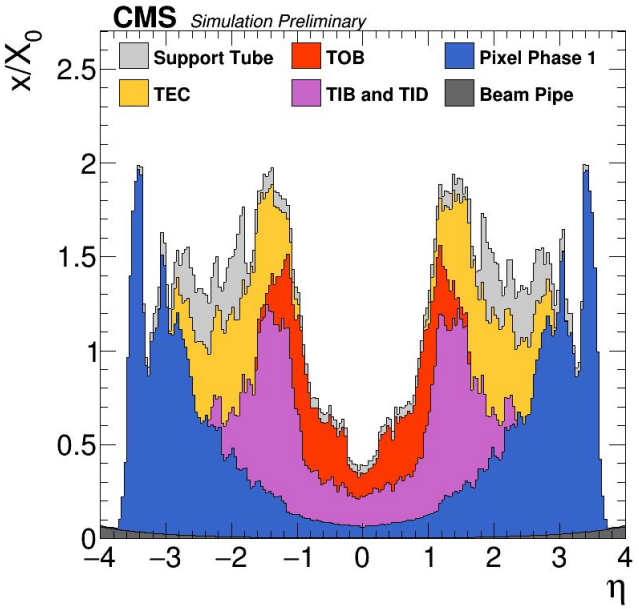

FIGURE 2.8: The silicon tracker material budget as a function of pseudorapidity, expressed in units of radiation length (X_0). Different material categories are shown: beam pipe, pixels, different parts of the strip system (TIB, TID, TOB, and TEC), and support tubes.

2.2.3 The Electromagnetic Calorimeter

The CMS electromagnetic calorimeter (ECAL) is a cornerstone for my search for $W \to \pi\gamma$, since the electrons used to tag the event and the photons arising from the

the rare decay must be detected with great resolution and efficiency.

The ECAL was indeed conceived to obtain an excellent energy resolution for electrons and photons. Its design [58] was inspired by the search for the Higgs boson through the $H \to \gamma\gamma$ decay channel [13] and resulted in a hermetic, homogeneous calorimeter, made up of 75000 lead-tungstate ($PbWO_4$) scintillating crystals.

Geometry

The ECAL has a cylindrical geometry and is composed of a barrel (EB), two endcaps (EE) and a preshower (ES). Approximately 80% of ECAL crystals are used in the central barrel, with a pseudorapidity coverage $|\eta| < 1.44$. EB is made of 36 identical supermodules. They are organized in a semi-projective geometry, forming a 3°-angle with respect to the line that connects them to the nominal interaction point, with the purpose to avoid particles to fall in the separation gap between two crystals. The barrel has a granularity of 360 crystals in the ϕ-direction and 2×85 crystals in the η-direction.

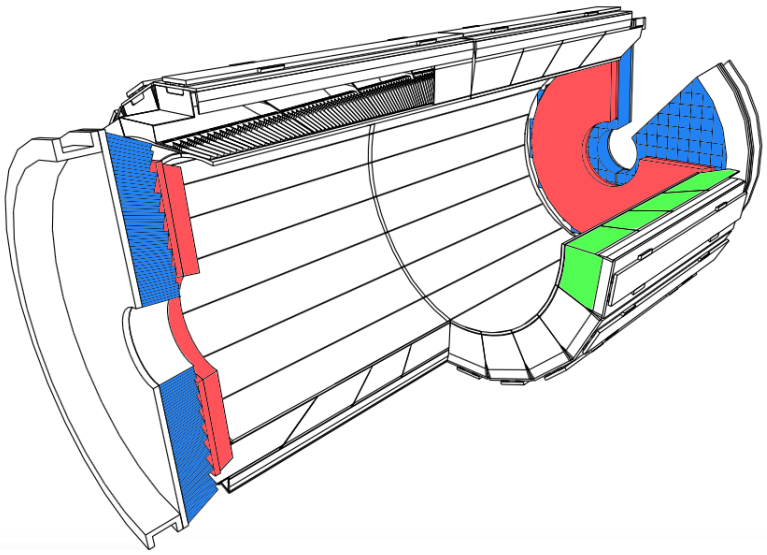

FIGURE 2.9: Schematic view of ECAL. The picture highlights its three components: central barrel EB (green), preshower ES (red), endcaps EE (blue)

The two endcaps are placed at a distance of 3 m from the nominal interaction point, covering up to $|\eta| = 3$. They are made of crystals with a length of $24.7\,X_0$. The endcaps are equipped with a preshower detector, which covers the region $1.7 < |\eta| < 2.6$. ES is a two-layered sampling calorimeter made of lead and silicon strip detectors. The thickness of the two lead absorbers is respectively 2 and $1\,X_0$. The function of the preshower is to provide a better spatial resolution in the endcaps, required to separate neutral pions from photons.

The Crystals

The choice of $PbWO_4$, whose main properties are shown in Table 2.1, has been driven by several factors:

- its short radiation length X_0 favors the construction of a compact calorimeter. This is of primary importance, since the presence of the magnet limits the available space;

- its small Molière radius R_{M} ensures lateral shower containment and, therefore, high granularity, required for $\pi^0 - \gamma$ separation and angular resolution;
- its very fast light emission process makes it suitable for the LHC environment, where bunch crossings are interspaced by 25 ns only.

Table 2.1: Main features of the ECAL crystals.

Parameter	Value
X_0	0.89 cm
R_{M}	2.2 cm
Light yield	100 γ/MeV
% of light emitted in 25 ns	80%

The crystals are trapezoidal, with a transverse dimension of $2.2 \times 2.2\,\mathrm{cm}^2$ at the largest base, namely a square with side corresponding to $1\,R_{\mathrm{M}}$; this characteristic, in addition to a length of 23 cm (about $26\,X_0$), ensures an excellent containment of the electromagnetic showers up to 1.5 TeV.

The light yield of $PbWO_4$ is low compared to other scintillating materials and necessitates the use of specific photodetectors, suitable to operate in the high-intensity magnetic field. Avalanche photodiodes (APDs) are used in the barrel. These photodetectors are made of semiconducting silicon, with a doping profile studied to generate an avalanche of electrons. The APDs operate under a 50 V electric field and are able to produce a very large signal in a short time, thus being suitable for the low light yield of lead tungstate. They can work at different gains, according to the magnitude of the energy released in the corresponding crystal.

On the other hand, endcaps are equipped with vacuum phototriodes (VPTs), because the radiation levels are too extreme for silicon photodiodes in the forward and backward areas of the detector. Each VPT contains three electrodes within a vacuum tube: one of them releases an electron when hit by a scintillation photon, one works as an anode, producing several electrons that are then accelerated to the third electrode (the dynode), which releases a second batch of electrons. As for the barrel case, the signal is amplified and digitized and sent along optic fibers outside the detector.

Energy resolution

The ECAL is a homogeneous calorimeter. Its energy resolution can be parametrized as:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{2.3}$$

where E is the energy of electrons and photons, expressed in GeV. The stochastic term S depends on fluctuations of the number of detected photons, the noise term N is due to the electronics noise and the constant term C depends on lateral containment, non-uniformity of response and intercalibration. These terms have been measured in test beams [59] at single-crystal level and were found to be:

$$S = 0.028\,\mathrm{GeV}^{\frac{1}{2}} \qquad N = 124\,\mathrm{MeV} \qquad C = 0.003. \tag{2.4}$$

The energy resolution for photons with $p_{\mathrm{T}} \simeq 60\,\mathrm{GeV}$ is around 1% over the solid angle of the ECAL barrel, and varies between 2.2% and 5% in the endcaps. For electrons

with $p_T \simeq 45\,\text{GeV}$ from $Z \to e^+e^-$ decays, the resolution is better than 2% in the central region of the ECAL barrel ($|\eta| < 0.8$), and is between 2% and 5% elsewhere. For low-bremsstrahlung electrons with 94% or more of the energy contained within a 3×3 array of crystals, the energy resolution improves to 1.5% for $|\eta| < 0.8$.

2.2.4 The Hadronic Calorimeter

The purpose of the hadronic calorimeter (HCAL) [60, 61] is to measure the energy of hadrons and jets. In particular, the HCAL is a crucial element for jet reconstruction. By contributing to the determination of the total visible energy of an event, it is also fundamental for neutrino detection via missing transverse momentum.

The HCAL is composed by four main parts. The barrel (HB) and the endcap (HE) detectors cover a pseudorapidity range up to $|\eta| < 3.0$, while the forward detector (HF) is designed for the high pseudorapidity region $3.0 < |\eta| < 5.0$. The HCAL is a sampling calorimeter, with 3.7 mm-thick active layers of plastic scintillators interspersed with 5–8 cm-thick brass absorbers. Light is collected by special optic fibers and the signal is then amplified by photodetectors. In the HF, plastic scintillators are replaced by quartz fibers, interspersed with layers of iron absorber. The total thickness of the HCAL of 7 interaction lengths (λ_0) is not always sufficient for a complete containment of hadronic showers. Therefore, an additional layer (HO) corresponding to $1\,\lambda_0$ is located outside the magnet, with the purpose of increasing the energy resolution.

The energy resolution of HCAL for pions can be parametrized as

$$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E(\text{GeV})}} \oplus 8\%, \tag{2.5}$$

while the resolution for the ECAL-HCAL combined system is

$$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E(\text{GeV})}} \oplus 5\% \tag{2.6}$$

2.2.5 The Muon System

Muons are the only charged particles that are not stopped by the calorimeters. Hence, a system of muon chambers [62] is hosted in between the iron layers of the magnet return yoke, so that the measurement of muon charge and momentum can exploit the large 1.5 T magnetic field in this region of the CMS detector. The muon identification is performed using these chambers, while the p_T measurement is obtained by combining the information of the chambers with that of the tracker. As it happens for other detectors, the muon system is divided into a barrel ($|\eta| < 1.2$) and two endcaps ($1.2 < |\eta| < 2.4$). A schematic view of the system is presented in Fig. 2.10. The apparatus consists of three kinds of sub-detectors, placed in different acceptance regions:

- *drift tubes* (DT), located in the barrel part. Each chamber is made of twelve 4 cm-wide tubes containing a stretched wire within a gas volume;
- *cathode strip chambers* (CSC) in the endcap. In this region, where a high-multiplicity of particles is expected, arrays of anode wires crossed with cathode strips are placed within a gas volume and used for muon detection;

- *resistive plate chambers* (RPC), located both in the central barrel and in the endcap regions. RPCs are fast detectors consisting of two parallel plates separated by a gas volume. Their excellent time resolution (≈1 ns) makes them suitable to be used also as fast, high-efficiency triggers.

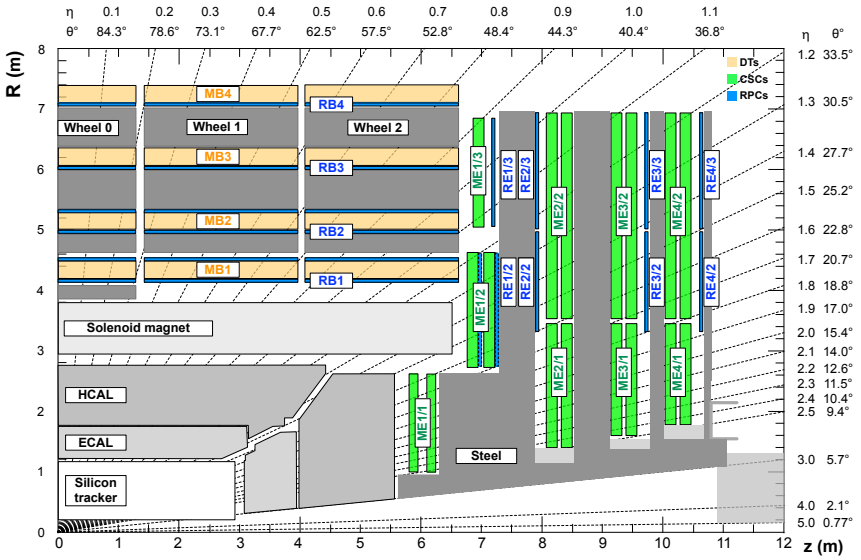

FIGURE 2.10: Schematic representation of a quarter of the CMS muon system in longitudinal view. The different subdetectors composing it are shown with different colors.

The muon system guarantees an excellent transverse momentum resolution, with a visible improvement with respect to the use of the tracker only, especially for a muon with $p_T > 200\,\text{GeV}$ [63]. Below this threshold, the performance in the momentum measurement is dominated by the tracker. This can be seen in Fig. 2.11, which shows the RMS of the residual distribution:

$$R(q/p_T) = \frac{1}{\sqrt{2}} \frac{(q/p_T)^{\text{upper}} - (q/p_T)^{\text{lower}}}{(q/p_T)^{\text{lower}}} \tag{2.7}$$

as a function of the muon p_T. q/p_T is the charge to momentum ratio, proportional to the muon trajectory curvature, measured in the upper and lower halves of CMS.

2.2.6 The Trigger System

With an average PU ≈ 30, the typical raw size of an "event" as seen by the CMS detector is about 1 MB. Since the bunch-crossings occur with a frequency of 40 MHz, not all of the enormous amount of generated data can be recorded and stored. Moreover, saving all the collision events would be counterproductive, since most of them only comprise soft processes that are not interesting for the CMS physics program. A trigger system [64] is therefore used, in order to reduce the event rate to ≈1 kHz, a value compliant with the maximum readout bandwidth of the CMS detector (≈2 GB/s). The trigger system is structured into two tiers:

- Level-1 Trigger (L1): a series of hardware processors performing simple logical operations directly on detector signals. The L1 trigger uses information from the calorimeters and muon detectors to select events at a rate of around 100 kHz within a fixed time interval of about 4 μs;

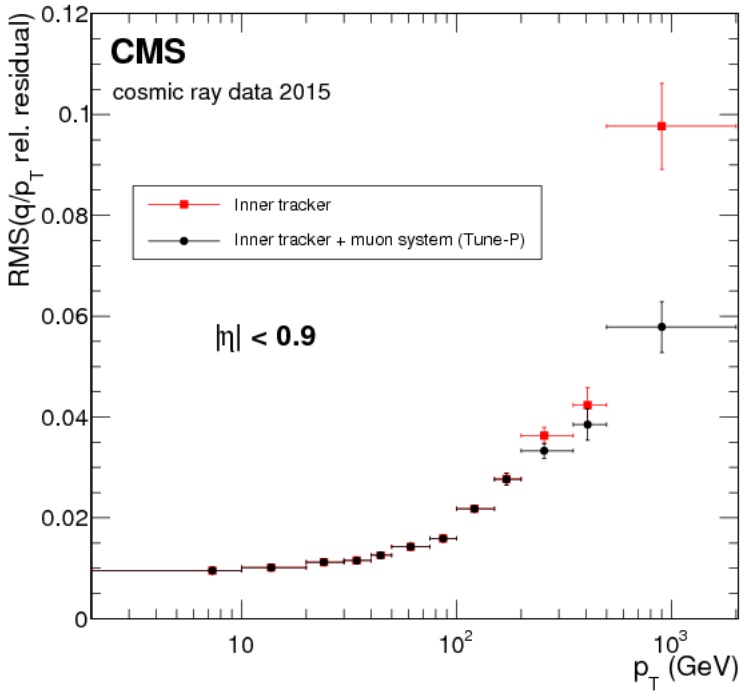

FIGURE 2.11: RMS of the distribution of $R(q/p_{\mathrm{T}})$ as a function of the p_{T} of muons from cosmic rays, using the inner tracker fit only (red squares) or including the muon system (black dots).

- High level trigger (HLT): a farm of processors running a version of the full event reconstruction software on the events selected at L1. Only those events containing objects in the useful regions and with specific kinematic properties are reconstructed, while the uninteresting events are rejected.

At a first stage, only the full information of the muon system and of the calorimeters is employed, since the track reconstruction algorithm is too slow to be used in the L1. On the contrary, the algorithms used by the L1 (called *L1 bits* or *seeds*) are fast and basic, and inspect the event for very general features, such as the presence of a muon, an electromagnetic particle or $p_{\mathrm{T}}^{\mathrm{miss}}$ above certain energy thresholds. Each event is therefore classified as an ensemble of several trigger bits. The HLT, on the other hand, implements more complex and flexible algorithms (referred to as *HLT paths*), thanks to the complete freedom in the selection of the data to access, and to the larger amount of time available.

In case a physical process of interest occurs with an unsustainable rate, prescales might be introduced at either L1 or HLT level to save only a fraction of the otherwise accepted events. Prescales reduce the effective amount of integrated luminosity available for analyses exploiting that specific process. Therefore, in this search for the rare decay $\mathrm{W} \to \pi\gamma$, no prescaled triggers were used at any level.

2.3 Physics objects identification and reconstruction

The work presented in this thesis relies on the capability of the CMS detector to identify and correctly reconstruct the final state particles involved in the search of the rare decay $\mathrm{W} \to \pi\gamma$. The presence of a charged lepton (muon or electron, possibly arising from a τ decay) is required to tag the $\mathrm{t\bar{t}}$ event. If the mentioned lepton is an

electron, it is of great importance for an experimental apparatus to distinguish it from the photon characterizing the rare decay. A key role is played by b jets, which help to confirm the $t\bar{t}$ topology and thus to reduce the background from the hadronization of light-flavor quarks and gluons. The following sections will briefly describe how these and other objects are identified and reconstructed at CMS. A section is dedicated to the particle flow-based event reconstruction.

2.3.1 The particle flow

The particle flow (PF) is a technique for event reconstruction that uses an optimized combination of the information from all the CMS subsystems [65]. The elementary information from the subdetectors (hits in the tracker or in the muon system, energy deposits in the calorimeters) is employed to build higher-level objects (such as tracks and energy clusters), which are then combined to form PF candidates. First, the PF algorithm tries to form muons by pairing segments in the muon chambers with tracks in the tracker. The tracks for which this pairing was not possible can then be used to form other PF candidates (or to form "tracker muons", as discussed in the next section). If they can be associated to energy clusters in the ECAL, they form electrons; on the contrary, ECAL clusters with no associated track are identified as photons. Similarly and with the addition of the information concerning HCAL clusters, charged and neutral hadrons are formed. Finally, jets can be built by clustering these PF objects.

2.3.2 Particle isolation

The PF technique also plays a major role in quantifying a particle's *isolation*. Each collision event that takes place inside the CMS detector is characterized by a great multiplicity of generated particles. The idea behind isolation is to consider a solid angle cone ΔR around one of them, and to calculate the energy contained within. According to the physical process that generated the particle, this energy is expected to be significantly different. For instance, a particle originating from electroweak processes tends to be more isolated than one from QCD processes, which usually exhibit a larger activity due to parton hadronization and gluon radiation.

In particular, with the PF-based isolation it is possible to separate and identify the different energy contributions that stem from charged and neutral hadrons, photons, and hadronic particles arising from PU.

2.3.3 Muons

The CMS muon system is able to detect muons in the pseudorapidity range $|\eta| < 2.4$. The energy of muons is obtained from the curvature of the corresponding track. If the segments in the muon system are matched with the tracks from the silicon tracker (reconstructed with a Kalman Filter algorithm [66]), this results in a momentum resolution of ≈1% in the barrel and ≈3% in the endcaps, for muons with momenta up to about 100 GeV [63]. The muon reconstruction leads to the definition of three typer of muons:

- *standalone muons:* exploit the information from the hits in the muon chambers only. The momentum resolution for a standalone-muon track is on average about one order of magnitude worse than that of a track produced in the tracker. This category of muons has then worse momentum resolution and higher admixture of cosmic-ray muons than the other two categories. Therefore, it is usually not

used in physics analysis, and mainly serves as a component of the global-muon reconstruction;

- *global muons:* combine standalone muons with tracks in the tracker;
- *tracker muons:* the reconstruction starts from inner tracks. Then, the algorithm tries to match them with compatible energy deposits in the calorimeters and hits in the muon system, where the reconstruction of a complete segment is not required. Well suited for low-p_{T} muons ($p_{\mathrm{T}} \lesssim 10\,\mathrm{GeV}$), which might barely reach the muon system due to the strong magnetic field, and therefore not leave enough hits in the muon stations for a standalone muon to be reconstructed.

The muons used in this analysis are selected by the PF algorithm and, at the same time, they need to be either tracker or global muons. In addition, the fraction of valid tracker hits must be >0.8 and either of the following requirements must be satisfied:

- be a "good" global muon:
  - global-track fit $\chi^2/\mathrm{ndf} < 3$;
  - tracker-standalone position match with a $\chi^2 < 12$;
  - kink finder < 20;
  - segment compatibility > 0.303;
- have a segment compatibility > 0.451.

The *kink finder* is an algorithm aiming at discriminating between muons originating from the PV and muons produced from the in-flight decay of K mesons. If a K meson does not decay in the tracker, though, the kink (angle between the initial direction of the K meson and the direction of the muon it generates) is not detected. In this case, the *segment compatibility*, which is the association method between the tracks in the tracker and the segment left by the charged particle in the muon system, permits to discriminate between the two scenarios of muons from PV and in-flight decay. All together, these features define a so-called *medium muon*. The efficiency for the medium muon identification in 2016 is shown in Fig. 2.12, as a function of the pseudorapidity.

2.3.4 Electrons

The electron identification procedure requires the association of an energy cluster in ECAL with a track in the silicon tracker. Since large radiation losses may occur when electrons traverse the tracker material, a track reconstruction via the Kalman Filter only can be affected by large changes of curvature due to bremsstrahlung. Therefore, a Gaussian-sum filter (GSF) algorithm [67] is used to improve the reconstruction. Using simulated events of $q\bar{q} \to \gamma^*/\mathrm{Z}^* \to \mathrm{e}^+\mathrm{e}^-$ (Drell–Yan), it is possible to estimate the number of reconstructed electron candidates that are not matched with real electrons (the so-called *fake rate*), as a function of the PU. This is presented in Fig. 2.13 for 2016 and 2017, with the 2017 performance showing improvements ascribable to the new pixel tracker (see Section 2.2.2).

In addition, multivariate techniques (MVA) are used to identify electrons [68]. Input variables include observables sensitive to bremsstrahlung along the electron trajectory, the geometrical and energy-momentum compatibility between the electron track and the associated energy cluster in the ECAL, the energy distribution of the electromagnetic shower, and discrimination against electrons originating from photon

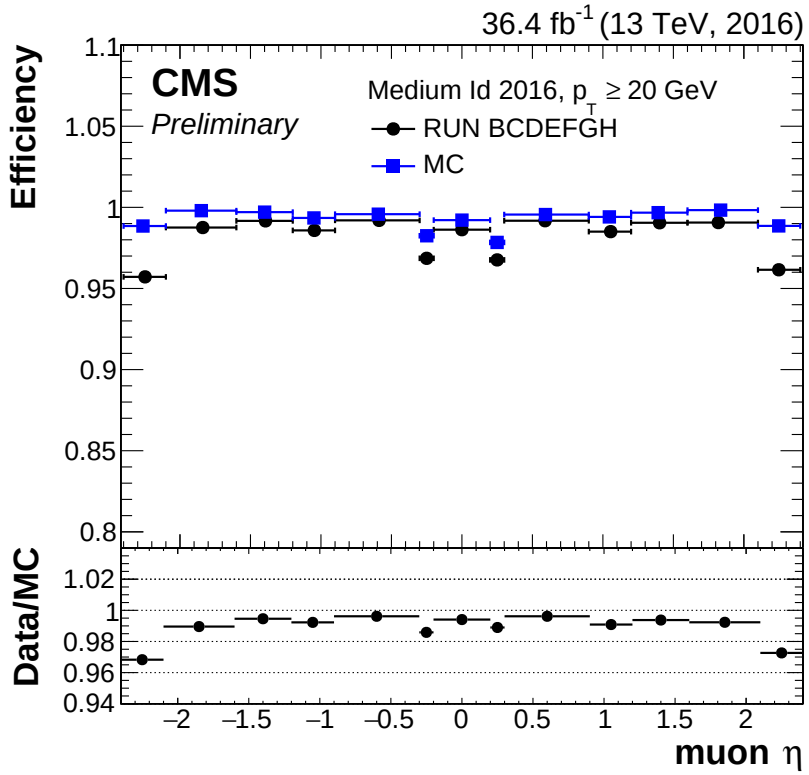

FIGURE 2.12: "Medium" identification efficiency for muons with $p_{\mathrm{T}} \geq 20\,\mathrm{GeV}$, as a function of the pseudorapidity η. A comparison is made between MC simulations and collision data corresponding to an integrated luminosity of $36.4\,\mathrm{fb}^{-1}$.

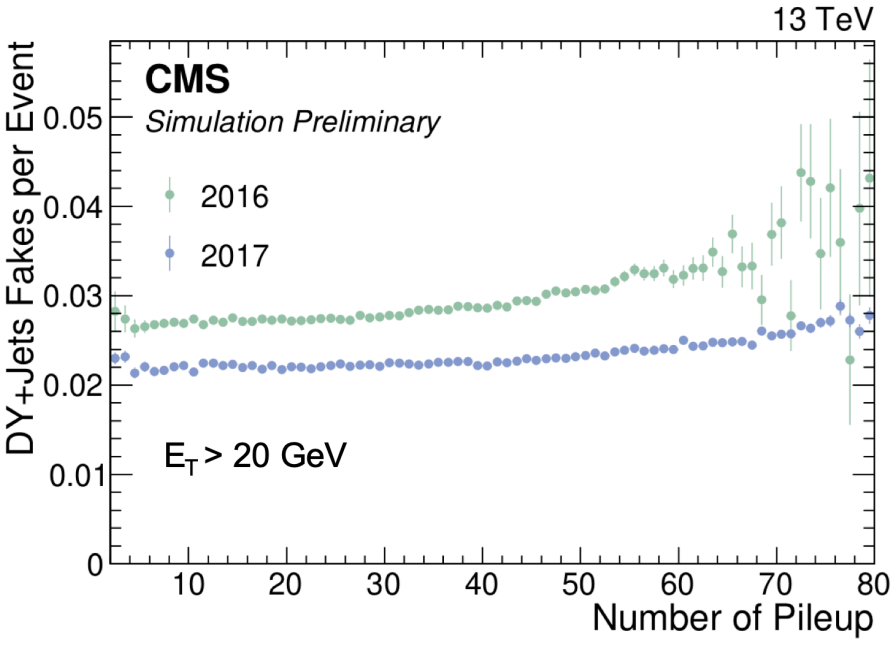

FIGURE 2.13: Electron fake rate as a function of the pileup in MC simulated events. Electrons with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.5$ are considered. The enhanced performance in 2017 is ascribable to the new pixel tracker. Note that the fake rate is further reduced by the electron identification step, which is not applied here.

conversion. Specific PF-based isolation variables are also employed in the multivariate classifier, including the p_T of charged hadrons, neutral hadrons, and photons in the event. As discussed in Section 2.3.2, these use a solid angle cone ΔR around the electron to evaluate the activity in its vicinity, helping to discriminate between electroweak processes and, for instance, parton hadronization and QCD-driven phenomena. In this analysis, the electron identification working point chosen corresponds to a 90% signal efficiency of the MVA, averaged on the different parts of the detector. Plots of electron reconstruction and identification efficiency are shown in Fig. 2.14, as an example, for the 2017 dataset.

2.3.5 Photons

Photons are reconstructed within the ECAL fiducial region $|\eta| < 2.5$, excluding the barrel-endcap transition region $1.44 < |\eta| < 1.57$, where the performance is suboptimal. A photon or electron entering the electromagnetic calorimeter produces an electromagnetic shower, which spreads over several ECAL crystals. Since material corresponding to $1 - 2\,X_0$ is placed before the ECAL, photons may undergo conversion into e^+e^- pairs, and electrons and positrons may generate bremsstrahlung photons. The strong magnetic field of the experiment tends to spread the radiated energy along ϕ within the tracker volume. Dynamic clustering algorithms are then used to sum together energy deposits in crystals belonging to the same electromagnetic shower and to recover the radiated energy. The crystals which participated in the shower are then merged into a "supercluster" (SC). The photon energy is computed using the signal recorded from ECAL crystals forming the supercluster. An electron veto is implemented to minimize photon misidentification, combining information from the ECAL and the pixel tracker. As in the electron case, photon identification also proceeds through a multivariate classifier that makes use of PF-based isolation variables [69]. The photon identification efficiency working point chosen in this analysis corresponds to a 90% signal efficiency of the MVA, averaged on the different parts of the detector. A plot of photon identification efficiency for the chosen MVA working point is shown in Fig. 2.15, as an example, for the 2017 dataset.

2.3.6 Charged hadrons

Once muons, electrons, and isolated photons are identified and removed from the list of PF candidates reconstructed in each event, the remaining particles to be identified are hadrons from jet fragmentation and hadronization. As any other charged particles, charged hadrons are first of all reconstructed as tracks in the silicon tracker [66]. A combinatorial track finder based on Kalman Filtering is used in three stages:

- initial seed generation with a few hits compatible with a charged-particle trajectory;
- trajectory building to gather hits from all tracker layers along this charged-particle trajectory;
- final fitting to determine the charged-particle properties: origin, transverse momentum, and direction.

Stringent track quality criteria are instrumental in keeping the misreconstructed track rate at the level of a few per cent, but, as inferred by simulations, they limit the reconstruction efficiency to only 70–80% for charged pions with $p_T > 1\,\text{GeV}$, compared to ≈99% for isolated muons. Below a few tens of GeV, the difference between pions

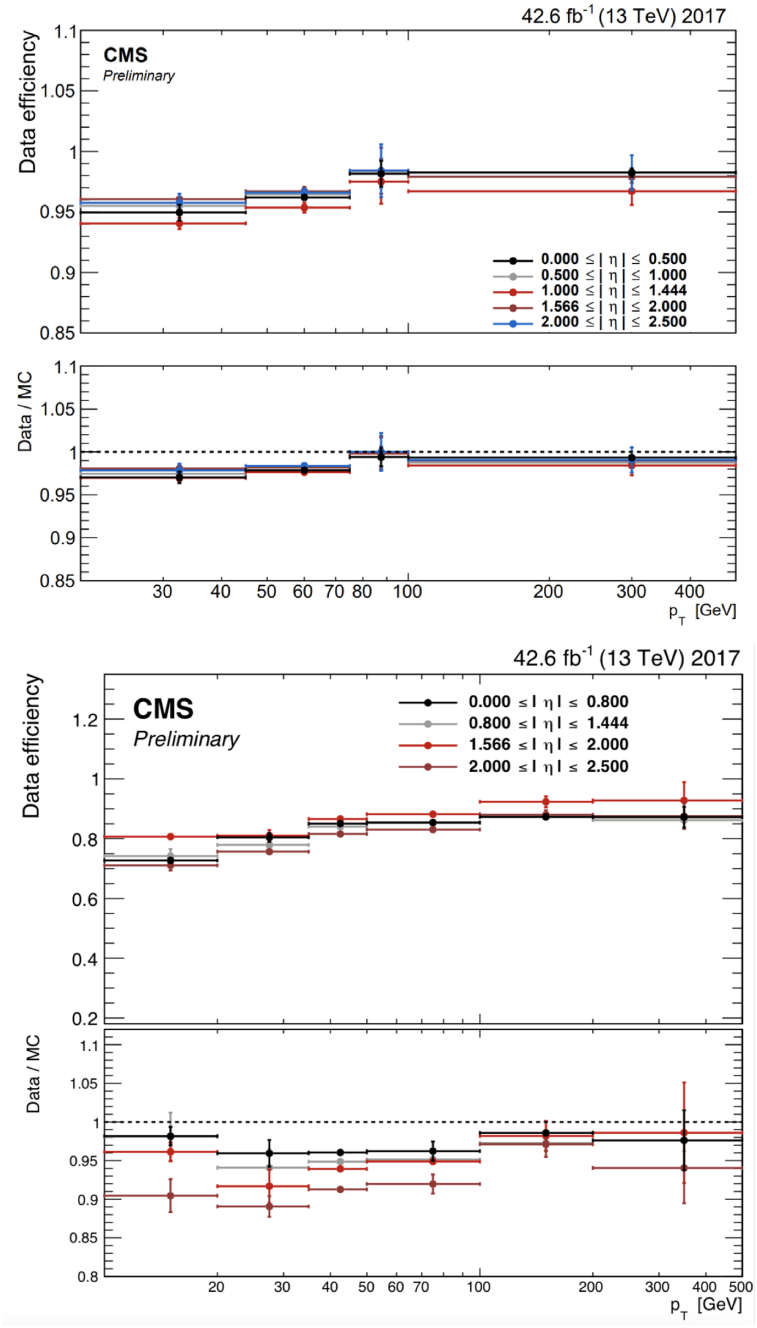

FIGURE 2.14: Electron reconstruction (upper) and identification (lower) efficiency, as a function of the electron p_{T} and for various pseudorapidity ranges. A comparison is made between MC simulations and collision data corresponding to an integrated luminosity of $42.6\,\mathrm{fb}^{-1}$. The identification efficiency is measured for the multivariate identification working point corresponding to a 90% signal efficiency, and the MVA includes the isolation variables.

and muons is almost entirely accounted for by the probability for pions to undergo a nuclear interaction within the tracker material before reaching the minimum threshold on the number of hits in the tracker layers to be reconstructed. This probability ranges between 10 and 30%, as a function of the pseudorapidity. The tracking efficiency is further reduced for p_{T} values above 10 GeV: these high-p_{T} particles are found mostly in collimated jets, in which the tracking efficiency is limited by the silicon detector pitch, that is by the capacity to disentangle hits from overlapping particles.

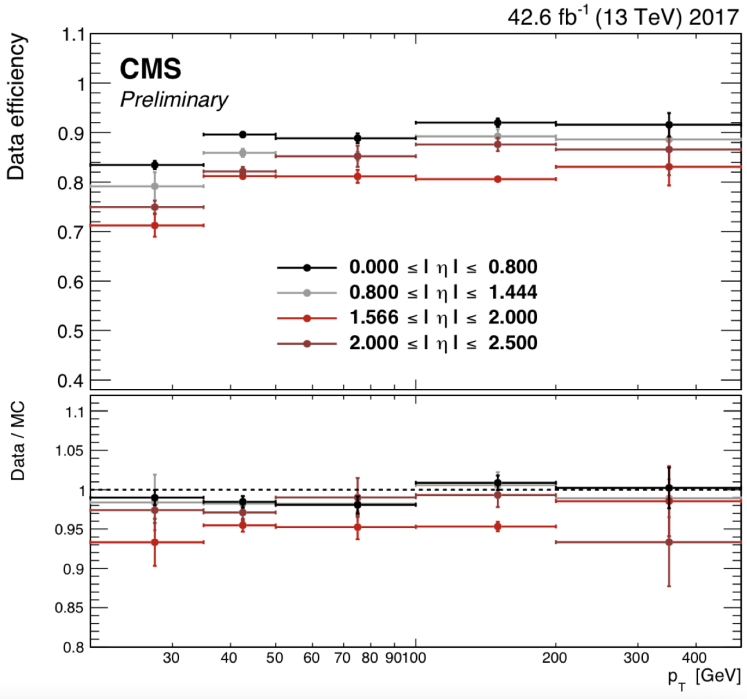

FIGURE 2.15: Photon identification efficiency, as a function of the photon p_{T} and for various pseudorapidity ranges. A comparison is made between MC simulations and collision data corresponding to an integrated luminosity of $42.6\,\mathrm{fb}^{-1}$. The identification efficiency is measured for the multivariate identification working point corresponding to a 90% signal efficiency, and the MVA includes the isolation variables.

To increase the tracking efficiency while keeping a low misreconstructed track rate, the combinatorial track finder is applied in several successive iterations ("iterative tracking"), each with moderate efficiency but with as high a purity as possible. At each step, the reduction of the misreconstruction rate is accomplished with quality criteria on the track seeds, on the track fit χ^2, and on the track compatibility with originating from one of the reconstructed primary vertices, adapted to the track p_{T}, $|\eta|$, and number of hits. The efficiency and the misreconstruction rates of the combinatorial track finder and of the iterative tracking method are shown in Fig. 2.16. Thanks to the iterative tracking, the reconstruction efficiency for charged hadrons exceeds 90% for $1 < p_{\mathrm{T}} < 10\,\mathrm{GeV}$, progressively degrading at larger transverse momenta.

In addition, hadrons generally deposit energy in both ECAL and HCAL. In particular, the ECAL has a substantially different response to hadrons and to photons. Its high granularity, together with the large magnetic-field strength, ensure excellent separation of charged hadron and photon energy deposits. The HCAL response, on the other hand, depends on the fraction of the shower energy deposited in the ECAL, and is not linear with energy. The ECAL and HCAL cluster energies therefore need to be carefully recalibrated to get an estimate of the true hadron energy.

The tracker measures the p_{T} of charged hadrons with a resolution of 1% for $p_{\mathrm{T}} < 20\,\mathrm{GeV}$. The relative resolution then degrades with increasing p_{T} to reach the calorimeter energy resolution for track momenta of several hundred GeV.

2.3.7 Jets

A jet is a collimated cascade of particles, and is seen by the detector in the form of energy clusters in the calorimeters (both ECAL and HCAL), associated with inner

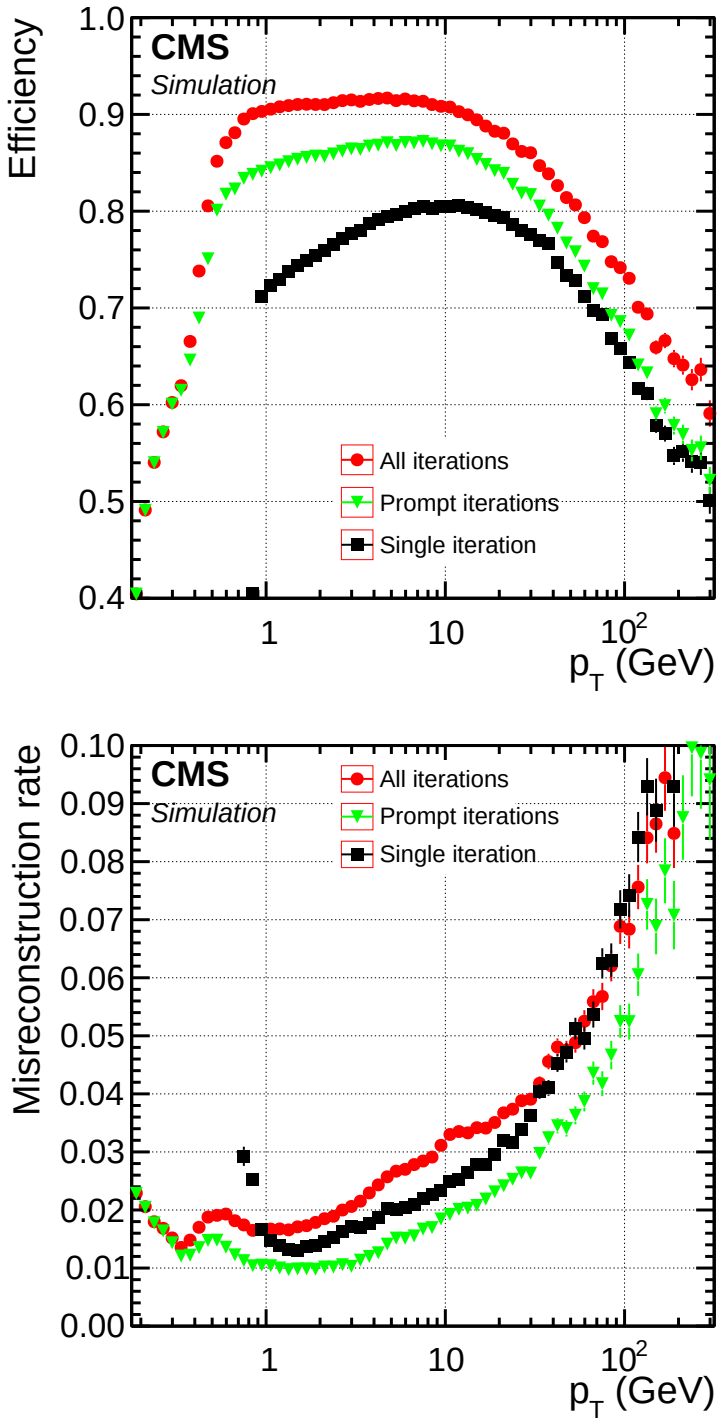

FIGURE 2.16: Efficiency (upper) and misreconstruction rate (lower) of the combinatorial track finder (black squares) and of the iterative tracking method (green triangles: prompt iterations based on seeds with at least one hit in the pixel detector; red circles: all iterations), as a function of the track p_{T}, for charged hadrons in simulated multijet events without pileup interactions. Only tracks with $|\eta| < 2.5$ are considered in the efficiency and misreconstruction rate determination. The efficiency is displayed for tracks originating from within 3.5 cm of the beam axis and ± 30 cm of the nominal centre of CMS along the beam axis.

tracks. Jets are reconstructed from the PF candidates using the anti-k_{T} clustering algorithm [70] with a distance parameter $R = 0.4$, as implemented in the FastJet

library [71]. PF candidates are used by the algorithm to build jets in a cone with radius ΔR around a given candidate, starting from particles with the largest p_{T} and proceeding by clustering softer ones until all candidates are used. Jet momentum is determined as the vectorial sum of all particle momenta in the jet, and is found from simulation to be, on average, within 5 to 10% of the true momentum over the whole p_{T} spectrum and detector acceptance. Additional pp interactions within the same or nearby bunch crossings can contribute with additional tracks and calorimetric energy depositions, increasing the apparent jet momentum. To mitigate this effect, tracks identified to be originating from PU vertices are discarded and an offset correction is applied to correct for remaining contributions. This is achieved through a technique referred to as *charged-hadron subtraction* (CHS) [72]. Figure 2.17 shows an estimate from simulation of the jet energy resolution as a function of the particle-level jet p_{T}, for two different pseudorapidity ranges.

Jet energy corrections are derived from simulation studies so that the average measured energy of jets becomes identical to that of particle level jets. In situ measurements of the momentum balance in dijet, photon+jet, Z+jet, and multijet events are used to determine any residual differences between the jet energy scale in data and in simulation, and appropriate corrections are made [73]. Furthermore, jets are required to have at least two PF constituents and more than 1% of their energy in either the ECAL or the HCAL, to reject fake jets arising from instrumental effects.

b jets

Jets originating from the hadronization of bottom quarks (b jets) are a special kind of jet, since b quarks hadronize in particles with a lifetime sufficient to travel a non-negligible distance before decaying. Thus, the tracks of the particles forming a b jet do not originate from the interaction vertex, but from a displaced secondary vertex. The presence of a secondary vertex and kinematic variables of the tracks associated with this vertex, such as their direction and distance from the primary vertex, can therefore be used to discriminate between b and non-b jets. Several algorithms have been developed by the high-energy physics community to correctly identify b jets. Among these, the *Combined Secondary Vertex* (CSV) uses the significance of the flight distance (the ratio of the flight distance to its estimated uncertainty) as a discriminating variable, together with track-based lifetime information. Thanks to the latter, it is possible to combine compatible tracks in "pseudo-vertices", or to define a "no vertex" category and attempt a jet reconstruction solely based on track information. This allows for the computation of a subset of secondary-vertex-based quantities even without an actual vertex fit. Therefore, discrimination can be provided also in cases when no secondary vertices are found. The variables utilized to discriminate between b jets, c jets, and light parton jets are:

- the vertex category (i.e. real, "pseudo" o "no vertex");
- the flight distance significance in the transverse plane ("2D") of the secondary vertex with the smallest uncertainty on its flight distance;
- the number of secondary vertices;
- the pseudorapidity of the track relative to the jet axis for the track with the highest 2D impact parameter significance;
- the corrected mass of the secondary vertex with the smallest uncertainty on its flight distance for jets with a reconstructed secondary vertex, or the invariant

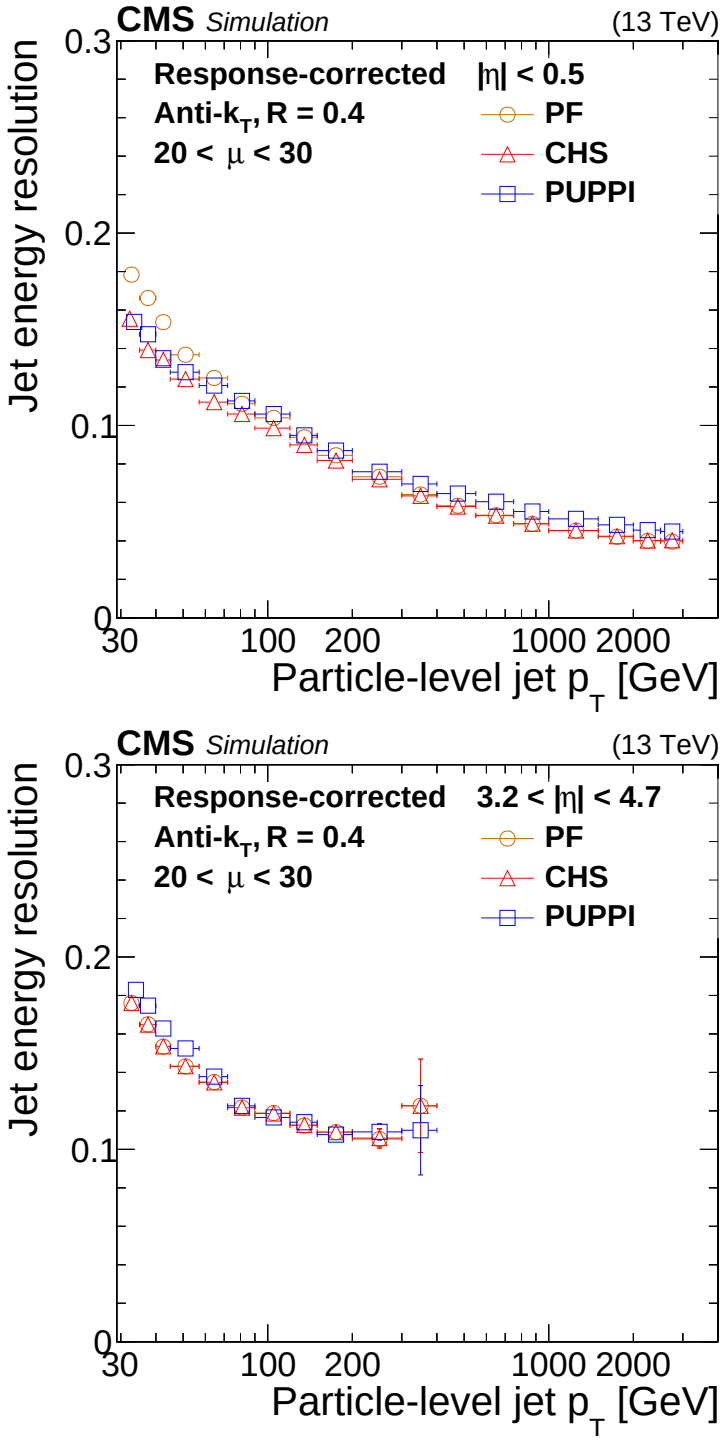

FIGURE 2.17: Jet energy resolution as a function of the particle-level jet p_T in a simulation of QCD multijet events, for $|\eta| < 0.5$ (upper) and $3.2 < |\eta| < 4.7$ (lower). The red triangles refer to the jets used in this analysis, namely PF jets with CHS applied. The number of concurrent pp interactions is required to be between 20 and 30.

mass obtained from the total summed four-momentum vector of the selected tracks for jets in the "pseudo-vertex" category;

- the number of tracks associated with the secondary vertex for jets with a reconstructed secondary vertex, or the number of selected tracks for jets in the

"pseudo-vertex" category;

- the energy of the secondary vertex with the smallest uncertainty on its flight distance divided by the energy of the total summed four-momentum vector of the selected tracks;
- the ΔR between the flight direction of the secondary vertex with the smallest uncertainty on its flight distance and the jet axis for jets with a reconstructed secondary vertex, or the ΔR between the total summed four-momentum vector of the selected tracks for jets in the "pseudo-vertex" category;
- the signed 3D impact parameter significances of the four tracks with the highest 2D impact parameter significance;
- the track p_{T} relative to the jet axis, i.e., the track momentum perpendicular to the jet axis, for the track with the highest 2D impact parameter significance;
- the ΔR between the track and the jet axis for the track with the highest 2D impact parameter significance;
- the track p_{T} relative to the jet axis divided by the magnitude of the track momentum vector for the track with the highest 2D impact parameter significance;
- the distance between the track and the jet axis at their point of closest approach for the track with the highest 2D impact parameter significance;
- the distance between the primary vertex and the track at the point of closest approach between the track and the jet axis for the track with the highest 2D impact parameter significance;
- the transverse momentum of the total summed four-momentum vector of the selected tracks divided by the transverse momentum of the jet;
- the ΔR between the total summed four-momentum vector of the tracks and the jet axis;
- the 2D impact parameter significance of the first track that raises the combined invariant mass of the tracks above 1.5 GeV (related to the c quark mass). This track is obtained by summing the four-momenta of the tracks adding one track at the time. Every time a track is added, the total four-momentum vector is computed. The 2D impact parameter significance of the first track that is added resulting in a mass of the total four-momentum vector above the aforementioned threshold is used as a variable;
- the number of selected tracks;
- the jet p_{T} and η.

Throughout this analysis, the b-tagging is performed using an advanced version of the CSV algorithm called DeepCSV [74], which exploits a deep neural-network architecture. The performance of the DeepCSV algorithm is shown in Fig. 2.18, where the probability for non-b jets to be misidentified as b jet is plotted as a function of the efficiency to correctly identify b jets. Throughout this analysis, a *loose* working point of the DeepCSV tagger is used, providing an efficiency to select b jets of about 90%, with a rate of incorrectly tagging jets originating from the hadronization of gluons or u, d, s quarks of about 10%. The minimum transverse momentum for reconstructed b jets is imposed to be 25 GeV.

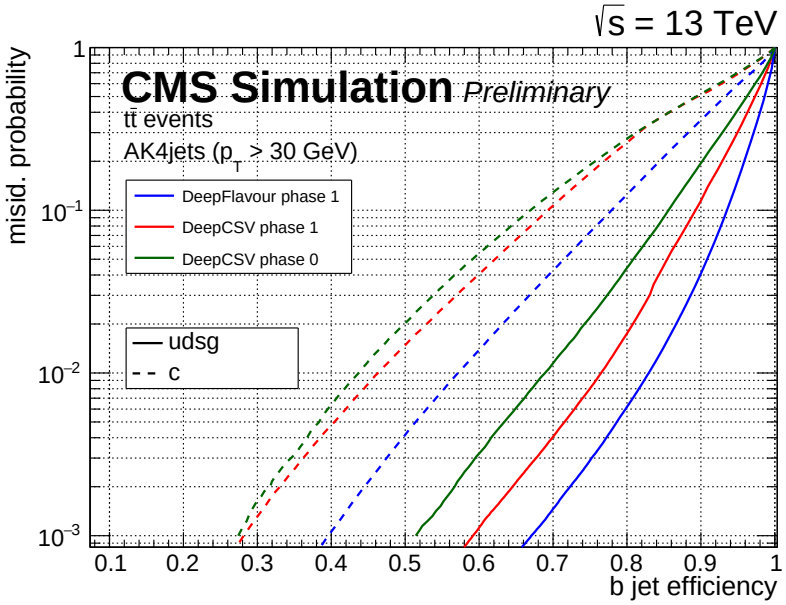

FIGURE 2.18: Performance of the DEEPCSV algorithm. The probability for non-b jets to be misidentified as b jet is shown as a function of the efficiency to correctly identify b jets. The curves are obtained on simulated $t\bar{t}$ events, using jets within $|\eta| < 2.5$ and with $p_T > 30\,\text{GeV}$. The performance is different for the 2016 (Phase 0) and 2017 (Phase 1) detector configuration.

2.3.8 Missing transverse momentum

The missing transverse momentum vector $\vec{p}_T^{\,\text{miss}}$ is computed as the negative vector sum of the transverse momenta of all the PF candidates in an event, and its magnitude is denoted as p_T^{miss} [75]. This vector is modified to account for corrections to the energy scale of the reconstructed jets in the event. It must be emphasized that PU mitigation is essential to ensure an accurate measurement of p_T^{miss}. This is mainly achieved by using the CHS-corrected jet transverse momentum in the aforementioned negative p_T-sum.

Chapter 3

Data samples and simulation

My search for the decay $W \to \pi\gamma$ follows the strategy outlined in Section 1.5.1, and is based on the data collected with the CMS experiment. Events simulated with MC techniques are also studied and used to tune the event selection procedures before they are applied to real data. The *blind analysis* approach, which has physical and historical motivations, has as its main goal to avoid to introduce experimental biases in the signal identification methods. In particular, it is a sound approach when investigating a phenomenon that was never observed before, since, in principle, it is impossible to know exactly how a signal will appear. For this reason, several distributions presented throughout this thesis were first observed in a restricted region of the phase space of the variable in exam. This is the case of the pion-photon invariant mass spectra once the event selection procedure is completed (Fig. 4.21), where a *blind window* between 65 and 90 GeV was applied to collision data before the analysis selection was tuned; or the data distribution in the BDT discriminant (Fig. 4.12), which was restricted to BDT discriminant < 0.2. The validation procedure culminated with the estimate of an expected upper exclusion limit on the branching fraction of $W \to \pi\gamma$, as will be discussed in Chapter 7, whereas the observed limit was only measured afterwards.

This chapter is intended to outline and describe both the collision and the simulated data analyzed in this work.

3.1 Collision data

The data used in this study correspond to an integrated luminosity of $\approx 137\,\text{fb}^{-1}$ recorded by the CMS experiment in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ during 2016 ($35.9\,\text{fb}^{-1}$), 2017 ($41.5\,\text{fb}^{-1}$), and 2018 ($59.7\,\text{fb}^{-1}$). The cumulative day-by-day integrated luminosity for the three years of data-taking is shown in Fig. 3.1. Before being declared suitable for physics analyses, each data sample has to pass strict quality criteria based on the correct functioning of all the CMS subsystems. For this reason, the total integrated luminosity recorded by CMS and displayed in these plots is slightly larger than $137\,\text{fb}^{-1}$.

For the purposes of this search for the decay $W \to \pi\gamma$, I use specific datasets containing events triggered by single, highly energetic muons or electrons.

3.1.1 Luminosity measurement at CMS

A precise measurement of the luminosity delivered to the CMS experiment by the LHC is essential for a variety of reasons. First of all, it provides realtime feedback on both the LHC performance and the CMS operations, such as measurements of trigger rates. Offline, it plays a fundamental role in the determination of the cross sections of observed processes and in the calculation of upper limits. Moreover, any comparison

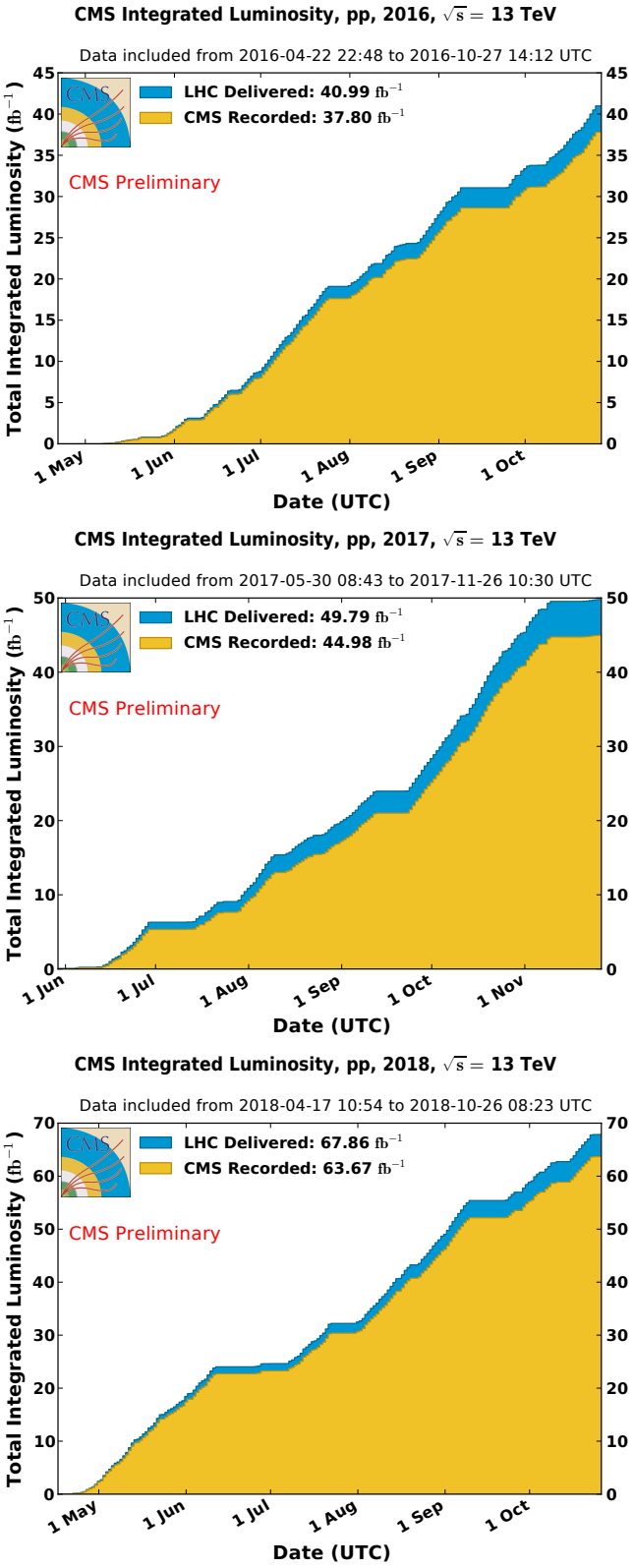

FIGURE 3.1: Cumulative day-by-day integrated luminosity in 2016 (upper), 2017 (central), and 2018 (bottom). The yellow fields correspond to the integrated luminosity recorded by the CMS experiment, to be compared with that delivered by the LHC (blue fields). The latter includes the time from the start of stable beams until the beam is dumped by the LHC (or stable beams end for other reasons, for example for beam studies).

between collision data and simulated events only makes sense if the latter are weighted to match the integrated luminosity of the former.

The luminosity measurement at CMS is carried out using information from dedicated systems, such as the Pixel Luminosity Telescope [76] and the Fast Beam Conditions Monitor [77], and from other subdetectors: the hadronic forward calorimeter, the drift tubes, and the silicon tracker. Each luminometer provides a reading of the rate R of the specific quantities observed in the detector: hits, tracks, clusters, etc. R is proportional to the instantaneous luminosity, $\mathcal{L}_{\text{inst}}$, with the constant of proportionality given by the visible cross section σ_{vis}:

$$R = \mathcal{L}_{\text{inst}}\,\sigma_{\text{vis}} \tag{3.1}$$

In practice, the luminometers usually exhibit some nonlinear dependence on the instantaneous luminosity, the pileup, or on external factors such as the LHC filling scheme; these nonlinearities need to be corrected to obtain an accurate measurement.

σ_{vis} is determined through van der Meer scans, performed with a dedicated LHC machine setup. In these scans, the two beams are separated and then moved across each other, and the resulting measurement of rate as a function of beam separation can be used to derive the beam overlap width. In particular, the instantaneous luminosity for a single bunch collision i, $\mathcal{L}^i_{\text{inst}}$ can be obtained as a function of beam parameters:

$$\mathcal{L}^i_{\text{inst}} = N^i_1 N^2_2 f \int \rho_1(x,y)\rho_2(x,y)\,\mathrm{d}x\,\mathrm{d}y = N^i_1 N^2_2 f \int \rho_{x1}(x)\rho_{x2}(x)\,\mathrm{d}x \int \rho_{y1}(y)\rho_{y2}(y)\,\mathrm{d}y, \tag{3.2}$$

where N^i_1 and N^i_2 are the number of protons in the two colliding bunches, $f = 11246\,\text{Hz}$ is the LHC orbit frequency, and the ρ terms are the normalized particle densities for each proton bunch. The rightmost term of Eq. (3.2) uses the assumption that ρ can be factorized into independent terms in x and y, $\rho_x(x)$ and $\rho_y(y)$, respectively. In addition, the expression above assumes that there is no crossing angle between the beams and that they collide with zero relative separation.

The two overlap integrals can be estimated by varying the beam separation and measuring the resulting rates:

$$\int \rho_{x1}(x)\rho_{x2}(x)\,\mathrm{d}x = \frac{R_x(0)}{\int R_x(\Delta)\,\mathrm{d}\Delta}, \tag{3.3}$$

where $R_x(\Delta)$ is the rate measured when the two beams are separated in x by a distance Δ; a similar equation can be written in y. The beam overlap width (in x, and similarly in y) is then:

$$\Sigma_x = \frac{1}{\sqrt{2\pi}} \frac{\int R_x(\Delta)\,\mathrm{d}\Delta}{R_x(0)}. \tag{3.4}$$

This yields to a final expression for the instantaneous luminosity for a bunch crossing i:

$$\mathcal{L}^i_{\text{inst}} = \frac{N^i_1 N^2_2 f}{2\pi\Sigma_x\Sigma_y}. \tag{3.5}$$

In practice, the integral in Eq. (3.4) is evaluated by performing two separate scans in the x and y directions, measuring the rate (normalized by N^i_1 and N^2_2) at a certain number of separation steps, fitting the resulting points with a functional form, and using the fitted function to determine the overall integral.

The measurement of Σ_x and Σ_y permits to derive σ_{vis} using Eq. (3.2). Then, this measurement can be used to determine the luminosity in regular physics fills of the

LHC, during which the beam overlap width cannot be measured with the necessary accuracy.

3.2 MC simulation

This analysis relies on two kinds of simulated events: a privately produced MC to describe the signal process, and a number of samples accounting for all the possible sources of background that might interfere with the search. The latter were centrally produced by the CMS Collaboration, the most relevant difference with the private production being the total number of generated events, generally larger for the central production. The number of events in the signal MC is nonetheless sufficient to constrain the uncertainties in the signal description to a level acceptable for the purposes of this analysis, as will be discussed in Section 6.2.4.

A second, outstanding distinction lies in the MC generator used for producing each sample. In general, a certain generator may be preferred over another depending on the accuracy needed in the description of a specific process. For instance, MC generators may differentiate based on the order in perturbative QCD achieved in the matrix element calculation, which describes the probability of transition from the initial state to a particular final state of particles, or on the way they model the parton shower (see Section 1.4.1).

Regardless of the generator used, all the events are subsequently processed through a simulation of the CMS detector based on GEANT4 [78] and reconstructed using the same algorithms as used for collision data, described in Section 2.3.

3.2.1 Signal

The signal process is simulated at leading order (LO) in perturbative QCD using the MC event generator PYTHIA 8.226 (8.230) for 2016 (2017 and 2018) [79]. About 80 thousand events per year of data-taking were generated, in which a pair of opposite-sign W bosons is produced from $\mathrm{t\bar{t}}$ decay, together with two b quarks. One of the two vector bosons is then forced to decay into either a muon, an electron (or positron) or a τ lepton, plus the relative (anti)neutrino. In the last case, the τ lepton decay may only proceed as $\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$ or $\tau^- \to \mathrm{e}^- \bar{\nu}_\mathrm{e} \nu_\tau$ (or the charge conjugated process). Finally, the other W boson in the event decays into a charged pion and a photon.

Each generated event is attributed the following cross section:

$$\sigma_{\mathrm{W}\to\pi\gamma} = \sigma_{\mathrm{t\bar{t}}} \times \mathcal{B}(\mathrm{W}^\mp \to \ell^\mp \nu) \times \mathcal{B}(\mathrm{W}^\pm \to \pi^\pm \gamma), \qquad \ell = \mu, \mathrm{e}, \tau_\ell \tag{3.6}$$

In the expression above, $\sigma_{\mathrm{t\bar{t}}} = 815\,\mathrm{pb}$ indicates the $\mathrm{t\bar{t}}$ production cross section in pp collisions at 13 TeV, as measured by the CMS Collaboration using a $\mathcal{L}_\mathrm{int} = 2.2\,\mathrm{fb}^{-1}$ in 2015 [80]; $\mathcal{B}(\mathrm{W}^\mp \to \ell^\mp \nu) = 0.1086\,(2{+}0.3521)$ uses the most up-to-date measurements of the branching fraction of the W boson into leptons [17], with the factor $(2+0.3521)$ accounting for the possible decay into a muon, an electron or a τ subsequently decaying into leptons (τ_ℓ); the branching ratio $\mathcal{B}(\mathrm{W}^\pm \to \pi^\pm \gamma)$ can be arbitrarily chosen to reproduce any desired scenario.

Given their relatively large production cross section at CMS and their similarity to the signal, it should not surprise that the main sources of background for this search are $\mathrm{t\bar{t}}$-based processes, typically with a charged lepton being misidentified as a pion and an energetic photon being produced via Bremsstrahlung or in the formation of a hadronic jet. As discussed in Section 3.2.2, these processes are simulated at NLO in the CMS central production, and thus a difference in the transverse momentum of

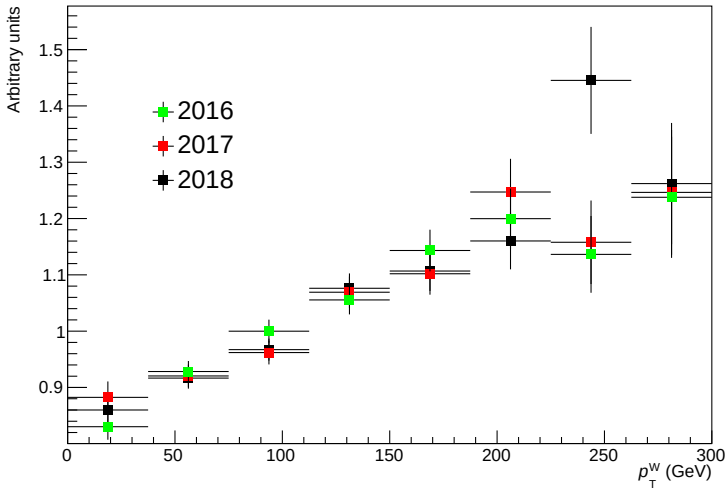

FIGURE 3.2: Sets of weights applied to the signal to make the p_T^W distribution match that of an NLO-simulated $t\bar{t}$ sample, for the three years of data-taking. For $p_T^W > 300\,\text{GeV}$, the value of the last bin is used.

the generated W boson (p_T^W) is expected. To make this important kinematic property match fairly between the different generators, the signal is reweighted so that the p_T of the W boson decaying into a pion and a photon matches that obtained in an NLO $W \to \ell\nu$ ($\ell = \mu, e$) simulation in POWHEG v2.0 [81–84], where the W boson originates from $t\bar{t}$ decay. Note that the decay into τ leptons is not considered, since the p_T^W is extracted at generation level (i.e., before the simulation of the interaction with the detector) and the statistics from the muon and electron channels is sufficient. The sets of weights, one per year of data-taking, are obtained as a ratio of the p_T^W from the NLO simulation to the p_T^W as modeled by PYTHIA, and are shown in Fig. 3.2.

At this point, it must be noted that the main reason why PYTHIA was chosen to simulate signal events is the relative simplicity of the configuration of this generator, and its straightforward integration with the CMS analysis framework. If a simulation which made use of POWHEG could have avoided the reweighting procedure just described, the systematic uncertainty associated to the corrections obtained with this method is very modest (see Section 6.2.3).

In addition, it must be stressed that W bosons arising from $t\bar{t}$ decay are produced in a mixture or longitudinal, left-, and right-handed polarization states (the latter being strongly suppressed in the Standard Model). This characteristic, which also reflects into the kinematic properties of their decay products, is not modeled by PYTHIA in my signal production. Therefore, a systematic uncertainty is evaluated and accounted for, as discussed in Section 6.2.3.

3.2.2 Backgrounds

A number of background sources must be considered for a proper description of the real situation to be found in data, carefully accounting for all the processes leading to final states that can fake a signal. From this accurate description depends the effectiveness of the event selection procedure described in the next chapter.

The main background processes in this analysis are non-signal $t\bar{t}$ production, Drell–Yan events ($q\bar{q} \to \gamma^*/Z^* \to \ell^+\ell^-$), associated production of a vector boson and a photon or jets ($W\gamma \to \ell\nu\gamma$, $Z\gamma \to \ell\ell\gamma$ and $W + \text{jets} \to \ell\nu + \text{jets}$), along with Standard

Model events comprised uniquely of jets produced through the strong interaction, referred to as QCD multijet events. As already mentioned, the $t\bar{t}$ backgrounds are simulated at NLO using the POWHEG v2.0 framework, while the Drell–Yan, $W\gamma \to \ell\nu\gamma$, $Z\gamma \to \ell\ell\gamma$, $W + \text{jets} \to \ell\nu + \text{jets}$, and QCD multijet processes are simulated at NLO using the MADGRAPH5_aMC@NLO2.2.2 (2.4.2) generator for 2016 (2017 and 2018) [85].

The NNPDF 3.0 NLO [86] (NNPDF 3.1 next-to-NLO [87]) parton distribution functions are used for generating all 2016 (2017 and 2018) MC samples. The generators are interfaced to PYTHIA for parton showering and parton fragmentation. The PYTHIA parameters affecting the description of the underlying event are set to the CUETP8M1 tune [88] (CP5 tune [89]) for the 2016 (2017 and 2018) simulation. For the 2016 $t\bar{t}$ backgrounds, the CP5 tune is used instead of CUETP8M1.

All the simulated background samples used, summarized in Table 3.1, are normalized to the cross sections obtained from the corresponding event generator. The equivalent integrated luminosity is also shown, calculated as the ratio between the number of generated events per sample and the cross section corresponding to the described process.

3.2.3 Event weight

Each of the MC samples considered in this analysis contains a certain number of generated events. In order to compare the different background processes, the signal, and the collision data, it is thus fundamental to properly renormalize the simulated events. This is achieved by multiplying their cross section ($\times$ branching fraction) by the following terms:

- $\mathbf{W_{MC}}$: the MC generation weight. Some of the generators (e.g., MADGRAPH5) implement a matching with the parton shower through perturbative matrix-element computations, which are carried out at NLO. As a drawback for the high precision achieved, the calculated NLO cross sections might not positive-definite locally in the phase space. This implies that some of the hard events that will eventually be showered have negative weights, and can only be "removed" via compensation with positive events. Therefore, $W_{MC} = +1$ for a number of positive-definite events (n_+) and $= -1$ for a number of negative-definite events (n_-);

- $\mathbf{1/n_{evts}}$, where $n_{evts} = n_+ - n_-$, referring to the total number of events in the sample;

- $\mathcal{L}_{\mathbf{int}}$: the reference integrated luminosity, essential in to compare the simulation with collision data;

- $\mathbf{W_{PU}}$: the pileup weight. This is necessary when an assumption is made regarding the PU distribution in the Monte Carlo generation, and it turns out to not correspond to the one measured in real pp collisions. Therefore, W_{PU} is simply obtained as the ratio between the PU distributions in data and in the MC. The PU distributions measured in 2016–2018 are shown in Fig. 3.3, and show an average of about 23 (32) interactions per bunch crossing in 2016 (2017 and 2018);

- $\mathbf{W_{PreFiring}}$. In 2016 and 2017, the gradual timing shift of the ECAL was not properly propagated to the L1 trigger, resulting in a significant fraction of triggering signals at large η being mistakenly associated to the previous bunch

TABLE 3.1: Background samples used in this analysis, together with their cross sections (times branching ratios) expressed in pb, and the equivalent integrated luminosity expressed in fb^{-1}. All the initial states must be considered as arising from a pp collision. Where the decay of an unstable particle is not specified, it means that it may decay via all known channels.

Process	$\sigma \times \mathcal{B}$ **(pb)**	**Equiv.** $\mathcal{L}_{\mathrm{int}}$ (fb^{-1})
$\mathrm{t\bar{t}} \to \mathrm{W^+bW^-\bar{b}}$, $\mathrm{W} \to q\bar{q}'$	378.0	819.5
$\mathrm{t\bar{t}} \to \mathrm{W^+bW^-\bar{b}}$, $\mathrm{W^+} \to q\bar{q}'$, $\mathrm{W^-} \to \ell\nu$	365.3	962.4
$\mathrm{t\bar{t}} \to \mathrm{W^+bW^-\bar{b}}$, $\mathrm{W} \to \ell\nu$	88.3	1604.2
$\mathrm{t\bar{t}W}$ + jets, $\mathrm{W} \to q\bar{q}'$	0.4	6200.0
$\mathrm{t\bar{t}W}$ + jets, $\mathrm{W} \to \ell\nu$	0.2	75935.2
$\mathrm{t\bar{t}Z}$, $\mathrm{Z} \to q\bar{q}$	0.5	22280.8
$\mathrm{t\bar{t}Z}$, $\mathrm{Z} \to \ell^+\ell^-$	0.3	116590.3
$\mathrm{t(\bar{t})W}$	35.9	5982.8
$q\bar{q} \to \gamma^*/\mathrm{Z}^* \to \ell^+\ell^-$, $10 < m_{\ell^+\ell^-} < 50\,\mathrm{GeV}$	18810.0	12.5
$q\bar{q} \to \gamma^*/\mathrm{Z}^* \to \ell^+\ell^-$, $m_{\ell^+\ell^-} > 50\,\mathrm{GeV}$	6225.4	84.3
γ + jets	4269.6	71.5
WW, $\mathrm{W} \to q\bar{q}'$	47.7	757.5
WZ	27.6	427.9
W + jets, $\mathrm{W} \to \ell\nu$	61731.0	22.1
$\mathrm{W}\gamma$ + jets, $\mathrm{W} \to \ell\nu$	191.1	421.6
$\mathrm{t\bar{t}}\gamma$ + jets	4.1	9511.6
$\mathrm{Z}\gamma$, $\mathrm{Z} \to \ell^+\ell^-$	123.8	475.0
QCD multijet	29648084.0	0.03

crossing. Because of the L1 rules, this led to several events self-vetoing when a significant amount of energy was found in the region of $2 < |\eta| < 3$ of the ECAL. This effect is not described by the simulation, and thus the $\mathrm{W_{PreFiring}}$ correction factor, based on the probability of the particles contained in an event to cause pre-firing, needs to be applied;

- $\mathbf{W_{z-Vtx}}$. There is an allowed region along the beam axis (z) for an electron to be accepted by the HLT. An issue during the 2017 data-taking caused this region along z to be too small, resulting in a reduction of the efficiency for events triggered by electrons. Since this effect is not described by the simulation, a fixed penalty $\mathrm{W_{z-Vtx}} = 0.991$ needs to be applied to all the MC events triggered by electrons;

- **SFs**: scale factors. With this general definition, all the additional correction factors with respect to the ones already mentioned, aiming to compensate the

dissimilarities between the conditions used for the MC generation and those observed in the data-taking. In particular, this analysis accounts for trigger, identification, and reconstruction efficiency SFs related to muons, electrons and photons, as will be discussed in Section 4.2.10.

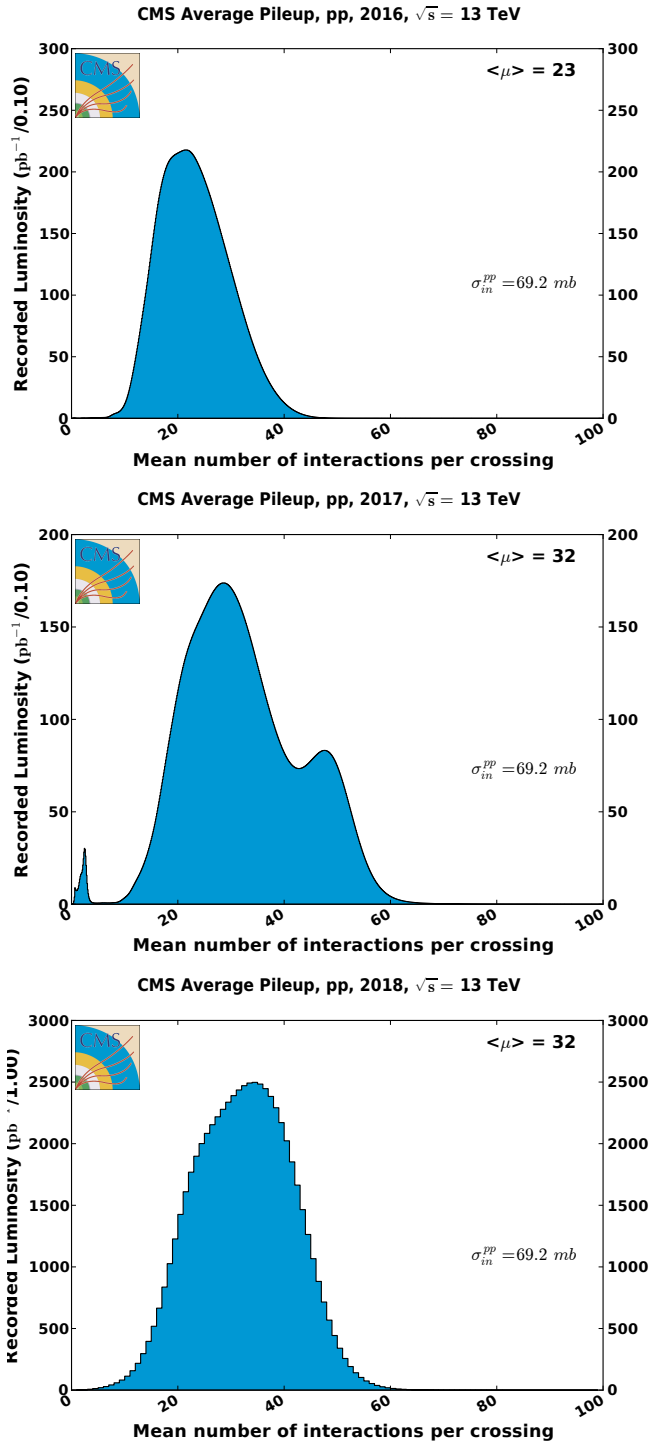

FIGURE 3.3: Mean number of interactions per bunch crossing (PU) for the 2016 (top), 2017 (central), and 2018 (bottom) pp run at 13 TeV. The values refer to a minimum bias cross section (i.e., measured using events selected with a minimally-biased trigger) of 69.2 mb, which is determined by finding the best agreement with data.

Chapter 4

Event selection

The event selection, which is designed and tuned on the MC simulation before inspecting the region sensitive to signal in collision data, relies on different stages:

- trigger selection;
- cut-based preselection;
- multivariate selection.

The final goal of these selection steps is to retain a number of good candidate events for the decay $W \to \pi\gamma$, by reconstructing the pion-photon invariant mass ($m_{\pi\gamma}$). A fit to $m_{\pi\gamma}$ is then used to extract $\mathcal{B}(W \to \pi\gamma)$. In this chapter the procedures adopted to identify candidate $W \to \pi\gamma$ events will be described in depth.

4.1 Trigger selection

The data acquisition uses unprescaled HLT paths for single muons and single electrons exceeding certain p_T thresholds. Such thresholds vary for different years, according to specific data-taking conditions. A full list of the triggers used is shown in Table 4.1.

TABLE 4.1: HLT paths used as a first event selection in the search for the decay $W \to \pi\gamma$.

2016 HLT path	**Request**
`HLT_IsoMu24`	Isolated muon with $p_T > 24\,GeV$
`HLT_IsoTkMu24`	Isolated muon (loose) with $p_T > 24\,GeV$
`HLT_Mu50`	Muon with $p_T > 50\,GeV$
`HLT_Ele27_WPTight_Gsf`	Electron with $p_T > 27\,GeV$
2017 HLT path	**Request**
`HLT_IsoMu27`	Isolated muon with $p_T > 27\,GeV$
`HLT_Mu50`	Muon with $p_T > 50\,GeV$
`HLT_Ele32_WPTight_Gsf`	Electron with $p_T > 32\,GeV$
2018 HLT path	**Request**
`HLT_IsoMu24`	Isolated muon with $p_T > 24\,GeV$
`HLT_Mu50`	Muon with $p_T > 50\,GeV$
`HLT_Ele32_WPTight_Gsf`	Electron with $p_T > 32\,GeV$

For what concerns the muon triggers, a logical *OR* is applied among the different paths. The path marked with `Tk` is based on the reconstruction of tracker muons, with an algorithm that only looks for a hit confirmation in the muon system (see the definition in Section 2.3.3). Electron triggers adopt a high-purity (where "purity" stands for probability that a particle identified as an electron be in fact an electron)

identification working point, and are reconstructed through the Gaussian-sum filter (GSF) algorithm described in Section 2.3.4. The efficiencies of all these HLT paths were measured by the CMS Collaboration yearly, and are shown in Fig. 4.1 and 4.2.

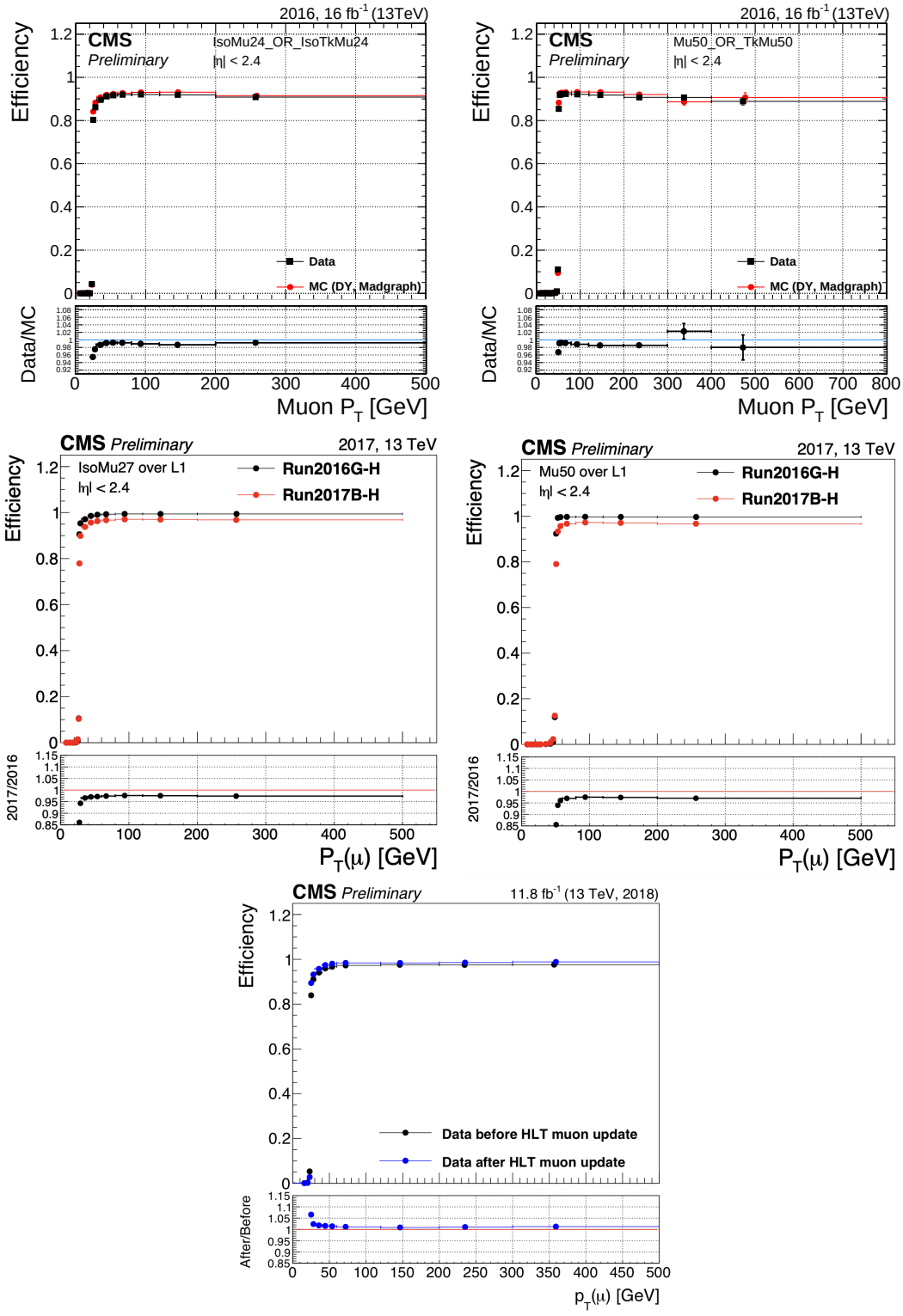

FIGURE 4.1: Muon trigger efficiency as a function of the muon p_{T} for the 2016 (upper), 2017 (central), and 2018 (lower) HLT paths used.

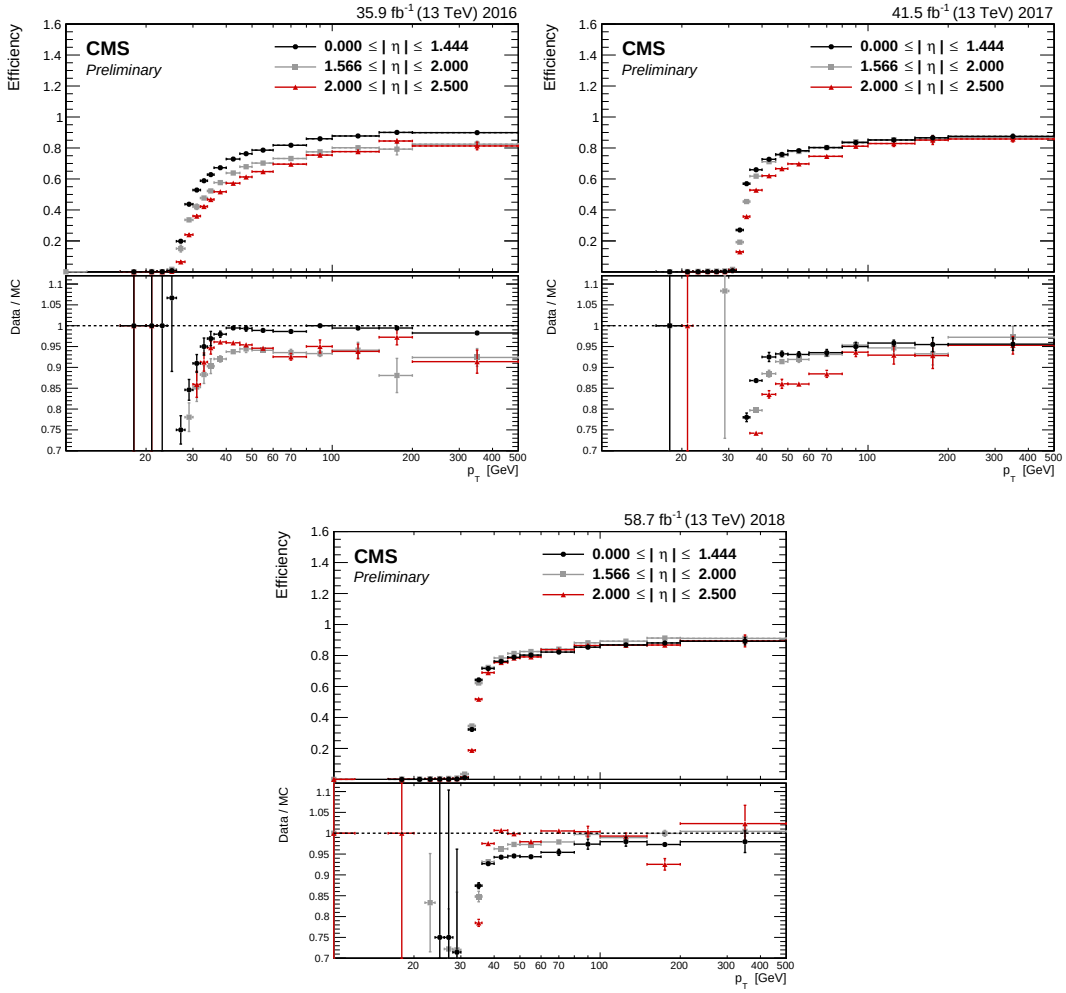

FIGURE 4.2: Electron trigger efficiency as a function of the electron p_{T} for 2016 (top left), 2017 (top right), and 2018 (bottom). The 2016 data points refer to HLT_Ele27_WPTight_Gsf, while the 2017 and 2018 data points refer to HLT_Ele32_WPTight_Gsf. Non-negligible differences in efficiency between the years, especially at large η, depend on the different p_{T} threshold (between 2016 and the other two years) and on modifications of the definition of the tight working point. The disagreement between data and MC, in particular at low p_{T}, is partially due to detector effects difficult to simulate, such as crystal transparency losses in the electromagnetic calorimeter and the evolution of dead regions in the pixel tracker.

4.2 Preselection

After the trigger selection, the events undergo a further skimming process: the preselection. This step aims at choosing the basic topological properties of signal-like events, and particles must fulfill a number of criteria in order to be considered as participants in the events of interest.

This procedure is applied to each particle expected in this topology. At first, a charged lepton must be identified, then a charged hadron and, eventually, a photon with certain properties. Ambiguities due to multiple candidates for the same object are removed with best-candidate criteria. Variables of interest related to these particles and general event properties are saved for further analysis.

4.2.1 Muons

Muons identified using the criteria described in Section 2.3.3 are considered if above a minimum p_{T} threshold of 20 GeV. The muon track should be within a distance $|d_{\mathrm{z}}| < 0.5\,\mathrm{cm}$ from the PV along the beam axis, and $|d_{\mathrm{xy}}| < 0.2\,\mathrm{cm}$ in the transverse plane. A muon originating from a W boson decay is expected to produce a track that is isolated from those of the other particles generated in the pp collision. Therefore, an isolation check that exploits the PF technique (see Section 2.3.2) is applied, with the requirement that:

$$\left[I_{\mathrm{h}^{\pm}} + \max\left(0,\ I_{\mathrm{h}^{0}} + I_{\gamma} - 0.5\, I_{\mathrm{h}^{\pm}_{\mathrm{PU}}}\right)\right] / p_{\mathrm{T}}^{\mu} < 0.25, \tag{4.1}$$

In the expression above, $I_{\mathrm{h}^{\pm}}$ indicates the sum over the p_{T} of the charged hadrons, $I_{\mathrm{h}^{0}}$ the neutral hadrons, I_{γ} the photons, and $I_{\mathrm{h}^{\pm}_{\mathrm{PU}}}$ the hadronic particles arising from PU. Because of the normalization to the muon p_{T}, this formula represents in fact a *relative isolation*, since the proper isolation and the transverse momentum were decoupled.

The search for a suitable muon candidate proceeds in parallel with that for an electron candidate: if no muon passes the selection, the event is not discarded, but the electron selection starts. It is important to underline that the latter takes place regardless that a suitable muon candidate is found. Since the topology of signal events is characterized by one and one only muon or one electron from W-boson decay (or from the decay of a τ lepton), events containing more than a charged lepton passing the muon or electron preselection cutoffs are subject to a veto procedure described in Section 4.2.3.

4.2.2 Electrons

Electron candidates identified as described in Section 2.3.4 are inspected for $|\eta| < 2.4$ and above a p_{T} threshold of 20 GeV. As in the muon case, the electron track should be within a distance $|d_{\mathrm{z}}| < 0.5\,\mathrm{cm}$ from the PV along the beam axis, and $|d_{\mathrm{xy}}| < 0.2\,\mathrm{cm}$ in the transverse plane. Residual energy corrections are applied to scale the data to the MC simulation and smear the latter to the resolution matched in data, using combined information from the tracker and the ECAL.

4.2.3 Multi-lepton veto

In principle, the largest-p_{T} muon or electron that satisfies all the aforementioned conditions would be selected as the best candidate to tag the leptonic decay of the W boson. This choice appears to be necessary if two or more charged leptons fulfill all the preselection criteria, and reasonable since the decay products of the W boson are expected to be highly energetic, due to its large mass. In order to understand whether this is a sound approach, a simulation-based test is performed to correlate the charged lepton originating from the W boson decay and the multiplicity of charged leptons in each signal event. The test uses information at generation level (the so-called *MC truth*) to know if a final state contains a muon or an electron from one of the decayed W bosons (or from the decay of a τ lepton arising from a W boson). Given this "tag lepton", the total number of electrons and muons passing the preselection cutoffs is counted. The results are shown in Fig. 4.3, where the signal MC samples corresponding to the three years of data-taking are merged. When the tag lepton is a muon (top plot), almost all events show only one muon in the final state, whereas the number of events containing electrons surviving the preselection is at least one order of magnitude smaller. Similarly, when the tag lepton is an electron (bottom plot),

the majority of the reconstructed events presents only one electron in the final state, while the number of events containing two or more electrons is at least one order of magnitude smaller. Moreover, a negligible number of electron-tagged events contains muons (one or more) surviving the preselection. This provides a clear guideline:

- events with more than one lepton of the same family must be discarded;
- events with both a muon and an electron passing the respective selection cutoffs must be discarded as well.

The combination of these two conditions leads to the following veto: one and only one charged lepton fulfilling the preselection requirements is allowed within an event. On an operational side, an important consequence is that any event can be classified as belonging to either the *muon* or the *electron channel*.

In addition, the electrons identified with a looser working point of the MVA described in Section 2.3.4 are counted. The signal efficiency for this latter working point is about 98%. In case the tag lepton is an electron, the event is discarded if the number of *loose* electrons is greater than one. This further skimming in the electron channel proved effective in the suppression of some background sources involving electron misidentification with photons, especially the Drell–Yan.

4.2.4 Trigger matching

In order to simplify the trigger efficiency measurement, a match is required between the reconstructed lepton surviving the multi-lepton veto and the particle that actually caused one of the triggers listed in Table 4.1 to be lit. The matching is successful if:

$$\Delta R(\ell_{\text{reco}}, \ell_{\text{trigger}}) < 0.1 \qquad \ell = \mu, \text{e}, \tag{4.2}$$

and the event is discarded otherwise.

4.2.5 Pion candidates

Because of the absence of a dedicated detector for particle identification within the CMS experiment, every track is in principle compatible with that of a charged pion. Nevertheless, the PF algorithm permits to use the energy deposits in the calorimeters and the hits in the tracker and the muon system to distinguish leptons from charged and neutral hadrons. The search for good candidate pions is thus performed among all the charged hadrons with $p_{\text{T}} > 20\,\text{GeV}$, reconstructed through the PF. Since the pion originating from the rare decay $\text{W} \to \pi\gamma$ carries opposite charge with respect to the lepton arising from the other W boson in the event, only candidates satisfying this charge sign requirement are considered. In order to reduce the fake rate, that is the fraction of reconstructed tracks that are not associated with a charged particle, the candidate must produce a *high-purity* track in the silicon tracker [66]. Such requirement implies that the track must fulfill a series of stringent criteria on the minimum number of layers hit in the silicon tracker, on the χ^2/ndof of the fit used for the track reconstruction, and on the compatibility with the primary vertex. These complex criteria are a function of the track η and p_{T}, and vary according to the iteration of the reconstruction algorithm discussed in Section 2.3.6.

In similarity with the muon and the electron cases, such track must be within $|d_{\text{z}}| < 0.5\,\text{cm}$ and $|d_{\text{xy}}| < 0.2\,\text{cm}$ from the PV, to help rejecting pions produced from in-flight decays of other particles. Finally, the candidate that carries the largest p_{T} and, at the same time, fulfills all the aforementioned requirements, is selected for further analysis.

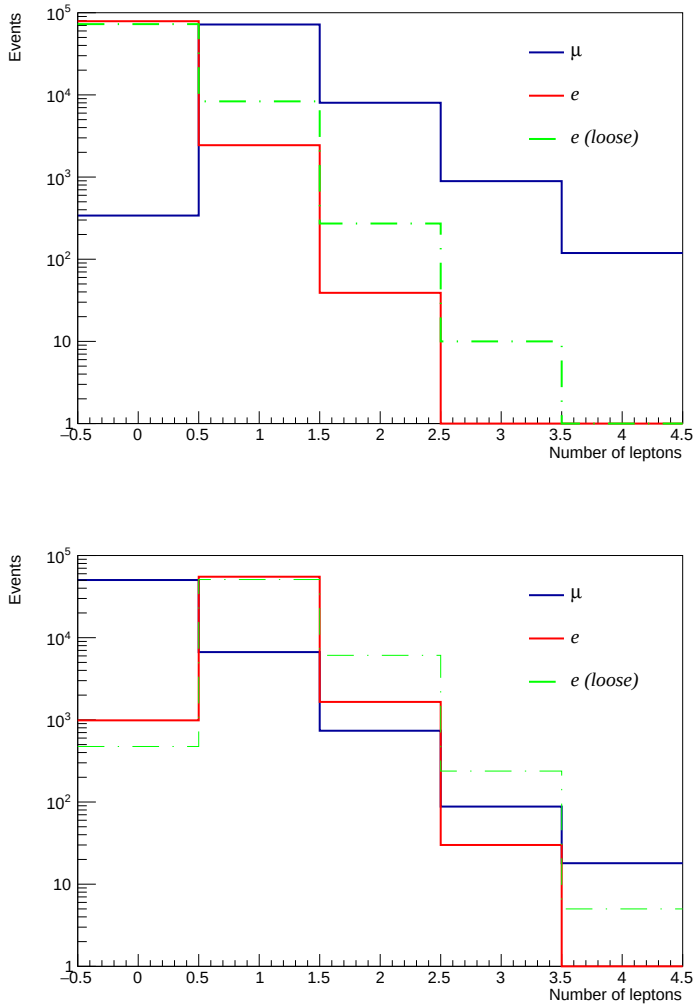

FIGURE 4.3: Charged lepton multiplicity in the case of a muon (top) and an electron (bottom) tag lepton. The plots contain the merging of the 2016–2018 signal generation with PYTHIA.

4.2.6 Pion isolation

Since the pion from the $W \to \pi\gamma$ decay originates from an electroweak process, any jet activity in its vicinity is expected to be minimal. This feature of the signal topology offers a solid handle for background discrimination, but no technique for estimating the isolation of charge hadrons exists in the standard CMS analysis framework. For this reason, this analysis implements a novel pion-isolation variable based on the sum of the p_T of all the PF candidates contained in a cone with radius ΔR around the candidate pion selected as discussed in the previous section, divided by the pion p_T^{π} ($\Sigma p_T / p_T^{\pi}$). The cone is defined to be $0.02 < \Delta R < 0.5$, with the lower bound helping to exclude from the p_T-sum the single charged particles generating two nearby reconstructed tracks (a phenomenon known as *split track*). To reduce the contribution from PU, charged particles are required to originate from the PV, with the same d_{xy} and d_z criteria as used for the pion. Figure 4.4 shows the distribution of the pion-isolation variable for signal events, which appear to be highly isolated, while background events indicate a clear presence of jet activity. The MC description is suboptimal for isolated background events. This discrepancy, independent of the data-taking period, does not affect the background characterization extracted from data, but somewhat reduces the

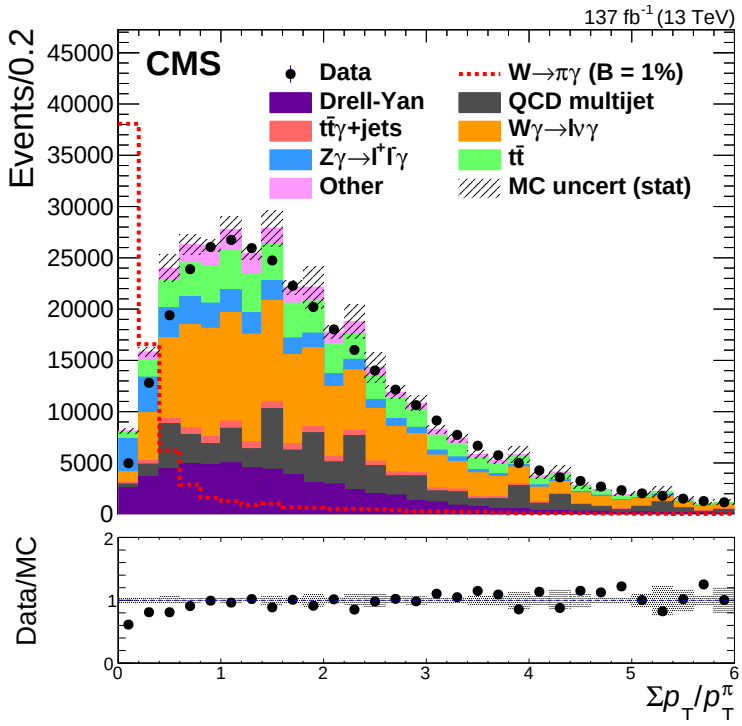

FIGURE 4.4: Event distribution as a function of the pion-isolation variable $\Sigma p_{\mathrm{T}}/p_{\mathrm{T}}^{\pi}$. The simulated MC distribution for the signal is given by the dashed red line and corresponds to a 1% branching fraction for the $\mathrm{W} \to \pi\gamma$ decay. The statistical uncertainties in the data are small and thus not visible. In the lower panel, the ratio between data and the background component of the MC is shown. The gray bands represent the statistical uncertainty in the MC background.

effectiveness of the multivariate selection technique described in Section 4.3, which is trained on simulated events. As far as the signal is concerned, this discrepancy is accounted for in the systematic uncertainty as discussed in Section 6.2.3.

Even though it is computed at preselection level, no direct cutoff is applied to the pion isolation, which is then included among the input variables of a multivariate classifier.

4.2.7 Photons

In similarity with the candidate pion selection, the p_{T} of the photon, which is identified using the criteria described in Section 2.3.5, must be >20 GeV. In order to discriminate between prompt and bremsstrahlung photons arising from the interaction of electrons with the inner tracker material, photons are removed if their calorimetric energy deposit matches with tracks in the silicon pixel detector. Residual energy corrections are applied to scale the data to the MC simulation and smear the latter to the resolution matched in data, using information from the ECAL. The photon that carries the largest p_{T} and, at the same time, fulfills all the aforementioned requirements, is selected for further analysis.

At this stage of the preselection, a suitable candidate pion and a photon have been identified. Thus, the event is retained only if the reconstructed value of the main observable is $50 < m_{\pi\gamma} < 100$ GeV.

4.2.8 b jets

The number n_{b} of b-tagged jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and identified as described in Section 2.3.7 is an important element for signal identification. Similarly to the pion isolation, no threshold is applied to this variable at preselection level, but it is preferred to use it as input of the multivariate classifier (see Section 4.3.2).

Due to an increased noise in some sectors of one of the ECAL endcaps in 2017, jets with $p_{\mathrm{T}} < 50\,\mathrm{GeV}$ and $2.65 < \eta < 3.14$ must be excluded from the count in the 2017 samples. In addition, a low performance in the calibration of the jet energy during part of the 2018 data-taking must be taken into account, due to the power supply failure of two HCAL sectors. In order to assess the its impact, the jet p_{T} in the 2018 MC simulation is scaled downwards by 20% if $-1.57 < \phi_{\mathrm{jet}} < -0.87$ and $-2.5 < \eta_{\mathrm{jet}} < -1.3$, and by 35% if $-1.57 < \phi_{\mathrm{jet}} < -0.87$ and $-3.0 < \eta_{\mathrm{jet}} < -2.5$. A comparison is then made between the number n_{b} computed with and without this energy scaling. The discrepancy is found to range from ≈1% for events with $n_{\mathrm{b}} = 0$, to ≈8% for events with $n_{\mathrm{b}} = 5$, as shown in Fig. 4.5. This observed effect is treated as a systematic uncertainty in this analysis, as discussed in Section 6.2.3.

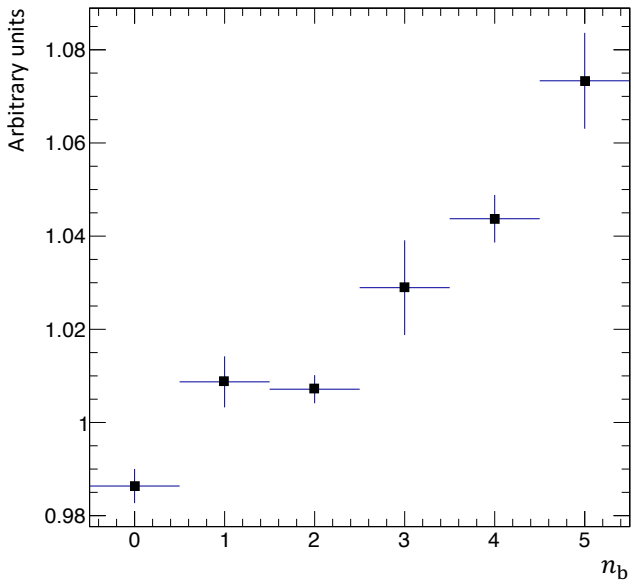

FIGURE 4.5: Ratio between the non-scaled and the scaled distributions of the number of b-tagged jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$.

4.2.9 Post-preselection cutoffs

After the preselection and before the multivariate selection, a further series of cutoffs is applied. First of all, the request on the minimum lepton p_{T} threshold, which was 20 GeV at preselection stage, is incremented to be above the *turn on* of the curves of muon and electron trigger efficiency (see Fig. 4.1 and 4.2). The choice of incrementing these thresholds only at this stage is meant to increase the effectiveness of the multi-lepton veto. A summary of the minimum muon and electron p_{T} thresholds (p_{T}^{μ} and $p_{\mathrm{T}}^{\mathrm{e}}$, respectively) after this step can be found in Table 4.2.

To reduce contamination from pion candidates misidentified as leptons, the requirements $\Delta\phi(\mu, \pi) > 0.09$ and $\Delta\phi(\mathrm{e}, \pi) > 0.05$ are imposed on the angle between the lepton and the candidate pion. In principle, the request a pion candidate's charge sign opposite to that of the lepton would make these additional cutoffs unnecessary,

TABLE 4.2: Minimum p_T thresholds for muons and electrons.

Year	p_T^{μ} (GeV)	p_T^{e} (GeV)
2016	25	30
2017	28	33
2018	25	33

but the $\Delta\phi$ thresholds become effective in case of lepton or hadron charge misidentification in the track refit. Similarly, the condition $\Delta\phi(\ell,\gamma) > 0.04$ is used to reduce the misidentification of the trigger lepton as the selected photon. The choice of these thresholds is made by inspecting the respective $\Delta\phi$ distributions and removing the regions where the lepton and the candidate pion or the photon overlap, as shown in Fig. 4.6.

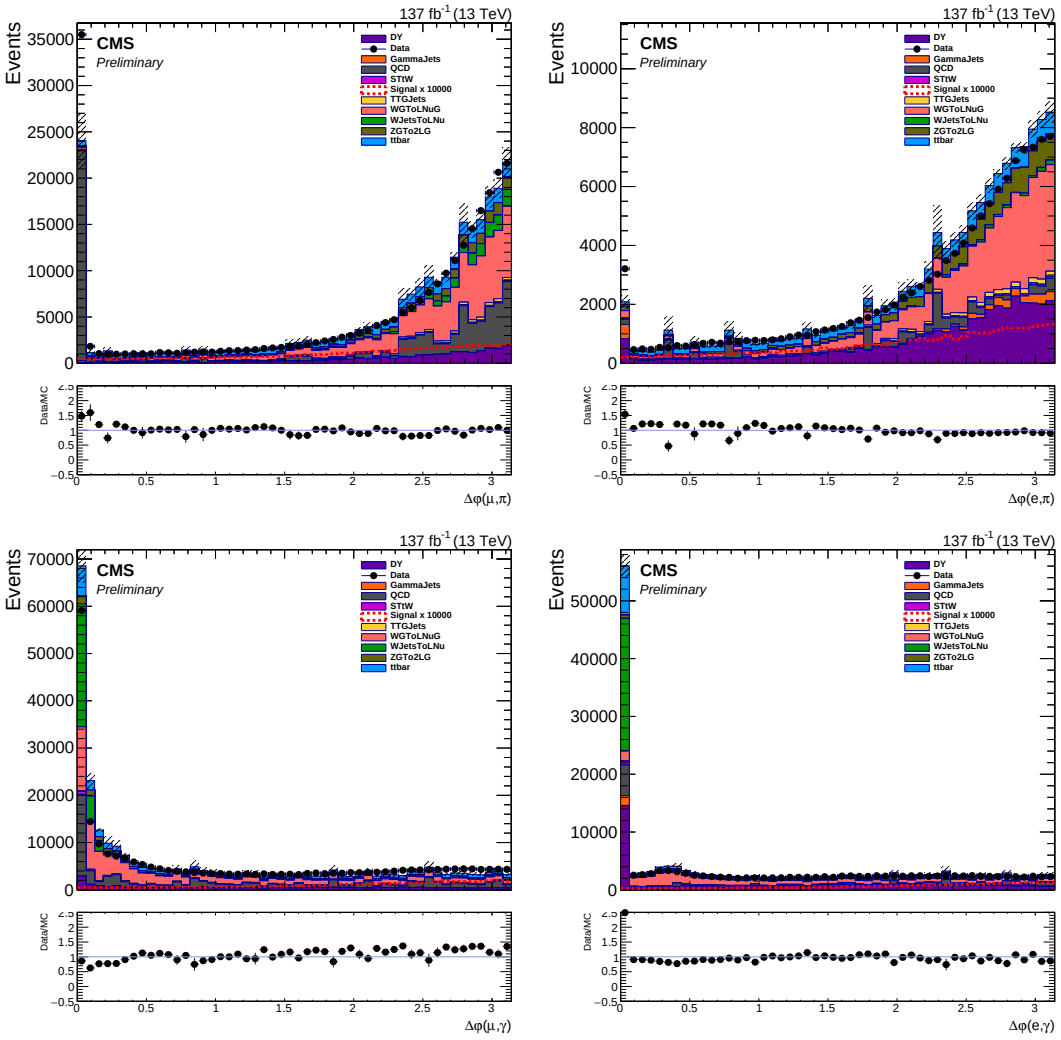

FIGURE 4.6: Distributions of $\Delta\phi(\mu,\pi)$ (top left), $\Delta\phi(e,\pi)$ (top right), $\Delta\phi(\mu,\gamma)$ (bottom left), $\Delta\phi(e,\gamma)$ (bottom right) after the preselection. In the top plots, a threshold $\Delta\phi(\ell,\gamma) > 0.04$ is accounted for, while in the bottom plots the cutoffs $\Delta\phi(\mu,\gamma) > 0.09$ and $\Delta\phi(e,\gamma) > 0.05$ are applied. The first bin in each plot clearly shows an overlap between the trigger lepton and the candidate pion or the photon. The signal corresponds to a $\mathcal{B}(W \to \pi\gamma) = 1\%$.

4.2.10 Scale factors

As anticipated in Section 3.2.3, the weight of the simulated events contains scale factors used to correct the conditions used in the MC production to those observed during the data-taking. Within the CMS Collaboration, the SFs are usually derived centrally by experts of the various physics objects they refer to, and account for data/MC discrepancies in the efficiency of the trigger system, the particle identification, reconstruction, and isolation techniques etc. As an exception, the electron trigger efficiency SFs (Fig. 4.7) are not centrally computed, and I have therefore derived them with a *Tag and Probe* method, which is discussed in Appendix A. In general, the SFs are p_{T} and η dependent, improving the accuracy of the correction. A full list of the SFs applied to the particles involved in the search for $\mathrm{W} \to \pi\gamma$ can be found in Table 4.3.

Table 4.3: Scale factors used in this analysis to correct for discrepancies in the data/MC conditions, divided by particle type.

Particle	Correction
μ	ID
	Reconstruction
	Isolation
	Trigger
e	ID
	Reconstruction
	Trigger
γ	ID
	Pixel Veto

4.2.11 Preselection overview

All the cutoffs applied to the kinematic variables and the characteristics of the particles analyzed in the preselection are summarized in Table 4.4. The p_{T} and η distributions of the leptons, candidate pions, and photons selected up to this stage are shown, for signal only, in Fig. 4.8 and 4.9.

The signal events surviving the preselection in the muon and the electron channels are about 30 and 20 thousand, respectively, considering the sum of the 2016–2018 periods. These can be compared with an initial number of signal events of ≈240000, summing the two channels and the three years, and the signal efficiency can be calculated as the ratio between the number of signal events surviving the preselection and the initial number of events in the MC sample. It must be stressed again that the two lepton channels receive a contribution from both events with one W boson decaying directly into a muon or an electron plus a neutrino, and from events with a W boson decaying into a τ lepton that subsequently decays leptonically. Therefore, any calculation of signal efficiency must be properly corrected to account for the different branching ratios of these two kinds of processes. This is achieved by multiplying the number of initial simulated events by a factor $(1 + 1 + 0.3521)/3$, where the two factors 1 correspond to the $\mathrm{W} \to \mu\nu$ and $\mathrm{W} \to \mathrm{e}\nu$ cases and 0.3521 follows from $\mathcal{B}(\tau \to \mu\nu\bar{\nu}) + \mathcal{B}(\tau \to \mathrm{e}\nu\bar{\nu}) \simeq 0.3521$ [17]. Based on these numbers, the product of the MC signal efficiency after the preselection and the detector acceptance is summarized in Table 4.5, divided per year and per lepton channel (ε_μ, ε_{e}). The uncertainty in the efficiency is considered to be binomial.

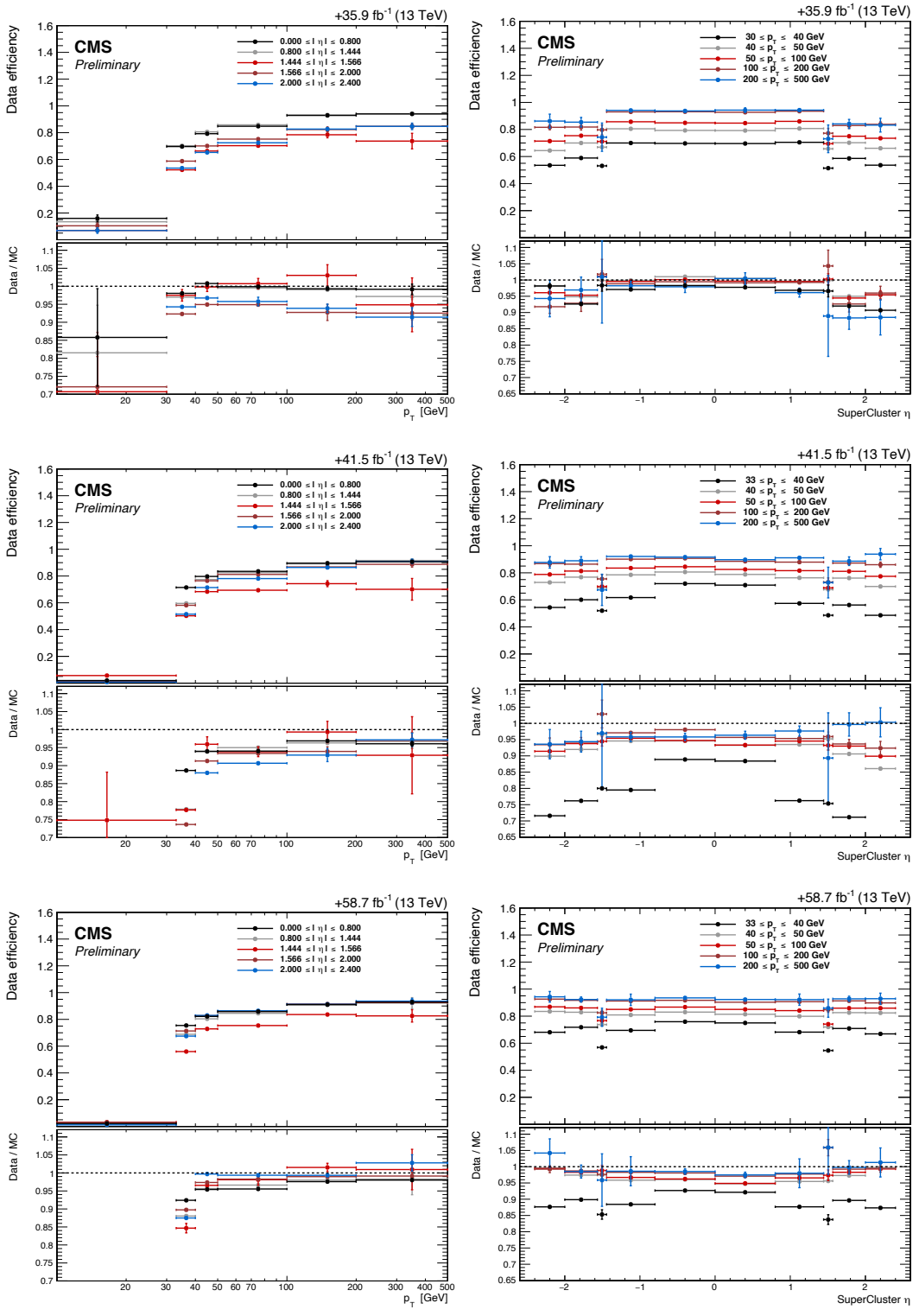

FIGURE 4.7: Electron trigger SFs versus p_{T} (left) and SuperCluster η (right), defined as the pseudorapidity computed with respect to the center of the CMS detector, and not with respect to the primary vertex of the pp interaction. The top row refers to 2016, the central to 2017, and the bottom to 2018. In each plot, the upper panel shows the efficiency observed in collision data for the HLT used, while in the lower panel the actual data/MC SFs are shown. The first bin in the plots vs p_{T} is unused, since it corresponds to an electron p_{T} range below the offline reconstruction thresholds for this particle.

TABLE 4.4: Preselection cutoffs concerning the analyzed particles.

μ	e
$p_{\mathrm{T}} > 25(28)\,\mathrm{GeV}$ for 2016, 2018 (2017)	$p_{\mathrm{T}} > 30(33)\,\mathrm{GeV}$ for 2016 (2017, 2018)
$\lvert\eta\rvert < 2.4$	$\lvert\eta\rvert < 2.4$
$\lvert d_{\mathrm{z}}\rvert < 0.5\,\mathrm{cm}$	$\lvert d_{\mathrm{z}}\rvert < 0.5\,\mathrm{cm}$
$\lvert d_{\mathrm{xy}}\rvert < 0.2\,\mathrm{cm}$	$\lvert d_{\mathrm{xy}}\rvert < 0.2\,\mathrm{cm}$
Relative isolation < 0.25	Isolation included in MVA ID
$\Delta R(\mu_{\mathrm{reco}}, \ell_{\mathrm{trigger}}) < 0.1$	
Only accept events with one lepton	
π	γ
$p_{\mathrm{T}} > 20\,\mathrm{GeV}$	$p_{\mathrm{T}} > 20\,\mathrm{GeV}$
High-purity track	$\lvert\eta\rvert < 2.5$
$\lvert d_{\mathrm{z}}\rvert < 0.5\,\mathrm{cm}$	Pixel veto against converted electrons
$\lvert d_{\mathrm{xy}}\rvert < 0.2\,\mathrm{cm}$	Isolation included in MVA ID
π charge $\neq \ell$ charge	$\Delta\phi(\ell, \gamma) > 0.04$
$\Delta\phi(\mu, \pi) > 0.09$	
$\Delta\phi(\mathrm{e}, \pi) > 0.05$	
$50 < m_{\pi\gamma} < 100\,\mathrm{GeV}$	

TABLE 4.5: Signal efficiency times acceptance per year and per lepton channel, after the event preselection.

Year	ε_{μ}	ε_{e}
2016	0.24 ± 0.01	0.15 ± 0.01
2017	0.23 ± 0.01	0.15 ± 0.01
2018	0.24 ± 0.01	0.15 ± 0.01

Figure 4.10 offers a graphical view of the reduction of the signal efficiency $\times$ acceptance at each step of the preselection, excluding the post-preselection cutoffs. The histogram, which is based on the merging of the two lepton channels, shows that the requiring a candidate pion and a photon above the respective p_{T} thresholds affects the signal efficiency more than the other requirements do. Table 4.6 reports the weighted event count for the signal and the main background processes after each step of the preselection (note that the steps correspond to those in Fig. 4.10). It must be stressed that the number reported in this table do not account for the post-preselection cutoffs described in Section 4.2.9 and, in particular, for the thresholds on the azimuthal angle ϕ between leptons and pions and leptons and photons. As shown in Fig. 4.6, these thresholds mainly affect the QCD multijet, W+jets, and Drell–Yan processes, removing a considerable portion of these samples and improving the agreement between collision data and MC simulation.

4.3 Multivariate selection

The event preselection involves somewhat minimal thresholds on the kinematic variables of the particles involved in the search for the decay $\mathrm{W} \to \pi\gamma$. Therefore, a further event selection is required to suppress the background sources and enhance the signal. This is performed through a multivariate analysis technique, namely a boosted decision tree (BDT). With respect to a cut-based analysis, which is able to select only one hypercube as region of phase space, the decision tree is able to split the

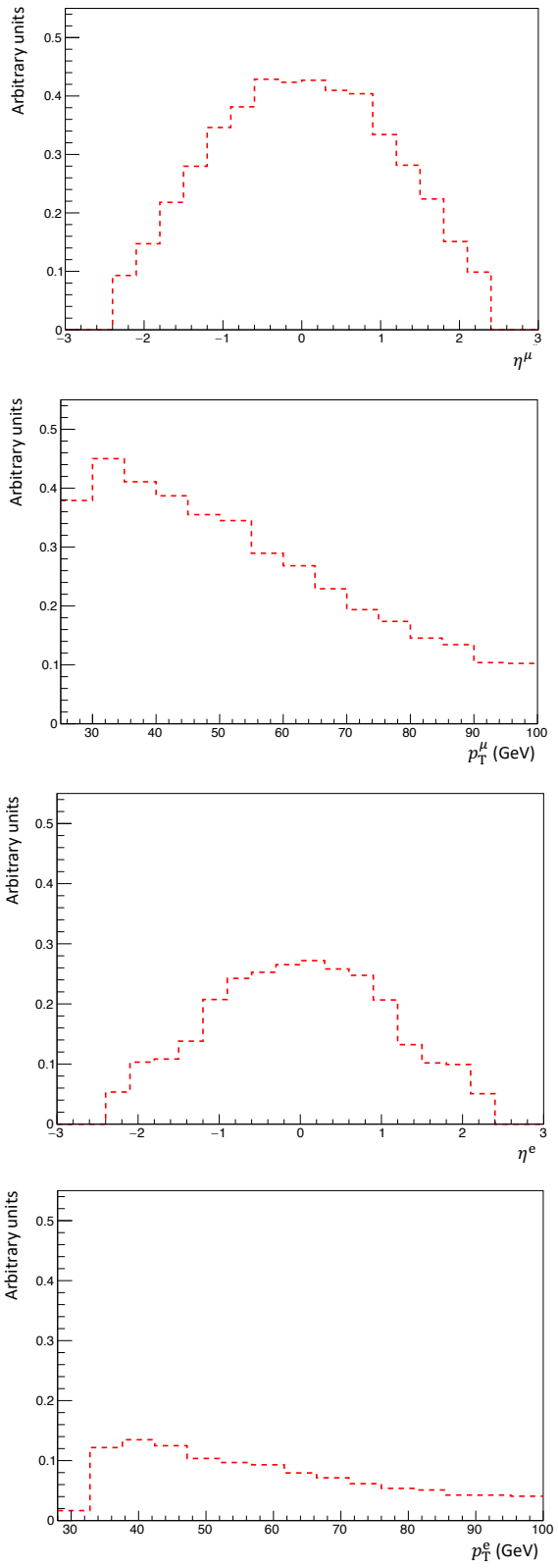

FIGURE 4.8: From top to bottom: distributions of η^{μ}, p_{T}^{μ}, η^{e}, and $p_{\mathrm{T}}^{\mathrm{e}}$ for signal only, after the event preselection.

phase space into a large number of hypercubes, each of which is identified as either "signal-like" or "background-like".

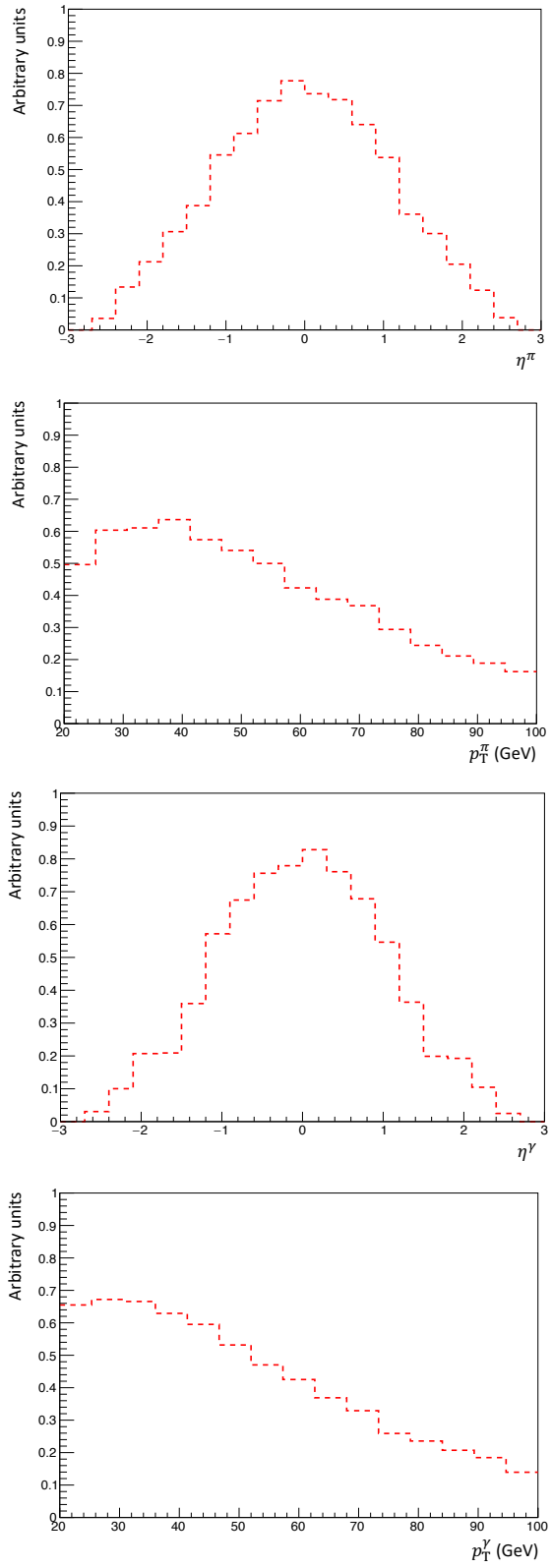

FIGURE 4.9: From top to bottom: distributions of η^{π}, p_{T}^{π}, η^{γ}, and p_{T}^{γ} for signal only, after the event preselection.

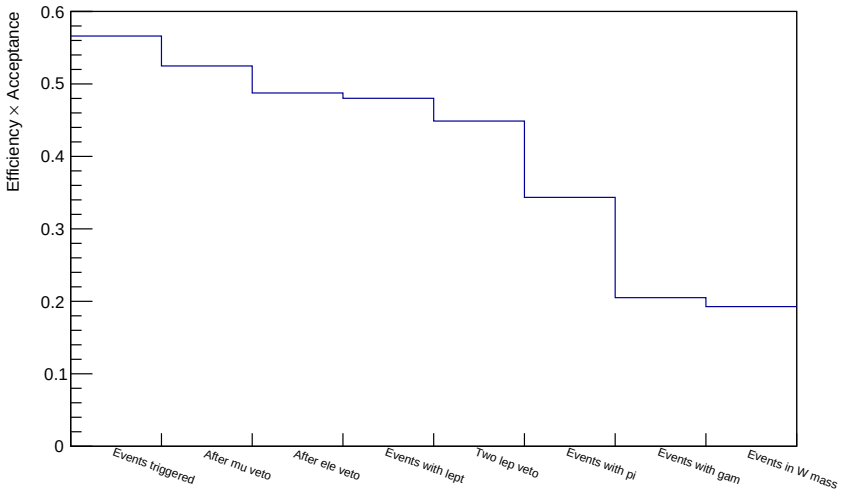

FIGURE 4.10: Reduction of the signal efficiency × acceptance after each step of the preselection. The particles involved are those that survived the respective thresholds, and the steps are defined as follows: number of events (n_{evts}) triggered, n_{evts} with at most one muon, n_{evts} with at most one electron, n_{evts} with at least a muon or an electron, n_{evts} that do not contain both a muon and an electron, n_{evts} with at least one pion, n_{evts} with at least one photon, n_{evts} in the $m_{\pi\gamma}$ valid range ($50 - 100\,\text{GeV}$). For each step, n_{evts} was divided by the total number of initial events ($\approx 240000 \times (1+1+0.3521)/3$), in order to obtain the corresponding signal efficiency.

TABLE 4.6: Weighted event count after each step of the preselection, for the signal and for the main background processes. The three years of data-taking are merged.

Process	triggered	μ veto	e veto	with ℓ	multi-ℓ veto	with π	with γ	in $m_{\pi\gamma}$ range
$t\bar{t}$	27×10^6	25×10^6	23×10^6	23×10^6	21×10^6	8×10^5	113176	78656
Drell–Yan	50×10^7	33×10^7	20×10^7	19×10^7	19×10^7	15×10^6	430951	216419
W+jets	27×10^8	27×10^8	27×10^8	25×10^8	25×10^8	13×10^7	12×10^5	714244
$W\gamma \to \ell\nu\gamma$	75×10^5	74×10^5	72×10^5	69×10^5	68×10^5	925546	395598	282755
$Z\gamma \to \ell\ell\gamma$	12×10^6	8×10^6	5×10^6	5×10^6	5×10^6	326564	104013	71443
$t(\bar{t})W$	24×10^5	22×10^5	21×10^5	21×10^5	19×10^5	627056	8988	5940
QCD multijet	62×10^7	60×10^7	49×10^7	36×10^7	36×10^7	58×10^6	293407	184881
$t\bar{t}\gamma$+jets	386979	345003	316504	311413	277859	126057	50924	33302
$W \to \pi\gamma$ ($\mathcal{B} = 10^{-6}$)	25	23	21	21	20	15	9	9

4.3.1 Boosted decision trees

A decision tree is a particular kind of multivariate classifier based on a sequence of binary splits of the data. The classifier needs to be trained using a set of known events (i.e., for which the distinction between signal and background is well-known) and of input variables, and its performance is measured using a separate set of testing events. The signal-to-background (S/B) discrimination process works as follows. Initially, all the training events are placed on one *node*. Afterwards, the variable x_{i} that provides the best S/B separation is identified, and the most effective threshold t_{i} is set on it so that two further nodes are created, containing the events for which $x_{\text{i}} > t_{\text{i}}$ and $x_{\text{i}} < t_{\text{i}}$, respectively. A variety of separation criteria may be used, but they all aim at maximizing the sample purity on each node. The same variable may be used at

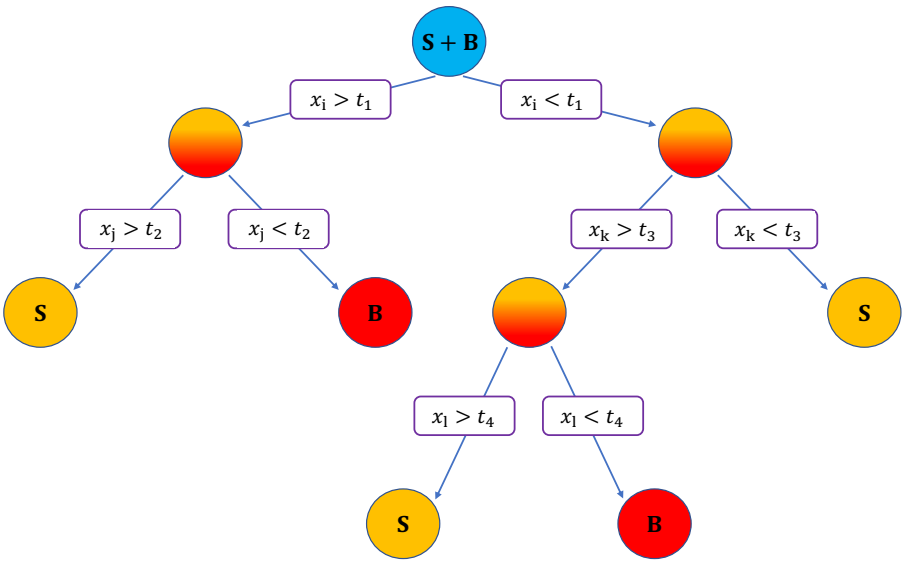

FIGURE 4.11: Schematic of a decision tree. Starting from a sample containing well-known signal and background events (S+B), a sequence of binary splits using the discriminating variables x_i is applied. The leaf nodes at the bottom end of the tree are labeled S or B depending on the majority of the events that end up in the respective nodes.

several nodes, while others might not be used at all. This procedure is iterated on these two nodes until a given number of final nodes (called *leaves*) is obtained, or until all leaves are pure-signal, or until a node contains too few events. The phase space is thus split into several regions that are eventually classified as signal or background, depending on the majority of training events that end up in the final leaf node. A sketch of the decision tree structure is shown in Fig. 4.11.

Boosting

Decision trees are a powerful and yet rather unstable tool, since small statistical fluctuations in the training sample can produce large effects in the tree. For instance, if two input variables exhibit similar S/B separation power, a fluctuation in the training sample may cause the tree growing algorithm to decide to split using one variable, whereas the other variable could have been selected without producing that fluctuation. In such a case the whole tree structure is altered below this node, possibly resulting in a substantially different classifier response. This problem is overcome by the use of *boosting*. With boosting, the training events that were misclassified (a signal event ended up on a background leaf or vice versa) see their weights increased (boosted), and a new tree is formed. This way, many trees are built up and the node-splitting procedure is repeated for each one of them. The final score is then taken as a weighted sum of the scores of the individual leaves. In summary, boosting increases the statistical stability of the classifier and can considerably enhance the S/B separation performance with respect to a single tree.

Different methods are available for boosting. The BDT used in this analysis exploits the *AdaBoost* [90], with a parameter $\beta = 0.25$ controlling the learning rate of the algorithm. Considering w_n as the ratio between the weight of the misclassified events and the total weight of the n^{th} tree, it is possible to define a variable

$$\alpha_n = \beta \log[(1 - w_n)/w_n]. \tag{4.3}$$

Then, the misclassified events have their weights multiplied by e^{α_n} and all the weights are renormalized so that their sum is one.

Configuration parameters

In addition to the boosting method, a number of other configuration options is available to set up the BDT classifier. In particular, in this analysis the following parameters are used:

- Number of grid points in the variable range used to find the optimal cutoff in node splitting: 20;
- Maximum depth of the decision tree: 3. A low value for this parameter, which indicates the maximum number of secondary nodes per tree, helps reducing the probability of overtraining (see the end of this section);
- Minimum percentage of training events required in a leaf node before interrupting the splitting: 2.5%. If this number is too large, detailed features in the parameter space are hard to be modeled. If it is too small, the risk to overtrain rises and boosting proved to be less effective. This parameter may implicitly set the maximum depth of the decision tree in case a node ends up containing less than 2.5% of training events before 3 splittings;
- Number of trees in the "forest" (i.e., number of trees generated in the boosting procedure): 800.

Several values and combinations of these configuration parameters were tested. The ones chosen have proved to provide a good separation between signal and background distributions in the BDT discriminant, and a generally better signal efficiency for a given background rejection efficiency with respect to the other configurations tested. At the same time, an attempt was made to limit the phenomenon of *overtraining* (that is a better signal efficiency in the training than in the validation sample, for a given background efficiency) by choosing "overtraining-safe" values of some of the parameters, such as the maximum depth of the decision tree.

4.3.2 BDT input variables

In this analysis, two BDT classifiers based on the ROOT library TMVA [91] are trained using half of the signal and background MC events, and validated through the other half. The requirement of a muon or an electron generates differences in the background distributions and the two cases are therefore handled as separate channels (from now, addressed as *muon* and *electron channels*). Nevertheless, the steps in the training and validation are carried out on the merged sample corresponding to the three years of data taking, for a larger statistics can enhance the performance of the classifier. The validity of this approach is further endorsed by the fact that the year-by-year performance does not show large differences, which was verified by comparing the signal efficiencies of BDTs tested and validated using the samples corresponding to a specific year, with those of BDTs trained using the samples of one year and validated using another. These efficiencies are compatible within the MC statistical uncertainty for any background rejection efficiency.

The BDT input variables with good signal-to-background discriminating power and whose distributions are well described by the simulation are:

- the transverse momentum of the lepton (p_{T}^{ℓ});

- the transverse momentum of the pion (p_{T}^{π});
- the transverse momentum of the photon (p_{T}^{γ});
- the missing transverse momentum in the event ($p_{\mathrm{T}}^{\mathrm{miss}}$);
- the number of b-tagged jets in the event (n_{b});
- the pion isolation ($\Sigma p_{\mathrm{T}}/p_{\mathrm{T}}^{\pi}$).

In the signal, the charged lepton, the pion and the photon arise from a W boson decay; therefore, their momenta are on average sufficiently large to make these particles distinguishable from others generated by the background processes. At the same time, the presence of a neutrino from one of the two W bosons in the event (or three neutrinos, in the case of $\mathrm{W} \to \tau\nu_{\tau},\ \tau \to \ell\bar{\nu}_{\ell}\nu_{\tau}$) guarantee a discrete amount of $p_{\mathrm{T}}^{\mathrm{miss}}$ in the event. The number of b jets associated to signal events is in principle two. If a direct threshold in this sense may be strongly effective in suppressing several background sources, it must be considered that n_{b} is also influenced by the efficiency and the purity of the b tagger (see Section 2.3.7). Therefore, it seems preferable to use n_{b} in the multivariate classifier. For what regards the pion-isolation variable, its remarkable discriminating power was already discussed in Section 4.2.6. A ranking of the BDT input variables is derived by counting how often the variables are used to split the decision tree nodes, and by weighting each split occurrence by the separation gain-squared it has achieved and by the number of events in the node. The results are presented in Table 4.7.

Table 4.7: Ranking of the BDT input variables for the two lepton channels.

Muon channel		*Electron channel*	
Variable	**Importance**	**Variable**	**Importance**
$\Sigma p_{\mathrm{T}}/p_{\mathrm{T}}^{\pi}$	0.23	$\Sigma p_{\mathrm{T}}/p_{\mathrm{T}}^{\pi}$	0.23
n_{b}	0.20	n_{b}	0.18
p_{T}^{γ}	0.19	p_{T}^{γ}	0.18
p_{T}^{π}	0.15	p_{T}^{π}	0.15
p_{T}^{μ}	0.12	$p_{\mathrm{T}}^{\mathrm{miss}}$	0.13
$p_{\mathrm{T}}^{\mathrm{miss}}$	0.11	$p_{\mathrm{T}}^{\mathrm{e}}$	0.13

Correlation matrices for the input variables are presented in Table 4.8, divided per lepton channel and for signal and background separately. Overtraining checks are performed and shown in Table 4.9. In general, the difference of signal efficiency between the training and the testing samples is at the per cent level.

TABLE 4.8: Correlation matrices for the BDT input variables.

Signal – muon channel

	$\Sigma p_{\mathrm{T}}/p_{\mathrm{T}}^{\pi}$	n_{b}	p_{T}^{γ}	p_{T}^{π}	p_{T}^{μ}	$p_{\mathrm{T}}^{\mathrm{miss}}$
$\Sigma p_{\mathrm{T}}/p_{\mathrm{T}}^{\pi}$	+1.000	+0.097	+0.121	−0.243	+0.069	+0.092
n_{b}	+0.097	+1.000	+0.015	−0.007	−0.011	+0.026
p_{T}^{γ}	+0.121	+0.015	+1.000	−0.009	+0.207	+0.232
p_{T}^{π}	−0.243	−0.007	−0.009	+1.000	+0.207	+0.324
p_{T}^{μ}	+0.069	−0.011	+0.207	+0.207	+1.000	+0.048
$p_{\mathrm{T}}^{\mathrm{miss}}$	+0.092	+0.026	+0.232	+0.324	+0.048	+1.000

Background – muon channel

	$\Sigma p_{\mathrm{T}}/p_{\mathrm{T}}^{\pi}$	n_{b}	p_{T}^{γ}	p_{T}^{π}	p_{T}^{μ}	$p_{\mathrm{T}}^{\mathrm{miss}}$
$\Sigma p_{\mathrm{T}}/p_{\mathrm{T}}^{\pi}$	+1.000	+0.095	+0.113	−0.300	+0.168	+0.128
n_{b}	+0.095	+1.000	+0.033	+0.116	−0.008	+0.115
p_{T}^{γ}	+0.113	+0.033	+1.000	−0.031	+0.070	+0.074
p_{T}^{π}	−0.300	+0.116	−0.031	+1.000	+0.212	+0.213
p_{T}^{μ}	+0.168	−0.008	+0.070	+0.212	+1.000	+0.055
$p_{\mathrm{T}}^{\mathrm{miss}}$	+0.128	+0.115	+0.074	+0.213	+0.055	+1.000

Signal – electron channel

	$\Sigma p_{\mathrm{T}}/p_{\mathrm{T}}^{\pi}$	n_{b}	p_{T}^{γ}	p_{T}^{π}	$p_{\mathrm{T}}^{\mathrm{miss}}$	$p_{\mathrm{T}}^{\mathrm{e}}$
$\Sigma p_{\mathrm{T}}/p_{\mathrm{T}}^{\pi}$	+1.000	+0.086	+0.070	−0.265	+0.054	+0.034
n_{b}	+0.086	+1.000	+0.026	−0.039	+0.001	+0.001
p_{T}^{γ}	+0.070	+0.026	+1.000	−0.003	+0.168	+0.175
p_{T}^{π}	−0.265	−0.039	−0.003	+1.000	+0.335	+0.219
$p_{\mathrm{T}}^{\mathrm{miss}}$	+0.054	+0.001	+0.186	+0.335	+1.000	−0.023
$p_{\mathrm{T}}^{\mathrm{e}}$	+0.034	+0.001	+0.175	+0.219	−0.023	+1.000

Background – electron channel

	$\Sigma p_{\mathrm{T}}/p_{\mathrm{T}}^{\pi}$	n_{b}	p_{T}^{γ}	p_{T}^{π}	$p_{\mathrm{T}}^{\mathrm{miss}}$	$p_{\mathrm{T}}^{\mathrm{e}}$
$\Sigma p_{\mathrm{T}}/p_{\mathrm{T}}^{\pi}$	+1.000	+0.118	+0.162	−0.282	+0.183	+0.225
n_{b}	+0.118	+1.000	+0.015	+0.072	+0.181	−0.003
p_{T}^{γ}	+0.162	+0.015	+1.000	−0.049	+0.076	+0.096
p_{T}^{π}	−0.282	+0.072	−0.049	+1.000	+0.222	+0.211
$p_{\mathrm{T}}^{\mathrm{miss}}$	+0.183	+0.181	+0.076	+0.222	+1.000	+0.063
$p_{\mathrm{T}}^{\mathrm{e}}$	+0.225	−0.003	+0.096	+0.211	+0.063	+1.000

TABLE 4.9: BDT overtraining checks.

	Sig efficiency from test (training) sample		
Channel	**at 0.01 Bkg eff**	**at 0.10 Bkg eff**	**at 0.30 Bkg eff**
Muon	0.589 (0.647)	0.889 (0.891)	0.975 (0.975)
Electron	0.662 (0.651)	0.904 (0.904)	0.978 (0.978)

4.3.3 BDT discriminant

Distributions of the BDT discriminant for collision data with simulated signal and background are shown in Fig. 4.12 for the two independent lepton channels. The

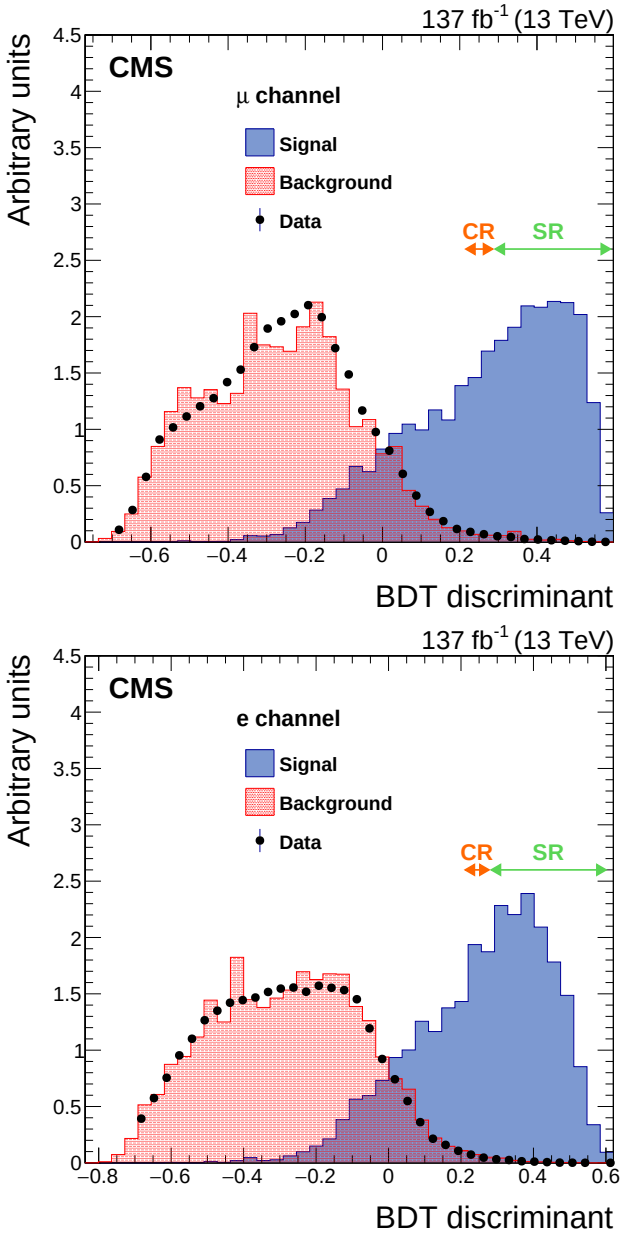

FIGURE 4.12: Event distributions as a function of the BDT discriminant for the muon (upper) and electron (lower) channels. The green and orange arrows indicate the intervals of the BDT discriminant used to define a signal (SR) and a control region (CR). The signal and background histograms are normalized to the same area. The statistical uncertainties in the data are small and thus not visible.

shape of the data is in good agreement with the MC background. Intervals of the BDT discriminant are chosen to define a *signal* (SR) and a *control region* (CR), which correspond to a signal- and a background-enriched range of the $m_{\pi\gamma}$ distribution, respectively. As will be discussed in Chapter 5, the $m_{\pi\gamma}$ distribution in the SR is used to extract the signal and background yields, while the CR, which contains roughly the same number of events as the SR, is used to estimate the functional form of

the background directly from data. A few aspects are considered when choosing the thresholds of the BDT discriminant corresponding to SR and CR. First of all, the significance Z is calculated as

$$Z = \frac{S\,\varepsilon_S^{\mathrm{BDT}}}{\sqrt{B\,\varepsilon_B^{\mathrm{BDT}}}}, \tag{4.4}$$

where S and B are the expected number of signal and background events after the preselection, and $\varepsilon_S^{\mathrm{BDT}}$ and $\varepsilon_B^{\mathrm{BDT}}$ are the signal and background efficiencies as a function of the BDT discriminant. As Fig. 4.13 shows, the significance tends to decrease as the signal efficiency increases. If aiming at the largest significance would seem natural, a second aspect must be taken into account: the SR needs to contain enough events in the $m_{\pi\gamma}$ distribution to be well fitted. In particular, it is desirable to keep the systematic uncertainty on the background parametrization, which is estimated with a fit on the data sidebands of the $m_{\pi\gamma}$ spectrum (see Section 6.2.5), at least one order of magnitude below the statistical uncertainty. This is achieved by retaining considerable background events in the sidebands. In summary, the following thresholds on the BDT discriminant and the corresponding regions are defined:

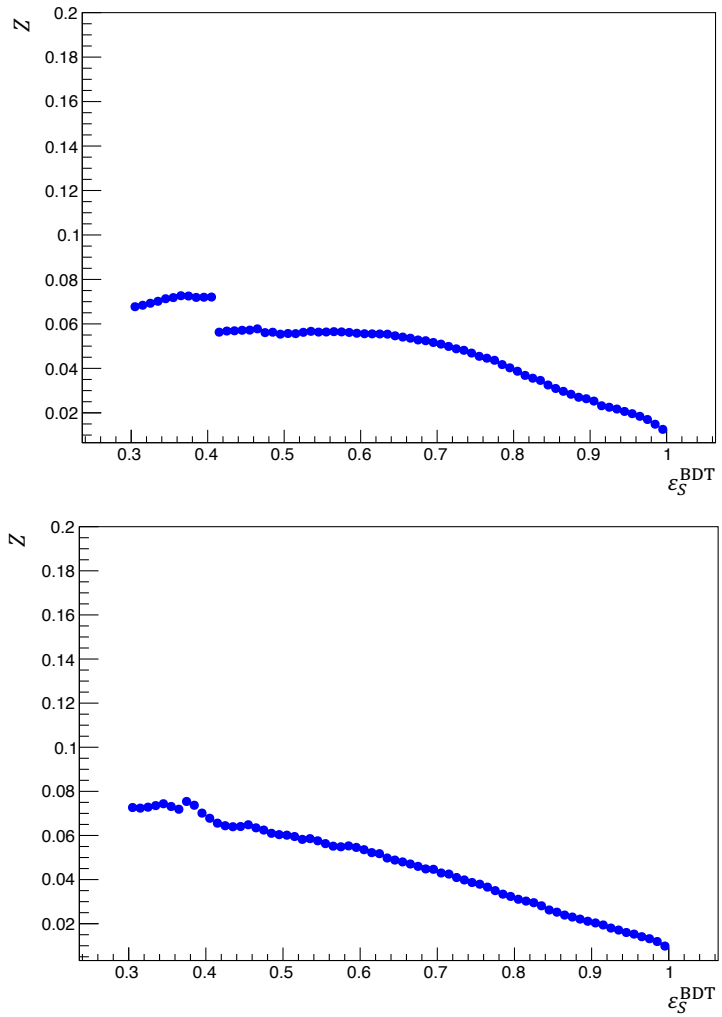

FIGURE 4.13: Significance as a function of the signal efficiency for the muon (upper) and the electron (lower) channels. The values corresponding to $\varepsilon_S^{\mathrm{BDT}} < 0.3$ are omitted, since strong fluctuations affect in the samples in that interval.

- SR: for a BDT discriminant > 0.281 in the muon channel and for a BDT discriminant > 0.269 in the electron channel. These cutoffs correspond to 0.55 and 0.54 signal efficiencies ($\varepsilon_S^{\mathrm{BDT}}$), respectively;
- CR: for $0.206 <$ BDT discriminant < 0.281 in the muon channel and $0.209 <$ BDT discriminant < 0.269 in the electron channel;
- $\mathrm{CR_{ext}}$: the *extended* control region, $-0.100 <$ BDT discriminant < 0.281 in the muon channel and $-0.100 <$ BDT discriminant < 0.269.

With respect to the CR, the $\mathrm{CR_{ext}}$ contains an even larger portion of background, and is used to test the dependence of the $m_{\pi\gamma}$ spectrum on the BDT discriminant (Section 4.3.4).

4.3.4 The $m_{\pi\gamma}$ dependence on the BDT discriminant

The dependence of the $m_{\pi\gamma}$ spectrum on the BDT discriminant can be observed in Fig. 4.14 for simulated signal and background events and in the two lepton channels separately. Qualitatively, the background shows a more uniform distribution over the BDT discriminant spectrum with respect to the signal.

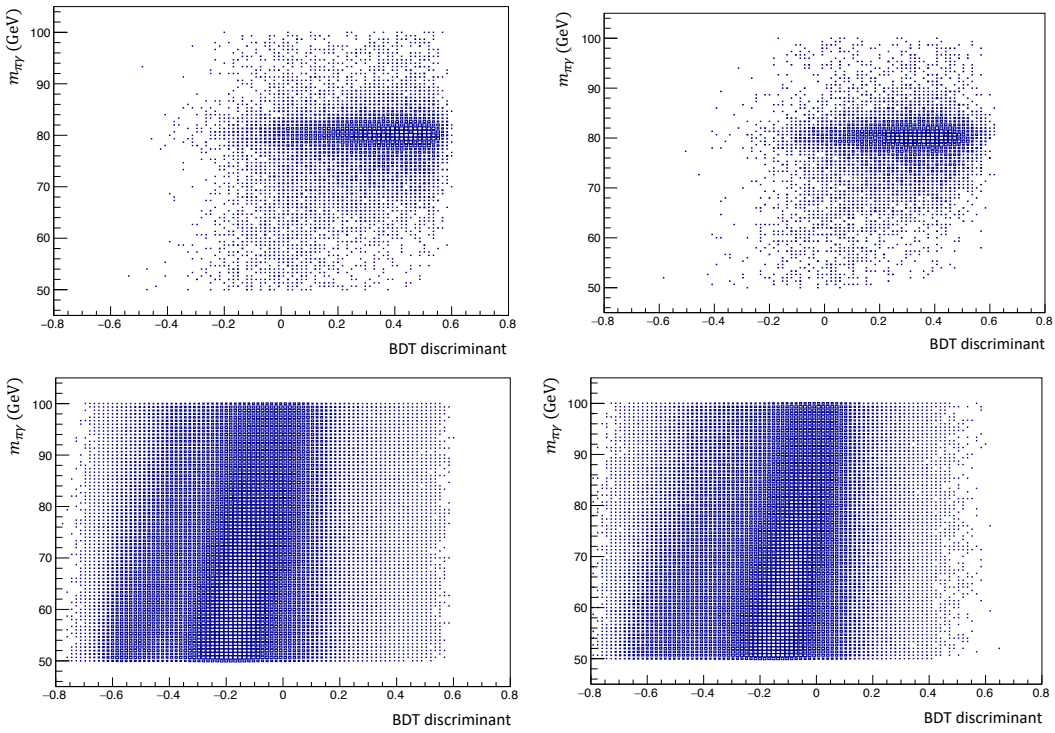

FIGURE 4.14: Simulated events in the $m_{\pi\gamma}$ distribution as a function of the BDT discriminant. The top plots contain signal events in the muon (left) and electron (right) channels; similarly, the bottom plots contain background events.

The inclusion of the pion and photon p_{T} among the BDT input variables could potentially generate distortions in the background distributions of the observable $m_{\pi\gamma}$. To ascertain if this is the case, a ratio between the $m_{\pi\gamma}$ spectrum in the SR and in the $\mathrm{CR_{ext}}$ is computed for background events only. The observed trend is flat within the uncertainties (Fig. 4.15, upper plots), thus not indicating any evidence of distortion effects induced by the pion and photon kinematic variables. As an additional test, two more BDT classifiers are trained using the input variables listed in 4.3.2, where

p_{T}^{π} and p_{T}^{γ} have been divided by $m_{\pi\gamma}$. This procedure can in principle further reduce the dependence of the $m_{\pi\gamma}$ distribution on p_{T}^{π} and p_{T}^{γ}. With respect to the nominal case, the ratio between the $m_{\pi\gamma}$ spectrum in the SR and in the $\mathrm{CR_{ext}}$ (Fig. 4.15, lower plots) does not appear to be more stable (note that the definition of the SR in this second case targets the same signal efficiency as the nominal case, leading to BDT discriminant > 0.273 and and BDT discriminant > 0.258 for the muon and the electron channels, respectively; the $\mathrm{CR_{ext}}$ are then chosen as $-0.1 <$ BDT discriminant < 0.273 and $-0.1 <$ BDT discriminant < 0.258).

At the same time, a comparison of the performance of the two BDT sets shows a reduction of the background rejection efficiency all over the signal efficiency range (Fig. 4.16), making the use of the nominal input variables preferable.

4.3.5 Multivariate selection overview

The agreement between data and MC simulation is verified in various distributions of interest, using events that fall inside the SR. As anticipated at the beginning of Chapter 3, this search for $\mathrm{W} \to \pi\gamma$ followed a blind analysis approach. Therefore, the selection procedures, including the evaluation of the consistency of the distributions in the kinematic variables between data and MC, were performed on the sidebands

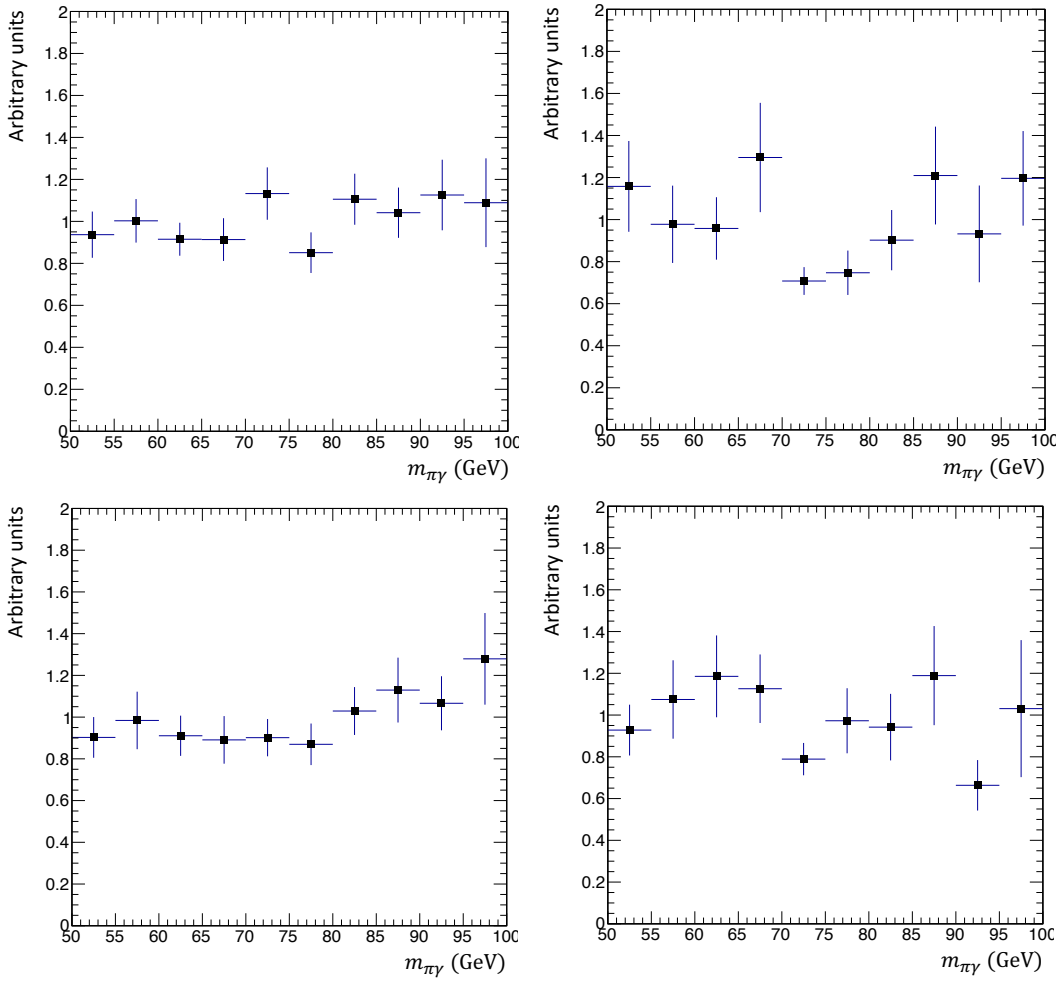

FIGURE 4.15: Ratios between the $m_{\pi\gamma}$ spectrum in the SR and in the $\mathrm{CR_{ext}}$. The upper plots refer to distributions produced with the nominal BDT input variables, whereas in the lower plots $p_{\mathrm{T}}^{\pi}/m_{\pi\gamma}$ and $p_{\mathrm{T}}^{\gamma}/m_{\pi\gamma}$ are used.

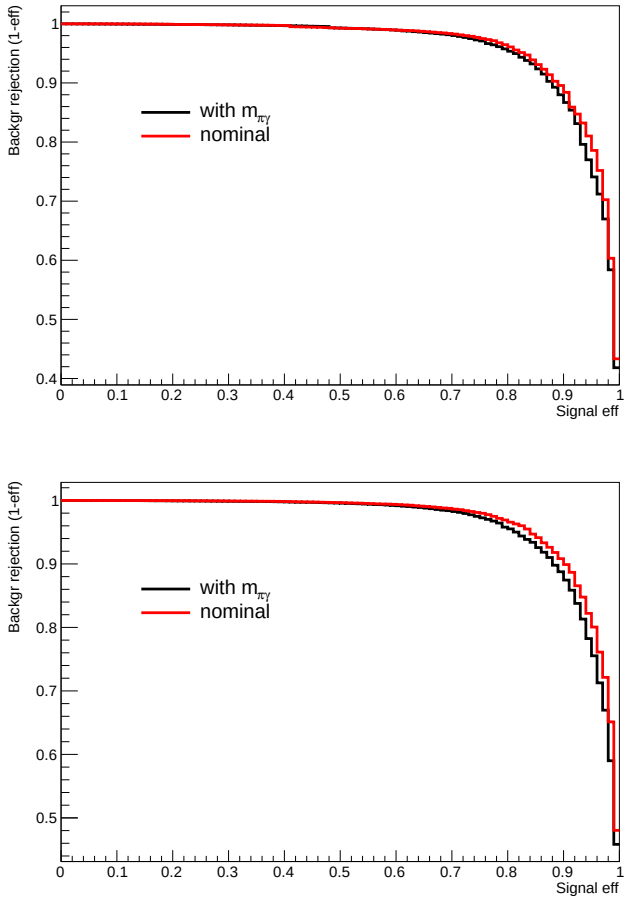

FIGURE 4.16: Observed trend of the background rejection efficiency as a function of the signal efficiency, for the muon (upper) and the electron (lower) channels. The red curves are obtained using nominal BDT input variables, whereas for the black curves $p_{\mathrm{T}}^{\pi}/m_{\pi\gamma}$ and $p_{\mathrm{T}}^{\gamma}/m_{\pi\gamma}$ are used.

of the SR in the distribution of the main observable ($m_{\pi\gamma}$), that is excluding a *blind window* in between 65 and 90 GeV. On the other hand, the other kinematic variables presented in this section could be inspected directly, since no observational bias can be translated from those to the $m_{\pi\gamma}$ spectrum. Distributions of lepton, pion, and photon η and p_{T} are shown in Figures 4.17, 4.18, and 4.19.

In general, a good agreement between data and MC is observed within the statistical uncertainties. Large fluctuations in some of the simulated background samples are due to the limited statistics in those samples. This is not a problem for the analysis, since the background shape in the main observable $m_{\pi\gamma}$ is derived directly from data. It is also important to emphasize that MC events are mainly used for the training of the BDT classifiers, and thus a possible poor modeling of the input variables in the simulation can at most lead to a suboptimal multivariate selection.

The event distribution as a function of the number of b-tagged jets with $p_{\mathrm{T}} > 25$ GeV is also shown in Fig. 4.20. As expected, the number of events with zero b-tagged jets is strongly suppressed by the multivariate selection, since this represents a strongly background-like topology.

Finally, a comparison between collision and simulated events is shown in the distribution of the observable $m_{\pi\gamma}$ (Fig. 4.21), both separately for the two lepton channels and in their sum. The shape of the background depends on the lepton channel, as

well as the relative composition of the background processes. This is expected, since the electron and the muon requests give rise to different background sources. The expected number of events in the range $65 < m_{\pi\gamma} < 90\,\text{GeV}$, that is in the range of the $m_{\pi\gamma}$ distribution where a signal, if present, is expected to be visible, is reported in Table 4.10 for the signal and for each surviving background source, divided per lepton channel. The product of the signal efficiencies per each lepton channel and the detector acceptance after the multivariate selection, with the associated statistical uncertainty, is summarized in Table 6.2.

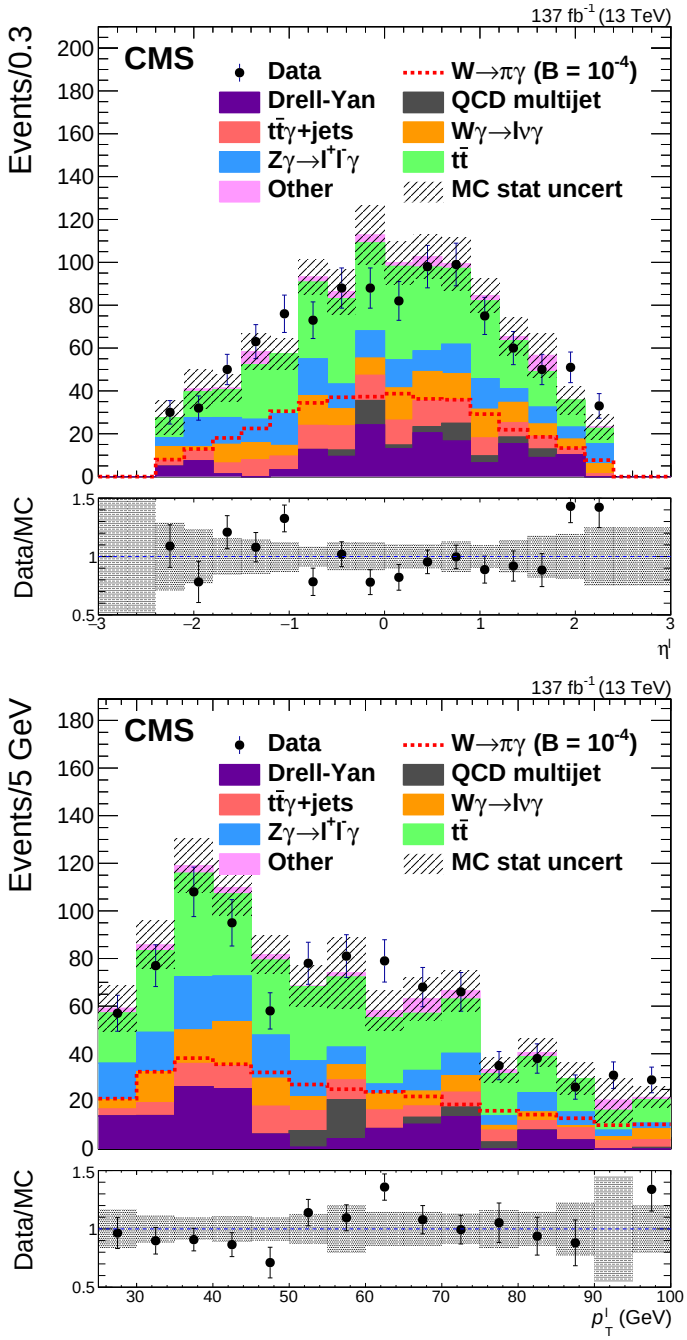

FIGURE 4.17: Event distributions as a function of η^{ℓ} (upper) and p_{T}^{ℓ} (lower) after the BDT selection. The simulated MC distribution for the signal is given by the dashed red line and corresponds to a 10^{-4} branching fraction for the $\mathrm{W} \to \pi\gamma$ decay. In the lower panels, the ratios between data and the background component of the MC are shown.

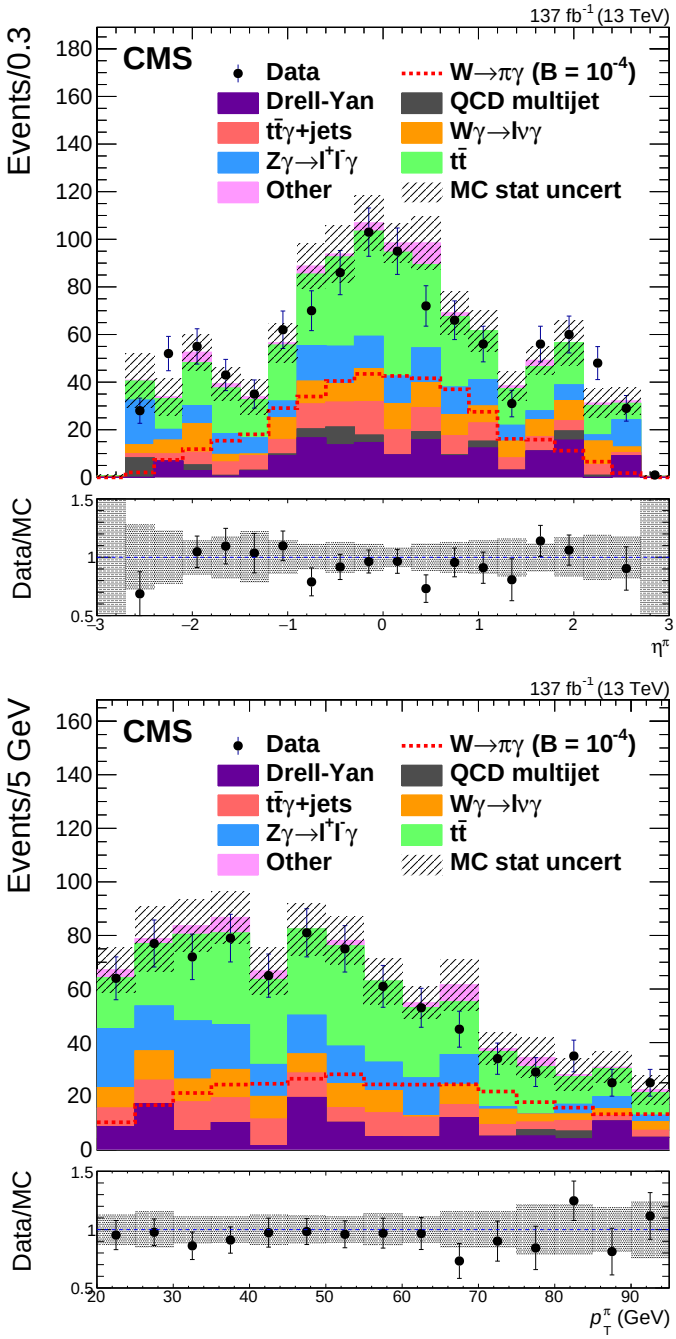

FIGURE 4.18: Event distributions as a function of η^{π} (upper) and p_{T}^{π} (lower) after the BDT selection. The simulated MC distribution for the signal is given by the dashed red line and corresponds to a 10^{-4} branching fraction for the W $\to \pi\gamma$ decay. In the lower panels, the ratios between data and the background component of the MC are shown.

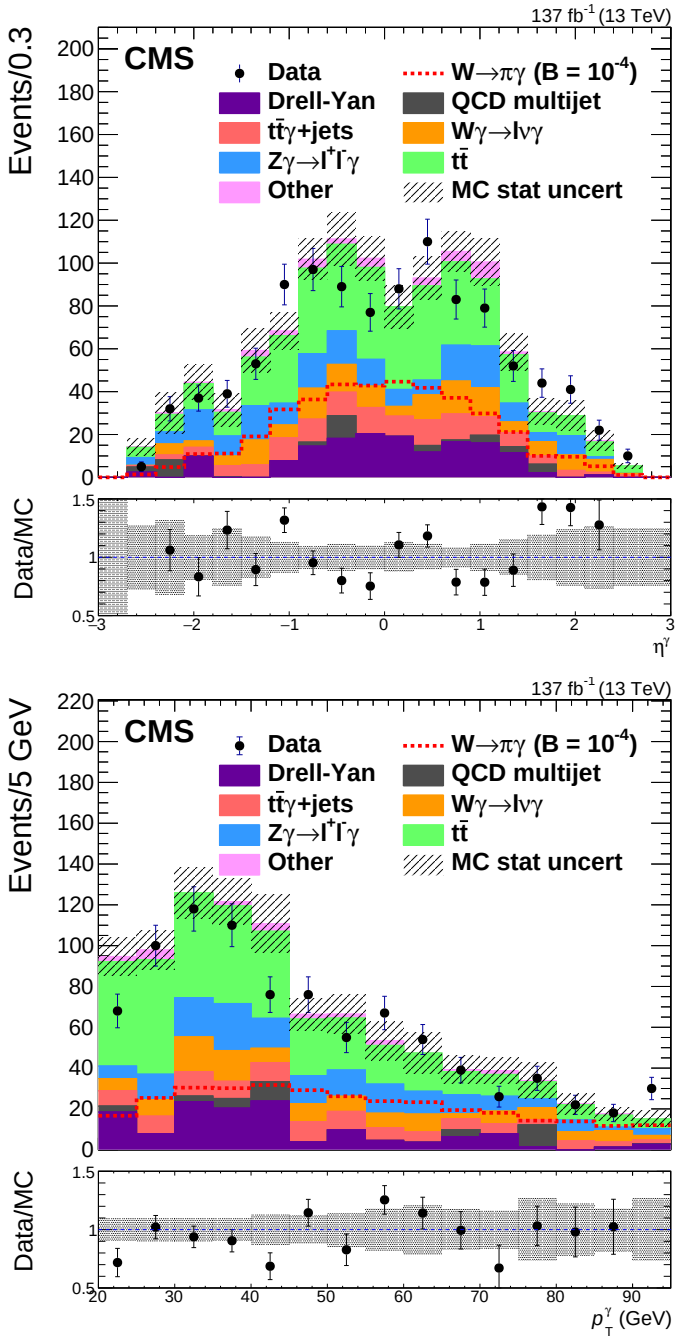

FIGURE 4.19: Event distributions as a function of η^{γ} (upper) and p_{T}^{γ} (lower) after the BDT selection. The simulated MC distribution for the signal is given by the dashed red line and corresponds to a 10^{-4} branching fraction for the $\mathrm{W} \to \pi\gamma$ decay. In the lower panels, the ratios between data and the background component of the MC are shown.

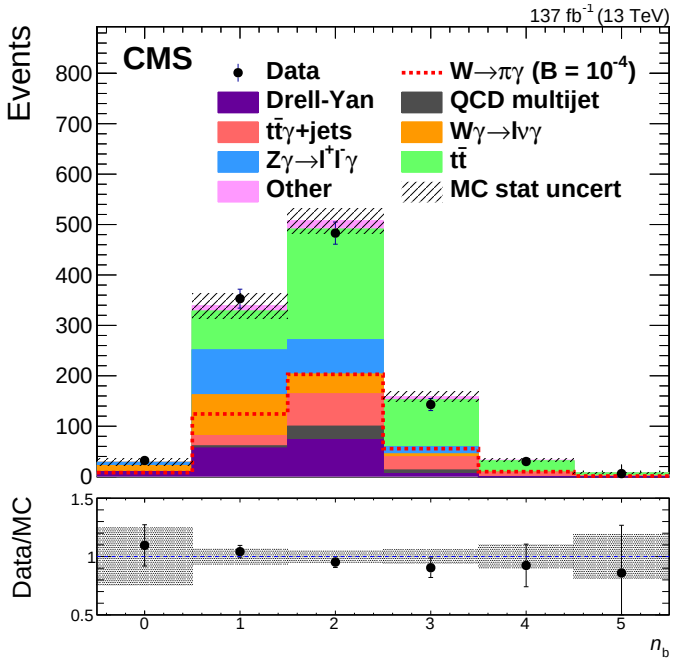

FIGURE 4.20: Event distributions as a function of n_{b} after the BDT selection. The simulated MC distribution for the signal is given by the dashed red line and corresponds to a 10^{-4} branching fraction for the $\mathrm{W} \to \pi\gamma$ decay. In the lower plot, the ratio between data and the background component of the MC is shown.

TABLE 4.10: Number of expected signal and background events per lepton channel, in a window $65 < m_{\pi\gamma} < 90\,\mathrm{GeV}$.

Process	**Muon channel**	**Elecron channel**	**Muon + Electron**
$\mathrm{t\bar{t}}$	156.10	60.35	216.45
$\mathrm{Z}\gamma \to \ell\ell\gamma$	46.15	37.09	83.24
$\mathrm{W}\gamma \to \ell\nu\gamma$	51.56	25.57	77.13
Drell–Yan	50.36	19.60	69.96
$\mathrm{t\bar{t}}\gamma + \mathrm{jets}$	43.95	22.64	66.59
QCD multijet	17.68	0.15	17.83
$\mathrm{t}(\bar{\mathrm{t}})\mathrm{W}$	8.69	4.98	13.67
$\gamma + \mathrm{jets}$	3.90	0.49	4.39
$\mathrm{W} \to \pi\gamma$ ($\mathcal{B} = 10^{-6}$)	2.45	1.39	3.84
Total	380.48	172.17	552.65

TABLE 4.11: Product of signal efficiency and acceptance per year and per lepton channel, after the multivariate event selection.

Year	ε_μ	ε_{e}
2016	0.12 ± 0.01	0.07 ± 0.01
2017	0.11 ± 0.01	0.07 ± 0.01
2018	0.12 ± 0.01	0.07 ± 0.01

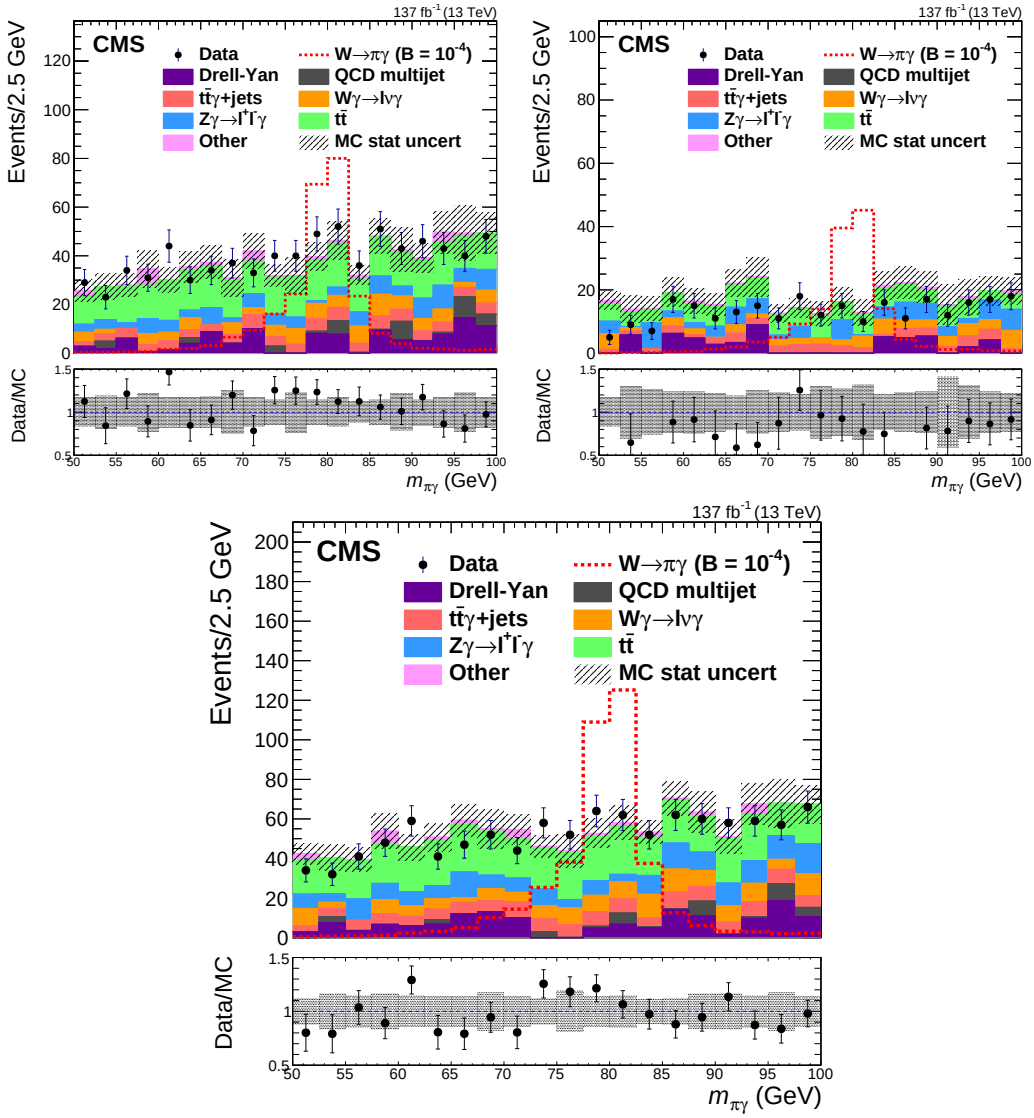

FIGURE 4.21: Event distributions as a function of $m_{\pi\gamma}$ for the muon (top left) and electron channel (top right), and for their sum (bottom) after the BDT selection. The simulated MC distribution for the signal is given by the dashed red line and corresponds to a 10^{-4} branching fraction for the $W \to \pi\gamma$ decay. In the lower panels, the ratios between data and the background component of the MC are shown.

Chapter 5

Signal and background description and yield extraction

After the events have been skimmed and classified as signal- or background-like using the selection procedures described in Chapter 4, the search for the rare decay $\mathrm{W} \to \pi\gamma$ proceeds analyzing the $m_{\pi\gamma}$ spectrum. Within this distribution, a signal, if present, is expected to be found around the value of the W boson mass [17]. This chapter discusses the extraction of the signal and background yields from the $m_{\pi\gamma}$ distribution, and the efforts in the description of the signal and background lineshapes. Section 5.3.1 is dedicated to the description of the principles of the maximum likelihood method for parameter estimation, and the computer tool used to perform it.

5.1 Signal and background yield extraction

The signal and background yields are extracted using an unbinned maximum likelihood fit to the $m_{\pi\gamma}$ distribution in the SR. The probability density function (pdf) used to perform the fit is defined as follows:

$$f(m_{\pi\gamma}) = \left[\frac{N_{\mathrm{sig}}}{N_{\mathrm{sig}} + N_{\mathrm{bkg}}(\mathrm{SR})} f_{\mathrm{sig}}(m_{\pi\gamma}) + \frac{N_{\mathrm{bkg}}(\mathrm{SR})}{N_{\mathrm{sig}} + N_{\mathrm{bkg}}(\mathrm{SR})} f_{\mathrm{bkg}}(m_{\pi\gamma}) \right] \mathcal{G}, \qquad (5.1)$$

where $N_{\mathrm{bkg}}(\mathrm{SR})$ is a floating parameter representing the number of background events in the SR, and $\mathcal{G}$ is a Gaussian pdf used to account for the nuisance parameters that model the systematic uncertainties (see Section 6.1). The determination of the functional form of the signal, $f_{\mathrm{sig}}(m_{\pi\gamma})$, and the background, $f_{\mathrm{sig}}(m_{\pi\gamma})$, is described in detail in Section 5.2 and 5.3, respectively. From now on, the dependance of f, f_{sig}, and f_{bkg} on $m_{\pi\gamma}$ will be omitted for the sake of brevity.

The $\mathrm{W} \to \pi\gamma$ branching fraction is extracted from the number of observed signal events N_{sig}, parametrized as:

$$N_{\mathrm{sig}} = \sigma_{\mathrm{t\bar{t}}}\, \mathcal{B}(\mathrm{W}^{\mp} \to \ell^{\mp}\nu)\, \mathcal{B}(\mathrm{W}^{\pm} \to \pi^{\pm}\gamma)\; \mathcal{L}_{\mathrm{int}}\, \varepsilon_{\ell}\, k^{\beta}, \quad \ell = \mu, \mathrm{e}, \tau_{\ell}. \qquad (5.2)$$

In the expression above, $\sigma_{\mathrm{t\bar{t}}} = 815\,\mathrm{pb}$ indicates the $\mathrm{t\bar{t}}$ production cross section in pp collisions at 13 TeV, as measured by the CMS Collaboration using a $\mathcal{L}_{\mathrm{int}} = 2.2\,\mathrm{fb}^{-1}$ in 2015 [80]; $\mathcal{B}(\mathrm{W}^{\mp} \to \ell^{\mp}\nu) = 0.1086\,(2{+}0.3521)$ uses the most up-to-date measurements of the branching fraction of the W boson into leptons [17], with the factor $(2 + 0.3521)$ accounting for the possible decay into a muon, an electron or a τ subsequently decaying into leptons (τ_{ℓ}); $\mathcal{L}_{\mathrm{int}} = 137\,\mathrm{fb}^{-1}$ is the integrated luminosity collected by CMS during Run 2 (2016–2018); ε_{ℓ} is the trigger, reconstruction, identification, and selection efficiency, computed as the ratio between the number of events surviving the selection process and the number of initial events in the MC simulation (see Table

6.2); the term k^{β} accounts for the systematic uncertainties, as will be discussed in Section 6.1.

N_{sig} is not sensitive to the contribution of W bosons that are not produced via $\mathrm{t\bar{t}}$ decay. To verify if this is the case, around 20000 events containing W bosons not produced via $\mathrm{t\bar{t}}$ decay were generated with PYTHIA using a cross section $\sigma_{\mathrm{pp}\to\mathrm{W}} \simeq 105000\,\mathrm{pb}$ [30], and the W bosons were forced to decay into a pion and a photon. Despite this cross section is much larger than that of the $\mathrm{t\bar{t}}$ signal contribution, such process is rejected by the requirement of $\mathrm{t\bar{t}}$ production with a large p_{T} lepton, yielding to less than 0.1 expected events after the preselection.

Another contribution to N_{sig} that is not considered arises from the associated production of a top quark and a W boson, which is expected to be two orders of magnitude smaller than that from $\mathrm{t\bar{t}}$ events.

5.2 Signal parametrization

Two intrinsic features of the signal component in the $m_{\pi\gamma}$ distribution must be described: the shape and the normalization. Since, in principle, the precise shape of the signal in this distribution in collision data is not known, the approach chosen is to derive it from the MC simulation by fitting the $m_{\pi\gamma}$ distribution in the SR. The signal lineshape is expected to be independent from the type of lepton arising from the decay of the other W in the event. Therefore, the signal samples corresponding to the two lepton channels are merged to reduce the statistical uncertainties in the signal lineshape parameters, after verifying their compatibility with a χ^2 test. Similarly, the signal samples corresponding to the three years of data-taking are checked for compatibility with a χ^2 test and then merged. Figure 5.1 shows a visual comparison of the signal event distribution as a function of $m_{\pi\gamma}$ in the SR in the two lepton channels and the three years of data-taking.

The pdf used to describe the signal lineshape, f_{sig}, is the sum of a double crystal ball (DCB) [92] and a Gaussian ($f_{\mathcal{G}}$) pdfs. In particular, the DCB is defined as

$$\mathrm{DCB}(x) = N \begin{cases} A\,(B+|x|)^{-n_{\mathrm{L}}}, & \text{for } x < \alpha_{\mathrm{L}} \\ A\,(B+|x|)^{-n_{\mathrm{R}}}, & \text{for } x > \alpha_{\mathrm{R}} \\ \exp\left(\frac{-x^2}{2\sigma_m^2}\right), & \text{for } \alpha_{\mathrm{L}} \leq x \leq \alpha_{\mathrm{R}} \end{cases} \tag{5.3}$$

In the expression above, $x = (m_{\pi\gamma} - m)/\sigma_m$; m, σ_m, n_{L}, n_{R}, α_{L}, and α_{R} are six parameters intended to capture both the Gaussian core of the $m_{\pi\gamma}$ resolution function and the tail al lower mass due to the energy loss of the reconstructed particles traversing the detector layers, well described by a power law; in particular, the α and n terms that define the power law are usually strongly correlated. A and B are nonindependent parameters, defined by requiring the continuity of the function itself and of its first derivative; N is the normalizing constant. The purpose of the Gaussian pdf

$$f_{\mathcal{G}}(m_{\pi\gamma}; \bar{x}, \sigma) = \frac{1}{\sqrt{2\pi}\sigma}\,\exp\left[-\frac{(m_{\pi\gamma}-\bar{x})^2}{2\sigma^2}\right] \tag{5.4}$$

is to model the natural width of the W boson.

The unbinned maximum likelihood fit to the MC signal $m_{\pi\gamma}$ distribution in the SR is shown in Fig. 5.2. The p_{T} and η dependent scale factors described in Section 4.2.10 are applied to each simulated event, since they have an impact on the lineshape. The values of the f_{sig} parameters extracted from the fit are listed in Table 5.1. The uncertainties are computed with the HESSE algorithm of the MINUIT minimization

tool, discussed in Section 5.3.1. When the fit to the $m_{\pi\gamma}$ distribution of the data is performed (see Section 7.1), these parameters are fixed, while the normalization can vary. Uncertainties in the position and the width of the DCB peak are accounted for as systematic uncertainties in the final result, as described in Section 6.2.4.

TABLE 5.1: Parameter estimation from the fit to the MC signal $m_{\pi\gamma}$ distribution in the SR.

pdf	**Parameter**	**Fit result**
DCB	m	$79.9 \pm 0.2\,\text{GeV}$
	σ_m	$3.3 \pm 0.2\,\text{GeV}$
	α_L	0.6 ± 0.1
	α_R	1.6 ± 0.1
	n_L	50.0 ± 35.1
	n_R	1.2 ± 0.2
$f_\mathcal{G}$	$\bar{x}$	$80.3 \pm 0.1\,\text{GeV}$
	σ	$1.3 \pm 0.1\,\text{GeV}$

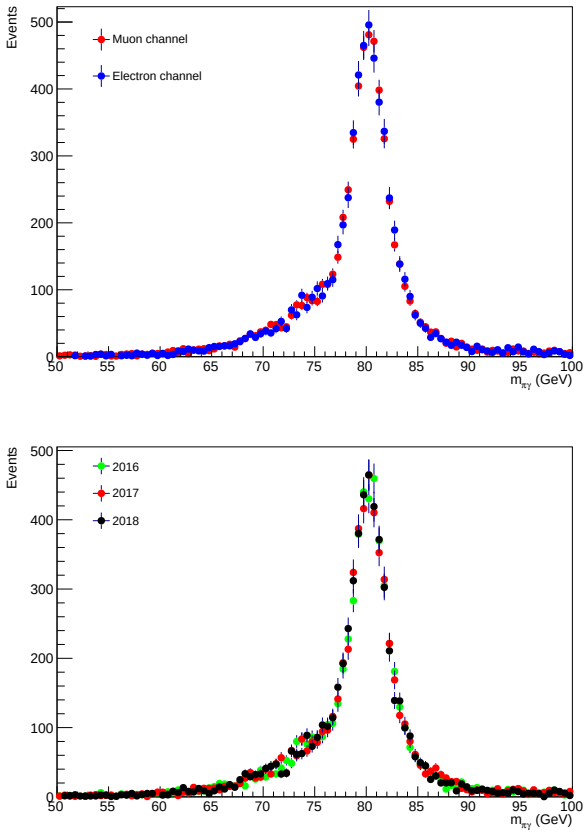

FIGURE 5.1: Visual comparison of the MC signal event distribution as a function of $m_{\pi\gamma}$ for the two lepton channels (top) and the three years of data-taking (bottom). In the left plot, the three years are merged and the number of events in the muon channel is normalized to that in the electron channel; in the right plot, the two lepton channels are merged and the number of events in the 2016 and 2017 samples is normalized to that in the 2018 channel.

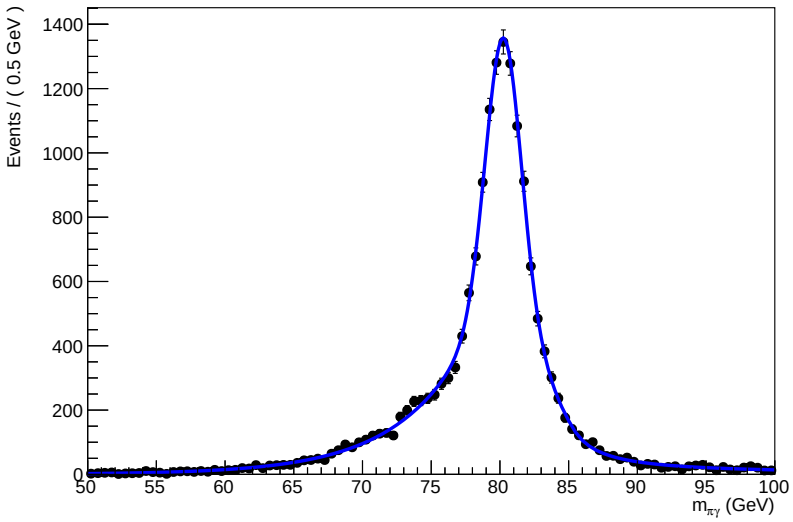

FIGURE 5.2: Unbinned maximum likelihood fit to the MC signal $m_{\pi\gamma}$ distribution in the SR.

5.3 Background parametrization

The study of the background lineshape is performed on the $m_{\pi\gamma}$ distribution in the CR of the data, which is chosen so it contains comparable number of events with respect to the SR. Such a strategy is an important item for this analysis, since it allows me to be independent from the MC description of the background processes when the background yield is extracted in the fit to the SR in data. A comparison between the data in the CR and in the SR can be observed in Fig. 5.3. The compatibility of the two distributions is rigorously verified using a χ^2 test.

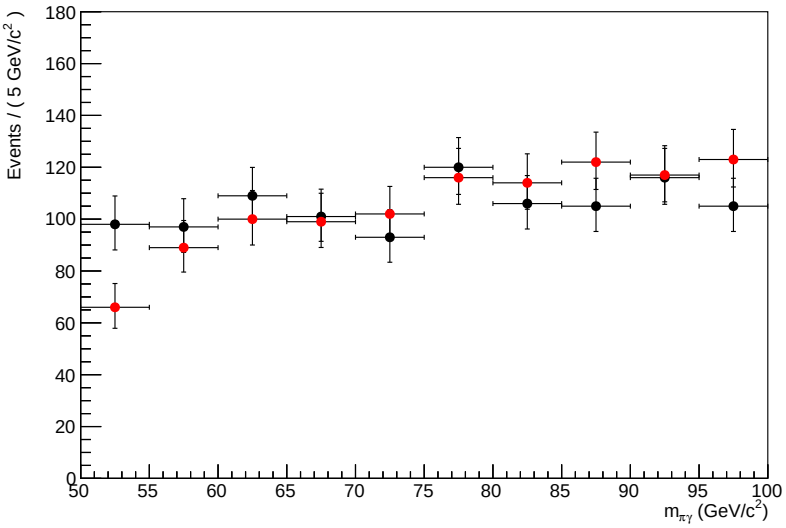

FIGURE 5.3: Comparison between the events in the data CR (black markers) and SR (red markers) of the $m_{\pi\gamma}$ distribution.

After testing several pdfs, including exponential and various polynomial functional forms, a first order Chebyshev polynomial (f_{bkg}, with one parameter b) is chosen, that is a linear function. The order of f_{bkg} is determined using a statistical test that

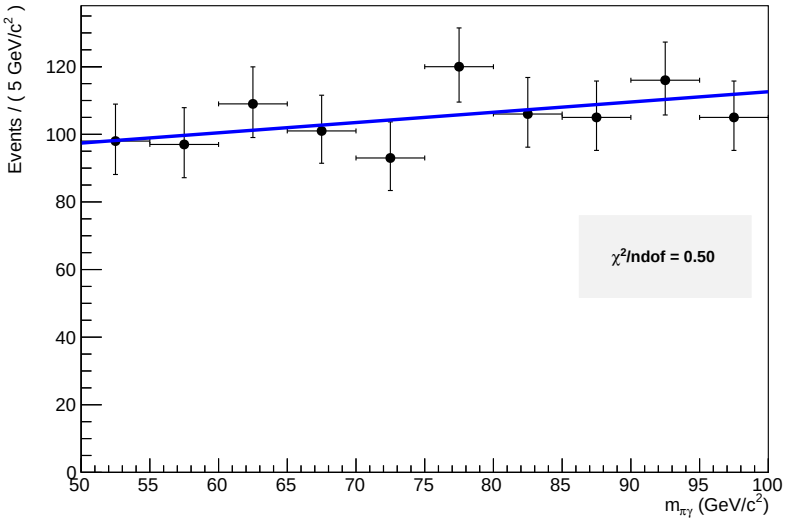

FIGURE 5.4: Fit to the data distribution of $m_{\pi\gamma}$ in the CR. The goodness-of-fit is verified with a minimum χ^2 method.

quantifies the goodness-of-fit gain at a given order. The test proceeds as follows. The fit to the $m_{\pi\gamma}$ distribution in the CR of the data is repeated for different polynomial orders (N) and the minimum negative log-likelihood value (NLL_N) is calculated. It can be shown that twice the difference between the NLL_N of two adjacent orders follows a χ^2 distribution with one degree of freedom. Fitting with the order $N+1$ polynomial instead of the order N is judged as a significant improvement if the measured $\chi^2_{N\to N+1}$ is such that the p-value $\mathrm{p}(\chi^2 > \chi^2_{N\to N+1}) > 0.05$. This is the case when $2\,(NLL_N - NLL_{N+1}) > 3.85$. For what regards the description of the data CR in this analysis, the transition from a first to a second order polynomial leads to a $2\,(NLL_1 - NLL_2) = 0.66$, suggesting that a linear function describes the data well.

The unbinned maximum likelihood fit to the data distribution of $m_{\pi\gamma}$ in the CR is shown in Fig. 5.4. The parameter estimation from the fit is reported in Table 5.2. The uncertainties are computed with the HESSE algorithm of the MINUIT minimization tool, discussed in Section 5.3.1. The values of both the slope and the intercept (i.e., the normalization) of f_{bkg} can then vary in the fit to the $m_{\pi\gamma}$ distribution of the data in the SR (see Section 7.1).

TABLE 5.2: Parameter estimation from the fit to the data distribution of $m_{\pi\gamma}$ in the CR.

pdf	**Parameter**	**Fit result**
f_{bkg}	b	$0.1 \pm 0.1\,\mathrm{GeV}^{-1}$

5.3.1 On the maximum likelihood estimation and MINUIT

One of the methods of estimating the parameters of a probability distribution is based on the construction of the combined probability distributions of all measurements in a sample, called a *likelihood function*. The likelihood function $\mathcal{L}$ is defined as the pdf f that characterizes the set of experimental observables $x_1, \cdots, x_n$, computed for the specific values they assume in the sample observed, and for given values of the

unknown parameters $\vec{\theta} = (\theta_1, \cdots, \theta_m)$. The central values of the parameters to estimate are obtained by finding the point in the parameter space that maximizes $\mathcal{L}$, which is referred to as *maximum-likelihood estimator*. This is valid under the conditions of the central limit theorem, which states that the average of n independent random variables $x_1, \cdots, x_n$, each distributed according to a pdf with finite variance, can be approximated to a Gaussian distribution in the limit of $n \to \infty$, regardless of the underlying pdf. It is also implicitly assumed that an absolute maximum exists, otherwise the determination of the maximum-likelihood estimator is not unique.

If N repeated measurements of the n observables $x_1, \cdots, x_n$ are performed, it is necessary to consider the probability density corresponding to the total sample

$$\mathbf{x} = (x_1^1, \cdots, x_n^1), \cdots, (x_1^N, \cdots, x_n^N). \tag{5.5}$$

Assuming that each event is independent, the likelihood function of a sample of N events can be written as the product of the pdfs corresponding to the measurement of each single event, that is:

$$\mathcal{L}(\mathbf{x}; \vec{\theta}) = \prod_{i=1}^{N} f(x_1^i, \cdots, x_n^i; \theta_1, \cdots, \theta_m). \tag{5.6}$$

The logarithm of the likelihood function is often computed, so that the product of many terms in the likelihood definition is turned into the sum of the logarithm of such terms:

$$-\ln \mathcal{L}(\mathbf{x}; \vec{\theta}) = -\sum_{i=1}^{N} \ln f(x_1^i, \cdots, x_n^i; \theta_1, \cdots, \theta_m). \tag{5.7}$$

The maximization of $\mathcal{L}$, equivalent to the minimization of the negative logarithmic likelihood $-\ln \mathcal{L}$, can be performed analytically only in the simplest circumstances. Most of the realistic cases require instead the use of numerical methods, implemented as computer algorithms such as MINUIT [93], a tool for physics analysis conceived to find the minimum of a multi-parameter function and analyze the shape of this function in its proximity. In this case, the minimization is based on the steepest descent direction in the parameter space, which is determined from a numerical evaluation of the gradient of $-\ln \mathcal{L}$.

Once the estimators $\hat{\theta}$ of the parameters of interest have been determined, it is necessary to define a *confidence interval* that corresponds to a statistical coverage of 68.27%, generally referred to as "$1\,\sigma$". In the presence of a large number of measurements, it is possible to use a Gaussian approximation to the distribution of the maximum-likelihood estimator. In this analysis, the estimate of the uncertainties in the parameters extracted from maximum likelihood fits is based on the numerical algorithm HESSE, which works under the assumption that the likelihood behaves like an n-dimensional Gaussian. The n-dimensional covariance matrix is then calculated as the inverse of the matrix containing the second-order partial derivatives of the negative logarithm of the likelihood function, i.e.:

$$C_{ij}^{-1} = -\frac{\partial \ln \mathcal{L}(x_1, \cdots, x_n; \theta_1, \cdots, \theta_m)}{\partial \theta_i \, \partial \theta_j}. \tag{5.8}$$

This covariance matrix gives the n-dimensional confidence contour with the correct coverage if the likelihood is approximately Gaussian. The variance relative to the

parameter θ is thus:

$$\frac{1}{\sigma_\theta^2} = \frac{\partial^2(-\ln \mathcal{L})}{\partial \theta^2}. \tag{5.9}$$

Therefore, errors computed using the HESSE algorithm are more reliable in the approximation of high statistics. This, however, could represent an issue when estimating the uncertainties in the branching ratio of $\mathrm{W} \to \pi\gamma$ by fitting the $m_{\pi\gamma}$ distribution in the data SR. For this reason, $\mathcal{B}(\mathrm{W} \to \pi\gamma)$ will be eventually calculated through a complete scan of the likelihood function (see Chapter 7).

Chapter 6

Systematic uncertainties

The measurement of $\mathcal{B}(\mathrm{W} \to \pi\gamma)$ presented in this thesis is affected by several sources of systematic uncertainty. This chapter outlines the way these uncertainties are estimated and how they affect the final result.

6.1 Systematic uncertainties as nuisance parameters

In Chapter 5, I described the extraction of the values of the parameters of the pdf used to fit the $m_{\pi\gamma}$ distribution in the data SR. This method is based on the maximization of the likelihood function, or, more precisely, on the minimization of the negative log-likelihood calculated upon the pfd outlined in Eq. (5.1). Systematic uncertainties are accounted for by treating them as nuisance parameters included in the f pdf in the form of *log-normal* pdfs.

In order to simplify the description of this procedure, I consider the likelihood function $\mathcal{L}(x;\theta_i)$ of an observable x, containing a single parameter θ_i. Supposing to know its value $\hat{\theta}_i$ that maximizes the likelihood, and the associated uncertainty $\sigma_{\hat{\theta}_i}$, it is possible to define an additional parameter:

$$k = 1 + \frac{\sigma_{\hat{\theta}_i}}{\hat{\theta}_i}. \tag{6.1}$$

The term above is included in the definition of the number of signal events, so that:

$$N_{\mathrm{sig}} = \sigma_{\mathrm{t\bar{t}}}\, \mathcal{B}(\mathrm{W}^{\mp} \to \ell^{\mp}\nu)\, \mathcal{B}(\mathrm{W}^{\pm} \to \pi^{\pm}\gamma)\; \mathcal{L}_{\mathrm{int}}\, \varepsilon_{\ell}\, k^{\beta}, \quad \ell = \mu, \mathrm{e}, \tau_{\ell}, \tag{6.2}$$

where β is a free parameter. The other terms of the equation are discussed following Eq. (5.2) in Chapter 5. Under these assumptions, a Gaussian pdf $\mathcal{G}(\beta; 0, 1)$ can be created with the following characteristics:

- β as the observable;
- 0 as the central value;
- 1 as the width.

The likelihood $\mathcal{L}(x;\theta_i)$ is then multiplied by $\mathcal{G}(\beta; 0, 1)$, so to obtain:

$$\mathcal{L}(x; \theta_i, \sigma_{\theta_i}, \beta) = \mathcal{L}(x;\theta_i)\, \mathcal{G}(\beta; 0, 1). \tag{6.3}$$

In practice, the new likelihood function undergoes a Gaussian penalty that models the systematic uncertainty in the knowledge of a given model parameter and its effect on the parameter of interest, the $\mathrm{W} \to \pi\gamma$ branching ratio. It must be emphasized that, with this approach, the value of the parameter θ_i that maximizes the likelihood function, namely $\hat{\theta}_i$, is unmodified. The way each of the nuisance parameters is inserted into the likelihood function is discussed in detail in the next sections.

6.2 Sources of systematic uncertainties

The sources of systematic uncertainties accounted for in this analysis are:

- the uncertainty in the $\mathrm{t\bar{t}}$ production cross section $\sigma_{\mathrm{t\bar{t}}}$ in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$;
- the uncertainty in the integrated luminosity $\mathcal{L}_{\mathrm{int}}$ collected by CMS in 2016 to 2018;
- the uncertainty in the product of the signal efficiency and the acceptance ε_ℓ (with $\ell = \mu, \mathrm{e}$);
- the uncertainty associated to the signal parametrization;
- the uncertainty associated to the background parametrization.

6.2.1 Top-antitop pair production cross section

The $\mathrm{t\bar{t}}$ production cross section in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ contributes in this analysis to the definition of the signal normalization (see Eq. (5.2)). This cross section was measured by the CMS Collaboration using data collected in 2015, corresponding to an integrated luminosity $\mathcal{L}_{\mathrm{int}} = 2.2\,\mathrm{fb}^{-1}$ [80]. Events were analyzed in which the final state included one electron, one muon, and two or more jets, at least one of which was identified as originating from hadronization of a b quark. The measured cross section is:

$$\sigma_{\mathrm{t\bar{t}}} = 815 \pm 9\,(\mathrm{stat}) \pm 38\,(\mathrm{syst}) \pm 19\,(\mathrm{lumi})\,\mathrm{pb}, \tag{6.4}$$

with the main sources of experimental uncertainties being of systematic nature. These include uncertainties in the knowledge of the PDFs of the proton, in the hadronization model employed, and in the parton shower scale. An uncertainty of 43 pb, corresponding to the quadratic sum of the statistical, systematic, and luminosity-related components, is the one used in the analysis.

The use of a more recent measurement of the $\mathrm{t\bar{t}}$ production cross section from the CMS Collaboration [94] was excluded, even though it exploits a larger integrated luminosity which leads to a slightly lower statistical uncertainty. Nevertheless, this latter measurement exploits the same 2016 collision data which constitutes part of the dataset this search for $\mathrm{W} \to \pi\gamma$ is based on. Therefore, there is a partial correlation between this $\mathrm{t\bar{t}}$ cross section measured in 2016 and the main background source of this analysis (Table 4.10).

The branching ratio $\mathcal{B}(\mathrm{W} \to \ell\nu)$ is also a term contributing to the signal normalization. Its value

$$\mathcal{B}(\mathrm{W} \to \ell\nu) = 0.1086 \pm 0.0009 \tag{6.5}$$

is an average on the three lepton flavors $\ell = \mu, \mathrm{e}, \tau$ [17]. Its associated uncertainty is considered negligible with respect to that in the $\mathrm{t\bar{t}}$ cross section. Similarly, the uncertainties in the leptonic branching ratios of the τ lepton [17]

$$\begin{aligned} \mathcal{B}(\tau \to \mu\bar{\nu}_\mu\nu_\tau) &= 0.1739 \pm 0.0004 \\ \mathcal{B}(\tau \to \mathrm{e}\bar{\nu}_\mathrm{e}\nu_\tau) &= 0.1782 \pm 0.0004 \end{aligned} \tag{6.6}$$

are neglected.

6.2.2 Integrated luminosity

The data sample analyzed in this search for the decay $W \to \pi\gamma$ corresponds to an integrated luminosity $\mathcal{L}_{\text{int}} \simeq 137\,\text{fb}^{-1}$, recorded by the CMS experiment during 2016 ($35.9\,\text{fb}^{-1}$), 2017 ($41.5\,\text{fb}^{-1}$), and 2018 ($59.7\,\text{fb}^{-1}$). The integrated luminosities of the 2016, 2017, and 2018 data-taking periods are individually known with uncertainties of 2.3, 2.5, and 2.3%, respectively [95–97], whereas the total 2016–2018 integrated luminosity has an uncertainty of 1.8%; the improvement in precision reflects the uncorrelated time evolution of some systematic effects. Therefore, the signal normalization (see Eq. (5.2)) is known with an uncertainty of about $2\,\text{fb}^{-1}$ in the integrated luminosity:

$$\mathcal{L}_{\text{int}} = 137 \pm 2\,\text{fb}^{-1}. \tag{6.7}$$

6.2.3 Signal efficiency and acceptance

The term ε_ℓ, which appears in the definition of N_{sig} in Eq. (5.2), indicates the product of several efficiency components and the detector acceptance A:

$$\varepsilon_\ell = \varepsilon_{\text{trigger}}\,\varepsilon_{\text{identification}}\,\varepsilon_{\text{reconstruction}}\,\varepsilon_{\text{selection}}\,A. \tag{6.8}$$

The subscript $\ell = \mu, \text{e}$ suggests that ε_ℓ is computed separately for the two lepton channels, and evaluated as the ratio between the number of simulated signal events N_ℓ that survive the whole selection procedure, and the initial number of signal events produced per channel ($N_{\text{T}}/2$). These numbers are reported in Table 6.1. As already

TABLE 6.1: For each year of data-taking, the initial number of signal events (N_{T}) and the number of signal events surviving the selection in the muon (N_μ) and in the electron (N_{e}) channels are reported.

Year	N_{T}	N_μ	N_{e}
2016	80000	3851	2193
2017	79985	3587	2234
2018	79905	3719	2170

discussed, the signal in the two lepton channels receives a contribution from both events containing a W boson that decays directly into a muon or an electron plus a neutrino, and from events where it decays into τ leptons, subsequently decaying into leptons too. Since the branching ratio for these two processes is different ($\mathcal{B}(\tau \to \mu\bar{\nu}_\mu\nu_\tau) + \mathcal{B}(\tau \to \text{e}\bar{\nu}_\text{e}\nu_\tau) \simeq 0.3521$), the term N_{T} must be corrected by a factor $(1 + 1 + 0.3521)/3$, with 1 being the weight associated to $W \to \mu(\text{e})\nu$. Consequently, the products of efficiency and acceptance per year of data taking and per lepton channel become:

$$\begin{aligned}
\varepsilon_\mu^{2016} &= 0.12 \pm 0.01, \\
\varepsilon_\text{e}^{2016} &= 0.07 \pm 0.01, \\
\varepsilon_\mu^{2017} &= 0.11 \pm 0.01, \\
\varepsilon_\text{e}^{2017} &= 0.07 \pm 0.01, \\
\varepsilon_\mu^{2018} &= 0.12 \pm 0.01, \\
\varepsilon_\text{e}^{2018} &= 0.07 \pm 0.01,
\end{aligned} \tag{6.9}$$

where the quoted uncertainties correspond to the statistical ones, which follow a binomial distribution:

$$\sigma_{\varepsilon_\ell} = \frac{\varepsilon_\ell\,(1-\varepsilon_\ell)}{N_\mathrm{T}}. \tag{6.10}$$

Nevertheless, the statistical component is not the only contribution to the total uncertainty in ε_ℓ, but others are included, concerning the BDT modeling, the modeling of kinematic variables for the signal in PYTHIA, the scale factors that correct for differences between detector conditions in collision data and simulation, and the charge-misidentification in the tracking algorithm.

BDT modeling

As discussed in Section 4.3.2, the two BDT classifiers used to perform the multivariate event selection in this analysis are trained using simulated events. Therefore, the way the BDT input variables are modeled by the MC generators affects the selection and has an impact on the signal efficiency. To assess the size of this impact and propagate it into an uncertainty in ε_ℓ, the following steps are taken. The performance of a BDT trained and validated with the nominal input variables is compared with that of a BDT trained with the same variables, and validated with variables to which a Gaussian smearing or a shift is applied. In particular, the smeared variables are p_T^ℓ, p_T^π, p_T^γ, and $p_\mathrm{T}^\mathrm{miss}$, and the width of the smearing corresponds to 5% of their values, that is the largest expected resolution on these variables. For charged particle tracks, it includes the uncertainties in the tracking efficiency too.

Contrarily to the lepton, pion, and photon momenta, the number of b-tagged jets, n_b, is an integer. Thus, it is increased by 1 for 5% of the events in the training samples of 2016 and 2017. For what concerns the 2018 training sample, the power supply failure of two HCAL sectors during part of the data-taking led to larger uncertainties in the jet energy calibration. In Section 4.2.8, a test was described that estimated the impact of these uncertainties in the computation of n_b, showing a maximum discrepancy of $\approx$8% for $n_\mathrm{b} = 5$ when the jet p_T was scaled down. Therefore, in the training samples of 2018, n_b is increased by one for 10% of the events, corresponding to the sum in quadrature of 5 and 8%.

The pion isolation is not smeared, but systematically shifted upwards by 10%. This choice stems from the observation of the event distribution as a function of the pion-isolation variable, as illustrated in Fig. 4.4, before this variable is used in the BDT training. The processes leading to an erroneous measurement of $\Sigma p_\mathrm{T}/p_\mathrm{T}^\pi$ may be multiple: for instance, excluding some good candidate particles from the p_T-sum would push the distribution of this variables to lower values; including in the p_T-sum particles originating from pile up would push it to higher values; measuring the wrong p_T for any kind of detector effects would potentially push it in both directions. Given these considerations, on an event-by-event basis it would seem reasonable to consider a smearing of $\Sigma p_\mathrm{T}/p_\mathrm{T}^\pi$. Nevertheless, the distribution of the pion isolation suggests that the MC simulation overestimates the data at low values of this variable more than it underestimates them at high values. To take into account and compensate this one-sided effect, the pion isolation is shifted upwards by 10% for all the MC simulated processes. The event distribution as a function of the shifted pion-isolation variable is presented in Fig. 6.1.

The performance of the BDTs trained and validated using nominal input variables and of those validated using smeared/shifted ones can be observed in Fig. 6.2, for the two lepton channels separately. For a given background rejection efficiency, the largest observed variation in terms of signal efficiency is 1% in the muon channel, and

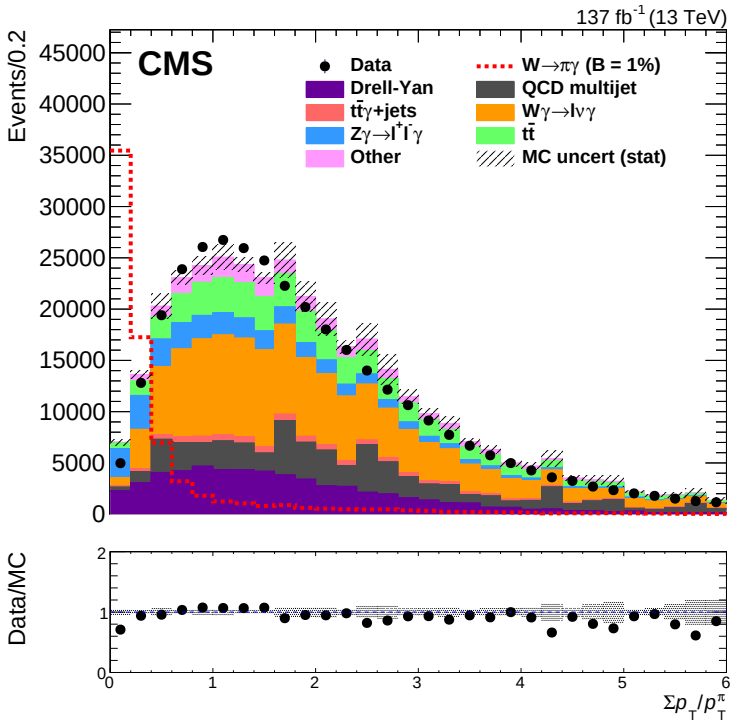

FIGURE 6.1: Event distribution as a function of the pion-isolation variable $\Sigma p_{\mathrm{T}}/p_{\mathrm{T}}^{\pi}$. All the simulated MC distributions are shifted by 10% towards higher values of this variable. The MC signal is given by the dashed red line and corresponds to a 1% branching fraction for the $\mathrm{W} \to \pi\gamma$ decay. The statistical uncertainties in the data are small and thus not visible. In the lower panel, the ratio between data and the background component of the MC is shown. The gray bands represent the statistical uncertainty in the MC background.

2% in the electron channel. Therefore, these uncertainties are summed in quadrature to the statistical uncertainty in ε_{ℓ}.

The procedures described in this section might look overly conservative, as the typical resolution on the BDT input variables is generally better than 5%. On the other hand, the final impact on the signal efficiency is rather small and has a minimal influence on the final result.

PYTHIA modeling – transverse momentum

The procedure for modeling a process in PYTHIA represents another source of systematic uncertainty in ε_{ℓ}, since inaccuracies of the model have an impact on the signal acceptance. As a first step in the evaluation of this uncertainty, the p_{T} distribution of the generated signal W bosons is considered. As discussed in Section 3.2.1, this spectrum is corrected to match that obtained from NLO $\mathrm{t\bar{t}}$ MC generated events in POWHEG v2, which includes the proper W boson polarization from top quark decays, otherwise not modeled in my MC signal sample. To assess the effect of a systematic change in the $p_{\mathrm{T}}^{\mathrm{W}}$ reweighting, the $p_{\mathrm{T}}^{\mathrm{W}}$ spectrum is shifted upwards and downwards by 5%, which roughly corresponds to the maximum uncertainty in the weights used, as shown in Fig. 3.2. By making the ratio between the shifted and the nonshifted distributions of $p_{\mathrm{T}}^{\mathrm{W}}$, various sets of weights are obtained that I use to reweight the signal. As an example, the $p_{\mathrm{T}}^{\mathrm{W}}$ spectrum of the 2018 MC signal sample is shown in Fig. 6.3 for the shifted and the nonshifted case, together with the two sets of weights produced. The distributions for the other years are similar.

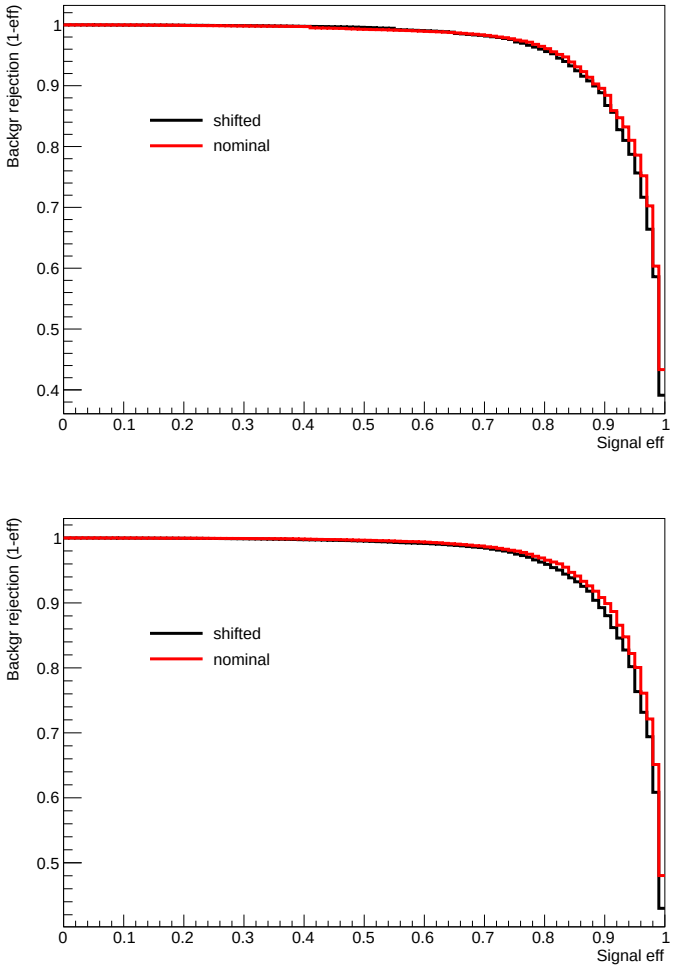

FIGURE 6.2: Background rejection efficiency as a function of signal efficiency, as obtained using the BDT in the muon (upper) and the electron (lower) channels. The red curves indicate that the BDT is trained and validated on the nominal input variables, whereas the black curves are made using the smeared/shifted variables in the BDT validation.

Afterwards, the performance of the BDT trained and validated through the nominal variables is compared to that of a BDT trained on the nominal variables, and tested using MC events where the $p_{\mathrm{T}}^{\mathrm{W}}$-based reweighting is shifted by 5%. The procedure is performed after merging the samples corresponding to the three years of data-taking, but separately for the muon (Fig. 6.4) and the electron (Fig. 6.5) channels. For a given background rejection efficiency, the largest observed variation in terms of signal efficiency is 2% in the muon channel, and 3% in the electron channel. Therefore, these uncertainties are summed in quadrature to the total uncertainty in ε_{ℓ}.

PYTHIA modeling – angular distribution

The signal acceptance also depends on the angular spectrum of the particles involved in the $\mathrm{W} \to \pi\gamma$ decay. To evaluate the uncertainty in ε_{ℓ} related to the PYTHIA modeling of the W polarization, I modify this angular spectrum. Specifically, I consider the generated distribution in θ^{*}, the helicity angle between the pion direction in the W boson rest frame and the direction of the W boson momentum in the laboratory. The

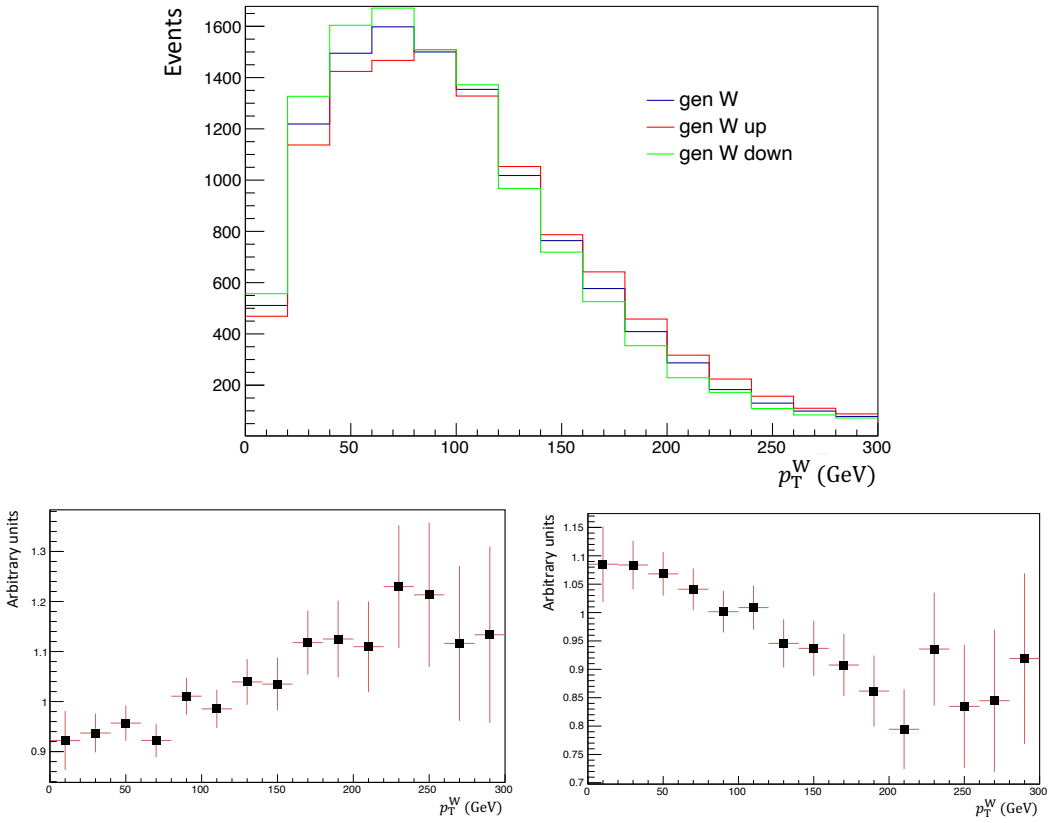

FIGURE 6.3: Event distribution as a function of p_T^W as simulated by PYTHIA for the 2018 signal, with and without the 5% shifting (upper plot). The lower plots show the weights obtained as the ratio between the red and the blue distributions in the upper plot (left), and the green and the blue distributions (right).

distribution in this variable is indeed expected to be dependent on the polarization of W bosons arising from top quark decays, which is not accounted for in the PYTHIA simulated events used in this analysis.

Let us consider the spherical harmonics

$$Y_\ell^m(\theta,\phi) = (-1)^m \left[\frac{2\ell+1}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}\right]^{\frac{1}{2}} P_\ell^m(\cos\theta)\, e^{im\phi}, \tag{6.11}$$

where P_ℓ^m are Legendre polynomials, and the two quantum numbers ℓ (orbital angular momentum) and m (projection of ℓ on one axis) must satisfy:

$$\begin{cases} \ell = 0, 1, 2, 3... \\ m = -\ell, -\ell+1, ..., \ell-1, \ell. \end{cases} \tag{6.12}$$

A W boson emerging from top quark decay may be observed in three possible eigenstates $|\,\ell; m>$:

$$\begin{aligned} &\mathrm{W}^{(1,2)} : |\,1; \pm 1> \\ &\mathrm{W}^{(3)} : |\,1; 0>, \end{aligned} \tag{6.13}$$

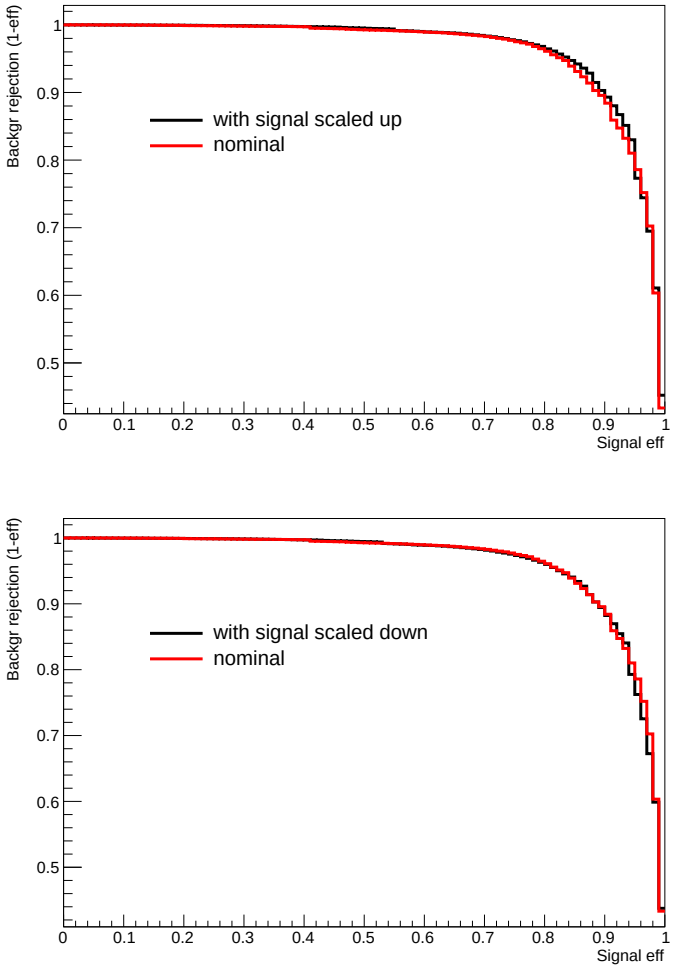

FIGURE 6.4: Background rejection efficiency as a function of signal efficiency, as obtained using the BDT with events falling in the muon channel. The red curves indicate that the BDT is trained and validated on the nominal input variables, whereas the black curves are obtained using the $p_{\mathrm{T}}^{\mathrm{W}}$-based signal reweighting in the BDT validation, with the $p_{\mathrm{T}}^{\mathrm{W}}$ shifted upwards (top) and downwards (bottom).

constraining the possible eigenstates of its decay products, the pion and the photon, to:

$$\begin{aligned} \pi^{(1,2)} &: |\, 0;0 > \quad \gamma^{(1,2)} : |\, 1;\pm 1 > \\ \pi^{(3)} &: |\, 1;\pm 1 > \quad \gamma^{(3)} : |\, 1;\pm 1 > \end{aligned} \tag{6.14}$$

For what regards the pion angular distribution in the the aforementioned variable θ^*, (1,2) lead to a spherical harmonic:

$$Y_0^0 = \frac{1}{2\sqrt{\pi}}, \tag{6.15}$$

whereas in (3):

$$Y_1^1 = -\frac{1}{2}\sqrt{\frac{3}{2\pi}}\, e^{i\phi} \sin\theta. \tag{6.16}$$

Given these considerations, the states (1,2) correspond to an isotropic distribution in

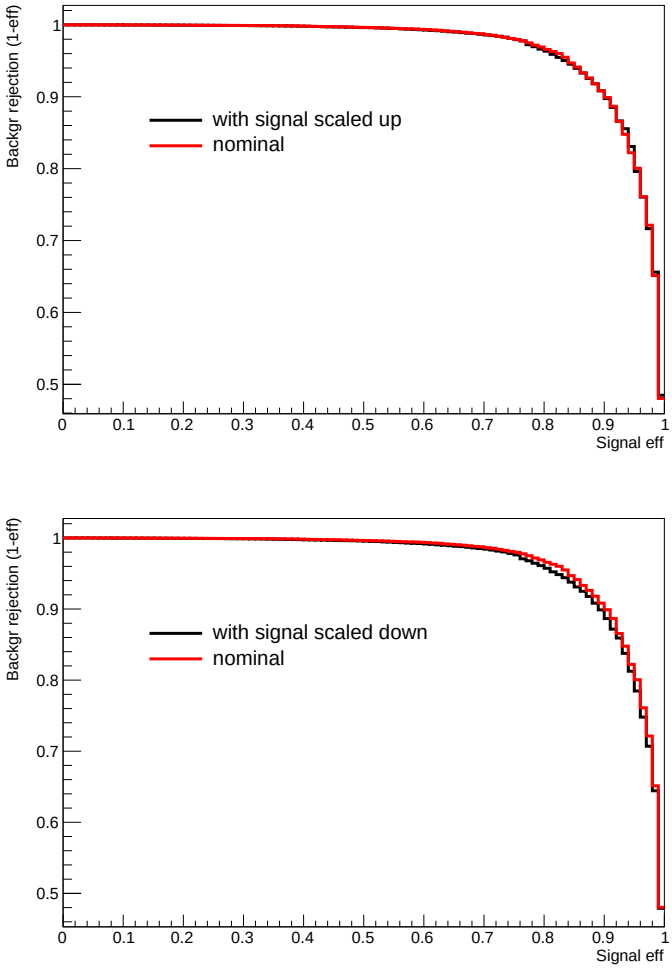

FIGURE 6.5: Background rejection efficiency as a function of signal efficiency, as obtained using the BDT with events falling in the electron channel. The red curves indicate that the BDT is trained and validated on the nominal input variables, whereas the black curves are obtained using the $p_{\mathrm{T}}^{\mathrm{W}}$-based signal reweighting in the BDT validation, with the $p_{\mathrm{T}}^{\mathrm{W}}$ shifted upwards (top) and downwards (bottom).

θ^*, whereas (3) yields to a sine squared distribution of the same angle. Therefore, isotropic and $\sin^2\theta^*$-based distributions are produced, corresponding to transverse and longitudinal polarizations of the W boson, respectively. These are shown in Fig. 6.6, together with the observed distribution, expressed in terms of $\cos\theta^*$. By taking the ratio of each of these alternative distributions to the observed one, two sets of weights are obtained and used to renormalize the signal. In similarity with $p_{\mathrm{T}}^{\mathrm{W}}$, a BDT is trained on a set of nominal events, and validated on a set that has the signal renormalization implemented. Its performance is compared with a BDT trained and validated on nominal events, as can be seen in Fig. 6.7 and in Fig. 6.8 for the muon and the electron channels, respectively. The most significant changes in terms of ε_ℓ for a given background rejection efficiency are about 3% and 5% for the muon and the electron channels, respectively. Therefore, these uncertainties are summed in quadrature to the total uncertainty in ε_ℓ.

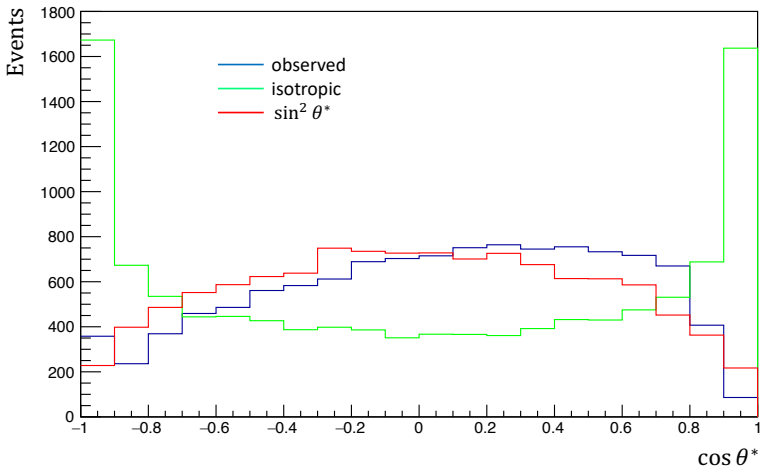

FIGURE 6.6: Event distribution as a function of $\cos\theta^*$ as simulated by PYTHIA (blue) for the 2018 signal. Alternative distributions are also shown, derived from an isotropic (green) and $\sin^2\theta^*$ (red) spectrum for θ^*.

Scale factors and charge misidentification

An additional source of uncertainty in the signal efficiency arise from the scale factors applied to simulated events (Section 4.2.10). The sum of statistical and systematic uncertainty associated to the SFs used for muons, electrons, and photons is 1% or lower, in the kinematic regions this search is sensitive to. Since in the same event both muon (or electron) and photon SFs are used, the total uncertainty is calculated as the quadratic sum of the uncertainty associated to each particle, that is 1.4%.

A further uncertainty of 1% in ε_ℓ, derived from detector performance studies [66], covers the effects of charge misidentification in the tracking algorithm. These uncertainties are summed in quadrature to the statistical and to the other systematic uncertainties, described above, in the signal efficiency.

The ε_ℓ for each year and lepton channel are summarized in Table 6.2, with the associated total uncertainties.

TABLE 6.2: Product of signal efficiency and acceptance per year and per lepton channel, after the multivariate event selection.

Year	ε_μ	ε_e
2016	0.12 ± 0.01	0.07 ± 0.01
2017	0.11 ± 0.01	0.07 ± 0.01
2018	0.12 ± 0.01	0.07 ± 0.01

6.2.4 Signal parametrization

The parameters m and σ_m represent the peak position and the width of the Gaussian component of the Double Crystal Ball pdf used to describe the signal lineshape. Their central values and uncertainties are determined in the fit to the MC signal $m_{\pi\gamma}$ distribution, as discussed in Section 5.2. When the fit to the same distribution in the data SR is performed, these parameters are fixed to their central values. To take into account the finite statistics in the MC sample, I consider the value of σ_m and its uncertainty as extracted from the signal MC fit. Moreover, I include a 1% uncertainty

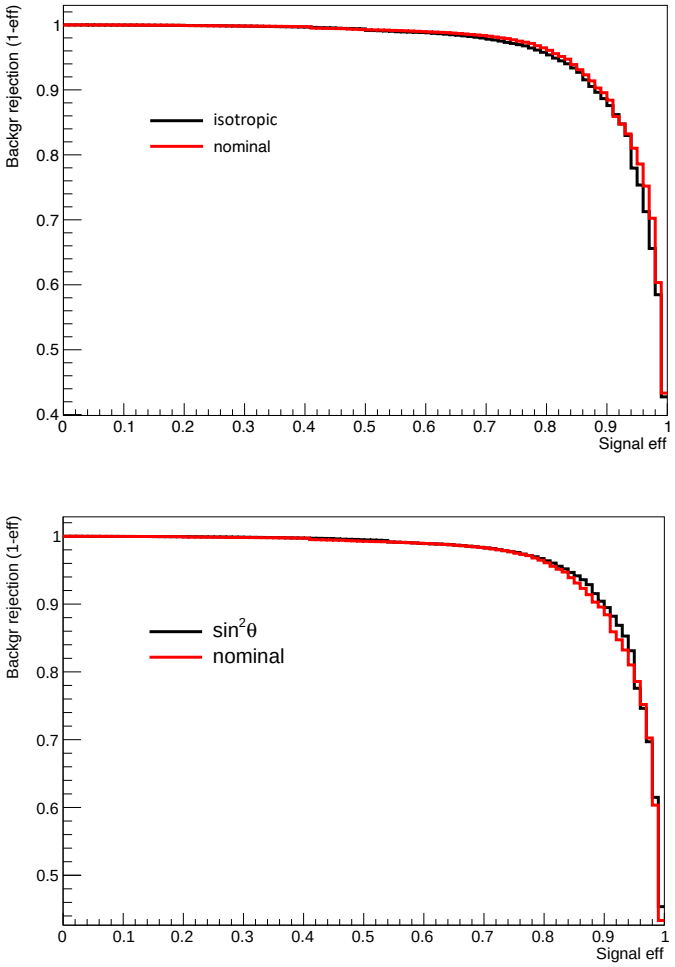

FIGURE 6.7: Background rejection efficiency as a function of signal efficiency, as obtained using the BDT with events falling in the muon channel. The red curves indicate that the BDT is trained and validated on the nominal input variables, whereas the black curves are obtained using the isotropic (upper) or $\sin^2\theta^*$-based (lower) signal reweighting in the BDT validation.

in m to account for the photon energy scale, which dominates the uncertainty in the pion-photon invariant mass:

$$\begin{aligned} m &= 79.9 \pm 0.8\,\text{GeV} \\ \sigma_m &= 3.3 \pm 0.2\,\text{GeV} \end{aligned} \tag{6.17}$$

The following procedure is performed: first, the data are fitted in the SR fixing the values of the peak and the width of the DCB function to their central values (nominal case); then, the fit is repeated after increasing and decreasing these parameters by their uncertainties (shifted parameters). The results are compared by calculating a set of pulls in the parameter of interest $\mathcal{B}(\text{W} \to \pi\gamma)$:

$$\frac{|\mathcal{B}_{\text{nominal}} - \mathcal{B}_{\text{shifted}}|}{\sigma_{\mathcal{B}_{\text{nominal}}}}, \tag{6.18}$$

where $\sigma_{\mathcal{B}_{\text{nominal}}}$ represents the uncertainty associated to $\mathcal{B}_{\text{nominal}}$. The largest pull is found to be 10.6% and is taken as an estimate of the systematic uncertainty associated

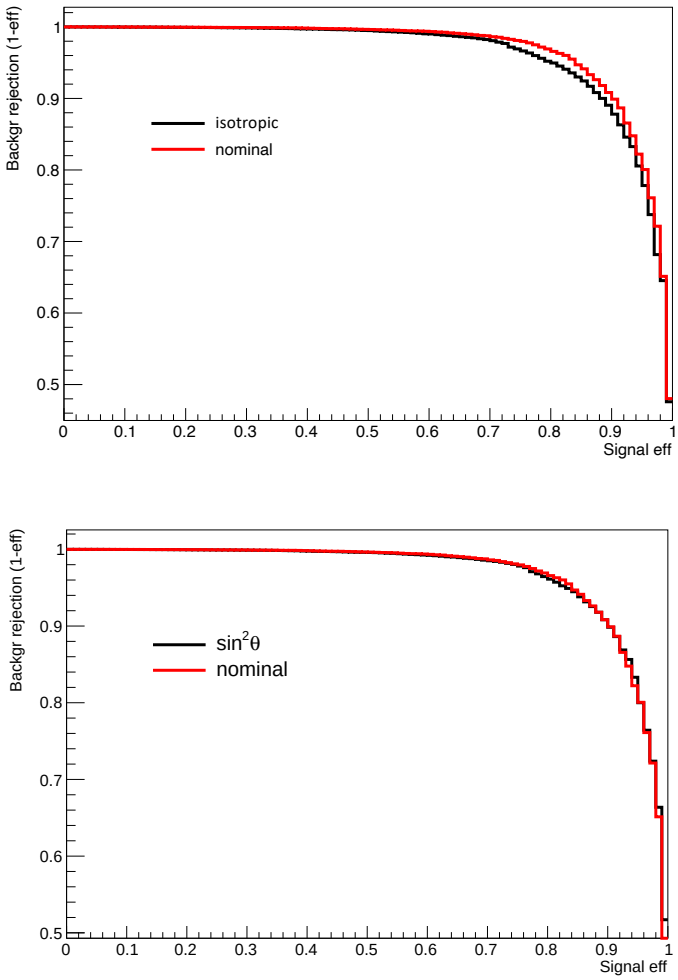

FIGURE 6.8: Background rejection efficiency as a function of signal efficiency, as obtained using the BDT with events falling in the electron channel. The red curves indicate that the BDT is trained and validated on the nominal input variables, whereas the black curves are obtained using the isotropic (upper) or $\sin^2\theta^*$-based (lower) signal reweighting in the BDT validation.

with the parametrization of the signal. Since the other parameters of the signal lineshape are highly correlated with each other, as well as with the peak and the width of the Gaussian component of the DCB function, it is not necessary to float them within their uncertainties.

It is important to remark that, in performing the procedure described above, the variables representing the parameter of interest in the fit, namely $\mathcal{B}_{\text{nominal}}$ and $\mathcal{B}_{\text{shifted}}$, are left free to assume negative values. From a physical point of view, a branching fraction is always a positive definite quantity, representing the probability for an unstable particle to decay into a specific final state. Indeed, it is imposed to be positive when performing the fit on the $m_{\pi\gamma}$ spectrum in the data SR described in the next chapter, where an underfluctuation of the background could make it assume negative values. Nevertheless, imposing a positive definite $\mathcal{B}(\mathrm{W} \to \pi\gamma)$ when estimating the systematic uncertainty in the signal parametrization would not guarantee a proper frequentist coverage.

6.2.5 Background parametrization

The choice of the pdf used to describe the background has an influence on the measured $W \to \pi\gamma$ branching fraction. To estimate this effect, a fit is performed to the data in the SR with the nominal background pdf, i.e., a linear function (see Section 5.3). The same distribution is then fitted again using an alternative pdf, namely an exponential. The goodness of the fit performed with the exponential pdf is verified with a minimum χ^2 method prior to performing this procedure, as can be observed in Fig. 6.9.

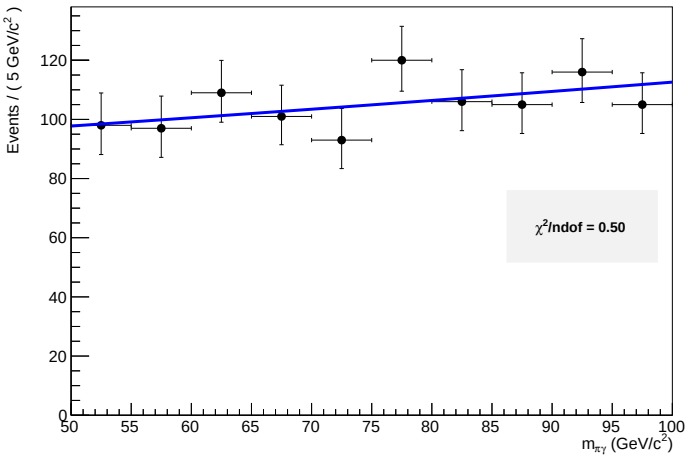

FIGURE 6.9: Fit to the $m_{\pi\gamma}$ spectrum in the data CR, using an exponential pdf.

The following pull is calculated for the branching fraction of $W \to \pi\gamma$:

$$\frac{|\mathcal{B}_{\text{lin}} - \mathcal{B}_{\text{exp}}|}{\sigma_{\mathcal{B}_{\text{lin}}}}, \tag{6.19}$$

where $\mathcal{B}_{\text{lin}}$ represents the value of the $W \to \pi\gamma$ branching fraction as extracted from the fit with a linear function, $\sigma_{\mathcal{B}_{\text{lin}}}$ is its uncertainty, and $\mathcal{B}_{\text{exp}}$ indicates the value of the branching fraction extracted from the fit using the exponential function. The pull amounts to 14.6% and is taken as an estimate of the systematic uncertainty associated with the background parametrization.

Chapter 7

Results

In this final chapter, the results obtained from the fit to the data in the SR are presented, together with the calculation of the upper limit on the branching fraction of $W \to \pi\gamma$.

7.1 Fit to the $m_{\pi\gamma}$ spectrum in the signal region

The $m_{\pi\gamma}$ distribution in the data SR, shown in Fig. 7.1 for the combination of the two lepton channels and the three data-taking periods, is fitted with the pdf outlined in Eq. (5.1). The goodness of the fit is verified using a minimum χ^2 method that leads to $\chi^2/\text{ndof} = 0.62$. Large uncertainties in the MC backgrounds are mainly of statistical nature and arise from samples with large cross sections and relatively few simulated events, such as the QCD multijet processes.

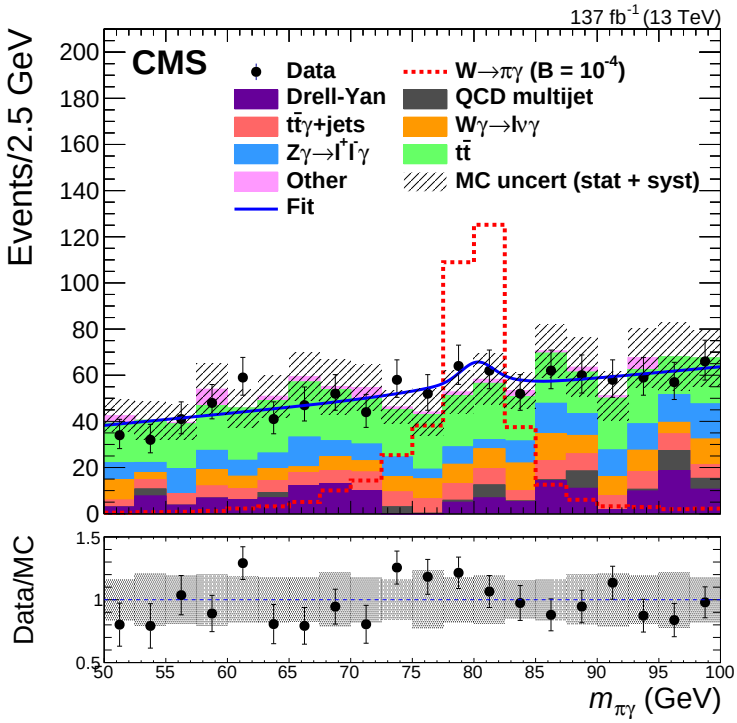

FIGURE 7.1: Event distribution as a function of $m_{\pi\gamma}$ for the combination of the lepton channels. The simulated MC distribution for the signal is given by the dashed red line and corresponds to a 10^{-4} branching fraction for the $W \to \pi\gamma$ decay. The uncertainties in the data are statistical only. The blue line represents the best fit to the data using the model described in Eq. (5.1). In the lower plot, the ratio between data and the background component of the MC is shown. The gray bands represent the uncertainty (statistical + systematic) in the MC background.

The values of the pdf parameters derived from the fit are listed in Table 7.1, including the background yield $N_{\text{bkg}}(\text{SR})$ and the slope b of the linear function used to describe the background (see Section 5.3). The signal yield N_{sig}, parametrized as in Eq. (5.2), is used to extract the parameter of interest:

$$\mathcal{B}(\text{W}^{\pm} \to \pi^{\pm}\gamma) = (5.6 \pm 4.1) \times 10^{-6}. \tag{7.1}$$

Separate fits to the same distribution in the muon and electron channels are also performed as a cross-check, as shown in Fig. 7.2. The pdfs used for the fit, f_{μ} and f_{e}, are similar to those used for the fit on the merged lepton channels, but contain a different definition of the channel-dependent parameters, such as the signal efficiency. The parameter estimation derived from the two fits is summarized in Table 7.2.

From a purely qualitative perspective, a small excess, which arises from the muon channel, is observed around the value of the W boson mass. To ascertain whether this is due to the presence of signal events or just to a statistical fluctuation of the background, the significance of the excess is calculated as the probability that the sample is inconsistent with the hypothesis that only background is present in the data.

7.2 On p-values and significance

To claim of discovery of a new signal within an observed data sample it is necessary to determine that the sample is sufficiently inconsistent with the hypothesis that only background is present in the data. The definition of "sufficiently inconsistent" is, of course, not unique. Nevertheless, the physics community has chosen to adopt some common criteria that are discussed in this section.

The inconsistency of the observation in the hypothesis of the presence of background only, typically assumed as a null hypothesis (H_0), can be evaluated using a test statistic t. The probability p that the considered test statistic t assumes a value greater or equal to that observed in the case of pure background fluctuation is called *p-value*. Implicitly, large values of t correspond to a more signal-like sample. The p-value has by construction a uniform distribution between 0 and 1 for the background-only hypothesis H_0, and tends to have small values in the presence of the signal (hypothesis H_1).

Alternatively to quoting a p-value, it is often preferred to report the equivalent number of standard deviations that correspond to an area equal to the p-value under the rightmost tail of a normal distribution. The significance (in units of $Z\sigma$) that

TABLE 7.1: Parameter estimation from the fit to the data distribution of $m_{\pi\gamma}$ in the SR. Uncertainties are computed with the HESSE algorithm implemented in MINUIT.

pdf	Parameter	Fit result
	b	$0.3 \pm 0.1\,\text{GeV}^{-1}$
f	$N_{\text{bkg}}(\text{SR})$	1018 ± 37
	$\mathcal{B}(\text{W}^{\pm} \to \pi^{\pm}\gamma)$	$(5.6 \pm 4.1) \times 10^{-6}$

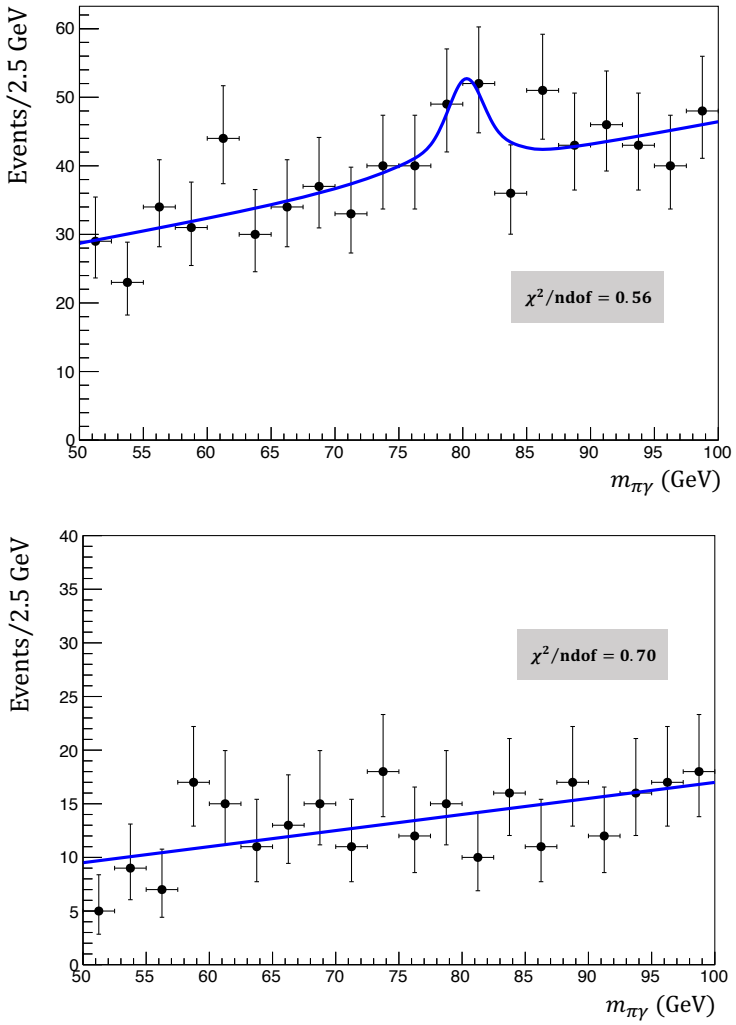

FIGURE 7.2: Event distribution as a function of $m_{\pi\gamma}$ for the SR in the muon (upper) and the electron (lower) channels. Only collision data are shown, with their statistical uncertainties.

TABLE 7.2: Parameter estimation from the fit to the data distribution of $m_{\pi\gamma}$ in the SR, for the two lepton channels separately. Uncertainties are computed with the HESSE algorithm implemented in MINUIT.

pdf	Parameter	Fit result
	b	$0.2 \pm 0.1\,\text{GeV}^{-1}$
f_μ	$N_{\text{bkg}}(\text{SR})$	750 ± 32
	$\mathcal{B}(\text{W}^\pm \to \pi^\pm\gamma)$	$(4.1 \pm 2.6) \times 10^{-5}$
	b	$0.3 \pm 0.1\,\text{GeV}^{-1}$
f_e	$N_{\text{bkg}}(\text{SR})$	265 ± 16
	$\mathcal{B}(\text{W}^\pm \to \pi^\pm\gamma)$	$(0.1 \pm 2.0) \times 10^{-5}$

corresponds to a given p-value can be obtained with the following transformation:

$$\text{p} = \int_Z^\infty \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\,\text{d}x = 1 - \Phi(Z) = \Phi(-Z) = \frac{1}{2}\left[1 - \text{erf}\left(\frac{Z}{\sqrt{2}}\right)\right], \qquad (7.2)$$

where $\Phi(-Z)$ indicates the cumulative function of a standard normal distribution. By convention, when the significance of the fluctuation under investigation is at least $3\,\sigma$ ($Z = 3$), which corresponds to a probability of background fluctuation of 1.35×10^{-3}, the *evidence* of the signal is claimed. The *observation*, or *discovery* of a signal typically requires a significance of $5\,\sigma$ ($Z = 5$), corresponding to a p-value of 2.87×10^{-7}.

For what regards this search for $\mathrm{W} \to \pi\gamma$, the excess observed above the expected background in Fig. 7.1 has a p-value of 0.055, corresponding to a significance of about $1.6\,\sigma$. It is therefore not possible to conclude that a signal has been observed.

7.3 Confidence intervals

Section 5.3.1 introduced the method implemented in MINUIT for the determination of the uncertainties in maximum likelihood estimates, based on the scan of the negative log-likelihood around its minimum. This method, however, guarantees coverage of the interval in the Gaussian approximation only. In many cases the provided level of approximation is sufficient, but not in this analysis, where the number of events considered is small and the likelihood function may exhibit significant deviations from the Gaussian approximation.

A more rigorous and general treatment of confidence intervals under the frequentist approach is due to Jerzy Neyman [98]. Neyman's procedure for the determination of confidence intervals is fundamentally composed by two steps:

- the construction of a *confidence belt*;
- the inversion of the confidence belt.

These two steps are discussed in the following.

7.3.1 Construction of the confidence belt

For simplicity, the case of a variable x is described, whose distribution follows a pdf which depends on a single unknown parameter θ. In order to determine the confidence belt, a scan of the parameter space is made, varying θ within its allowed range. For each fixed value of $\theta = \theta_0$, the pdf $f(x;\theta_0)$ describing the distribution of x is uniquely determined. According to $f(x;\theta_0)$, an interval $[x_1(\theta_0), x_2(\theta_0)]$ is defined, whose corresponding probability is equal to the desired *confidence level*: $\mathrm{CL} = 1 - \alpha$, usually equal to 68.27% ($1\,\sigma$), 90% or 95%:

$$1 - \alpha = \int_{x_1(\theta_0)}^{x_2(\theta_0)} f(x;\theta_0)\,\mathrm{d}x. \tag{7.3}$$

Figure 7.3 shows a graphical representation of the Neyman confidence belt construction.

The choice of the interval defined by $x_1(\theta_0)$ and $x_2(\theta_0)$ is referred to as *ordering rule*. It is somewhat arbitrary, since different intervals may provide the same probability.

7.3.2 Inversion of the confidence belt

This is the second phase of Neyman's procedure. Considering a measurement of $x = x_0$, the confidence interval for the parameter θ is determined by inverting the Neyman belt. Two extreme values $\theta_1(x_0)$ and $\theta_2(x_0)$ are identified by intersecting the vertical line at $x = x_0$ with the two boundary curves of the belt. An interval

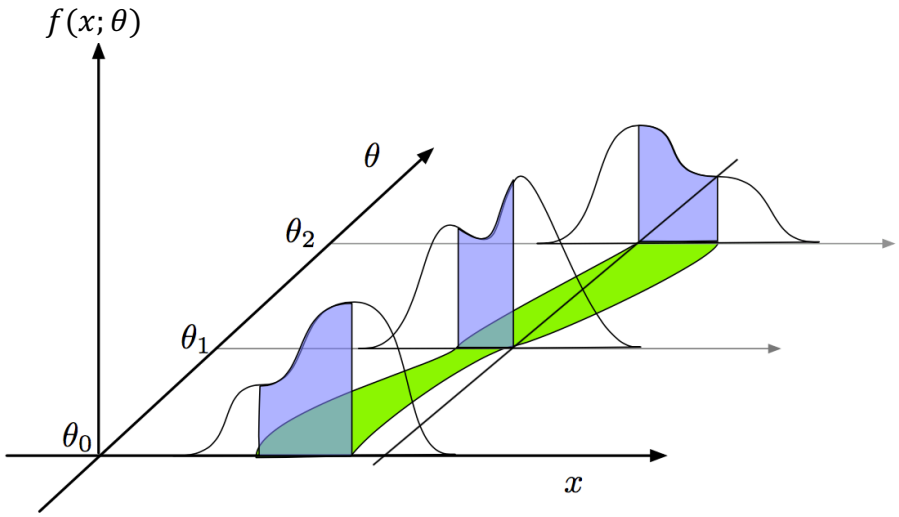

FIGURE 7.3: Confidence belt construction for a pdf $f(x;\theta)$. The belt (green area) is defined as a consequence of the ordering rule that selects the desired confidence level $1-\alpha$ (violet regions).

$[\theta_1(x_0), \theta_2(x_0)]$ is obtained which has, by construction, a coverage equal to the desired confidence level, that is $1-\alpha$. This means that, if θ is equal to the true value θ^{true} and $x = x_0$ is extracted randomly according to the pdf $f(x;\theta^{\text{true}})$, θ^{true} will be included in the determined confidence interval $[\theta_1(x_0), \theta_2(x_0)]$ in a fraction $1-\alpha$ of the cases (in the limit of a large number of extractions). Neyman's belt inversion is graphically illustrated in Fig. 7.4.

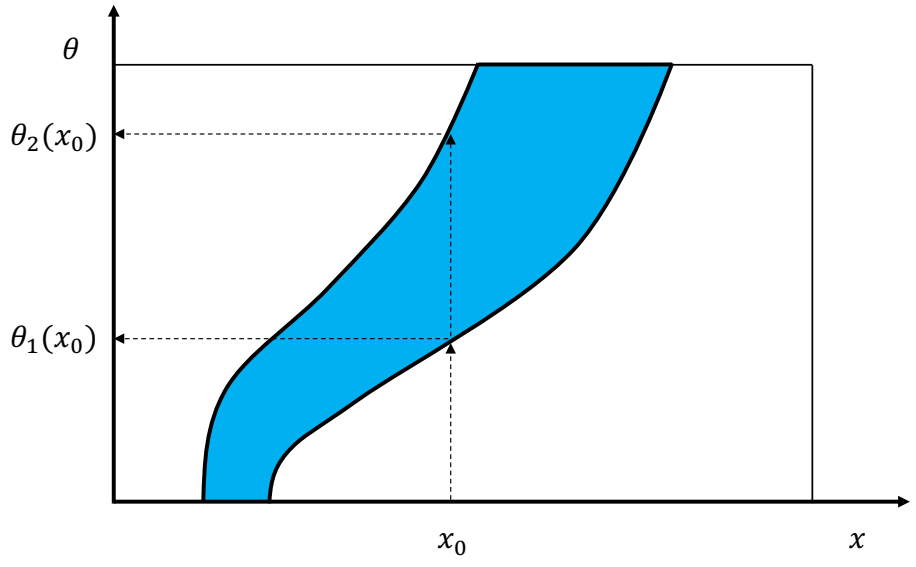

FIGURE 7.4: Graphical illustration of the Neyman belt inversion for a pdf $f(x;\theta)$.

7.4 Upper limits

The procedure of setting an upper limit on an unknown parameter is, in the frequentist approach, a special case of confidence interval determination. In this case, the

central interval for the parameter is replaced by a fully asymmetric interval $[0, \theta^{\mathrm{up}})$, corresponding to the desired confidence level $1 - \alpha$, usually 95%. It must be underlined that, for the purpose of signal exclusion, the requirements applied in terms of p-value are milder than for the claim of a signal discovery. In fact, a 95% confidence level corresponds to $p < 0.05$, where p indicates the probability of a signal underfluctuation (namely, null and alternative hypotheses are inverted with respect to the case of a discovery).

The computation of a frequentist upper limit is achieved by inverting the Neyman belt for a parameter θ, with asymmetric intervals for the observable x (in similarity to what was shown in Fig. 7.4 for symmetric intervals). Assuming that the belt is monotonically increasing with x, the choice of the interval $(x^{\mathrm{low}}(\theta_0), +\infty)$ for x as a function of θ_0 leads to a confidence interval $[0, \theta^{\mathrm{up}}(x_0))$ for θ (see Fig. 7.5), corresponding to the upper limit:

$$\theta < \theta^{\mathrm{up}}(x_0) \tag{7.4}$$

at the desired confidence level.

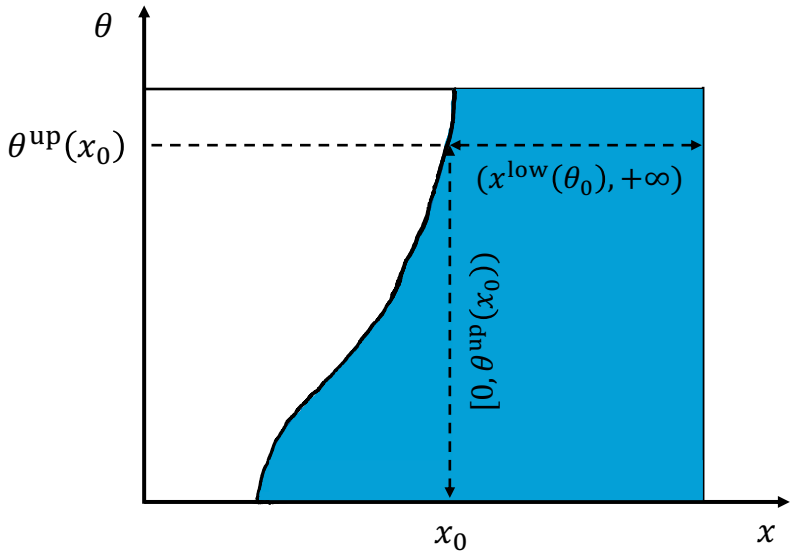

FIGURE 7.5: Graphical illustration of the Neyman belt construction in the case of upper limit determination.

The use of frequentist upper limits may raise some concerns, though. First of all, if the data exhibit large underfluctuations, they could exclude certain regions in the space of the parameter of interest where the experiment should be insensitive, or that are even unphysical. Moreover, for some of the upper limit calculation methods, such as the Feldman–Cousins' [99], if no signal event is observed, the limit depends on the expected amount of background, and tends to be more stringent for larger values of the latter. This is somehow counterintuitive, since it means that a better result can be obtained with an experiment with a lower background rejection performance. These concerns were addressed by defining a novel approach, which will be discussed in the next section.

7.4.1 The CL_s modified frequentist approach

The CL_s method [100, 101] derives from a modification of the pure frequentist approach. Its novelty stands in the introduction of a conservative corrective factor to the p-value that aims at curing the aforementioned peculiarities arising from the pure frequentist methods. The CL_s requires a test statistic. In the work described in this thesis, I consider the profile likelihood ratio:

$$\lambda(\mu, \vec{\theta}) = \frac{\mathcal{L}(\mathbf{x}; \mu, \hat{\vec{\theta}}_\mu)}{\mathcal{L}(\mathbf{x}; \hat{\mu}, \hat{\vec{\theta}})}, \tag{7.5}$$

where $\hat{\mu}$ and $\hat{\vec{\theta}}$ are the best fit values of the parameter of interest μ and the nuisance parameters $\vec{\theta}$ in the observed data sample, whereas $\hat{\vec{\theta}}_\mu$ is the best fit value of $\vec{\theta}$ obtained for fixed μ. The test statistic for the CL_s is then defined using the profile likelihood ratio evaluated under two different hypotheses: H_1, or the presence of both signal and background (s+b), and H_0, corresponding to the presence of background only (b):

$$\lambda(\mu, \vec{\theta}) = \frac{\mathcal{L}_{\mathrm{s+b}}(\mathbf{x}; \mu, \hat{\vec{\theta}}_\mu)}{\mathcal{L}_{\mathrm{b}}(\mathbf{x}; \hat{\mu}, \hat{\vec{\theta}})}, \tag{7.6}$$

Following a frequentist approach, in order to quote an upper limit the distribution of the test statistic $\lambda(\mu, \vec{\theta})$ must be known in the hypothesis of signal plus background, and the p-value corresponding to the observed value $\lambda = \hat{\lambda}$, denoted below as $\mathrm{p_{s+b}}$, must be determined as a function of the parameter of interest μ and nuisance parameters $\vec{\theta}$.

The proposed modification to the purely frequentist approach consists in finding *two* p-values corresponding to both the H_1 and H_0 hypotheses:

$$\begin{aligned} \mathrm{p_{s+b}}(\mu, \vec{\theta}) &= P_{\mathrm{s+b}}(\lambda(\mu, \vec{\theta}) \geq \hat{\lambda}), \\ \mathrm{p_b}(\mu, \vec{\theta}) &= P_{\mathrm{b}}(\lambda(\mu, \vec{\theta}) \leq \hat{\lambda}). \end{aligned} \tag{7.7}$$

The following quantity can be obtained from these two probabilities:

$$\mathrm{CL_s}(\mu, \vec{\theta}) = \frac{\mathrm{p_{s+b}}(\mu, \vec{\theta})}{1 - \mathrm{p_b}(\mu, \vec{\theta})}. \tag{7.8}$$

Upper limits can then be determined excluding the range of the parameter of interest for which $\mathrm{CL_s}(\mu, \vec{\theta})$ is lower than the desired confidence level (e.g., 95%). In most cases, it is not trivial to obtain the probabilities $P_{\mathrm{s+b}}$ and P_{b} in Eq. (7.7) analytically, and thus they are determined numerically using pseudo-experiments or with asymptotic approximations [102].

The intervals obtained using the CL_s method are "conservative", from a frequentist point of view, since $p_{\mathrm{b}} \leq 1$ and therefore $\mathrm{CL_s}(\mu, \vec{\theta}) \geq \mathrm{p_{s+b}}$. This results in less stringent limits than the purely frequentist ones, although without the downsides discussed at the end of the previous section.

7.5 Upper limit on $\mathcal{B}(\mathrm{W} \to \pi\gamma)$

The $\mathrm{CL_s}$ method described in the previous sections of this chapter is used to set an upper limit at 95% CL on the branching fraction of the rare decay $\mathrm{W}^{\pm} \to \pi^{\pm}\gamma$:

$$\mathcal{B}(\mathrm{W}^{\pm} \to \pi^{\pm}\gamma) < 1.50 \times 10^{-5}. \tag{7.9}$$

The expected upper limit corresponds to $0.85^{+0.52}_{-0.29} \times 10^{-5}$. The latter is obtained using pseudo-experiments, and is useful to understand the actual sensitivity of the analysis. In this case, the difference with respect to the observed upper limit is due to the statistical fluctuation of the background in the signal-sensitive window of the $m_{\pi\gamma}$ distribution.

A graphical representation of the scan of the parameter of interest $\mathcal{B}(\mathrm{W}^{\pm} \to \pi^{\pm}\gamma)$ and the corresponding expected and observed p-values ($\mathrm{CL_s}$) is shown in Fig. 7.6. For what regards the observed $\mathrm{CL_s}$, a drastic change in the slope of the curve appears

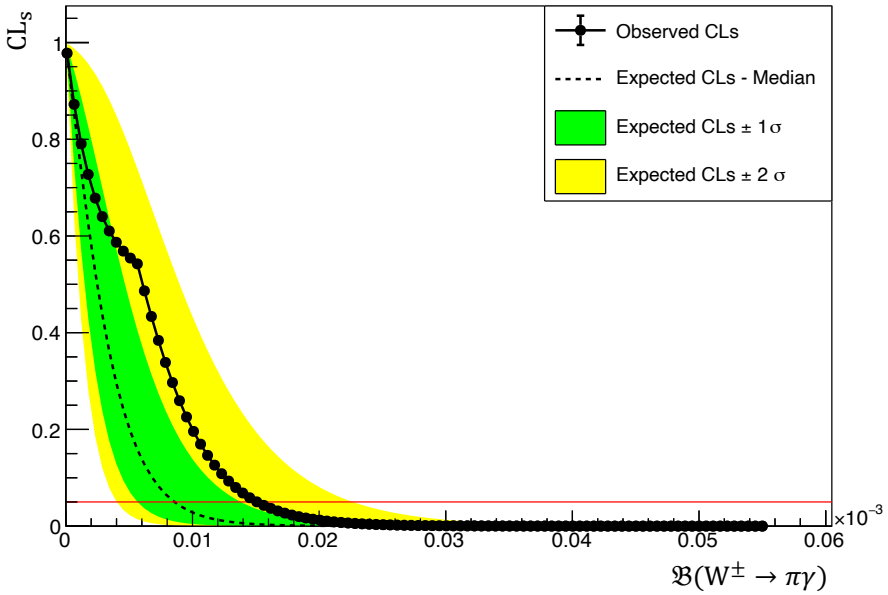

FIGURE 7.6: $\mathrm{CL_s}$ values for a scan of the parameter of interest $\mathcal{B}(\mathrm{W}^{\pm} \to \pi^{\pm}\gamma)$. The horizontal red line corresponds to a 0.05 p-value.

around the central value of the measured branching fraction, reported in Eq. (7.1). Such a behavior is consistent with the observation of a small excess around the value of the W boson mass in the $m_{\pi\gamma}$ distribution.

The total uncertainty in this final result is dominated by the statistical contributions, which account for more than 80%. The largest systematic uncertainties arise from the measurements of the $\mathrm{t\bar{t}}$ cross section (≈5%) and the integrated luminosity (≈2%).

7.6 Perspectives

The result in Eq. (7.9) can be compared with the state of art measurement of this rare decay branching ratio, set by the CDF Collaboration at Tevatron [33]:

$$\mathcal{B}(\mathrm{W}^{\pm} \to \pi^{\pm}\gamma) < 7.0 \times 10^{-6}, \quad \mathrm{CL} = 95\%. \tag{7.10}$$

That search was based on events collected in proton-antiproton collisions at $\sqrt{s} = 1.96\,\text{TeV}$, using an inclusive photon trigger. Unfortunately, the same approach is unsuitable at CMS, where the trigger thresholds for single photons are too high (>170 GeV). Therefore, the search I presented in this thesis exploits a novel strategy, focusing on top quark-antiquark pair events where one of the two W bosons decays into leptons. If, on the one hand, this strategy provides a relatively clear signature for the rare decay process, on the other the signal cross section is significantly reduced by including the $\mathrm{t\bar{t}}$ cross section and the branching fraction of $\mathrm{W} \to \ell\nu$. In particular, at CMS:

$$\sigma_{\mathrm{t\bar{t}}}\,(\mathcal{B} \to \ell\nu) \simeq 815\,\text{pb} \times (2 + 0.3521)\,0.1086 \simeq 208\,\text{pb}, \tag{7.11}$$

and therefore around 30 million events of $\mathrm{t\bar{t}}$ with consequent leptonic decay of one of the W bosons are expected to be produced for a $\mathcal{L}_{\text{int}} = 137\,\text{fb}^{-1}$. At CDF, on the other hand, the inclusive W boson production cross section in proton-antiproton collisions at $\sqrt{s} = 1.96\,\text{TeV}$ was about 25 nb [103]. Therefore, their search for the decay $\mathrm{W} \to \pi\gamma$, which exploited a $\mathcal{L}_{\text{int}} = 4.3\,\text{fb}^{-1}$, could rely on more than 100 million initial events containing a W boson.

Nevertheless, the result I obtained demonstrates the feasibility of a search for rare hadronic decays of the W boson at the LHC using a new search strategy. With the increased amount of data that will be available in the next phases of the LHC operation, it will be possible to enhance the sensitivity of this analysis to the rare decay $\mathrm{W} \to \pi\gamma$. In particular, by the end of Run 3, currently estimated for 2024, CMS is expected to collect approximately the same amount of integrated luminosity as in the period 2016–2018. In the realistic perspective of very similar data-taking conditions with respect to Run 2, and thus assuming a control over the systematic uncertainties comparable to what obtained in this case, it is plausible to foresee an improvement on the measurement of $\mathcal{B}(\mathrm{W}^{\pm} \to \pi^{\pm}\gamma)$ of a factor $\approx\sqrt{2}$.

In addition, improvements on some of the techniques and strategies adopted in this analysis might be possible. For instance, the use of data-driven methods for a more accurate estimate of some of the background contributions (e.g., QCD multijet processes) could lead to an enhanced agreement between collision data and simulation in some key variables, such as the pion isolation. This might increase the performance of the BDT or any multivariate selection procedure that makes use of these variables. Another viable option would consist in measuring the ratio $\mathcal{B}(\mathrm{W} \to \pi\gamma)/\mathcal{B}(\mathrm{W} \to \ell\nu)$, thus obtaining the cancellation of the (already small) systematic uncertainties related to the normalization, such as the uncertainty in the integrated luminosity.

These considerations provide very encouraging perspectives regarding this study at the LHC and at future hadron colliders.

Conclusions

The work presented in this thesis describes the first search at the LHC for the rare decay of the W boson into a pion and a photon, using the data collected in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the CMS experiment. The full integrated luminosity recorded by CMS in 2016 to 2018 is exploited, corresponding to $137\,\text{fb}^{-1}$. Because of the high trigger thresholds for single photons, which make an inclusive search unsuitable at CMS, a novel measurement technique is used, which focuses on top quark-antiquark pair events, where one of the W bosons originating from the top quark decays into leptons. The charged lepton, either a muon or an electron, is used to select the events of interest at trigger level. The presence of b quark jets helps reducing the background from the hadronization of light-flavor quarks and gluons. The W boson originating from the other top quark is then used to search for the rare decay $\text{W} \to \pi\gamma$, characterized by an isolated track, for which a specific pion-isolation variable is designed, and an isolated photon with large transverse momentum.

The analysis relies on Monte Carlo simulated events for the tuning of the event selection and for the estimation of the signal lineshape. In particular, I have produced a simulation of the signal process using the PYTHIA 8 generator, whereas the several background sources accounted for were produced centrally by CMS MC experts. All the simulated events are weighted to match the integrated luminosity of collision data.

The analysis selection criteria were designed before inspecting the signal region in the collision data, so to avoid any observational bias. First of all, a cut-based selection is applied, exploiting the basic features of the signal topology. Afterwards, several variables with good signal-to-background discriminating power are used to train a multivariate classifier (boosted decision tree) with simulated events. Upon the output of this classifier, a signal- and a background-enriched regions of the reconstructed pion-photon invariant mass are defined. The background-enriched region is employed to determine the functional form of the background directly from data, whereas the functional form of the signal is derived from the MC simulation. The signal-enriched region in collision data is then fitted to extract the branching ratio of $\text{W} \to \pi\gamma$.

No significant excess is observed above the expected background. I set an upper limit $\mathcal{B}(\text{W}^{\pm} \to \pi^{\pm}\gamma) < 1.50 \times 10^{-5}$ at 95% confidence level. This is not the most stringent upper limit on this rare decay branching ratio, but the result obtained demonstrates the feasibility of a search for rare hadronic decays of the W boson at the LHC using a new strategy. With the increased amount of data that will be available in the next phases of the LHC operation, it will be possible to enhance the sensitivity of this search. Moreover, there might be room for improvement in some of the techniques adopted in this analysis. For instance, the use of data-driven methods for a more accurate estimate of some of the background contributions enhance the agreement between collision data and simulation in some key variables, and possibly increase the performance of the BDT or any multivariate selection procedure that makes use of these variables.

With these considerations in mind, the perspectives regarding the study of this decay and, in general, of rare decays of massive bosons at the LHC and at future hadron colliders are very encouraging.

Appendix A

The Tag and Probe method

In high energy physics experiments, the Tag and Probe is a method that utilizes a known mass resonance (e.g., J/ψ, Υ, Z boson) to select particles of the desired type, and to measure the efficiency of a particular selection criterion on those particles.

The *tag* is generally an object that passes a number of very tight identification criteria, specifically designed to isolate the required particle - typically an electron or muon, though in principle the method is not strictly limited to these. Therefore, the fake rate for passing the tag selection cutoffs must be very small.

A set of the desired particle type, known as *probes*, is selected by pairing these particles with tags so that the invariant mass of the combination is consistent with the mass of the resonance. Depending on whether they pass or not the specific selection criterion whose efficiency needs to be determined, the probes are then split into two subsamples: the *passing probes* and the *failing probes*. The efficiency may then be measured as:

$$\varepsilon = \frac{P_{\text{pass}}}{P_{\text{pass}} + P_{\text{fail}}}, \tag{A.1}$$

where P_{pass} and P_{fail} are the number of passing and failing probes, respectively, estimated by fitting the resonance peak. In order to perform the count, combinatoric backgrounds must be accounted for and eliminated. Therefore, it is essential to possess a good knowledge of the background contributions and their shape in the observed distribution.

A.1 Tag and Probe for electron scale factors

The Tag and Probe method is used to derive the electron trigger efficiency scale factors presented in Section 4.2.10. Electrons are chosen as tags if they have $p_{\text{T}} > 35\,\text{GeV}$ and $|\eta| < 2.1$. Such electron must fulfill an MVA-based identification criterion similar to that described in Section 2.3.4, but with a working point corresponding to a 80% signal efficiency of the MVA, averaged on the different parts of the detector.

The list of probes comprises electrons whose pairing with the tags generates an invariant mass compatible with that of a Z boson. Passing probes must trigger the year-dependent electron HLT paths presented in Table 4.1, whose efficiency has to be measured. In addition, they must fulfill the preselection criteria:

- $p_{\text{T}} > 20\,\text{GeV}$;
- $|\eta| < 2.4\,\text{GeV}$;
- $|d_{\text{z}}| < 0.5\,\text{cm}$;
- $|d_{\text{xy}}| < 0.2\,\text{cm}$,

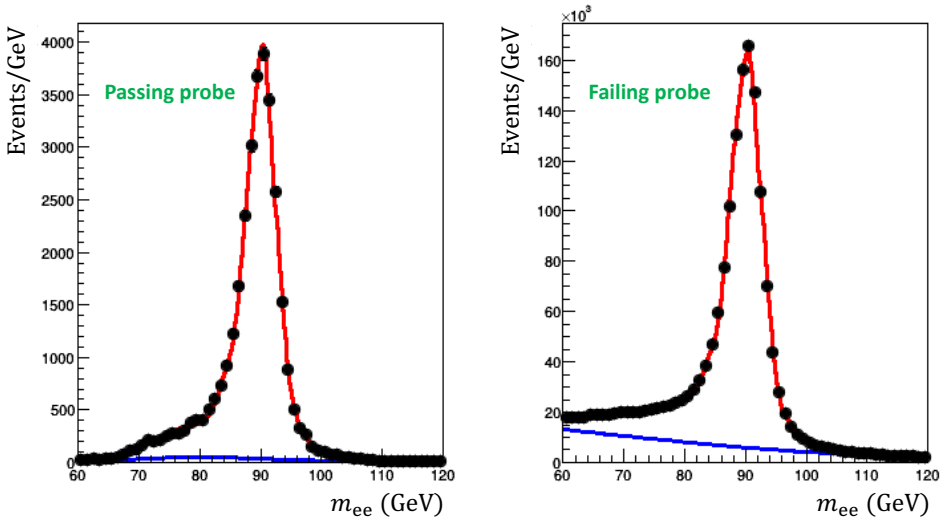

FIGURE A.1: Fit to the m_{ee} distribution for passing and failing probes.

as discussed in Section 4.2.2.

Distributions of e^+e^- invariant mass (m_{ee}) are reconstructed for passing and failing probes, separately for several bins of p_T and η of the probes themselves. The number of passing and failing probes is determined by fitting the respective m_{ee} distributions using a double crystal ball pdf (introduced in Section 5.2) to describe the signal lineshape, and a CMSShape pdf to parametrize the background. The latter pdf is defined as:

$$\text{CMSShape}(m_{ee}; x_0, \alpha, \beta, \gamma) = \begin{cases} \text{erf}\left(\beta(\alpha - m_{ee})\right) \times 10^{20} & \text{for } u < -70 \\ \text{erf}\left(\beta(\alpha - m_{ee})\right) e^{-u} & \text{for } -70 \leq u \leq 70 \\ 0 & \text{for } u > 70, \end{cases} \quad \text{(A.2)}$$

where $u = \gamma(m_{ee} - x_0)$. Examples of the performed fits are shown in Fig. A.1, for the passing and failing probes distributions of m_{ee}. The efficiency is then calculated as in Eq. (A.1).

This procedure is iterated for collision data and for MC samples that simulate Drell–Yan processes. The scale factors can thus be calculated as the ratio between the bin-by-bin efficiency measured in data and in simulated events. Finally, systematic uncertainties in the scale factors are estimated by comparing the results obtained using MC samples produced with the MADGRAPH5_*a*MC@NLO generator or the POWHEG framework.

Bibliography

[1] James Chadwick, "The existence of a neutron", *Proc. R. Soc. A* **136** (1932) 692. DOI: 10.1098/rspa.1932.0112.

[2] Enrico Fermi, "Tentativo di una teoria dei raggi β", *Nuovo Cim.* **11** (1934) 1. DOI: 10.1007/BF02959820.

[3] Bruno Pontecorvo, "Nuclear capture of mesons and the meson decay", *Phys. Rev.* **72** (1947) 246. DOI: 10.1103/PhysRev.72.246.

[4] Jack Steinberger, "On the Range of the Electrons in Meson Decay", *Phys. Rev.* **75** (1949) 1136. DOI: 10.1103/PhysRev.75.1136.

[5] C. S. Wu et al., "Experimental test of parity conservation in beta decay", *Phys. Rev.* **105** (1957) 1413. DOI: 10.1103/PhysRev.105.1413.

[6] T. D. Lee and C. N. Yang, "Parity nonconservation and a two-component theory of the neutrino", *Phys. Rev.* **105** (1957) 1671. DOI: 10.1103/PhysRev.105.1671.

[7] Sheldon L. Glashow, "Partial-symmetries of weak interactions", *Nucl. Phys.* **22** (1961) 579. DOI: 10.1016/0029-5582(61)90469-2.

[8] Steven Weinberg, "A Model of Leptons", *Phys. Rev. Lett.* **19** (1967) 1264. DOI: 10.1103/PhysRevLett.19.1264.

[9] Abdus Salam, "Weak and Electromagnetic Interactions", *Conf. Proc. C* **680519** (1968) 367. DOI: 10.1142/9789812795915_0034.

[10] François Englert and Robert Brout, "Broken symmetry and the mass of gauge vector mesons", *Phys. Rev. Lett.* **13** (1964) 321. DOI: 10.1103/PhysRevLett.13.321.

[11] Peter W. Higgs, "Broken symmetries and the masses of gauge bosons", *Phys. Rev. Lett.* **13** (1964) 508. DOI: 10.1103/PhysRevLett.13.508.

[12] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Physics Letters B* **716** (2012) 1. DOI: https://doi.org/10.1016/j.physletb.2012.08.020.

[13] CMS Collaboration, "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV", *JHEP* **06** (2013) 081. DOI: 10.1007/JHEP06(2013)081. arXiv: 1303.4571.

[14] Nicola Cabibbo, "Unitary Symmetry and Leptonic Decays", *Phys. Rev. Lett.* **10** (1963) 531. DOI: 10.1103/PhysRevLett.10.531.

[15] Makoto Kobayashi and Toshihide Maskawa, "CP violation in the renormalizable theory of weak interaction", *Prog. Theor. Phys.* **49** (1973) 652. DOI: 10.1143/PTP.49.652.

[16] S. L. Glashow, J. Iliopoulos, and L. Maiani, "Weak interactions with lepton-hadron symmetry", *Phys. Rev. D* **2** (1970) 1285. DOI: 10.1103/PhysRevD.2.1285.

[17] Particle Data Group, P. A. Zyla, et al., "Review of Particle Physics", *Prog. Theor. Exp. Phys.* **2020** (2020) 083C01. DOI: 10.1093/ptep/ptaa104.

[18] David J. Gross and Frank Wilczek, "Ultraviolet Behavior of Non-Abelian Gauge Theories", *Phys. Rev. Lett.* **30** (1973) 1343. DOI: 10.1103/PhysRevLett.30.1343.

[19] H. David Politzer, "Reliable perturbative results for strong interactions?", *Phys. Rev. Lett.* **30** (1973) 1346. DOI: 10.1103/PhysRevLett.30.1346.

[20] Simon van der Meer, *Stochastic damping of betatron oscillations in the ISR*. Tech. rep. CERN-ISR-PO-72-31. 1972. URL: https://cds.cern.ch/record/312939.

[21] UA1 Collaboration, "Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV", *Phys. Lett. B* **122** (1983) 103. DOI: 10.1016/0370-2693(83)91177-2.

[22] UA2 Collaboration, "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN $\bar{\mathrm{p}}\mathrm{p}$ collider", *Phys. Lett. B* **122** (1983) 476. DOI: 10.1016/0370-2693(83)91605-2.

[23] UA1 Collaboration, "Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider", *Phys. Lett. B* **126** (1983) 398. DOI: 10.1016/0370-2693(83)90188-0.

[24] UA2 Collaboration, "Evidence for $\mathrm{Z}^0 \to \mathrm{e}^+\mathrm{e}^-$ at the CERN $\bar{\mathrm{p}}\mathrm{p}$ collider", *Phys. Lett. B* **129** (1983) 130. DOI: 10.1016/0370-2693(83)90744-X.

[25] CDF Collaboration, "Precise measurement of the W-boson mass with the CDF II detector", *Phys. Rev. Lett.* **108** (2012) 151803. DOI: 10.1103/PhysRevLett.108.151803.

[26] ATLAS Collaboration, "Measurement of the W-boson mass in pp collisions at $\sqrt{s} = 7\,\mathrm{TeV}$ with the ATLAS detector", *EPJC* **78** (2018) 110. DOI: 10.1140/epjc/s10052-017-5475-4.

[27] Matteo Cacciari et al., "Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation", *Physics Letters B* **710** (2012) 612. DOI: 10.1016/j.physletb.2012.03.013.

[28] Michal Czakon and Alexander Mitov, "Top++: a program for the calculation of the top-pair cross-section at hadron colliders", *Comput. Phys. Commun.* **185** (2014) 2930. DOI: 10.1016/j.cpc.2014.06.021. arXiv: 1112.5675.

[29] CMS Collaboration, "Measurement of the $\mathrm{t\bar{t}}$ production cross section, the top quark mass, and the strong coupling constant using dilepton events in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$", *Eur. Phys. J. C* **79** (2019) 368. DOI: 10.1140/epjc/s10052-019-6863-8.

[30] CMS Collaboration, "Measurement of inclusive W and Z boson production cross sections in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$", CMS-PAS-SMP-15-004 (2015). URL: https://cds.cern.ch/record/2093537.

[31] CDF Collaboration, "First search for exotic Z boson decays into photons and neutral pions in hadron collisions", *Phys. Rev. Lett.* **112** (2014) 111803.

[32] ATLAS Collaboration, "Search for the lepton flavor violating decay $\mathrm{Z} \to \mathrm{e}\mu$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", *Phys. Rev. D* **90** (2014) 072010. DOI: 10.1103/PhysRevD.90.072010.

[33] CDF Collaboration, "Search for the rare radiative decay: $W \to \pi\gamma$ in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV", *Phys. Rev. D* **85** (2012) 032001. DOI: 10.1103/PhysRevD.85.032001. arXiv: 1104.1585.

[34] CDF Collaboration, "Search for the rare decay $W^{\pm} \to D_s^{\pm}\gamma$ in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV", *Phys. Rev. D* **122** (1998) 091101. DOI: 10.1103/PhysRevD.58.091101.

[35] CMS Collaboration, "Search for W boson decays to three charged pions", *Phys. Rev. Lett.* **122** (2019) 151802. DOI: 10.1103/PhysRevLett.122.151802. arXiv: 1901.11201.

[36] Michelangelo Mangano and Tom Melia, "Rare exclusive hadronic W decays in a $t\bar{t}$ environment", *Eur. Phys. J. C* **75** (2015) 258. DOI: 10.1140/epjc/s10052-015-3482-x. arXiv: 1410.7475 [hep-ph].

[37] Aneesh V. Manohar, "The decays $Z \to W\pi$, $Z \to \gamma\pi$, and the pion form factor", *Phys. Lett. B* **244** (1990) 101. DOI: 10.1016/0370-2693(90)90276-C.

[38] Lampros Arnellos, William J. Marciano, and Zohreh Parsa, "Radiative decays: $W^{\pm} \to P^{\pm} + \gamma$ and $Z \to P^0 + \gamma$", *Nucl. Phys. B* **196** (1982) 378. DOI: 10.1016/0550-3213(82)90496-5.

[39] Y. Y. Kneum and X. Y. Pham, "Possible huge enhancement in the radiative decay of the weak W boson into a pion or a charmed D_s meson", *Mod. Phys. Lett. A* **09** (1994) 1545. DOI: 10.1142/S0217732394001386.

[40] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands, "A brief introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852. DOI: 10.1016/j.cpc.2008.01.036.

[41] G. Peter Lepage and Stanley J. Brodsky, "Exclusive processes in quantum chromodynamics: Evolution equations for hadronic wavefunctions and the form factors of mesons", *Phys. Lett. B* **87** (1979) 359. DOI: 10.1016/0370-2693(79)90554-9.

[42] G. Peter Lepage and Stanley J. Brodsky, "Exclusive processes in perturbative quantum chromodynamics", *Phys. Rev. D* **22** (1980) 2157. DOI: 10.1103/PhysRevD.22.2157.

[43] Gino Isidori, Aneesh V. Manohar, and Michael Trott, "Probing the nature of the Higgs-like boson via $h \to V\mathcal{F}$ processes", *Phys. Lett. B* **728** (2014) 131. DOI: 10.1016/j.physletb.2013.11.054.

[44] Geoffrey Bodwin et al., "Higgs boson decays to quarkonia and the $H\bar{c}c$ coupling", *Phys. Rev. D* **88** (2013) 053003. DOI: 10.1103/PhysRevD.88.053003.

[45] Alexander L. Kagan et al., "Exclusive Window onto Higgs Yukawa Couplings", *Phys. Rev. Lett.* **114** (2015) 101802. DOI: 10.1103/PhysRevLett.114.101802.

[46] Dao-Neng Gao, "A note on Higgs decay into Z boson and $J\Psi(\Upsilon)$", *Phys. Lett. B* **737** (2014) 366. DOI: 10.1016/j.physletb.2014.09.019.

[47] Bhubanjyoti Bhattacharya, Alakabha Datta, and David London, "Probing new physics in Higgs couplings to fermions using an angular analysis", *Phys. Lett. B* **736** (2014) 421. DOI: 10.1016/j.physletb.2014.07.065.

[48] Wai-Yee Keung, "Decay of the Higgs boson into heavy-quarkonium states", *Phys. Rev. D* **27** (1983) 2762. DOI: 10.1103/PhysRevD.27.2762.

[49] "LHC Design Report Vol.1: The LHC Main Ring", (2004). Ed. by Oliver S. Bruning et al. DOI: 10.5170/CERN-2004-003-V-1.

[50] "LHC Design Report Vol. 3. The LHC Injector Chain", (2004). Ed. by M. Benedikt et al. DOI: 10.5170/CERN-2004-003-V-3.

[51] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[52] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[53] ALICE Collaboration, "The ALICE Experiment at the CERN LHC", *JINST* **3** (2008) S08002. DOI: 10.1088/1748-0221/3/08/S08002.

[54] CMS Collaboration, "The CMS Experiment at the CERN LHC", *JINST* **3** (2008) S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[55] CMS Collaboration, "CMS, the magnet project: Technical design report", (1997).

[56] CMS Collaboration, "The CMS tracker system project: Technical Design Report", (1997). Ed. by M. Mannelli et al.

[57] "CMS Technical Design Report for the Pixel Detector Upgrade", (2012). Ed. by David Aaron Matzner Dominguez et al. DOI: 10.2172/1151650.

[58] "The CMS electromagnetic calorimeter project: Technical Design Report", (1997).

[59] Q. Ingram et al., "Energy resolution of the barrel of the CMS Electromagnetic Calorimeter", *JINST* **2** (2007) P04004. DOI: 10.1088/1748-0221/2/04/p04004.

[60] "The CMS hadron calorimeter project: Technical Design Report", (1997).

[61] "CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter", (2012). Ed. by J. Mans et al. DOI: 10.2172/1151651.

[62] "The CMS muon project: Technical Design Report", (1997).

[63] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *JINST* **13** (2018) P06015. DOI: 10.1088/1748-0221/13/06/P06015.

[64] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017) P01020. DOI: 10.1088/1748-0221/12/01/p01020.

[65] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017) P10003. DOI: 10.1088/1748-0221/12/10/p10003. arXiv: 1706.04965.

[66] The CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014) P10009. DOI: 10.1088/1748-0221/9/10/p10009.

[67] W Adam et al., "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC", *J. Phys. G* **31**.9 (2005) N9. DOI: 10.1088/0954-3899/31/9/n01.

[68] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$", *JINST* **10** (2015) P06005. DOI: 10.1088/1748-0221/10/06/P06005. arXiv: 1502.02701.

[69] CMS Collaboration, "Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$", *JINST* **10** (2015) P08010. DOI: 10.1088/1748-0221/10/08/P08010. arXiv: 1502.02702 [physics.ins-det].

[70] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez, "The anti-k_T jet clustering algorithm", **04** (2008) 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189.

[71] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez, "FastJet user manual", *Eur. Phys. J. C* **72** (2012) 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[72] CMS Collaboration, "Pileup mitigation at CMS in 13 TeV data", *JINST* **15** (2020) P09018. DOI: 10.1088/1748-0221/15/09/p09018. arXiv: 2003.00503.

[73] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017) P02014. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex].

[74] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *JINST* **13** (2018) P05011. DOI: 10.1088/1748-0221/13/05/p05011. arXiv: 1712.07158.

[75] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector", *JINST* **14** (2019) P07004. DOI: 10.1088/1748-0221/14/07/P07004. arXiv: 1903.06078 [hep-ex].

[76] Paul Lujan, "Performance of the Pixel Luminosity Telescope for luminosity measurement at CMS during Run 2", *PoS* **314** (2017) 504. DOI: 10.22323/1.314.0504.

[77] Maria Hempel, "Development of a novel diamond based detector for machine induced background and luminosity measurements". PhD thesis. DESY, Hamburg, 2017. DOI: 10.3204/PUBDB-2017-06875.

[78] GEANT4 Collaboration, "GEANT4—a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250. DOI: 10.1016/S0168-9002(03)01368-8.

[79] Torbjörn Sjöstrand et al., "An Introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[80] CMS Collaboration, "Measurement of the $\mathrm{t\bar{t}}$ production cross section using events in the eμ final state in pp collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **77** (2017) 172. DOI: 10.1140/epjc/s10052-017-4718-8. arXiv: 1611.04040.

[81] Paolo Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2014) 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146.

[82] Stefano Frixione, Paolo Nason, and Carlo Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method", *JHEP* **11** (2017) 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092.

[83] Simone Alioli et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581.

[84] John M. Campbell et al., "Top-pair production and decay at NLO matched with parton showers", *JHEP* **04** (2015) 114. DOI: 10.1007/JHEP04(2015)114. arXiv: 1412.1828.

[85] Johan Alwall 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** (2014) 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301.

[86] NNPDF, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[87] NNPDF, "Parton distributions from high-precision collider data", *Eur. Phys. J. C* **77** (2017) 663. DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 [hep-ph].

[88] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016) 155. DOI: 10.1140/epjc/s10052-016-3988-x. arXiv: 1512.00815.

[89] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements", *Eur. Phys. J. C* **80** (2020) 4. DOI: 10.1140/epjc/s10052-019-7499-4. arXiv: 1903.12179.

[90] Robert E. Schapire and Yoram Singer, "Improved boosting algorithms using confidence-rated predictions", *Machine Learning* **37** (1999) 297. DOI: 10.1023/A:1007614523901.

[91] Andreas Hoecker et al., "TMVA: toolkit for multivariate data analysis", *PoS* **ACAT** (2007) 040. arXiv: physics/0703039.

[92] Mark Joseph Oreglia, "A study of the reactions $\psi' \to \gamma\gamma\psi$". SLAC Report SLAC-R-236. PhD thesis. Stanford University, 1980. URL: http://www.slac.stanford.edu/cgi-wrap/getdoc/slac-r-236.pdf.

[93] Fred James et al., "MINUIT: function minimization and error analysis reference manual", (1998). CERN Program Library Long Writeups. URL: https://cds.cern.ch/record/2296388.

[94] CMS Collaboration, "Measurement of the $\mathrm{t\bar{t}}$ production cross sectionm the top quark mass, and the strong coupling constant using dilepton events in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$", *Eur. Phys. J. C* **79** (2019) 368. DOI: 10.1140/epjc/s10052-019-6863-8. arXiv: 1812.10505.

[95] CMS, "CMS Luminosity Measurement for the 2016 data-taking period", CMS-PAS-LUM-17-001 (2017). URL: https://cds.cern.ch/record/2138682.

[96] CMS, "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13\,\mathrm{TeV}$", CMS-PAS-LUM-17-004 (2017). URL: https://cds.cern.ch/record/2621960.

[97] CMS, "CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13\,\mathrm{TeV}$", CMS-PAS-LUM-18-002 (2018). URL: https://cds.cern.ch/record/2676164.

[98] J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability", *Phil. Trans. Roy. Soc. Lond.* **A236** (1937) 333. DOI: 10.1098/rsta.1937.0005.

[99] Gary J. Feldman and Robert D. Cousins, "A Unified approach to the classical statistical analysis of small signals", *Phys. Rev. D* **57** (1998) 3873. DOI: 10.1103/PhysRevD.57.3873. arXiv: physics/9711021.

[100] Thomas Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth. A* **434** (1999) 435. DOI: 10.1016/S0168-9002(99)00498-2. arXiv: hep-ex/9902006 [hep-ex].

[101] A. L. Read, "Presentation of search results: the CL_s technique", *J. Phys. G* **28** (2002) 2693. DOI: 10.1088/0954-3899/28/10/313.

[102] G. Cowan et al., "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727.

[103] CDF Collaboration, "First measurements of inclusive W and Z cross sections from Run II of the Fermilab Tevatron collider", *Phys. Rev. Lett.* **94** (2005) 091803. DOI: 10.1103/PhysRevLett.94.091803. arXiv: hep-ex/0406078.