Doctoral Thesis

Measurement of differential cross sections for diphoton production in pp collisions with the CMS experiment

Author(s):
Peruzzi, Marco

Publication Date:
2015

Permanent Link:
https://doi.org/10.3929/ethz-a-010558087

Rights / License:
In Copyright - Non-Commercial Use Permitted
MEASUREMENT OF DIFFERENTIAL CROSS SECTIONS FOR DIPHOTON PRODUCTION IN PP COLLISIONS WITH THE CMS EXPERIMENT

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

Marco Peruzzi

Laurea Magistrale in Fisica, Università di Pisa

born on 17.09.1987

citizen of Italy

accepted on the recommendation of

Prof. Günther Dissertori, examiner

Prof. Rainer Wallny, co-examiner

Prof. Luigi Rolandi, co-examiner

2015
Abstract

A measurement of differential cross sections for diphoton production in proton-proton collisions using data collected by the CMS experiment at the LHC is presented. This analysis represents a stringent test of quantum chromodynamics, the theory of strong interactions, because the cross section is sensitive to next-to-next-to-leading order corrections in restricted regions of the phase space. Moreover, diphoton production represents the major source of background to the study of the Higgs boson in the diphoton decay channel, as well as to searches for several new physics models. Models used for background predictions in these analyses are thus validated.

The measurement is based on the statistical separation of the diphoton signal from the jet background by means of a template fit. The photon isolation is used as the discriminating variable. A detailed understanding of the photon reconstruction and identification is therefore an essential prerequisite.

After an illustration of the main features of the CMS electromagnetic calorimeter, a study of the effect of out-of-time pileup and electronic noise on photon observables is presented. This work has led to a more accurate simulation of these phenomena, with important benefits to all photon analyses.

The performance of the photon isolation variable in the particle-flow approach to event reconstruction is then illustrated. A novel technique used to avoid double counting of the photon energy deposit and the energy flow associated to other particles in the vicinity of the photon is presented. This method allows building the isolation templates entirely from data, to perform the measurement with small systematic uncertainty.

The analysis is then extended to final states where the photon pair is produced in association with at least one or two jets. In particular, the cross section is studied as a function of the variables commonly used to select the Higgs boson produced through vector boson fusion and decaying to a photon pair.

Finally, performance evolution studies of the electromagnetic calorimeter are presented. An enduring performance of the calorimeter is required for studying several physics processes during the high-luminosity running of the LHC. As the current detector was not designed to withstand the radiation levels expected in that period, an upgrade of the most affected parts is envisaged. Beam tests of a sampling calorimeter prototype channel based on radiation-hard CeF$_3$ scintillating crystals are illustrated. Measurements of response linearity, uniformity and energy resolution demonstrate that this is a promising upgrade design.
Riassunto

Si presenta una misura di sezioni d’urto differenziali per la produzione di coppie di fotoni in collisioni protone-protone, utilizzando i dati registrati dall’esperimento CMS all’LHC. Questa analisi rappresenta un test stringente della cromodinamica quantistica, la teoria delle interazioni forti, perché la sezione d’urto è sensibile a correzioni del second’ordine in alcune regioni dello spazio delle fasi. Inoltre, la produzione di coppie di fotoni rappresenta la maggior fonte di fondo per lo studio del bosone di Higgs nel canale di decadimento in due foton e per varie ricerche di nuova fisica. I modelli usati per predire il fondo in queste analisi vengono così verificati.

La misura si basa sulla separazione statistica del segnale di coppie di foton dal fondo di getti adronici, ottenuta per mezzo di un fit templato. L’isolamento del fotone è usato quale variabile discriminante. Una comprensione dettagliata della ricostruzione e identificazione dei fotoni è quindi un prerequisito essenziale.

Dopo un’illustrazione delle principali caratteristiche del calorimetro elettromagnetico di CMS, si presenta uno studio dell’effetto del pileup fuori tempo e del rumore elettronico sulle osservabili dei foton. Questo lavoro ha portato ad una simulazione più accurata di questi fenomeni, con importanti benefici per tutte le analisi su foton.

Si illustra poi il rendimento della variabile di isolamento dei foton nell’approccio particle-flow alla ricostruzione degli eventi. Si utilizza una nuova tecnica per evitare di conteggiare il deposito di energia del fotone nel flusso di energia associato ad altre particelle prodotte in prossimità di esso. Questo metodo permette di costruire le distribuzioni dell’isolamento interamente dai dati, e di effettuare la misura con una bassa incertezza sistematica.

L’analisi è poi estesa a stati finali in cui la coppia di foton è prodotta in associazione con almeno uno o due getti adronici. In particolare, la sezione d’urto è misurata in funzione delle variabili comunemente usate per la selezione del bosone di Higgs prodotto per fusione di bosoni vettori e decaduto in due foton.

Infine si presentano studi relativi all’evoluzione delle performance del calorimetro elettromagnetico. Un rendimento duraturo del calorimetro è necessario per studiare svariati processi di fisica durante la presa dati di LHC ad alta luminosità. Dato che il rivelatore attuale non è stato progettato per sopportare i livelli di radiazione attesi per quel periodo, è previsto un aggiornamento delle sue parti più degradate. Si illustrano i test su fascio di un canale prototipo di un calorimetro a campionamento basato su cristalli scintillanti di CeF₃. Misure di linearità, uniformità di risposta e risoluzione in energia dimostrano che questo è un progetto di aggiornamento promettente.
Contents

Introduction 1

1 Theoretical framework for diphoton production 3
  1.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3
  1.2 Theory of strong interactions . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
    1.2.1 The strong coupling constant . . . . . . . . . . . . . . . . . . . . . . . 4
  1.3 Proton-proton collisions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
    1.3.1 Hard scattering . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
    1.3.2 Non-perturbative effects . . . . . . . . . . . . . . . . . . . . . . . . . . 6
    1.3.3 Jet clustering . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7
  1.4 Diphoton production . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
  1.5 Event generators . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 9

2 The CMS detector and its electromagnetic calorimeter 15
  2.1 The CMS detector at the LHC . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
  2.2 The CMS electromagnetic calorimeter (ECAL) . . . . . . . . . . . . . . . . . 19
  2.3 Tuning of the out-of-time pileup and noise simulation in ECAL . . . . . . . . 23
    2.3.1 Comparison of energy sums in data and simulation . . . . . . . . . . . . 24
    2.3.2 Transverse energy sums in special runs . . . . . . . . . . . . . . . . . . 25
    2.3.3 Extended out-of-time pileup simulation . . . . . . . . . . . . . . . . . 29
    2.3.4 Improvements of the ECAL electronic noise model . . . . . . . . . . . . 32

3 Photon reconstruction and identification 39
  3.1 Clustering . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 39
  3.2 Supercluster energy corrections . . . . . . . . . . . . . . . . . . . . . . . . . . 41
    3.2.1 Factorized energy corrections . . . . . . . . . . . . . . . . . . . . . . . . 43
    3.2.2 Tuning of the calibration and simulated resolution . . . . . . . . . . . . 56
  3.3 Photon identification . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 63
    3.3.1 Electron rejection . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 64
    3.3.2 Identification variables . . . . . . . . . . . . . . . . . . . . . . . . . . . 65
    3.3.3 Selection requirements . . . . . . . . . . . . . . . . . . . . . . . . . . . 68
  3.4 Removal of the photon footprint from the isolation sum . . . . . . . . . . . . . 70
    3.4.1 Energy leakage cleaning based on veto-cones . . . . . . . . . . . . . . . . 73
    3.4.2 Supercluster footprint removal algorithm . . . . . . . . . . . . . . . . . 78

4 Measurement of the diphoton production cross section 85
  4.1 Data samples and event selection . . . . . . . . . . . . . . . . . . . . . . . . . . 85
    4.1.1 Event trigger . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 85
    4.1.2 MC samples . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 85
Introduction

The Standard Model (SM) of particle physics is a theory that has been shown to accurately predict a very large set of experimental observations. Quantum chromodynamics (QCD), in particular, describes strong interactions in the framework of the SM. It plays a key role in the study of hadron collisions, such as those taking place at the Large Hadron Collider at CERN in Geneva, Switzerland.

The measurement of the differential diphoton production cross section offers an important test of both perturbative and non-perturbative QCD. Theoretical predictions for this process are challenging, especially in restricted regions of the phase space where the relative contribution from higher-order QCD corrections is the largest.

The study of diphoton production in the SM is also motivated by the fact that this is the major source of background to the diphoton decay channel of the Higgs boson. New physics may also appear in diphoton final states as a non-resonant deviation from the predicted spectrum in gauge-mediated supersymmetry breaking (GMSB) or universal extra dimensions (UED) models, or as a narrow resonance in the case of Randall-Sundrum graviton models. Validating the event generators used to model the background in these analyses is therefore of crucial importance.

This thesis presents the measurement of differential cross sections for diphoton production in proton-proton collisions using data collected by the CMS detector at a center-of-mass energy of 7 TeV.

In Chapter 1, the theoretical framework for diphoton production in the context of the Standard Model is presented. The main phenomenological aspects of quantum chromodynamics are first illustrated, focusing on the most relevant ones for describing hadron collisions. The Monte Carlo generators used to obtain theoretical predictions for diphoton production are then presented, and the main features of each of them briefly discussed.

Chapter 2 first presents the Large Hadron Collider. A general description of the CMS experimental apparatus is then given, focusing on its crystal electromagnetic calorimeter (ECAL), whose performance is crucial for the measurement of photons. The layout of the ECAL and its readout are illustrated. Finally, a detailed discussion of the methods used to tune the simulation of the out-of-time pileup and electronic noise in the ECAL is given.

The algorithms used to reconstruct and identify photons are discussed in Chapter 3. After a brief review of the clustering of energy deposits in the ECAL, cluster energy corrections used to improve the photon energy resolution are introduced. The corrections are based on a novel, more adequate way of treating the correlation between the position of the photon and the probability of photon conversion in the material upstream of
ECAL. The variables used to identify photons, rejecting the jet and electron backgrounds, are then presented. The last part of the Chapter describes a novel technique, denoted as footprint removal, that avoids double counting between the photon ECAL energy deposit and the energy flow associated to other particles in the vicinity of the photon.

The measurement of the differential diphoton cross section is described in Chapter 4. The diphoton signal is separated from the jet background by a template fit using particle-flow isolation as the discriminating variable. The data-driven methods used to build the templates are described in detail, and their performance assessed using simulated events. The footprint removal technique plays a key role in this step. Then, the fit procedure is illustrated and the selection efficiency is measured in data using an electron control sample. The sources of systematic uncertainty are discussed, and the results are finally compared to several theoretical predictions.

Chapter 5 describes an extension of this analysis, where the differential cross section is studied as a function of jet observables in events with at least one or two jets in the final state. In the case of the two jet final state, the differential observables include those commonly used to select diphoton decays of Higgs bosons produced through vector boson fusion (VBF).

Finally, performance evolution and upgrade studies of the ECAL are presented in Chapter 6. The expected radiation levels during the High-Luminosity (HL) LHC running exceed the ones that the present detector was designed to withstand. A large class of physics analyses, especially those involving photons, relies on an enduring ECAL performance. An upgrade of the ECAL is therefore justified.

After a brief presentation of the upgrade options currently under consideration, beam tests of a sampling calorimeter prototype channel based on radiation-hard CeF$_3$ scintillating crystals are illustrated. Studies of response linearity, uniformity and energy resolution are performed using low and high energy electron data. They demonstrate that this design is a promising candidate for forward electromagnetic calorimetry in HL-LHC conditions.
Chapter 1

Theoretical framework for diphoton production

This chapter presents the theoretical framework for diphoton production in the Standard Model. After a brief introduction recalling the main phenomenological aspects of QCD and hadronic collisions, theoretical predictions obtained with Monte Carlo event generators are introduced. Only the key aspects of each generator are discussed here; more detailed information can be found in the given references.

1.1 Introduction

The theory describing the properties of the fundamental constituents of the universe is known as Standard Model (SM). Three forces are described within this framework: the strong, weak and electromagnetic interactions. The particle content of the SM can be divided into fermion matter particles (quarks and leptons) and boson force carriers.

The weak and electromagnetic interaction have a unified description, known as electroweak theory [1, 2], based on the SU$_2^L \times$ U$_1^Y$ gauge group. The electromagnetic interaction, mediated by a massless force carrier called photon, corresponds to the U$_1^Q$ gauge group. The weak force is mediated instead by the W$^\pm$ and Z massive gauge bosons, and is therefore short-ranged.

The mass of the W$^\pm$ and Z bosons is accommodated in the framework of a renormalizable quantum field theory by the Brout-Englert-Higgs mechanism [3–5]. A detailed review of this topic can be found in Ref. [6]. In short, a doublet of complex scalar fields is introduced:
\[ \Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \]

The SM Lagrangian is modified with a kinetic term, and a potential of the following form:
\[ V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \]

where \( \lambda \) must be positive for the potential to be bounded from below. If \( \mu^2 < 0 \), the minimum of the potential is defined by the following condition:
\[ \Phi^\dagger \Phi = \frac{v^2}{2} = -\frac{\mu^2}{2\lambda} \]
By choosing a vacuum configuration of the form
\[ \langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu \end{pmatrix} \]

the SU(2)_L × U(1)_Y symmetry is spontaneously broken because the Lagrangian is still invariant under the symmetry, but the expansion of the potential around the ground state is not. Inserting \( \langle \Phi \rangle \) into the kinetic term of the Lagrangian, the masses of the W± and Z bosons are expressed in terms of the vacuum expectation value of the Higgs field and the weak coupling constants. The excitations of the field around the vacuum state give rise to a massive and interacting state known as the Higgs boson, recently discovered [7, 8].

1.2 Theory of strong interactions

Strong interactions are described by quantum chromodynamics (QCD) [9–13]. QCD is based on the SU(3) gauge group, and the corresponding charge is called color. No free color charges have been observed so far. The color-neutral bound states of quarks, called hadrons, can be divided in mesons (quark-anti-quark pairs with integer spin) and baryons (quark triplets with half-integer spin).

The QCD Lagrangian can be written as:
\[ \mathcal{L}^{\text{QCD}} = \bar{q}_j (i \delta_{jk} \gamma^\mu \partial_\mu - g_s T^a_{jk} \gamma^\mu G^a_{\mu}) q_k - \frac{1}{4} F_{\mu\nu}^a F^{a}_{\mu\nu} \]

where \( q_i \) and \( G^a_{\mu} \) represent the quark and gluon fields, \( g_s \) the strong coupling constant, \( F_{\mu\nu}^a = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f^{abc} G_b^\mu G_c^\nu \) the field strength tensor, and \( f^{abc} \) the structure constants of SU(3).

1.2.1 The strong coupling constant

A fine structure constant for the QCD theory can be defined as a function of the strong coupling constant:
\[ \alpha_s = \frac{g_s^2}{4\pi} \]

Contributions of high order in perturbative QCD lead to a decrease of the effective \( \alpha_s \) value when the theory is probed at high momentum transfer (\( \mu^2 \)). This behaviour is encoded in the QCD beta function, defined as:
\[ \beta = \mu^2 \frac{\partial \alpha_s(\mu^2)}{\partial \mu^2} = -\frac{\alpha_s}{4\pi} \left[ \beta_0 \frac{\alpha_s}{4\pi} + \beta_1 \left( \frac{\alpha_s}{4\pi} \right)^2 + \mathcal{O}(\alpha_s^3) \right] \]

The \( \beta_i \) coefficients are determined by the gauge group. The value of \( \beta_0 \) is connected to the number of quark flavors by \( \beta_0 = 11 - 2N_f/3 \), and is positive for QCD. Solving the above equation at leading order yields:
\[ \alpha_s(\mu^2) = \frac{\alpha_s(\mu_0^2)}{1 + \alpha_s(\mu_0^2) \beta_0 \ln(\mu^2/\mu_0^2)} \]

as a function of the value of \( \alpha_s \) at a reference scale \( \mu_0^2 \), that has to be determined experimentally (Fig. 1.1).
The decrease of the strong coupling at high energies implies that quarks and gluons can be regarded as free in the short-distance regime. This feature of QCD is called asymptotic freedom [15], and allows to treat QCD processes taking place at high-energy hadron colliders by using perturbative methods.

### 1.3 Proton-proton collisions

#### 1.3.1 Hard scattering

The parton model assumes that each constituent of the proton, called parton, carries a fraction $x$ of the proton’s momentum. At the high momentum transfer typically reached by pp collisions at the LHC, the hard scattering involves partons that can be regarded as approximately free from the other constituents.

In these conditions, the hard scattering and the long-distance structure of the proton factorize and the interaction cross section can be written as:

$$
\sigma(pp \to X) = \int_0^1 dx_1 \int_0^1 dx_2 \sum_{a,b} f_a(x_1, \mu_F^2) f_b(x_2, \mu_F^2) \hat{\sigma}_{ab \to X}(Q^2, \mu_F^2)
$$

where $f_a(x)$ is the parton distribution function (PDF) describing the number density for parton $a$ to carry a fraction $x$ of the proton’s momentum, $\hat{\sigma}_{ab \to X}$ the hard scattering cross section, and $\mu_F$ an energy scale below which collinear soft emissions are absorbed into the PDF definition (factorization scale). The evolution of PDFs as a function of $\mu_F^2$ is described by the DGLAP equations [16–18]. The shape of the PDFs for valence quarks, sea quarks and gluons is shown in Fig. 1.2.

The emission of gluons or splitting of gluons into quarks (“parton shower”) that can take place before or after the hard scattering is referred to as initial or final state radiation, respectively. The probability for a final state parton to evolve from an initial energy scale...
Figure 1.2: Distributions of $x$ times the unpolarized parton distribution function $f(x)$ obtained in the NNPDF 2.3 global analysis at scales $\mu^2 = 10 \text{ GeV}^2$ and $\mu^2 = 10^4 \text{ GeV}^2$, with $\alpha_s(M_Z^2) = 0.118$. [14]

to a lower one without splitting (Sudakov factor) can be calculated in the perturbative theory, down to the scale where non-perturbative effects start to appear.

1.3.2 Non-perturbative effects

When the final state partons have such a low virtuality that $\alpha_s$ becomes large, they build color-neutral bound states. This process is called hadronization, and cannot be described using perturbative QCD. Two phenomenological approaches exist to treat it: the string model [19] and the cluster model [20, 21] (Fig. 1.3).

In the string model, the color field at the end of hadronization is described as a string of uniform energy density. The string then breaks into hadrons by creation of quark-anti-quark pairs. In the cluster model, on the other hand, all gluons are split in quark-anti-quark pairs. The quarks and anti-quarks are finally clustered to form color singlets that can decay to stable hadrons if the energy of the cluster field allows it.

Another aspect of the final state in pp collisions is the possibility of secondary interactions between other partons of the protons. This process is called underlying event (UE), and cannot be described by perturbative QCD because the momentum transfer is usually very small. The description of the UE is tuned using measurements of soft inelastic pp collisions. The rare cases where a secondary hard interaction takes place are referred to as multi parton interaction (MPI). A pictorial representation of all these phenomenological aspects of pp collisions is given in Fig. 1.4.
2.3. Summary and experimental tests of the Standard Model

In sections 2.1 and 2.2, the underlying theories of the SM were reviewed. The total description of the SM is the combination of these theories using a SU(3)$_C \times$SU(2)$_L \times$U(1)$_Y$ gauge group.

The hard interaction can be described using the total Lagrangian, which is

$$L_{SM} = L_{EW} + L_{Higgs} + L_{QCD} \tag{2.27}$$

with the single terms described by equations (2.1), (2.11), (2.16), and (2.19).

The properties of the bosons and fermions were already summarized in tables 2.1 and 2.2. As right-handed neutrinos have no quantum numbers at all, they are considered as non-existing in the SM.

1.3.3 Jet clustering

The hadronization of boosted quarks and gluons leads to collimated sprays of energetic hadrons, called jets [22]. Jets are defined by a prescription for grouping particles and assigning an estimate of the original parton momentum (jet algorithm). A comprehensive review of jet clustering can be found in Ref. [23]. A brief summary focusing on the most important aspects for pp collisions is given here.

A first class of jet algorithms is based on cones. The rationale behind this choice is that QCD hadronization does not alter dramatically the general structure of the energy flow in the event (more specifically, the energy flow in a cone).

A second class of algorithms is based on sequential recombination. In this case, a distance measure is defined between every pair of particles in the event. Particles are then grouped to form jets, starting from those with minimal separation according to the chosen distance.

An important and desirable property of a jet algorithm is infrared and collinear safety. This means that the set of hard jets should not change if a particle is subject to collinear splitting, or if a soft emission is added to the event.

If an algorithm is infrared- and collinear-safe, the interpretation of the event is not affected by the sequence of non-perturbative collinear splittings and soft emissions during hadronization, nor by the limited resolution and low energy thresholds of experimental detectors. Moreover, in fixed-order perturbative calculations, the cancellation of divergences that appear with opposite sign in tree-level and loop diagrams relies on this property.

The majority of cone algorithms is not infrared- and collinear-safe, because they rely on a seed to start the clustering procedure. Sequential recombination algorithms with a distance measure of the following form\textsuperscript{1}:

\textsuperscript{1}In the formula, the distance between two particles $i$ and $j$ is expressed as a function of their $p_T$ and their distance in the $(\eta, \phi)$ plane. See Section 2.1 for a definition of the coordinate system.
Diphoton production offers an important test of both perturbative and non-perturbative QCD. Prompt photons can be categorized in direct photons, that are radiated by quark lines in the hard scattering process, and fragmentation photons, that are produced by the collinear fragmentation of a quark or gluon in the final state.

At leading order, direct photon pairs are produced via quark-anti-quark annihilation $q \bar{q} \to \gamma \gamma$. This is commonly referred to as Born diagram (Fig. 1.5(a)). At next-to-leading order (NLO), direct diphoton production also includes the quark-gluon channel, while next-to-next-to-leading order (NNLO) predictions add the gluon-gluon channel, which includes a box diagram (Fig. 1.5(b)) and represents a non-negligible fraction of the total cross section.

At leading order, fragmentation photons (Fig. 1.5(c)) can be thought of as photon radiation from a quark in the final state. At higher orders, fragmentation includes all processes...
Figure 1.5: Examples of Feynman diagrams for diphoton production.

where a final state parton fragments to produce a photon in association with remnants. Such contributions are absorbed into photon fragmentation functions \( D_{qg}(z, \mu_F^2) \), where \( z \) represents the fraction of the initial parton momentum carried by the photon, and \( \mu_F \) the fragmentation scale. It should be noted here that the photon can be collinear to the fragmented parton. The quark-to-photon fragmentation function has been measured from LEP data [29].

The above argument implies that, beyond the leading order, the distinction between direct and fragmentation photons is not physical. Therefore, it will not be used to define the fiducial phase space of the measurement presented in this thesis.

Because of this rich phenomenology, theoretical predictions for diphoton production are challenging, especially in restricted regions of phase space where higher-order corrections give a sizable contribution to the cross section. A general description of the tools used to derive differential cross section predictions is given in the next Section.

### 1.5 Event generators

Before presenting the theoretical calculations and event generators used for diphoton predictions, it is worth defining the notion of isolated photon at particle level.

In collider experiments, the signal (isolated photons) is separated from a large background of neutral mesons using the energy flow around the object as a discriminating variable. The approach used in most cases is to apply a requirement on the sum of transverse energies of final state particles in a cone centered on the candidate. This method is called cone isolation. The threshold value can be chosen either as a function of the photon transverse energy (relative isolation), or as a fixed value (absolute isolation).

In both cases, the isolation requirement partially rejects fragmentation photons and soft gluon emissions in proximity to photons. A detailed discussion on the consequences of adding an isolation requirement to the phase space definition can be found in Ref. [30].
Figure 1.6: Invariant mass distribution of the photon pair at LO (dots), NLO (dashes) and NNLO (solid). The distributions of the box and NLO+box contributions are also shown. The inset plot presents the corresponding k-factors. The MSTW 2008 PDF set [33] has been used. [32]

An alternative approach to completely reject the fragmentation contribution is known as Frixione isolation [31]. The isolation requirement is defined as:

\[ E_{\text{T}}^{\text{iso}}(\Delta R) < \epsilon \left( \frac{1 - \cos(\Delta R)}{1 - \cos(\Delta R_0)} \right)^n \]

where \( E_{\text{T}}^{\text{iso}} \) is the isolation sum in a cone of size \( \Delta R \), and \( \Delta R_0 \) and \( n \) are parameters. The above inequality has to be satisfied for every \( \Delta R < \Delta R_0 \). This definition allows for the presence of energetic hadrons at the boundaries of the isolation cone, but becomes tighter in the collinear region, suppressing the fragmentation component. While the cone isolation has been already used in experimental measurements, the Frixione isolation is still under study. However, this prescription is theoretically well motivated and has been implemented in theoretical calculations.

2\( \gamma \)NNLO

The differential diphoton production cross section has been calculated at next-to-next-to-leading order (NNLO) in Ref. [32], and implemented in the 2\( \gamma \)NNLO parton-level Monte Carlo program. The numerical calculation is affected by infrared divergences that occur at intermediate stages. A detailed discussion of the method that was used to overcome them can be found in the reference. It is based on the \( \eta_T \) subtraction formalism, and is restricted to the treatment of colorless systems in the final state. Therefore, the fragmentation component cannot be calculated and is rejected using the Frixione isolation requirement.
Table 1.1: Cross section predictions at LO, NLO and NNLO for diphoton production in proton-proton collisions at $\sqrt{s} = 14$ TeV. The phase space is the same as for Fig. 1.6. The values are given for different choices of the factorization and renormalization scales. [32]

<table>
<thead>
<tr>
<th>Cross section (fb)</th>
<th>LO</th>
<th>NLO</th>
<th>NNLO</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_F = \mu_R = M_{\gamma\gamma}/2$</td>
<td>5045 $\pm$ 1</td>
<td>26581 $\pm$ 23</td>
<td>45588 $\pm$ 97</td>
</tr>
<tr>
<td>$\mu_F = \mu_R = M_{\gamma\gamma}$</td>
<td>5712 $\pm$ 2</td>
<td>26402 $\pm$ 25</td>
<td>43315 $\pm$ 54</td>
</tr>
<tr>
<td>$\mu_F = \mu_R = 2M_{\gamma\gamma}$</td>
<td>6319 $\pm$ 2</td>
<td>26045 $\pm$ 24</td>
<td>41794 $\pm$ 77</td>
</tr>
</tbody>
</table>

Figure 1.6 shows the diphoton cross section for a phase space defined by asymmetric transverse momentum ($p_T$) thresholds: $p_T > 40$ (25) GeV$^2$ for the leading (sub-leading) photon, and absolute rapidity$^3$ of less than 2.5. The cross section increases significantly with the perturbative order of the calculation (Table 1.1). At LO both photons have the same $p_T$. The region of the phase space where the sub-leading photon has a $p_T$ of less than 40 GeV is accessible only at NLO. The effect is further enhanced by the opening of a new production channel at the next perturbative order, up to NNLO. This also implies that the scale uncertainty on an NLO prediction cannot properly account for the theoretical uncertainty on diphoton production.

The k-factor (ratio of cross sections) between the NNLO and the NLO prediction is about 1.55. The box diagram gives a sizable contribution because of the large gluon partonic luminosity at the LHC, but the difference is mainly due to next-order corrections to the quark-gluon production channel. Actually, the k-factor between the NNLO prediction and the NLO calculation complemented by the box diagram is about 1.35.

Figure 1.7 presents the $p_T$ distributions for the leading and sub-leading photon. The higher-order corrections are largest in the region where the $p_T$ of the sub-leading photon is low. In the high-$p_T$ region, on the other hand, the effect of the asymmetric $p_T$ selection is less relevant, and the impact of corrections less dramatic. The distribution also exhibits an instability in correspondence of the leading photon $p_T$ threshold, that is due to the step-like behaviour of the LO prediction. A more detailed discussion of this effect can be found in Refs. [32, 34].

A comparison of the NNLO theoretical calculation with the results of a previous CMS measurement is shown in Fig. 1.8. A clear discrepancy between the data and the NLO prediction is observed in the region of low azimuthal angle separation between the two photons. In this region, forbidden by transverse momentum conservation at LO, the NNLO is the first effective higher-order correction to the cross section. The NNLO prediction is in agreement with data.

SHERPA

SHERPA [28] is a general-purpose event generator for simulation of collisions at high-energy colliders. For this analysis, version 1.4 is used. It is based on a tree-level multi-leg

$^2$Natural units are used throughout the dissertation.

$^3$Rapidity is defined as $\ln((E + p_L)/(E - p_L))/2$, where $E$ and $p_L$ are the energy and the component of momentum along the axis of the collision, respectively.
Figure 1.7: Transverse momentum distribution of the harder (left) and softer (right) photon for the same phase space as in Fig. 1.6. [32]

Figure 1.8: Differential diphoton cross section as a function of the azimuthal separation of the two photons, $\Delta\phi_{\gamma\gamma}$, at NNLO (blue) and at NLO (dotted black) calculated with a preliminary result from the $2\gamma$NNLO program, superimposed on results from CMS data (points). [35, 36]
matrix element generator (AMEGIC++/COMIX) for several processes described in the SM, as well as in new physics models.

The emission of additional partons from initial or final state legs\(^4\) is described by a parton shower model, and different jet multiplicities are merged using an extended version [37] of the CKKW [38] approach. Hadronization is described by a cluster model implemented in the AHADIC++ software package. Multi parton interaction is also described using the AMISIC++ routine.

For the diphoton predictions used in this thesis, up to three additional emissions in the final state are calculated at the matrix element level. Moreover, the box diagram is included at tree level. The factorization scale is set to the lowest invariant mass or negative virtuality in the core \(2\rightarrow2\) configuration clustered using a \(k_T\) algorithm (METS scale). More detailed information on this prescription can be found in the software documentation.

**Other event generators**

**PYTHIA** [39] is a complete event generator that accounts for initial and final state radiation, fragmentation, hadronization and underlying event. Several SM processes can be predicted. For the analysis described in this thesis, **PYTHIA** is used to describe the photon + jet final state, as well as the multi-jet final state. These calculations are based on a \(2\rightarrow2\) tree-level matrix element. Extra partons are added by the parton shower. The hadronization is performed using the Lund string model. A peculiar feature of **PYTHIA** is that it can be interfaced with the output of different matrix element generators to perform only the parton shower and hadronization steps.

**MADGRAPH** [40] is an event generator that computes tree-level multi-leg predictions at the matrix element level. Its output has to be interfaced with another program to perform the parton shower and hadronization steps. Different methods based on the MLM [40, 41] and CKKW [38] approaches are supported to match jets originating from the hard scatter described by a matrix element or from the parton shower [42]. **MADGRAPH**, complemented with **PYTHIA**, is used in this thesis to model the diphoton signal. Up to two jets are calculated at the matrix element level.

The **aMC@NLO** [43] package includes **MADGRAPH**, and complements it with the **MC@NLO** method for matching the parton shower to NLO matrix element calculations. Different jet multiplicities can be merged at NLO using the FxFx approach [44]. In this way, parton-level matrix element calculations at fixed order, parton shower matching and merging of different parton multiplicities are provided in a unique tool. The **aMC@NLO** diphoton sample used in the analysis includes up to two jets in the final state.

The **GOSAM** diphoton prediction is based on Ref. [45]. It features an NLO calculation of diphoton production in association with one or two jets, that also includes the contribution from quark and gluon fragmentation to photons. The virtual amplitude has been calculated with the **GOSAM** automated one-loop program. The output is not interfaced to a parton shower generator.

---

\(^4\)Lines of the Feynman diagram.
The DIPHOX [46] generator computes predictions for diphoton production using a full NLO calculation of the matrix element. Moreover, single and double fragmentation contributions to the cross section are included at NLO accuracy, as well as the box diagram. The prediction used for this thesis is complemented with GAMMA2MC [47] to include NLO corrections to the box contribution.

RESBOS [48, 49] also provides an NLO prediction for diphoton production, including transverse momentum resummation at next-to-next-to-leading-log accuracy.
Chapter 2

The CMS detector and its electromagnetic calorimeter

This chapter first provides a general presentation of the CMS experimental apparatus at the LHC. A more detailed description of the CMS electromagnetic calorimeter (ECAL) follows. Finally, the procedures used to tune the simulation of the out-of-time pileup and electronic noise on ECAL measurements are discussed. These methods have been documented in Ref. [50]. The discussions, text and illustrations which can be found in that reference are closely followed here.

2.1 The CMS detector at the LHC

The Large Hadron Collider [51] is a superconducting hadron collider designed to work at a center of mass energy \( \sqrt{s} = 14 \) TeV, with a bunch crossing every 25 ns. The collider is installed in the 26.7 km tunnel that was previously hosting the CERN Large Electron Positron (LEP) collider. During 2010 and 2011, the machine was operated at \( \sqrt{s} = 7 \) TeV. In 2012, the center of mass energy was increased to 8 TeV.

The machine is composed by two rings (Fig. 2.1), where counter-rotating proton beams are accelerated. Heavy ions, such as lead ions, can also be accelerated. The rings are connected to the rest of the CERN accelerator complex by two transfer lines. Protons are first accelerated in the LINAC2, then sequentially injected in the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS). This last accelerator acts as an injector for the LHC, at a beam energy of 450 GeV. Further details on the complex can be found in Ref. [51].

The tunnel is organized in eight arcs and eight straight sections, where experiments or other facilities (such as RF cavities for particle acceleration) are located. The centripetal force to keep particles in the orbit is provided by 1232 main dipole magnets, generating a magnetic field up to 8.3 T. The magnets use niobium-titanium superconducting coils cooled to a temperature below 2 K by thermal contact with superfluid helium.

In the approximation of small beam crossing angle and Gaussian-shaped beams, the instantaneous luminosity of the collider is expressed by the following formula [14]:

\[
\mathcal{L} = \frac{f n^2}{4 \pi \sigma^2}
\]
where \( f \) is the collision frequency, \( n \) the number of protons in each bunch, and \( \sigma \) the bunch transverse spread. In the LHC design, each of the 2808 injected bunches contains about \( 10^{11} \) protons and the luminosity can reach a peak value of \( 10^{34} \text{cm}^{-2}\text{s}^{-1} \).

**Overall concept of the CMS detector**

The Compact Muon Solenoid (CMS) [52] is a general purpose experiment designed to study a wide range of processes and search for new physics. The high instantaneous collider luminosity and the short spacing between bunch crossings pose challenging requirements on the detector design in terms of radiation tolerance, online data selection and time resolution. Moreover, a very high detector granularity is required to measure the large number of particles produced at each bunch crossing.

The detector (shown in Fig. 2.2) has an overall length of 22 m, a diameter of 15 m, and a weight of about 12500 tonnes. Its central feature is a superconducting solenoid providing a field of 3.8 T. The solenoid has a diameter of about 6 m and a length of about 13 m. A silicon pixel and strip tracker, a crystal electromagnetic calorimeter (ECAL) and a brass/scintillator hadron calorimeter (HCAL) are installed within the field volume. A muon system consisting of drift tubes (DT), cathode strip chambers (CSC) and resistive plate chambers (RPC) is embedded in the steel yoke of the magnet outside the superconducting coil, and exploits the 1.8 T return flux. Forward calorimeters extend the coverage at small deflection angles.

The coordinate system adopted by CMS is right-handed, with the origin at the nominal interaction point located at the center of the detector. The \( y \) axis points upward in the vertical direction, while the \( x \) axis points radially inward in the direction of the center of the LHC ring. The \( z \) axis is therefore oriented in the anticlockwise-beam direction. The azimuthal angle \( \phi \) is measured from the \( x \)-axis, in the \( x - y \) plane. The polar angle \( \theta \) is
measured from the positive z-axis. The pseudorapidity \( \eta \) is defined as:

\[
\eta = -\ln \tan \left( \frac{\theta}{2} \right)
\]

The components of momentum and energy transverse to the beam direction are denoted as \( p_T \) and \( E_T \), respectively.

A detailed description of CMS can be found in Ref. [52]. In the following, the most important features of the main detector components will be highlighted. As the ECAL plays a crucial role in the analysis presented in this thesis, a more detailed description of it will instead be given in Section 2.2.

**Silicon pixel and strip tracker**

In order to deal with the very large charged particle multiplicity at the LHC design luminosity (about 1000 particles produced every 25 ns), the CMS inner tracker is characterized by a very fine granularity. The detector covers the region \( |\eta| < 2.5 \), and consists of more than 200 m\(^2\) of active silicon sensors. A schematic view of the detector is shown in Fig. 2.3.

In the innermost region, about 66 million pixels of size 100×150 \( \mu m^2 \), oriented in the \( r - \phi \) and \( z \) directions, are used to accurately measure the track impact parameter and reconstruct the position of primary and secondary vertices. Three cylindrical pixel layers are located in the barrel region at a radius of about 4, 7 and 10 cm from the beam axis. Two disks are used to cover the forward region.
The pixel detector is surrounded by silicon strips, arranged in the same geometry at radii between 20 cm and 110 cm and organized in several layers. This guarantees a large hit redundancy to perform an efficient track pattern recognition and track fitting.

The calorimeter performance is directly affected by the amount of material that particles have to traverse in the innermost region of the detector. The thickness of the CMS tracker ranges from 0.4 radiation lengths in the central region, up to about 1.8 radiation lengths at $|\eta| \sim 1.4$.

**Hadron calorimeter**

The design of the hadron calorimeter has been influenced by the volume available to install the detector inside the magnet coil. The barrel part of the calorimeter (HB) is installed between the ECAL and the solenoid, and has a radial width of about 1.2 m. It is complemented by an extension located outside the solenoid (HO) that works as shower tail catcher using the solenoid as an additional absorber.

The HCAL consists of non-magnetic brass absorber interleaved with scintillating tiles. The light is extracted with optical fibers and readout by hybrid photo diodes (HPD). The total thickness of the calorimeter amounts to about 11.8 nuclear interaction lengths.

The calorimeter granularity is a multiple of the ECAL one. Up to $|\eta| = 1.6$, HCAL towers have a size of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. At larger pseudorapidities, the size increases to about $0.17 \times 0.17$.

**Muon system**

The main goal of this detector is the identification of muons and the precise determination of their momentum. The view of a quarter of the muon system is shown in Fig. 2.4.

In the barrel region, chambers with drift cells (DT) are used. They cover the region up to $|\eta| = 1.2$, and are arranged in four stations interleaved by the iron of the magnet.
return yoke. Cathode strip chambers (CSC) are used in the region between $|\eta| = 0.9$ and $|\eta| = 2.4$, where the magnetic field is less uniform and the particle rates are higher. Both detector designs provide an excellent spatial resolution for the measurement of the muon momentum. The fast resistive plate chamber (RPC) detectors, on the other hand, are used for triggering the readout of events containing muons with high transverse momentum. They cover the pseudorapidity range up to $|\eta| = 2.1$.

**Trigger and data acquisition**

The high LHC collision rate cannot be sustained by the storage system if an online data reduction is not performed. The needed reduction factor is of the order of $10^6$.

The CMS trigger system performs this task. It is organized in two levels: the Level-1 trigger (L1), based on custom-made programmable electronic boards, and the High-Level-Trigger (HLT), implemented at software level in a computer filter farm.

The L1 trigger is designed to reduce the event rate to 100 kHz. It uses data collected with a coarse segmentation, and has no access to tracking information. It is able to identify basic muon candidates and calorimetric deposits, as well as events with missing transverse energy. When an event passes the L1 filter, an accept signal is generated and data are transferred from the readout buffers to the HLT.

At this stage, a more complex and flexible processing of the information is performed, including the reconstruction of tracks. Algorithms similar to those used in the offline analysis are run to identify interesting signatures. The HLT output rate can reach 1 kHz.

### 2.2 The CMS electromagnetic calorimeter (ECAL)

The CMS ECAL is a homogeneous and hermetic electromagnetic calorimeter. It consists of a barrel (EB) region with 61200 lead tungstate (PbWO$_4$) crystals, closed at each end by
an endcap (EE) region containing 7324 crystals and extending the coverage up to $|\eta| = 3$. A global view of ECAL is shown in Fig. 2.5.

In this Section, a summary of the ECAL layout and signal readout from Ref. [52] is presented. A more complete set of information can be found in Refs. [52, 55–57].

**Properties of lead tungstate crystals**

Lead tungstate ($\text{PbWO}_4$) crystals are a suitable choice for calorimetry at the LHC, thanks to their high density (8.3 g/cm$^3$), short radiation length (0.89 cm) and small Molière radius ($R_M = 2.2$ cm). These characteristics allowed building a compact calorimeter with fine granularity. The scintillation decay time of $\text{PbWO}_4$ is comparable with the LHC bunch crossing time: about 80% of the light is emitted within 25 ns from the energy deposition. The emitted light is of blue-green color, with a broad wavelength maximum at about 420 nm. Thanks to high efficiency photodetectors, 4.5 photoelectrons per MeV are collected. The light yield dependency on temperature is about 2.1% at 18 °C. As the calorimeter response has to be stable at the level of per mil for precision measurements, the operating temperature is kept under control within ±0.05 °C by a specially regulated cooling circuit.

The radiation tolerance of $\text{PbWO}_4$ is sufficient to ensure good performance throughout the LHC operation, up to an integrated luminosity of about 500 fb$^{-1}$. A detailed discussion of the mechanisms of radiation damage to the $\text{PbWO}_4$ scintillation and light transmission properties will be presented in Chapter 6, in the context of ECAL upgrade studies for longer term operation.
The ECAL barrel

The barrel part of the ECAL covers the region $|\eta| < 1.479$. The crystals have a front face of $22 \times 22$ mm$^2$ and a length of 230 mm, corresponding to a depth of 25.8 radiation lengths. They have the shape of a truncated pyramid and are arranged in a $\eta - \phi$ geometry at a radius of 1.29 m from the detector axis.

The barrel granularity is 360-fold in the $\phi$ direction, and $2 \times 85$-fold in the $\eta$ direction. Each crystal covers a region of about $\Delta \eta \times \Delta \phi = 0.0174 \times 0.0174$. The crystals of each half of the barrel are grouped in 18 supermodules. Each supermodule subtends an angle of 20° in $\phi$, and comprises four modules covering different $\eta$ regions. Inside modules, crystals are kept in place by a very thin wall ($\sim 100$ µm) alveolar structure, and are oriented in an off-pointing geometry with respect to the center of the detector. The tilt angle is of about 3° in both $\eta$ and $\phi$.

In the barrel region, the scintillation light is detected by avalanche photo diodes (APD). Two APDs, each with an active area of $5 \times 5$ mm$^2$, are coupled to the rear face of each crystal. They are normally operated at a gain of 50 and read out in parallel. As the APD gain varies by about 3.1%/V at this gain, the bias voltage is kept stable within a few 10 mV.

The ECAL endcaps

The endcap part of the calorimeter covers the region between $|\eta| = 1.48$ and $|\eta| = 3$. Each endcap detector is divided in two Dee-shaped sections. The crystals are tapered and oriented in an quasi-projective geometry. Their faces have 28.6 mm sides and are arranged in a $x - y$ grid. They are grouped in $5 \times 5$ mechanical structures called supercrystals. Each crystal is readout by a vacuum photo triode (VPT) with a diameter of 25 mm and an active area of about 280 mm$^2$. 
The preshower detector

In the region $1.653 < \vert \eta \vert < 2.5$, the crystal calorimeter is complemented by a preshower silicon detector. The preshower is designed to improve the discrimination between isolated photons and neutral meson decays, and to estimate the energy loss in the material upstream of ECAL. Two pairs of silicon sensor layers are interleaved with two lead radiator plates with a depth of 2 and 1 radiation lengths, respectively. The silicon detector is segmented in strips with a pitch of 1.9 mm. Photons typically deposit about 5% of their energy in the lead radiators.

Amplitude reconstruction

The analog signal from the photodetectors is pre-amplified, shaped and then processed by a multi-gain amplifier with nominal gains of 1, 6 and 12. Each output is digitized by a 12-bit analog-to-digital converter (ADC) operating at 40 MHz. An integrated logic selects the highest non-saturated digital signal as output.

The signal pulse (see Fig. 2.7) is expected to start from the fourth out of ten recorded samples. The first three samples are used to estimate the value of the baseline pedestal. A linear combination of the first eight samples is used to estimate the pulse amplitude, using weights that take into account the expected pulse shape and the correlation between the samples introduced by the electronic noise.

The pulse amplitude, expressed in ADC counts, is then multiplied by an ADC-to-GeV conversion factor and weighted by the correction factors $S_i(t)$ and $C_i$. $S_i(t)$ accounts for the time variation of the channel response, while $C_i$ is used to equalize the response of all channels (inter-calibration coefficient). The calibration factors are extracted using a laser monitoring system and collision events. A detailed description of the calibration procedure can be found in Refs. [57, 59].

Energy resolution

The ECAL energy resolution can be parametrized as:

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

where $S$, $N$ and $C$ represent the stochastic, noise and constant term, respectively, and $E$ is expressed in GeV. Additional terms due to longitudinal shower losses, relevant only at energies above 500 GeV, are not included here.

The resolution was measured in beam tests (Fig. 2.8) [52]. For electrons between 20 and 250 GeV impinging in an area of $4 \times 4$ mm$^2$ around the point of maximum containment of a crystal, the following values were measured:

$$S = 2.8\%, \; N = 0.12, \; C = 0.3\%$$
Figure 2.7: Typical pulse shape measured in the ECAL, as a function of the difference between the time ($T$) of the ADC sample and the time ($T_{\text{max}}$) of the maximum of the pulse. The dots indicate ten discrete samples of the pulse, from a single event, with pedestal subtracted and normalized to the maximum amplitude. The solid line is the average pulse shape, as measured with a beam of electrons triggered asynchronously with respect to the digitizer clock phase. [58]

2.3 Tuning of the out-of-time pileup and noise simulation in ECAL

A significant discrepancy has been observed between data and simulation in sums of calorimeter energy deposits when comparing 2012 pp data to the corresponding simulated Monte Carlo (MC) samples. The discrepancy first came to light in Level 1 trigger studies, where the simulation significantly overestimated the rates of total calorimeter energy triggers compared to 2012 data. For some of the calorimetric trigger paths, the disagreement between data and MC rates reached values of MC/data of about 4. This was subsequently confirmed by studies of HCAL quantities, which observed significant $\eta$-dependent discrepancies between data and MC for offline reconstructed quantities, such as the summed transverse energy of Calorimeter Towers (CaloTowers).

The studies [50] presented in this Section have lead to a more accurate simulation of the effect of out-of-time pileup and electronic noise on ECAL observables. Both of these improvements will be used in the generation of MC samples for LHC Run II analyses, as they provide a better data/MC agreement in various electron/photon and jet-related quantities relevant for physics analyses. Moreover, selected samples with Run I conditions have been reproduced for specific analyses, such as the Run I legacy measurement.
of the Higgs boson properties in the diphoton decay channel.

2.3.1 Comparison of energy sums in data and simulation

A reference data run from the 2012 pp running period was chosen (Run 202299) and events were selected from the ZeroBias dataset using the HLT trigger path HLT_ZeroBias_v7. The ZeroBias trigger requires a coincidence of two dedicated beam position monitors in the same bunch crossing. This dataset was compared against a ZeroBias Monte Carlo dataset (neutrino gun plus pile-up) generated with pp collisions at $\sqrt{s} = 8$ TeV and 2012 pile-up conditions. Both in-time and out-of-time pile-up were simulated, with interactions generated in a (-50ns,+50ns) time window, centered on the time of the true interaction.

The digitized signals in the ECAL were analyzed using the standard ECAL pulse reconstruction algorithm [56] and the resulting reconstructed hits (RecHits) were calibrated according to [57]. Hits with transverse energy greater than a threshold value were summed to form the total transverse energy ($\Sigma E_T$) recorded per event in the ECAL barrel and endcaps respectively. The dependence of the $\Sigma E_T$ on the transverse energy threshold of the contributing hits is shown in Fig. 2.9. The high energy tail of the $\Sigma E_T$ is dominated by the lower energy hits that contribute in the sum. In order to decouple the study of the pileup, from effects related to the accuracy of the ECAL noise model which is studied elsewhere, a lower threshold of 0.5 GeV is applied to all hits that contribute to the $\Sigma E_T$ observable.
Figure 2.9: Spectrum synthesis of $\Sigma E_T$ for different RecHit transverse energy thresholds, from data taken during the 2012 reference run (Run 202299). [50]

Figure 2.10 shows the profile distribution of the mean value of $\Sigma E_T$ versus the number of reconstructed vertices (used as a proxy for the number of simultaneous interactions per bunch crossing) for EB (left) and EE (right). The distributions for data (filled circles) and Monte Carlo (open squares) are compared. For the barrel, the distributions are in reasonable agreement. However, in the endcaps a significant discrepancy, growing with pileup, is observed - the transverse energy sum in Monte Carlo for a given number of pileup interactions is greater than the one observed in data.

Figure 2.11 shows the distribution of $\Sigma E_T$ for events with twenty reconstructed vertices in data and Monte Carlo, for EB (left) and EE (right) respectively. The distributions are normalized to the same number of events and the two bottom panels show the ratio of data to Monte Carlo. The shift towards higher values of $\Sigma E_T$ in Monte Carlo is particularly evident in EE.

A similar trend can be observed in Fig. 2.12, which shows the transverse energy sum of ECAL RecHit clusters (called SuperClusters) in various ranges of pseudorapidity, for events with twenty reconstructed vertices. The discrepancy between data (points) and Monte Carlo (histograms) is strongly $\eta$-dependent. These results are in qualitative agreement with those observed in HCAL quantities, which also showed a strongly $\eta$-dependent discrepancy between data and Monte Carlo.

2.3.2 Transverse energy sums in special runs

The dependence of the $\Sigma E_T$ discrepancy with event pileup was explored by comparing the following special data runs taken in 2012 with the corresponding Monte Carlo samples:

- **Very low pileup**: LHC fill 2576, CMS run 193092, mean event pileup $\sim 0.03$. 


Figure 2.10: Profile distribution of $\Sigma E_T$ vs the number of reconstructed vertices showing run 2012 data and the default Monte Carlo used in 2012 physics analyses. [50]

Figure 2.11: 1D $\Sigma E_T$ distribution for a fixed PU value (PU=20). [50]
• High pileup: LHC fill 2824, CMS run 198603, mean event pileup ∼ 45.

Figures 2.13 and 2.14 show the comparison of $\Sigma E_T$ (using RecHits with transverse energy greater than 0.5 GeV) for data and Monte Carlo, for these two special runs. In both cases, the agreement between data and Monte Carlo for EE is substantially better than that shown in Fig. 2.11, although some discrepancies still remain.

The profile distribution of the mean $\Sigma E_T$ versus the number of reconstructed vertices is shown in Fig. 2.15 for the high pileup run and the reference run 202299. The agreement between the data and MC samples in EB is satisfactory for both runs, while in EE this is true only for the high pileup run.

At this point, it is worth noting that the composition of in-time and out-of-time pileup is quite different for run 198603 (high pileup run) and run 202299 (nominal pileup, run 2012C conditions). For the former, the LHC filling scheme consisted of two colliding bunch pairs, with a large separation between the two (approximately 1000 bunch crossings, or 25 µs). As a result, collisions in this run will experience in-time pileup, but no
Figure 2.13: 1D RecHit $\Sigma E_T$ distributions (data and MC) for the low PU ($\mathrm{PU} \simeq 0$) run. \cite{50}

Figure 2.14: 1D RecHit $\Sigma E_T$ distributions (data and MC) for the high PU run. \cite{50}
Figure 2.15: Profile plot of $\Sigma E_T$ versus the number of reconstructed vertices, showing run 2012C data, a special high PU data run (average pileup $\approx 45$) and default MC (with in-time and OOT PU for runC conditions and only IT PU for comparison with the special high PU run). [50]

out-of-time PU. On the other hand, the LHC filling scheme for run 202299 consists of 1368 colliding bunch pairs, separated into 38 bunch trains. Each bunch train consisted of 36 colliding bunch pairs, separated by 50 ns. For this run, collisions contain both in-time and out-of-time pileup.

The relevance of this can be deduced from the right hand plot of Fig. 2.15. Here, the $\Sigma E_T$ profile distributions are in good agreement between data and MC when the collisions include only in-time PU. The significant discrepancy between data and MC only occurs when both in-time and out-of-time PU are present.

Furthermore, one can observe that the effect of out-of-time PU on the mean $\Sigma E_T$ in data is minimal in EE (compare the two data distributions in the right-hand plot of Fig. 2.15). This is expected, and will be explained further in the next section. Looking at the same plot, it is clear that this is not the case for the Monte Carlo, which shows a substantial effect of out-of-time PU on the mean $\Sigma E_T$ (compare the two data distributions in the right-hand plot of Fig. 2.15). This indicates that the data/MC discrepancy observed in EE is strongly related to the simulation of out-of-time PU in the Monte Carlo.

### 2.3.3 Extended out-of-time pileup simulation

Pileup is simulated by generating individual minimum bias events and overlaying these on top of the hard scattering event prior to event digitization, during which a simulation of the detector noise is added. The number of pileup interactions is sampled randomly from a Poisson distribution. To simulate out-of-time pileup, the signals from additional minimum bias interactions are shifted in time relative to the primary bunch crossing of the hard scattering event and are digitized with this time shift applied.
The simulation used for most LHC Run I analyses simulates OOT PU in a [-50,+50] ns window. For collisions with 50 ns bunch spacing, this means that three bunch pairs are simulated: the in-time bunch, a bunch at -50 ns and a bunch at +50 ns.

**ECAL amplitude reconstruction in presence of out-of-time pileup**

The ECAL amplitude reconstruction (described in [56]) uses a linear series of weights, that multiply the signal amplitude observed in each of the 10 digitized bunch crossings read out by the ECAL front-end electronics. The weights are optimized to provide the best estimate of the energy of an in-time signal. The weights are pedestal subtracting, meaning that the first three weights (used for the pulse pre-samples) are negative, providing an event-by-event subtraction of the pedestal. Separate weights are derived for EB and EE, using pulse shapes measured from test beam data. For offline amplitude reconstruction, eight out of the ten weights are used: three for pedestal subtraction and five to estimate the amplitude of the signal.

Figure 2.16 shows the result of applying the weights to in-time and out-of-time signals. The left-hand plot shows the reconstructed amplitude of a signal with a time shift of \( \Delta t \) with respect to an in-time signal. The shapes have been normalized to the amplitude for an in-time signal.

The features of the left-hand plot of Fig. 2.16 can be explained qualitatively. Firstly, the reconstructed amplitude for an out-of-time signal is reduced relative to an in-time signal. This happens because the weights are derived from the pulse shape of an in-time signal, with the pulse maximum in the sixth digitized sample. If the phase of the out-of-time pulse is sufficiently large and positive, the reconstructed pulse amplitude will tend to zero. This is because the peak of the pulse falls outside of the digitization window. This can clearly be seen for positive values of \( \Delta t \). For negative values of \( \Delta t \), the pulse amplitude in fact becomes negative. This is because the peak of the pulse falls within the three pre-samples, to which negative weights are applied in order to perform the dynamic
pedestal subtraction.

The right-hand plot of Fig. 2.16 shows the integral of the left-hand plots, assuming an integration window starting from a particular value of $\Delta t$ on the x axis and extending to $+50$ ns, with pulses spaced at 50 ns intervals. In other words, it shows the contribution to the reconstructed energy sum of a pulse train, similar to the colliding bunch structure used for 2012 CMS pp data. The properties of the weights are such that, for a bunch train whose length is comparable or larger than the digitization time window (10 samples or 250 ns), the average amplitude, summing positive and negative components, is close to zero. This means that a long train of out-of-time pulses forms a “white-noise” which is effectively subtracted by the pulse reconstruction algorithm. This can clearly be observed in the right-hand plot of Fig. 2.16, which shows that the integral value of a long train of pulses, spaced at 50 ns intervals, is close to zero when considering an integration window that extends from $\Delta t \approx -300$ ns to $+50$ ns.

If one now examines the right-hand plot of Fig. 2.15, this feature can in fact be observed in the 2012 data. The mean value of $\Sigma E_T$ for a given event pileup is almost the same for runs with long bunch trains consisting of in-time PU and out-of-time PU and those with only in-time PU. Returning to the right-hand plots of Fig. 2.16, one can see that if a narrow integration window is used (-50 to $+50$ ns) the effect of out-of-time signals will cause a positive bias in the mean value of $\Sigma E_T$. This can also be observed in the right-hand plot of Fig. 2.15, where the Monte Carlo simulation of out-of-time PU uses a window of -50 to $+50$ ns, and the resulting value of $\Sigma E_T$ for a given number of reconstructed vertices is significantly larger than the simulation with only in-time PU. The conclusion of this discussion is that, in order to correctly simulate the effect of OOT PU on data taken in 2012 with long (72 BX) bunch trains, it is necessary to generate out-of-time collisions in a window extending back to at least -300 ns.

Comparison of data and improved simulation

Figure 2.17 shows data/MC comparisons of $\Sigma E_T$ and the mean value of $\Sigma E_T$ versus the number of reconstructed vertices using MC generated with an extended OOT PU window. In these plots, the MC window for OOT PU is extended from (-50,$+50$ ns) to (-300,$+50$ ns). There is a significant improvement in the level of agreement between data and MC when the extended time window is used. The effect of pileup (both IT and OOT) on the ECAL reconstruction is much greater in EE since the pileup density (average energy deposited per crystal by the multiple minimum bias interactions in a typical event) is strongly $\eta$-dependent.

Extending the OOT PU window in the Monte Carlo to (-300,$+50$ ns), larger than the time window used for the digitization of ECAL pulses, simulates the effect of an infinite length bunch train. The effect on $\Sigma E_T$ of finite length bunch trains can be seen in Fig. 2.18. The plots in this figure show, separately for EB and EE, the mean value of $\Sigma E_T$ versus the position of the colliding bunch pair within a bunch train. For the data, each bunch train consists of 36 colliding bunches with 50 ns spacing. There is a minimum gap of 10 BX between bunch trains. Bunch to bunch intensity variations are removed by dividing $\Sigma E_T$ by the number of reconstructed vertices on an event-by-event basis. Events from all 38 bunch trains within each LHC orbit are overlaid on the same plot.
For EE, one can see a clear dependence of the value of $\Sigma E_T$ on the bunch position within the train. There is a positive bias at the front of the train, which reduces to an equilibrium value after approximately 6 colliding bunches, and a negative bias at the end of the train, where only out-of-time interactions from earlier bunches in the train contribute. These effects can be understood by referring to the behaviour of the amplitude reconstruction algorithm as a function of the signal jitter as shown in Fig. 2.16. Note that the Monte Carlo simulation, represented by the solid and dashed lines, does not include these end-of-train effects.

### 2.3.4 Improvements of the ECAL electronic noise model

An important improvement to the Monte Carlo simulation was in the simulation of the single-channel noise in EB and EE. This improvement, which is described in more detail in [60] accounts for time-dependent changes in the energy equivalent noise due to increases in APD dark current in EB [61], as well as amplification of the noise due to loss, and subsequent correction, of crystal transparency. The noise increased from about 40 MeV to 50 MeV in 2012.

The implementation of the noise model in 2012 consists of identifying three separate data-taking periods: A and B (April-June), C (July-September), and D (October-December). The average noise per channel per period, measured from dedicated pedestal runs, has been injected in the simulation. During the digitization step, the simulated energy for each channel is divided by a factor which corresponds to the average transparency correction for that channel in a given run period. Electronic noise is then applied by smearing the signal amplitude (in ADC counts) in each of the ten digitized samples with the noise level measured in data for the same channel in the same run period. Fi-
Figure 2.18: Mean $\Sigma E_T$ vs the position of the colliding bunch within a bunch train (top - EB, bottom - EE). The value of $\Sigma E_T$ for each bunch position is normalized to the number of vertices in each event. The data is represented by the black points. The mean value of $\Sigma E_T$ per reconstructed vertex for the default [-50,+50 ns] and extended window [-300,+50 ns] Monte Carlo simulation is shown by the dashed and solid lines. [50]
nally, the resultant signal amplitude is corrected for transparency loss by multiplying by the same transparency correction used above. In this way, the amplification of the noise relative to the simulated energy of signal pulses due to transparency loss is correctly taken into account.

In the following, the data recorded in run 200473 are used, because the detector conditions for this run are closer to those used in the run-dependent MC for modelling the Run2012C data-taking period. Figures 2.19-2.20 show the effect of these simulation improvements on various low-level ECAL RecHit and cluster-related quantities. An improvement in the data/MC agreement is visible when moving from the standard MC sample with a PU simulation window of [-50 ns, +50 ns] to a MC sample with the wider window of [-300 ns, +50 ns]. A further substantial improvement is observed when the new noise model is applied in addition to the extended PU simulation window. The overall good agreement between data and MC demonstrates the maturity of the understanding of the detector and reconstruction algorithms, and the importance of these two model improvements.

The improved data/MC agreement is also observed in high-level quantities used directly in many analyses. Figures 2.21 and 2.22 shows the considerable improvement seen in ECAL isolation and shower shape variables (see Chapter 3 for their definition) when the new MC is used. This directly reduces the systematic errors associated with dealing with residual data/MC discrepancies, such as MC to data smearing and event reweighting procedures. A significant improvement in data/MC agreement for the pileup contribution to jets is observed when an extended window for OOT PU generation is used. This will substantially reduce the systematic uncertainty in the jet energy corrections.

The shower shape variables are also sensitive to the amount of material upstream of ECAL. Removing the disagreement due to out-of-time pileup allows to use them for tuning the material budget in the simulation.

Finally, the level of agreement between the data and simulated $\Sigma E_T$ profiles is re-examined with a reduced threshold on the RecHit transverse energy of $E_T > 0.1$ GeV. Figure 2.23 shows the profile distribution of $\Sigma E_T$ versus the number of reconstructed vertices for EB and EE comparing data to the default MC, the MC with the extended OOT PU window, and with the addition of the new ECAL noise model. The improvements obtained with the combination of these two model improvements are striking, and the plot shows the importance of the new noise model for this lower RecHit threshold, especially in EB.
Figure 2.19: This set of plots shows an improved agreement between the MC and the data when the PU simulation window is extended from [-50 ns, +50 ns] to [-300 ns, +50 ns]. The agreement improves further when, on top of the MC samples with the extended PU window, the new noise model described in the text is applied. The plots show the RecHit multiplicity for EB (top) and for EE (bottom). [50]
Figure 2.20: As Fig. 2.19, but showing the supercluster multiplicity in EE (top) and occupancy versus \( \eta \). [50]
Figure 2.21: Plots of the ECAL isolation variable for data vs the default MC and a MC sample with an extended OOT PU window (in this case extending from [-200,+50] ns) and with the new noise model. The left-hand plot is for the ECAL barrel, and the right-hand plot is for the ECAL endcaps. [50]

Figure 2.22: Plots of the R9 shower shape variable for data vs the default MC and a MC sample with an extended OOT PU window (extending from [-300,+50] ns) and with the new noise model. The left-hand plot is for the ECAL barrel, and the right-hand plot is for the ECAL endcaps. [50]
Figure 2.23: $\Sigma E_T$ profiles using MC with the improved ECAL noise model and the extended OOT PU window and $E_T > 0.1$ GeV single crystal threshold. [50]
Chapter 3

Photon reconstruction and identification

This chapter first describes the reconstruction of photons from the measured energy deposits in the electromagnetic calorimeter and the estimation of their energy. The variables used for photon selection are then presented, with a focus on the shape of the electromagnetic shower and the isolation calculated with the particle-flow approach to event reconstruction. The methods used are thoroughly documented in Refs. [62, 63]. The discussions, text and illustrations which can be found in those references are closely followed, to provide a general overview of the photon reconstruction and identification performance with LHC Run I data\(^1\).

The performance of a new energy correction scheme, originally developed [64] for the analysis presented in this thesis and used to describe the effect of the material upstream of the ECAL on the cluster energy, is described in more detail. A novel technique [65] used to subtract the photon footprint from the isolation sum is also presented. This method, known as footprint removal, plays a crucial role for building the isolation templates used in the diphoton cross section measurement.

3.1 Clustering

Photons are reconstructed from energy deposits in the ECAL crystals by grouping them in clusters compatible with the shape of energy deposition expected for a photon or electron. The clustering algorithm does not assume that the reconstructed particle is a photon or an electron at any stage. Therefore, electrons from Z decays can be used to test the photon trigger, reconstruction and identification efficiencies, as well as to tune the energy scale in data and the energy resolution in the simulation.

The clustering of ECAL energy deposits is designed to collect the energy that is spread when electrons radiate and photons convert as a result of the interaction with the CMS tracker material. The fraction of energy lost by an electron interacting with the material upstream of ECAL ranges from 33% to 86%. The conversion probability for photons from H decays is shown in Fig. 3.1. This energy is mainly spread along the φ direction because of the bending of electron trajectories in the solenoidal magnetic field. The spread in the η direction is negligible for \( p_T \gtrsim 5\) GeV.

\(^1\)Some of the results presented in this Chapter are obtained for a center of mass energy of 8 TeV. When not stated otherwise, the performance in 7 TeV data can be assumed to be comparable.
The “hybrid” clustering algorithm is used in the ECAL barrel to exploit the cylindrical symmetry of that sub-detector and the fact that crystals are arranged along the \( \eta \) and \( \phi \) directions. It aims at constraining the cluster shape to a small window in the \( \eta \) direction, while allowing for a larger extension in the \( \phi \) direction.

The algorithm starts from a seed crystal that is chosen using a minimum transverse energy (\( E_T \)) threshold of 1 GeV. Linear arrays of 5 crystals in the \( \eta \) direction are added in a range of 17 crystals in both directions of \( \phi \), if the energy of the array exceeds 0.1 GeV. Contiguous arrays are collected in the final cluster only if they contain a seed array of at least 0.35 GeV. The final cluster has a fixed width of 5 crystals along \( \eta \) and a variable extension along the \( \phi \) direction. It is called the “supercluster” (SC).

The transition region between the ECAL barrel and endcaps is affected by detector services shadowing the crystals, and is excluded from the reconstruction acceptance. The fiducial region is defined by the \( |\eta| < 1.4442 \) requirement for the barrel region of the ECAL, and by \( 1.566 < |\eta| < 2.5 \) for the ECAL endcaps.

In the ECAL endcaps crystals are arranged along the \( x \) and \( y \) directions, therefore a different algorithm based on \( 5 \times 5 \) crystal matrices is used (“multi-\( 5 \times 5 \)”). Crystals whose \( E_T \) is maximal relative to their four direct neighbors and exceeds 0.18 GeV are used as seeds. A \( 5 \times 5 \) matrix centered on each seed is constructed in decreasing order of seed \( E_T \). The matrices associated to different seeds can partially overlap. These \( 5 \times 5 \) clusters are grouped into a SC if the total \( E_T \) exceeds 1 GeV, within a range of 0.07 in the \( \eta \) direction and 0.3 rad in the \( \phi \) direction around each seed crystal.

The weighted positions of all \( 5 \times 5 \) clusters in the SC are then extrapolated to the planes of the preshower sub-detector. The extrapolated position of the most energetic cluster is used as reference point to define the preshower clustering window. This window extends in the \( \phi \) direction by the maximum distance between the reference point and the extrapolated positions of all clusters, and in the \( \eta \) direction by \( \pm 0.15 \). The window is then
Figure 3.2: Comparison of the distributions of the ratio of reconstructed over generated energy for simulated electrons from Z boson decays in a) the barrel, and b) the endcaps, for energies reconstructed using superclustering (solid histogram) and a matrix of $5 \times 5$ crystals (dashed histogram). No energy corrections are applied. [62]

extended by an additional $\pm 0.15$ rad in the $\phi$ direction. The preshower energy deposits within this range are added to the SC ECAL energy.

The effect of the clustering on the recovery of radiated energy in $Z \rightarrow e^+e^-$ Monte Carlo events is shown in Figure 3.2. The electron energy reconstructed using the algorithms described above is compared with the simple sum of energy deposits running on a matrix of $5 \times 5$ crystals centered on the seed crystal. It is seen that the left tail of the distribution of the reconstructed energy over the generated one ($E/E_{gen}$) is reduced using the SC.

The $R_9$ variable is defined as the energy sum of the $3 \times 3$ crystals centered on the SC seed divided by the SC energy. Unconverted photons deposit about 97% of their energy in a region of $5 \times 5$ crystals. Their SC usually consists of this single region both in barrel and endcaps, and has a value of $R_9$ close to 1. On the other hand, photons that convert before reaching the calorimeter surface have a wider shower profile along the $\phi$ direction and lower values of the $R_9$ variable. Figure 3.3 shows the distribution of the $R_9$ variable for simulated Higgs decay photons that convert in the tracker material at a radius of less than 85 cm, and for those that convert later or do not convert at all before reaching the ECAL surface.

3.2 Supercluster energy corrections

The photon energy is estimated on the basis of the associated SC energy, corrected for several effects that degrade its resolution. The energy deposited in the ECAL is affected by variations of the energy lost in the tracker by showers starting before the ECAL surface, and of the fraction of this energy that is collected by the clustering algorithm.
Figure 3.3: Distributions of the $R_9$ variable for photons in the ECAL barrel that convert in the material of the tracker before a radius of 85 cm (solid filled histogram), and those that convert later, or do not convert at all before reaching the ECAL (outlined histogram). [63]

Moreover, the sum of crystal energies in the supercluster is affected by variations of the shower containment due to the imperfect hermeticity of the ECAL. This is due to a leakage of the electromagnetic shower energy in non-instrumented regions between the ECAL modules, as well as smaller gaps between the ECAL crystals. This effect is mitigated by the relative angle between the shower axis and the crystal sides for photons originating in the interaction region at the center of CMS. The equivalent scintillating crystal depth seen by such showers is decreased by $0.75X_0$ for inter-crystal gaps, and by about half a crystal depth for inter-module cracks.

The supercluster energy is corrected through several steps [66, 67]:

- laser calibrations to correct for crystal transparency losses, applied only on data;
- crystal inter-calibration constants to obtain a uniform calorimeter response [66];
- at the cluster level, a geometrical correction (denoted as $C(\eta)$) is applied in the barrel to compensate for the energy leakage from the lateral faces of the crystals;
- for electrons and low $R_9$ photons (mostly converted) a correction factor is applied to correct for the energy not recovered by the superclustering algorithm because of material effects. For photons with large $R_9$ this correction is not applied and the energy of the photon is estimated simply from the energy in a $5 \times 5$ matrix around the supercluster seed;
- a geometrical correction is applied in the barrel to correct for energy losses in the cracks [68] between ECAL modules.
3.2.1 Factorized energy corrections

A factorized energy correction scheme is adopted as a default to correct for material budget effects. It is based on three variables:

- $\sigma_\phi/\sigma_\eta = \text{“brem”}$: a measurement of the asymmetry between the $\eta$ and $\phi$ development of the electromagnetic shower. Non-converted photons and electrons with low bremsstrahlung have an almost circular transverse shower profile, hence $\text{brem} \sim 1$. On the other hand, converted photons and electrons with large bremsstrahlung have a transverse shower profile more elongated along $\phi$ because of the magnetic field. The normalization with $\sigma_\eta$ also reduces the energy dependence of the brem variable.

- $\eta$ : pseudorapidity of the supercluster.

- $E (E_T)$ : (transverse) energy of the supercluster after a first correction based on brem has been applied.

The expression for the factorization of the corrections before 2012 [67] was:

$$F = f(\sigma_\phi/\sigma_\eta) \cdot F(E_T, \eta)$$

The first factor $f(\sigma_\phi/\sigma_\eta)$ was called “material” correction. The second factor, $F(E_T, \eta)$, was referred to as “residual” correction and was used to correct for all the remaining effects not taken into account by the first correction.

However, this approach has the disadvantage of neglecting the correlation between $\sigma_\phi/\sigma_\eta$ and $\eta$. The broadening of the shower (then spread by the magnetic field along $\phi$) is due to the interaction of the photons and electrons with the material upstream of the ECAL. Therefore, it depends on the material distribution in the tracker. The amount of material expressed in radiation lengths (see Fig. 3.4) varies from about 0.4 $X_0$ in the central region up to about 1.8 $X_0$ at $|\eta| \approx 1.5$. This induces an $\eta$ dependence on the broadness of the transverse shower profile. Moreover, correction factors derived on simulated electrons were applied to both electrons and photons. Because of the higher probability of radiative energy losses for electrons, this lead to a systematic overestimation of the photon energy, especially at low $E_T$.

A novel factorization scheme [64] is presented in the following. This method properly takes into account the correlation between the shower broadening and the material in the tracker, and it has now been adopted as default in the event reconstruction. In the new formulation:

$$F_{e,\gamma} = f(\sigma_\phi/\sigma_\eta, \eta)_{e,\gamma} \cdot F(E_T)_{e,\gamma}$$

The corrections for photons and electrons are extracted from photon and electron particle gun samples separately. Photons or electrons are generated with a flat distribution in $\eta$, $\phi$ and $E_T$ in the range $2 < E_T < 250$ GeV.

The correction factors are then validated in the same simulated samples, and crack corrections are included to reduce the tails in the energy measurement. The resulting scheme is further validated on $H \rightarrow \gamma\gamma$ simulation and $Z \rightarrow e^+e^-$ simulation and data.

Finally, the energy scale in the fourth module of the ECAL barrel (the region where the tracker material budget is largest and less accurately known) is more closely investigated and the uncertainties on the energy measurement are computed.
Energy corrections for photons

As the supercluster energy is corrected only for photons with $R_9 < 0.94$ (0.95) in the ECAL barrel (endcaps), the extraction of the correction factors is limited to this range. The possible use of $R_9$ in place of brems as the input variable for the "material" correction has been investigated, and found not to improve the performance of the correction scheme.

The algorithm used to extract the $f(\sigma_\phi/\sigma_\eta, \eta)$ correction bins the events in a two-dimensional matrix $\sigma_\phi/\sigma_\eta \times \eta$. The $\eta$ binning has been chosen by referring to Fig. 3.4, in order to obtain a material distribution as flat as possible within one $\eta$-bin and excluding the regions between the ECAL barrel modules and the barrel/endcap gap region:

<table>
<thead>
<tr>
<th>$\eta_{\text{min}}$</th>
<th>0.02</th>
<th>0.25</th>
<th>0.46</th>
<th>0.81</th>
<th>0.91</th>
<th>1.01</th>
<th>1.13</th>
<th>1.44</th>
<th>1.65</th>
<th>1.8</th>
<th>2.0</th>
<th>2.2</th>
<th>2.3</th>
<th>2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta_{\text{max}}$</td>
<td>0.25</td>
<td>0.42</td>
<td>0.77</td>
<td>0.91</td>
<td>1.01</td>
<td>1.13</td>
<td>1.44</td>
<td>1.65</td>
<td>1.8</td>
<td>2.0</td>
<td>2.2</td>
<td>2.3</td>
<td>2.4</td>
<td>2.5</td>
</tr>
</tbody>
</table>

The bin $[1.653, 1.8]$ covers the endcap regions not instrumented with the preshower detector. A slice of $\pm 2$ degrees in $\phi$ has been removed around each module boundary to avoid inter-module cracks. After deriving the corrections outside of the cracks, the function $f(\sigma_\phi/\sigma_\eta, \eta)$ is extrapolated in the cracks.

A minimum cut of $E_T > 10$ GeV is required for all photon superclusters. This corresponds to the lowest photon transverse energy selection applied at reconstruction level. For each of the $(\sigma_\phi/\sigma_\eta, \eta)$ bins, a histogram of the variable $E_{\text{rec}}/E_{\text{gen}}$ is constructed, where $E_{\text{rec}}$ is the reconstructed supercluster energy and $E_{\text{gen}}$ the generated energy. $E_{\text{rec}}$ is the raw energy (energy sum of all the crystals in the supercluster) multiplied by the geometrical
correction $C(\eta)$ in the barrel. In the endcap, it is equal to the raw ECAL energy summed to the associated preshower energy.

The $E_{\text{rec}}/E_{\text{gen}}$ distribution is then fitted with a Crystal Ball function\(^2\). The inverse of the position of the most probable value (mpv) of the distribution, which is usually shifted below unity, is used as a multiplicative correction factor for all events in the bin. This implies that the correction only brings the mpv of the distribution to 1, without reducing the tails in one bin. An effective improvement of the resolution is observed only when integrating over multiple bins.

The dependency of the mean of the Crystal Ball fit on the brem variable before any correction is shown on Fig. 3.5 (left). The different colors represent different $\eta$ bins. The larger the amount of material is in the $\eta$ bin, the steeper the brem dependence is. The slopes in the endcaps are larger than the slopes in the barrel, and even at low brem (where the transverse shower profile is almost circular) the most probable value of $E_{\text{rec}}/E_{\text{gen}}$ for the endcap showers only reaches about 97%.

After correcting with $f(\sigma_{\phi}/\sigma_{\eta}, \eta)$ a residual dependence of $E_{\text{rec}}/E_{\text{gen}}$ is still observed as a function of the transverse energy. An extra multiplicative correction is therefore derived. The same Crystal Ball fit is applied on the $E_{\text{rec}}/E_{\text{gen}}$ distribution. A parametrization of the fit results as $F(E_T)$ in the barrel and $F(E)$ in the endcap gives the best results. Finally, $f(\sigma_{\phi}/\sigma_{\eta}, \eta), F(E_T)$ and $F(E)$ are fitted to smooth out bin-by-bin fluctuations.

\(^2\)The Crystal Ball function is defined as:

$$f_{\text{CB}}(m) = \begin{cases} \frac{N}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(m-m_0)^2}{2\sigma^2}\right), & \text{for } \frac{m-m_0}{\sigma} > -\alpha; \\
\frac{N}{\sqrt{2\pi}\alpha} \left(\frac{n}{|n|}\right)^n \exp\left(-\frac{|n|^2}{2}\right) \left(\frac{n}{|n|} - |n| - \frac{m-m_0}{\sigma}\right)^{-n}, & \text{for } \frac{m-m_0}{\sigma} \leq -\alpha. 
\end{cases}$$
Validation on simulated photons

The energy corrections described above are applied on the same photon gun sample used to extract them. The most probable value of the $E_{\text{rec}} / E_{\text{gen}}$ distribution is studied as a function of $E_T$ (Fig. 3.6), $E$ (Fig. 3.7), $\eta$ (Fig. 3.8) and $\sigma_{\phi}/\sigma_{\eta}$ (Fig. 3.9).

The following correction schemes are compared:

- Before any correction is applied (i.e. raw energy multiplied by $C(\eta)$ in the barrel, raw energy plus the preshower energy in the endcap).
- The old correction scheme $f(\text{brem}) \times F(E_T, \eta)$, where the corrections are derived from an electron MC sample.
- The old correction scheme $f(\text{brem}) \times F(E_T, \eta)$, where correction factors have been re-extracted from a photon MC sample.
- The new correction scheme $f(\text{brem}, \eta) \times F(E_T)$ that is used in the barrel.
- The new correction scheme $f(\text{brem}, \eta) \times F(E)$ that is used in the endcaps.

Figure 3.6 shows that the previously used corrections with the scheme $f(\text{brem}) \times F(E_T, \eta)$ derived from electrons over-correct the photon energy especially at low $E_T$, by about 4%. Moreover, a trend as a function of $\sigma_{\phi}/\sigma_{\eta}$ is observed. Using the corrections derived from simulated photons, $E_{\text{rec}} / E_{\text{gen}}$ is brought back to 1 on average, but the energy is systematically over-corrected at high $E_T$ in the barrel and under-corrected by about 1% in the endcap. In the new correction scheme, the two approaches for the residual corrections ($F(E_T)$ and $F(E)$) give consistent results. The parametrization of the residual corrections as $F(E_T)$ in the barrel and $F(E)$ in the endcap performs slightly better than just using $F(E_T)$ everywhere.

After corrections, $E_{\text{rec}} / E_{\text{gen}}$ as a function of $E_T$ is centered around unity within a few per mil. The remaining effects are consistent with statistical fluctuations in the photon MC sample. The corrections perform as expected also as a function of the $\eta$ variable, where only a mild structure in the fourth ECAL barrel module ($1.16 < \eta < 1.44$) is observed.
Figure 3.7: $E_{\text{rec}}/E_{\text{gen}}$ most probable value as a function of $E$ for different correction schemes, evaluated using a photon MC sample. [64]
Energy corrections for electrons

The energy corrections are derived for electrons without applying any requirement on \( R_9 \). The same approach used for photons is applied to single electron gun Monte Carlo samples. The minimum threshold on the supercluster transverse energy is lowered to 5 GeV, as low-energy electrons are of interest for \( H \to ZZ \) and \( H \to W^+W^- \) analyses. The energy estimate derived in this scheme is used as supercluster energy in the High Level Trigger. Offline analyses apply more refined factorized corrections taking into account the bremsstrahlung profile of the electron.

The dependence of the most probable value of the fitted Crystal Ball function on the brem variable before any correction is shown in Fig. 3.10. It is very similar to the one observed with photons. The slopes of brem as a function of \( \eta \) are steeper for electrons than for photons. This can be explained by the different electron interaction with matter, that makes electrons more sensitive to the amount of material traversed. The residual dependence is at the percent level for electrons. This suggests that even after correcting for the material with \( f(\sigma_\phi/\sigma_\eta, \eta) \) a sizable amount of energy is still lost. In other words, early bremsstrahlung losses are not fully recovered and the brem variable does not describe this clustering inefficiency.

The same validation procedure described before is performed using the electron gun MC sample used to derive the electron corrections. Figures 3.11-3.14 show that the old scheme over-corrects, as a function of \( E_T \), over the entire range. In the barrel, \( E_{\text{rec}}/E_{\text{gen}} \) after the material correction is offset by a few per mil, while in the endcap it shows an almost linear dependence as a function of \( E_T \), reaching about 1.5% at low \( E_T \). The new corrections bring \( E_{\text{rec}}/E_{\text{gen}} \) to 1 on average.
Figure 3.10: (left) Projection of $f(\sigma_\phi/\sigma_\eta, \eta)$ in bins of $\eta$, for an electron Monte Carlo sample with no correction applied; (left) 2D representation of the $f(\sigma_\phi/\sigma_\eta, \eta)$ correction after smoothing. [64]

Figure 3.11: Most probable value of $E_{\text{rec}}/E_{\text{gen}}$ as a function of $E_T$ for different correction schemes, for an electron MC sample. [64]
Figure 3.12: Most probable value of $E_{\text{rec}}/E_{\text{gen}}$ as a function of $E$ for different correction schemes, for an electron MC sample. [64]

Figure 3.13: Most probable value of $E_{\text{rec}}/E_{\text{gen}}$ as a function of $\sigma_{\phi}/\sigma_{\eta}$ for different correction schemes, for an electron MC sample. [64]
Energy corrections for losses in gaps

The energy corrections derived in the previous Sections were extracted using events where the photon/electron does not hit any of the inter-module or supermodule cracks, to avoid mixing up material effects and ECAL geometry effects. The corrections for the leakage of shower energy into cracks between ECAL channels are of the order of a few percent for energy deposits in the vicinity of a crack boundary. They are parametrized using local ECAL coordinates (i.e., coordinates connected with the ECAL geometry) and are applied only in the barrel. Crack corrections are not applied in the endcaps because of a residual Monte Carlo mismodeling of the ECAL geometry after alignment. They are derived using a sample of unconverted photons [68].

Every cluster inside the supercluster is corrected, taking into account both cracks in $\eta$ (between ECAL modules and at $\eta=0$) and in $\phi$ (between ECAL supermodules). The same correction is applied to the $5 \times 5$ energy sum used in the case of high-$R_9$ photons.

Local containment corrections finally correct for the dependence of the energy response on the impact position of the particle with respect to the center of the crystal. They are found to be of the order of a few per mil.

Validation using $H \rightarrow \gamma\gamma$ and $Z \rightarrow e^+e^-$ events

The diphoton mass resolution in $H \rightarrow \gamma\gamma$ simulated events is studied with the different correction schemes. A striking performance improvement is observed using the new corrections with respect to the old ones tuned on electrons: the effective sigma\(^3\) of the diphoton resonance is decreased by about 17%.

\(^3\)The effective sigma ($\sigma_{\text{eff}}$) of a distribution is defined as the half-width of the tightest window containing at least 68% of the integral of the distribution.
The $H \rightarrow \gamma\gamma$ analysis [69] categorizes events based on their mass resolution to maximize the analysis sensitivity. As photons reconstructed in the ECAL barrel feature the best energy resolution, the performance of the corrections has been tested in the following diphoton categories:

- Category 0: Both photons in the barrel, with $R_9 > 0.94$
- Category 1: Both photons in the barrel, with at least one having $R_9 < 0.94$

The mass resolution is improved by about 8% in category 0, due to the crack corrections being applied also at high $R_9$ to the $5 \times 5$ energy in the new scheme, and by about 14% in category 1, thanks to the new factorization. Local containment corrections have a marginal effect on the mass resolution.

The performance of the electron corrections is tested in $Z \rightarrow e^+e^-$ events in data and simulation. Standard electron selection requirements are applied to select a sample with high purity. The study is conducted on electrons with a $p_T$ of at least 30 GeV. The resolution of the di-electron peak is studied using an unbinned maximum-likelihood fit. The signal is described by a Breit-Wigner lineshape convoluted with a Crystal Ball function to account for the smearing due to energy mismeasurement. The residual background is described with an exponential function. The goal of the study is to extract the $\sigma$ parameter of the Crystal Ball ($\sigma_{CB}$) that represents the electron energy resolution.

The $n$ parameter of the Crystal Ball function is fixed to $n = 5$. The fit is first performed in simulated events. The $\alpha$ parameter of the Crystal Ball function is fixed to the value fitted in MC, and then data events are fitted. This allows to compare the mass scale and resolution consistently between data and MC, as several parameters of the fit could otherwise be correlated in a non-trivial way. Events are divided in three categories: the results of the fit are shown in Fig. 3.15 for barrel-barrel, Fig. 3.16 for barrel-endcap and Fig. 3.17 for endcap-endcap events separately, for the old and the new correction schemes.

The comparison shows that the mass resolution is improved by about 20% both in data and MC using the new corrections. In the endcap-endcap category, the mass scale calibration is also significantly improved. The performance degradation in data with respect to simulation, both in terms of resolution and mass scale, is roughly the same for both correction schemes.
Performance in the ECAL barrel fourth module

The performance of energy corrections in the fourth module of the ECAL barrel (1.13 < |η| < 1.44) requires a more careful study. The results of the fits in data and MC are shown on Fig. 3.18. The old corrections give a small over-correction of the mass scale. The new corrections show a mass shift of -1.6 GeV in MC and -2.8 GeV in data. The 1.2 GeV difference between data and MC is visible for both old and new corrections.

In photon and electron gun samples only a mild structure of the energy scale as a function of η in the fourth module is visible (Fig. 3.8 and 3.14). The invariant mass distribution is sensitive to the shape of the tails of $E_{\text{rec}} / E_{\text{gen}}$. Checking the behaviour of the energy scale for $H \rightarrow \gamma\gamma$ events with both photons in the fourth module, it is observed that the mass scale is more accurate than for electrons.

The description of the material upstream of the ECAL might be imperfect. This is consistent with the presence of pixel, TIB and TOB strip tracker services in the η slice corresponding to the ECAL barrel fourth module. A very accurate description of these elements is difficult. The problem might reside in the description of the material at low
radii, which yields a stronger effect on low $p_T$ electrons. Material at large radii also has a smaller effect on low $p_T$ electrons because the photons from bremsstrahlung and subsequent conversions won’t have enough space to spread out. The material at low radii, on the other hand, will have a larger effect on low $p_T$ electrons because the bremsstrahlung photons and subsequent conversions could escape the clustering algorithm.

These arguments lead to an improved scheme for factorizing the energy corrections. The dependence on $\eta$ of the residual correction is reintroduced, leading to a 2D×2D scheme:

$$F_{e,\gamma} = f(\sigma_\phi/\sigma_\eta, \eta)_{e,\gamma} \cdot F(E_T, \eta)_{e,\gamma}$$

The $F(E_T, \eta)$ correction for photons and electrons is shown Fig. 3.19. As expected, only for electrons the $\eta$ dependence of $F(E_T, \eta)$ is sizable, especially at low $p_T$, in the fourth barrel module and in the region of the endcaps not covered by the preshower. In these regions the clustering algorithm has a lower efficiency to recover very early bremsstrahlung emissions into low energy superclusters. Figure 3.20 shows that this parametrization decreases the mass shift in the simulation. The disagreement between data and simulation is still present, and is investigated using alternative descriptions of the detector geometry in the context of the measurements of the tracker material budget.

Energy measurement uncertainties

The novel energy correction scheme presented above leads to improvements in the energy scale and resolution with respect to the previous scheme. A related question is to study if it also improves the uncertainty in the energy measurement. A correct estimation of the uncertainty in the ECAL energy measurement for electrons is particularly important, since it is used together with the estimate of the track momentum uncertainty to compute the uncertainty in the combined electron energy measurement.

The effective sigma of the $E_{rec}/E_{gen}$ is compared for the two schemes in Figs. 3.21-3.25 as a function of supercluster variables and separately for electrons, low-$R_9$ and high-$R_9$ photons. It is interesting to note that high-$R_9$ photons can still have a value of $\sigma_\phi/\sigma_\eta$ up
Figure 3.19: 2D representation of the $F(E_T, \eta)$ correction for low $R_9$ photons (left) and for electrons (right), after $f(\sigma_\phi/\sigma_\eta, \eta)$ is applied. [64]

Figure 3.20: Reconstructed $Z \rightarrow e^+e^-$ invariant mass in data and simulation for events where both electrons are reconstructed in the fourth module. [64]
to 3, which means that in some cases $R_9$ does not discriminate against a significant spread of the supercluster energy along the $\phi$ direction. As expected, high-$R_9$ photons have the lowest energy uncertainty. The uncertainty is not affected much by the change in the factorization scheme. The most significant improvement is observed for low-$R_9$ photons in the barrel. The uncertainty for high-$R_9$ photons is also decreased by the use of crack corrections.

Uncertainties are parametrized as a function of bremsstrahlung and $E_T$. The particle gun samples are divided in 6 $\eta$ bins and 6 bremsstrahlung bins. For each $(\eta, \sigma_{\phi}/\sigma_\eta)$ bin, the events are further split in 7 $E_T$ bins and the $E_T$ dependence of the effective sigma is fitted. Figure 3.26 shows an example of this procedure. The fit bias is negligible (Fig. 3.27).

Finally, the energy uncertainty is validated in $Z \rightarrow e^+e^-$ events from data and simulation. Figure 3.28 shows that the simulation is in agreement with the data, except for a small discrepancy in the endcap region.

### 3.2.2 Tuning of the calibration and simulated resolution

The energy scale calibration and the simulated energy resolution are tuned using electron showers from $Z$ decays to obtain the best possible agreement between data and simulation. This is of crucial importance, for instance, in the study of the Higgs boson in the...
Figure 3.23: Effective sigma of $E_{\text{rec}}/E_{\text{gen}}$ distributions as a function of $\sigma_{\phi}/\sigma_{\eta}$ in the endcap for electrons and low $R9$ photons. [64]

Figure 3.24: Effective sigma of $E_{\text{rec}}/E_{\text{gen}}$ distributions as a function of $E_T$ in the barrel for electrons, low and high $R9$ photons. [64]

Figure 3.25: Effective sigma of $E_{\text{rec}}/E_{\text{gen}}$ distributions as a function of $E_T$ in the endcap for electrons and low $R9$ photons. [64]
Figure 3.26: Effective sigma of $E_{\text{rec}}/E_{\text{gen}}$ distributions for an electron MC sample as a function of $\sigma_\phi/\sigma_\eta$ and $E_T$ for $0 < |\eta| < 0.7$ (left), and as a function of $E_T$ for $0.8 < \sigma_\phi/\sigma_\eta < 1.0$ and $0 < |\eta| < 0.7$ (right). [64]

Figure 3.27: Bias of the scale uncertainty introduced by the fit, evaluated using an electron MC sample. [64]
diphoton decay channel [69]. This Section presents the state-of-the-art calibration methods for photon energy using 8 TeV data collected in 2012. For these results, multivariate analysis techniques have been used to further improve the energy resolution. These methods are not discussed here; more detailed information can be found in Ref. [63].

A very pure sample of Z bosons is selected and electrons are reconstructed as photons, using only the ECAL information. Invariant mass distributions are compared in data and simulation, and residual corrections are applied to maximize their agreement.

A multiplicative correction is first applied to data to match the energy scale in the simulation. Data events are split according to the data-taking epoch. An analytic fit to the invariant mass peak is performed in each sub-sample using a convolution of a fixed-width Breit-Wigner distribution with a Crystal Ball function. The fit treats separately the different pseudorapidity regions (two regions in the barrel and two in the endcaps). This correction accounts for imperfections in the crystal transparency corrections. No significant dependence on $R_9$ is observed, while different detector regions exhibit a different evolution of the residual corrections with time.

At the same time, a smearing correction is applied to simulated electrons to match the di-electron mass resolution observed in data. The correction takes the form of a Gaussian distributed random addition to the reconstructed energy. Showers are categorized in two $R_9$ bins in each of four absolute pseudorapidity regions. Combining different pairs of categories, di-electron events are divided in 36 sub-samples. The likelihood of the joint fit to the data for all sub-samples is maximized to extract the smearing amplitudes. Finally, $E_T$-dependent corrections to the energy scale are applied as a function of $|\eta|$ and $R_9$. Figure 3.29 shows the values extracted in this final step for different $(|\eta|, R_9)$ categories.

The total smearing correction applied to simulated photons ranges from about 0.7% to 1% for the high-$R_9$ barrel category, from about 1% to 2% for the low-$R_9$ barrel category, and from 1.6% to 2% in the endcaps. It is extracted with an uncertainty of about 0.05-0.1%
Figure 3.29: Residual discrepancies in the photon energy scale obtained in the final step of the fine-tuning procedure, as a function of $E_T$, for different $\eta$ and $R_\eta$ categories. The reciprocals of these values are applied as corrections to the energy scale. The horizontal error bars indicate the ranges of the $E_T$ bins. Some of the error bars have been deflected vertically to avoid overlap with others. [63]

depending on the category. Figure 3.30 shows the reconstructed invariant mass distribution in $Z \to e^+e^-$ events after all scale and resolution corrections have been applied. The core of the distribution shows an excellent data/MC agreement both in the barrel and in the endcaps. The low mass tail of the endcap distribution is affected by a slight discrepancy. This is due to the fact that the Gaussian smearing cannot account for some non-Gaussian effects that are important there. The peak of the distribution does not coincide with the Z boson mass, because electrons have been reconstructed here as photons. This implies that the lower fraction of energy collected in the supercluster for electrons is not properly corrected for using photon-based corrections.

The single-electron energy resolution is measured using methods similar to those used for extracting the smearing correction, and is shown in Fig. 3.31. The simulated resolution matches the one observed in data, except for low-$R_\eta$ endcap electrons where the non-Gaussian effects on the degradation of the energy resolution are most important. The accurate description of the electron energy resolution in the simulation builds up confidence in an accurate description of the photon energy resolution as well. Nevertheless, electron and photon showers differ because of the different probability of radiative losses already discussed in Section 3.2.1. Therefore, the extrapolation of the resolution measurements obtained on electrons to photons requires a careful treatment.

The imperfect modeling of the tracker material budget is the main source of uncertainty in the electron to photon extrapolation. The amount of material upstream of the ECAL has been measured with two complementary methods. The first method studies the energy loss by bremsstrahlung of electrons from Z decays while they cross the mate-
Figure 3.30: Reconstructed invariant mass distribution of electron pairs in $Z \rightarrow e^+e^-$ events in data (points) and in simulation (histogram). The electrons are reconstructed as photons and the full set of photon corrections and smearings are applied. The comparison is shown for (left) events with both showers in the barrel and (right) the remaining events. For each bin, the ratio of the number of events in data to the number of simulated events is shown in the panels beneath the main plot. [63]

Material of the tracking system. The transverse momentum of the electron is first measured ($p_{in}$) with a subset of hits located in the central region of the detector. The same measurement is then repeated ($p_{out}$) using the hits located in the outer layers, closest to the ECAL surface. The relative energy loss ($(p_{in} - p_{out}) / p_{in}$) measured comparing these two values is correlated with the amount of material crossed by the electron track.

Simulations where the amount of tracker material is increased uniformly by 10, 20 and 30% are used to calibrate the relative energy loss due to the material budget in the simulation. The calibration is then applied on data to measure the amount of material. A similar strategy is used to measure the material budget from the energy loss of low-\(p_T\) charged-hadron tracks (0.9 GeV < \(p_T\) < 1.1 GeV). For \(|\eta| < 1.6\) the two methods give compatible results, that are shown in Fig. 3.32.

These results are based on a calibration where the assumption of uniform material mismodeling has been made. Nevertheless, the difference between reality and simulation is most likely due to incorrect description of specific structures in localized regions, such as the tracker services (cooling, cables). This measurement is used to assign a systematic uncertainty in the photon energy scale. Localized structures can affect in a different way electrons and photons. Therefore, the uncertainty is calculated in a conservative way, studying the scale variation induced by a 10% uniform deficit of material for \(|\eta| < 1.0\), and a 20% deficit for \(|\eta| > 1.0\). It ranges for 0.03% in the central ECAL barrel to 0.3% in the outer endcap regions.
Figure 3.31: Relative photon energy resolution measured in small bins of absolute supercluster pseudorapidity in $Z \rightarrow e^+e^-$ events, for data (solid black circles) and simulated events (open squares), where the electrons are reconstructed as photons. The resolution is shown for (upper plot) showers with $R_9 > 0.94$ and (lower plot) $R_9 < 0.94$. The vertical dashed lines mark the module boundaries in the barrel, and the vertical gray band indicates the range of $|\eta|$, around the barrel/endcap transition, removed from the fiducial region. [63]
Electron and photons also differ in the average depth of the energy deposition in the ECAL scintillating crystals. The ECAL features an adequate uniformity of light collection efficiency as a function of the depth of emission, thanks to one face of each barrel crystal being depolished. This effect is described in the simulation. The uncertainty in the uniformity has to be propagated to the photon energy scale.

Moreover, the loss of crystal transparency induced by radiation affects the light collection uniformity. A dedicated study has been setup to assess the magnitude of this effect, using laboratory measurements of transparency loss in irradiated crystals. The uncertainty in the energy scale is found to be 0.04% for unconverted photons, and 0.06% for converted ones.

Finally, the linearity of the energy scale has been verified studying the energy-momentum ratio $E/p$ in electrons from $W$ and $Z$ decays, and the dependence of the reconstructed $Z$ mass as a function of the scalar sum of the $E_T$ of its decay electrons. Figure 3.33 shows the result of these studies. The uncertainty is found to be at the per mil level.

Further minor uncertainties in the photon energy scale and resolution measurements are described in [63]. In summary, the uncertainty in the photon energy scale at $p_T \simeq m_Z/2$ is about 0.1% in the central ECAL barrel, 0.15% in the outer barrel, and 0.3% in the endcaps.

### 3.3 Photon identification

The experimental background to prompt photon signals is mainly due to boosted neutral mesons produced in jets. Neutral pions produced with high transverse momentum decay to a pair of collimated photons, that are reconstructed as a single photon candidate in the
Figure 3.33: Residual discrepancy of the energy response in data relative to that in simulated events as a function of transverse energy (for the $E/p$ analysis, squares) and of $H_T/2$ (for the dielectron mass analysis, circles) in four $\eta$ and $R_9$ categories. $H_T$ is the scalar sum of transverse energies of jets in the event. The dielectron analysis is restricted to events where both the electron showers fall in the same $\eta, R_9$ category. The uncertainties in the fit parameters of a linear response model are shown by bands. [63]
Table 3.1: Fractions of photons and electrons, in the ECAL barrel and endcap, passing the two different electron vetoes. The statistical uncertainties in the values given for electrons are negligible. Candidates are preselected using the loose working point of the cut-based photon selection described in Section 3.3.3 [63].

<table>
<thead>
<tr>
<th></th>
<th>Barrel</th>
<th></th>
<th>Endcap</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>γ</td>
<td>e</td>
<td>γ</td>
<td>e</td>
</tr>
<tr>
<td>Conversion-safe veto</td>
<td>99.1 ± 0.1%</td>
<td>5.3%</td>
<td>97.8 ± 0.2%</td>
<td>19.6%</td>
</tr>
<tr>
<td>Pixel track seed veto</td>
<td>94.4 ± 0.2%</td>
<td>1.4%</td>
<td>81.0 ± 0.6%</td>
<td>4.3%</td>
</tr>
</tbody>
</table>

material of the beam pipe or within the first layer of the pixel detector.

The electron rejection can be improved by using a more stringent set of requirements, at the price of a loss in efficiency for photons. This is the case for the “pixel track seed veto” requirement, where photon candidates are rejected if at least two hits in the pixel detector suggest the presence of a charged particle pointing at the supercluster. The geometrical matching is performed using a window that is computed starting from the position of the supercluster and extrapolating to the central region of the detector. The expected bend of the trajectory due to the magnetic field is taken into account. This mechanism is the same used to seed the Gaussian-Sum-Filter track reconstruction for electrons. A more detailed description is available in Ref. [62].

The efficiencies for photons and electrons to pass these requirements have been measured in 8 TeV data, and are illustrated in Table 3.1 separately for the barrel and the endcap. Photon efficiencies are extracted from $Z \rightarrow \mu^+ \mu^- \gamma$ events, while for electrons a sample of $Z \rightarrow e^+ e^-$ events is used.

3.3.2 Identification variables

The identification of prompt photons is based on two main categories of observables: those describing the shape of the electromagnetic shower in the ECAL, and those describing the energy flow around the candidate (isolation).

Shower shape variables

The $\sigma_{\eta\eta}$ variable is used to measure the energy weighted spread in the $5 \times 5$ matrix centered on the seed crystal of the supercluster. It is defined as:

$$
\sigma_{\eta\eta}^2 = \frac{\sum (\eta_i - \bar{\eta})^2 w_i}{\sum w_i},
$$

where the sum runs over all elements of the $5 \times 5$ matrix around the most energetic crystal in the supercluster, and $\eta_i = 0.0174 \, \hat{\eta}_i$ in EB, $\eta_i = 0.0447 \, \hat{\eta}_i$ in EE with $\hat{\eta}_i$ denoting the index of the $i$th crystal along the $\eta$ direction. The individual weights $w_i$ are given by $w_i = \max(0, 4.7 + \ln(E_i/E_{5 \times 5}))$, where $E_i$ is the energy of the $i$th crystal and $\bar{\eta} = \frac{\sum \eta_i E_i}{\sum E_i}$ is the weighted average pseudorapidity. Distances are measured here by counting crystals. In this way, the differences in the size of the gaps between the crystals, especially at the module boundaries, are not taken into account.
Figure 3.34: Distribution of the shower-shape variable, $\sigma_{\eta\eta}$, for FSR photons in $Z \rightarrow \mu^+\mu^-\gamma$ events in data (solid circles) and simulation (histogram), and for background-dominated photon candidates in dimuon triggered events (open circles). The barrel and endcaps are shown separately. The simulated signal and background distributions are normalized to the number of signal photons in the data. The ratios between the photon signal distributions in data and simulation are shown in the bottom panels. [63]

Figure 3.34 shows how signal can be separated from background using this variable. The photon candidates are selected by requiring a $p_T$ of at least 20 GeV and a hadronic fraction of the energy deposit of less than 5%. The matching between the shape of the $\sigma_{\eta\eta}$ distribution in data and simulation is not perfect, especially in the barrel region. Corrections are derived from $Z \rightarrow \mu^+\mu^-\gamma$ events to improve the data/MC agreement.

Other variables of frequent use for photon identification are:

- $q_{\eta\phi}$: the diagonal component of the covariance matrix constructed from the energy-weighted crystal positions within the $5 \times 5$ crystal array centered on the crystal containing the largest energy.

- the ratio $E_{2 \times 2}/E_{5 \times 5}$, where $E_{2 \times 2}$ is the maximum energy sum collected in a $2 \times 2$ crystal array that includes the largest energy crystal in the supercluster, and $E_{5 \times 5}$ is the energy collected in a $5 \times 5$ crystal matrix centered around the same crystal.

- the preshower variable $\sigma_{RR} = \sqrt{\sigma_{xx}^2 + \sigma_{yy}^2}$, where $\sigma_{xx}$ and $\sigma_{yy}$ measure the lateral spread in the two orthogonal sensor planes of that detector.

- $H/E$: the ratio between the energy collected by the HCAL towers behind the supercluster and the energy of the supercluster, describing the hadronic leakage of the shower in the HCAL.

- $\sigma_{\eta}$, $\sigma_{\phi}$ and $R_\theta$, already introduced previously.
**Isolation variables**

The photon isolation is best measured using the particle-flow approach to event reconstruction [70, 71]. The particle-flow algorithm combines information from the tracker, the calorimeters, and the muon detectors, and aims to reconstruct the four-momenta of all particles in the event. Particles are classified as charged and neutral hadrons, photons, electrons and muons by a series of identification requirements. In this way, the output of the reconstruction is a unique interpretation of the event in terms of particle candidates.

Isolation sum variables are obtained by summing the transverse momenta of charged hadrons, $I_\pi$, photons, $I_\gamma$, and neutral hadrons, $I_n$, inside an isolation region of radius $\Delta R$ in the $(\eta, \phi)$ plane around the photon direction.

An important step for calculating the photon isolation is removing the energy deposit due the photon itself from the isolation cone. This is done by applying geometrical requirements on the particle candidates entering the isolation sum. When calculating $I_\gamma$, the particle-flow photons falling in a pseudorapidity slice of size $\Delta \eta = 0.015$ are excluded from the sum. For the charged isolation $I_\pi$, the charged hadrons in the region $\Delta R < 0.02$ are equally excluded.

A novel, more refined cleaning algorithm has been developed in the context of this thesis and is described in Section 3.4. This alternative method improves the effectiveness of the cleaning for analyses that are particularly sensitive to the shape of the isolation distribution, such as the measurement of the isolated diphoton cross section.

For charged hadrons, a clear association between the candidate and a reconstructed primary vertex exists by means of the reconstructed track. Therefore, $I_\pi$ is calculated using only charged hadrons coming from a chosen vertex in the event. This fact can be problematic when using the charged isolation to discriminate prompt photons from jets. In events with two photons having balanced momenta in the transverse plane, for instance, the true interaction vertex is often characterized by a little amount of hadronic activity. In this situation, the identification of the true vertex can be inefficient. If a wrong vertex is chosen among the reconstructed ones, $I_\pi$ in background events will be consistent with the one of an isolated photon. For this reason, an alternative sum usually denoted as “worst charged isolation” ($I_{\pi}^{\text{max}}$) is used. $I_{\pi}^{\text{max}}$ is defined as the largest $I_\pi$ among those calculated for all reconstructed primary vertices.

The charged-hadron component of the isolation $I_\pi$ is independent on the number of pileup events as shown in the left plot of Fig. 3.35. This plot is populated with photons from $\gamma + \text{jet}$ events with a recoil transverse momentum of at least 50 GeV. This selection ensures a high probability of correct vertex choice. The neutral and photon components of the isolation, on the other hand, need to be corrected to remove the pileup contribution.

The contribution from pileup energy in the isolation region is estimated as $\rho \times A_{\text{eff}}$, where $\rho$ is the median of the transverse energy density per unit area in the event [72] and $A_{\text{eff}}$ is the area of the isolation region weighted by a factor that takes into account the dependence of the pileup transverse energy density on pseudorapidity. If the estimated pileup contribution is larger than the isolation sum, the isolation sum is set to zero.

When the extra contribution due to pileup, calculated using $\rho$, is subtracted from the photon and neutral hadron sums, their dependence on the number of vertices is significantly reduced (Fig. 3.35, right).
Figure 3.35: Mean value of the isolation variables for photons with $p_T > 50$ GeV in $\gamma + \text{jet}$ events, as a function of the number of reconstructed primary vertices, for events (left) before and (right) after being corrected for pileup using the $\rho$ variable. [63]

Figure 3.36 illustrates the distribution of the three isolation components for signal and background for barrel photons. The signal photons shown are from $Z \rightarrow \mu^+\mu^-\gamma$ events in data and simulation, while the background-dominated candidates are obtained from dimuon triggered events in data, as in Fig. 3.34. Data and simulation are in good agreement using an isolation cone with radius $\Delta R = 0.3$. Similar results are found for the endcap region.

### 3.3.3 Selection requirements

The set of selection requirements most commonly used in photon analyses acts on $\sigma_{\eta\eta}$, $H/E$, and the isolation sums. In most cases, the isolation thresholds are expressed as a constant term added to a term proportional to the candidate photon transverse momentum. A set of standard identification requirements is presented in Table 3.2 for three different working points. The working points correspond to increasing prompt photon efficiency targets, at the price of increasing background contamination. Figure 3.37 shows the signal efficiency as a function of the candidate’s $\eta$, $E_T$ and the number of pileup interactions.

The efficiency for prompt photons to pass the identification working points is measured with a tag-and-probe [73] method using $Z \rightarrow e^+e^-$ events. This measurement is used to correct the efficiency in simulation. Electrons from $Z$ decays are triggered in data requiring an electron with $p_T$ larger than 27 GeV in the High-Level-Trigger. The “tag” candidates are required to have $p_T > 30$ GeV and satisfy tight electron identification requirements [62]. The dielectron invariant mass is required to be in the range $60 < m_{ee} < 120$ GeV. The “probe” candidates are electron showers reconstructed as photons. The minimum $p_T$ of the probe is 15 GeV. The efficiency for the probe to pass the identification criteria, with the exception of the electron veto, is studied as a function of $p_T$ and separately for the barrel and endcap regions. The efficiency for passing the elec-
Figure 3.36: Distributions of the isolation variables: (top) $I_γ$, (bottom left) $I_π$, and (bottom right) $I_n$, constructed from particle-flow objects. The distributions are shown for FSR photons from $Z \rightarrow \mu^+ \mu^- \gamma$ events in data (solid circles) and simulation (histogram) and for background-dominated photon candidates in dimuon triggered events (open circles). The simulated signal and background distributions are normalized to the number of signal photons in data. The ratios between the photon signal distributions in data and simulation are shown in the bottom panels. [63]
Table 3.2: Photon identification requirements for three working points corresponding to selections of different tightness. [63]

<table>
<thead>
<tr>
<th></th>
<th>Loose</th>
<th>Medium</th>
<th>Tight</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_{\gamma}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>barrel</td>
<td>1.3 GeV + 0.005 $p_T^\gamma$</td>
<td>0.7 GeV + 0.005 $p_T^\gamma$</td>
<td>0.7 GeV + 0.005 $p_T^\gamma$</td>
</tr>
<tr>
<td>endcap</td>
<td>—</td>
<td>1.0 GeV + 0.005 $p_T^\gamma$</td>
<td>1.0 GeV + 0.005 $p_T^\gamma$</td>
</tr>
<tr>
<td>$I_{\pi}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>barrel</td>
<td>3.5 GeV + 0.04 $p_T^\pi$</td>
<td>1.0 GeV + 0.04 $p_T^\pi$</td>
<td>0.4 GeV + 0.04 $p_T^\pi$</td>
</tr>
<tr>
<td>endcap</td>
<td>2.9 GeV + 0.04 $p_T^\pi$</td>
<td>1.5 GeV + 0.04 $p_T^\pi$</td>
<td>1.5 GeV + 0.04 $p_T^\pi$</td>
</tr>
<tr>
<td>$\sigma_{\eta\eta}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>barrel</td>
<td>0.012</td>
<td>0.011</td>
<td>0.011</td>
</tr>
<tr>
<td>endcap</td>
<td>0.034</td>
<td>0.033</td>
<td>0.031</td>
</tr>
<tr>
<td>$H/E$</td>
<td></td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Electron veto</td>
<td>conversion-safe</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

...tron veto cannot be studied with electrons, and is measured instead using $Z \to \mu^+\mu^-\gamma$ events.

The invariant mass distribution is built separately for the cases in which the probe satisfies or not the identification requirements. The distributions are then fitted to extract the identification efficiency. The di-electron signal is modeled with a template extracted from simulated $Z \to e^+e^-$ events and convoluted with a Gaussian function to account for the slightly worse resolution in data with respect to the simulation. The background is modeled with a template extracted from simulated $Z \to \mu^+\mu^-$ events and convoluted with a Gaussian function to account for the slightly worse resolution in data with respect to the simulation. The background is modeled with an exponential function times a Gauss error function. In this way, the shaping of the distribution induced by the minimum $p_T$ requirements can be correctly described.

Figure 3.38 shows an example of the fit for the central barrel region. The left tail of the failing probes distribution is populated by radiating electrons for which a fraction of energy is not clustered. Figure 3.39 illustrates the selection efficiency in data and simulation for the medium working point. The ratio shows a good agreement for $p_T > 20$ GeV. The systematic uncertainty, represented by the shaded band in the plot, is evaluated by replacing the functional form used to model the background with simple exponential and polynomial functions. The statistical uncertainty is negligible.

### 3.4 Removal of the photon footprint from the isolation sum

The shower energy deposit (“footprint”) can be removed from the isolation cone by excluding the central part of the cone from the calculation. This Section describes a novel procedure to perform this task more effectively [65]. The method can be applied to photons or electrons. Its performance is validated in data. Moreover, the discrimination power of the isolation variable is increased with respect to what has been presented in the previous Section for endcap photons. The algorithm is based on purely geometrical considerations, and was designed to use a minimal amount of information about the reconstructed particle-flow candidates. In this...
Figure 3.37: Efficiency of photon identification based on sequential requirements in simulated $\gamma +$ jet events for three different working points, as a function of the (top) photon pseudorapidity, (middle) number of pileup vertices, and (bottom) photon transverse momentum. The efficiencies shown include the electron veto requirement. A minimum $p_T$ threshold of 15 GeV is applied. [63]
Figure 3.38: Study of the efficiency of the medium working point detailed in Table 3.2. Example of fits to the $Z \rightarrow e^+e^-$ invariant mass distribution for (left) passing and (right) failing probes, in the transverse momentum range $20 < p_T < 30$ GeV and $|\eta| < 0.8$. [63]

Figure 3.39: Comparison of the selection efficiency as a function of photon transverse momentum in data (circles) and simulation (triangles) for the identification based on sequential requirements for (left) $|\eta| < 0.8$ and (right) $1.6 < |\eta| < 2$. Statistical and systematic uncertainties are respectively shown by the error bars and shaded bands. The horizontal error bars mark the full width of the $p_T$ bins in which the measurements are made, and the data points are plotted at the center of each bin. The ratios of efficiencies in data and simulation are shown in the bottom panels. [63]
way, it can be applied at analysis level on datasets with a reduced amount of stored event information, saving computational power.

3.4.1 Energy leakage cleaning based on veto-cones

The isolation energy of an electromagnetic object (photon or electron) is properly defined only if the bias coming from the leakage of the electromagnetic shower inside the isolation cone is removed. Moreover, very low energy deposits typically coming from noise also have to be rejected.

This is a general issue valid for both the detector-based isolation (computed as the sum of the track transverse momenta, or the sum of the RecHit transverse energies in the ECAL and HCAL within an isolation cone around the supercluster) and Particle-Flow-based isolation (computed as the sum of transverse momenta of reconstructed PF candidates). The approach followed by detector-based isolation consists in applying low energy thresholds on the RecHits and tracks in the isolation cone, and to remove the energy of the core of the supercluster by vetoing the energy deposits in a small cone around it ("veto-cone"). In the case of the ECAL isolation, a $\Delta \eta$ strip requirement is also introduced to suppress the energy deposited by photon conversions with a spread along the $\phi$ direction.

The particle-flow isolation has the advantage to be more noise-resistant, due to the relatively high thresholds applied on the RecHit, cluster and particle candidate energies. Actually, PF clusters are reconstructed separately from the main reconstruction workflow described in Section 3.1. They are built aggregating all contiguous crystals with energies of more than two standard deviations above the electronic noise level. The aggregation is performed around local maxima, used as seed crystals. Seed crystals are required to have an energy of at least 230 MeV in the barrel, and an energy of at least 600 MeV and a transverse energy of at least 150 MeV in the endcaps. An important difference relative to the general workflow is that it is possible to share the energy of one crystal among two or more clusters.

A cleaning of the supercluster energy leakage inside the isolation cone is applied by default to the high $E_T$ photons and electrons satisfying tight identification criteria. In this case, particle-flow candidates are removed from the isolation sums if they share a fraction of their energy with the photon/electron supercluster. This procedure guarantees full cleaning of the shower leakage, as long as the clustering algorithm is able to collect the whole shower energy.

The main limitation of this procedure is the fact that high-$E_T$ photons or electrons often fail to satisfy the identification criteria and undergo the cleaning, as shown in Fig. 3.40. The inefficiency for photons has evolved through different versions of the reconstruction software (about 40% in 2011, 20% in 2012), and is explained by two characteristics of the PF reconstruction.

First, PF gives priority to electrons over photons during the disambiguation needed to provide an unique interpretation of the events. In this way, converted photons can be reconstructed as electrons, especially in the endcaps. In the second place, the supercluster often fails to satisfy the tight requirements on isolation variables required to be identified as a PF photon.
In any of these cases, the result is an isolation cone where part of the core shower energy enters the isolation sum. Figures 3.41-3.46 show the \((\Delta \eta, \Delta \phi)\) distribution of the relative distance between the PF candidates and the supercluster position. The distributions for charged-hadron, neutral-hadron and photon PF candidates are shown separately for the barrel and endcap regions. The PF candidate position is the one obtained by extrapolating its trajectory to the ECAL surface. The extrapolation is necessary because the relevant position is the one of the ECAL impact point measured in the detector reference system. The default PF candidate orientation is instead given with respect to the reconstructed interaction vertex.

A very clear difference is observed between the identified and not-identified photon superclusters. In the case of cleaned photons, the PF candidates overlapping with the cluster are effectively removed and a deficit of residual candidates in the region covered by the cluster is observed. For non-identified photons, on the other hand, a concentration of candidates is observed in correspondence with the cluster center, with a spread in the \(\phi\) direction.

Veto-cones have been developed in analogy to what is done for detector-based isolation. Their definition is summarized in Table 3.3 for electrons and Table 3.4 for photons. The dependence of the veto-cone size in the endcaps aims at adapting the tightness of this requirement to the varying \(\eta\) range covered by each crystal at different pseudorapidities.
Figure 3.41: 2D $(\Delta \eta, \Delta \phi)$ profile between the center of the photon supercluster and the photon PF candidates at the surface of ECAL in the barrel, for photons identified by the particle-flow (left) and for photons not identified by the particle-flow (right). [65]

Figure 3.42: 2D $(\Delta \eta, \Delta \phi)$ profile between the center of the photon supercluster and the photon PF candidates at the surface of ECAL in the endcap, for photons identified by the particle-flow (left) and for photons not identified by the particle-flow (right). [65]

Table 3.4: Photon isolation veto-cones

<table>
<thead>
<tr>
<th></th>
<th>photons</th>
<th>neutral hadrons</th>
<th>charged hadrons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barrel</td>
<td>$\Delta \eta &gt; 0.015$</td>
<td>-</td>
<td>$\Delta R &gt; 0.02$</td>
</tr>
<tr>
<td>Endcap</td>
<td>$\Delta R &gt; 0.00864 \times</td>
<td>\sinh(\eta_{SC})</td>
<td>\times 4$</td>
</tr>
</tbody>
</table>
Figure 3.43: 2D \((\Delta \eta, \Delta \phi)\) profile between the center of the photon supercluster and the neutral hadron PF candidates at the surface of ECAL in the barrel, for photons identified by the particle-flow (left) and for photons not identified by the particle-flow (right). [65]

Figure 3.44: 2D \((\Delta \eta, \Delta \phi)\) profile between the center of the photon supercluster and the neutral hadron PF candidates at the surface of ECAL in the endcap, for photons identified by the particle-flow (left) and for photons not identified by the particle-flow (right). [65]
Figure 3.45: 2D $(\Delta \eta, \Delta \phi)$ profile between the center of the photon supercluster and the charged hadron PF candidates at the surface of ECAL in the barrel, for photons identified by the particle-flow (left) and for photons not identified by the particle-flow (right). [65]

Figure 3.46: 2D $(\Delta \eta, \Delta \phi)$ profile between the center of the photon supercluster and the charged hadron PF candidates at the surface of ECAL in the endcap, for photons identified by the particle-flow (left) and for photons not identified by the particle-flow (right). [65]
In some cases, however, the veto-cones lead to an under-cleaning (especially at large $\Delta \eta$ or $\Delta \phi$ from the center of the supercluster) or over-cleaning (for example, in the case of unconverted photons) of the leakage. In order to address these issues, the novel method presented in the following Section takes into account the supercluster size on an event-by-event basis. It mimics the procedure used by the PF reconstruction for identified photons or electrons, but can be applied to every supercluster.

### 3.4.2 Supercluster footprint removal algorithm

This algorithm is based on purely geometrical considerations:

- Extrapolate the PF candidate trajectory from its starting point to the surface of the ECAL. For the charged hadrons, the effect of the magnetic field is taken into account. Only the distance $dz$ (along the $z$ direction) with respect to the primary vertex is considered, while the transverse distance $dxy$ is neglected.

- If the extrapolated candidate hits the surface of a crystal that belongs to the supercluster, the candidate is removed from the isolation sum. A tolerance of 25% of the crystal size is allowed, to account for the fact that the PF candidate energy deposit has a finite extension in the ECAL. This has a sizable effect at the edges of the supercluster.

In this way, the supercluster shape defines the region excluded from the isolation cone in each event.

The performance is verified using simulated samples of $H \to \gamma\gamma$ (for signal) and $\gamma + \text{jet}$ (for background) at $\sqrt{s} = 7$ TeV. Figures 3.47-3.51 illustrate, in analogy with what has been presented before, the distribution of $\Delta R$ and $\Delta \eta$ between the PF candidates and the supercluster position. The distribution of the ratio of transverse momenta of each PF candidate entering the isolation sum and the supercluster is illustrated in Figures 3.52-3.54.

The plots for the charged component show that the leakage along the $\phi$ direction due to converted photons is successfully cleaned. In the barrel, deposits inside the isolation cone with $|\Delta \phi| > 0.28$ are not cleaned, because they lie outside of the supercluster (recall here that the $\phi$ road of the barrel “hybrid” reconstruction algorithm is $\pm 17$ crystals around the supercluster seed).

The behaviour of the neutral component of PF isolation is also improved. Photons hitting the cracks in the ECAL leave very little energy in it, but create a cluster in the HCAL. The PF algorithm reconstructs them as neutral hadrons, and given their lack of ECAL hits they cannot share any RecHit with photon or electron superclusters. The new procedure is more effective in removing them, because it is based on a purely geometric matching.

At high values of $\Delta R$, the distribution of all components is expected to grow linearly after cleaning because each bin corresponds to a thin ring in $\Delta R$, the area of each ring increases linearly with $R$ and the expected distribution of PF candidates coming from pile-up and the underlying event is flat.

The peak present at $\Delta R \approx 0$ for the photon component is due to the photon PF candidates that are duplicates of the supercluster. The same events populate the peak around
Figure 3.47: $\Delta R$ profile between the center of the photon supercluster and the charged hadrons PF candidates at the surface of ECAL for photons not identified by the particle-flow, barrel (left) and endcap (right). Different veto-cones are compared. The $y$ axis is in arbitrary units. [65]

Figure 3.48: $\Delta R$ profile between the center of the photon supercluster and the neutral hadrons PF candidates at the surface of ECAL for photons not identified by the particle-flow, barrel (left) and endcap (right). The $y$ axis is in arbitrary units. [65]
Figure 3.49: $\Delta R$ profile between the center of the photon supercluster and the photon PF candidates at the surface of ECAL for photons not identified by the particle-flow, barrel (left) and endcap (right). Different veto-cones are compared. The $y$ axis is in arbitrary units. [65]

Figure 3.50: $\Delta \eta$ profile between the center of the photon supercluster and the charged hadrons PF candidates at the surface of ECAL for photons not identified by the particle-flow, barrel (left) and endcap (right). Different veto-cones are compared. The $y$ axis is in arbitrary units. [65]
Figure 3.51: $\Delta \eta$ profile between the center of the photon supercluster and the photon PF candidates at the surface of ECAL for photons not identified by the particle-flow, barrel (left) and endcap (right). Different veto-cones are compared. The $y$ axis is in arbitrary units. [65]

Figure 3.52: Distribution of the transverse momentum, $p_T$, of charged hadron PF candidates, divided by the supercluster $E_T$ for photons not identified by the particle-flow, barrel (left) and endcap (right). Different veto-cones are compared. The $y$ axis is in arbitrary units. [65]
Figure 3.53: Distribution of the transverse momentum, $p_T$, of neutral hadron PF candidates, divided by the supercluster $E_T$ for photons not identified by the particle-flow, barrel (left) and endcap (right). Different veto-cones are compared. The $y$ axis is in arbitrary units. [65]

Figure 3.54: Distribution of the transverse momentum, $p_T$, of photon PF candidates, divided by the supercluster $E_T$ for photons not identified by the particle-flow, barrel (left) and endcap (right). Different veto-cones are compared. The $y$ axis is in arbitrary units. [65]
1 in the plot of the ratio of PF candidate and supercluster $p_T$. This contribution is suppressed by both the standard veto-cones and the new procedure. The standard veto-cones, however, show an under-cleaning in the region around $\Delta R$ close to 0.25 in the endcap, as well as an over-cleaning in the region around $\Delta \eta < 0.015$ in the barrel. The new procedure is shown to perform correctly everywhere.

The simulated performance of the new procedure in terms of signal and background efficiency is shown in Fig. 3.55, and compared to that of the old cleaning procedure using veto-cones. Photons with $p_T > 25$ GeV from $H \rightarrow \gamma \gamma$ are considered here as signal if matched to generator level photons, while reconstructed candidates in a $\gamma$ + jet sample are considered as background if not matched to a generator level photon. An isolation cone radius of $\Delta R = 0.4$ is used.

The new method shows a 4% improvement in background rejection at 90% of signal efficiency in the endcaps. In any case, it is essential to note that the more efficient cleaning will be crucial for building the isolation templates used for the measurement presented in this thesis. A detailed description of the template construction will be given in Chapter 4.

In the newer reconstruction software envisaged for the LHC Run 2 data taking, the superclusters described in Section 3.1 will be replaced by superclusters based on PF clusters. In this way, the association between PF candidates and supercluster components will be straightforward. A cleaning strategy analogous to the footprint removal, and based on PF cluster sharing, will be applied by default on all superclusters.
Chapter 4

Measurement of the diphoton production cross section

This chapter presents the measurement of differential cross sections for the production of an isolated photon pair in proton-proton collisions. A fit based on isolation templates is used to statistically separate the signal from the background. The analysis is documented in Refs. [74, 75]. The discussions, text and illustrations which can be found in those references are closely followed here. After a description of the event selection, the procedures used to build the templates from data are presented. Then, the fit is described and the sources of systematic uncertainty are discussed. Finally, the results are compared to theoretical predictions for QCD diphoton production.

4.1 Data samples and event selection

The data sample used for this analysis consists of proton-proton (pp) collision events collected at the LHC with the CMS detector in the year 2011, at a center-of-mass energy ($\sqrt{s}$) of 7 TeV and corresponding to an integrated luminosity of 5.0 fb$^{-1}$.

4.1.1 Event trigger

The event readout is triggered by the presence of two photon candidates with asymmetric transverse energy thresholds. The trigger system is organized in two stages: a Level-1 (L1) trigger filter performs a first reduction of the event rate, and is followed by a High-Level-Trigger (HLT) where a more refined set of requirements (denoted as “trigger path” in the following) can be applied.

The $E_T$ thresholds at trigger level are 26 (18) and 36 (22) GeV on the leading (sub-leading) photon, depending on the running period. Each candidate is required to satisfy either loose calorimetric identification requirements, based on the shape of the electromagnetic shower, or loose isolation conditions.

4.1.2 MC samples

Several samples (Table 4.1) of simulated events are used in the analysis to model signal and background processes. Drell-Yan+jets and $\gamma\gamma$+jets signal events are generated with MADGRAPH 1.4.8 [40]. The $gg \rightarrow \gamma\gamma$ box signal process, $\gamma$+jet, and QCD dijet background processes are generated with PYTHIA 6.4.24 [39].
Table 4.1: List of Monte Carlo simulation samples. The cross section used in the generation of each process, and an inclusive $k$-factor used to improve the data/MC agreement, are also given. A generator-level filter is applied on the $\gamma$+jet and QCD processes to select events where the jet contains a boosted neutral meson.

<table>
<thead>
<tr>
<th>Simulated process</th>
<th>Cross section (pb)</th>
<th>$k$-factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma\gamma$ + jets</td>
<td>336.0</td>
<td>1.15</td>
</tr>
<tr>
<td>$\gamma\gamma$, box diagram</td>
<td>12.37</td>
<td>1.15</td>
</tr>
<tr>
<td>$\gamma$+ jet enriched in e.m. component, $p_T &gt; 20$ GeV</td>
<td>493.4</td>
<td>1.3</td>
</tr>
<tr>
<td>QCD enriched in e.m. component, $p_T &gt; 30$ GeV</td>
<td>54.441</td>
<td>1.3($\gamma$), 1.0 (jj)</td>
</tr>
<tr>
<td>Drell-Yan dilepton + jets</td>
<td>30.48</td>
<td>1.0</td>
</tr>
</tbody>
</table>

For all simulated samples the CTEQ6L1 [76] parton distribution functions (PDFs) are used. All generated events are then processed with PYTHIA (Z2 tune) [77] for hadronization, showering of partons and the underlying event; a detailed simulation of the CMS detector based on GEANT 4 [78] is performed, and the simulated events are finally reconstructed using the same algorithms as used for the data. The normalization of each simulated sample is adjusted by a multiplicative factor (indicated as $k$-factor in Table 4.1) to improve the data/MC agreement ([79], App. B).

The simulation includes the effects of in-time pileup (overlapping pp interactions within a bunch crossing) and out-of-time pileup (overlapping pp interactions from interactions happening in the previous or following bunch crossing). The number of in-time and out-of-time pile-up events are generated with the same distribution. In order to better account for the correlation between the number of in-time and out-of-time pileup in data, a two-step procedure is used to sample the pileup distribution. First, a value is sampled from a distribution corresponding to the per-bunch-crossing luminosity. Then, the number of in-time and out-of-time simulated interactions is sampled from a Poissonian distribution having the previously sampled value as mean.

All samples are re-weighted to match the pileup distribution observed in data. The target pileup distribution is derived using the values of the instantaneous luminosity recorded by the Pixel Luminosity Telescope for every luminosity section and the total pp inelastic cross section (68.0 mb), and is averaged over the entire data-taking period. The pileup reweighting technique is validated comparing the number of reconstructed interaction vertices in data and simulation. The result of such a comparison is shown in Fig. 4.1.

4.1.3 Photon selection

The photon candidates are first required to pass a sequence of filters that aim to remove beam backgrounds or identified detector issues and to satisfy more stringent criteria than the trigger requirements, to reduce the effect of the trigger turn-on on the selection efficiency.

A preselection is then applied. It is based on the shape of the electromagnetic shower in the ECAL and on the degree of isolation of the photon (i.e. the amount of energy deposited in the vicinity of the photon). The variables used are:

- Photon supercluster raw energy $E_{SC}^{raw}$: the sum of the calibrated crystal energies;
Figure 4.1: Distribution of the number of reconstructed vertices for events passing the diphoton preselection in data (black points) and MC (red histogram) after applying the pileup reweighting procedure. [75]

- Preshower energy $E_{ES}^{SC}$: the sum of the energy deposits reconstructed in the preshower detector (ES) and associated with the supercluster;

- $R_9$: the energy sum of $3 \times 3$ crystals centered on the most energetic crystal in the supercluster divided by the raw energy of the supercluster;

- H/E: the ratio of the energy deposited in HCAL that is inside a cone of size $\Delta R = \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2} = 0.15$ centered on the photon direction, to the supercluster energy;

- $\sigma_{\eta\eta}$: the shower transverse extension along $\eta$ defined in Section 3.3;

- $\text{Iso}_{0.3}^{ECAL}$ (ECAL isolation): the scalar sum of the $E_T$ of the deposits in the electromagnetic calorimeter lying inside a cone of size $\Delta R = 0.3$, centered on the direction of the supercluster but excluding an inner cone of size 3.5 crystals and an $\eta$-slice region of 2.5 crystals;

- $\text{Iso}_{0.3}^{HCAL}$ (hadronic calorimeter isolation): the scalar sum of the $E_T$ of the deposits in the hadron calorimeter that lie inside a hollow cone of outer radius of size $\Delta R = 0.3$ and inner radius of size $\Delta R = 0.15$ in the $\eta$-$\phi$ plane, centered on the direction of the supercluster;

- $\text{Iso}_{0.3}^{TRK}$ (tracker isolation): the scalar sum of the $p_T$ of the tracks that are consistent with originating from the primary vertex in the event, and lie inside a hollow cone of outer radius of size $\Delta R = 0.3$ and inner radius of size $\Delta R = 0.04$ in the $\eta$-$\phi$ plane, centered around a line connecting the primary vertex with the supercluster but excluding an $\eta$-slice region ($\Delta \eta = 0.015$).

The shower shape variables in the simulation are corrected to compensate for their imperfect modeling, using factors extracted from a sample of photons in $Z \rightarrow \mu^+ \mu^- \gamma$ events. The list of preselection criteria is presented in Table 4.2. The isolation requirements are kept loose because the isolation is used as the discriminating variable in the
Table 4.2: List of requirements that a candidate has to satisfy to pass the analysis preselection.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photon raw + preshower energy</td>
<td>$E^\text{raw}<em>{SC} + E^\text{ES}</em>{SC} &gt; 20 \text{ GeV}$</td>
</tr>
</tbody>
</table>
| H/E | if $(R_9 > 0.9)$: $H/E < 0.082$ (EB), $0.075$ (EE)  
| | if $(R_9 < 0.9)$: $H/E < 0.075$ |
| $\sigma_{\eta\eta}$ | $0.001 < \sigma_{\eta\eta} < 0.014$ (EB), $0.034$ (EE) |
| ECAL isolation in a $\Delta R=0.3$ cone | $\text{Iso}_{\text{ECAL}}^0 < 4$ GeV (only if $R_9 < 0.9$) |
| HCAL isolation in a $\Delta R=0.3$ cone | $\text{Iso}_{\text{HCAL}}^0 < 4$ GeV (only if $R_9 < 0.9$) |
| TRK isolation in a $\Delta R=0.3$ cone | $\text{Iso}_{\text{TRK}}^0 < 4$ GeV (only if $R_9 < 0.9$) |

Table 4.3: List of additional requirements applied in the photon candidate selection.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matched pixel measurements</td>
<td>False</td>
</tr>
<tr>
<td>H/E</td>
<td>$H/E &lt; 0.05$</td>
</tr>
<tr>
<td>$\sigma_{\eta\eta}$</td>
<td>$\sigma_{\eta\eta} &lt; 0.011$ (EB), $0.030$ (EE)</td>
</tr>
</tbody>
</table>

signal extraction procedure.

The preselected photons must satisfy additional requirements to be considered as photon candidates. These consist of the absence of reconstructed electron track seeds in the pixel detector which match the candidate’s direction (“pixel seed veto”), and a tighter selection on the hadronic leakage of the shower and the $\sigma_{\eta\eta}$ shower shape variable. The list of additional selection criteria is shown in Table 4.3.

In the simulation, prompt photons are defined as candidates satisfying the analysis selection requirements and geometrically matched to an isolated generator-level photon, defined by the following requirements:

- the generated photon is a final-state particle (status-1 in the generator);
- the reconstructed and generated photons are geometrically matched (the distance in the $\eta - \phi$ plane between them is required to be less than 0.1), and have compatible transverse energy (within a factor of 2);
- the mother particle of the generated photon is either a quark or gluon (fragmentation photon), or a partonic photon (status-3 in the generator, i.e. from the hard scattering);
• the cone isolation around the generated photon (calculated as the scalar sum of transverse energies of all stable generated particles in a cone of \( \Delta R = 0.4 \)) does not exceed 5 GeV.

4.1.4 Diphoton selection

Events are retained only if at least two photon candidates satisfying the requirements described above are found. If more than two photon candidates are selected, the two with highest transverse energy \((E_T)\) are retained. A transverse energy greater than 40 (25) GeV for the leading (sub-leading) photon is required, together with a separation of \(\Delta R > 0.45\) between the two candidates. The minimum separation requirement ensures that the energy deposit of one photon does not enter the isolation cone centered on the other one.

The signal fraction is statistically separated from jets mis-identified as photons by means of a binned maximum likelihood fit that uses the photon component of the PF isolation as the discriminating variable. The procedure for building the templates used in the fit is described in the next Section.

4.2 Isolation templates

The diphoton signal is extracted through a two-dimensional binned maximum likelihood fit that uses the isolation of the two selected photon candidates as discriminating variables. Different templates are built for the prompt-prompt \((f_{pp})\), prompt-non-prompt \((f_{pn})\), non-prompt-prompt \((f_{np})\), and non-prompt-non-prompt \((f_{nn})\) components in the \((Iso_1, Iso_2)\) plane, where \(Iso_1\) and \(Iso_2\) represent the isolation variables for the two selected photon candidates in the event.

The pile-up introduces a spurious correlation between the two candidate photons’ isolation sums. For this reason, the PF isolation sums for both photons are corrected, event by event, for the presence of pile-up with a factor proportional to the average pile-up energy density \((\rho)\) calculated with FASTJET [72]. This procedure is not sufficient to completely decorrelate the isolation sums, because it cannot remove the local fluctuations of the pile-up energy density within each event.

Moreover, the shape of the isolation distribution depends on the photon’s position in the detector. Therefore, kinematic correlations between the two photons are propagated into a correlation between their isolation sums.

In summary, a factorized probability distribution function of the following form:

\[
P_{1D \times 1D}(Iso_1, Iso_2) = f_{pp} \cdot T_{pp}(Iso_1, Iso_2) + f_{pn} \cdot T_{pn}(Iso_1, Iso_2) + f_{np} \cdot T_{np}(Iso_1, Iso_2) + f_{nn} \cdot T_{nn}(Iso_1, Iso_2)
\] (4.1)

where \(T_{pp}(Iso)\) are the functions describing the isolation distribution (template) for prompt and non-prompt photons respectively, cannot describe the data in a satisfactory way.

The correlation has to be taken into account in the probability distribution, as follows:

\[
P_{2D}(Iso_1, Iso_2) = f_{pp} \cdot T_{pp}(Iso_1, Iso_2) + f_{pn} \cdot T_{pn}(Iso_1, Iso_2) + f_{np} \cdot T_{np}(Iso_1, Iso_2) + f_{nn} \cdot T_{nn}(Iso_1, Iso_2)
\] (4.2)
where $T_{kk}(\text{Iso}_1, \text{Iso}_2)$ is the bi-dimensional template for the component $f_{kk}$.

Two-dimensional templates are extracted from data to avoid possible biases coming from an imperfect modeling of the events in the simulation, mainly related to fine detector effects such as the electronic noise in the ECAL readout. These imperfections affect the softest part of the isolation distribution, which is also the most discriminating one against non-prompt photons. The residual differences in the simulation between the isolation distribution and the templates defined, as it will be detailed below, with the random cone and the sideband techniques are accounted for as systematic uncertainties in the template shapes.

Samples of events where at least one photon passes the photon selection are used to create prompt-prompt, prompt-non-prompt, non-prompt-prompt and non-prompt-non-prompt templates with high statistical precision. The following Sections describe how this is done. The basic methods used for building one-dimensional prompt and non-prompt templates will be presented first. Then, the procedures used to model the correlation between the isolation sums will be described.

### 4.2.1 Signal template with the random cone method

The prompt photon template is built with the "random cone" technique. This method relies on the assumption that a prompt photon is isolated. In this case, the activity recorded in the detector around it (excluding the region affected by its own energy deposit in the calorimeter, see Section 3.4) can only come from pile-up, the underlying event and ECAL electronic noise.

These contributions do not vary - on average - if the isolation sum is calculated in a region separated from the candidate photon, at the same pseudorapidity (Fig. 4.2). The following steps lead to the selection of a suitable orientation of the isolation cone:
• Consider a direction (named “random-cone axis” in the following) obtained from the photon direction by rotating around the $z$ axis of the detector reference system (e.g. rotating in the $\phi$ direction while keeping $\eta$ unchanged) by a random angle $\phi_{R\text{Cone}}$ between 0.8 and $2\pi - 0.8$ radians. In this way, an isolation cone of $\Delta R = 0.4$ centered on the random-cone axis would not overlap with another isolation cone centered on the photon, or with an object produced back-to-back with respect to it.

• Check that neither a AK5\footnote{Anti-$k_T$ jet clustering algorithm, defined in Section 1.3.3, with a size parameter $R = 0.5$.} Particle-Flow jet with a $p_T$ of at least 20 GeV at $\Delta R < 0.8$, nor a photon with $p_T$ of at least 10 GeV at $\Delta R < 0.8$, nor a muon at $\Delta R < 0.4$ with respect to the random-cone axis are reconstructed in the event. No further requirement is applied on these objects. This check is performed to avoid choosing a region of the detector where a substantial energy flow is present which is not due to pile-up, underlying event or ECAL electronic noise. If such an object is found, another random angle $\phi_{R\text{Cone}}$ is generated as before, and the check repeated until a suitable direction is found. In this way, no jet or photon present in the event will overlap by chance with the random cone. Typically the procedure is not repeated more than two times before finding a suitable direction.

The isolation sum is calculated in a cone centered on the chosen direction. As in this new direction no SuperCluster exists along the cone axis, the footprint removal is performed in the following way: an area corresponding to the region covered by the photon SuperCluster, and rotated by the same $\phi_{R\text{Cone}}$ to be aligned with the random-cone axis, is excluded from the calculation. In this way, the area considered for the isolation calculation, as well as the area excluded in the central part of the cone, has exactly the same extension and shape both in the photon and in the random-cone direction.

The performance of the method is checked (Fig. 4.3) in simulated events by comparing the template obtained from the random-cone procedure with the one obtained by computing the isolation along the direction of a reconstructed photon which is matched to a prompt photon at generator level. Moreover, these MC templates are compared with the one built using the random-cone technique in data.

The isolation distribution depends on the pseudorapidity, because the conditions and the response of the detector are position-dependent. Therefore, the pseudorapidity is reweighted in order to equalize its distribution for the events used to build the different templates.

This study confirms that the footprint removal technique successfully removes the photon energy deposit from the isolation cone, as no large discrepancy exists between the distributions obtained from random cones and generated prompt photons.

Moreover, the simulation can qualitatively predict the shape of the isolation distribution for prompt photons in the barrel, while in the endcap a discrepancy between data and MC is observed. This discrepancy has the form of an excess mostly present in the softest part of the isolation distribution. The reason for this behaviour is investigated in the next paragraph.

**Effect of ECAL electronic noise**

The discrepancy in the signal template shape between data and simulation in the endcap region is even more evident if no subtraction of the pile-up energy is performed, as
Figure 4.3: Comparison of prompt photon templates in data and simulation: prompt photons in the simulation (squares), prompt photon templates extracted with the random cone technique from simulation (triangles) and from data (dots); Top: candidates in the ECAL barrel, Bottom: candidates in the ECAL endcaps. All histograms are normalized to unit area. [74, 75]
Figure 4.4: Distribution of the template variable before pileup subtraction in the endcap region. [75]

presented in Fig. 4.4. The disagreement takes the form of an excess of events just above the lower end of the distribution. In that region, the shape of the template is determined by energy and transverse momentum thresholds applied to ECAL energy deposits and clusters in the reconstruction of PFCandidates. Furthermore, an additional peak around 0.4 GeV is found in data only. This suggests the presence of an increased effective noise in the measurement of the ECAL crystal energy, which causes more PFCandidates to be reconstructed just above the minimum energy threshold in data than in the simulation.

Studies of measured pulse shapes in the ECAL endcaps show that the electronic noise in the readout channels does not change significantly with time if measured in units of ADC counts. Nevertheless, ECAL crystals lose transparency because of radiation, and the energy response is corrected back to a constant value in data by the ECAL calibration sequence. This is done by multiplying the pulse amplitude, measured in ADC counts, by a calibration factor that increases with time. Therefore, the effective value of the electronic noise, measured in units of energy, also grows as more integrated luminosity is collected.

No simulation of the crystal loss of transparency exists in the detector simulation used for this analysis. An approximate way of studying the effect of an increased effective noise in the simulation is to artificially increase the simulated electronic noise in units of ADC counts. The result of this procedure is reported in Fig. 4.5. With the default noise model the spectrum of reconstructed photon PFCandidates in the endcap region is qualitatively different between data and simulation. An increased simulated noise reduces the discrepancy. This observation confirms that the imperfect modeling of the ECAL effective noise is responsible for the difference in shape between the endcap signal templates in data and simulation. The origin of that discrepancy is therefore understood. It should be kept in mind that, in any case, the techniques used to build the templates used for the measurement rely only on data, and are therefore insensitive to the noise modeling in the simulation. The purpose of this study is to ensure that data do not present any feature that is not qualitatively modeled by the simulation, and thus possibly not understood.
Figure 4.5: Spectrum of photon PF Candidates in the endcap region in data and in MC with default and increased noise models. [75]

Effect of ECAL selective readout (zero-suppression)

Each ECAL channel is readout only if its measured energy reaches a certain value or if it is in proximity of other large energy deposits. This mechanism acts as a zero-suppression, rejecting energy measurements compatible with the electronic noise and helping to reduce the bandwidth of the data acquisition stream.

In the reconstruction of PF objects further energy thresholds are applied at different stages of clusters and PF Candidates formation. These requirements are tuned to be tighter than those applied by the online zero-suppression. In this way, the response of the detector is uniform across fully-readout and zero-suppressed regions of the ECAL.

The random-cone is built in an empty region of the detector, which has very likely been readout with zero-suppression applied. On the other hand, in the vicinity of the photon candidate the detector has been fully readout. Therefore, this analysis is potentially sensitive to an imperfect matching of the online and offline energy thresholds. This hypothesis has been tested in the simulation, that features an accurate description of the ECAL selective readout. No dependence of the PF Candidates energy spectrum nor of the template shape on the online zero-suppression thresholds has been observed.

Template cross-check with electrons in data

The cone isolation distribution for an isolated object, such as prompt photons, is in principle not dependent on the nature of the object itself. Isolated electrons from Z decays provide a sufficiently pure and high-statistics sample for this purpose.

The footprint removal technique decouples the isolation sum from the characteristics of the energy deposit in the ECAL. Nevertheless, it should be taken into account that electrons and photons have a different behaviour while crossing the tracker material upstream of the ECAL. Electrons are more subject to energy loss by bremsstrahlung than photons. One can expect that electron templates reproduce, to a good extent, the ones built with the random-cone technique, but show an increased leakage of energy from the
Figure 4.6: Signal templates in MC, compared with the ones from the random-cone technique and electrons from Z decays in data. [75]

hard object into the isolation cone. This is observed in the distributions shown in Fig. 4.6, especially for the endcap region.

Template shape for fragmentation photons

In this analysis, isolated photons originating from the fragmentation of a quark are considered part of the signal. The distribution of the template variable for fragmentation and direct photons has been compared in the simulation, separating the two cases using generator information. Figure 4.7 shows an isolation distribution slightly shifted to higher values for fragmentation photons. This effect is taken into account when estimating the systematic uncertainty in the signal purity.

4.2.2 Background template with the \( \sigma_{\eta\eta} \) sideband method

The background template is defined by selecting candidates that fail one selection requirement (we define this new set of requirements as “sideband” selection). In this way, it is possible to select a sample highly enriched in “photon-like” jets, i.e. jets where the hadronization process fluctuates into a boosted neutral meson with limited activity around. It is important to keep the sideband selection as close as possible to the photon selection, to keep rejecting genuine jets that have very different energy profiles.

The sideband selection is defined by replacing the requirements on the \( \sigma_{\eta\eta} \) shower shape variable (sideband variable) with the following ones:

<table>
<thead>
<tr>
<th>Sideband definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 0.011 &lt; \sigma_{\eta\eta} &lt; 0.014 ) (ECAL barrel)</td>
</tr>
<tr>
<td>( 0.030 &lt; \sigma_{\eta\eta} &lt; 0.034 ) (ECAL endcaps)</td>
</tr>
</tbody>
</table>
Figure 4.7: Signal templates in MC for direct and fragmentation photons, compared with the ones from the random-cone technique. [75]

The key point for maximizing the performance of sideband templates is choosing a sideband variable that shows little correlation with the template variable. In this way, the shape of the sideband template will be close to the isolation distribution for non-prompt photons passing the photon selection.

The $\sigma_{\eta\eta}$ shower shape is calculated only from energy deposits in a $5 \times 5$ matrix centered on the most energetic crystal in the supercluster. The footprint removal technique excludes the supercluster area from the calculation of the isolation sum. Therefore, no common input exists between these two variables. A residual correlation between them is motivated by physical arguments related to the detailed properties of jet hadronization.

Among the different components (charged, neutral, photon) of the PF isolation, the photon one is observed to have the minimal correlation with the electromagnetic shower shape. This justifies its choice as template variable in this analysis. The choice of a combined isolation (charged plus photon) as template variable would have spoiled the performance of the sideband technique, resulting in a larger systematic uncertainty in the final result.

The performance of the sideband method has been checked in simulation. Templates obtained from reconstructed photons passing the sideband selection and from non-prompt generated photons have been compared (Fig. 4.8). In this case, the template shape shows a dependency on both $p_T$ and $\eta$. Therefore, a reweighting is applied in order to equalize their distributions for the candidates entering this comparison.

The agreement between the templates in simulation is good. This confirms that the correlation described above is small enough for the method to yield isolation templates for non-prompt photons. Moreover, the distribution in data is correctly modeled by the simulation, and does not show any qualitatively unexpected features.
Figure 4.8: Comparison of non-prompt photon templates in data and simulation: (red) non-prompt photons in the simulation, (blue) non-prompt photon templates extracted with the sideband technique from simulation and (black) from data; Top: candidates in the ECAL barrel, Bottom: candidates in the ECAL endcaps. All histograms are normalized to unit area. [74, 75]
Dependence of the template shape on the sideband variable

The estimation of the systematic uncertainty on the purity measurement from the shape of the template for fake photons relies on a representative description of the correlation between the template and sideband variables in the simulation. While in data it is not possible to select a pure sample of non-prompt photons passing the photon selection, this correlation can be studied in the sideband region. The same can be done in simulated events.

In this study, the sideband region is split in two $\sigma$ sub-regions (0.011-0.0125, 0.0125-0.014 for EB, 0.030-0.032, 0.032-0.034 for EE). The isolation distributions in each sub-region are compared in data and simulation to check that no correlation between isolation and $\sigma$ is observed in data, which is not accounted for in the simulation. Figure 4.9 shows that the modeling of this correlation is qualitatively correct in the simulation.

4.2.3 Two-dimensional templates with event matching technique

The rationale for modeling the isolation correlations in two-dimensional templates is to select candidates with the same kinematics as the diphoton events to be fitted. Let us treat first the case where the isolation cones do not overlap.

Two large samples of random-cone isolation sums ("random-cone candidates" in the following) and sideband candidates are first built as described in the previous Sections. Let us denote these samples as $T_R$ and $T_S$ respectively. For each template candidate, the following quantities are considered:

- $\eta, \phi$ of the candidate
- FastJet $\rho$ of the parent event
- only for sideband candidates, the $p_T$ of the candidate

These properties are inserted in a $kD$-tree [80] structure.

Templates are built by examining each diphoton event to be fitted. The $kD$-Tree is used to retrieve in a computationally efficient way the 5 template candidates whose properties are closest to those of each photon candidate. Let us denote as $e_{11}, \ldots, e_{15}$ the template candidates for photon 1, and as $e_{21}, \ldots, e_{25}$ those for photon 2. In this way, the statistics of events used to build templates (that serve as probability distribution functions in the fit) largely exceeds that of fitted events. This is beneficial to the fit stability.

When building the prompt-prompt template, template candidates are sampled from $T_R$ for both photons. When building the prompt-fake template, one set of candidates is extracted from $T_R$ and the other one from $T_S$. For the non-prompt-non-prompt template, both are extracted from $T_S$. Templates are populated using the isolation of template candidates $e_{1i}$ and $e_{2j}$ for the two axes of the $(\text{Iso}_1, \text{Iso}_2)$ plane.

4.2.4 Templates for overlapping isolation cones

The construction of templates used to fit events where the two photons are close to each other requires more care. The analysis selection applies a minimum requirement of 0.45 on the $\Delta R$ separation between the photons. This implies that even if the energy deposit
Figure 4.9: Signal templates in data and MC for the two regions of the sideband. The “left” sideband sub-region extends from 0.011 to 0.0125 in EB, from 0.030 to 0.032 in EE. The “right” sideband sub-region extends from 0.0125 to 0.014 in EB, from 0.032 to 0.034 in EE. [75]
Figure 4.10: Photon candidates with overlapping isolation cones.

Figure 4.11: Procedure used to build the prompt-not-prompt template.

of one photon cannot enter the isolation cone of the other one, the two cones can still partially overlap. When this happens, PFCandidates located in the overlap region enter both isolation sums, increasing the correlation between them (Fig. 4.10).

For prompt-prompt and prompt-fake templates, it is straightforward to deal with this situation. In the case of the prompt-prompt template, isolation sums are calculated in an empty region of the detector, along directions oriented as the photon candidates. In the case of the prompt-fake template, a sideband candidate is first selected from the $T_5$ sample. Its isolation is used for the template axis corresponding to the non-prompt hypothesis. Then, another isolation sum is calculated in the same parent event along the direction of the photon candidate in the event to be fitted. This second isolation sum is used to populate the template axis corresponding to the prompt hypothesis (Fig. 4.11).

The case of the non-prompt-non-prompt template is more involved, because the overlap of the energy flows surrounding the two candidates has to be taken into account. Ideally, one should build the template from events containing two sideband candidates close to each other. This situation happens too rarely to ensure a sufficient statistical pop-
ulation of the template.

An alternative technique based on event merging is used in those bins of the differential variables where at least 10% of the events to be fitted present a ΔR separation between the photons of less than 1.0. This method fully exploits the rationale of the Particle-Flow event reconstruction. Its output is a set of particle candidates (PFCandidates) that provides a unique interpretation of the event. One should note that, to a first approximation, additional pp interactions happening in the same bunch crossing (pile-up) yield additional PFCandidates in the final state.

In analogy to the previous case of well separated photons, two sideband candidates \( e_{1i}, e_{2i} \) are extracted from \( T_S \) requiring their orientation to match the one of the photon candidates. However, in this case the pile-up energy densities \( \rho_{1i}, \rho_{2i} \) of their parent events do not necessarily match the one \( \rho_{fit} \) of the event to be fitted. Therefore, this requirement is modified to \( \rho_{1i} + \rho_{2i} \simeq \rho_{fit} \).

Then, the merging of the parent events of \( e_{1i} \) and \( e_{2i} \) is performed. Thanks to the property of PF reconstruction described above, it is possible to simply merge the set of reconstructed PFCandidates in the two events. In this way, the merged event has a pile-up energy density of about \( \rho_{1i} + \rho_{2i} \) and contains sideband candidates oriented as the photon candidates in the event to be fitted. The template is populated by calculating the isolation sums in this merged event (Fig. 4.12).

Low-energy thresholds in the PF reconstruction break the assumption used to merge the events. This effect is taken into account as a systematic uncertainty in the template shape, but it is sub-dominant with respect to the correlation between isolation and shower shape variables for the 2011 pile-up conditions.

### 4.3 Purity fit

The diphoton signal is studied as a function of the diphoton invariant mass \( m_{\gamma\gamma} \), the diphoton transverse momentum \( p_{T\gamma\gamma} \), the azimuthal angle difference \( \Delta \phi_{\gamma\gamma} \) between the two photons, and the cosine of the polar angle \( \theta^* \) in the Collins-Soper frame of the diphoton system [81]. A maximum likelihood fit is used to study each bin of the distributions in the above variables.

The fraction of selected events containing two prompt photons is fitted separately for the cases where both candidates are reconstructed in the ECAL barrel, one in the ECAL barrel and one in the ECAL endcaps, or both in the ECAL endcaps. This is necessary because the template shapes are different for the barrel and endcap regions. If both candidates are in the same detector region (EB-EB and EE-EE categories), the leading selected photon is assigned randomly to axis 1 or 2 of the two-dimensional plane, and the prompt-non-prompt \( (f_{pn}) \) and non-prompt-prompt \( (f_{np}) \) fractions are constrained to have the same value.

The fit is restricted to the region where the isolation of both candidates is smaller than 9 GeV. To guarantee its stability even in the less populated bins, the fit is performed in steps. The size of the bins in the \( (\text{Iso}_1, \text{Iso}_2) \) plane is optimized to reduce statistical fluctuations of the template shape in the tails. A first fit is performed using one-dimensional templates on the projections of the isolation distribution on the two axes of the plane.
In a subsequent step, the fractions of prompt-prompt, prompt-non-prompt, non-prompt-prompt, and non-prompt-non-prompt candidates, which are constrained to sum up to unity, are fit in the two-dimensional isolation plane using as a constraint the results of the previous fit. The final likelihood maximization is performed after removing all constraints and using as initial values of the parameters those found in the previous step.

An example of the result of the first fitting step is shown in Fig. 4.13. Figure 4.14 shows instead an example of the final two-dimensional fit, projected on the axes for the sake of clarity. The fractions of prompt-prompt, prompt-non-prompt, and non-prompt-non-prompt photons are shown in Fig. 4.15 as a function of the differential observables.

### 4.3.1 Goodness of fit

The purity measurement is subject to statistical fluctuations non only of the fitted data, but also of the templates used in the fit. Statistical fluctuations of the templates are not considered in the likelihood used in the fit. They are taken into account when estimating the systematic uncertainty in the fit results, as described in Section 4.5.

Figure 4.12: Procedure used to build the non-prompt-non-prompt template.
Figure 4.13: Result of the first step of the fitting procedure, for the $90 \text{ GeV} < m_{\gamma\gamma} < 95 \text{ GeV}$ bin in the EB-EE category: isolation distribution for the photon reconstructed in the (Left) ECAL barrel, (Right) ECAL endcaps. [74]

The goodness of the binned likelihood fit is assessed with a $\chi^2$ test. For this test the fluctuations of both data and templates are considered. The standard $\chi^2$ variable:

$$\chi^2 = \sum_{\text{bins}} \left( \frac{\text{data} - \text{fit}}{\sigma(\text{data})} \right)^2$$

can be generalized to the case of this purity fit using the following assumptions:

- fluctuations in data and in each template are uncorrelated, because data and templates are built from different selections;
- fluctuations can be approximated to be distributed as a Gaussian in both data and templates.

These conditions are verified in each bin of the $(\text{Iso}_1, \text{Iso}_2)$ plane where both the data and the templates have sufficiently high statistics. Therefore, let us restrict the calculation of the $\chi^2$ to bins where the population of data and templates is of at least 20 events.

Under these hypotheses, a new quantity $\chi_{\text{new}}^2$ is defined as follows:

$$\chi_{\text{new}}^2 = \sum_{\text{bins}} \chi_{\text{bin,new}}^2 = \sum_{\text{bins}} \left( \frac{\text{data} - \text{fit}}{\sqrt{\sigma^2(\text{data}) + \sigma^2(\text{fit})}} \right)^2 = \chi_{\text{new}}^2$$
Figure 4.14: Projections of the result of the final step of the fitting procedure, for the
90 GeV < m_{\gamma\gamma} < 95 GeV bin in the EB-EE category: isolation distribution for the photon
reconstructed in the (Left) ECAL barrel, (Right) ECAL endcaps. [74]

\[ \chi^2_{\text{bin,new}} = \sum_{\text{bins}} \left( \frac{\text{data} - \text{fit}}{\sqrt{\sigma^2(\text{data}) + \sigma^2(\text{pp}) + \sigma^2(\text{pn}) + \sigma^2(np) + \sigma^2(nn)}} \right)^2 \]

where the denominator contains the contributions from the statistical fluctuations
propagated from each template to the fitted prompt-prompt, prompt-non-prompt, or non-
prompt-non-prompt yield in each bin.

The number of degrees of freedom is the number of \((I_{\text{iso}1}, I_{\text{iso}2})\) bins considered in the
sum, reduced by 1 for the normalization constraint and by the number of free parameters
in the fit (2 for EB-EB and EE-EE categories, 3 for the EB-EE category).

The expected distribution of \(\chi^2_{\text{bin,new}}\) is not a Gaussian with zero mean and a standard
deviation of 1, because of the minimal population required in each bin. Therefore, the
distribution of the \(\chi^2_{\text{bin,new}}\) variable is not a \(\chi^2\) and the goodness of the fit cannot be evalu-
ated rigorously using the \(\chi^2\) cumulative distribution function. Nevertheless, it is possible
to study qualitatively the pull \(\chi_{\text{bin,new}}\) for each bin of the \((I_{\text{iso}1}, I_{\text{iso}2})\) plane. Figure 4.16
shows the distribution of this variable for all bins of the differential variables in each
category (EB-EB,EB-EE,EE-EE). The distribution is bell-shaped, and no large tails nor sig-
nificant populations far from the core are observed.

4.4 Efficiency correction and unfolding

4.4.1 Subtraction of electron contamination

The prompt diphoton yield extracted from the purity fit is contaminated by electrons com-
ing predominantly from the Drell-Yan e^+e^- process and incorrectly identified as photons.
Figure 4.15: Fractions of prompt-prompt, prompt-non-prompt and non-prompt-non-prompt components as a function of $m_{\gamma\gamma}$, $p_T^{\gamma}$, $\Delta\phi_{\gamma\gamma}$, $|\cos \theta^*|$ in the whole acceptance of the analysis. Uncertainties are statistical only. [74]

This contamination is most significant in the Z peak region, where it reaches about 25% of the total yield.

All photon identification requirements have similar efficiency on isolated electrons and photons, with the exception of the absence of hits in the pixel detector in the case of photons. Therefore, an inefficiency in reconstructing track segments (“tracklets”) in the pixel detector leads to the mis-identification of an electron as a photon. Technically, the track segments used for this identification step coincide with those used for seeding the Gaussian-Sum-Filter track reconstruction used for electrons. They are referred to as “GSF seeds”. [62]

The diphoton yield is corrected by subtracting the Drell-Yan $e^+e^-$ contamination yield in simulated events after applying a data/simulation scale factor to it. This scale factor for GSF seeding inefficiency is measured using a tag-and-probe procedure with electrons from Z decays.

The correlation between the inefficiencies for the two legs of the Z decay cannot be neglected. This is due to the presence of failure mechanisms in the iterative tracking that are due to mis-reconstruction of interaction vertices using only pixel tracklets. If the correct vertex is not reconstructed using only the pixel tracklets, no GSF seed will be created for either electron from the Z decay. Therefore, a standard tag-and-probe technique using
Figure 4.16: Distribution of the fit pull corrected for template fluctuations ($\chi_{\text{bin,new}}$). [75]

one electron as the tag and the other one as the probe cannot be used.

The alternative method used here relies on the fact that redundant mechanisms are in place in the reconstruction software to seed the track fitting with the Kalman-Filter (KF) algorithm. The procedure applied can be separated in the following steps:

- The requirement of the absence of a GSF seed is removed from the photon selection. Only events with two candidates passing this relaxed selection and with invariant mass compatible with the Z mass are retained.

- Only events where two KF tracks are reconstructed in the direction of the photon candidates are retained. A loose compatibility requirement between the $p_T$ of the track and the transverse energy of the candidate’s supercluster is applied.

- The KF tracks are required to have at least 3 hits in the TIB or TID strip tracker sub-detectors. This requirement ensures that these tracks are reconstructed even in absence of a GSF seed, thanks to a dedicated seeding step in the iterative tracking algorithm. Therefore, the reconstruction efficiency of these KF tracks is not correlated with the GSF seed efficiency.
• The fraction of $Z$ events where both legs lack the GSF seed is studied by means of a fit to the invariant mass distribution, as done in the standard tag-and-probe technique.

The data/simulation scale factor for GSF seed inefficiency on both legs of the $Z$ decay is found to be $0.7 \pm 0.3$. The systematic uncertainty in the scale factor is estimated varying the track selection requirements.

### 4.4.2 Trigger efficiency

The trigger efficiency is evaluated using a tag-and-probe technique on $Z \rightarrow e^+e^-$ events [73], with electrons treated as photons, and is found to be constant over the data taking period. The contributions to the trigger inefficiency are evaluated using a set of trigger paths running concurrently with the diphoton ones.

The leading source of inefficiency in the HLT diphoton paths is due to the differences between online and offline track finding algorithms. Tracks are used to calculate a cone isolation variable used for photon identification at HLT in some of the paths used for this analysis. Each HLT diphoton path is required to be initiated by at least one L1 EG candidate, which consists of an energy deposit in ECAL compatible with an electron or photon. The reason for requiring at least one L1 candidate and not two is to reduce the effect of the L1 inefficiency due to energy thresholds, identification requirements and inoperative trigger towers on the final efficiency of the trigger path. Since any of the two photons in the event can contribute a L1 seed, the overall L1 trigger path efficiency for a pair of photon candidates is given by $\epsilon_{\text{diphoton}}^{\text{L1}} = (1 - (1 - \epsilon_{\text{L1}}) \times (1 - \epsilon_{\text{L1}}))$.

Once a L1 seed is available, ECAL electromagnetic clusters are formed in the region in the vicinity of the L1 seed. ECAL information is processed only from the readout units overlapping with a rectangle centered on the L1 candidate with a size $\Delta \eta \times \Delta \phi = 0.25 \times 0.4$. The resulting cluster is required to have a position matching the one of the L1 candidate and a transverse energy satisfying the requirements of the given HLT path. Furthermore, candidates that overlap with a significant energy deposit in the hadronic calorimeter (HCAL) are rejected.

The trigger efficiency is measured using events that pass a trigger path requiring a single tagged electron plus an ECAL cluster. This unprescaled electron trigger (HLT_Ele32_Iso_SC17) applies tight isolation and identification requirements on one leg and minimal ECAL cluster reconstruction requirements on the other leg. The efficiency for a photon to pass such minimal requirements is nearly 100% for $E_T > 5$ GeV. The events are required to pass the analysis offline selection with the exception of the pixel seed veto requirement. The offline photon candidates are geometrically matched to the candidates at HLT level. In order to further enrich the sample in electrons, the photon candidates are required to form an invariant mass compatible with the $Z$ mass (between 70 and 110 GeV).

Differences in the probability of interaction with the tracker material between electrons and photons are taken into account. The $R_9$ variable is used as an indication of the amount of interaction. Electron candidates are reweighted such that their $R_9$ distribution equals the one of photons in the simulation. This has the net effect of increasing the measured efficiency due to the migration of events towards higher $R_9$ values.
Table 4.4: Trigger efficiency for $Z \rightarrow e^+e^-$ events with two electromagnetic showers passing the analysis selection, corrected for electron/photon differences.

<table>
<thead>
<tr>
<th>Both photons in barrel</th>
<th>At least one in endcap</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\min(R_9) &gt; 0.94$</td>
<td>$\min(R_9) &lt; 0.94$</td>
</tr>
<tr>
<td>100.00±0.01%</td>
<td>99.3±0.11%</td>
</tr>
<tr>
<td>$\min(R_9) &gt; 0.94$</td>
<td>$\min(R_9) &lt; 0.94$</td>
</tr>
<tr>
<td>100.00±0.02%</td>
<td>98.8±0.4%</td>
</tr>
</tbody>
</table>

This sample of events is finally used to measure the efficiency of the HLT diphoton paths with a tag-and-probe method. Events are categorized depending on the photon position (both photons in the ECAL barrel, or at least one in the endcaps) and the amount of interaction with the tracker material ($R_9 > 0.94$ for both photons, or not). Results are presented in Table 4.4.

The systematic uncertainty on the efficiency is derived from the difference in the results obtained by using different triggers for selecting the tag-and-probe sample. These diphoton triggers (HLT_Photon26_CaloId_Loosy_Photon18 and HLT_Photon26_Photon18) are prescaled and apply relaxed requirements on one or both photon candidates.

### 4.4.3 Selection efficiency

The selection efficiency in data can be separated into the trigger efficiency and reconstruction/selection efficiency:

$$\epsilon_{\gamma\gamma} = \epsilon_{\text{trig}} \times \epsilon_{\text{sel}} \times C_{Z\rightarrow e^+e^-} \times C_{Z\rightarrow e^+e^-} \times C_{Z\rightarrow \mu^+\mu^-} \times C_{Z\rightarrow \mu^+\mu^-},$$

(4.3)

where $\epsilon_{\text{trig}}$ is the trigger efficiency and $\epsilon_{\text{sel}}$ is the diphoton reconstruction and selection efficiency from simulation (“raw” efficiency in the following). The factors $C_{Z\rightarrow e^+e^-}$ and $C_{Z\rightarrow e^+e^-}$ are the corrections to the efficiency for each photon candidate to pass all the selection requirements except the electron veto; $C_{Z\rightarrow \mu^+\mu^-}$ and $C_{Z\rightarrow \mu^+\mu^-}$ are the corrections to the electron veto efficiency.

The central value of the raw efficiency is extracted from the MADGRAPH signal sample. The systematic uncertainty is evaluated using the PYTHIA box signal sample instead. The values of the correction factors are determined from the ratio of the efficiency in data to that in the simulation, measured with a tag-and-probe method using (i) samples of $Z \rightarrow e^+e^-$ for the full selection except the electron-veto requirement, and (ii) samples of photons from the final-state-radiation of $Z \rightarrow \mu^+\mu^-\gamma$ for the electron-veto requirement.

### Efficiency scale factors from $Z \rightarrow e^+e^-$ events

The tag-and-probe method is used to measure the efficiency for an isolated electron to pass the photon selection requirements, with the exception of the GSF seed veto which has been discussed already. The study is performed separately for barrel and endcap, and as a function of the probe $p_T$.

$Z \rightarrow e^+e^-$ events are selected using the same HLT paths used for evaluating the trigger efficiency, as described in Section 4.1.1. The invariant mass shape in $Z$ events is described with a template derived on simulated events, smeared with a Gaussian function...
Table 4.5: Selection efficiency from $Z \rightarrow e^+e^-$ data and MC samples in each $E_T$ bin. [75]

<table>
<thead>
<tr>
<th>$E_T$ bin (GeV)</th>
<th>$\epsilon_{data}$</th>
<th>$\epsilon_{MC}$</th>
<th>$\epsilon_{data}/\epsilon_{MC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25-35</td>
<td>0.948±0.001(stat.)±0.007(syst.)</td>
<td>0.956±0.004(stat.)±0.007(syst.)</td>
<td>0.991±0.008(tot.)</td>
</tr>
<tr>
<td>35-40</td>
<td>0.949±0.001(stat.)±0.007(syst.)</td>
<td>0.961±0.002(stat.)±0.007(syst.)</td>
<td>0.988±0.007(tot.)</td>
</tr>
<tr>
<td>40-45</td>
<td>0.966±0.001(stat.)±0.007(syst.)</td>
<td>0.972±0.001(stat.)±0.007(syst.)</td>
<td>0.993±0.007(tot.)</td>
</tr>
<tr>
<td>45-50</td>
<td>0.974±0.001(stat.)±0.007(syst.)</td>
<td>0.977±0.001(stat.)±0.007(syst.)</td>
<td>0.996±0.007(tot.)</td>
</tr>
<tr>
<td>&gt;50</td>
<td>0.981±0.002(stat.)±0.007(syst.)</td>
<td>0.985±0.005(stat.)±0.007(syst.)</td>
<td>0.996±0.009(tot.)</td>
</tr>
</tbody>
</table>

Probe object in ECAL endcap

<table>
<thead>
<tr>
<th>$E_T$ bin (GeV)</th>
<th>$\epsilon_{data}$</th>
<th>$\epsilon_{MC}$</th>
<th>$\epsilon_{data}/\epsilon_{MC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>25-35</td>
<td>0.935±0.007(stat.)±0.008(syst.)</td>
<td>0.934±0.004(stat.)±0.008(syst.)</td>
<td>1.001±0.012(tot.)</td>
</tr>
<tr>
<td>35-40</td>
<td>0.949±0.002(stat.)±0.008(syst.)</td>
<td>0.936±0.007(stat.)±0.008(syst.)</td>
<td>1.014±0.011(tot.)</td>
</tr>
<tr>
<td>40-45</td>
<td>0.968±0.001(stat.)±0.008(syst.)</td>
<td>0.958±0.002(stat.)±0.008(syst.)</td>
<td>1.010±0.008(tot.)</td>
</tr>
<tr>
<td>45-50</td>
<td>0.978±0.001(stat.)±0.008(syst.)</td>
<td>0.967±0.003(stat.)±0.008(syst.)</td>
<td>1.011±0.008(tot.)</td>
</tr>
<tr>
<td>&gt;50</td>
<td>0.989±0.001(stat.)±0.008(syst.)</td>
<td>0.979±0.002(stat.)±0.008(syst.)</td>
<td>1.010±0.008(tot.)</td>
</tr>
</tbody>
</table>

to account for the slightly worse electron energy resolution in data than in the simulation. The smooth background is modeled with an exponential shape smeared with a Gaussian function and complemented with a turn-on function to account for the shaping of the distribution by the $p_T$ cuts applied on the electrons.

Figure 4.17 shows an example of the invariant mass fits. The results of the efficiency measurement are summarized in Table 4.5 and Fig. 4.18. The uncertainty includes the statistical uncertainty from the fit and the systematic uncertainty from the choice of the signal and background models.

**Efficiency scale factors from $Z \rightarrow \mu^+\mu^-\gamma$ events**

$Z \rightarrow \mu^+\mu^-\gamma$ events are used to measure the efficiency for a photon to pass the GSF seed veto. Muon candidates are first preselected requiring that both the muon detectors and the silicon tracker have measured the candidate and applying loose isolation criteria. Then, the invariant mass of the muon pair is required to lie in the range between 40 and 80 GeV. Events are retained further only if a photon candidate with $p_T$ larger than 10 GeV is found. The invariant mass of the $\mu^+\mu^-\gamma$ system should also be compatible with the $Z$ mass and one of the muons should be close to the photon.

This set of requirements provides a sample of $Z \rightarrow \mu^+\mu^-\gamma$ events with purity higher than 99%. The residual contamination from other processes is estimated with a tag-and-probe method, using the $\mu^+\mu^-$ pair as the tag object and the photon as the probe. The signal is modeled with a template built from the simulation, while the background is described by an exponential shape. Figure 4.19 shows an example of the fitted invariant mass distributions in data.

The efficiency for the probe to pass the GSF seed veto is studied separately in barrel and endcaps. Because of the limited statistics of the sample, it is not possible to perform a differential measurement as a function of the photon $p_T$. Table 4.6 presents the results of this study.

In summary, the total diphoton selection efficiency $\epsilon_{sel}$ is about 85% when both photons are in the barrel, 75% when one photon is in the barrel and the other in one of the
Figure 4.17: Probe photons from $Z \rightarrow e^+e^-$ in data (a) and MC (b) samples are used for the selection efficiency study excluding the electron veto. These plots show the fit results from the tag and probe method, for electrons in the bin $35 \text{ GeV} < E_T < 40 \text{ GeV}$ in the ECAL barrel. Passing candidates are shown in green, failing ones in red, all candidates in blue. [75]
Figure 4.18: Data/MC correction factors of the selection efficiency, excluding the electron veto. The error bars indicate the statistical uncertainty only. Left for the ECAL barrel, right for the ECAL endcaps. [75]

Figure 4.19: Fit result for probe photons from $Z \rightarrow \mu^+ \mu^- \gamma$ in data, for the barrel region. Passing candidates are shown in green, failing ones in red, all candidates in blue. [75]

Table 4.6: Electron veto efficiencies from $Z \rightarrow \mu^+ \mu^- \gamma$ data and MC samples.

| $|\eta|$ bin | $\epsilon_{data}$ | $\epsilon_{MC}$ | $\epsilon_{data}/\epsilon_{MC}$ |
|-----------|------------------|------------------|-------------------------------|
| 0-1.4442  | 0.963 ± 0.006(stat.) | 0.959 ± 0.003(stat.) | 1.004 ± 0.009(total) |
| 1.566-2.5 | 0.871 ± 0.017(stat.) | 0.880 ± 0.011(stat.) | 1.025 ± 0.021(total) |
endcaps, and 64% when both photons are in the endcaps. The data/MC scale factors range from 0.99 to 1.03, depending on the photon $E_T$ and $\eta$.

### 4.4.4 Unfolding

The detector effects are unfolded from the measured yields in order to enable a direct comparison of the experimental measurements with theoretical predictions. The unfolded diphoton yield is obtained correcting the number of reconstructed events in data using an unfolding response model obtained from simulation.

An unfolding matrix $M$ is built with MADGRAPH simulated events, such that $\mathbf{N}_{MC}^{RECO} = M \times \mathbf{N}_{MC}^{GEN}$. Figure 4.20 shows examples of the $M$ matrix for some of the differential observables and event categories. Thanks to the excellent resolution, and to the chosen binning, the unfolding matrix is very close to a diagonal matrix, therefore limiting the impact of the detector effects and making the unfolding rather straightforward. The D’Agostini iterative technique [82, 83] based on Bayes’ theorem is then applied on data using $M$ as response model. The largest unfolding correction amounts to 7% of the reconstructed diphoton yield, for the bins where the slope of the kinematic distributions is the steepest.

### 4.5 Systematic uncertainties

Table 4.7 summarizes the main sources of systematic uncertainty in the measurement of the integrated cross section. Each of them is described in detail in the following.

**Template shape uncertainty**

The dominant uncertainty in the template shapes arises from the difference in shape between the templates built with the techniques described in Section 4.2 and the distributions of the isolation variable for prompt or non-prompt isolated photons for simulated events. The latter are used to generate toy data samples for each bin of the differential variables, with the prompt-prompt, prompt-non-prompt and non-prompt-non-prompt fractions measured in data. Each of these datasets is fitted with templates obtained from simulated events with the same techniques used on data.

The average difference between the prompt-prompt fraction obtained from the fit and used for the generation of the toy dataset is quoted as a systematic uncertainty. It amounts
to about 3% (barrel template) and 5% (endcap template) for the prompt component, and between 5% (barrel template) and 10% (endcap template) for the non-prompt component.

In the case of the non-prompt-non-prompt template, for bins where a significant fraction of the diphoton candidates have overlapping isolation cones, an additional uncertainty ranging from 3% to 5% is introduced to account for the imperfect template shape description due to the effect of ECAL noise and PF low-energy thresholds on the merging of two different events to build the template. The uncertainty is derived with the same method described in the previous paragraph, using background templates built with the event mixing technique.

The template shape is also affected by the uncertainty in the fraction of fragmentation photons in the simulation. The probability of the photon fragmentation process is doubled (by reweighting the affected events) before building the templates, and the toy study

Figure 4.20: Response matrix for the unfolding analysis of (a) $m_{\gamma\gamma}$, (b) $p_{T\gamma}$, (c) $\Delta\phi_{\gamma\gamma}$ and (d) $|\cos\theta^*|$ with both photons in the ECAL barrel. [75]
described above is repeated with the modified shapes. This yields an additional 1.5% uncertainty in the measured cross section.

The systematic uncertainty arising from the statistical uncertainty in the shape of the templates is evaluated with a different toy study. The content of each isolation bin of the template is smeared by a Gaussian distribution centered on the nominal bin value and with standard deviation equal to the statistical uncertainty of the bin. The root mean square of the distribution of the diphoton purity values fitted with modified templates is used as systematic uncertainty in the purity and amounts to about 3%.

Other uncertainties
A possible bias associated with the fitting procedure is evaluated using pseudo-experiments. Pseudo-data samples are generated with given fractions of prompt-prompt, prompt-non-prompt, and non-prompt-non-prompt contributions, using the templates from simulation as generator probability density functions. Each data sample is then fitted with the same templates used for the generation. The average bias is negligible in all bins.

The systematic uncertainty associated with the subtraction of Drell-Yan $e^+e^-$ events is evaluated by propagating the uncertainty in the electron to photon mis-identification probability to the subtracted yield. The uncertainty in the fraction of such events that is fitted as prompt-prompt is also taken into account. This contribution is maximal for $m_{\gamma\gamma}$ close to the Z-boson mass. The relative contribution to the total systematic uncertainty is below 0.5%.

The systematic uncertainty in the trigger efficiency is below 0.5%. The systematic uncertainty in the reconstruction and selection efficiencies is obtained from the uncertainty in the data-simulation correction factors from the $Z \to e^+e^-$ and $Z \to \mu^+\mu^-\gamma$ control samples, and it ranges from 2% in the barrel to 4% in the endcap. The systematic uncertainty in the integrated luminosity is 2.2% [84].

In summary, the total systematic uncertainty in the measurement amounts to approximately 8% when both candidates are reconstructed within the ECAL barrel, and to 11% for the full acceptance of the analysis.

4.6 Results and comparison with theoretical predictions

The unfolded differential cross sections are compared with the following generators for QCD diphoton production: SHERPA 1.4.0 [28], DIPHOX 1.3.2 [46] supplemented with GAMMA2MC 1.1 [47], RESBOS [48, 49], and $2\gamma$NNLO [32].

Predictions with SHERPA are computed at LO for the Born contribution with up to three additional real emissions (three extra jets) and with the box contribution at the matrix element level.

The DIPHOX NLO generator includes the direct and fragmentation contributions and uses a full fragmentation function for one or two partons into a photon at NLO. The direct box contribution, which is formally part of the NNLO corrections, since it is initiated by gluon fusion through a quark loop, is computed at NLO with GAMMA2MC.
The RESBOS NLO generator features resummation for Born and box contributions, and effectively includes fragmentation of one quark/gluon to a single photon at LO. The latter process is regulated to avoid divergences and does not include the full fragmentation function. The RESBOS $p_T^{\gamma\gamma}$ spectrum benefits from a soft and collinear gluon resummation at next-to-next-to-leading-log accuracy.

$2\gamma\text{NNLO}$ predicts the direct $\gamma\gamma+X$ processes at NNLO.

The SHERPA sample is used after hadronization while DIPHOX + GAMMA2MC, RESBOS, and $2\gamma\text{NNLO}$ are parton-level generators only and cannot be interfaced with parton shower generators.

The predictions have been computed for the phase space $E_T^{\gamma_1} > 40$ GeV, $E_T^{\gamma_2} > 25$ GeV, $|\eta_1| < 1.44$ or $1.57 < |\eta_1| < 2.5$, $\Delta R(\gamma_1, \gamma_2) > 0.45$. An isolation requirement is applied at the generator level. In SHERPA, the $E_T$ sum of stable particles in a cone of size $\Delta R = 0.4$ has to be less than 5 GeV (after hadronization). In DIPHOX, GAMMA2MC, and RESBOS the $E_T$ sum of partons in a cone of size $\Delta R = 0.4$ is required to be less than 5 GeV. In $2\gamma\text{NNLO}$, the Frixione isolation [31] is applied to the photons to suppress the fragmentation component:

$$E_T^{\text{iso}}(\Delta R) < \epsilon \left( \frac{1 - \cos(\Delta R)}{1 - \cos(\Delta R_0)} \right)^n,$$

where $E_T^{\text{iso}}$ is the $E_T$ sum of partons in a cone of size $\Delta R$, $\Delta R_0 = 0.4$, $\epsilon = 5\text{GeV}$, and $n = 0.05$. This criterion, tested with DIPHOX, is found to have a similar efficiency as that used for the other generators.

A non-perturbative correction is applied to DIPHOX, GAMMA2MC, and $2\gamma\text{NNLO}$ predictions to correct for the fact that those generators do not include parton shower or underlying event contributions to the isolation cone. The fraction of diphoton events not selected due to underlying hadronic activity falling inside the isolation cone is estimated using the PYTHIA 6.4.22 [39] event generator with tunes Z2, D6T, P0, and DWT [77]. A factor of $0.95 \pm 0.04$ is applied to the parton-level cross section to correct for this effect.

Theoretical predictions use the CT10 [85] NLO PDF set for SHERPA, DIPHOX + GAMMA2MC, and RESBOS, and the MSTW2008 [33] NNLO PDF set for $2\gamma\text{NNLO}$. The DIPHOX and GAMMA2MC theoretical uncertainties are computed in the following way: the factorization and renormalization scales in GAMMA2MC are varied independently up and down by a factor of two around $m_{\gamma\gamma}$ (configurations where one scale has a factor of four with respect to the other one are forbidden). In DIPHOX, the factorization, renormalization and fragmentation scales are varied in the same way. In RESBOS, the factorization and renormalization scales are varied coherently by a factor of two. The maximum and minimum values in each bin are used to define the uncertainty.

In DIPHOX, GAMMA2MC, and RESBOS, the 52 CT10 eigenvector sets of PDFs are used to build the PDF uncertainty envelope, also considering the uncertainty in the strong coupling constant $\alpha_S$, determined according to the CT10 $\alpha_S$ PDF set. In $2\gamma\text{NNLO}$, because of CPU limitations, only the renormalization and factorization scales are varied coherently by a factor of two up and down around $m_{\gamma\gamma}$, and no PDF uncertainty is computed. The same procedure is used in SHERPA, using the internal METS scale, where scales are defined as the lowest invariant mass or negative virtuality in the core $2\to2$ configuration.
clustered using a $k_T$-type algorithm.

The total cross section measured in data for the phase space defined above is:

$$\sigma = 17.2 \pm 0.2 \text{ (stat.)} \pm 1.9 \text{ (syst.)} \pm 0.4 \text{ (lum.) pb},$$

to be compared with the following theoretical predictions:

$$\sigma_{\text{NNLO}}(2\gamma_{\text{NNLO}}) = 16.2^{+1.5}_{-1.3} \text{(scale) pb},$$
$$\sigma_{\text{NLO}}(\text{DIPHOX + GAMMA2MC}) = 12.8^{+1.6}_{-1.5} \text{(scale)}^{+0.6}_{-0.8} \text{(pdf+\alpha_s)} \text{ pb},$$
$$\sigma_{\text{NLO}}(\text{RESBOS}) = 14.9^{+2.2}_{-1.7} \text{(scale)} \pm 0.6 \text{(pdf+\alpha_s)} \text{ pb},$$
$$\sigma_{\text{LO}}(\text{SHERPA}) = 13.8^{+2.8}_{-1.6} \text{(scale)} \text{ pb}.$$

Figures 4.21, 4.22, 4.23 and 4.24 show the comparisons of the differential cross section between data and the SHERPA, DIPHOX + GAMMA2MC, RESBOS, and 2$\gamma$NNLO predictions for the four observables.

The NLO predictions of DIPHOX + GAMMA2MC are known to underestimate the data [36], because of the missing higher-order contributions. Apart from an overall normalization factor, the phase space regions where the disagreement is the largest are at low $m_{\gamma\gamma}$, and the corresponding low $\Delta\phi_{\gamma\gamma}$. The RESBOS generator shows a similar trend, with a cross section closer to the data than DIPHOX + GAMMA2MC; its prediction is improved at high $\Delta\phi_{\gamma\gamma}$ due to soft gluon resummation.

With higher-order diagrams included, 2$\gamma$NNLO shows an improvement for the overall normalization. It also shows a better shape description, especially at low $\Delta\phi_{\gamma\gamma}$, but it still underestimates the data in the same region. SHERPA generally reproduces rather well the shape of the data, to a similar level as 2$\gamma$NNLO. It should be noted that 2$\gamma$NNLO and SHERPA predict the $p_T^{\gamma\gamma}$ shoulder near $E_T^{\gamma_1} + E_T^{\gamma_2} \sim 65 \text{ GeV}$ observed in the data. This is expected since SHERPA includes up to three extra jets at the matrix element level.

In conclusion, this measurement is providing clear evidence for the need of including higher-order corrections, up to at least NNLO, for obtaining a satisfactory description of diphoton production over a large region of the phase space.
Figure 4.21: The comparisons of the differential cross section between data and the SHERPA, DIPHOX + GAMMA2MC, RESBOS, and 2γNNLO predictions for $m_{γγ}$. Black dots correspond to data with error bars including statistical and systematic uncertainties. Only the scale uncertainty is included for the SHERPA prediction. Scale, PDF and $α_S$ uncertainties are included for DIPHOX + GAMMA2MC and RESBOS. Only statistical and scale uncertainties are included for the 2γNNLO prediction. [74]
Figure 4.22: The comparisons of the differential cross section between data and the SHERPA, DIPHOX + GAMMA2MC, RESBOS, and 2γNNLO predictions for $p_T^{\gamma\gamma}$. Black dots correspond to data with error bars including statistical and systematic uncertainties. Only the scale uncertainty is included for the SHERPA prediction. Scale, PDF and $\alpha_S$ uncertainties are included for DIPHOX + GAMMA2MC and RESBOS. Only statistical and scale uncertainties are included for the 2γNNLO prediction. [74]
Figure 4.23: The comparisons of the differential cross section between data and the SHERPA, DIPHOX + GAMMA2MC, RESBOS, and 2γNNLO predictions for Δφγγ. Black dots correspond to data with error bars including statistical and systematic uncertainties. Only the scale uncertainty is included for the SHERPA prediction. Scale, PDF and αS uncertainties are included for DIPHOX + GAMMA2MC and RESBOS. Only statistical and scale uncertainties are included for the 2γNNLO prediction. [74]
Figure 4.24: The comparisons of the differential cross section between data and the SHERPA, DIPHOX + GAMMA2MC, RESBOS, and 2γNNLO predictions for |cos θ∗|. Black dots correspond to data with error bars including statistical and systematic uncertainties. Only the scale uncertainty is included for the SHERPA prediction. Scale, PDF and αS uncertainties are included for DIPHOX + GAMMA2MC and RESBOS. Only statistical and scale uncertainties are included for the 2γNNLO prediction. [74]
Chapter 5

Measurement of diphoton production in association with jets

The measurement of differential cross sections for the production of a photon pair in association with jets is described. This analysis represents an extension of the inclusive measurement presented in the previous Chapter. These results are presently going through the last steps of the CMS internal approval process, and are being documented in Refs. [86, 87]. The discussions, text and illustrations which can be found in those references are closely followed here.

5.1 Introduction

Diphoton production in association with jets represents the major source of background to the study of the Higgs boson [7, 8, 69] in the diphoton decay channel, when produced via vector boson fusion (VBF). In this production channel (Fig. 5.1), the Higgs is produced in association with two hard jets in the forward and backward regions of the detector. Moreover, gluon radiation in the central rapidity region is suppressed by the fact that vector bosons do not carry color charge. These characteristics are exploited by the event selection typically used to study this channel, allowing to reject the large QCD background, as well as the Higgs produced via gluon fusion in association with two jets.

New physics processes may also appear as deviations from the predicted diphoton spectrum in events with jets in the final state, as in gauge-mediated SUSY breaking [88]. Validating the event generators used to model the background in these analyses is there-

Figure 5.1: Gluon fusion (a) and vector boson fusion (b) Higgs production channels in proton-proton collisions. The VBF channel is characterized by two forward jets produced in association with the Higgs boson. This feature is used in the event selection to reject the QCD background and the Higgs produced via gluon fusion in association with two jets. [14]
fore of crucial importance. Theoretical predictions for this process have been derived by the GoSam [45], BlackHat [89] and NJET [90] collaborations at next-to-leading order in perturbative QCD.

This chapter presents an extension of the inclusive diphoton measurement. The cross section is measured here as a function of several photon and jet observables, separately in events with at least one or two jets in the final state. Both transverse momentum spectra and angular distributions are studied to probe the dynamics of diphoton production. Corrections to data for selection efficiency and detector effects are obtained using simulated events, yielding distributions at the level of individual particles. The results are compared to predictions obtained using the SHERPA [28], aMC@NLO [43] and GoSam event generators.

In events with at least one jet in the final state, the cross section is measured as a function of the number of jets in the final state (Njets), the $\Delta R$ separation between the leading jet and the leading ($\Delta R_{\gamma j}$) and sub-leading ($\Delta R_{\gamma j}$) photon, the minimum ($\Delta R_{\gamma j}$) and maximum ($\Delta R_{\gamma j}$) $\Delta R$ separation between the leading jet and the two leading photons, and the transverse momentum of the leading jet ($p_T,j$).

In events with at least two jets in the final state, the cross section is measured as a function of the transverse momenta, invariant mass, $\Delta R$, $\Delta \eta$ and $\Delta \phi$ separation of the two leading jets ($p_{T,\gamma j}, p_{T,\gamma j}, m_{jj}, \Delta R_{jj}, \Delta \eta_{jj}, \Delta \phi_{jj}$), the $\Delta \phi$ separation between the systems composed by the two leading photons and two leading jets ($\Delta \phi_{\gamma \gamma jj}$), and the Zeppenfeld variable [91] of the event. These variables are especially motivated by their use to discriminate VBF-produced Higgs events from the diphoton QCD background.

The phase space has been defined such that the presence of jets does not bias the photon isolation sums. Therefore, the analysis strategy pursued here is the same as that of the inclusive measurement described in Chapter 4. The jet information is only used to calculate the differential observables. Section 5.2 presents the definition of jets used in the analysis, the corrections applied to their energy and the systematic uncertainties induced on the cross section. Data and theoretical predictions are finally compared in Section 5.3.

This measurement benefits from an updated simulation and reconstruction software release with respect to the one used for the inclusive measurement. The simulated samples include the tuning of the ECAL out-of-time pileup and noise models obtained through the work described in Section 2.2.

The performance of the methods used to build the isolation templates has been checked for the new release. The same procedure as described in Section 4.2 is used. Figures 5.2 and 5.3 show that no large discrepancies are observed between the distributions obtained with the data-driven template building techniques and the isolation distributions for prompt and non-prompt photons. The systematic uncertainty in the cross section due to the template shape mismodeling does not change significantly with respect to the inclusive measurement.

Moreover, the diphoton signal modeling is improved by using SHERPA 1.4.2 [28]. Events are generated with the same configuration for scales and PDF choice as described in Section 4.6. The detector response is then simulated as described in Section 4.1.2.
Figure 5.2: Comparison of prompt photon templates in data and simulation, for the reconstruction software release used in the diphoton+jets analysis: prompt photons in the simulation, prompt photon templates extracted with the random cone technique from simulation and from data; Left: candidates in the ECAL barrel, Right: candidates in the ECAL endcaps. All histograms are normalized to unit area. [86]

Figure 5.3: Comparison of non-prompt photon templates in data and simulation, for the reconstruction software release used in the diphoton+jets analysis: non-prompt photons in the simulation, non-prompt photon templates extracted with the sideband technique from simulation and from data; Left: candidates in the ECAL barrel, Right: candidates in the ECAL endcaps. All histograms are normalized to unit area. [86]
This represents the most accurate diphoton signal model available. \textsc{Sherpa} has been shown to agree very well with data in the inclusive diphoton phase space, and in the configuration used for this analysis the tree-level matrix element calculation includes up to three additional partons in the final state.

5.2 Inclusion of jet information

Jet definition and energy corrections

Jets are constructed by clustering particles using the anti-$k_T$ algorithm with a size parameter $R = 0.5$. In the generator, all stable\footnote{Stable particles are defined by the $ct > 1 \text{ cm}$ requirement, where $\tau$ is the particle lifetime. Neutrinos are excluded.} particles are used. At reconstruction level, jets are clustered from the particle-flow candidates. This provides a greatly improved momentum and spatial resolution with respect to jets constructed from calorimeter deposits alone. The PF algorithm, actually, allows resolving individual charged hadrons and photons inside the jet, which constitute on average about 85% of the jet energy.

Jet energy corrections (Fig. 5.4) are applied both in data and simulation to correct the jet response at reconstruction level to the jet energy in the generator. The raw jet four-momentum is first multiplied by a factor that accounts for the excess energy due to electronic noise and pileup. This is known as offset correction. The average response is then calibrated using simulated QCD multi-jet events, differentially in transverse momentum and pseudorapidity. A residual correction derived from a data/simulation comparison in $\gamma/Z+$jets events is finally applied. Further corrections are applied to match the energy resolution in the simulation to the one observed in data. Balanced di-jet and photon+jet events are used to measure the jet resolution, that is found to be about 10% better in the simulation than in data. More information on jet clustering and energy correction may be found in Ref. [92].

The origin of energy deposits from pileup collisions is mainly related to low-$p_T$ jet production. The combination of overlapping soft jets of this type to form a single high-$p_T$ jet constitutes a relevant source of background for this analysis.

Pileup jets exhibit two characteristic features: the energy flow in the jet is more diffuse than for genuine jets, and a large fraction of its charged constituents does not come from the primary interaction vertex. Two variables used to reject pileup jets are shown in Fig. 5.5. A detailed description of the requirements applied on these and other similar variables can be found in Ref. [93].

Fiducial phase space of the analysis

The jet acceptance is defined by the $p_T > 25 \text{ GeV}$, $|\eta| < 4.7$ requirements. Jets within a cone of $\Delta R < 1.0$ from one of the selected photon candidates in the event are also rejected.

This requirement ensures that photons are not counted as jets. It also makes the definition of the phase space insensitive to jets close to photons, that are harder to model in the Monte Carlo simulation and are affected by the parton-level phase space definition used in the matrix element calculation. Finally, it avoids the selection of jets that could bias the isolation sum of photons, because the isolation cone and the jet area do not overlap.
Figure 5.4: Left: Absolute jet energy scale uncertainty as a function of jet $p_T$ for PF jets. Right: The ratio of jet $p_T$ resolutions in data and MC samples vs. $p_T$, in $|\eta| < 1.1$, from di-jet and $\gamma$+jet samples. A combined fit to both datasets is also shown. [92]

Figure 5.5: Left: $\beta^*$, defined as the sum of the $p_T$ of all PF charged particles associated to an interaction vertex different from the primary one, divided by the sum of the $p_T$ of all charged particles in the jet. Right: average of the square of the $\Delta R$ separation between each particle in the jet and the jet axis, weighted by the particle $p_T^2$. This variable can be used also outside of the tracker acceptance. [93]
Systematic uncertainties due to jet observables

Systematic uncertainties on the jet observables mainly arise from the imperfect description of the jet energy scale and resolution. The impact on the final cross section depends on the variable under consideration. In general, it becomes more important at high jet multiplicity and for variables directly connected to the jet momentum, ranging from 3% to about 15%. Angular observables are also affected through the transverse momentum requirements used to define the jet acceptance, but typically to a more limited extent.

Imperfections in the pileup description also affect the measured cross section. The uncertainty is evaluated by changing by ±5% the inelastic pp cross section used for calculating the pileup distribution in data. The simulation is reweighted to the different target distributions obtained, and the difference in the cross section quoted as a systematic uncertainty. It amounts to about 1.5%. A comparison of the jet uncertainties with those related to the template shape is shown as a function of the jet multiplicity in Fig. 5.7.

5.3 Results and comparison with theoretical predictions

The results of the purity fit for each differential observable are presented in Figs. 5.8, 5.9 and 5.10. Thanks to a more efficient identification of pileup vertices in the reconstruction software release used for this analysis compared to what was used for the diphoton cross section shown in Chapter 4, the contamination of electrons coming from \(Z \rightarrow e^+e^-\) decays and mis-reconstructed as photons is reduced. The pileup vertices can now be reconstructed using only pairs of pixel hits, instead of triplets, in a broader spectrum of situations where one of the pixel modules traversed by the electron is not efficient. The
The differential cross sections measured in data are compared to predictions obtained with SHERPA 1.4.2 [28], aMC@NLO [43] and GoSAM [45]. The SHERPA prediction is obtained as described in Section 4.6 for the inclusive analysis.

The aMC@NLO event generator describes diphoton production in association with up to 2 jets at NLO. The FxFx merging technique [44] is used to merge different jet multiplicities. Events are generated using the NNPDF 3.0 NLO PDF set [94], and hadronized with PYTHIA 8 [95] using the CUETP8M1 tune.

The scale uncertainty is evaluated as the envelope of the predictions obtained varying independently the factorization and renormalization scales up and down by a factor of 2. The PDF uncertainty is evaluated as the rms of the predictions obtained with all NNPDF set replicas. The aMC@NLO prediction is complemented by the box diagram contribution calculated with PYTHIA 8.

The parton-level GoSAM prediction describes diphoton production in association with at least one or two jets at NLO, without parton shower. The predictions for the 1-jet and 2-jet selections are provided separately. Therefore, GoSAM cannot be used to predict the cross section as a function of jet multiplicity.

The scale uncertainty is obtained by varying the factorization and renormalization scale up and down by a factor 2 at the same time. The GoSAM prediction is corrected for the fact that it does not include parton shower or underlying event contributions. The fraction of events not selected due to underlying hadronic activity is estimated using the PYTHIA 8 event generator with tunes CUETP8M1 and 4C [96]. A correction factor of 0.95 ± 0.05 is applied to the predicted cross section.

The SHERPA and aMC@NLO diphoton predictions, without any requirement on the jets possibly present in the event, have been rescaled to the value of the inclusive diphoton

Figure 5.7: Summary of the systematic uncertainties in the diphoton+jets cross section, as a function of the jet multiplicity.
Figure 5.8: Fractions of prompt-prompt (red), prompt-non-prompt (green) and non-prompt-non-prompt (black) components as a function of the differential variables. Uncertainties are statistical only. The differential variable under study is indicated on the horizontal axis label. [86]
Figure 5.9: Fractions of prompt-prompt (red), prompt-non-prompt (green) and non-prompt-non-prompt (black) components as a function of the differential variables. Uncertainties are statistical only. The differential variable under study is indicated on the horizontal axis label. [86]
cross section calculated at NNLO with $2\gamma_{NNLO}$ \cite{32} and presented in Section 4.6. Multiplicative factors of 1.17 and 0.88 have been applied, respectively.

The integrated cross sections measured in data for the 1-jet and 2-jet selections are:

\[ \sigma_{data}^{1\text{-jet}} = 7.4 \pm 0.2 \text{ (stat.)} \pm 1.0 \text{ (syst.)} \pm 0.2 \text{ (lumi)} \text{ pb} \]
\[ \sigma_{data}^{2\text{-jet}} = 2.3 \pm 0.1 \text{ (stat.)} \pm 0.4 \text{ (syst.)} \pm 0.1 \text{ (lumi)} \text{ pb} \]

These measured cross sections are to be compared to the following predictions:

\[ \sigma_{SHERPA}^{1\text{-jet}} = 6.9^{+2.2}_{-1.5} \text{ (scale+stat.) pb} \]
\[ \sigma_{aMC@NLO}^{1\text{-jet}} = 7.3^{+1.1}_{-1.1} \text{ (scale+pdf+stat.) pb} \]
\[ \sigma_{GOSAM}^{1\text{-jet}} = 5.8^{+0.9}_{-0.8} \text{ (scale+stat.) pb} \]
\[ \sigma_{SHERPA}^{2\text{-jet}} = 2.0^{+0.9}_{-0.5} \text{ (scale+stat.) pb} \]
\[ \sigma_{aMC@NLO}^{2\text{-jet}} = 2.2^{+0.4}_{-0.4} \text{ (scale+pdf+stat.) pb} \]
\[ \sigma_{GOSAM}^{2\text{-jet}} = 1.9^{+0.4}_{-0.4} \text{ (scale+stat.) pb} \]

Figures 5.11-5.17 show the comparison of the differential cross sections between data and the theoretical predictions. As expected, aMC@NLO is in general less affected by the scale uncertainty with respect to SHERPA, that is a leading-order generator.

The aMC@NLO prediction for the jet multiplicity is in excellent agreement with data, even for events with three jets or more. SHERPA also agrees with data within the uncertainties, but tends to under-predict the cross section in events with at least two jets.
SHERPA and aMC@NLO are in agreement with data for the jet spectrum and for the angular observables in 1-jet events. GoSAM is not able to predict the angular observables to a similar level of agreement. It should be considered here that GoSAM is a parton-level generator that does not account for additional jet emissions and hadronization. SHERPA and aMC@NLO also agree with data for all observables studied in 2-jet events. GoSAM is in agreement with data, with the exception of $\Delta \phi_{i,j,j}$ where it shows a steeper distribution.

Figure 5.11: Comparison of the differential cross sections between data and the theoretical predictions, as a function of number of jets (left) and jet transverse momentum (right). The black error bars indicate the statistical uncertainty in data. The gray band indicates the full (stat.+syst.) uncertainty in data. The color shaded areas indicate the theoretical uncertainty. [86]
Figure 5.12: Comparison of the differential cross sections between data and the theoretical predictions, as a function of the $\Delta R$ separation between the leading jet and the closer (left) or more distant (right) photon. The black error bars indicate the statistical uncertainty in data. The gray band indicates the full (stat.+syst.) uncertainty in data. The color shaded areas indicate the theoretical uncertainty. [86]
Figure 5.13: Comparison of the differential cross sections between data and the theoretical predictions, as a function of the $\Delta R$ separation between the leading jet and the leading (left) or subleading (right) photon. The black error bars indicate the statistical uncertainty in data. The gray band indicates the full (stat.+syst.) uncertainty in data. The color shaded areas indicate the theoretical uncertainty. [86]
Figure 5.14: Comparison of the differential cross sections between data and the theoretical predictions, as a function of the leading (left) and subleading (right) jet transverse momentum. The black error bars indicate the statistical uncertainty in data. The gray band indicates the full (stat.+syst.) uncertainty in data. The color shaded areas indicate the theoretical uncertainty. [86]
Figure 5.15: Comparison of the differential cross sections between data and the theoretical predictions, as a function of the di-jet invariant mass (left) and $\Delta R$ separation (right). The black error bars indicate the statistical uncertainty in data. The gray band indicates the full (stat.+syst.) uncertainty in data. The color shaded areas indicate the theoretical uncertainty. [86]
Figure 5.16: Comparison of the differential cross sections between data and the theoretical predictions, as a function of the $\Delta \eta$ (left) and $\Delta \phi$ (right) between the two leading jets. The black error bars indicate the statistical uncertainty in data. The gray band indicates the full (stat.+syst.) uncertainty in data. The color shaded areas indicate the theoretical uncertainty. [86]
Figure 5.17: Comparison of the differential cross sections between data and the theoretical predictions, as a function of the $\Delta \phi$ separation between the diphoton and dijet systems (left), and the Zeppenfeld event variable (right). The black error bars indicate the statistical uncertainty in data. The gray band indicates the full (stat.+syst.) uncertainty in data. The color shaded areas indicate the theoretical uncertainty. [86]
Chapter 6

Beam tests of a CeF$_3$ sampling calorimeter prototype channel

In this chapter, the performance evolution of the ECAL endcaps is first presented, and their upgrade in view of the High-Luminosity (HL) LHC running justified. The text and illustrations which can be found in Ref. [97] are used. A sampling calorimeter prototype channel based on CeF$_3$ radiation-hard scintillating crystals, readout through wavelength-shifting fibers, is then described [98]. Results from beam tests with low and high energy electrons are finally illustrated. They demonstrate that such a design is a promising candidate for electromagnetic calorimetry in HL-LHC conditions.

I have personally contributed to the initial design of the CeF$_3$ channel, to its operation in both beam tests, including the development of the DAQ system, and to prompt data analysis. These studies have been performed in the framework of the W-CeF$_3$ collaboration [99, 100].

6.1 Motivation for an upgrade of the ECAL endcaps

The current CMS electromagnetic calorimeter and its electronics were designed to withstand a total integrated luminosity of up to 500 fb$^{-1}$, and a maximum instantaneous luminosity of about $1 \cdot 10^{34}$ cm$^{-2}$s$^{-1}$. Very different operating conditions are expected for the High-Luminosity running of the LHC, that is scheduled to take place between about 2025 and 2035, up to a final delivered integrated luminosity of about 3000 fb$^{-1}$. The instantaneous luminosity is expected to reach a value of $5 \cdot 10^{34}$ cm$^{-2}$s$^{-1}$, with about 140 simultaneous pp interactions per bunch crossing. The radiation levels in the endcap region are strongly position-dependent. At high values of pseudorapidity ($|\eta| \sim 2.6$), the ionizing dose rate will be about 30 Gy/h and the fluence of energetic hadrons will be of the order of $1.8 \cdot 10^{14}$ cm$^{-2}$. These values exceed by a factor of at least four the present ECAL specifications.

The ECAL has played a crucial role in many CMS analyses, including the discovery of the Higgs boson [8]. Also for a large class of physics analyses that will be performed with HL-LHC data, the ECAL will be a key detector component. Therefore, the long-term physics reach of the CMS experiment will strongly depend on the endurance of the ECAL performance. It is crucial to understand in detail how the properties of PbWO$_4$ crystals evolve under irradiation and to study the performance of the calorimeter in high pileup conditions.
6.1.1 Performance evolution

A loss of transparency of the PbWO$_4$ crystals under irradiation has been established with the LHC collision data collected so far [61]. A laser light injection system has been used to continuously monitor this effect during the data-taking. The loss of transparency has been observed to be strongly correlated with the LHC instantaneous luminosity and with the crystal position in the detector, being the largest at high values of $|\eta|$ (Fig. 6.1). These monitoring data have been used to correct the crystal response as a function of time and to stabilize the energy scale of objects whose energy measurement is based on ECAL.

The main mechanism leading to the transparency loss observed in the detector so far is the formation of color centers due to ionizing radiation. This type of damage is not cumulative. Under irradiation at a constant rate, the crystal damage evolves to an equilibrium between the formation of new color centers and the recovery of the previously formed ones. Only the recovery takes place when the irradiation stops (as can be seen in Fig. 6.1). The loss of transparency measured in the detector with collision data correlates with the results obtained from $\gamma$ irradiation during the crystal quality control tests before installation (Fig. 6.2).

However, another component of transparency loss is observed to arise after hadron irradiation [102–105]. It consists of an induced absorption length due to interactions of energetic hadrons with the crystal lattice. It does not recover at room temperature when irradiation stops and therefore builds up during the data-taking. At the large values of integrated luminosity expected at the HL-LHC, this ageing component will become the dominant one. The transparency loss extends throughout the PbWO$_4$ transmission band and causes the lower band edge to shift towards higher wavelengths (Fig. 6.3). The residual light output is expected to be about 10% of the nominal one at $|\eta| \sim 2$ after 3000 fb$^{-1}$, leading to a contribution to the Higgs diphoton invariant mass peak resolution of the order of several percent (Fig. 6.4) [106]. On the other hand, the transparency loss will remain acceptable in the barrel region, where the radiation levels are much lower.
Figure 6.3: Light transmission curves for crystals with various degrees of proton-induced radiation damage. The scintillation emission spectrum (dotted line) has arbitrary normalization. [102]

Figure 6.4: Contribution to the Higgs mass resolution due to the ECAL ageing, for the two photon decay of the Higgs boson in the endcap regions. [106]
Another effect due to crystal transparency loss is the amplification of the effective noise in the ECAL readout electronics. The signal pulse from the ECAL photosensors is sampled by an ADC at 40 MHz. A conversion factor from ADC counts to energy is then applied. While the electronic noise in the endcap channels stays constant in units of ADC counts, the loss of transparency will lead to larger conversion factors, effectively amplifying the energy equivalent of the noise. This effect will strongly degrade the trigger performance in the endcap region, increasing the energy thresholds required on electron and photon triggers. The ECAL resolution will also be worsened by non-uniformity of light collection and non-linearity.

Moreover, the large number of pileup interactions expected at the HL-LHC will significantly increase the detector occupancy. In these conditions, energy deposits from adjacent bunch crossings will bias the crystal energy readout. The main handle available to fight this effect is to increase the detector granularity. In conclusion, all terms (stochastic, noise, constant) of the endcap ECAL resolution are worsened by the crystals’ hadron radiation damage [107].

6.1.2 Upgrade options

The proper functioning of the ECAL barrel will be ensured by upgrades of the readout electronics and modification of their operating conditions. The APD sensors manifest an increase of the dark current when exposed to ionizing radiation. This induces an effective noise on the energy deposition measurement. A reduction of the sensor temperature from 18 °C to 8 °C is envisaged to decrease the dark current by a factor of about 2.5 and the noise by about 30%. The front-end readout electronics will also be replaced, in order to meet more stringent trigger latency requirements, and to directly deal with the anomalous isolated signals from particles interacting directly in the APD (“spikes”).

The loss of performance justifies a more extensive upgrade of the ECAL endcaps for operation under HL-LHC conditions. Several options have been studied for replacing the CMS electromagnetic and hadronic (HCAL) calorimeters in that region with sampling calorimeters based on new technologies, with finer granularity than the existing devices. The two main design schemes under consideration are described in the following.

High-granularity design

A replacement of both ECAL and HCAL endcaps with a highly segmented (longitudinally and transversely) system, based principally on silicon sensors, has been proposed. The aim of this design is to obtain an excellent three-dimensional shower profile reconstruction and particle identification. The current proposal consists of a total of about 600 m$^2$ of silicon sensors of different thickness, segmented into about 9 million channels and interleaved with varying amounts of heavy absorber material. Current R&D studies aim at a highly integrated design of sensors, absorber material and cooling infrastructure.

Sampling design using inorganic scintillating crystals

A sampling calorimeter based on radiation-hard inorganic scintillating crystals interleaved with heavy absorber material plates has also been proposed to replace the ECAL endcaps. The sampling design is motivated by the fact that a shorter light path through the crystal reduces the effect of the transparency loss induced by radiation. The light produced in
the crystals is wavelength-shifted and extracted towards the photosensors by capillaries or fibers running along the longitudinal direction of the channel (Fig. 6.6). R&D studies are ongoing to find the best wavelength-shifting and radiation-hard material.

Two scintillating materials under consideration are LYSO and CeF₃. Both have been studied in irradiation tests and have been shown to be able to withstand larger hadron fluences than PbWO₄ (Fig. 6.7) [108]. CeF₃ even shows a spontaneous recovery of the hadron damage [109] when irradiation stops (Fig. 6.5). The proposed absorber material is tungsten. This design aims at a stochastic term in the calorimeter resolution of about 10%. The small radiation length and Molière radius of this configuration lead to a very compact and granular design.

Two prototypes have been realized and successfully tested in beams. One uses LYSO crystals with fibers running through the channel; the other uses CeF₃ crystals with fibers running along chamfers located at the channel corners [98]. The following Sections provide an extensive description of the prototype channel using CeF₃ crystals.

### 6.2 Description of the CeF₃ prototype channel

The prototype channel is made of ten layers. Each layer consists of a 10 mm-thick CeF₃ scintillating crystal, a reflective foil on each side of the crystal, and a tungsten plate of 3.1 mm. Figure 6.8 (left) shows the stack of layers during the assembly. The total longitudinal depth corresponds to about 17 radiation lengths, that are sufficient to contain electromagnetic showers from electrons of up to 500 MeV. This design is adequate for a beam test with low energy electrons; for an application in CMS, more layers would be needed to ensure the longitudinal containment of high energy particles.

The Molière radius of this configuration is 23.1 mm. The transverse dimensions of the channel are 24 × 24 mm, approximately matching the Molière radius. Therefore, it is expected that a large portion of the shower energy is contained in the channel, though lateral losses are not negligible.
Figure 6.6: Sketch of a Shashlik configuration based upon interleaved W and LYSO scintillating crystal layers.

Figure 6.7: Induced absorption at the peak-of-emission wavelength for PbWO$_4$, LYSO and CeF$_3$, measured longitudinally through the crystals, as a function of integrated proton fluence, for various producers. [108]
At the edges of the crystals and tungsten layers, four 3 mm-wide chamfers (Fig. 6.8, right) allow for positioning wavelength-shifting fibers in a longitudinal orientation. The fibers used for the beam tests are of plastic type. They are characterized by a 0.02 mm-thick cladding layer surrounding the fiber core. The radiation hardness of these fibers is not adequate to sustain the conditions expected at the HL-LHC. Nevertheless, they are chosen as a temporary option for the beam tests, as their excitation wavelength matches the emission wavelength of CeF$_3$. The fibers are aluminized at the front end. The rear end of each fiber is coupled to a photomultiplier tube (PMT). Optical grease is used to maximize the optical coupling efficiency. The whole channel is then fixed inside an aluminum support structure. Figure 6.9 illustrates the setup.

In order to contain the lateral losses of shower energy, eight Bismuth Germanium Oxide (BGO) scintillating crystals from the L3 [110] experiment are placed around the CeF$_3$ channel to form a 3×3 matrix. The BGO crystal dimensions (22 mm at the front, 30 mm at the rear) approximately match the CeF$_3$ channel size. Each crystal’s rear face is optically coupled to a PMT of the same model used for the CeF$_3$ fibers.

A scintillating fiber hodoscope is placed in front of the matrix, at a distance of about 15 mm from it. The relative position of these two elements is fixed. The hodoscope consists of 8×8 pairs of plastic scintillating fibers of 1 mm diameter. Each fiber pair is separately readout by a multi-anode PMT. The transverse displacement between the CeF$_3$ axis and the hodoscope center is known with an accuracy of 0.1 mm. This allows to use the hodoscope for accurately aligning the CeF$_3$ to the beam axis.

Water cooled metal plates are used to stabilize the temperature of the matrix to 18 °C. The matrix and the hodoscope are housed in a light-tight box, equipped with a patch panel for connection to external readout electronics and power supplies.
6.3 Beam test with low energy electrons

6.3.1 The LNF beam test facility

The DAΦNE Beam Test Facility (BTF) [111] is located at the Laboratori Nazionali di Frascati, Italy. During the beam test of the CeF$_3$ prototype, electrons or positrons with nominal energies between 97 MeV and 497.4 MeV were provided in a bunched beam structure. Pulses with a duration of 10 ns are repeated 25 times per second. Each bunch contains from one up to a few electrons or positrons.

The structure of the facility is shown in Fig. 6.10. Electrons are first accelerated in a linear beamline, then they hit a copper target whose depth can be tuned from 1.7 to 2.3 radiation lengths. The target is used as an energy spreader. The secondary beam is extracted towards the beam test area by means of a bending dipole, and collimators are used to select the secondary beam energy with a precision of about 5 MeV at 497.4 MeV nominal energy. The transverse spread of the beam is measured to be about 4 mm in the horizontal direction and 3 mm in the vertical direction.

A periodic signal in phase with the LINAC radio-frequency is provided in the beam test area, and is used to trigger the data readout in coincidence with the delivery of a bunch. The readout uses NIM and VME crates and is based on charge-integrating ADC boards directly fed with the PMT signals. The ADC integration gate is 300 ns for the PMTs connected to the CeF$_3$ fibers, and about 1.5 $\mu$s for those connected to the BGO PMTs. This is done to take into account the different scintillation decay times of the two crystal types.

6.3.2 Calibration of the channel

A preliminary study of each fiber channel response has been performed with cosmic muons. The coincidence of light signals in the top and bottom BGO crystals in the $3 \times 3$ ma-
Figure 6.10: Sketch of the LNF beam test facility complex. [111]
The energy deposited by a minimum ionizing particle crossing few cm of CeF$_3$ is of a few MeV. Taking into account the CeF$_3$ light yield, the efficiency of the wavelength-shifting process (less than 1%) and the PMT quantum efficiency (about 7%), this leads to only one or two photoelectrons in the PMT. The distribution of the number of photoelectrons is Poissonian.

These events are used to calibrate the single photoelectron peak in units of ADC channels for the operational PMT high voltage. The ADC spectrum is fitted with the following function:

$$
\sum_{n=0}^{\infty} \frac{\mu^n e^{-\mu}}{n!} \frac{1}{\sigma_1 \sqrt{2\pi n}} \exp\left(\frac{-(x-nQ_1)^2}{2n\sigma_1^2}\right)
$$

(6.1)

where $Q_1$ and $\sigma_1$ represent the mean and width of the Gaussian response to a single photoelectron, $x$ is the ADC value, $\mu$ is the average number of photoelectrons (Poisson distributed). An example of the fit is shown in Fig. 6.11.

The relative calibration of the four fibers is then performed in beam events. The ADC spectrum is iteratively fit with a Gaussian distribution. The peak positions are measured for each fiber and aligned by correcting the ADC value with a multiplicative factor. Figure 6.12 shows that the spectra of the four fiber channels have compatible shapes after the correction has been applied. The secondary peaks corresponding to bunches with more than one electron also result to be well aligned. Compatible calibration factors are derived at different beam energies.

### 6.3.3 Event selection

In order to study the linearity of the response and the energy resolution, it is necessary to identify the beam bunches containing one electron hitting the central region of the CeF$_3$ channel front face. The front plastic scintillator and the fiber hodoscope are used together to maximize the background rejection efficiency. Figure 6.13 (left) shows that the magni-
tude of the integrated scintillator signal ($A_{SC}$) clearly separates single and double electron bunches. A requirement of $500 < A_{SC} < 2000$ ADC channels is thus introduced.

The signals from the hodoscope fibers are individually integrated. A threshold is applied on each channel readout to discriminate signals from the noise. Finally, adjacent fibers where a signal is detected are clustered. The cluster size is limited to four fibers. Figure 6.13 (right) shows the distribution of cluster centers of gravity for the horizontal and vertical fiber arrays. Only events with at least one reconstructed cluster are retained. In this way, electrons impinging on the central $8.5 \times 8.5 \text{ mm}^2$ region of the CeF$_3$ channel are selected.

Figure 6.14 shows the sum of the four CeF$_3$ channel readouts before and after the event selection is applied. Without selection, bunches with two and three electrons significantly contaminate the distribution. The selection requirement based on the scintillator amplitude removes these events, but an anomalous low tail is still present. This is due to events where the electron hits the channel close to its boundary. For these events, lateral containment fluctuations are larger and they are rejected by the requirement on the central hodoscope clusters.

### 6.3.4 Measurement of the response linearity

The response linearity of the prototype channel is tested with five different nominal beam energies: 98, 147, 197, 295 and 491 MeV. The energy spectrum is iteratively fitted using a Gaussian distribution with a limited fitting range in single-electron events. An alternative fitting method using Eq. 6.1 in multi-electron events is also used, and the difference in the results is quoted as a systematic uncertainty. The fitted response in data is compared to values obtained with a GEANT4 simulation. The simulation takes into account the relative light yield of each CeF$_3$ layer, preliminarily measured with a $^{22}$Na source. No light collection inefficiency is simulated.
Figure 6.13: Left: CeF$_3$ response as a function of the integrated signal in the front scintillator. Right: distribution of the centers of gravity of fiber hodoscope clusters. [99]

Figure 6.14: CeF$_3$ energy spectrum without any event selection (black), with single electron selection using the front scintillator (red), and with the requirement of at least one fiber hodoscope cluster (blue). [99]
Figure 6.15: Left: CeF$_3$ response as a function of the beam energy, in data and simulation. Right: comparison of the CeF$_3$ response in simulation for the following geometries: $1 \times 1$, $5 \times 5$ and $5 \times 5$ with a crystal in front. No selection is applied on the number of hodoscope clusters for this study. [99]

Figure 6.15 (left) shows the results of this study. The measured channel response as a function of the beam energy is fitted with a straight line. Data and simulation are found to be in good agreement. The intercept of the fit is not compatible with zero, even in simulation alone (Fig. 6.15, right). This is due to lateral containment losses, which has been verified by repeating the simulation using a $5 \times 5$ CeF$_3$ matrix instead of a single channel. Moreover, the channel is assembled with an absorber layer in front. If the channel is simulated instead with a CeF$_3$ crystal in front, the expected intercept is compatible with zero, because even very low energy deposits can be detected.

### 6.3.5 Measurement of the energy resolution

The resolution is extracted from the width of the Gaussian fit in single-electron events. Figure 6.16 shows the comparison of data and $1 \times 1$ CeF$_3$ simulation. The resolution obtained with a $5 \times 5$ CeF$_3$ simulation is also shown. The simulated resolution is corrected to account for statistical fluctuations in the number of photons produced in the crystals and wavelength-shifted in the fibers (photostatistics effect). The data and simulation are in agreement.

This builds up confidence on the extrapolation to the $5 \times 5$ CeF$_3$ matrix and that it can be used to remove the lateral containment fluctuations and predict the expected performance of a calorimeter based on this design. Because of the low beam energy, only the stochastic term of the energy resolution is measured. It is found to be $9.70 \pm 0.02\%$.

Finally, the energy resolution is studied as a function of the beam impact angle on the channel front face. This study is motivated by the expected arrangement of the channels in the upgraded ECAL endcaps. In the present design, they are parallel to the beam axis, thus not pointing to the interaction region at the center of CMS. As a consequence, the
impact angles for particles produced in the collisions are expected to reach 10° in the region at lowest pseudorapidity.

Data with impact angles up to 5° are obtained by rotating the light-tight box on the beamline. The pivot point corresponds to the position of the hodoscope in front of the channel. Figure 6.17 shows that the simulation is in agreement with the measured values. The main reason for the worsening of energy resolution at higher angles is the increase of lateral shower energy losses. The 5×5 matrix simulation is immune to this effect, showing a negligible dependency up to 10°.

6.4 Beam test with high energy electrons

6.4.1 The CERN SPS H4 beamline and experimental setup

The CERN Super Proton Synchrotron (SPS) accelerates protons up to an energy of 450 GeV. The proton beam is extracted and directed on a series of beryllium and lead targets. The electrons produced in the fixed target interactions are selected in momentum using dipole magnets and collimators. Further beam optical elements, such as quadrupoles, are used to focus the beam in the H4 beam test area.

During the test of the CeF₃ prototype channel, electrons at a rate of a few kHz have been used. Data have been taken at the nominal energies of 10, 15, 20, 50, 100, 150 and 200 GeV. The electron beam is delivered with a cycle of two spills with a duration of about 5 s, and is not bunched. As a consequence, the trigger relies on fast plastic scintillators placed at the end of the beamline. The probability of two electrons being delivered within one ADC integration gate is negligible.

In order to ensure a sufficient longitudinal containment for high energy electrons, the channel was extended by five additional layers. In this way, the channel depth reaches approximately 25 radiation lengths. For the same reason, the lateral containment is improved by extending the matrix to a 5×5 geometry (from the 3×3 used at low energy).
Figure 6.17: CeF₃ resolution as a function of the beam impact angle on the front face of the channel, in data and simulation. [99]

with the addition of 16 BGO crystals.

The light-tight box is placed on a remotely controlled movable table. The beamline is also equipped with four delay wire chambers (two for the horizontal, two for the vertical direction). They are connected to a time-to-digital converter (TDC) VME board. They provide a beam impact point determination with a nominal resolution of less than 0.2 mm. Moreover, two fiber hodoscopes are located at a distance of about 3 m and 6 m from the experimental box. Each of them consists of 64×64 scintillating fibers of 0.5 mm diameter. Their signals are acquired using a pattern unit VME board.

The DAQ infrastructure has been deeply revisited with respect to the LNF beam test to accommodate the readout of all the elements described above at high rate. It is based on three VME readout crates and one NIM crate used for triggering. The CeF₃ channel fiber signals are readout by a digitizer VME board at a sampling rate of up to 5 GHz. A measured pulse shape is shown in Fig. 6.18. The same board is used to read the signals from the central 2×2 fibers of the hodoscope placed in front of the channel. The VME crate hosting the digitizer board is placed inside the experimental area, to minimize the distortion of signals in the connecting cables. All BGO signals are integrated in charge using an ADC board.

Each VME crate is controlled by a dedicated computer readout node connected to a communication network. Each node is controlled by a finite-state machine (FSM). The logic of trigger veto during the readout time ("busy logic") is implemented using NIM boards that receive signals from each readout node. The busy time is dominated by the readout time of the digitizer board, which is of the order of ms.

An event builder node receives and processes data fragments from every readout node at the end of the SPS extraction spill. An online data quality monitoring routine is then
run, and the results made available to the operator by means of a graphical user interface (GUI). The GUI also features a visible and audible alarm system, and is interfaced with:

- a run control process, to supervise the data acquisition;
- a database, to keep track of the data taking conditions;
- the moving table, to set the position of the box;
- temperature sensors and webcams installed in the experimental area.

6.4.2 Measurement of the response uniformity

Figure 6.19 shows the map of the sum of the four fiber channel signals as a function of the impact point. This study is made possible by the beam tracking detectors (wire chambers and hodoscopes). The nominal dimensions of the channel are overlaid in yellow. The map is centered on the nominal beam axis, and this confirms that the alignment of the channel with the beam reference system is performed with an accuracy better than 1 mm.

The response is uniform in the central region, while it decreases close to the boundaries because of lateral losses. The distribution of the average response in a centered 6 mm-wide strip along the $x$ or $y$ direction is shown in Fig. 6.20, as a function of the impact point. The shape in data is compared to the one predicted by a GEANT4 simulation analogous to the one used for the LNF beam test. It does not account for the propagation of light inside the crystal, thus it cannot model the light collection inefficiency. Nevertheless, it agrees with data within 5% even in proximity of the module boundaries. This suggests that the effect of lateral energy losses is dominant with respect to the light collection inefficiency.

When studying the response map of a single fiber, the dependence of the light collection efficiency on the impact point has a larger effect. Actually, Fig. 6.21 shows that the response increases when the impact point gets closer to the fiber. Further studies are envisaged to determine if this information can be used to constrain the impact point position from calorimeter measurements alone.
Figure 6.19: CeF$_3$ response map as a function of the impact point on the front face of the channel, at 50 GeV (left) and 100 GeV (right) beam energy. The nominal channel dimensions are overlaid in yellow. [100]

Figure 6.20: CeF$_3$ response as a function of the $x$ (left) and $y$ (right) coordinate. For each plot, events are selected requiring that their impact point lies within a horizontal (resp. vertical) strip of 6 mm centered on the channel axis. [100]
Figure 6.21: CeF$_3$ single fiber response at 100 GeV beam energy, as a function of the impact point. The fiber under consideration is located in the lower right corner. [100]

6.4.3 Measurement of the resolution at high energy

The study of resolution as a function of the beam energy relies on the linearity of the channel response. The PMTs exhibit a strong non-linearity at energies of more than 50 GeV and high voltage of 950 V (Fig. 6.22). A scan of the channel response as a function of beam energy and PMT high voltage is performed at settings of 600, 700 and 950 V. Linearity is mostly recovered with the lowest gain, but this dramatically decreases the signal to noise ratio. The contribution of the electronic noise cannot be neglected, and has to be subtracted from the measured resolution to obtain the intrinsic channel performance.

Only events in the central 6×6 mm region are used, to reduce the fluctuations from lateral containment. The resolution is modeled with an electronic noise term plus an intrinsic performance term. Figure 6.23 shows the intrinsic resolution as a function of energy, after the noise has been subtracted. At the high beam energies explored here, both the stochastic and constant term of the resolution can be measured. Data and simulation are in agreement.

In analogy with what has been done for the LNF beam test, the simulation is validated and can be used to predict the resolution of a 5×5 CeF$_3$ matrix. This yields a stochastic term of 9.54 ± 0.13%, in agreement with the previous results, and a constant term of 0.33 ± 0.04%.

In conclusion, this test proves that a sampling calorimeter based on CeF$_3$ scintillating crystals and readout with wavelength-shifting fibers has the potential to achieve an energy resolution given by a stochastic term below 10%/√E(GeV), and a constant term below 1%. Therefore, it represents a very promising candidate for the upgrade of the CMS ECAL endcaps.
Figure 6.22: CeF₃ linearity curve at different high voltages. [100]

Figure 6.23: CeF₃ resolution as a function of the beam energy. The noise due to the readout electronics has been subtracted. Data and simulation (1×1 and 5×5 matrices) are compared. The 5×5 simulated points are fitted with a stochastic plus a constant term. [100]
Summary

A measurement of differential cross sections for diphoton production in proton-proton collisions has been presented. The dataset used has been collected with the CMS experiment at the LHC in 2011, and corresponds to an integrated luminosity of $5.0 \text{fb}^{-1}$. The analysis is based on a template fit to statistically separate the signal from the jet background. The particle-flow photon isolation is used as discriminating variable. A detailed understanding of photon reconstruction and identification is required to perform the measurement with a small systematic uncertainty.

The out-of-time pileup and electronic noise have been shown to have a non-negligible effect on several photon observables measured using the CMS electromagnetic calorimeter. Their simulation has been tuned using soft inelastic collision data. An accurate description of these phenomena was required for this analysis and yielded strong benefits to all CMS photon analyses.

Isolation templates were built from data to minimize the systematic uncertainty in the measured cross section. This was made possible by a novel technique, denoted as footprint removal, used to remove the double counting of the photon energy deposit and its isolation sum. All sources of correlation between the isolation sums of the two photons were taken into account. The measurement was extended to the interesting region of the phase space where the isolation cones partially overlap.

These results represent a stringent test of quantum chromodynamics. In the region of low diphoton invariant mass, and corresponding low azimuthal angular separation, the contribution of higher-order QCD corrections to the cross section is significant. In this region of phase space, the $2\gamma_{\text{NNLO}}$ prediction at next-to-next-to-leading order in perturbative QCD is needed to reach an agreement with data, while NLO calculations fail to describe the differential cross section. Predictions obtained with the multi-leg matrix element plus parton shower generator SHERPA are also in good agreement with data. Furthermore, the modeling of diphoton production in association with jets was tested. This is a final state of interest for studying the VBF production channel of the Higgs boson.

Finally, the evolution of the calorimeter performance under irradiation and the plans for an upgrade of its endcaps was presented. The results obtained from beam tests of a CeF$_3$ sampling calorimeter prototype channel show that this is a promising design for electromagnetic calorimetry under High-Luminosity LHC conditions.
Vorrei innanzitutto esprimere la mia gratitudine al Prof. Günther Dissertori per la sua guida, i suoi consigli e la fiducia che ha riposto in me durante il periodo che ho trascorso come suo studente di dottorato presso il Politecnico Federale di Zurigo.

Ringrazio poi in modo davvero speciale il Dr. Mauro Donegà, per il suo supporto costante e gli innumerevoli preziosi insegnamenti che mi ha fornito giorno per giorno. Lavorare insieme ha contribuito in modo decisivo a fare di questi anni un’esperienza fruttuosa sia sul piano professionale che su quello umano.

Vorrei inoltre ringraziare gli altri membri del gruppo dell’ETH, in modo particolare il Prof. Rainer Wallny (co-referente di questo lavoro di tesi) e coloro i quali hanno lavorato a più stretto contatto con me (Nicolas, Pasquale, Milena, Francesca, Kostas, Andrea), per la disponibilità che hanno dimostrato nei miei confronti e i rapporti di amicizia che si sono creati tra noi. Estendo lo stesso ringraziamento ai membri del gruppo di Roma e degli altri istituti che fanno parte della collaborazione CMS.

Esprimo la mia riconoscenza verso tutti i professori che fin dai tempi del liceo, e in seguito alla Scuola Normale, mi hanno trasmesso la passione per la Fisica, e in modo particolare al Prof. Gigi Rolandi.

Infine, non posso dimenticare che il mio soggiorno a Ginevra è stato possibile grazie a tutti coloro che mi sono stati vicini. Ringrazio quindi la mia famiglia, che mi supporta sempre, e tutti i miei amici, quale che sia il posto che hanno scelto per vivere: sotto il mio stesso tetto, in una stanza forse simile alla mia, ma in una nazione diversa, oppure dall’altra parte dell’oceano.
References


167


[79] CMS $H \rightarrow \gamma \gamma$ working group. Search for a Higgs boson decaying into two photons in proton-proton collisions recorded by the CMS detector at the LHC. CMS internal Analysis Note 2011/129, 2011.


