SAPIENZA
UNIVERSITÀ DI ROMA

Facoltà di Scienze Matematiche Fisiche e Naturali

Ph.D. Thesis

Measurement of the mass of the Higgs Boson in the two photon decay channel with the CMS experiment.

Candidate

Shervin Nourbakhsh

Matricola

1106418

Thesis advisors

Dott. Riccardo Paramatti

Dott. Paolo Meridiani

Year 2012/2013

Contents

Introduction

The Standard Model (SM) of particle physics describes elementary particles and their interactions in the contest of the Quantum Field Theory. It has been very successful in describing high energy measurements at the actual experimental limits. The electro-weak sector of the theory is spontaneously broken by an additional scalar field (the Higgs field) with a non void expectation value for the ground state. The mass of the particles is given by their interaction with the Higgs field, whose quantum is the Higgs boson. The Higgs boson of the SM is a scalar particle and with spin 0 with a coupling to the other particles proportional to their masses. The Higgs boson mass is instead a free parameter of the theory.

The 4th July 2012, at the Large Hadron Collider (LHC) the A Toroidal LHC ApparatuS experiment (ATLAS) and the Compact Muon Solenoid (CMS) collaborations have announced the discovery of a new boson.

In this thesis, the measurement of the properties of the new boson in the di-photon decay channel, with the CMS data, is presented. The objective of the analysis is the study of the new boson properties in order to assess the compatibility with the SM predictions.

The objective of this thesis is to present the measurement of the parameters of the boson in the di-photon final state, with a special highlight to the close relation between the analysis sensitivity and the Electromagnetic Calorimeter (ECAL) performance, and illustrate in details the mass measurement and its uncertainties.

The data collected in 2011 at 7 TeV center-of-mass energy and in 2012 at 8 TeV by the CMS experiment are used in this analysis.

A theoretical introduction to the Higgs boson, its production mechanism at LHC and decay modes is given in the first chapter. An overview of the Higgs boson properties that is possible to study in the di-photon final state is presented. Notions about the LHC and the proton-proton physics environment are also given.

Chapter 2 is focused on the CMS detector. Particular attention is devoted to the ECAL description given its central role in the $H \to \gamma\gamma$ analysis sensitivity.

The Higgs signal is expected as a resonance in the di-photon invariant mass over a smoothly falling background. According to the SM predictions, the natural width of the Higgs boson is negligible with respect to the experimental resolution. An excellent energy and direction resolution of the photons is therefore crucial. sensitivity.

The electron and photon reconstruction is described in Chapter 3, starting from the ECAL energy measurement algorithms, the ECAL *in situ* calibration and the energy corrections for electrons and photons.

My personal contribution to the ECAL calibration activities in the LHC Run1 is to validate, through the evaluation of the resolution improvements in $Z \to e^+e^-$ events, each step of the calibration procedure. The ECAL conditions for the legacy reprocessing of 7 and 8 TeV data have been validated with the tools I developed during this thesis.

Chapter 4 is focused on the measurement of the energy scale and resolution for electrons and photons. The tools I developed are described and the results I obtained with the legacy reprocessing for the paper in preparation are reported. The origin of discrepancies between the data and the simulation are also discussed, and the corrections needed in the $H \to \gamma\gamma$ analysis to compensate such discrepancies are shown.

The strategy and the main details of the $H \to \gamma\gamma$ analysis are presented in Chapter 5 with the corresponding results.

Chapter 6 is dedicated to the mass measurement and the discussion of the results and uncertainties. I've contributed to estimate the uncertainties due to the energy scale and resolution and to reduced them improving the correction derivation described in Chapter 4.

In this thesis, the most recent CMS public results on the $H \to \gamma\gamma$ analysis are shown, whilst at the moment of writing, the reprocessing of the data with final calibration is being used in the $H \to \gamma\gamma$ analysis for final results to be published. The systematic uncertainties presented in this thesis are however the one related to the most recent public results.

Chapter 1

Standard Model Higgs and LHC physics

1.1 Standard Model Higgs boson

The Standard Model (SM) [1, 2, 3] has been very successful in explaining high-energy experimental data. A question that remains open is the origin of the masses of the W and Z bosons – the electroweak symmetry breaking mechanism. The solution in the SM is the Higgs mechanism [4, 5] introducing an additional scalar field whose quantum, the Higgs boson, should be experimentally observable. Indeed, the discovery of a new boson with a mass of 125 GeV has been announced by the A Toroidal LHC ApparatuS experiment (ATLAS) and the Compact Muon Solenoid (CMS) collaborations on the 4th of July 2012 using the data from the proton-proton collision of the Large Hadron Collider (LHC) at 7 and 8 TeV center of mass energy. The properties of the new boson are compatible with the ones expected for the SM Higgs boson with the precision currently reached by both experiments. Its properties are now under deeper investigation in order to verify or exclude deviations from the predictions and clarify its role in the electroweak symmetry breaking mechanism.

1.1.1 Overview

The SM of particle physics describes all known elementary particles and the electromagnetic, weak and strong interactions between them [1, 2, 3]. It is a mathematical model developed in the context of the Quantum Field Theory unifying under a common theoretical framework three of the four fundamental forces of Nature: the strong, the electromagnetic and the weak forces. The SM does not describe gravity. In such a representation, all particles are interpreted as excitations of relativistic quantum fields. Ordinary matter is made of six spin $\frac{1}{2}$ particles divided into leptons interacting only by electromagnetic and weak forces and six spin $\frac{1}{2}$ particles called quarks subjected also to the strong force. Fermions (leptons and quarks) are classified in three generations and grouped in doublets; each quark is also repeated in three "colours". Elementary particles interact with each other via exchanges of gauge bosons, also called the mediators of the corresponding fields.

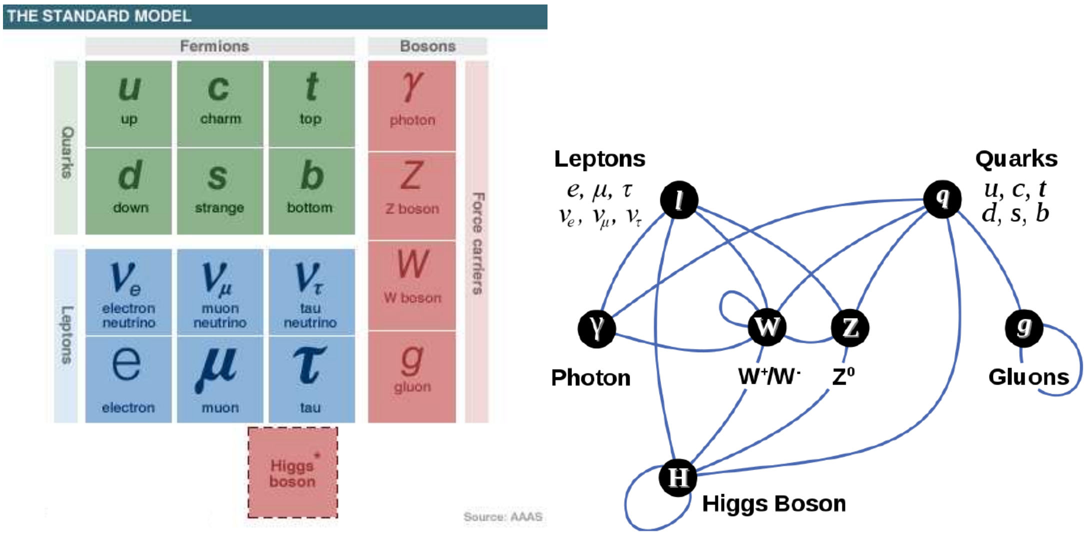
The SM elementary particles are shown in Fig. 1.1a and their interactions are schematically shown in Fig. 1.1b

The motion and interaction of particles are derived from the Lagrangian of the model using the Minimal Action Principle. The Standard Model Lagrangian is required to be invariant to the following gauge transformation symmetry groups:

$$SU(3)_c \times SU(2)_L \times U(1)_Y \tag{1.1}$$

The $SU(2)_L \times U(1)_Y$ invariance group represent the theoretical unification of the electromagnetic and the weak forces in the electroweak interaction. The Glashow[1], Weinber[2] and Salam[3] theory derives from this symmetry group the existence and behaviour of the electroweak boson mediators (the Z^0 and $W^{\pm}$ bosons) and the electromagnetic carrier, the photon (γ).

The strong force carriers are derived by the $SU(3)_c$ term. The quark model was proposed by Gell-Mann[6] in 1964. The idea of the "colour" quantum number was introduced by Han and Nambu[7] in 1965 to avoid the apparent paradox that the quark model seemed to require a violation of the Pauli exclusion principle in order to describe the hadron spectroscopy. The Quantum Chromo-Dynamics (QCD), the strong interaction sector of the SM, was then quantized as a gauge theory with $SU(3)_c$ symmetry in 1973 by Fritzsch[8].



Source: AAAS

(a) Elementary particles

(b) Elementary particle interactions

Figure 1.1: The Standard Model elementary particles and associated fields are divided into ordinary matter particles which are fermions (left side) and bosonic force carriers (right side). Fermions are divided into quarks (subjected to the strong force) and leptons (neutral for the strong force). Quarks and leptons are grouped in three generation of doublets. The particles masses are given by the interaction with the Higgs boson shown in the bottom.

Due to the very high magnitude of the strong force compared to the others, it's not possible to find free quarks even at very high energies. Thus, the production of quarks in the collisions is revealed as jets of particles originated by the hadronization process.

1.1.2 Spontaneous symmetry breaking and the Higgs boson

The spontaneous symmetry breaking mechanism [4, 5, 9] provides the way to give mass to particles preserving the gauge invariance of the Lagrangian.

The symmetry of the electroweak (EW) sector of the theory ($SU(2) \times U(1)$) is broken by additing a complex scalar Higgs doublet

$$\phi \equiv \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{1.2}$$

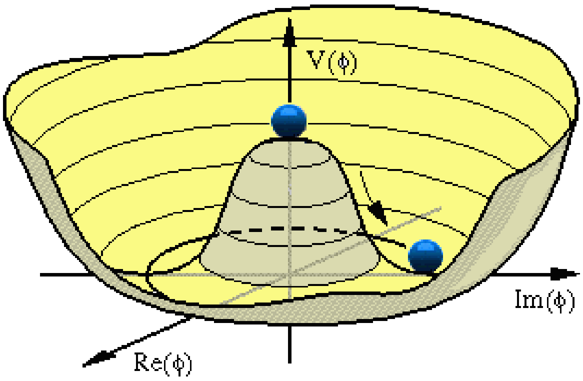


Figure 1.2: Qualitative representation of Higgs potential

and a potential term in the lagrangian:

$$V(\phi) = \mu^2 \phi^\dagger \phi + \frac{\lambda^2}{2}(\phi^\dagger \phi)^2 \tag{1.3}$$

A degenerate set of minima in this potential are produced for $\mu^2 < 0$ and a positive quartic coupling $\lambda > 0$ as illustrated by Fig. 1.2.

The Higgs field has a non-zero vacuum expectation value:

$$\langle 0|\, \Phi \,|0\rangle = \frac{v}{\sqrt{2}}; \qquad v = \sqrt{\frac{-\mu^2}{\lambda}} \tag{1.4}$$

breaking the EW gauge symmetry.

The boson masses in the EW sector are given (at the lowest order in perturbation theory) by

$$\begin{aligned} M_H &= \lambda v \\ M_W &= \frac{v}{2} g = \frac{ev}{2\sin\theta_W} \\ M_Z &= \frac{v}{2}\sqrt{g^2 + g'^2} = \frac{M_W}{\cos\theta_W} \\ M_\gamma &= 0 \end{aligned} \tag{1.5}$$

where g and g' are the couplings respectively with the $SU(2)$ and with the $S(1)$ sectors.

The mass of the fermions are obtained introducing a Yukawa interaction term which couples a left-handed fermionic doublet ψ^L, a right-handed singlet ψ^R and the Higgs doublet.

1.1.3 Higgs boson production

In the Standard Model, Higgs boson production in proton-proton collisions include four main mechanisms: gluon fusion ($gg \to H$), vector boson fusion ($qq \to H + 2jets$), associated production of a Higgs boson with a W or Z boson (Higgs strahlung), and associated production with a $t\bar{t}$ pair ($t\bar{t}$ fusion)[10, 11]. The gluon fusion is the dominant production mode at the LHC with a cross section 10 times higher then the others. Two gluons from the colliding protons interact with a quark loop producing the Higgs boson in the final state. In the other production modes, the Higgs boson is produced in association with other particles: two quarks boosted in the proton beam direction in the vector boson fusion, a W o a Z boson in the Higgs strahlung or a pair of top quarks in the $t\bar{t}$ fusion. Figure 1.3 shows the Feynman diagrams for these four production mechanisms.

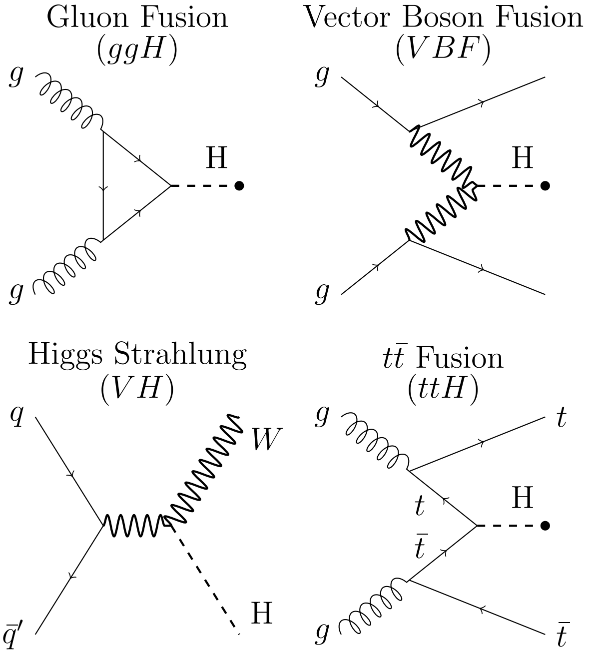


Figure 1.3: Feynman diagrams for the four main production mechanisms of the Standard Model Higgs boson at the LHC.

Figure 1.4 shows the cross section of each mechanism as a function of the Higgs boson mass at $\sqrt{s} = 7$ TeV and 8 TeV, the center of mass energy of LHC Run 1 data.

The single production mechanisms can be studied at the LHC exploiting the peculiarity of their final states.

1.1.4 Higgs boson decays

At leading order, the Standard Model Higgs boson can decay into pairs of fermions and W or Z bosons. Since the coupling of the Standard Model Higgs boson to fermions is proportional to the fermion mass, the decay branching ratio to any fermion is proportional to the square of its mass. Therefore, for m_{H} below the $2m_t$ threshold, the primary

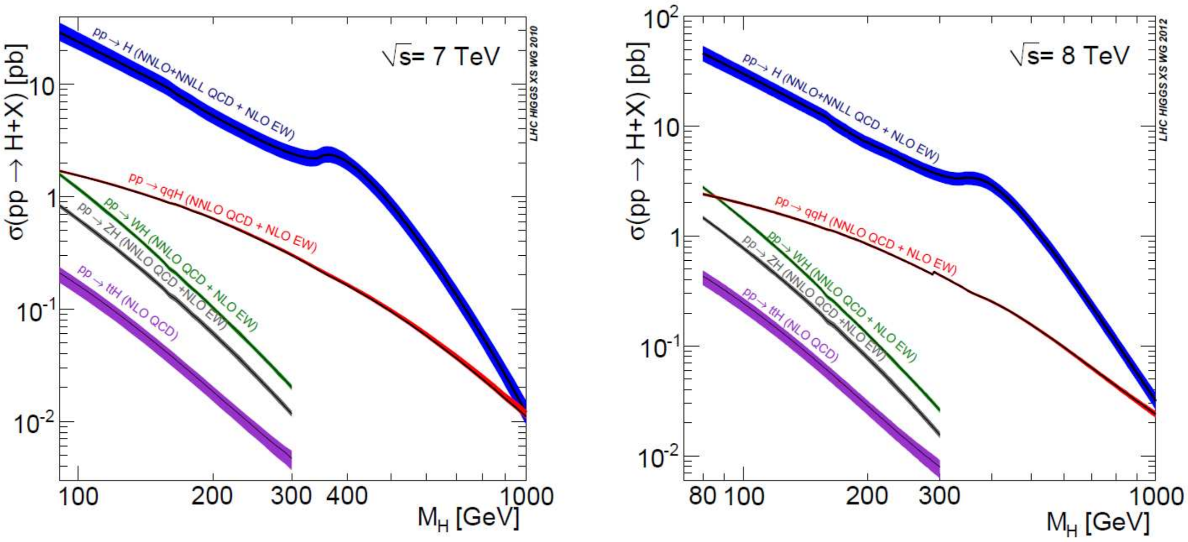


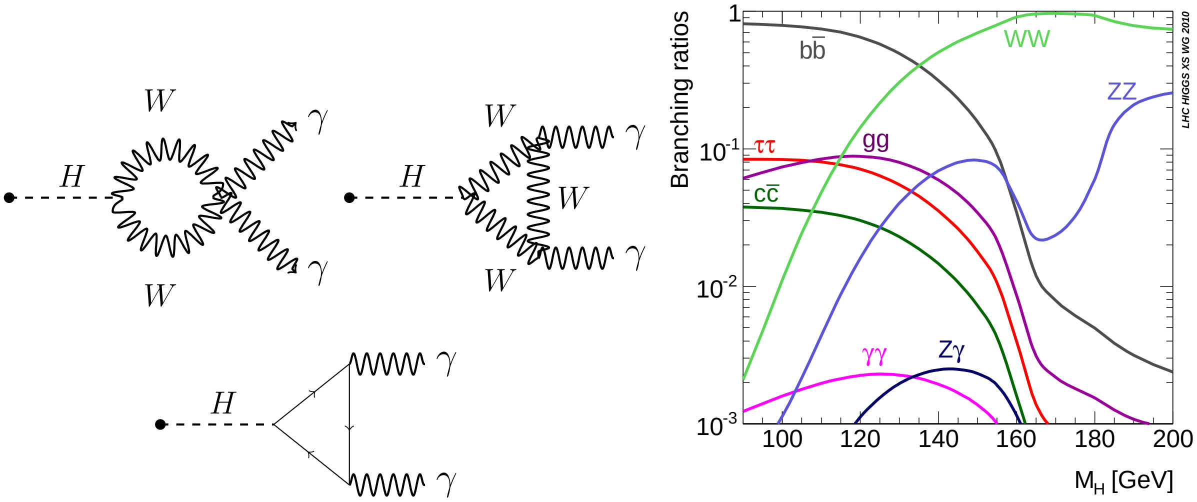
Figure 1.4: Theoretical predictions for the Higgs boson production cross sections in proton proton collisions at the LHC center of mass energies: $\sqrt{s} = 7$ TeV (left) and 8 TeV (right).

fermionic decay products are $b\bar{b}$ pairs, with smaller contributions from $c\bar{c}$ and $\tau\bar{\tau}$ pairs. For the bosonic decays, for m_{H} below the $2m_W$ or $2m_Z$ threshold, the Higgs boson can decay into a pair of off-shell Ws or Zs, which subsequently decay into two pair of fermions. Loop-induced decay into a pair of photons or gluons are also important in the low-mass region. Although the low branching ratio, the di-photon decay mode represent one of the most important channel both for discovery and study of the properties because of its very clear experimental signature represented by a narrow peak in the di-photon invariant mass over a smoothly falling background. The Higgs boson can decay into two photons via a fermion loop, dominated by the top quark contribution, and via a W loop (Figure 1.5a). The interference of the 2 terms is destructive in the SM assuming the standard positive coupling to the top quark.

The various Higgs boson decay branching fractions as functions of m_{H} are shown in Figure 1.5b.

1.1.5 Higgs boson property study in the two photon final state

The discovered Higgs boson is studied at the LHC in all the main decay modes ($b\bar{b}$, WW, $\tau\bar{\tau}$, ZZ^*, $\gamma\gamma$).



(a) Leading order Feynman diagrams for a Standard Model Higgs boson decaying into two photons.

(b) Theoretical predictions for the Higgs boson decay branching fractions.

Figure 1.5

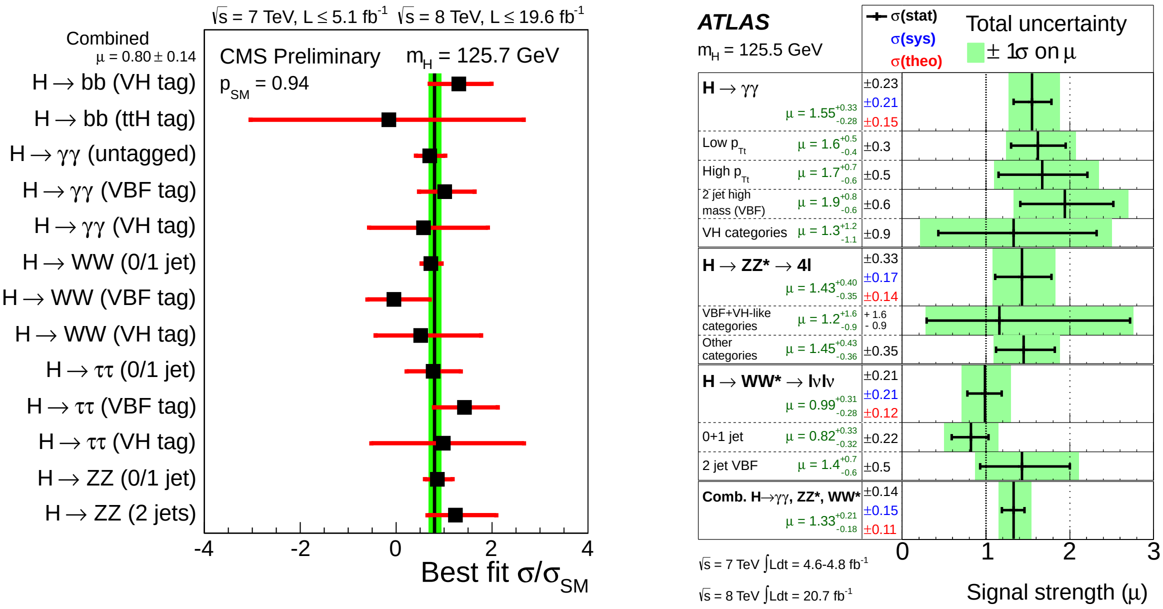
The Higgs couplings to massive particles is measured in all the decay channels.

The The Higgs mass is measured in the most sensitive channels ($H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ \ell$) and limits on its intrisic width are also set.

The signal strength ($\mu = \frac{\sigma}{\sigma_{SM}}$, ratio between the measured cross-section and the SM expectation) is shown in Fig. 1.6 for the different decay channels studied by ATLAS and CMS with 7 and 8 TeV data. The analysis in the single decay modes are performed defining high purity categories for the different production modes exploiting further their peculiar final state topologies (discussed in Sec. 1.1.3).

Higgs couplings

One important aspect to investigate in order to assess the compatibility of the discovered Higgs boson with the SM expectations is the study of the couplings with the other massive particles. The couplings to fermions and boson are studied using all the possible decay modes (coupling with different particles in the decay) and exploiting the additional information on the production given by the exclusive categories (coupling with particles in the production). With the available data, the limits on the couplings are set using simplified



(a) (b)

Figure 1.6: Values of $\mu = \frac{\sigma}{\sigma_{SM}}$ for single channels. The CMS results are shown on the left. The vertical band shows the result obtained combining all the channels. On the right the ATLAS results.

models [12], scaling, for example, the coupling with all the fermions by the same factor k_F and the coupling with all the bosons by the same factor k_V. Deviations from the SM prediction ($k_V = 1, k_F = 1$) can furthermore discriminate between models Beyond the Standard Model or constraint their free parameters.

The study in the two photon final state is able to discriminate between the relative sign of k_V and k_F because of the decay loop where quarks (almost t) and bosons (W) enter with opposite sign in the amplitude calculation.

With new data will be possible to split further the coupling study with different scaling factors for up and down quarks, etc.

Higgs mass and width

The Higgs boson mass is measured at the LHC with high precision (better than 1%) both in the di-photon final state and in the $H \to ZZ^* \to 4\ \ell$ decay mode.

The Higgs mass depends on the unknown coupling λ in the Higgs potential, and

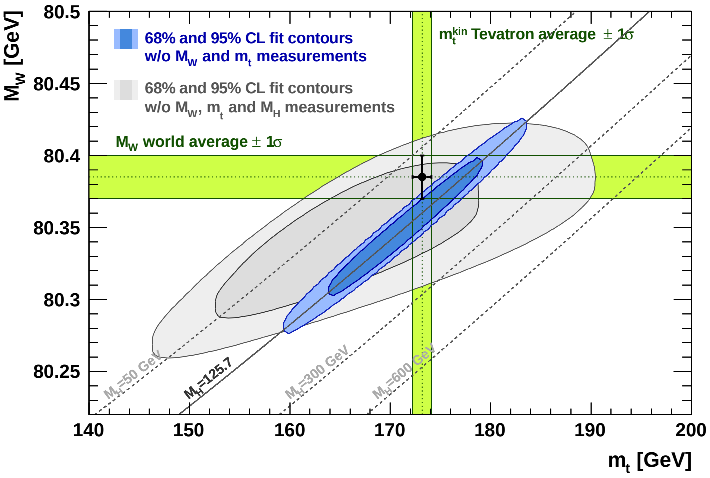


Figure 1.7: Contours of 68% and 95% CL obtained from scans of fixed M_W and m_t. The blue (grey) areas illustrate the fit results when including (excluding) the new M_H measurements. The direct measurements of M_W and m_t are always excluded in the fit. The vertical and horizontal bands (green) indicate the 1σ regions of the direct measurements.

therefore cannot be predicted. However some constraints can be fixed on a theoretical basis. Indirect information on the Higgs mass was extracted from Higgs loops affecting the values of Z boson asymmetry observables and the W mass. Assuming the new discovered particle to be the SM Higgs boson, all fundamental parameters of the SM are known allowing, for the first time, to overconstrain the SM at the electroweak scale and assert its validity.

Figure 1.7 displays CL contours of scans with fixed values of M_W and m_t, where the direct measurements of M_W and m_t were excluded from the fit. The contours show agreement between the direct measurements (green bands and data point), the fit results using all data except the M_W, m_t and M_H measurements (grey contour areas), and the fit results using all data except the experimental M_W and m_t measurements (blue contour areas). The observed agreement again demonstrates the consistency of the SM. A more accurate measurement of the mass can then also be used to have an indirect measurement on the top or the W mass.

The mass measurements by ATLAS and CMS agree within 1.3σ with the indirect determination $M_H = 94^{+25}_{-22}\,\text{GeV}$ [13].

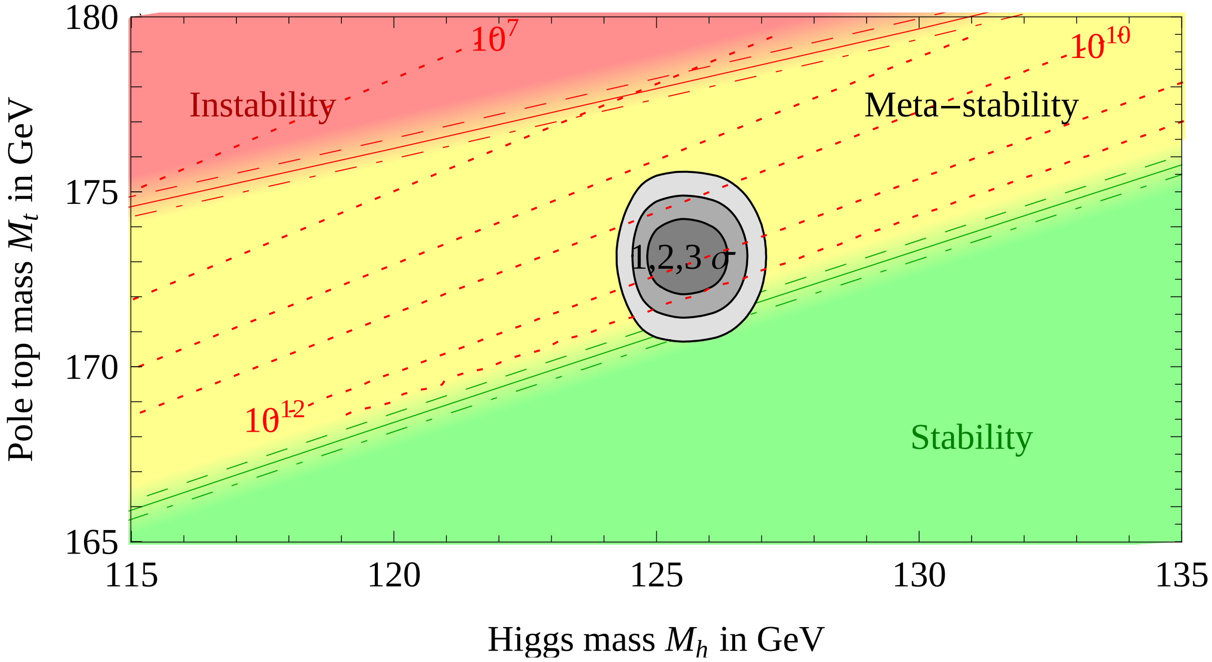
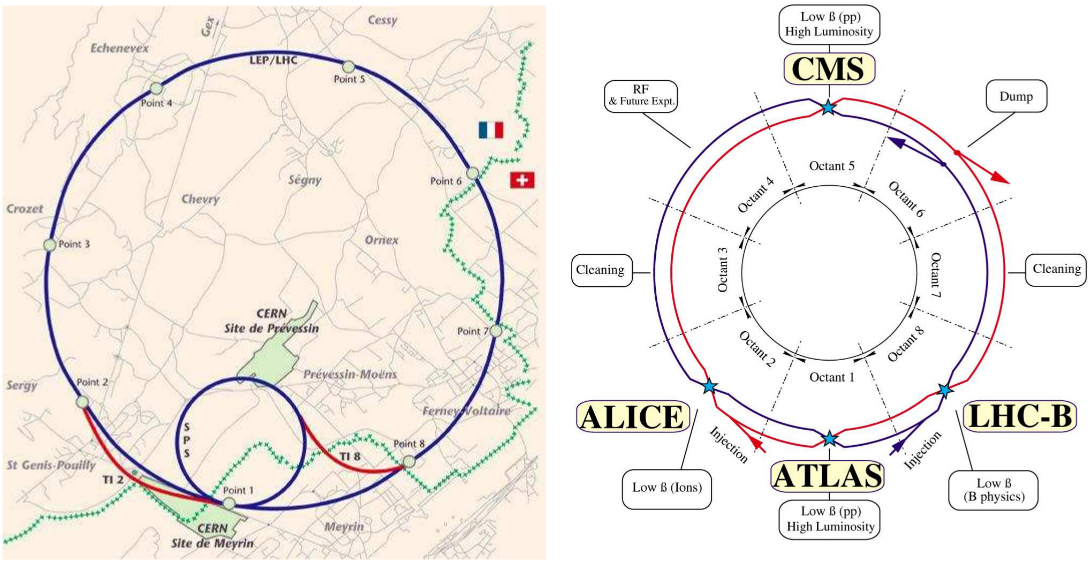


Figure 1.8: Regions of absolute stability, meta-stability and instability of the SM vacuum in the (M_t, M_H) plane. The gray areas denote the allowed region at 1, 2, and 3σ by the experimental data on M_H and M_t. The three boundary lines correspond to $\alpha_s(M_Z) = 0.1184 \pm 0.0007$, and the grading of the colors indicates the size of the theoretical error.

A precise measurement of the Higgs boson mass gives also information on the vacuum stability at the Plank scale [14]. The regions of absolute stability, meta-stability and instability of the SM vacuum in the (M_t, M_H) plane are shown in Fig. 1.8. With the most recent measurement by the ATLAS and CMS a meta-stable vacuum is preferred, but a stable vacuum is not excluded. A precise measurement of the Higgs boson mass can discriminate between a meta-stable and a stable vacuum.

The SM Higgs boson natural width for a Higgs boson of 125 GeV is of the order of $O(10\text{ MeV})$, well below the experimental resolution. The interference between the Higgs resonance in gluon fusion and the continuum background amplitude for gluon pair to photon pair[15] produces an apparent shift of the Higgs mass by around 100 MeV in the SM in the leading order approximation. The apparent mass shift can be experimentally observable and provides a way to measure, or at least bound, the Higgs boson width at the LHC through “interferometry”. At Higgs width above 30 MeV the mass shift is over



(a) View of the LHC location (b) View of the four LHC experiments positions and LHC layout

Figure 1.9

200 MeV and increases with the square root of the width. The apparent mass shift could be measured by comparing the $H \to \gamma\gamma$ mass measurement with the ZZ^* channel, where the shift is much smaller.

1.2 LHC

The LHC [16] is a circular hadron accelerator designed to accelerate protons at the energy up to 7 TeV per beam (14 TeV in the center of mass) and ions (Pb) at the energy of 2.76 TeV per nucleon in the center of mass. The main driving reason to build the LHC was the unveiling of the electroweak simmetry breaking mechanism. The LHC allows to produce particles up to an energy of 4/5 TeV when the design center of mass energies will be reached, thus allowing to explore completely the TeV energy scale range, where new physics beyond the SM is expected to be found. LHC and its experiments have been designed to fully explore this energy scale range.

The main target of particle physics for these years and those to come will be the comprehension of the electroweak simmetry breaking mechanism and the search for possible

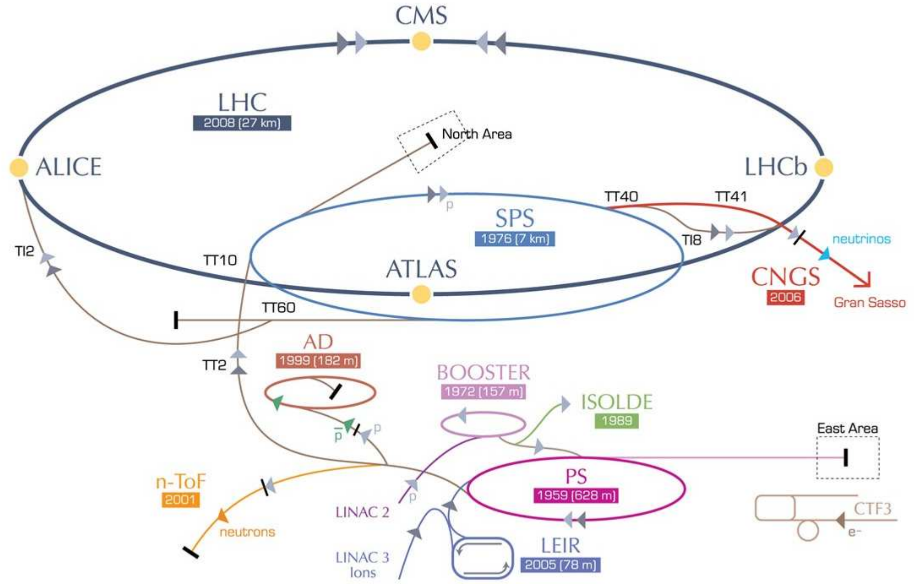


Figure 1.10: Proton (ions) acceleration and injection system to the LHC: LINAC2, PSB, PS, SPS and LHC

new physics. These are the reasons that led the particle physics community to design and build a new and more powerful accelator, the LHC.

1.2.1 The LHC layout

The main idea behind the LHC design was to install a new hadron collider into the existing 27 km long tunnel previously occupied by the Large Electron Positron collider (LEP) (sited 100 m underground at CERN laboratories in Geneva). This gave also the possibility to reuse part of the existing infrastructures, including preaccelerators. In the LHC design, 1232 main dipoles operating at 1.9 K and generating a magnetic field up to 8.33 T are used to steer the particles into curvilinear trajectories together with 386 quadrupoles, 360 sextupoles and 336 octupoles for stability control.

Prior to the injection in the LHC the particles are prepared by a series of injective systems that successively accelerate them from 1.4 up to 450 GeV (see Fig. 1.10).

The protons in the accelerator are inserted in bunches. Up to 2800 bunches of 10^{10} protons can be inserted in the machine.

The main accelerator characteristics are summarized in Table 1.1.

	Design target	Currently achieved
length	27 km	
center mass energy	14 TeV	8 TeV
luminosity	10^{34} cm^{-2}s^{-1}	$7.8 \cdot 10^{33}$ cm^{-2}s^{-1}
proton per bunch	10^{10}	$1.4 \cdot 10^{11}$
collision rate	40 MHz	20MHz
MAGNETS		
number	1600	
working temperature	1.9 K	
magnetic field	8 T	

Table 1.1: Summary table of design characteristics of the LHC for proton-proton collisions.

The experimental caverns are located in the four points where the two beam pipes intersect to produce collisions. Four experiments are installed there, with different designs and aims: ALICE for the study of quark-gluon plasma in heavy ion collisions, LHCb for the study of the CP violation in heavy flavour quarks physics (b-quark), ATLAS and CMS, which are two general purpose experiments.

1.2.2 Machine operation

The event rate R_i of a physics channel i occurring with cross section σ_i can be defined as the number of events per unit of time:

$$\frac{dN_i}{dt} = R_i = \sigma_i \mathcal{L} \tag{1.6}$$

and it is proportional to cross section σ_i via the constant $\mathcal{L}$, luminosity, which depends only on the machine parameters. Assuming a small crossing angle between the beams and gaussian-shaped beam bunches, the luminosity $\mathcal{L}$ can be expressed as

$$\mathcal{L} = f \frac{n_b N_1 N_2}{4\pi \sigma_x \sigma_y} \tag{1.7}$$

where f is the revolution frequency of the n_b bunches, N_1 and N_2 number of protons in the two colliding bunches, σ_x and σ_y the beam profiles in horizontal (bend) and vertical directions at the interaction point.

Physics proton-proton collisions started in 2010 at 7 TeV center-of-mass energy. In 2010 the instantaneous luminosity ($\mathcal{L}$) increased by 6 order of magnitute from 10^{26} $\text{cm}^{-2}\text{s}^{-1}$ as a demonstration of the excellent performance of the LHC and the control of the transversal size of the beams. As a consequence of the raise of the instantaneous luminosity, the number of multiple proton-proton interecations (pile-up) in the same collision grew up considerably up to an average of 20 pile-up in 2012 (see Fig. 1.11b). The maximum instantaneous luminosity reached in ATLAS and CMS interaction points is $7.8 \cdot 10^{33}$ $\text{cm}^{-2}\text{s}^{-1}$.

The machine operated until beginning of 2013 with a bunch crossing of 50 ns. The 50 ns bunch crossing limits the number of bunches circulating in the machine at 1400; this means that the only handle to reach higher instantaneous luminosities is the reduction of the β^* and the increase of the number of protons per bunch. Both possibilities have the drawbrack of increasing the number of collisions per bunch crossing, thus rendering the experimental conditions for the experiments more difficult (50 pile-up interactions are to be expected for instantaneous luminosities $> 10^{34}$). After the first long shut-down (lasting until 2015), the LHC should restart operations with increased center-of-mass energy and 25 ns bunch spacing doubling the instantaneus luminosity.

The center-of-mass-energy has already been increased in 2012 to 8 TeV in order to further enhance the Higgs discovery possibility (the Higgs production cross section increases by $\sim 20\%$ from 7 to 8 TeV).

The amount of data collected during 2010, 2011 and 2012 by the CMS experiment are shown in Fig. 1.11a and summarized in the following table.

Year	Data collected
2010	36 pb^{-1}
2011	5.3 fb^{-1}
2012	19.6 fb^{-1}

The peak luminosity provided by the LHC during proton-proton collisions in CMS is shown in Fig. 1.11c for 2010, 2011 and 2012.

1.2.3 LHC physics (proton-proton collision)

The collision of two protons A and B, at the LHC center-of-mass energy, involve the their constituents, called "partons" (quarks and gluons). Each parton carries only a fraction x of the proton momentum.

The collision cross section is then expressed by the sum of the probabilities of all the possible parton-parton interactions:

$$\sigma_{pp} = \sum_{a,b} \int dx_a dx_b \, P_a(x_a, Q^2) \cdot P_b(x_b, Q^2) \cdot \hat{\sigma}_{ab}(x_a, x_b) \tag{1.8}$$

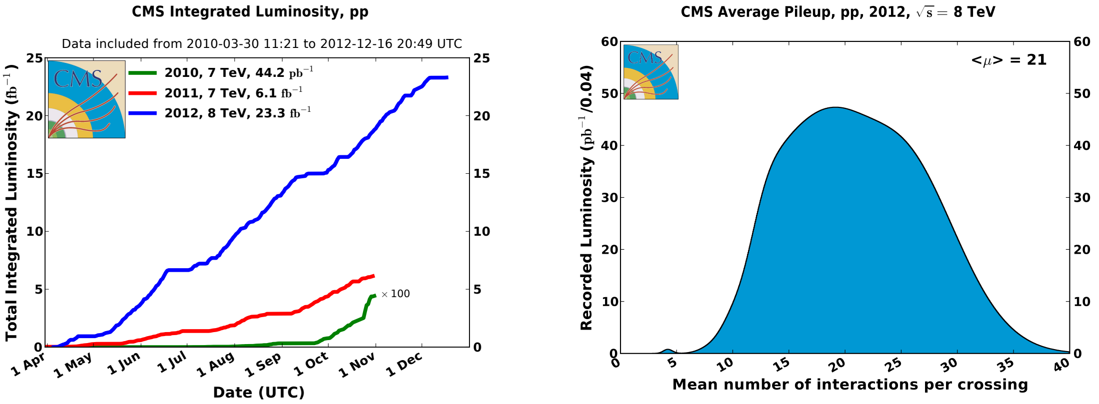
where $\hat{\sigma}_{ab}$ is the cross section of the interaction between a parton a of proton A and a parton b of proton B, $f(x, Q^2)$ is the Parton Density Function that is the probability density of having a parton bringing a fraction x of the proton momentum at a given exchanged four-momentum Q^2 between the two partons.

Inelastic proton-proton interactions (about 70% of the total proton-proton cross-section) are generally divided into:

- long range collisions with small trasfered momentum during the interaction,
- head-on collisions with high Q^2 exchanged between the partons. The particles produced by the interaction have a large transverse momentum with respect to the beam direction.

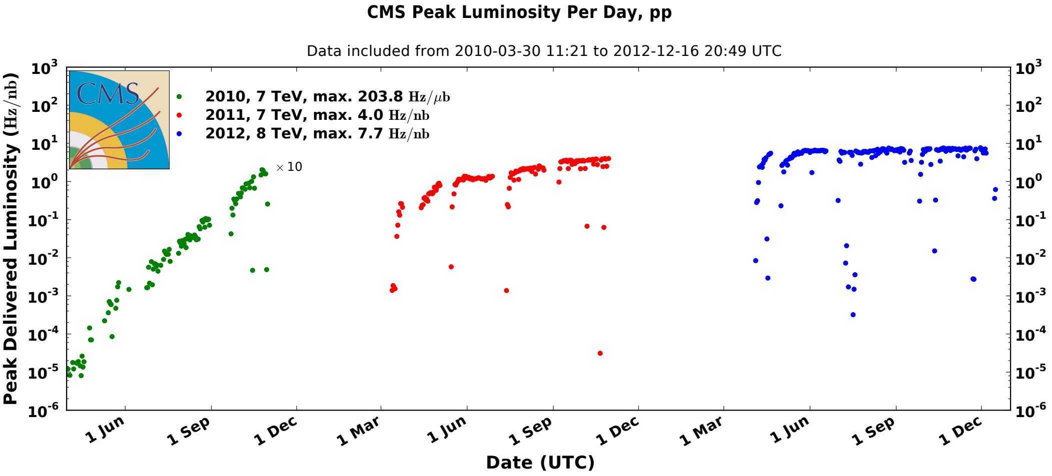
The second type of collisions are those we are mainly interested in.

In proton-proton collisions the kinematic of the process is closed in the trasverse plane where in fact particles not interacting with the detectors are reconstructed as an imbalance of the energy in the transverse plane.



(a) Integrated luminosity collected by the CMS experiment in 2010 (green) multiplyed by factor 10 for visibility, 2011 (red) and 2012 (blue)

(b) Pile-up distribution in 2012 seen by the CMS experiment



(c) Peak luminosity versus day delivered to CMS during stable beams and for p-p collisions. This is shown for 2010 (green), 2011 (red) and 2012 (blue) data-taking.

Figure 1.11

Chapter 2

Compact Muon Solenoid Experiment

2.1 Detector overview

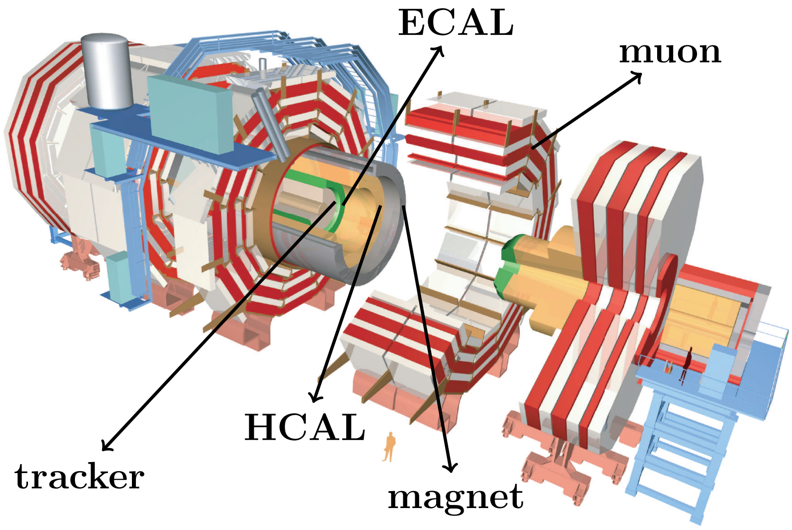


Figure 2.1: The CMS detector

The CMS[17] experiment (shown in Figure 2.1) is made of a large superconducting solenoid containing a full silicon tracker, a crystal Electromagnetic Calorimeter (ECAL) and a hadron calorimeter (HCAL). Muon chambers are embedded in the iron return yoke of the magnet. Overall the experiment measures 21.6 m in lenght, 14.6 m in height and weighs 12500 t.

After an overview of all the CMS subdetectors, a more detailed description of the Electromagnetic Calorimeter will be given, playing a major role in the analysis carried on in this thesis.

2.1.1 The coordinate system

The CMS experiment uses a right-handed coordinate system, with the origin at the nominal interaction point in the centre of CMS, the x axis pointing to the centre of the LHC ring, the y axis pointing vertically up (perpendicular to the LHC plane), and the z axis along the anticlockwise beam direction. The azimuthal angle ϕ is measured from the x-axis in the x-y plane and the radial coordinate in this plane is denoted by r. The polar angle θ is measured from the z-axis. Pseudorapidity is defined as $\eta = -\log\left(\tan\frac{\theta}{2}\right)$. Thus, the momentum and energy transverse to the beam direction, denoted by p_T and E_T respectively, are computed from the x and y components. The imbalance of energy measured in the transverse plane is called missing E_T and denoted by $\not{E}_{\mathrm{T}}$.

2.1.2 The magnet

Magnetic length	12.5 m
Cold bore diameter	6.3 m
Central magnetic induction	3.8 T
Nominal current	19.14 kA
Inductance	14.2 H
Stored energy	2.6 GJ
Operating temperature	1.8 K
Yoke magnetic induction	1.8 T

Table 2.1: The principal characteristics of the CMS solenoid magnet.

An important aspect driving the detector design and layout is the choice of the magnetic field configuration for the momentum measurement. Large bending power is needed to measure precisely the momentum of high-energy charged particles (the relative uncertainty on the momentum is $\frac{\delta p}{p} \propto \frac{p}{B}$).

The CMS magnet [18] is a superconducting solenoid providing a very uniform field in the inner tracking reaching 3.8 T along its axis, which is parallel to the beam axis. The tracker system, ECAL and HCAL are hosted within the magnet.

The magnet return yoke of the barrel has 12-fold rotational symmetry and consist of three sections along the z-axis; each is split into 4 layers (holding the muon chambers in the gaps). Most of the iron volume is saturated or nearly saturated, and the field in the yoke is around 1.8 T.

2.1.3 Tracker system

The silicon tracker [19, 20] is the innermost part of the CMS detector. At design luminosity about 1000 tracks/event are expected, therefore high detector granularity is needed.

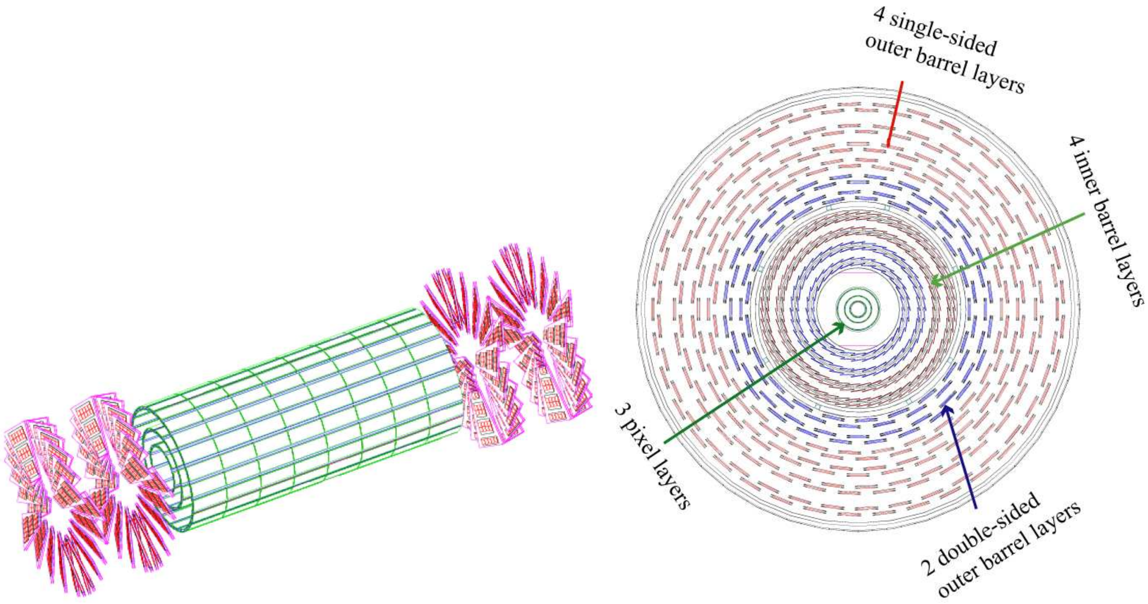
Speed and radiation hardness are other two requirements for the tracker because of the high intense flux of charged particle expected by the interactions at the LHC design luminosity. Moreover, the minimum material budget is required in order to minimize the multiple scattering, photon conversion and bremsstrahlung emission.

In the barrel region the tracking system consists of a cylindrical detector of 5.5 m in length and 1.1 m in radius. It is equipped with 3 silicon pixel detector layers (66 million channels) for the innermost part (for radii 4.4 cm $< R <$ 10.2 cm and for $|z| <$ 50 cm) and 10 silicon strip detector layers (2.8 million channels) for the outer part ($R <$ 110 cm, $|z| <$ 275 cm).

The tracker silicon strip detector consists of four inner barrel (TIB) layers assembled in shells with two inner endcaps (TID), each composed of three small discs. The outer barrel (TOB) consists of six concentric layers. Finally two endcaps (TEC) close off the tracker.The tracker layout is drawn in Fig. 2.2.

The tracker system provides a very precise measurement of particle momentum.

For high p_T track (100 GeV) the p_T resolution is about $1-2\%$ in the central region and a bit worse in the endcaps due to the lower lever arm. At this p_T the multiple scattering contribution is about $20-30\%$ and it increases for lower transverse momentum. The efficiency is above 99% in most of the acceptance. There is a small drop at high η due to the lack of coverage of the pixels.



(a) Pixel detector layout (b) Trasversal view of the CMS tracker

Figure 2.2: a) Layout of the pixel detector in CMS tracker. b) Layout of the barrel tracker.

Pixel

The pixel detector is the closest one to the interaction region, and is mostly used to provide a precise measurement of the primary and secondary vertices. The pixel detector provides in general 2 or 3 hits per track, each with a three-dimensional resolution of about 10 μm in the transverse plane and 15 μm in z. The good impact parameter resolution is important for good secondary vertices reconstruction.

Strip detector

The silicon strip detectors can provide up to 14 hits per track, with a two-dimensional precision ranging from 10 μm to 60 μm in R. Some of the silicon strip layers are double-sided to provide a longitudinal measurement with a similar accuracy. The tracker acceptance for a minimum of 5 collected hits extends up to pseudorapidities η of about $|\eta| < 2.5$.

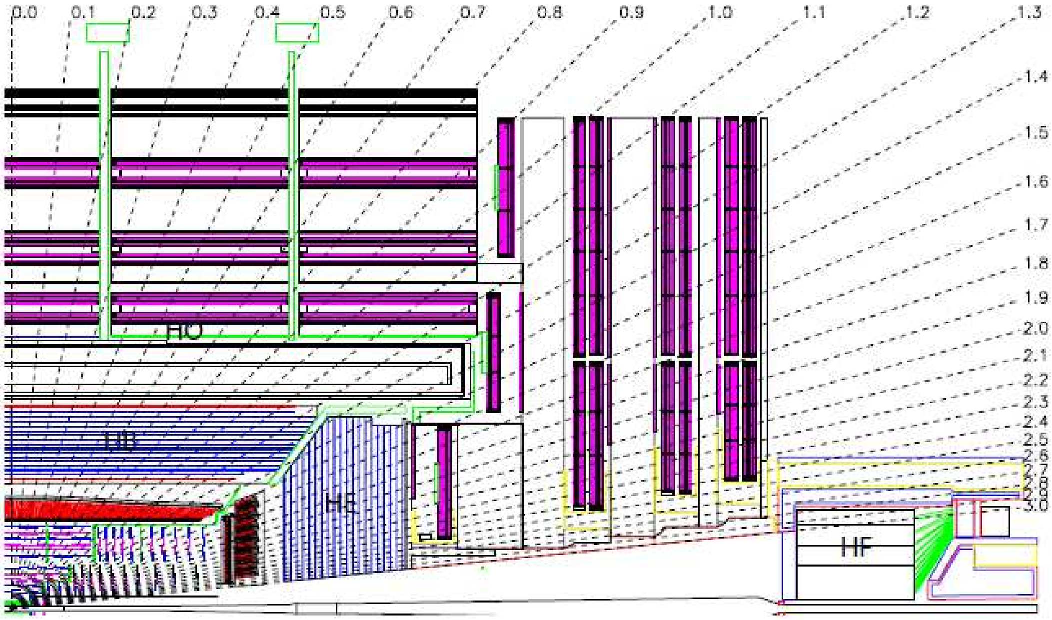


Figure 2.3: Longitudinal view of the CMS hadron calorimeter. The HCAL covers up to $|\eta| < 5$.

2.1.4 Hadron Calorimeter

The hadron calorimeter (HCAL) [21] is used together with ECAL to measure energy and direction of jets, and the energy imbalance in the transverse plane $\not{E}_{\mathrm{T}}$. It provides good segmentation, moderate energy resolution and angular coverage up to $|\eta| < 5$. HCAL is made of four subdetectors (Fig. 2.3):

- the Barrel Hadronic Calorimeter (HB) is placed inside the magnetic coil and it covers the central pseudorapidity region, up to $|\eta| = 1.3$
- the Endcap Hadronic Calorimeter (HE) is inside the magnetic coil as well and it is made of two endcaps extending the angular coverage up to $|\eta| = 3$
- the Outer Hadronic Calorimeter (HO, or Tail Catcher) is placed in the barrel region, outside the magnetic coil and is needed to enhance the depth of the calorimeter in terms of nuclear interaction length λ
- the Forward Hadronic Calorimeter (HF) consists of two units placed outside the magnetic coil, at ±11 m from the interaction point along the beams direction. It

extends the pseudorapidity coverage up to $|\eta| = 5$.

HB and HE are made with layers of $4 \div 7.5$ cm thick brass or stainless steel absorber plates interleaved with 3.7 mm thick plastic scintillators. The signal is readout through wavelengthshift fibres and hybrid photodiodes (HPD). The granularity ($\Delta\phi \times \Delta\eta$) is 0.087×0.087 in the central part and 0.17×0.17 at high η. The minimum depth is about $5.8\,\lambda$. In order to increase the calorimeter depth in the barrel region a tail catcher (HO) has been added outside the magnetic coil. HO is made of two scintillator layers, with the same granularity as HB; the total depth in the central region is thus extended to about $11.8\,\lambda$, with an improvement in both linearity and energy resolution. HE has a minimum depth of $10\,\lambda$. The two HFs are made of steel absorbers with embedded radiation hard quartz fibers. The fast Cherenkov light produced is collected with photomultipliers. The granularity is 0.175×0.175.

HB has an energy resolution for single pions of approximately $120\%/\sqrt{E}$.

The energy resolution of the ECAL-HCAL combined system was evaluated with a combined test beam with high energy pions [22] and it is given by:

$$\frac{\Delta E}{E} = \frac{84.7\%}{\sqrt{E}} \oplus 7.4\%$$

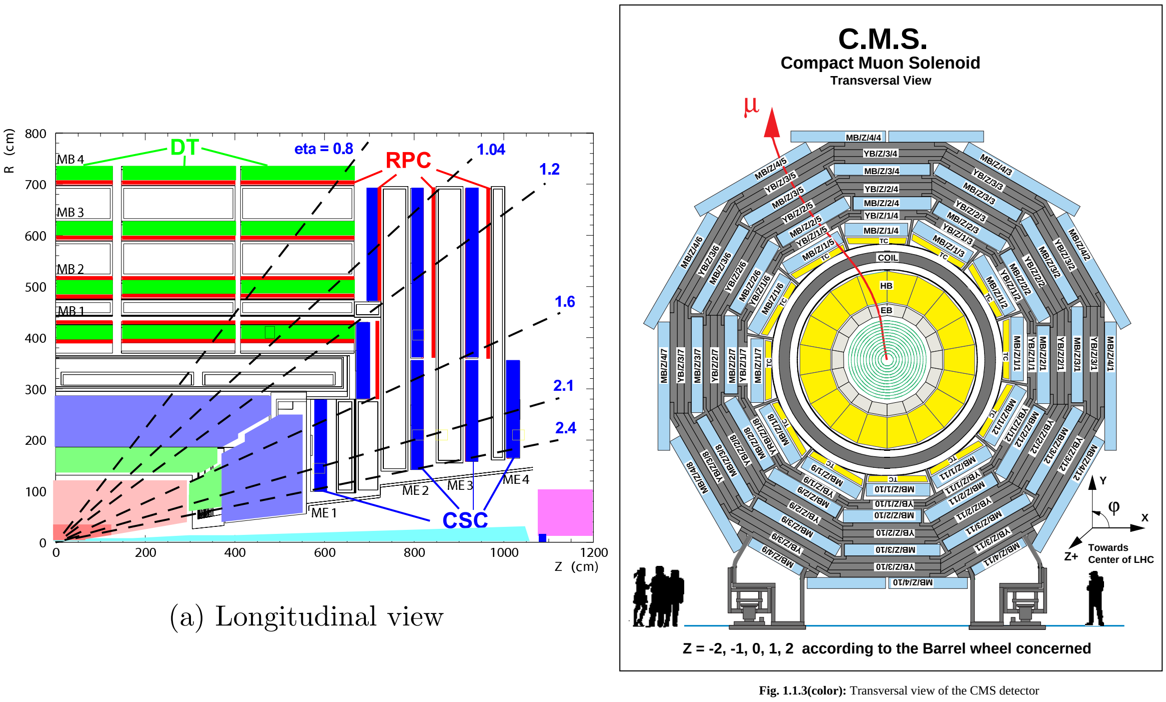
2.1.5 The muon system

The CMS muon system [23] is dedicated to the triggering, identification and momentum measurement of high p_T muons, the latter in combination with the tracker. The system is placed outside the magnetic coil, embedded in the return yoke, to fully exploit the 1.8 T return flux. The system consists of three independent subsystems (Fig. 2.4a):

- Drift Tubes (DT) are choosen for the barrel region, where the occupancy is relatively low (< 10 Hz/cm^2), the magnetic field uniform and the hadron flux low;

- Cathode Strip Chambers (CSC) are used in the endcaps, where the occupancy is higher (> 100 Hz/cm^2);

- Resistive Plate Chambers (RPC) are both in the barrel and in the endcaps.



(a) Longitudinal view

(b) Trasversal view

Figure 2.4: Longitudinal and trasversal view of the CMS muon system

The Drift Tube system is made of chambers consisting of twelve layers of drift tubes each, packed in three independent substructures called super-layers, for a total of four chambers with three super-layers per chamber. In each chamber two super-layers have anode wires parallel to the beam axis, and one has perpendicular wires. Thus, each chamber can provide two measurements along the $r - \phi$ coordinate and one measurement along z. Each chamber is made of two parallel aluminium plates jointed with “I” shaped spacer cathodes. Chambers are filled with a gas mixture of Ar(85%) and CO_2(15%). The position resolution is about 100 μm in both $r - \phi$ and rz.

Cathode Strip Chambers are multi-wire proportional chambers with segmented cathodes. Each chamber can provide both hit position coordinates. Chambers are filled with a gas mixture of Ar(40%), CO_2 (50%), CF_4(10%). The chamber spatial resolution is about $80 - 85$ μm.

Resistive Plate Chambers are made of parallel bakelite planes, with a bulk resistivity of $10^{10} \div 10^{11}$ Ωcm. They are operated in avalanche mode. These chambers have limited spatial resolution, but they have excellent timing performances; they are used for bunch crossing identification and for trigger purposes.

2.2 Electromagnetic Calorimeter

2.2.1 Physics requirements and design goals

The main driving criteria in the design of the Electromagnetic Calorimeter (ECAL) [24, 25] was the capability to detect the decay into two photons of the Higgs boson. This capability is enhanced by the excellent energy resolution and good angular resolution provided by a homogeneous crystal calorimeter.

2.2.2 ECAL design

Layout

The CMS ECAL (Fig. 2.5) is a homogeneous and hermetic calorimeter containing 61200 lead tungstate ($PbWO_4$) scintillating crystals mounted in the ECAL Barrel (EB), closed at each end by ECAL Endcap (EE) each containing 7324 crystals. A preshower detector (ES), made of silicon strip sensors interleaved by lead absorbers, is placed in front of the endcap crystals to enhance photon identification capabilities. The high-density ($8.28\,\mathrm{g/cm^3}$), short radiation length ($X_0 = 0.89$ cm), and small Molière Radius ($R_\mathrm{M} = 2.2$ cm) of $PbWO_4$ allow the construction of a compact calorimeter with fine granularity. Part of the lead tungstate crystals have been produced by BTCP in Bogoroditsk and part by SIC in Shanghai.

The scintillation decay time of these crystals is of the same order of magnitude as the LHC bunch crossing time: about 80% of the light is emitted in 25 ns. In order to compensate the low light yield (100 photons per MeV), the lead tungstate crystals are coupled to photodetectors with a high gain: two Avalanche Photo-Diodes (APDs) per crystal are used in the barrel, characterized by a higher magnetic field, and Vacuum Photoriodes (VPTs) in the endcaps because insensitive to the high hadron flux.

The ECAL layout is shown in Fig. 2.5.

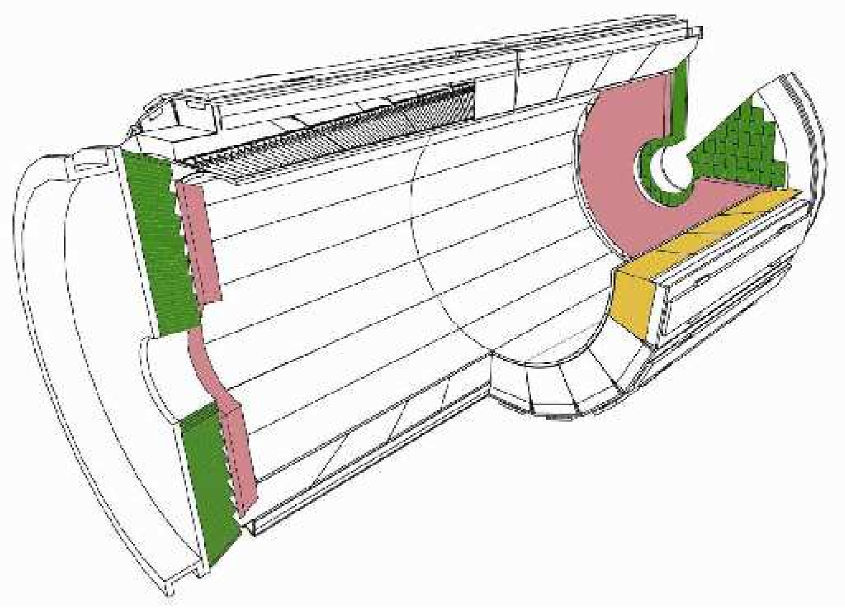


Figure 2.5: Layout of the CMS ECAL, showing the barrel supermodules, the two endcaps and the preshower detectors. The ECAL barrel coverage is up to $|\eta| = 1.48$; the endcaps extend the coverage to $|\eta| = 3.0$; the preshower detector fiducial area is approximately $1.65 < |\eta| < 2.6$.

The ECAL Barrel

The ECAL Barrel covers the region $|\eta| < 1.479$. The barrel is made of 61200 trapezoidal and quasiprojective crystals of approximately 1 $\mathrm{R_M}$ in lateral size and about $25.8\,X_0$ in depth. The barrel inner radius is of 124 cm.

Viewed from the nominal interaction vertex, the individual crystals appear tilted (off-pointing) by about 3° both in polar (η) and azimuthal angles (ϕ). The barrel is divided in two halves, each made of 18 supermodules containing 1700 crystals.

The ECAL Endcap

The endcaps consist of two detectors, a preshower device followed by a $PbW0_4$ calorimeter (Fig. 2.6). The preshower is made of silicon strips placed in a 19 cm sandwich of materials including about $2.3\,X_0$ of Pb absorber. It covers inner radii from 45 cm to 123 cm, corresponding to the range $1.6 < |\eta| < 2.6$. Each endcap calorimeter is made of 7324 rectangular and quasi-projective crystals of approximately 1.3 Molière Radius ($\mathrm{R_M}$) in

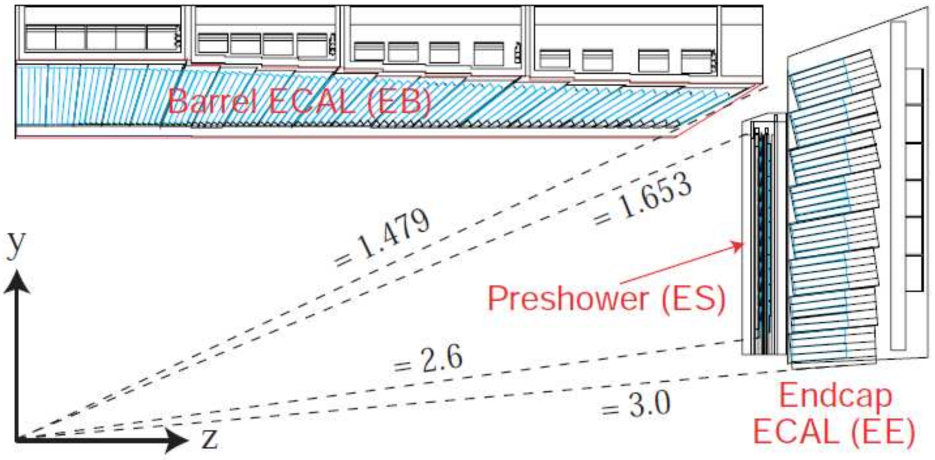


Figure 2.6: Schematic view of ECAL layout

lateral size and about 24.7 X_0 in depth. The crystal front faces are aligned in the (x, y) plane but, as for the barrel, the crystal axes are off-pointing from the nominal vertex in the polar angle by about 3°.

Electron and photon separation is possible up to $|\eta| = 2.5$, the limit of the region covered by the silicon tracker.

2.3 Trigger and data acquisition

When the LHC is running there are about one billion proton-proton interactions taking place every second. It is impossible for CMS to read out and record all these data. Furthermore, many of these events will not be interesting since they might be low-energy glancing collisions instead of head-on hard collisions. In order to select the most interesting events, triggers are employed.

A two level trigger system has been designed to reduce the event rate provided by LHC according to the allocated bandwidith for the data acquisition [26].

At the first level (L1), the trigger consists of custom designed, largely programmable electronics, taking information directly from subdetectors and provided a reduction of event rate of about 1000.

The Level-1 Trigger (L1) triggers can identify basic muon, calorimetric deposits in ECAL and HCAL and missing transverse energy candidates by accessing rough segmented data from the detector and storing all the high-resolution data in pipeline memories in the front-end electronics. L1 trigger rate is limited by the speed of the detector electronics readout, its selection criteria should utilize the most distinctive signatures of the particle objects. The L1 trigger is comprised of several subcomponents associated with the different subdetectors: the bunch crossing timing, the L1 muon systems, the L1 calorimetry and the global trigger (GT). The GT has the ability to provide up to 128 trigger algorithms to select an event based on logical combinations of L1 objects.

The second level or Higher Level Trigger (HLT) is a software system implemented in a filter farm of about one thousand commercial processors running an event filter that is a faster version of the offline reconstruction algorithms. The HLT trigger can access the complete data to identify particles in greater detail and accuracy. The HLT provides a reduction of the event rate to the required ~ 400 Hz.

The trigger flow is shown in Fig. 2.7.

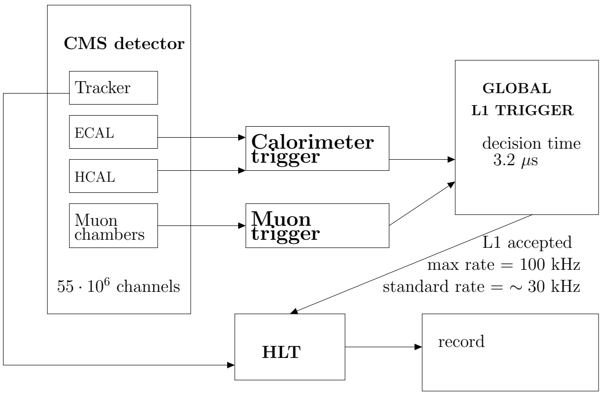


Figure 2.7: Workflow of the CMS trigger system

2.3.1 Calorimetric Trigger

At L1, electromagnetic candidates are formed from the sum of the transverse energy in two adjacent trigger towers (i.e., arrays of 5×5 crystals in EB). Coarse information on the lateral extent of the energy deposit inside each trigger tower is exploited to suppress spurious triggers, such as those arising from direct ionization in the APD sensitive region [27]. This feature has allowed the single-e/γ L1 trigger to be operated unprescaled at a low threshold of $E_\mathrm{T} = 15$ GeV in 2011 and 2012. From data analysis, this trigger has been verified to be fully efficient (>99%) for $E_\mathrm{T} > 20$ GeV, causing no inefficiencies to, e.g., the H $\rightarrow \gamma\gamma$ analysis, for which events are retained if the leading photon has transverse energy $E_\mathrm{T} > 35$ GeV [28, 29].

Once the L1-seeding requirement has been satisfied, ECAL electromagnetic clusters are formed in the region in the vicinity of the L1 seeds. ECAL information is unpacked only from the readout units overlapping with a rectangle centered on an L1 candidate with a size $\Delta\eta \times \Delta\phi = 0.25 \times 0.4$ to save processor time in the HLT. The resulting cluster should have a position matching the L1 candidate, a transverse energy satisfying the requirements of the given HLT path and show little energy in the hadronic calorimeter (HCAL) region just behind it. Electrons are selected as the e/γ candidates with hits in pixel detectors matching in energy and position the ECAL deposit. The ECAL energy weighted position is propagated back through the field to obtain an estimate of the direction of the electron at the vertex and the hit positions expected in the pixel detector. For the electron candidates, the full track is then reconstructed (using information also from the silicon strip detector) and further selection criteria are used as the ratio between the ECAL energy and the track momentum.

2.4 CMS simulation

In high energy physics experiments, Monte Carlo (MC) simulations are extensively used to better understand the collision dynamics, the detector response to final state particles, to tune proper algorithmic corrections to the particle properties (energy, direction), to optimize event selections.

The whole event is simulated starting from the parton-parton fundamental interaction (generation step) up to the interaction of final state particles with the matter of the detector (simulation step). The simulated detector information is then processed by the reconstruction framework as the real data are.

The generation step can be ideally subdivided into

1) the simulation of the hard scattering

2) the parton showering of the final state particles and the decay of the unstable

particles

The general structure of the physics process modeling and event generation procedure is shown in Fig. 2.8.

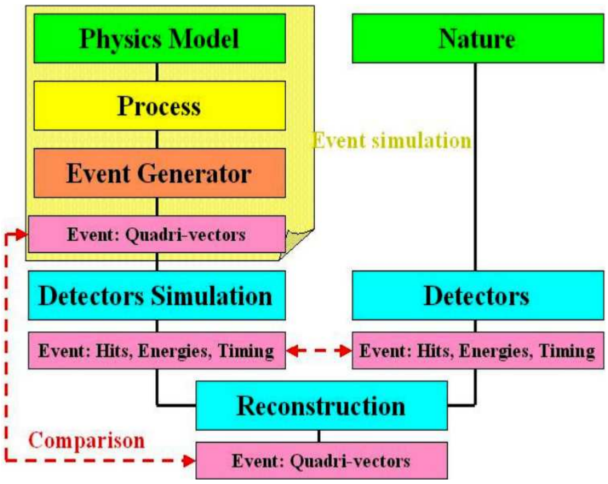


Figure 2.8: Basic steps in event generation, simulation and data analysis

Event generators are intended to generate complete events by subdividing the task into simpler steps. For generating a given hard process, the basic steps are as follows: generation of the Feynman diagrams involved in the process, construction of the matrix elements which after being integrated over whole phase space provides the total and differential cross-section. Finally, events are randomly generated according to the full differential cross-section and provides a set of four momentum vectors each associated with one of the final state particle.

The hard scattering final states often contain partons-(quarks and gluons), which cannot exist in the bare state. These partons get hadronized to produce hadrons as they move apart. Thus a general scheme of the event generation assumes the evaluation of the hard process, then evolve the event through a parton showering and hadronization step and the decay of the unstable particles. The event information contains the four momenta of all the final state particles (hadrons, leptons and photons) and the position of their decay vertices.

There are specific generator packages such as PYTHIA [30, 31] which simulate the transformation of partons into hadrons through the parton showering and hadronization algorithms. At hadron colliders like LHC, the hadronic processes are even more complex due to non-elementary structure of proton. There will be large number of possible initial hard scattering states. Moreover, the multiple interaction between the partons, not involved in the hard scattering, must be taken into account. During the collision of two proton bunches in LHC, more than 20 inelastic events are superimposed on the single possible interesting event. These events must also be simulated in order to come as close as possible to the real situation in the CMS detector.

The simulated events used in this thesis have been produced using several MC generators summarized in the following table:

Generator	Matrix element generator	parton showering
PYTHIA	✓	✓
MadGraph	✓	PYTHIA
POWHEG	✓	PYTHIA
SHERPA	✓	✓

The PYTHIA program is frequently used for event generation in high-energy physics. The emphasis is on multiparticle production in collisions between elementary particles. This in particular means hard interactions in e^+e^-, pp and ep colliders, although other applications are also envisaged. The PYTHIA generator is optimized in processes with one or two particles in the final state. PYTHIA has also a parton showering interface that is used in order to simulate the hadronization process and also to generate additional coloured particles in the final state.

MadGraph [32, 33] is a general purpose matrix-element based event generators, a tool for automatically generating matrix elements for High Energy Physics processes, such as decays and $2 \rightarrow n$ scatterings. MadGraph generates all Feynman diagrams for the process and evaluate the matrix element at a given phase space point. MadGraph is then interfaced with PYTHIA's parton showering in order to produce the complete final state.

The Leading Order (LO) calculations, implemented in the context of general purpose Shower Monte Carlo (SMC) programs i.e. PYTHIA, have been the main tools used in the various analysis. The SMC programs generally include dominant QCD effects at the leading logarithmic level, but do not enforce Next to Leading Order (NLO) accuracy. These programs were routinely used to simulate background processes and signals in physics searches. When a precision measurement was needed, to be compared with an NLO calculation, one could not directly compare the experimental results with the output of SMC program, since the SMC does not have the required accuracy. In view of the positive experience with the QCD calculations at the NLO level in the phenomenological study at electron and hadron colliders, it has become clear that SMC programs should be improved, when possible, with NLO results. The problem of merging NLO calculations with parton shower simulations is basically that of avoiding overcounting, since the SMC programs do implement approximate NLO corrections already. Several proposals have appeared in the literature that can be applied to both e^+e^- and hadronic collisions. One of them is "Positive Weight Hardest Emission Generator" called as POWHEG [34]. In the POWHEG method, the hardest radiation is generated first, with a technique that yields only positive weighted events using the exact NLO matrix elements. The POWHEG output can then be interfaced to any SMC program that is either p_T-ordered, or allows the implementation of a p_T veto.

SHERPA [35] is a general purpose matrix-element based event generators. It is capable of generating parton-level events at NLO precision. SHERPA also provides the possibility to generate hadron-level events at NLO accuracy using the POWHEG algorithm to combine NLO matrix elements with SHERPA parton shower.

More details about MC event generators for LHC can be found in [36].

Chapter 3

Electrons and photons reconstruction and identification

3.1 Energy measurement in ECAL

The front-end electronics of the EB, EE, use a 12-bit analogue-to-digital converters (ADC) to sample the analogue signals from the detectors (APDs and VPTs) at 40 MHz.

Ten consecutive samples are read out and stored for each crystal. The signal pulse is expected to start from the fourth sample and the baseline pedestal value is estimated from the first three samples [37]. In Fig. 3.1 an illustrative example of ECAL pulse shape is shown.

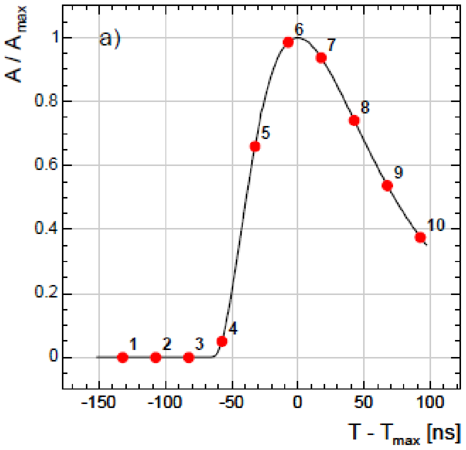


Figure 3.1: Illustrative example of ECAL pulse shape. Three of the ten samples is used to estimated the channel pedestal to be subtracted in the amplitude reconstruction.

The pulse amplitude A_i, in ADC counts, of each ECAL channel i is multiplied by an ADC-to-GeV conversion factor G weighted with channel dependent coefficients to correct for time response variation ($S_i(t)$), and to equalize the channel response (C_i, hereafter referred to as inter-calibration coefficients).

In general an electromagnetic shower speads over a few crystal: an ideal electromagnetic shower, like for example what can be obtained in the test beam conditions where no inert material is places in front of ECAL, releases aboout 90% of its energy within a matrix of 5×5 crystals. In CMS the situation is complicated by the presence of material upstream CMS, which in some region reaches almost 2 radiation lengths, causing bremsstrahlung for electrons and positrons and conversions for photons. The presence of the intense 3.8 T magnetic field furthermore spreads in the bending plane electrons and positrons from the conversions of photons, effectively spreading over a large region the initial energy of the electromagnetic particle. In general clusters which extends beyond a 5×5 matrix of crystals are used especially in the bending direction ϕ (as we will see in Sec. 3.3 a dynamic algorithm is exploited). Furthermore the energy of the e/γ candidate is corrected for imperfect clustering and geometry effects ($F_{e/\gamma}$). For endcap clusters the preshower energy E_{ES} is also added:

$$E_{e/\gamma} = F_{e/\gamma} \times (E_{ES} + G \cdot \sum_i (C_i \cdot S_i(t) \cdot A_i)) \tag{3.1}$$

The single channel response time variation and the channel-to-channel response variation affect directly the energy resolution, as shown in Fig. 3.2, where the Z boson invariant mass is reconstructed without any single channel calibration corrections, with only the intercalibration corrections, with full set of corrections.

3.2 Energy calibration

The ECAL energy calibration workflow exploit different methods and physics channels to measure, by several successive steps, the terms entering in the energy reconstructions expressed by Eq. 3.1. The first step consists of correcting for the single channel response time variation. Once the response of the channel is stable in time, collision events are used to derive the intercalibration corrections. As final step of the calibration procedure, the absolute energy scale G is tuned.

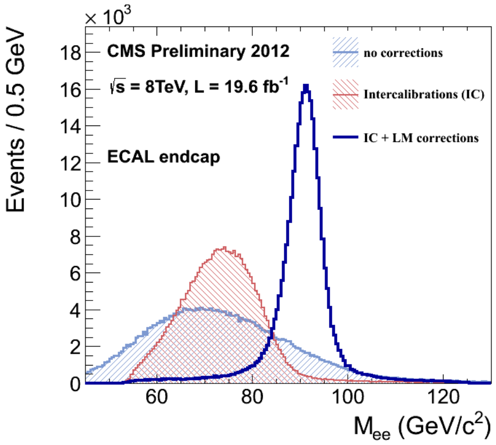


Figure 3.2: Z invariant mass reconstructed with two electrons in the ECAL endcap, without any single channel correction (violet), without response time variation correction S_i (red), with full corrections (blue).

3.2.1 Response time variation

The main reasons for variation of the response of ECAL channels as a function of time is due to creation of colour centres in the lead tungstate under irradiation [38]. This reduces the crystals transparency, and is followed by a spontaneous recovery due to thermal annealing of the colour centres when the irradiation stops (during inter-fill periods, technical stops or shutdowns). Another effect present only in the ECAL endcaps is the conditioning of the VPT [39] which depends on the total accumulated charge, thus is an incremental effect as a function of integrated luminosity. Other possible reasons are variation of temperature, which affects both the crystals light yield and the APD gains, and the bias voltage of the APD, affecting the APD gains. Thanks to the operational stability during Run 1 these sources of instability are to be considered negligible [40, 41, 42].

To correct in particular the variation of crystals' transparency, a light monitoring system is installed in-situ, with the capability to inject light into each crystal. Light produced from a laser with a wavelength close to the emission peak of the $PbWO_4$ (447nm) is used [43].

The laser light is injected through optical fibres in each EB and EE crystal through the front and rear face respectively. The spectral composition and the path for the collection of laser light at the photodetector are different from those for scintillation light.

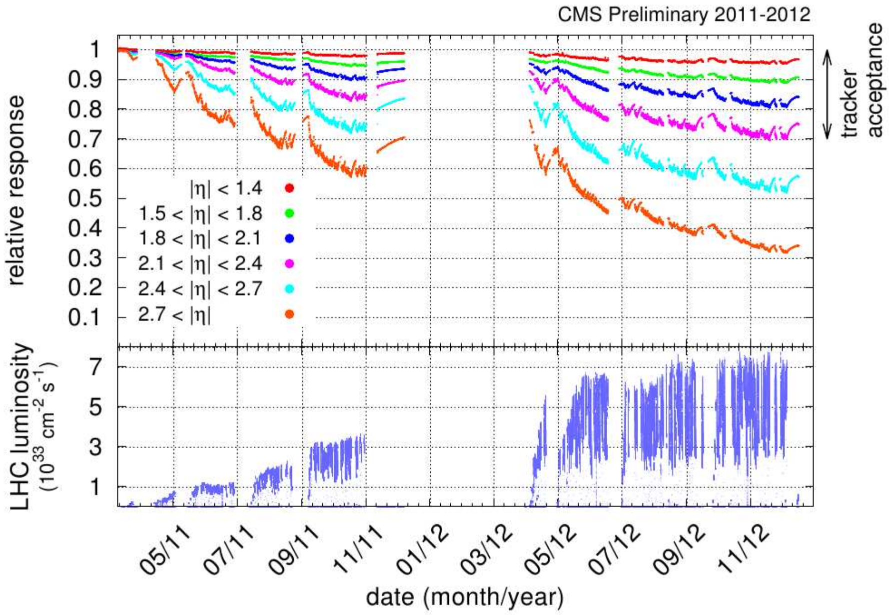


Figure 3.3: Relative response variation measured by the laser monitoring system in 2011 and 2012. The response is averaged over the pseudorapidity ranges listed in the legend. The LHC luminosity varied from 10^{33} cm^{-2}·s^{-1} in April 2011 to 7×10^{33} cm$^{-2}\cdot$ s^{-1} at the end of 2012. Heavy ion collisions took place in November 2011.

A conversion factor is required to relate the changes in the ECAL response to laser light to the changes in the scintillation signal. The relationship is described by a power law [24]:

$$\frac{S(t)}{S_0} = \left(\frac{R(t)}{R_0}\right)^{\alpha}, \tag{3.2}$$

where $S(t)$ is the channel response to scintillation light at a particular time t, S_0 is the initial response, and $R(t)$ and R_0 are the corresponding response to laser light. The exponent α is independent of the loss for small transparency losses.

The value of α has been measured in a beam test for a limited set of crystals under irradiation. Average values of 1.52 and 1.0 were found for crystals from the two producers, BTCP and SIC, respectively [44, 45]. The spread in α was found to be 10%, which arises from residual differences in transparency and different surface treatments of the crystals. The average α values are used *in situ* for all the crystals from the two producers.

The laser monitoring system provides one monitoring point per crystal every 40 minutes with a single point precision of better than 0.1% and long-term instabilities of about

0.2%. Following quasi-online processing of the monitoring data, response variation corrections are delivered in less than 48 h for prompt reconstruction of CMS data and are cross-checked by monitoring the stability of the π^0 invariant mass peak.

The evolution of the ECAL response to the laser light in 2011 and 2012 is shown in Fig. 3.3, as a function of time for several $|\eta|$ intervals corresponding to increasing levels of irradiation.

The data are normalized to the measurements at the beginning of 2011. The corresponding instantaneous luminosity is also shown. The response drops during periods of LHC operation and recovers during LHC stops due to thermal annealing of the colour centres in the $PbWO_4$ crystals. The smooth recovery in November 2011 occurred during heavy ion collisions at low luminosity. The observed losses are consistent with expectations and reach 5% in the barrel and about 30% at the end of the CMS acceptance region for e/γ ($|\eta| < 2.5$) at the end of 2012.

Given the response loss to laser light, the spread in α limits the precision of the response correction by the end of 2011 running for a single channel to 0.3% in EB, and between 0.5% and a few percent at high pseudorapidity in EE.

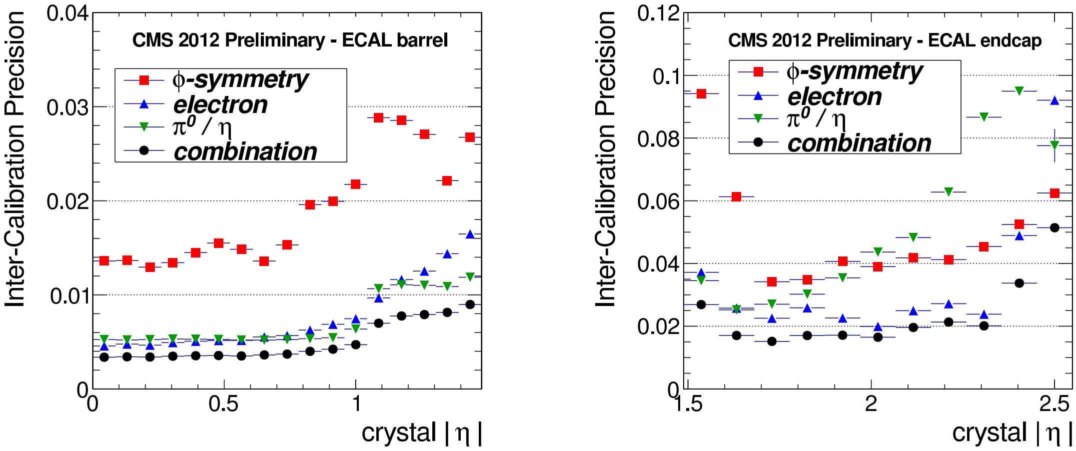
The gradual loss in VPT response in EE [39] due to the radiation environment at the LHC contribution to the observed response variations is not disentangled from the transparency loss of the crystals by the current monitoring system.

3.2.2 Intercalibrations

The main sources of the channel-to-channel response variations are the crystal light yield spread in EB, about 15%, and the gain spread of the photodetectors in EE (about 25%). These response variations contribute to the *in situ* constant term of the energy resolution and are corrected by the inter-calibration procedure.

The CMS ECAL has been calibrated prior to installation with laboratory measurements of crystal light yield and photo-detector gain during the construction phase (all EB and EE channels), with test-beam electrons (nine out of 36 EB supermodules and about 500 EE crystals) and with cosmic ray muons for all EB channels [46].

Refined inter-calibration has been derived *in situ* with several techniques exploiting



(a) EB

(b) EE

Figure 3.4: Precision of the various calibration sets used in 2012 in EB (left) and EE (right).

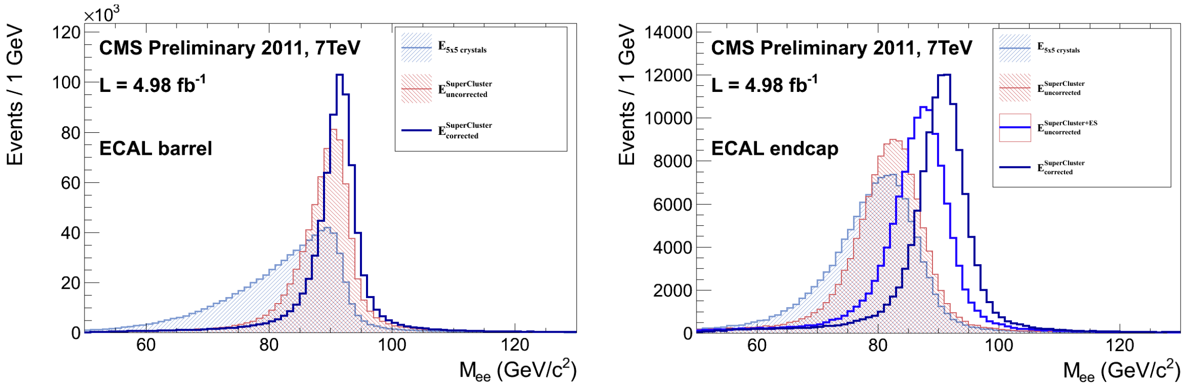
the properties of collision events [47]. These include the invariance around the beam axis of the energy flow in minimum bias events (ϕ-symmetry method), π^0/η mass constraint on the energy of the two photons from the $\pi^0/\eta \to \gamma\gamma$ decays, the momentum constraint on the energy of isolated electrons from Z and W decays, and the Z mass constraint on the energy of the two electrons from the $Z \to e^+e^-$ decays.

The precision of each method has been estimated from the cross-comparison of the individual results, and via cross-checks against pre-calibration constants derived from test beam campaigns. Results obtained in the 2012 run are shown in Fig. 3.4 as a function of pseudorapidity for EB and EE.

The inter-calibration with electron accuracy (still statistically dominated at $|\eta| > 1$) profited from the higher statistics of 2012, while the π^0/η accuracy degraded in 2012 in the EE due to the higher background coming from pileup events with respect to 2011.

The precision achieved by combining the results from the different methods is also displayed.

The levels of the residual errors on the channel response ensures a contribution to the energy resolution below 0.5% in the central part of the barrel ($|\eta| < 1$), and below 2% in the endcaps on both 2011 and 2012 data.



(a) ECAL Barrel (b) ECAL Endcap

Figure 3.5: Reconstructed invariant mass of electron pairs from $Z \to e^+e^-$ events, using the energy reconstructed in fixed arrays of 5×5 crystals (blue shaded), in the SC without algorithmic corrections (red shaded) and in the SC with algorithmic corrections. In the endcaps 3.5b, the ES energy is added before to apply $F_{e/\gamma}$.

3.3 Clustering and energy corrections

In test beams, the best energy estimate is obtained by summing the energy deposited in fixed arrays of crystals. In CMS, dynamic "clustering" algorithms are used to recover additional clusters of energy deposits due to secondary emission in the tracker material by bremsstrahlung or photon conversions, spread azimuthally by the intense magnetic field of CMS, and merge them into Super Clusters (SCs) [48, 49]. In the barrel region, clustering is performed with Hybrid algorithm. SCs are formed from windows 5 crystals wide in η around the most energetic crystal and a variable window in ϕ (up to 35 crystals wide). In the endcap region, clustering is performed with Multi5x5 algorithm. Matrices of 5×5 crystals around the most energetic crystals are merged if contiguous.

The effect on the resolution due to the clustering is shown in Fig. 3.5: the invariant mass of electron pairs in $Z \to e^+e^-$ decays is shown at different levels of the energy reconstruction, using fixed arrays of 5×5 crystals, as in the test beam, the SC energy without algorithmic corrections, and the SC energy including algorithmic corrections ($F_{e/\gamma}$ described later). The figure also suggests that the ratio of the energy in a fixed array of crystals over the energy in the SC is a convenient way to identify electrons with little radiation in the tracker or unconverted photons, for which a better energy resolution is

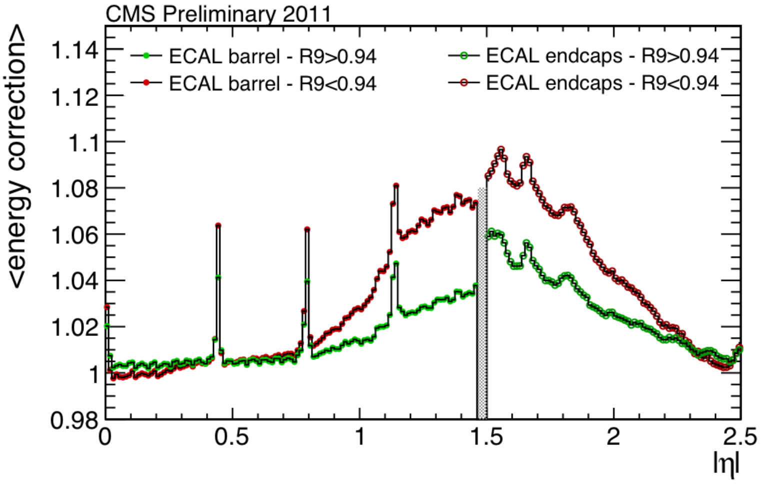


Figure 3.6: Energy corrections as function of $|\eta|$. The profile of the corrections closely matches the material upstream of ECAL. The barrel module boundaries are visible. Corrections are shown for electrons with small interaction with the upstream material ($R_9 > 0.94$) and electrons that have have emitted high energetic bremsstrahlung photons ($R_9 < 0.94$).

expected. For this purpose the variable $R_9 = E_{3\times3}/E_{raw}$, defined as the ratio of the energy in a 3×3 array to the energy in the SC before algorithmic corrections, is introduced.

3.4 Algorithmic corrections to electron and photon energies

Particle interactions with the material upstream of ECAL spread the particle energy over more crystals, also generating additional clusters in ECAL, that are recovered by the dynamic clustering algorithm. The energy lost due to algorithm inefficiencies or lost in the tracker (soft charged particles do not reach ECAL because of the magnetic field) is taken into account by particle, energy and position dependent corrections expoiting the different interaction mechanisms of e/γ upstream of ECAL and the CMS geometry.

In CMS, two different sets of cluster corrections are both derived on MC simulations of the interactions of electrons and photons with the CMS detector. A parametric set of corrections based on few shower shape quantities, and a more refined version using a multivariate regression technique which exploits the correlation among several shower shape variables and non-ECAL variables as the number of reconstructed primary vertices in the event. The parametric set of corrections is used also in the high level trigger reconstruction, while the multivatiare regression is used mostly when the ultimate ECAL energy resolution wants to be achived (like for example in the $H \rightarrow \gamma\gamma$ or $H \rightarrow ZZ^* \rightarrow 4\ \ell$ analysis)

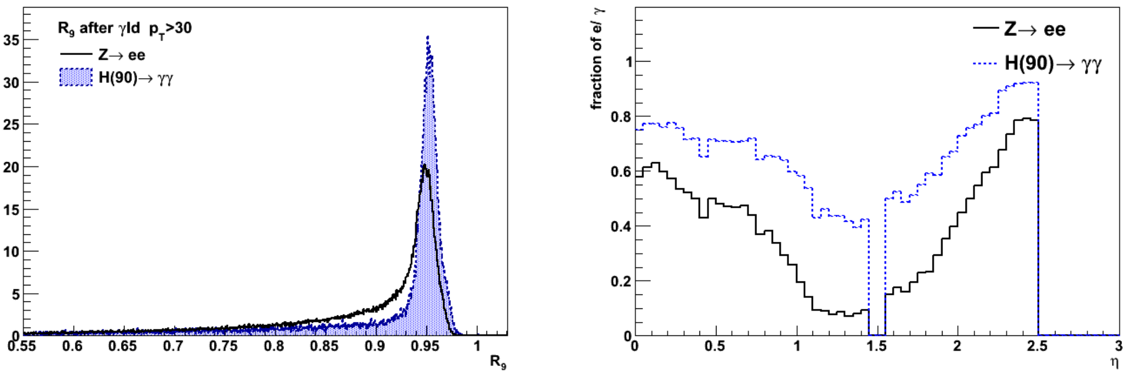
3.4.1 Parametric electron and photon energy corrections

The parametric version of the energy corrections factorizes three terms: one accounting for the variation of the cluster containment due to the position of the impinging particle on the central crystal front face, a second one for the interaction with the upstream material, and the third for the residual average E_T corrections. These corrections are optimized for electrons. For photons with large R_9 (almost not interacting with the upstream material) only the local containment correction is applied and the energy of the photon is reconstructed in a 5 × 5 matrix around the SC seed; this was proved in test-beam to give the best estimate of the photon energy. Converted photons and electrons are instead correcting using this schema.

$$\begin{gathered} F_{e,\gamma} = f\left(\frac{\sigma_\phi}{\sigma_\eta}, \eta\right) \cdot F(E_T) \\ \sigma_\phi = \sqrt{\frac{\sum_i E_i \cdot (\phi_i - \phi_{SC})}{E_{SC}^{raw}}} \qquad \sigma_\eta = \sqrt{\frac{\sum_i E_i \cdot (\eta_i - \eta_{SC})}{E_{SC}^{raw}}} \end{gathered} \tag{3.3}$$

$(\frac{\sigma_\phi}{\sigma_\eta})$ is the ratio between the shower spread in the bending direction ϕ and the spread in η; E_i, η_i, ϕ_i are the energy, the η and ϕ coordinate of the i-th crystal belonging to the SC. The η_{SC} and ϕ_{SC} are the energy weighted average of the η and ϕ of the SC crystals. In order to properly take into account the correlation between the shower broadening as function of the pseudorapidity and the material in the tracker, the first term $(f(\frac{\sigma_\phi}{\sigma_\eta}, \eta))$ is a function of $\frac{\sigma_\phi}{\sigma_\eta}$ and η.

A comparison between the R_9 distributions for electrons and photons is shown in Fig. 3.7 (left). The fraction of $R_9 > 0.94$ electrons and photons is shown in Fig. 3.7



(a) (b)

Figure 3.7: (left) R_9 distribution of electrons in $Z \rightarrow e^+e^-$ events compared to photons from H(90 GeV)$\rightarrow \gamma\gamma$, (right) the fraction of $R_9 > 0.94$ electrons and photons as a function of pseudorapidity respectively in $Z \rightarrow e^+e^-$ and H(90 GeV)$\rightarrow \gamma\gamma$ events.

(right) as a function of pseudorapidity respectively in $Z \rightarrow e^+e^-$ and H(90 GeV)$\rightarrow \gamma\gamma$ events.

The algorithm used to extract the $f(\frac{\sigma_\phi}{\sigma_\eta})$ correction bins the events in a two-dimensional matrix. For each of bin, the distribution of the ratio between the reconstructed SC energy and the generated energy is taken. The reconstructed SC energy is the uncorrected energy E_{SC}^{raw} with the associated energy deposited in preshower for the endcap.

The peak of the distribution is fixed to 1 by the correction factor.

3.4.2 MultiVariate (MVA) electron and photon energy corrections

The ultimate ECAL energy resolution is achieved using a multivariate regression technique based on a Boosted Decision Tree (BDT) implementation.

The electron/photon energy is computed starting from the raw SC energy E_{SC}^{raw} (adding also the preshower energy in the ECAL Endcap). These corrections are computed using a multivariate regression technique originally based on the TMVA Gradient Boosted Decision Tree implementation, though substantially optimized.

The regression is trained separately for electrons and photons. The photon-tuned

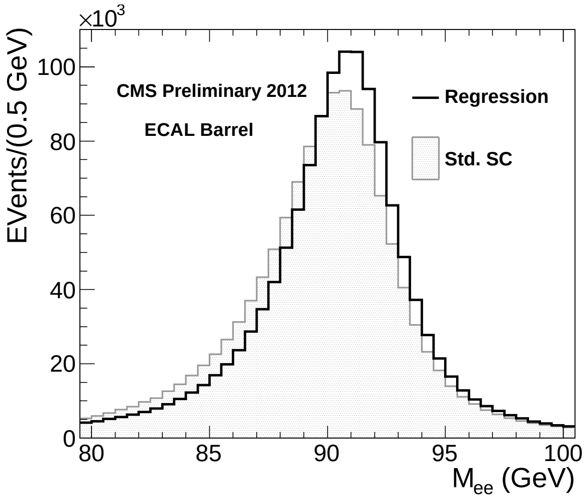


Figure 3.8: Reconstructed invariant mass of electron pairs from $Z \to e^+e^-$ events, using the energy reconstructed with MVA energy corrections (solid line) and standard energy corrections (pattern filled histogram).

regression is trained on prompt photons in Monte Carlo (from the photon + jets sample), while the electron-tuned regression is trained on a Drell-Yan (DY) sample. The ratio of generator level photon/electron energy to the raw SC (+ preshower) energy has been used as the target variable. The input variables are the SC position and a collection of shower shape variables. In the EE, the ratio of preshower energy to raw SC energy is additionally included. These variables provide information on the likelihood and location of a photon conversion and the degree of showering in the material, and together with their correlation with the global position of the SC, drive the degree of global containment correction predicted by the regression. Finally the number of reconstructed vertices and median energy density ρ in the event are included in order to correct residual energy scale effects from pileup.

The better performance of the MVA electron and photon energy corrections with respect to the standard corrections is shown Fig. 3.8. The Z lineshape is reconstructed with two electrons in the EB using the MVA regression energy (solid line) and the standard SC energy (pattern filled histogram).

3.5 Photon reconstruction

The photon is the simplest electromagnetic object. Any reconstructed SC is considered as a photon candidate. At this stage each electron is also reconstructed as a photon candidate, since a veto with respect to reconstructed prompt electrons is applied only later at the photon selection step. Photon candidates which comes from a conversion in the material upstream ECAL are tagged using an algorithm which searches for conversion tracks matching the ECAL SC as will be describe in the following section.

3.5.1 Reconstruction of conversions

Conversion reconstruction is used in the $H \to \gamma\gamma$ analysis to help with the identification of the correct Higgs interaction vertex as explained in Section 5.4. About one quarter of the $H \to \gamma\gamma$ events have at least one of the photons reconstructed as a converted photon.

The search for a pair of tracks which are compatible to a conversion candidate starts searching for tracks seeded by the ECAL SCs above a certain threshold in order to minimize the combinatorial background ($p_T > 10$ GeV). These list of tracks is merged to the electron specific tracks (removing duplicate tracks sharing a large fraction of hits) which are obtained from a special algorithm taking into account the large radiative Bethe-Heitler tail of the electron tracks (Gaussian Sum Filter (GSF) algorithm [50]) that will be described later and is more efficient to reconstruct tracks associated to early conversions.

In case of multiple conversions matching the geometrically the same SC, the one with the minimum $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ distance between the supercluster direction and the conversion direction is chosen. The direction of the conversion is taken from the momentum of the track pair refitted with the conversion vertex constraint.

3.6 Electron reconstruction

In CMS there are two electron reconstruction algorithms which differ by the seeding procedure: EcalDriven electrons [51] (inward procedure starting from ECAL and associating then a track) and TrackerDriven electrons [52] (outward procedure starting from a track and searching energy deposits in ECAL).

The EcalDriven algorithm is the standard CMS electron reconstruction, most suitable for energetic and isolated electrons. Low-energy (below 10 GeV) and/or non-isolated electrons, are reconstructed more efficiently by the TrackerDriven electron algorithm.

Electron seeds selected by both algorithms are collected together, keeping track of the seeding algorithm originating them. In this thesis we will refer only to the EcalDriven electrons, because they are reconstructed starting from the same list of SCs of the photons. Only 0.15% of $Z \rightarrow e^+e^-$ events passing the selection described later in Sec. 4.2.4 and with $E_T > 25$ GeV are discarded requiring both electrons to be EcalDriven.

The ECAL driven electron algorithm, starts from the already reconstructed SCs. The SC seeds a track matching algorithm with simple geometrical cuts aimed to associate the best matching track to the SC.

In the CMS reconstruction software (CMSSW) the electron reconstruction procedure is divided into steps [53]:

1) electron seeding
2) electron tracking
3) electron preselection (track - ECAL matching)
4) Bremsstrahlung recovery

Electron seeding

The EcalDriven approach starts finding suitable energy deposits in the ECAL represented by SCs.

Hit pairs and triplets in the pixel layers constitute a trajectory seed. The trajectory seeds in a z and ϕ window along the electron helix trajectory starting from the SC are selected as geometrically matching trajectory seeds. The ϕ matching is performed propagating the SC ϕ direction backward to the first pixel layers using both charge hypotesis.

Electron tracking

Electron seeds are then used to initiate a dedicated electron track building and fitting procedure in order to best handle the effect of bremsstrahlung energy loss. The track

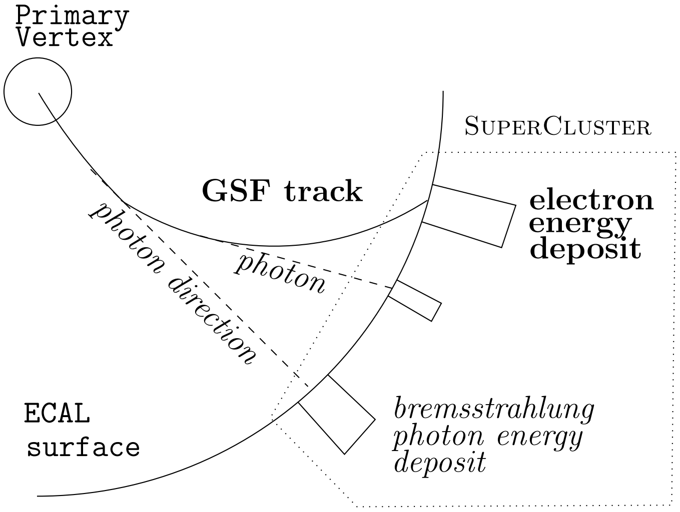


Figure 3.9: Schematical view of an electron reconstructed in CMS. The track is reconstructed by the GSF algorithm taking into account the trajectory kinks due to energetic bremsstrahlung photon emission. The energy deposits belonging to the emitted photons are collected together to the electron cluster by the clustering algorithm.

finding is based on a combinatorial Kalman Filter (KF), requiring a very loose χ^2 compatibility. The combinatorics is limited by requiring at most 5 candidate trajectories at each tracker layer and at most one layer with a missing hit. The hits collected by the track finding are then fitted by the GSF [50].

Track reconstruction

The KF algorithm [54] is the standard track reconstruction algorithm in CMS. It was developed for particles below the critical energy (as muons) and is the faster track algorithm. Because of the bremsstrahlung emission, electron track reconstruction needs a dedicated tracking algorithm that can take into account the kinks given by the bremsstrahlung energy loss to the electron direction, this is the GSF track algorithm (see Fig. 3.9).

Track - ECAL matching

In the case of EcalDriven electrons, the GSF track-ECAL matching requires:

- $|\Delta\eta_{in}| = |\eta_{\text{SC}} - \eta_{in}^{extr}| < 0.02$
- $|\Delta\phi_{in}| = |\phi_{\text{SC}} - \phi_{in}^{extr}| < 0.15$

where η_{SC} and ϕ_{SC} are the energy weighted position of SC crystals in η and ϕ respectively and where η_{in}^{extr} and ϕ_{in}^{extr} are the η and ϕ coordinate position of the closest approach to the SC position, extrapolating from the innermost track position and direction.

Removal of Conversions from Bremsstrahlung Photons

Once preselected, a further selection step is applied to remove ambiguous electron candidates that arise from the reconstruction of conversion legs from photon(s) radiated by primary electrons. In the case of an emitted photon taking more than half of the original electron p_T, the predicted position in the next layer is closer to the photon than to the electron after emission. If the photon converts, the hits from its conversion legs will likely be efficiently found by the electron track reconstruction algorithm. In such bremsstrahlung conversion patterns, the reconstruction often leads to electron candidates constituted by closeby tracks associated to the same or closeby SCs, hereafter defined as ambiguous candidates.

Firstly, electron candidates having SCs “in common” are identified. Therefore, two SCs are considered “in common” if a minimum energy is shared. Then, the ambiguity is solved keeping the electron candidate that satisfies the first of the following creteria:

1) the innermost first hit
2) the best E/p ratio

3.7 ECAL noise and simulation

The single channel noise has been measured using the laser monitoring data in 2011 and 2012. The pulse shape generated by the injected laser light is recorded as such the first

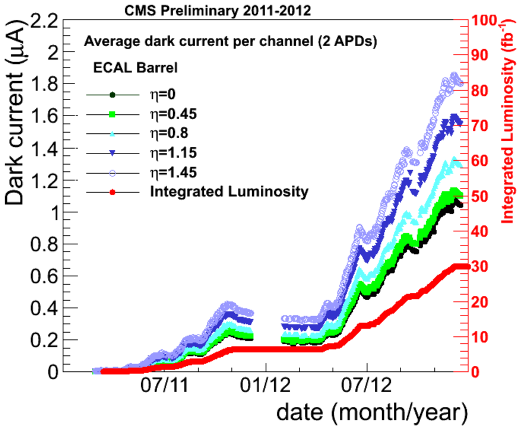


Figure 3.10: ECAL Barrel high-voltage dark current measured in 2011 and 2012.

three samples contains no signal, thus allowing to estimate the high frequency noise using these samples.

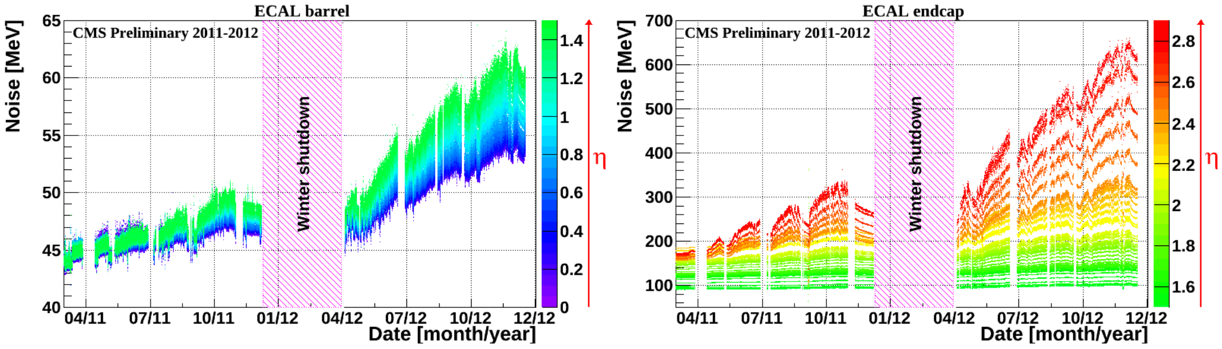
ECAL Barrel

In the ECAL EB, the major contribution to the variation in the single channel noise is the increasing dark current in the APDs. APDs are silicon devices and they are sensitive to neutron damage. The neutrons create defects in the silicon lattice, which cause an increase in the dark current. The neutron fluence at high η is expected to be larger by a factor of 2 with respect to the central barrel ($\eta = 0$), and the measured current in the ECAL EB high voltage system is observed to scale accordingly as shown in Fig. 3.10. Some thermal annealing inducing a partial recovery is visible during technical stops and shutdown between 2011 and 2012.

In test beam measurement, the single channel noise is ~ 40 MeV, at the beginning of 2011 it is ~ 45 MeV and it increased by 50% at the end of 2012. The noise as function of time is shown in Figg. 3.11a.

ECAL Endcap

The main contribution to single channel noise variation in the EE is indirectly due to the high radiation damage. The single channel noise is almost constant during all data taking, but the response loss correction (the response loss is shown in Fig. 3.3) amplify the



(a) ECAL Barrel

(b) ECAL Endcap

Figure 3.11: ECAL noise in barrel (left) and endcaps (right) measured in MeV

equivalent noise. The η profile of the noise and its time variation therefore follow closely the response corrections. In EB where the response losses are smaller, the contribution due to the radiation damage is negligible with respect to the APD dark current.

3.7.1 Time dependent simulation

High energy physics MC simulated samples are generally produced with realistic detector conditions which take into account the precision of the calibration, the alignment and dead channels (ECAL has been operating efficiently since installation, with about 1% of non-operational channels in EB and EE and 4% in ES). .

In CMS a simulation representing the evolving detector conditions has been adopted in order to account for aspects like the noise evolution that cannot be corrected.

The more realistic simulation have been then produced with three runs, the first with conditions close to the first part of the data taking (from April to Jun, RUN 2012 A+B), the second with conditions matching the data taken from Jun to September (RUN C) and the last reproducing the data taken in RUN D (from September to December).

The simulation has been then further improved including also an extended simulation of Out-of-Time (OOT) pileup events. These events do not occur in the same collision of the simulated hard scattering, but in the previous collisions. Energy deposits from OOT pileup events, contribute to the first three samples of the pulse shape, generating a bias in the amplitude measurement. More detailed studies have demonstrated that the agreement between data and simulation deteriored in 2012 and is restored introducing in the simulation the extended OOT pileup up to 300 ns before the bunch crossing of the simulated hard scattering and a noise level evolving in time as it do in data.

Chapter 4

Measurement of the energy scale and energy resolution

This chapter is focused on the measurement of the ECAL energy scale and resolution and the evaluation of the discrepancies between data and simulation.

Firstly, the results of the energy resolution measurement performed at test beam are presented with the relevant contributions to the intrinsic ECAL energy resolution. Then the additional contributions to the *in situ* energy resolution are summarized.

A pure sample of electrons from $Z \rightarrow e^{+}e^{-}$ decay have been used to measure the *in situ* energy scale and resolution.

A first technique, based on the fit of the Z lineshape had been developed in 2010. From the fit parameters the Z peak shift with respect to the simulation and the electron experimental resolution are estimated. The description of this technique is given in Sec. 4.2.8.

A more powerful method have been developed, exploiting the increased data collected by the CMS experiment in 2012. Results on the energy scale and energy resolution corrections obtained by the second method are then used in the $H \rightarrow \gamma\gamma$ analysis.

4.1 Intrinsic ECAL energy resolution

The ECAL barrel energy (E) resolution for electrons in beam tests has been measured to be [55]:

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E(\mathrm{GeV})}} \oplus \frac{12\%}{E(\mathrm{GeV})} \oplus 0.3\% \tag{4.1}$$

where the three contributions correspond to the stochastic, noise, and constant terms. This result was obtained reconstructing the showers in a matrix of 3×3 crystals where the electron impact point on the calorimeter was tightly localized in a region of 4 mm×4 mm to give maximum containment of the shower energy within the 3×3 crystal matrix. The noise term of 12% at 1 GeV corresponds to a single-channel noise of about 40 MeV, giving 120 MeV in a matrix of 3×3 crystals. In Sec. 3.7, the evolution of the single channel noise during 2011 and 2012 has already been discussed. The constant term, which dominates the energy resolution for high-energy electron and photon showers, depends on the non-uniformity of the longitudinal light collection, energy leakage from the back of the calorimeter, single-channel response uniformity and stability. The beam test setup was without magnetic field, no inert material in front of the calorimeter, and accurate equalization and stability of the single-channel response (better than 0.3%) [46]. The specification for the ECAL barrel crystals was chosen to ensure that the non-uniformity of the longitudinal light collection and the energy leakage from the back of the calorimeter contributed less than 0.3% to the constant term. The beam test resolution studies show that this target was met.

4.2 Measurement of the *in situ* energy resolution

4.2.1 Contributions to the *in situ* energy resolution

The *in situ* contributions to the energy resolution relative to the ECAL calibration have been introduced in the previous Chapter and can be summarized as: stability of the time response corrections, accuracy of the inter-calibrations, accuracy of the algorithmic corrections to the energy that have been derived on the simulation.

A further contribution to the constant term of the energy resolution [41, 42] comes from the general environmental stability . The main instability sources are fluctuations in temperature, that directly affect the light yield of the crystals ($-2\%/^{\circ}$ C) and the gain of the APDs ($-2.3\%/^{\circ}$C), and fluctuations in the bias voltage supply to the photodetectors. A cooling system utilising water flow [40] has maintained a stable operating temperature (to within 0.02°C in EB and 0.05°C in EE), ensuring a contribution of less than 0.1% to the energy resolution.

In the following, the evidence of further contributions to the energy resolution are discussed.

Stability of the time response corrections

Response time corrections have been described in Sec. 3.2.1.

In the response corrections as function of time, the variation of the response to a physical signal is related to crystal response loss measured by the laser monitoring system throught the parameter α.

As discussed in Sec. 3.2.1, the α value has been tuned *in situ* separately for all the barrel crystals and for all the endcap crystals to effective fixed values. The expected effect on the resolution is given by the product of the spread of the values of α and the average signal loss. In the endcaps, where the irradiation and the transparency loss are more important, the spread of the true α values with respect to the effective α led to a worsening of the resolution as function of time (irradiation) as an effective mis-calibration of the crystals. In the refined calibrations adopted in the 8 TeV data used in this thesis, the inter-calibration has been performed in time bins, in order to partially correct this effective mis-calibration. The result is a constant resolution as function of time in the endcaps. In Fig. 4.1 the di-electron resolution as function of time is shown with the calibrations used in the prompt reconstruction during the 2012 and with the refined calibrations. The di-electron resolution has been estimated with the *fit method* (described later in Sec. 4.2.8). In each time bin, the residual mis-calibration as function of time contributes to the *in situ* energy resolution and has a higher effect in regions with larger response corrections, i.e. increasing as function of η (see Fig. 3.3).

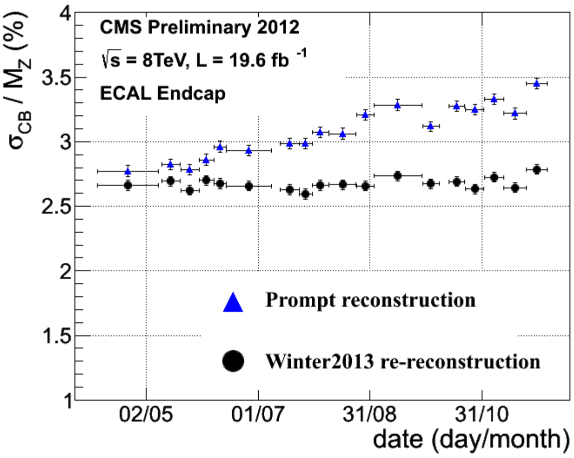


Figure 4.1: $Z \rightarrow e^+e^-$ experimental width as a function of time in EE with 2012 data. Data reconstructed with “prompt” calibration conditions are shown with black dots and with blue triangles data with refined calibrations (“Winter13” conditions).

The crystal response loss is not implemented in the simulation therefore the additional contribution to the energy resolution due to the spread of the crystal α values is missing in the MC samples.

Inter-calibration precision

The uncertainty on the measurement of the inter-calibration corrections have been shown in Sec. 3.2.2 in Fig. 3.4. In the central part of the ECAL Barrel, the contribution to the constant term of the energy resolution is the energy seed energy fraction × the inter-calibration accuracy ($0.6 \times 0.5\% \sim 0.3\%$). The accuracy of the inter-calibration is implemented in the simulation with an effective crystal mis-calibration.

Interaction with the upstream material

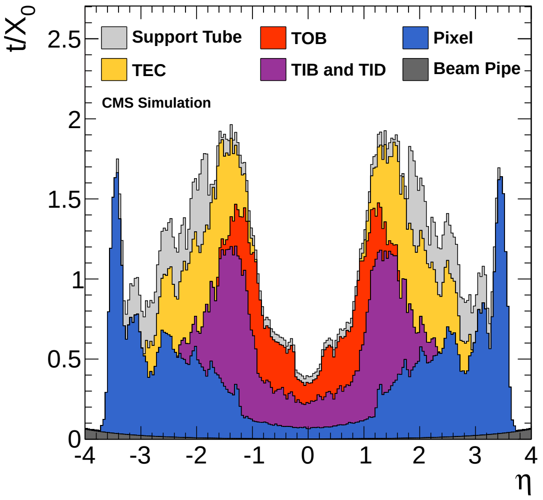


Figure 4.2: Upstream material in front of ECAL in units of radiation lengths as function of the pseudo-rapidity

A detailed description of the material in front of ECAL (shown in Fig. 4.2 as function of the pseudo-rapidity in units of radiation lengths) has been included in the full CMS simulation.

The minimum amount of material is in the central part of the barrel, up to $|\eta| = 0.8$, then it increases until the end of the barrel.

The average ratio $\frac{E_{rec}}{E_{MC}}$ between the electron energy after the algorithmic corrections and the energy at generator level in the simulation is shown in Fig. 4.3 as function of the pseudorapidity. Electrons almost not interacting with the material (with $R_9 > 0.94$) have an almost flat distribution, confirming that the R_9 variable is able to discriminate between interacting and non interacting electrons. Electrons with $R_9 < 0.94$ have a large drift of the average of the $\frac{E_{rec}}{E_{MC}}$ distribution due to the increasing left tail, whilst the mode of the distribution is fixed at 1 by the algorithmic corrections. In the small pad the distributions for $R_9 < 0.94$ and $R_9 > 0.94$ electrons are shown for the outer part of the barrel ($1.0 < |\eta| < 1.4442$).

4.2.2 Energy scale and resolution with $Z \rightarrow e^+e^-$ events

In the previous Section, the *in situ* energy resolution has been shown to depend on the pseudorapidity and the strength of the interaction with the upstream material. The electrons and photons are therefore classified by SC pseudorapidity and R_9.

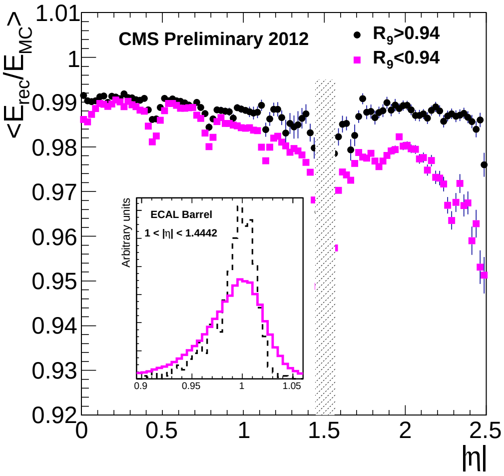


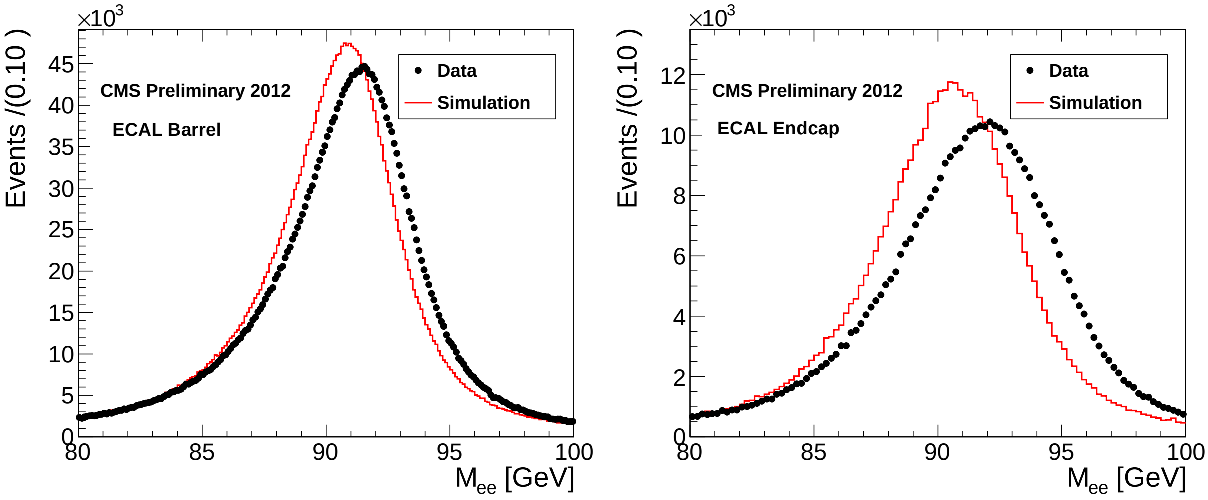
Figure 4.3: The mean of the distribution of the ratio between the reconstructed energy (E_{rec}) and the generator level energy (E_{MC}) of electrons is shown as function of the pseudorapidity. Electrons with low interaction with the upstream material (black points) have an almost flat distribution. The mode of the $\frac{E_{rec}}{E_{MC}}$ distribution is fixed to 1 by the algorithmic corrections. The mean of the distribution drifts in the region with higher material due to the larger left tail as shown in the small pad for electrons in the $1.0 < |\eta| < 1.4442$ range.

The adopted classification is:

- pseudorapidity (four regions, two in EB and two in EE):
  - $|\eta| < 1.0$, $1.0 < |\eta| < 1.4442$, $1.566 < |\eta| < 2.0$, $2.0 < |\eta| < 2.5$.
- R_9 (2 categories):
  - $R_9 < 0.94$: high interaction with upstream material
  - $R_9 > 0.94$: low level of interaction with the upstream material

The following study is performed using a pure sample of electrons produced by the decay of Z bosons. A large number of $Z \to e^+e^-$ events are produced at the LHC (cross section ~ 1 nb).

In Fig. 4.4 the reconstructed Z invariant mass is shown requiring two electrons in the barrel region (4.4a) or two electrons in the endcap regions (4.4b). The Z peak results



(a) ECAL Barrel

(b) ECAL Endcap

Figure 4.4: Z lineshape from $Z \rightarrow e^+e^-$ decay, reconstructed with two electrons in ECAL Barrel (left) or two electrons in ECAL Endcap (right)

shifted in data with respect to the simulation, the Z width results larger in the data than in the simulation. The Z peak position is related to the energy scale; the Z width is the convolution of the Z intrinsic width ($\Gamma_Z = 2.4952$ GeV [56]) with the experimental energy resolution

With the increased number of $Z \rightarrow e^+e^-$ events with the 8 TeV data (almost a factor 4 with respect to 7 TeV), a detailed study of the energy scale and resolution had been possible.

A more powerful method described in Sec. 4.3 has then been developed.

In the following, the selection of the $Z \rightarrow e^+e^-$ events and the two methods will be described. The methods are used to estimate the corrections to the energy scale and the energy resolution to be used in the $H \rightarrow \gamma\gamma$ analysis to correct for the discrepancies between the data and the simulation.

The main sources of discrepancy that can affect the energy of the reconstructed electrons and photons are:

1) mostly due to additional material budget in front of ECAL

   An evidence of an underestimation in the simulation of the amount of material in front of ECAL comes from studies with photon conversions,

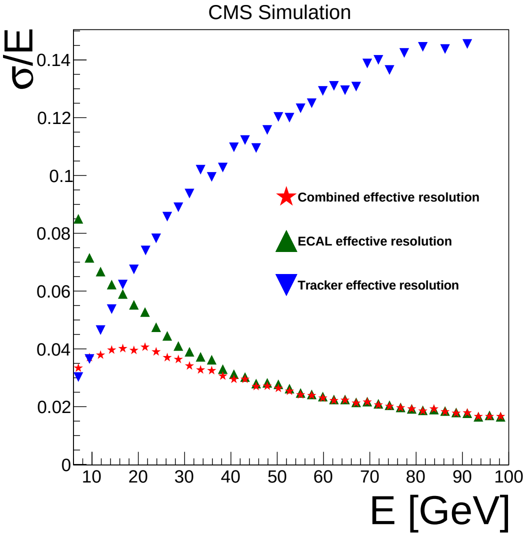


Figure 4.5: The electron energy resolution is estimated using the ECAL energy measurement only (green upward triangle), the momentum measurement from the track curvature in the magnetic field (blue downward triangle) and the combination of the two (red stars).

nuclear interactions and from data/simulation discrepancies in electron studies. The lack of material is concentrated in the high η region of the ECAL Barrel.

2) α parameter in response loss corrections

A more accurate level of calibration would require a knowledge of the α parameter at single the crystal level. The spread of the "true" α of the crystals translates into a worse resolution in data with respect to the simulation were the crystals have no transparency correction simulated.

4.2.3 Definition of electron energy for ECAL energy scale and resolution studies

The standard electron energy is obtained from the combination of the ECAL and the tracker measurements, so to take advantage of the better resolution of the ECAL at high energies and the better accuracy of the tracker momentum estimation at low energies.

The energy resolution as function of the electron energy is shown in Fig. 4.5 for the

ECAL only, the track momentum only, and the combined energy measurement.

Above 20 GeV the ECAL measurement dominates the energy combination.

Hereafter we will refer to electron energy has the energy obtained with the ECAL measurement only.

4.2.4 $Z \rightarrow e^+e^-$ event selection

Trigger

At level-1 the trigger requires two electromagnetic deposit in the ECAL with transverse energy $E_T > 8$ GeV The two level-1 seeds are then combined at the HLT level requiring two electromagnetic deposits with $E_T > 17$ GeV and $E_T > 8$ GeV respectively and with a pixel triplet defining a trajectory pointing to the electromagnetic deposit.

Offline selection

At offline level the events are required to have fired the selected trigger and both electrons have to fulfill the following requirements:

- have an $E_T > 25$ GeV,
- pass loose electron identification and isolation criteria
- be in the ECAL fiducial region: in the barrel away from the barrel-endcap transition region ($0 < |\eta| < 1.4442$); in the endcap away from the barrel-endcap transition region and within the tracker coverage ($1.566 < |\eta| < 2.5$).

The CMS standard electron identification criteria have been adopted in this analysis. This electron selection has been tuned on a simulated DY sample.

The selection is performed cutting on both electrons the following variables:

$\Delta\eta$: difference in pseudorapidity between the measured track associated to the electron at the vertex and the barycenter of the cluster position of the energy deposit in ECAL extrapolated to the vertex.

$\Delta\phi$: difference angle in the transverse plane with respect to the beam axis between the measured track associated to the electron at the vertex and the barycenter of the cluster position of the energy deposit in ECAL extrapolated to the vertex.

H/E: ratio between the energy deposit in ECAL and the deposit in HCAL in a cone of 0.15 centered around the seed of the ECAL cluster.

$\sigma_{i\eta,i\eta}$:

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

$$\text{where} \;\; \bar{\eta} = \frac{\sum w_i \eta_i}{\sum w_i} \qquad \text{and} \qquad w_i = \max\left[0 \quad ; \quad 4.7 + \log\left(\frac{E_i}{E_{5\times5}}\right)\right]$$

where the sum runs over the 5 × 5 crystal matrix around the most energetic crystal in the supercluster, and the η distances are measured in units of the crystal size in the η direction.

PF iso: isolation variable, based on the particle flow algorithm [1] [57], defined has a combination of the activity of charged, neutral particles around the considered electron after that the electron deposit has been properly removed.

conversion rejections: in order to reject electrons coming from conversions, the di-electron vertex position is compared with the closer primary vertex both in the transverse plane with respect to the beam axis ($d0vtx$) and in the beam axis direction ($dzvtx$). Furthermore, a tight requirement is done on the number of hits used for the track reconstruction: it is required that the number of hits missing in the pixel layers to be below a cut threshold.

The full set of electron identification and isolation criteria is shown in Tab. 4.1.

---

[1] The particle flow algorithm aims at reconstructing all stable particles in the event, i.e., electrons, muons, photons and charged and neutral hadron from the combined information from all CMS sub-detectors, to optimize the determination of particle types, directions and energies. The resulting list of individual particles can then be used, as if it came from a Monte Carlo event generator, to construct a variety of higher-level objects and observables such as jets, missing transverse energy (E^T_{miss}), taus, charged-lepton and photon isolation, b-jet tagging, etc.

Variable	Cut thresholds					
	EB			EE		
	loose	medium	tight	loose	medium	tight
	Electron identification variables					
$\Delta\eta$	0.007	0.004	0.004	0.009	0.007	0.005
$\Delta\phi$	0.15	0.06	0.03	0.10	0.03	0.02
$\frac{H}{E}$	0.12			0.10		
$\sigma_{\eta,\eta}$	0.01			0.03		
	Electron isolation variable					
PF iso	0.15		0.10	0.15(0.10)		0.10(0.07)
	Conversion rejection variables					
$d0vtx$	0.02			0.02		
$dzvtx$	0.2	0.1	0.1	0.2	0.1	0.1
miss. hits	1	1	0	1	1	0

Table 4.1: Electron selection criteria and cut thresholds are set differently in EB and EE due to the different amount of background as a function of the pseudorapidity and due to the constructive difference between the two parts. Three level of selection have been set with higher level of purity.

The main source of background for this analysis are electroweak processes like ($W \to e\nu$, $W \to \tau\nu$, $Z \to \tau^+\tau^-$), $t\bar{t}$ events, and QCD events.

The loose electron selection has an efficiency close to 90% and the residual background is negligible.

4.2.5 Simulation

Several large MC simulated samples are used for the signal:

Simulated process	generator	parton shower	$\sqrt{s}$
Drell-Yan decaying into leptons	MadGraph	pythia + tauola	8 TeV, 7 TeV
Drell-Yan decaying into electrons	powheg	pythia	8 TeV, 7 TeV
Drell-Yan decaying into electrons	sherpa	pythia	8 TeV

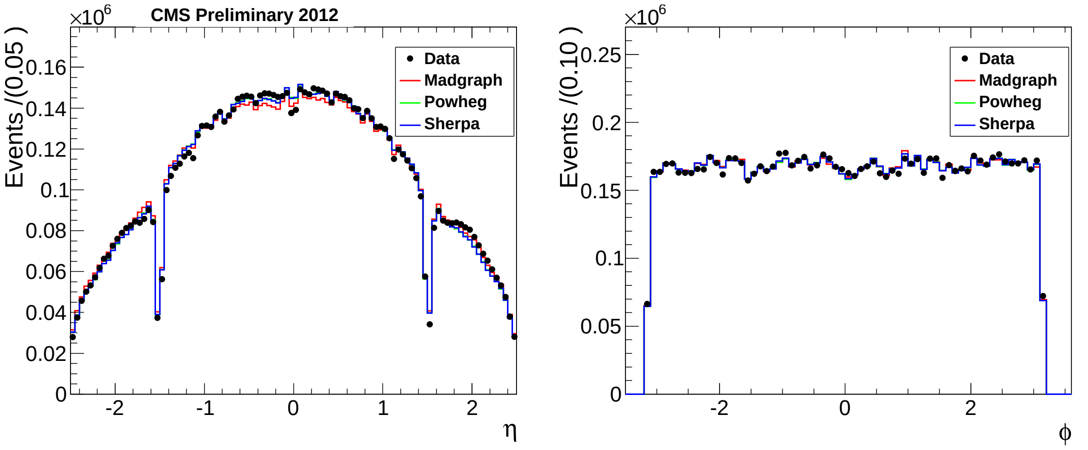
The 8 TeV MC samples have been produced with the run dependent conditions because of the ECAL noise increasing discussed in Sec. 3.7.1. The 7 TeV MC samples does not have run dependent conditions because the variation of the noise in 2011 have been smaller and not relevant for the analysis.

4.2.6 Comparison between data and MC samples

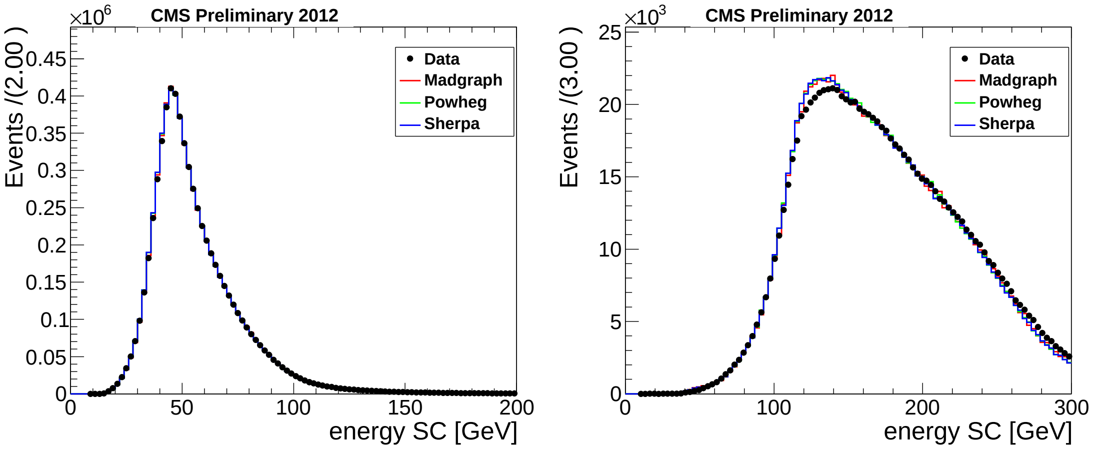
In order to use the three $Z \to e^+e^-$ simulated samples, a preliminary check on the most important variables used in the following analysis has been conducted.

The first set of variables are the electron kinematic variables, shown in Fig. 4.6: pseudorapidity (η), ϕ, energy and transverse energy (E_T) for events passing the selection described in Sec. 4.2.4.. For all the kinematic variables the MadGraph, the powheg and the sherpa samples are in excellent agreement with the data (except for the small but important discrepancy in the energy scale and resolution which is the subject of this chapter).

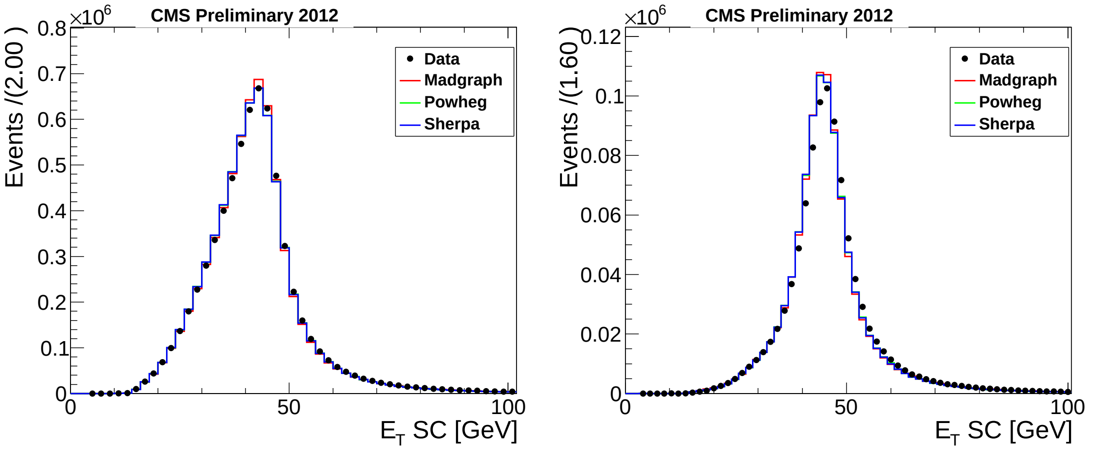
The run dependent MC samples, thanks also to the extended OOT pileup simulation, have almost cured the discrepancies in the EE present with the standard simulation, while a small shift in EB is still visible in Fig. 4.7.



(a) η of the electrons (b) ϕ of the electrons



(c) SC energy of the electrons in EB (d) SC energy of the electrons in EE



(e) E_T of the electrons in EB (SC energy) (f) E_T of the electrons in EE (SC energy)

Figure 4.6: Kinematic distributions of $Z \rightarrow e^+e^-$ electrons.

The effect on the energy scale and smearings due to this uncorrected shift has been taken into account as a systematic uncertainty.

No relevant discrepancies have been found between the different MC simulations.

4.2.7 Pile-up re-weighting

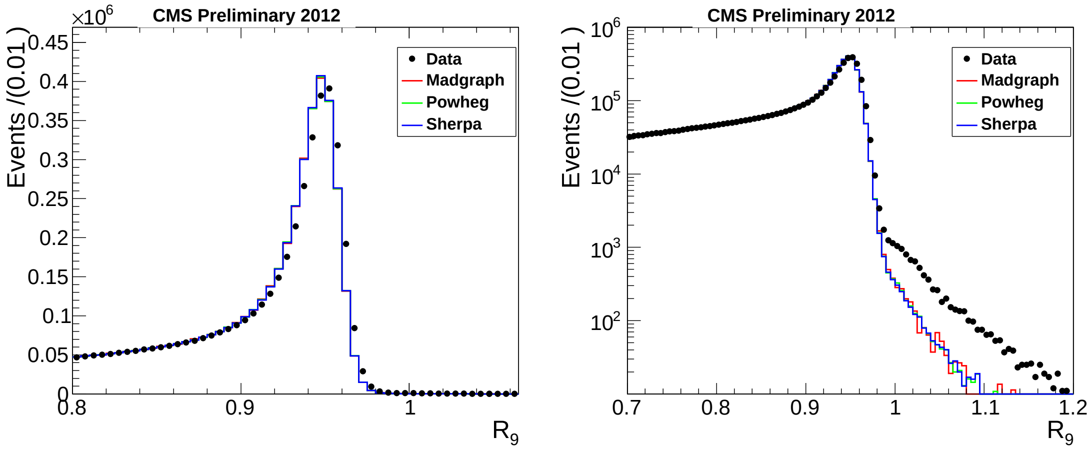
The simulation includes an accurate distribution of the number of interactions taking place in each bunch crossing.

Although the Deterministic Annealing primary vertex reconstruction [58] has been shown well-behaved up to the levels of pileup observed in 2012 data, the final distribution for the number of reconstructed primary vertices is still sensitive to the details of the primary vertex reconstruction and to the underlying event. Additionally, the distribution for the number of reconstructed vertices can be biased by the offline event selection criteria and even by the trigger. In order to factorize these effects, instead of re-weighting the Monte Carlo by the number of reconstructed Primary Vertices, we re-weight instead the number of pileup interactions from the simulation truth. The target pileup distribution for data is derived by using the per bunch-crossing-per-luminosity section instantaneous luminosity together with the total pp inelastic cross-section (69.4 mb) to generate an expected pileup distribution, correctly weighted by the per-bunch-crossing-per-lumi section integrated luminosity over the entire data-taking period.

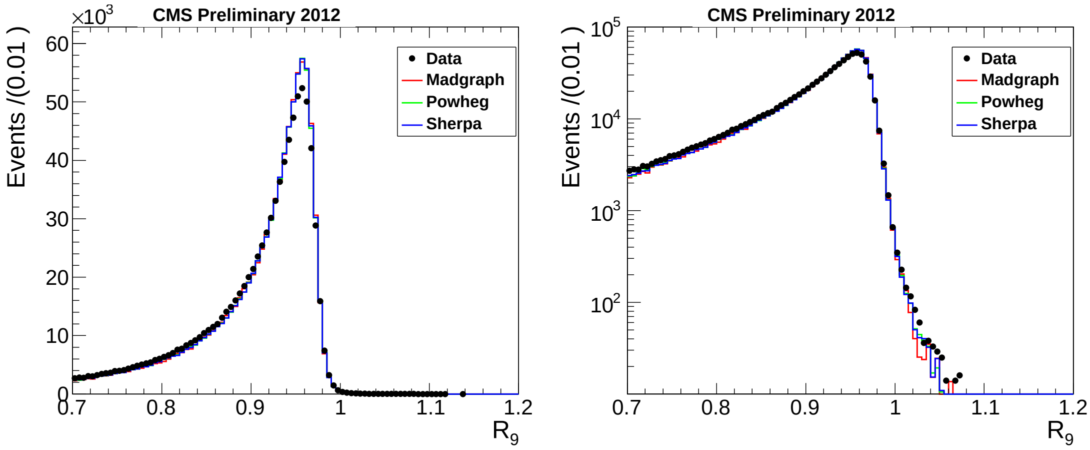
To validate the weighting technique that is applied to the MC in order to match the actual number of pile-up events distribution observed in the data, the comparison of the number of reconstructed vertices using the Deterministic Annealing algorithm is compared between the data and the MC. The distribution is shown in Fig. 4.8.

Although in the run dependent MC, the pileup has been tuned to match the data in each of period, a pileup re-weighting is in any case performed in order to fix the small residual differences.

In Fig. 4.8a is reported the number of reconstructed vertices in data and MC after the reweighting for the different samples. The agreement is excellent, indicating also no not-simulated bias due to the trigger and offline selection.

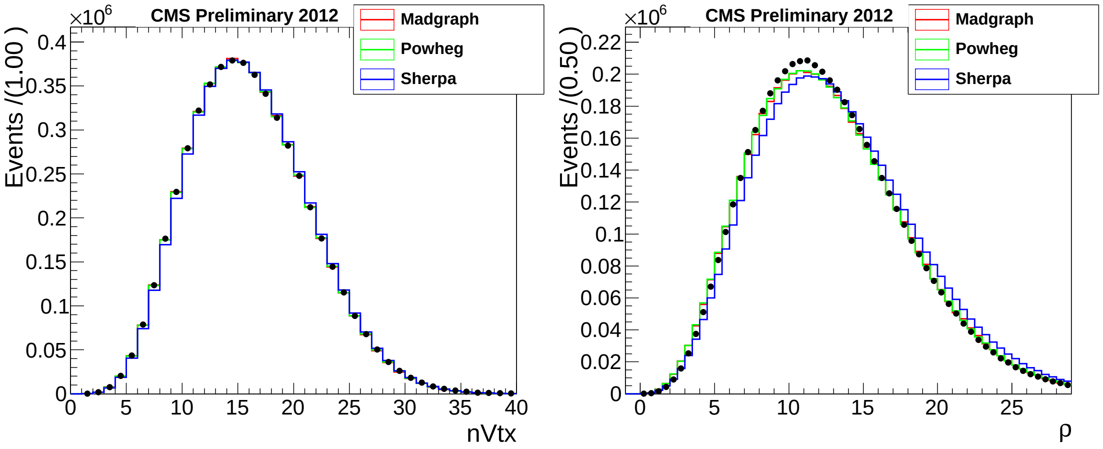


(a) R_9 of the electrons in EB (b) R_9 of the electrons in EB



(c) R_9 of the electrons in EE (d) R_9 of the electrons in EE

Figure 4.7: R_9 distribution for $Z \rightarrow e^+e^-$ electrons after full selection in linear and logarithmic scale. Data are shown by black dots and MC simulations are shown by solid lines.



(a) nPV of the electrons

(b) ρ of the electrons

Figure 4.8: Number of reconstructed primary vertices (4.8a) and average energy deposits in the calorimeters due to the underline event and OOT pileup (4.8b).

4.2.8 Fit Method

The *fit method* provides a fast and reliable method to have information about the shift of the Z peak position with respect to the simulation and about the width of the invariant mass distribution.

The *fit method* consists in fitting the Z lineshape with a proper function to obtain the energy scale and the experimental resolution from the fit parameters. Crucial is the choice of a proper function that can take into account not only the “ideal” Z lineshape represented by a Breit-Wigner (BW) distribution, but also the effect of the experimental resolution.

The fit function chosen is a convolution of a Breit-Wigner distribution with a Crystal Ball (CB) [59, 60, 61] that is a Gaussian with a power-law low-mass tail. The Crystal Ball function is considered a good resolution function since it describes well the low energy tail of the electron energy distribution mostly due to the bremsstrahlung photon emission. Furthermore, the integral of the CB function is known analytically, making the fit of the convoluted function faster.

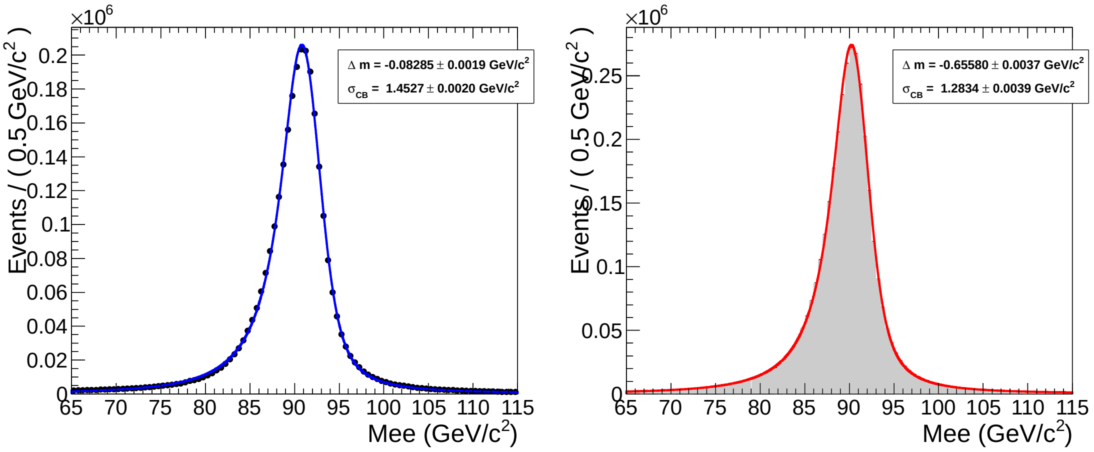
The CB function is defined in Eq. 4.2 where the α parameter defines the transaction between the Gaussian with mean Δm and standard deviation of σ_{CB} and the parameter

n is the order of the power-law of the low mass tail.

$$f(x;\alpha,n,\Delta m,\sigma_{CB}) = N\begin{cases} gaus(\Delta m,\sigma_{CB}) & \text{if } \frac{x-\Delta m}{\sigma_{CB}} > -\alpha \\ A\cdot(B-\frac{x-\Delta m}{\sigma_{CB}})^{-n} & \text{if } \frac{x-\Delta m}{\sigma_{CB}} \leq -\alpha \end{cases}$$
$$\text{where} \tag{4.2}$$
$$A = \left(\frac{n}{|\alpha|}\right)^{n}\cdot\exp\left(-\frac{|\alpha|^{2}}{2}\right)$$
$$B = \frac{n}{|\alpha|} - |\alpha|$$

The BW parameters are fixed to the nominal Z mass (M_Z) and width as reported in the Particle Data Group (PDG) [56]. The peak shift of the CB with respect to zero (Δm) is related to the energy scale correction.

Since the width of the Z invariant mass distribution is the convolution of the intrinsic Z width and the energy resolution, the *fit method* with the BW-CB convoluted pdf is a proper choice to extract the experimental energy resolution that is represented in a first approximation only by the σ_{CB} parameter of the CB.



(a) Data events

(b) Simulated events

Figure 4.9: Z line shape fit examples with *fit method* for events with both electrons in EB. The fits are performed with the *fit method* using a BW-CB convoluted *pdf*.

4.2.9 Energy scale correction and experimental resolution estimation

The energy scale correction is defined as the correction to be applied to the electron energy in data in order to have the Z peak to the expected position in the simulation.

The peak shift of the CB function (Δm) should then be the same in data and simulation. The relative shift between data and simulation is defined as

$$\Delta P = \frac{\Delta m_{data} - \Delta m_{MC}}{m_Z} \tag{4.3}$$

When both electrons belong to the same category, $1 - \Delta P$ is the per-electron scale correction in that category (assuming one constant correction for all the electrons belonging to the same category), otherwise it is the geometrical mean of the two electron corrections.

The relative resolution $\frac{\sigma_{CB}}{peak_{CB}}$ of the di-electron category is related to the single electron energy resolution when both electrons belong to the same category:

$$\frac{\sigma_{CB}}{peak_{CB}} = \sqrt{2}\frac{\sigma_E}{E}$$

The difference in energy resolution between data and simulation is defined in terms of additional smearing to be added to the single electron energy in the simulation in order to reproduce the resolution in data.

The single electron energy additional smearing factor is:

$$\Delta\sigma = \sqrt{2 \cdot \left(\left(\frac{\sigma_{CB}}{peak_{CB}}\right)^2_{data} - \left(\frac{\sigma_{CB}}{peak_{CB}}\right)^2_{MC} \right)} \tag{4.4}$$

4.2.10 Uncertainties on peak position and experimental resolution

The estimation of the peak shift Δm and the experimental resolution σ_{CB} have been checked against variation of the electron selection (loose vs medium), binning of the invariant mass distribution using 3 different binnings, 0.5 GeV, 0.25 GeV and 0.75 GeV, varying the invariant mass window for the fit $(60 - 120, 65 - 115, 70 - 110)$. The uncertainties have been evaluated for in $|\eta| \times R_9$ categories defined in Sec. 4.2.2.

The larger contribution to the Δm measurement is due to the particular choise of the invariant mass range for the fit, whilst no uncertainty is associated to the electron selection. The systematic uncertainty is 0.01 GeV for $R_9 < 0.94$ electrons and 0.02 GeV for $R_9 > 0.94$ electrons.

Injection of artificial scale shifts

The MC sample has been used to check the linearity of the correction defined by Eq. 4.3. The electron energies have been shifted by $\Delta P = \pm 1\%, \pm 2\%, \pm 5\%$ and $\pm 10\%$, giving a new invariant mass equal to $M_{new} = m_0 \cdot \Delta P$. The Δm of the CB is expected to be linearly correlated with the injected scale shifts.

$$\begin{aligned} M_{new} = M_Z + \Delta m_{new} = (M_Z + \Delta m_0) \cdot \Delta P \rightarrow \\ \rightarrow \Delta m_{new} = (M_Z + \Delta m_0) \cdot \Delta P - M_Z \end{aligned} \tag{4.5}$$

The Δm measured for the eight $|\eta| \times R_9$ categories after the injection of the artificial energy scales is shown in Fig. 4.10. For each category the values of Δm have been shifted for a better readability of the results.

In the following table, the slopes normalized by the $M_Z + \Delta m_0$ are shown.

Category	slope / $M_Z + \Delta m_0$
$\lvert\eta\rvert < 1.0$ $R9 < 0.94$	0.994 ± 0.000
$\lvert\eta\rvert < 1.0$ $R9 > 0.94$	0.990 ± 0.001
$\lvert\eta\rvert > 1.0$ $R9 < 0.94$	0.990 ± 0.001
$\lvert\eta\rvert > 1.0$ $R9 > 0.94$	0.994 ± 0.016
$\lvert\eta\rvert < 2.0$ $R9 < 0.94$	0.993 ± 0.002
$\lvert\eta\rvert < 2.0$ $R9 > 0.94$	0.994 ± 0.005
$\lvert\eta\rvert > 2.0$ $R9 < 0.94$	0.996 ± 0.002
$\lvert\eta\rvert > 2.0$ $R9 > 0.94$	0.997 ± 0.001

The maximum bias among the different categories is 1%, that implies a bias in the corrected electron energy of 0.01% for a correction of 1%. The bias is of the same order of the statistical uncertainty on the energy scale corrections derived for the $H \rightarrow \gamma\gamma$ analysis.

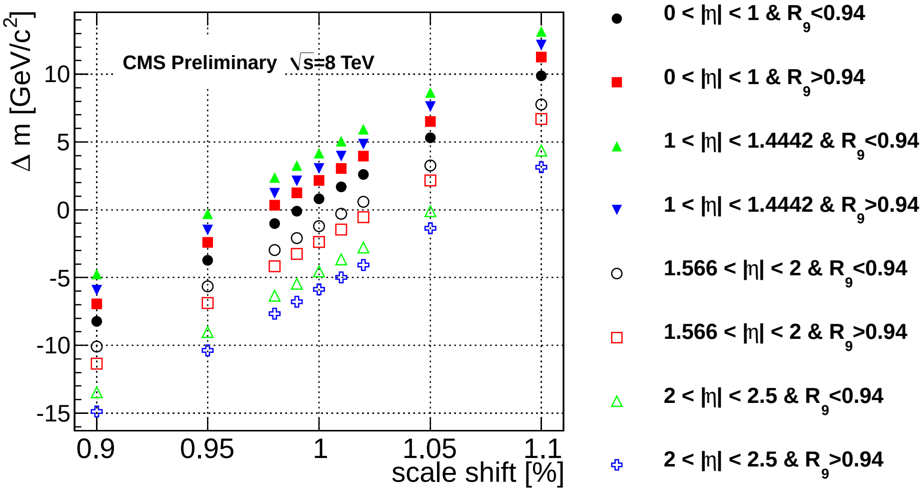


Figure 4.10: The electron energy have been shifted by a scale factor of $\pm1\%, \pm2\%, \pm5\%, \pm10\%$ in the MC sample. The correlation of the scale shift and the peak of the CB function (Δm) is checked in various electron categories. The different categories have been shifted in Δm for a better readability of the figure.

4.3 Smearing method

The smearing method represents an alternative method to extract directly the energy scale and the additional smearing for single electron categories.

The *smearing method* consists in using the invariant mass shape in the MC simulation as a *pdf* for a maximum likelihood fit instead of using an a-priori chosen function as in the *fit method*. The main advantage is that all the known detector effects, reconstruction inefficiencies and Z kinematic behaviour are already taken into account in the simulation. The main hypothesis behind this procedure is that the discrepancy between the data and the simulation can be described by a smearing function. A Gaussian smearing is shown in the following to adequately reproduce the additional smearing needed in the simulation to describe the data in all the categories. In addition, another advantage of the method consists in the possibility to define a large number (n) of single electron categories with arbitrary selection cuts.

The Z events are consequently grouped in $N = \frac{n\cdot(n+1)}{2}$ di-electron invariant mass distributions that can be divided into “diagonal” ($H_{i,i}$) with both electrons belonging to

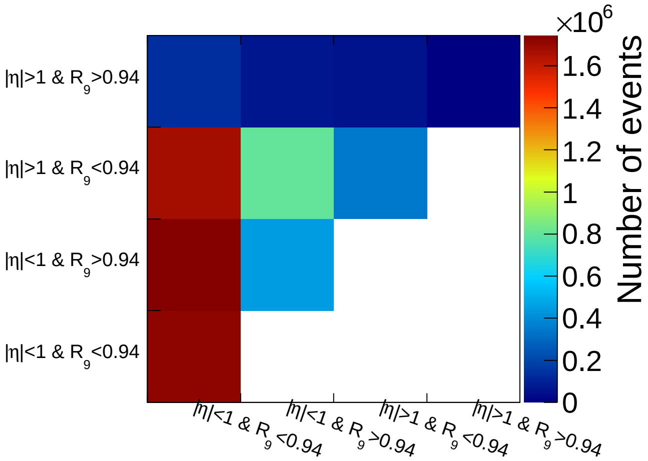


Figure 4.11: Distribution of the number of events between 4 barrel categories: $|\eta| < 1.0$ $R9 < 0.94$ $|\eta| < 1.0$ $R9 > 0.94$ $|\eta| > 1.0$ $R9 < 0.94$ $|\eta| > 1.0$ $R9 > 0.94$. A very low number of events have electrons in the outer part of the barrel with low interaction with the material upstream ECAL.

the same category and "off-diagonal" ($H_{i,j}$ with $i \neq j$) with $H_{i,j} = H_{j,i}$. For each single electron category one scale shift (ΔP) and one additional smearing ($\Delta\sigma$) are defined. As a consequence, the "off-diagonal" invariant mass distributions are fully exploited entering in the likelihood calculation. The relevance of the "off-diagonal" distributions grows with the increasing number of defined single electron categories because only a small fraction of the events falls in the "diagonal" distributions. In Fig. 4.11 the number of events for all the "diagonal" and "off-diagonal" categories are shown. The statistical uncertainty on the energy scale and the additional smearing for the $|\eta| > 1.0$ $R9 > 0.94$ is considerably reduced with the *smearing method* with respect to the *fit method* thanks to the additional usage of events belonging to "off-diagonal" distributions.

All the di-electron invariant mass distributions are built for data and for the simulation. In the latter case, the single electron energy used in the invariant mass calculation is corrected by a Gaussian smearing function: $E_{new} = E_{old} \cdot gaus(1 + \Delta P, \Delta\sigma)$. where ΔP is the relative energy scale shift between data and simulation defined in Eq. 4.3. The Gaussian is shifted by $1 + \Delta P$ being $1 - \Delta P$ the scale correction to be applied to the data to match the simulation. The width of the Gaussian is the additional smearing to be measured.

Starting from the single contributions to the energy resolution (Eq. 4.1), it results natural to parametrize the additional smearing as the sum of three terms: additional noise term, additional stochastic term and additional constant term as shown in Eq. 4.6

$$\Delta\sigma = \frac{\Delta N}{E} \oplus \frac{\Delta S}{\sqrt{E_T}} \oplus \Delta C \tag{4.6}$$

The stochastic term is scaled by $\sqrt{E_T}$ instead of $\sqrt{E}$ for consistency with the choice done later in Sec. 4.4.2. The additional smearing can be parametrized as arbitrary functions of the energy or other additional variables, but the actual sensitivity is not yet enough to discriminate between different parametrizations (as the one function of E and the one as function of E_T). The single contributions to the additional smearing and the energy scale can then be defined also as functions of further variables.

The determination of these parameters is performed in a multidimensional space N minimizing the negative log likelihood (NLL) defined as the likelihood of having $d_{i,k}$ events in the k-th category and in the i-th bin of the invariant mass distribution. The bin-by-bin likelihood is defined by the binomial distribution:

$$\mathcal{L}^{i,k} = p_{i,k}^{d_{i,k}} \cdot (1 - p_{i,k})^{N_k - d_{i,k}} \tag{4.7}$$

where $p_{i,k}$ is the number of events in the MC histogram of the k-th category in the i-th bin after having normalized the histogram area to 1 (the MC histogram is a *pdf*). The expected number of events $p_{i,k}$ is given by the scale and smearing.

For each evaluation of the likelihood, it is necessary to run over the whole MC events, scale and smear the electron energy, build the invariant mass distributions and evaluate the bin-by-bin likelihood. Contrary to the *fit method*, the statistical power is not spoiled out when a high number of categories is defined, but the main limitation is the likelihood computation time.

4.3.1 Mitigation of the likelihood fluctuations

As already mentioned, in the *smearing method* procedure, the *pdf* in the likelihood calculation is given by the MC invariant mass distribution. As a consequence, the likelihood evaluated for the same scale and smearing values for a particular category can fluctuate due to the different possible realization of the MC invariant mass distribution. This effect is intrisic in the method due to its stochastic nature.

The limited number of events in the MC simulation becomes more and more important with the higher granularity in the category definition because the invariant mass distributions in the MC become poorly populated.

In order to use the MC invariant mass distribution as a *pdf*, the lineshape has to be as smooth as possible. The statistical uncertainty on the MC invariant mass bins have to be negligible with respect to the statistical uncetainty in data for the same bin. For this reason, in the simulation, the number of simulated events should be at least one order of magnitude higher than the data. So much large DY MC samples have not been produced so far in CMS. An over-sampling of the MC sample has been adopted to reduce likelihood fluctuations.

Over-sampling

For each simulated electron, the energy is smeared by a random number extracted from the Gaussian distribution. The over-sampling consists in smearing the same electron many times with different extraction of the random number, given the same scale and smearings. The invariant mass histogram is then filled by the same electrons many times and the result is an invariant mass distribution averaged over all the possible smearing realizations. The likelihood fluctuations discussed above are considerably reduced by the over-sampling technique adopted.

Anyway, the dependency on the first realization of the MC (the original unsmeared invariant mass) is not cancelled but only mitigated as shown in Fig. 4.12 where the fluctuation around 95 GeV present in the original MC distribution is still present after the smoothing with the over-sampling.

The likelihood ($\mathcal{L}$) fluctuations represent one of the most difficult part to tune in the

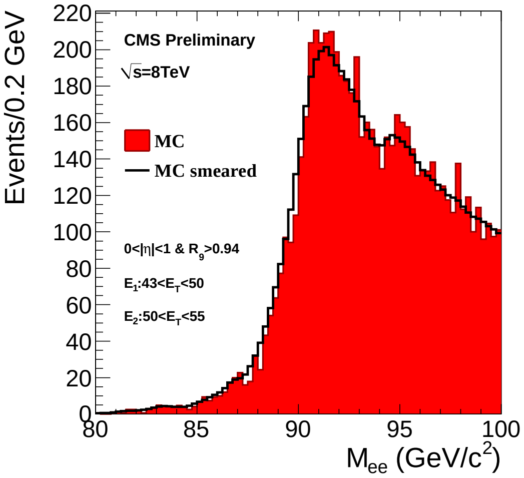


Figure 4.12: Invariant mass distribution for two electrons in the central barrel region ($|\eta| < 1.0$), one electron with $43 < E_T < 50$ GeV, the second with $50 < E_T < 55$ GeV and both with $R_9 > 0.94$. The peak around 95 GeV in the original invariant mass distribution (filled red histogram) is smoothed by the smearing, but still present in the smeared histogram (black line).

smearing method. The noise represented by these fluctuations should be kept much below the $\Delta(-\ln\mathcal{L}) = 0.5$ in order to neglect it in the statistical uncertainty determination of the scale and smearings. Unfortunately, this is not possible due to the high computational power required. A trade-off between the need to scan more points in the multidimensional phase space and the required time to perform the minimization of the $-\ln\mathcal{L}$ (NLL) has been found using a deterministic smearing described in the following.

Deterministic smearing

The random numbers extracted for the Gaussian smearing for each electron are generated only once. For each electron, a set of random numbers is generated once and then reused at each likelihood calculation. With the deterministic smearing, the time needed to evaluate the likelihood has been drastically reduced and the likelihood calculated for the same point in the N multidimensional parameter space has the same value by construction. The likelihood fluctuations described so far are not removed, but only masked. The number of over-sampling needed to keep $NLL < 1$ was greater than 200, with the deterministic smearing a smooth invariant mass distribution can be obtained with 20 over-samplings with a 90% reduction in the execution time.

4.3.2 E_T dependent energy scale

The introduction of an E_T categorization would not be possible with the *fit method*, because of the effect of the E_T cut on the Z lineshape. Fig. 4.12 shows the distortion of the lineshape in one category with a boosted Z and two high E_T electrons. The Z peak is still visible, but the non resonant component of the DY formes the high shoulder in the high invariant mass region due to the E_T threashold. The *smearing method* is perfectly capable to deal with anomalous invariant mass shapes, also in small region of the Z kinematic phase space because it uses the MC invariant mass distribution as the *pdf* in the likelihood calculation.

In the case of the E_T binning, not all the possible di-electron categories are used in the likelihood calculation: the ones with very low statistics and/or with a shoulder in the invariant mass distribution are excluded. These categories have a very low sensitivity to the scale and the smearing and they would have introduced additional fluctuations in the likelihood without any gain in the parameter determination.

A threshold of 2000 events has been set as the minimum number of events needed to have a smooth invariant mass distribution, therefore categories with less than 2000 events are excluded from the likelihood calculation. Diagonal categories are accepted if they have at least 1000 events because they add more sensitivity to the determination of the parameters and because they are less correlated with other categories (both electrons belong to the same category).

The non resonant component of the DY process is shaped in the di-electron invariant mass distribution by the E_T bin cuts. The categories where this component alters significantly the peaked shape of the Z lineshape are removed from the likelihood calculation. The exclusion is performed whenever the difference in height between the right and the left tails of the di-electron invariant mass distribution exceed 20% of the peak height. With this requirement, the migration of electrons in and out of the invariant mass window considered in the analysis is considerably reduced and consequently also the likelihood fluctuations.

4.3.3 Minimization algorithm

Given the long time needed for each likelihood evaluation (∼ 30 s) it is mandatory that the algorithm performs the minimum number of steps to reach the convergence and it should be robust against the likelihood fluctuations. None of the tested algorithms (MIGRAD tool, genetic algorithm, Markov Chain MC) resulted both robust and fast enough. An iterative minimization procedure have been developed: the algorithm performes successively a 1D likelihood scan for a given parameter and minimize it. An example of NLL 1D scan is shown in Fig. 4.13.

The minimization is performed iteratively until no new minimum is found. The correlation between the categories is taken into account by the iterative minimization.

The main assumptions of the method are:

- the presence of one global minimum and no other local minima
- an almost parabolic behaviour of the minimum in the NLL

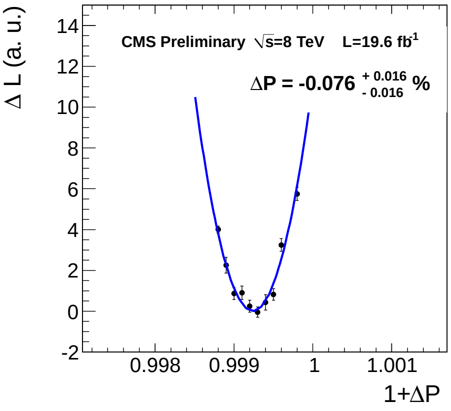


Figure 4.13: Example of NLL 1D scan.

Both this assumption are reasonably true for the scale and smearing derivation.

In order to increase the efficiency of the minimization, it has divided into two steps, the first fixing the smearing values to some reasonable initial values and minimizing against the scales only, in the second step minimizing all the parameters together. It has been verified that the minima for the scale parameters converges stably to the final value already after the first step because of the higher sensitivity to the scales with respect to the smearings.

The iterative minimization though 1D likelihood scans present a big advantage also in minimizing the time need for each likelihood calculation. It is in fact possible to store the

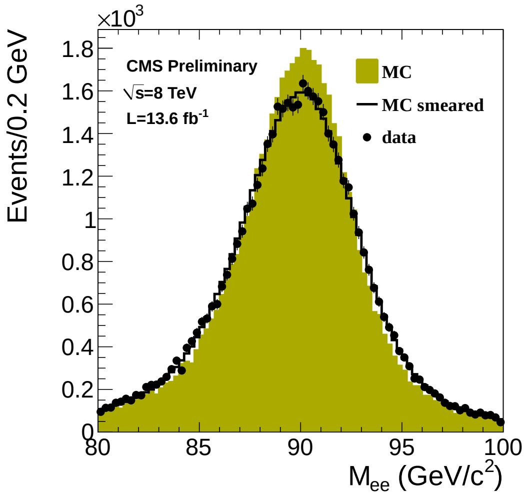


Figure 4.14: Z invariant mass distribution with both electrons in $2.0 < |\eta| < 2.5$, one with $R_9 < 0.94$ and the other with $R_9 > 0.94$. The simulation is shown in solid filled histogram, the data by points and the simulation after the application of the additional smearing by the black line.

previously calculated NLL for each category and update it only for those categories whose parameters have been changed. Moving only one parameter has then the advantage that the NLL is updated only for few categories. The time needed to calculate the NLL for each point of the scan results reduced to only few seconds.

In Fig. 4.14 the invariant mass distribution of two electrons in $2.0 < |\eta| < 2.5$, one with $R_9 < 0.94$ and the other with $R_9 > 0.94$ is shown for the simulation (solid filled histogram), the data (points) and the simulation after the application of the additional smearing.

4.4 Energy scale corrections and additional smearing derivation

The derivation of the energy scale corrections and the additional smearings is discussed in this section. Two approaches have been followed: one we will refer to as "traditional approach" and a second one we will refer to as "improved approach".

The energy scale corrections in the traditional approach are fuction of the time, the pseudorapidity and the level of interaction of the electron with the upstream material (R_9); in the improved approach they are function also of the E_T of the electrons in order to correct for non-linearity effects on the energy scale.

The additional smearings are derived in $|\eta| \times R_9$ categories in the both approaches, but they have a different parametrization: in the traditional approach they are measured as additional constant terms to the energy resolution, whilst in the improved approach the additional smearing includes also an additional stochastic term contribution.

The improved approach has been used only for the 8 TeV data where the number of events allows the study with finer categorization of the electron sample. Due to the limited number of events in the ECAL Endcap also for the 8 TeV data, the study with the improved approach is limited to the EB region.

4.4.1 Energy scale corrections

The energy scale correction is derived in two steps, at step i the corrections of step $i-1$ are applied:

1) Correction of residual time dependence of the scale for η regions: c_1

   The aim of this step is to remove the residual variations as a function of the time due to imperfect transparency corrections. Given the different radiation doses, this step is done in the $|eta|$ bins.

2) Correction of energy scale for $\eta \times R_9$ categories: c_2

   energy scale corrections for residual data/MC difference depending on the level of interaction with the upstream material. This step is performed using the full data sample, without time binning.

3) Correction of energy scale for $\eta \times R_9 \times E_T$ categories: c_3 (only for electrons in the EB in 8 TeV data)

The final energy scale corrections are then function of runNumber, η, R_9: $c = c_1 \times c_2$ for 7 TeV data (EB and EE) and 8 TeV data (EE), and runNumber, η, R_9, E_T: $c = c_1 \times c_2 \times c_3$ for 8 TeV data (EB). The third step c_3 is not possible on 7 TeV data because such granularity in the categorization need at least 3 times the available number of events.

Scale corrections c_1

The corrections c_1 are derived using the *fit method*. The dataset is divided into a large number of bins in time, all with almost the same number of $Z \rightarrow e^+e^-$ events. The fit is performed in each time bin for four regions: $|\eta| < 1.0$, $1.0 < |\eta| < 1.4442$, $1.566 < |\eta| < 2.0$, $2.0 < |\eta| < 2.5$.

Once the c_1 corrections are applied, the Z peak is stable in time in the four regions.

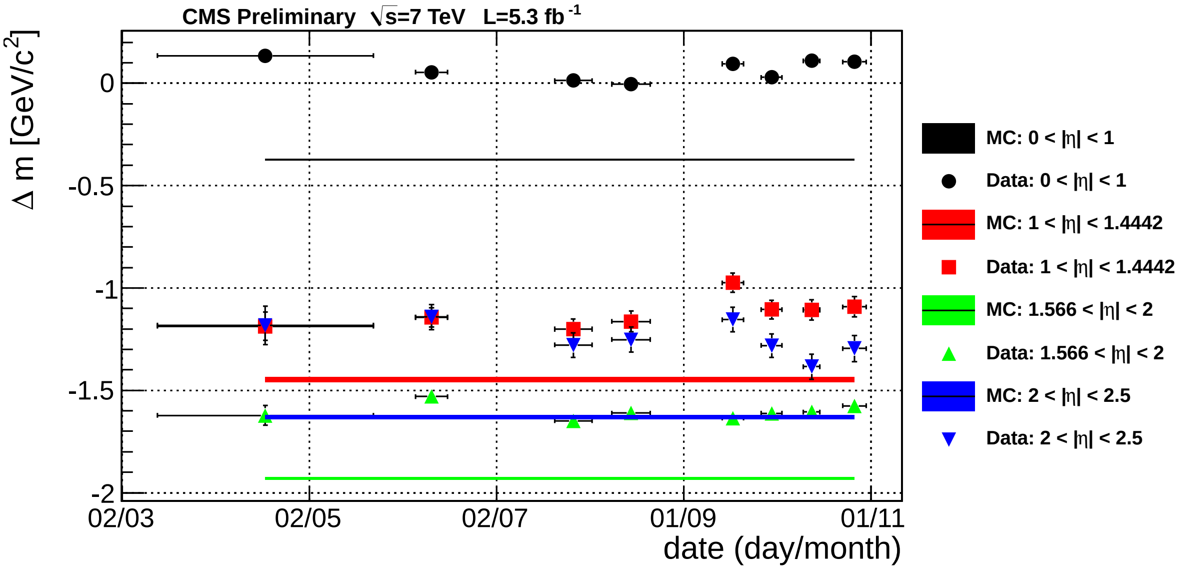
The Z peak stability before any scale correction is shown in Fig. 4.15a for the 7 TeV data and in Fig. 4.16a for the 8 TeV data in the four $|\eta|$ regions.

The absolute ECAL scale in the last reconstruction of the 8 TeV data has been tuned on $Z \rightarrow e^+e^-$ events using the standard CMS SC energy (defined in Sec. 3.4.1) with the same time granularity reported in Fig. 4.16. Since the standard SC energy is not pileup robust, in the absolute scale has been introduced a pileup dependence. The energy corrected by the regression (described in Sec. 3.4.2) is on the contrary more pileup safe. The regression energy results then double corrected for the pileup effect as function of time, with the trend shown in Fig. 4.16a.

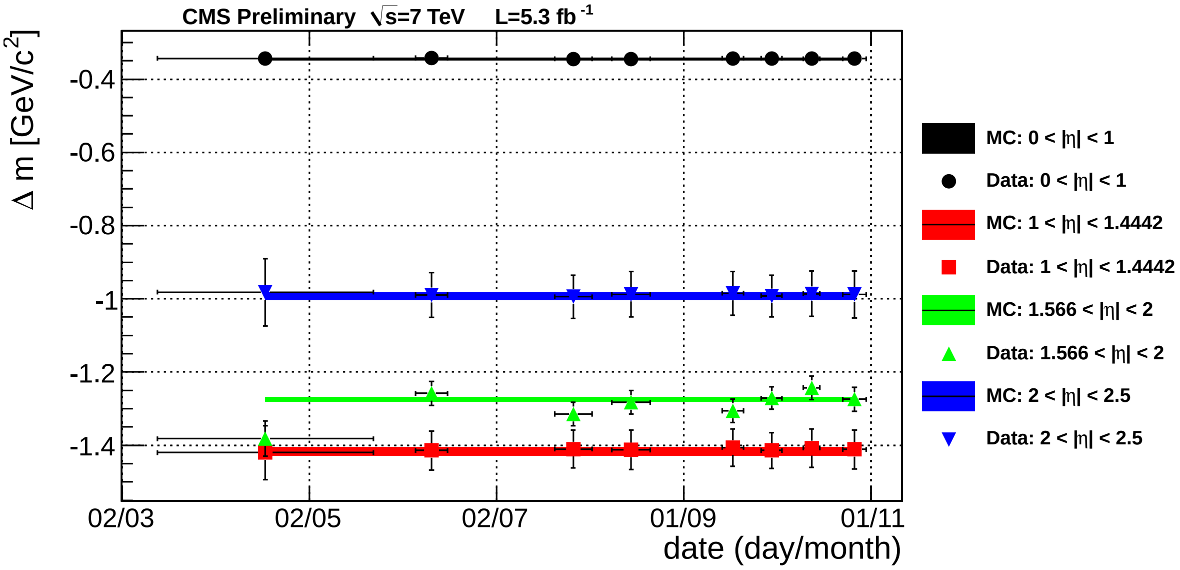
After the application of the c_1 corrections, the second step in the scale correction derivation and the additional smearing measurement are performed simultaneously with the *smearing method*.

Scale corrections c_2

The c_2 corrections are derived separately for EB and EE. There is no advantage in adding the EE categories for the determination of the EB scale corrections due to the worse resolution of the EE categories. The sensitivity to the scale and smearings is in fact enhanced with the use of high resolution categories. For the same reason, in the determination of the EE scales and smearings, the best resolution category of the barrel is added. “Off-

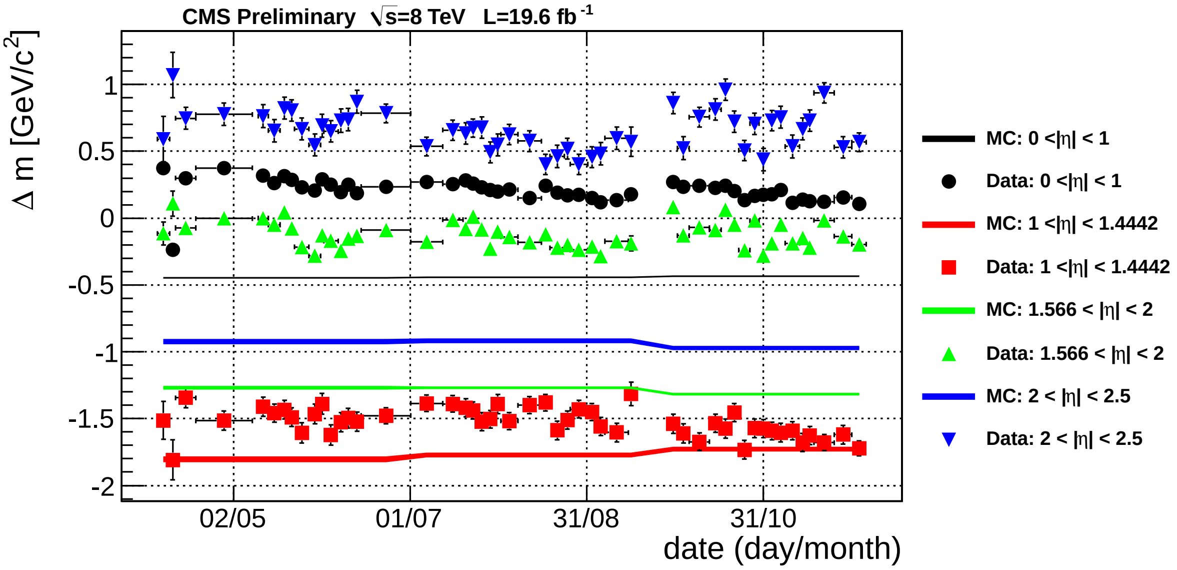


(a) Z peak stability before energy scale corrections in four $|\eta|$ regions. The Z peak in the simulation is shifted in different regions because the electron energy has been corrected by the algorithmic corrections tuned for photons (the same corrections used in $H \to \gamma\gamma$ analysis).

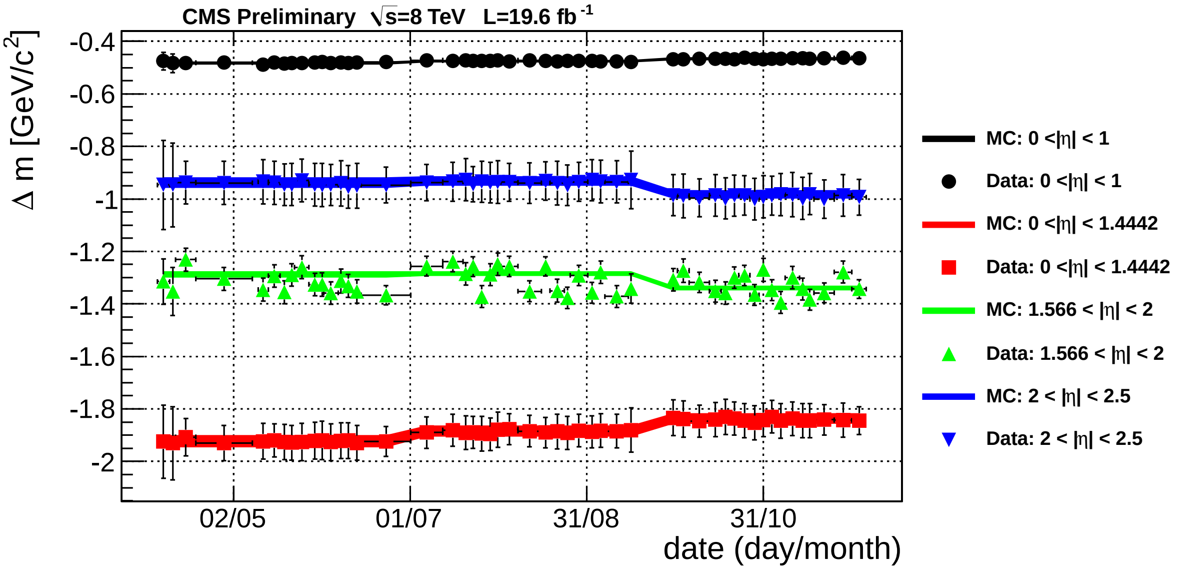


(b) Z peak stability after the c_1 corrections obtained after the first step of the energy scale correction derivation.

Figure 4.15: Z peak position stability before (top) and after (bottom) energy scale corrections derived in the first step of the procedure (c_1) with 7 TeV data (2011)



(a) Z peak stability before energy scale corrections in four $|\eta|$ regions. The Z peak in the simulation is shifted in different regions because the electron energy has been corrected by the algorithmic corrections tuned for photons (the same corrections used in $H \to \gamma\gamma$ analysis).



(b) Z peak stability after the c_1 corrections obtained after the first step of the energy scale correction derivation.

Figure 4.16: Z peak position stability before (top) and after (bottom) energy scale corrections derived in the first step of the procedure (c_1) with 8 TeV data (2012). The Z peak in simulation has few steps due to the different conditions that have been used in the run dependent simulation.

diagonal" categories with one electron in EB and the other in EE are then improving the statistical precision of the measurement. In Fig. 4.17 the distribution of the number of events in each di-electron category for the c_2 corrections is shown.

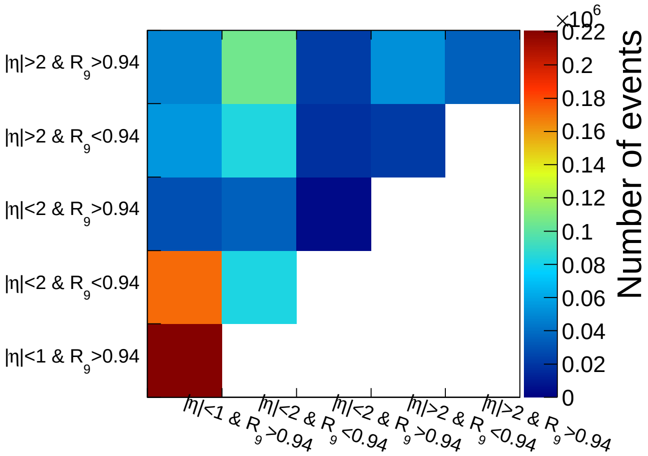


Figure 4.17: Distribution of the number of events with electrons belonging to the high resolution barrel category and the four endcap categories: $|\eta| < 1.0$ $R9 > 0.94$, $|\eta| < 2.0$ $R9 < 0.94$, $|\eta| < 2.0$ $R9 > 0.94$, $|\eta| > 2.0$ $R9 < 0.94$, $|\eta| > 2.0$ $R9 > 0.94$.

In Fig. 4.18 the data are shown before and after the c_2 corrections for the off-diagonal category with one electron in $|\eta| < 1.0$ $R9 > 0.94$ and the other in $1.0 < |\eta| < 1.4442$ $R9 > 0.94$. The Z peak is corrected to the position in the simulation. The Z peak position before and after the c_2 corrections is shown in Fig. 4.19 as function of the electron pseudorapidity obtained with the *smearing method*.

Scale correction c_3

The scale corrections in the improved derivation scheme are defined in the same categories of the traditional corrections but with further E_T categorization in EB:

- EB $|\eta| < 1.0$ & $R_9 < 0.94$: $20 < E_T < 33$, $33 < E_T < 39$, $39 < E_T < 45$, $45 < E_T < 50$, $50 < E_T < 58$, $58 < E_T < 100$
- EB $|\eta| < 1.0$ & $R_9 > 0.94$: $20 < E_T < 35$, $35 < E_T < 43$, $43 < E_T < 50$,

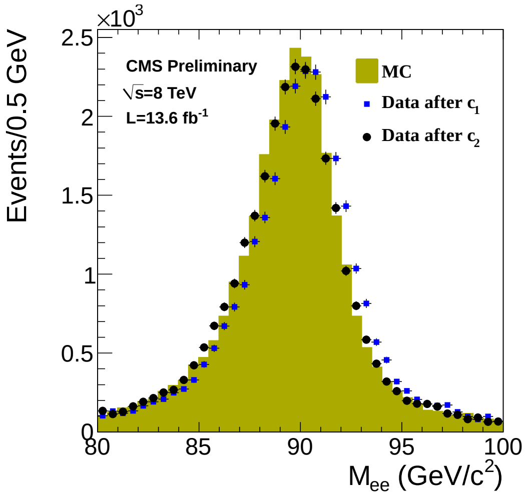


Figure 4.18: Invariant mass distibution for off-diagonal category with one electron in $|\eta| < 1.0$ $R9 > 0.94$ and the other in $1.0 < |\eta| < 1.4442$ $R9 > 0.94$. The additional smearing to the simulation (solid filled histogram) is not applied. Data before (squares) and after (circles) the c_2 corrections are shown.

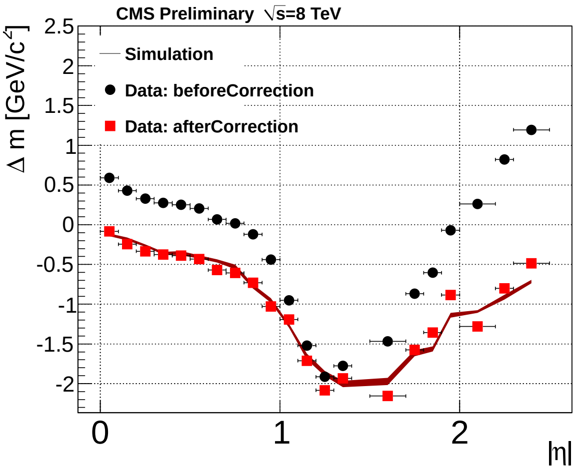


Figure 4.19: Z peak position as function of $|\eta|$ before (black circles) and after (red squares) c_2 energy scale corrections.

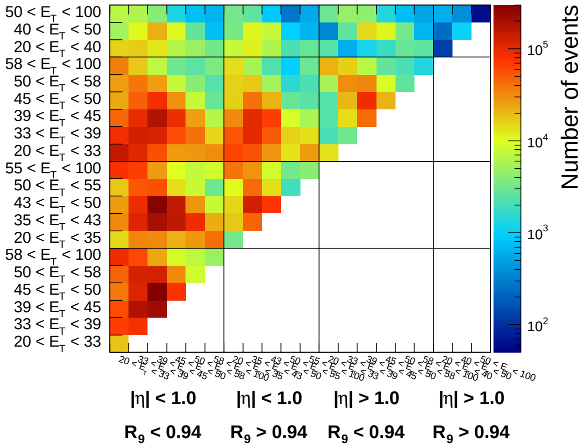


Figure 4.20: Distribution of the number of events for the di-electron categories defined in the improved approach scale and smearing derivation. A very small number of events have electrons in the outer part of the barrel with low interaction with the material upstream ECAL.

$50 < E_T < 55$, $55 < E_T < 100$

- EB $1.0 < |\eta| < 1.4442$ & $R_9 < 0.94$: $20 < E_T < 33$, $33 < E_T < 39$, $39 < E_T < 45$, $45 < E_T < 50$, $50 < E_T < 58$, $58 < E_T < 100$
- EB $1.0 < |\eta| < 1.4442$ & $R_9 > 0.94$: $20 < E_T < 40$, $40 < E_T < 50$, $50 < E_T < 100$
- EE $1.566 < |\eta| < 2.0$ & $R_9 < 0.94$
- EE $1.566 < |\eta| < 2.0$ & $R_9 > 0.94$
- EE $2.0 < |\eta| < 2.5$ & $R_9 < 0.94$
- EE $2.0 < |\eta| < 2.5$ & $R_9 > 0.94$

The energy scale corrections as function of the single electron E_T are obtained with high accuracy with the *smearing method*.

In Fig. 4.20 the number of events for the defined di-electron categories is shown.

In Fig. 4.21 the energy scale corrections with the improved approach are derived on top of the c_2 corrections as they would be with the traditional approach. The measured

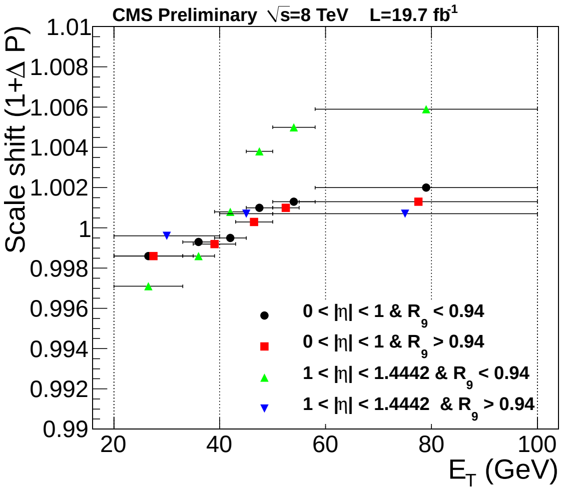


Figure 4.21: Residual energy scale corrections as function of E_T on top of the c_2 energy scale corrections.

energy scale shifts are then the residual shifts as function of E_T with respect to the $|\eta| \times R_9$ categories.

4.4.2 Additional smearings

The additional smearings are derived in the same categories used for c_2 energy scale corrections:

- $|\eta| < 1.0$ & $R_9 < 0.94$
- $|\eta| < 1.0$ & $R_9 > 0.94$
- $1.0 < |\eta| < 1.4442$ & $R_9 < 0.94$
- $1.0 < |\eta| < 1.4442$ & $R_9 > 0.94$
- $1.566 < |\eta| < 2.0$ & $R_9 < 0.94$
- $1.566 < |\eta| < 2.0$ & $R_9 > 0.94$
- $2.0 < |\eta| < 2.5$ & $R_9 < 0.94$
- $2.0 < |\eta| < 2.5$ & $R_9 > 0.94$

Traditional approach smearings

In the traditional approach the additional smearing ($\Delta\sigma$) is parametrized as a simple additional constant term:

$$\Delta\sigma = \Delta C \tag{4.8}$$

The results are obtained from the 1D scan of NLL for each category, fixing scales and smearings for all the other categories to the minimum found by the minimization. According to deterministic smearing (described in Sec. 4.3.1) the random numbers used to smear the electron energy are fixed for each electron. This can generate small biases. For this reason, after the minimization, the NLL scan for each parameter is performed many times varying the seed for the random number's generator. For each parameter, n NLL scans are performed. Each point k of the final profile is given by the mean value of that point over n profiles and the uncertainty on the point is given by the uncertainty of the mean.

The best fit value and the statistical uncertainty is obtained fitting the points with an asymmetric parabola:

$$f(x; y_0, x_{min}, \sigma_L, \sigma_R) = \begin{cases} \sigma_L \cdot (x - x_{min})^2 & x < x_{min} \\ \sigma_R \cdot (x - x_{min})^2 & x > x_{min} \end{cases} \tag{4.9}$$

where x_{min} is the best fit value of the parameter x, the statistical uncertainty is estimated from σ_L and σ_R as

$$err_{L,R} = \frac{1}{\sqrt{2 \cdot \sigma_{L,R}}}$$

If the values of $\sigma_{L,R}$ are compatible, the average uncertainty is assigned for the single electron category smearing. In the traditional approach this condition is always satisfied.

The additional smearings for the 7 and 8 TeV data are shown in Tab. 4.2

Improved approach: stochastic term in energy smearing

In the improved approach, the additional smearing is parametrized in two terms: one additional constant term and one additional stochastic term. The idea is to exploit further the E_T categorization introduced for the c_3 scale correction derivation in order

Electron category	additional smearing: $\Delta\sigma$ (%)	
	7 TeV	8 TeV
EB $\|\eta\| < 1.0$ $R9 < 0.94$	0.96 ± 0.03	0.83 ± 0.02
EB $\|\eta\| < 1.0$ $R9 > 0.94$	0.68 ± 0.04	0.72 ± 0.03
EB $\|\eta\| > 1.0$ $R9 < 0.94$	1.85 ± 0.04	1.87 ± 0.02
EB $\|\eta\| > 1.0$ $R9 > 0.94$	1.01 ± 0.14	1.07 ± 0.09
EE $\|\eta\| < 2.0$ $R9 < 0.94$	1.85 ± 0.07	1.98 ± 0.04
EE $\|\eta\| < 2.0$ $R9 > 0.94$	1.58 ± 0.18	1.56 ± 0.08
EE $\|\eta\| > 2.0$ $R9 < 0.94$	1.83 ± 0.09	1.94 ± 0.05
EE $\|\eta\| > 2.0$ $R9 > 0.94$	2.01 ± 0.06	1.87 ± 0.04

Table 4.2: Energy smearings with traditional approach for 7 and 8 TeV data.

to measure the additional smearing as function of the energy. The choice of a particular functional form for the additional smearing is arbitrar. The sources of *in situ* resolution discussed in Sec. 4.2.1 suggest that the spread of the single crystal α parameter in the response corrections affect mostly the constant term of the resolution, whilst the imperfect knowledge of the material in front of ECAL affect mostly the stochastic term of the energy resolution due to the stochastic nature of the interaction of electrons and photons with the material.

The parametrization chosen is then:

$$\Delta\sigma = \frac{\Delta S}{\sqrt{E_T}} \oplus \Delta C \tag{4.10}$$

There is not a strong motivation for scaling the stochastic term by $\sqrt{E_T}$ instead of $\sqrt{E}$ and there is not yet sensitivity to discriminate between the two parametrization. The $\sqrt{E_T}$ result more natural because the electron categories are divided into E_T bins, so a slightly larger sensitivity is expected.

The strong correlation between the two terms in the additional smearing is taken into account in order to simplify the minimization. With a coordinate transformation the correlation in the $(\Delta S, \Delta C)$ plane is decoupled into polar coordinates (ρ, ϕ) where ρ is the magnitude of the additional smearing $\Delta\sigma$, ϕ discriminates between the contribution of the two terms.

The transformation is defined as:

$$\begin{aligned} \Delta C &= \rho \cdot \sin\phi \\ \Delta S &= \rho \cdot \sqrt{< E_T >} \cdot \cos\phi \\ \sqrt{< E_T >} &= \frac{\Delta C|_{\Delta S=0}}{\Delta S|_{\Delta C=0}} \end{aligned} \tag{4.11}$$

where the scaling factor $\sqrt{< E_T >}$ is the root square of the average E_T for the category. The scaling factor is measured imposing that the additional smearing is the same if we measure it as a pure constant term or a pure stochastic term:

$$\begin{aligned} \Delta\sigma|_{\Delta S=0} &= \Delta C \\ \Delta\sigma|_{\Delta C=0} &= \frac{\Delta S}{\sqrt{< E_T >}} \\ \Delta\sigma &= \Delta\sigma|_{\Delta S=0} = \Delta\sigma|_{\Delta C=0} \end{aligned} \tag{4.12}$$

With this factorization, the minimization is still possible with 1D likelihood scans in the three parameters for each category $\Delta P, \rho, \phi$.

In Fig. 4.22 the likelihood profiles for the ρ and ϕ parameters are respectively shown in the left and the right plots. In all the four categories, the additional smearing is dominated by the stochastic term. In Tab. 4.3a the best fit values and uncertainties are shown in the ρ and ϕ parametrization.

The uncertainties on the ρ and ϕ are given by the likelihood profiles as illustrated for the traditional smearing derivation. From the NLL profiles, with a flat prior hypothesis, the posterior for the ρ and ϕ are built and then transformed into ΔC and ΔS. The uncertainties on ΔC and ΔS are the standard deviations of the *pdf*s. In Tab. 4.3b the values of ΔC and ΔS are summarized. The best fit values are obtained by analytical transformation of the ρ and ϕ best fit values. The mean and the standard deviation of the ΔC and ΔS *pdf*s are also shown.

(a) $|\eta| < 1.0 \,\& R_9 < 0.94$ (b) $|\eta| < 1.0 \,\& R_9 < 0.94$

(c) $|\eta| < 1.0 \,\& R_9 > 0.94$ (d) $|\eta| < 1.0 \,\& R_9 > 0.94$

(e) $|\eta| > 1.0 \,\& R_9 < 0.94$ (f) $|\eta| > 1.0 \,\& R_9 < 0.94$

(g) $|\eta| > 1.0 \,\& R_9 > 0.94$ (h) $|\eta| > 1.0 \,\& R_9 > 0.94$

Figure 4.22: The NLL profiles are shown on the left for the ρ parameter and on the right for ϕ in the four electron categories in EB. The best values and the uncertainties are estimated fitting the NLL profiles with an asymmetric parabola.

Category	$\rho[\%]$	$\phi[rad]$	$\sqrt{< E_T >}$
EB $\|\eta\| < 1.0\,\&R_9 < 0.94$	0.79 ± 0.01	0.00 ± 0.12	7.18
EB $\|\eta\| < 1.0\,\&R_9 > 0.94$	0.78 ± 0.02	0.16 ± 0.16	7.05
EB $\|\eta\| > 1.0\,\&R_9 < 0.94$	1.31 ± 0.02	0.01 ± 0.08	8.71
EB $\|\eta\| > 1.0\,\&R_9 > 0.94$	1.18 ± 0.07	$0.72^{+0.39}_{-0.72}$	8.08

(a) Best fit values for ρ and ϕ parameters of single electron additional smearings for 8 TeV data. Both ρ and ϕ are defined positve.

Category	$\Delta C[\%]$			$\Delta S[\%]$		
	best fit	mean	std. dev.	best fit	mean	std. dev.
EB $\|\eta\| < 1.0\,\&R_9 < 0.94$	0.00	0.07	0.06	5.67	5.65	0.19
EB $\|\eta\| < 1.0\,\&R_9 > 0.94$	0.12	0.17	0.10	5.43	5.33	0.24
EB $\|\eta\| > 1.0\,\&R_9 < 0.94$	0.01	0.07	0.06	11.41	11.38	0.18
EB $\|\eta\| > 1.0\,\&R_9 > 0.94$	0.77	0.63	0.34	7.16	7.19	2.20

(b) Additional smearings for 8 TeV data in terms of additional constant term (ΔC) and an additional stochastic term ΔS.

Table 4.3

4.4.3 Validation with toy MC study

The *smearing method* has been validated with a toy MC study, injecting artificial scales and smearings to a fraction of the simulated sample used as data.

The toys are generated starting from the sum of the three MC samples (MadGraph, POWHEG and SHERPA). In order to profit from the full statistical power, no pileup reweighting is applied. Part of the simulated events is used as pseudo-data, the remaining part as MC.

The pseudo-data are shifted and smeared without any over-sampling and with a fixed seed for the random generator used. On the contrary, for the MC part, the over-sampling is used and the random generator seed is randomly choosed for each toy.

The sensitivity of the *smearing method* to the constant term and the stochastic term in the additional smearing are demostrated by the toy study. The magnitude of the injected in data is $\sim 1\%$. Two toy studies have been performed, one with a purely constant term, the second with a purely stochastic term. The expected values for ϕ are:

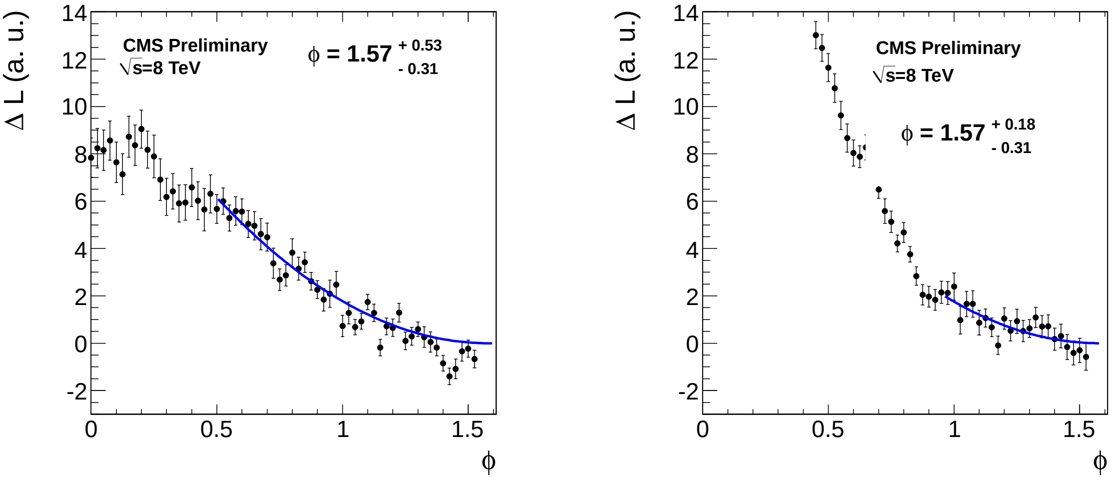
- $\phi = \frac{\pi}{2}$ for a purely constant term
- $\phi = 0$ for a a purely stochastic term

The results for the ϕ parameter are shown in Fig. 4.23 and Fig. 4.24 and confirm the validity of the method.

4.4.4 Systematic uncertainties on additional smearings

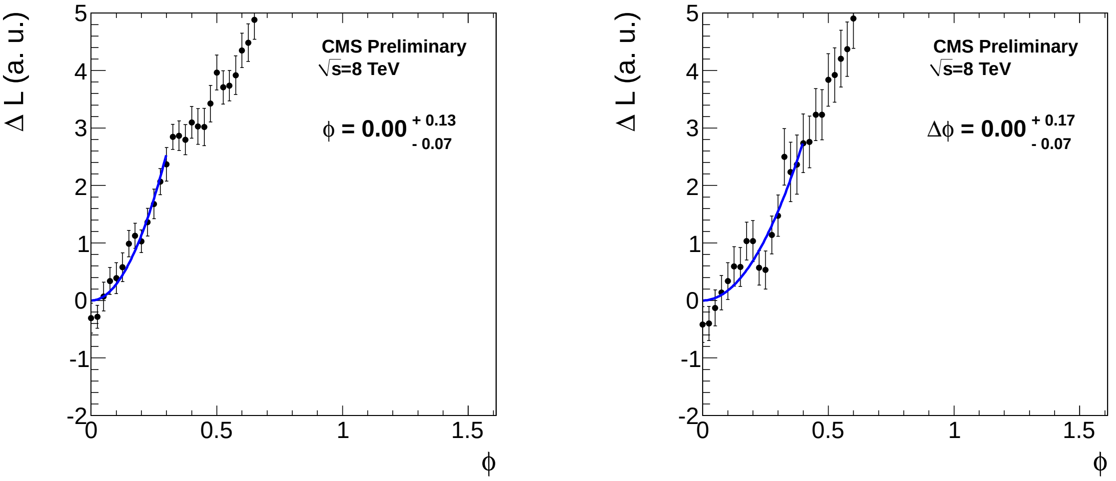
The systematic uncertainties associated to the smearing method have been evaluated with toy MC studies. The scale and the additional smearings have been derived in all the electron categories after the injection of artificial scale shifts and additional smearings of respectively ±1%, ±2%, ±5%; 0.5%, 1%, 2% and 3% as constant term.

The bias for the parameter x is defined as the mean of the distribution of $x - x_{true}$ where x is the measured value and x_{true} is the injected one. The spread of the same distribution (standard deviation) is due to the statistical uncertainty on the measurement of the variable x.



(a) $|\eta| < 1.0\,\&R_9 < 0.94$ (b) $|\eta| < 1.0\,\&R_9 > 0.94$

Figure 4.23: NLL profiles of ϕ parameter of a toy MC study. The injected 1% additional smearing is a pure constant term ($\phi = \frac{\pi}{2}$).



(a) $|\eta| < 1.0\,\&R_9 < 0.94$ (b) $|\eta| < 1.0\,\&R_9 > 0.94$

Figure 4.24: NLL profiles of ϕ parameter of a toy MC study. The injected 1% additional smearing is a pure stochastic term ($\phi = 0$).

The uncertainty on scale and smearing has been defined in conservatively way as the maximum between the bias and the standard deviation. The bias and the standard deviation of the parameter x has been evaluated injecting separately the energy scale or the additional constant smearing and the larger uncertainty is considered.

In Tab. 4.4 the systematic uncertainties are shown for the different categories against the injection of energy scales and additional smearings.

The systematic uncertainty on the energy scale does not exceed 0.01%. The uncertainty on the additional smearing is 0.01% for all the categories.

Scale		
Category	Injected scales [%]	Injected smearings [%]
$\lvert\eta\rvert < 1.0 \,\& R_9 < 0.94$	0.001	0.000
$\lvert\eta\rvert < 1.0 \,\& R_9 > 0.94$	0.010	0.010
$1.0 < \lvert\eta\rvert < 1.4442 \,\& R_9 < 0.94$	0.003	0.003
$1.0 < \lvert\eta\rvert < 1.4442 \,\& R_9 > 0.94$	0.009	0.009
$1.566 < \lvert\eta\rvert < 2.0 \,\& R_9 < 0.94$	0.001	0.000
$1.566 < \lvert\eta\rvert < 2.0 \,\& R_9 > 0.94$	0.010	0.010
$2.0 < \lvert\eta\rvert < 2.5 \,\& R_9 < 0.94$	0.003	0.003
$2.0 < \lvert\eta\rvert < 2.5 \,\& R_9 > 0.94$	0.009	0.009
Additional smearing		
Category	Injected scales [%]	Injected smearings [%]
$\lvert\eta\rvert < 1.0 \,\& R_9 < 0.94$	0.001	0.000
$\lvert\eta\rvert < 1.0 \,\& R_9 > 0.94$	0.010	0.010
$1.0 < \lvert\eta\rvert < 1.4442 \,\& R_9 < 0.94$	0.003	0.003
$1.0 < \lvert\eta\rvert < 1.4442 \,\& R_9 > 0.94$	0.009	0.009
$1.566 < \lvert\eta\rvert < 2.0 \,\& R_9 < 0.94$	0.001	0.000
$1.566 < \lvert\eta\rvert < 2.0 \,\& R_9 > 0.94$	0.010	0.010
$2.0 < \lvert\eta\rvert < 2.5 \,\& R_9 < 0.94$	0.003	0.003
$2.0 < \lvert\eta\rvert < 2.5 \,\& R_9 > 0.94$	0.009	0.009

Table 4.4: Systematic uncertainties on energy scale and additional smearing with traditional categorization.

Chapter 5

Search for a Higgs boson in the $H \to \gamma\gamma$ channel

5.1 Introduction

The analysis has been conducted using two approaches: an MVA-based analysis (main analysis) and a cut-based analysis (cross-check analysis). In the cut-based analysis (discussed in Sec. 5.5.1) the photon selection and the event categorization are both performed with cuts on reconstructed quantities aimed to discriminate the $H \to \gamma\gamma$ photons from background events. In the MVA-based the photon identification and the event categorization are performed using MultiVariate (MVA) techniques with a BDT that exploits also the correlation between the variables used in the selection.

The analysis presented in this chapter is performed on the data collected by the CMS experiment in 2011 at 7 TeV center-of-mass energy (5.3 fb^{-1}) and in 2012 at 8 TeV center-of-mass energy (19.6 fb^{-1}).

At the moment of writing, the analysis for the final results on the full dataset (7 and 8 TeV) is ongoing. Therefore the analysis results will not be public until the end of the year. For this reason, the analysis and the results described in this chapter and in the following are the most recent agreed to be public by the CMS collaboration. The energy scale corrections and the additional smearings described in Chapter 4 are included in the ongoing re-analysis of the data.

In the $H \to \gamma\gamma$ search, an excellent di-photon invariant mass resolution is a crucial element of the analysis. The energy resolution has been extensively discussed in Sec. 4 and the energy scale corrections in data and the additional smearings to be added to the simulation to match the resolution in data were subject of the previous Chapter 4. The energy resolution is the dominant term when the di-photon vertex is identified within 1 cm from the true vertex. The vertex identification will be discussed in Sec. 5.4.

The general analysis strategy relies on an efficient event selection logically divided into single photon identification and di-photon identification criteria. The selected di-photon events are then classified according to the Higgs production mode (exclusive mode categories) with tight requirements on the additional particles in the final state. The events not assigned to any exclusive category are subdivided according to the amount of background and di-photon expected resolution into 4 “inclusive” categories. Search sensitivity is then increased by the classification and the combination of the results in each class.

The di-photon events are firstly compared to the exclusive categories and then if not matched their criteria, it is assigned to the inclusive categories. The classification order is: lepton-tagged (muon-tag and electron-tag), MET-tagged (two categories), dijet-tagged (two categories), ttH-tagged events and inclusive categories.

A preselection common to all the categories and both analysis approaches is applied to the di-photon events.

The goal of the analysis is the study of the new boson properties in order to assess the compatibility or eventual deviations from the expected SM Higgs boson. The $H \to \gamma\gamma$ channel is still largely dominated by the statistical uncertainty, for that reason the analyis is optimized for the best statistical significance.

5.2 Trigger

The events retained for the $H \to \gamma\gamma$ search are those selected by di-photon triggers at HLT with a threshold $E_T > 26$ GeV on the most energetic electromagnetic candidate and $E_T > 18$ GeV on the second one. The di-photon triggers are seeded at L1 by two electromagnetic energy deposits with $E_T > 13$ GeV on the first candidate and $E_T > 7$ GeV on the second one. We will refer to these triggers as Type 1 triggers.

In order to measure the trigger efficiencies at L1 and HLT, a second type of triggers is used. Type 2 di-photon triggers have a tighter E_T cut both at L1 and HLT level but seeded by just one e/γ candidate. For these single-L1-seeded di-photon paths, another candidate is reconstructed by unpacking the rest of the ECAL information. Then, at least two of the reconstructed clusters are required to pass the relevant HLT criteria that are grouped into general, isolation plus calorimeter identification and R_9 cuts.

To measure the trigger efficiencies, one needs to evaluate separately the efficiency of the L1-seeding and the efficiency of the HLT filters provided that the L1 requirement has been satisfied. This has been performed using the tag-and-probemethod on $Z \rightarrow e^+e^-$ data. The data are required to fulfill the preselection criteria described later in Sec. 5.4.3.

Compatible efficiencies have been measured both for the cut-based and the MVA-based analysis with an efficiency $> 99\%$ for Type 1 triggers at L1. The L1 efficiencies for Type 1 and Type 2 triggers are listed in Table 5.1.

	MVA analysis	cut-based analysis
L1 Type 1	99.75±0.01%	99.78±0.01%
L1 Type 2	97.14±0.02%	97.65 ±0.02%

Table 5.1: L1 efficiencies for events passing the MVA analysis preselection and the cut-based analysis selections.

For events passing the preselection, the HLT efficiency of the OR of the triggers mentioned above has been found to be $99.68 \pm 0.03\%$. Similar efficiency is found for the cut-based selection.

5.3 Simulated samples

The $H \rightarrow \gamma\gamma$ analysis is the first analysis using the most updated 8 TeV simulation with the evolving ECAL conditions, extended OOT pileup simulation that has been discussed in Sec. 3.7.1. The mentioned ECAL simulation plays an important role in the 2012 data analysis, whilst the standard simulation describes with a sufficient accuracy the 7 TeV data conditions.

The signal samples for the gluon fusion and vector boson fusion production processes were generated with POWHEG at NLO [62, 63], whereas the associated production process was simulated with PYTHIA [30] at LO.

The cross-sections and branching ratios recommended by the LHC Cross-Section Working-Group are used [64]. In Tab. 5.2 and Tab. 5.3 the cross section and $\gamma\gamma$ Branching Fraction for different Higgs masses are shown.

Table 5.2: SM Higgs cross-sections at 8 TeV (pb) for different production mechanisms and H $\rightarrow \gamma\gamma$ branching ratios for different Higgs masses.

m_{H} (GeV)	Gluon Fusion	Vector Boson Fusion	$W \rightarrow WH$, $Z \rightarrow ZH$	$t\bar{t} \rightarrow H$	Branching Fraction
90	36.2	2.19	1.99, 1.09	0.32	$1.22 \cdot 10^{-3}$
95	32.7	2.08	1.70, 0.938	0.28	$1.39 \cdot 10^{-3}$
100	29.7	1.98	1.45, 0.810	0.24	$1.58 \cdot 10^{-3}$
105	27.0	1.89	1.24, 0.702	0.21	$1.77 \cdot 10^{-3}$
110	24.7	1.80	1.07, 0.613	0.19	$1.95 \cdot 10^{-3}$
115	22.7	1.73	0.927, 0.536	0.17	$2.11 \cdot 10^{-3}$
120	20.9	1.65	0.805, 0.471	0.15	$2.23 \cdot 10^{-3}$
125	19.3	1.58	0.705, 0.415	0.13	$2.28 \cdot 10^{-3}$
130	17.9	1.51	0.617, 0.367	0.12	$2.24 \cdot 10^{-3}$
135	16.6	1.44	0.542, 0.326	0.10	$2.12 \cdot 10^{-3}$
140	15.4	1.39	0.477, 0.290	0.09	$1.93 \cdot 10^{-3}$
145	14.5	1.33	0.422, 0.258	0.08	$1.67 \cdot 10^{-3}$
150	13.6	1.28	0.373, 0.231	0.07	$1.36 \cdot 10^{-3}$

Signal samples were produced for Higgs mass values ranging from 90 to 150 GeV.

The background MC samples used in this analysis were generated with PYTHIA, with the NLO matrix-element generator POWHEG interfaced with PYTHIA and the MADGRAPH

Table 5.3: SM Higgs cross-sections at 7 TeV (pb) for different production mechanisms and H $\rightarrow \gamma\gamma$ branching ratios for different Higgs masses.

m_{H} (GeV)	Gluon Fusion	Vector Boson Fusion	$W \rightarrow WH$, $Z \rightarrow ZH$	$t\bar{t} \rightarrow H$	Branching Fraction
90	29.03	1.723	1.654, 0.8959	0.2162	$1.22 \cdot 10^{-3}$
95	26.10	1.639	1.404, 0.7678	0.1880	$1.39 \cdot 10^{-3}$
100	23.64	1.557	1.195, 0.6616	0.1637	$1.58 \cdot 10^{-3}$
105	21.45	1.478	1.029, 0.5724	0.1432	$1.77 \cdot 10^{-3}$
110	19.56	1.410	0.8847, 0.4978	0.1257	$1.95 \cdot 10^{-3}$
115	17.89	1.344	0.7626, 0.4345	0.1105	$2.11 \cdot 10^{-3}$
120	16.43	1.279	0.6617, 0.3808	0.09758	$2.23 \cdot 10^{-3}$
125	15.13	1.222	0.5785, 0.3351	0.08632	$2.28 \cdot 10^{-3}$
130	13.98	1.168	0.5059, 0.2957	0.07660	$2.24 \cdot 10^{-3}$
135	12.95	1.117	0.4431, 0.2616	0.06816	$2.12 \cdot 10^{-3}$
140	12.02	1.069	0.3839, 0.2322	0.06079	$1.93 \cdot 10^{-3}$
145	11.24	1.023	0.3437, 0.2068	0.05429	$1.67 \cdot 10^{-3}$
150	10.51	0.980	0.3034, 0.1842	0.04867	$1.36 \cdot 10^{-3}$

matrix-element generator interfaced with PYTHIA.

All MC samples are re-weighted to match the pile-up distribution in the data as already described in Sec. 4.2.7.

The SM background samples are di-photon + jets (MADGRAPH), di-photon by box processes (PYTHIA), $\gamma +$ jet (PYTHIA), QCD di-jet (PYTHIA) and DY (MADGRAPH) events.

A “double EM-enriched” filter at generator level was applied during the production of the QCD di-jet and γ+jet samples in order to improve the production efficiency. This filter requires a potential photon signal – electromagnetic activity, coming from photons, electrons or neutral hadrons, with $E_T > 15$ GeV within a small region. In addition it is required that this potential photon signal have no more than one charged particle in a cone $\Delta R < 0.2$, thus mimicking a tracker isolation.

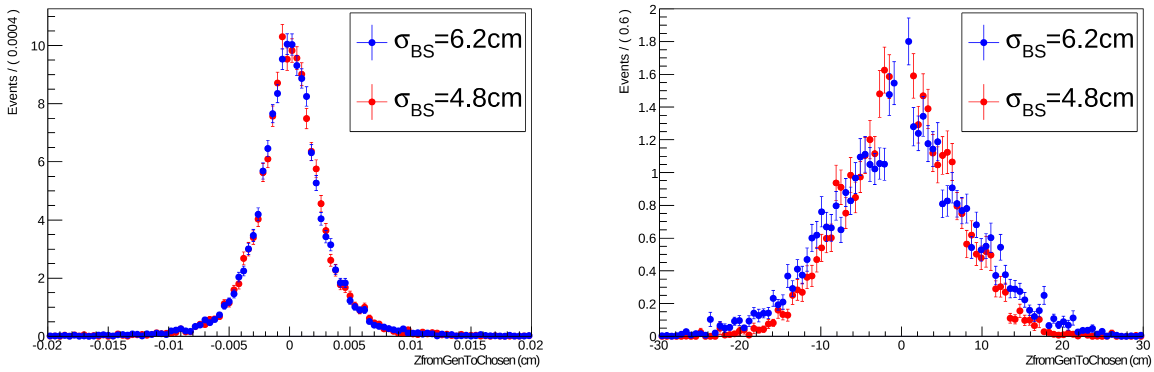
Beamspot correction in MC

The 8 TeV production of MC used a simulated beamspot width (σ_{BS}) of ~ 6.2 cm. The actual σ_{BS} in data is closer to 5 cm. Gluon fusion $H \to \gamma\gamma$ samples were produced with a σ_{BS} of ~ 4.8 cm in order to study the effects of this large discrepancy between data and MC.

The distribution of the distance in z between the selected reconstructed vertex and the generated vertex ($\Delta Z_{ChosenToGen}$) is compared for MC samples with $\sigma_{BS} = 6.2$ cm and $\sigma_{BS} = 4.8$ cm.

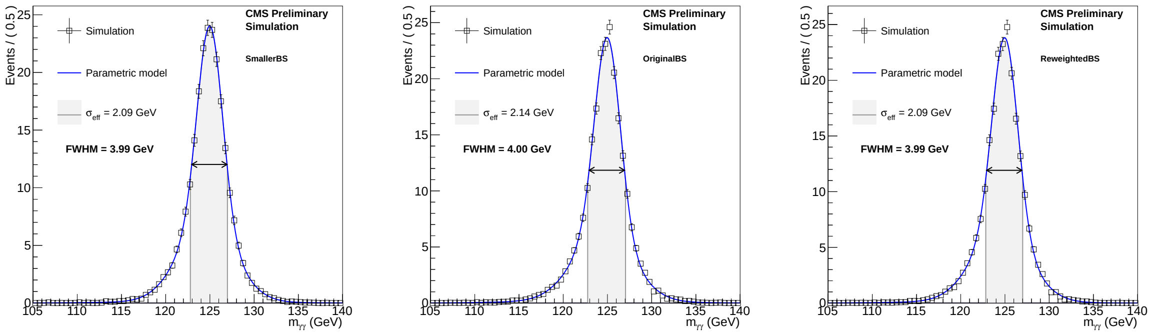
A reweighting is performed if $\Delta Z_{ChosenToGen} > 0.1$ cm because of the large discrepancies between data and simulation, whilst no correction is applied otherwise. The reweighting restores the mass resolution to nearly the level of the more realistic beamspot sample, as shown in Fig. 5.2.

The reweighting procedure is validated in data using $Z \to \mu\mu$ events (Fig. 5.3), where the vertices are refitted after removing the muon tracks from the track collection to mimic the $H \to \gamma\gamma$ topology.



(a) $\Delta Z_{ChosenToGen} < 0.1$ cm (b) $\Delta Z_{ChosenToGen} > 0.1$ cm

Figure 5.1: Distribution of $\Delta Z_{ChosenToGen}$ for MC samples with $\sigma_{BS} = 6.2$ cm and $\sigma_{BS} = 4.8$ cm.



(a) $\sigma_{BS} = 4.8$ cm (b) $\sigma_{BS} = 6.2$ cm (c) Reweighted Beamspot

Figure 5.2: Effect of beamspot reweighting on mass resolution

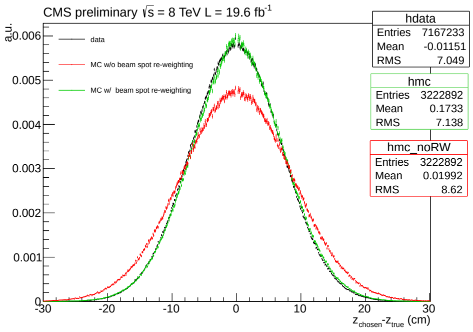


Figure 5.3: Distribution of ΔZ between the chosen vertex and the true vertex for data, MC and MC after beam spot reweighting in $Z \rightarrow \mu\mu$ events.

5.4 Diphoton vertex identification

The mass resolution of a narrow resonance decaying into two photons is driven by two factors: the photon energy resolution and the resolution in measuring the opening angle between the two photons. The latter is determined by two factors: SC position and vertex location. The SC position is well measured due to the fine granularity of the ECAL. So, it is important to select the correct primary vertex from all the pileup vertices. In the 8 TeV data there are in average $\sim$ 20 pileup vertices spread over 5 cm along the beam direction (z) due to the high number of interactions per bunch crossing.

The resolution on the photon opening angle makes a negligible contribution to the mass resolution, as compared to the ECAL energy resolution, when the interaction point is known to within about 1 cm. The mass resolution can be preserved by correctly assigning the reconstructed photons to one of the interaction vertices reconstructed from the charged tracks using the standard CMS algorithm [58].

The method used in this analysis to select the best primary vertex candidate is based on a multivariate approach exploiting the kinematic properties of the vertex tracks and their correlation with the diphoton kinematics, and adding the tracker information for converted photons. Details can be found in [65].

5.4.1 Base algorithms

The variables used as input of the MVA are the sum of the squared transverse momentum of all the tracks ($sumpt2 = \sum_i |\vec{p}_T^{\,i}|^2$), the balance of the transverse momentum ($ptbal = -\sum_i (\vec{p}_T^{\,i} \cdot \frac{\vec{p}_T^{\,\gamma\gamma}}{|\vec{p}_T^{\,\gamma\gamma}|})$) and the p_T asymmetry ($ptasym = \frac{|\sum_i \vec{p}_T^{\,i}| - p_T^{\gamma\gamma}}{|\sum_i \vec{p}_T^{\,i}| + p_T^{\gamma\gamma}}$).

In events with at least one photon conversion, the longitudinal coordinate of the estimated primary vertex position is derived exploiting the additional information by the photon conversion tracks. In the BDT a fourth variable is then used as input: $pull_{conv} = |z_{vertex} - z_{conv}|/\sigma_{conv}$ where z_{conv} is the estimated primary vertex position and σ_{conv} is the resolution measured in data. The BDT algorithm was trained on simulated $H \rightarrow \gamma\gamma$ events.

The resolution on the primary vertex z as determined from conversions from data and

σ (cm)	Pix Barrel	TIB	TOB	Pix Fwd	TID	TEC
Data 2012 Run	0.011	0.289	1.389	0.060	0.311	1.031
Simulated $\gamma + \text{jet}$	0.011	0.305	1.452	0.045	0.344	0.924
Ratio data/simulation	1.	0.948	0.957	1.333	0.904	1.116

Table 5.4: Resolution on the primary vertex z as determined from conversions.

simulated $\gamma + \text{jet}$ sample are listed in Table 5.4. The results obtained in data are used in the analysis.

5.4.2 Per-event probability of correct diphoton vertex choice

Since the fraction of events where the vertex is not correctly identified is significant (the vertex efficiency is 75% in the current 2012 pile-up conditions), the estimation of an event-by-event probability for the vertex assignment is important in order to be able to profit from the calorimeter energy resolution.

A second, vertex-related multivariate discriminant was designed for this purpose. More specifically, a BDT is trained to tag events where the chosen vertex lies within 10 mm of the generated one using simulated $H \rightarrow \gamma\gamma$ events. The inputs of the BDT are the following:

- p_T of the diphoton system.
- number of reconstructed primary vertices in each event.
- value of the per-vertex BDT discriminant for the best three vertices in each event.
- Δz between the best vertex and the second and third choices.
- number of photons with associated conversion tracks (0, 1, or 2).

The dependence of the vertex identification probability on the per-event BDT discriminant is shown in Fig. 5.4. In order to estimate the vertex identification on an event-by-event basis, a linear function of the per event BDT output is used. The comparison between the true vertex identification efficiency and the average estimated vertex

	DATA			MC		R	
	Eff.	Stat. Err.	Syst. Err.	Eff.	Stat. Err.	Eff.	Err.
Barrel; R_9 >0.90	0.9864	0.0002	0.0030	0.9897	0.0001	0.997	0.003
Barrel; R_9 <0.90	0.9406	0.0007	0.0055	0.9614	0.0003	0.978	0.006
Endcap; R_9 >0.90	0.9880	0.0003	0.0090	0.9824	0.0002	1.006	0.009
Endcap; R_9 <0.90	0.9368	0.0012	0.0170	0.9460	0.0004	0.990	0.018

Table 5.5: Photon preselection efficiencies measured in the 4 photon categories using tag-and-probewith $Z \to e^+e^-$ events (for all cuts except electron rejection).

probability is shown in Fig. 5.5 as a function of the reconstructed di-photon p_T and of the number of reconstructed vertices. The estimation is accurate within 3-5%.

5.4.3 Preselection

All prompt and non-prompt photons within ECAL acceptance are required to pass a preselection. This selection allows to keep the common phase space between data passing the trigger and the MC where no trigger requirement is applied.

The preselection cut thresholds are set tighter for converted photons ($R_9 < 0.9$) and looser for unconverted photons and different for photons in the barrel and in the endcaps. The variables used in the preselection are the ratio between the energy in the ECAL seed cluster and the HCAL tower behind it, the covariance of the energy spread over the η direction and relative isolation variables calculated for energy deposits in HCAL around the photon and tracks in the cone around the photon direction.

Table 5.5 shows preselection efficiencies measured using tag-and-probewith $Z \to e^+e^-$ events for data, simulation and the ratio between data and simulation, for the four photon categories. In the computation of the scale factors the inefficiency of the tag-and-probetrigger path used to select Z events is taken into account as well as the photon/electron R_9 difference.

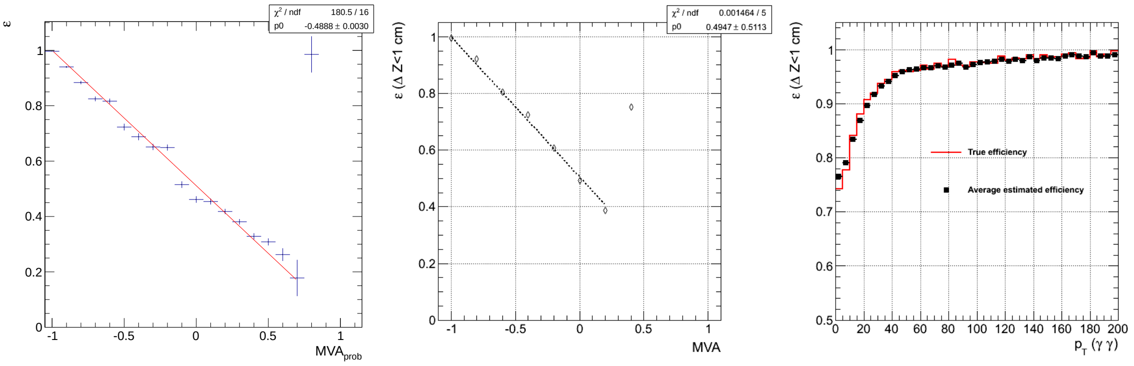


Figure 5.4: Dependence of the vertex identification probability on the per-event BDT discriminant.

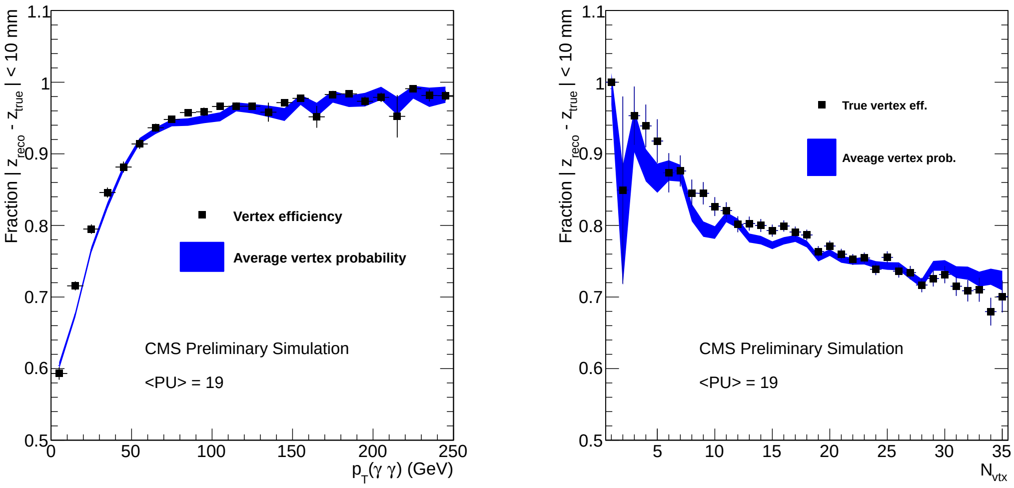


Figure 5.5: Comparison the true vertex identification efficiency and the average estimated vertex probability as a function of the reconstructed diphoton p_T and the number of reconstructed vertices. The plots are obtained on signal events selected by the MVA analysis.

5.5 Cut-based selection and categorization: untagged categories

5.5.1 Single photon identification

In the cut-based analysis, photon identification cut values are optimized separately in four categories defined in terms of pseudorapidity and R_9. These categories have significantly different levels of background and mass resolution and their use provides increased sensitivity.

The cuts are set to get the highest efficiency of selected photons for a fixed value of the signal over background ratio (S/B). Meaning that the cuts are tighter in the low R_9 category than in the high R_9 category and in the endcap than in the barrel.

The chosen photon identification working point is applied to both legs of the diphoton pair and the cuts are listed in Table 5.6.

	barrel		endcap	
	$R_9 > 0.94$	$R_9 < 0.94$	$R_9 > 0.94$	$R_9 < 0.94$
PF isolation sum, chosen vertex	< 6	< 4.7	< 5.6	< 3.6
PF isolation sum worst vertex	< 10	< 6.5	< 5.6	< 4.4
Charged PF isolation sum	< 3.8	< 2.5	< 3.1	< 2.2
$\sigma_{i\eta i\eta}$	< 0.0108	< 0.0102	< 0.028	< 0.028
H/E	< 0.124	< 0.092	< 0.142	< 0.063
R_9	> 0.94	> 0.298	> 0.94	> 0.24

Table 5.6: Photon ID selection cut values. The cuts are applied to both the leading and subleading photons.

5.5.2 Di-photon event selection

Events are required to contain two reconstructed photons passing the cut based photon ID selection described in Section 5.5.1 and satisfying the kinematic requirements:

$p_T/m_{\gamma\gamma}$ >1/3 (lead) and $p_T/m_{\gamma\gamma}$ >1/4 (sublead), where $m_{\gamma\gamma}$ is the diphoton invariant mass. This "sliding cut" corresponds to a cut of p_T >40 GeV (leading) and p_T >30 GeV (subleading) for $m_{\gamma\gamma}$=120 GeV. If more than two photons pass the selection, the photon pair with the highest scalar sum of the photon transverse momenta is considered

5.5.3 Event classification

Two classifiers are used to differentiate diphotons with good mass resolution from those with less good resolution and in separating events with a higher S/B probability from those with a lower S/B probability: the minimum R_9 of the two photons, and the maximum pseudorapidity of the two photons. The event class definitions are shown in Table 5.7.

Table 5.7: Inclusive event classes for the cut-based analysis.

class 0	Both photons in barrel	min(R_9)>0.94
class 1	Both photons in barrel	min(R_9)<0.94
class 2	At least one photon in endcap	min(R_9)>0.94
class 3	At least one photon in endcap	min(R_9)<0.94

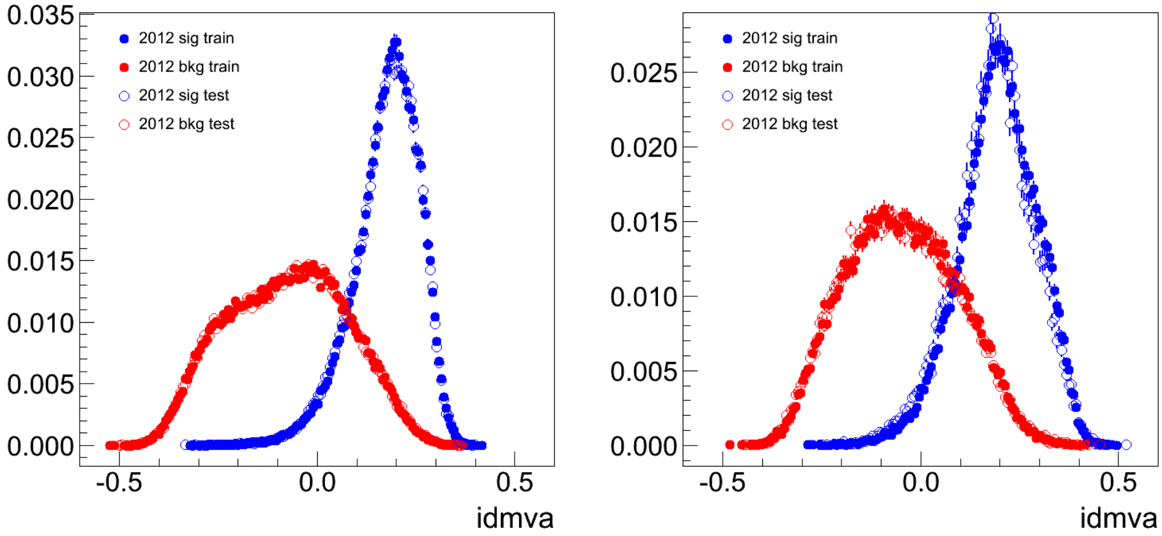


Figure 5.6: BDT Output for Barrel (left) and Endcap (right).

5.6 MVA-based selection and categorization: untagged categories

5.6.1 Single photon identification

In the MVA analysis, the di-photon event selection is performed using MVA techniques. First BDT is trained to distinguish prompt photons from non-prompt photons, primarily from high momentum neutral mesons decaying to two photons, where both photons are reconstructed as a single SC and mimic a single photon. The output BDT value of a single photon is used as a photon identification input variable for the diphoton event level classification.

The training inputs of the BDT are a set of shower topology variables, R_9, isolation variables, ρ (the median energy density per solid angle) and the pseudorapidity of the reconstructed SC.

The BDT output distributions for Barrel and Endcap are shown in Fig. 5.6, where the signal prompt photon is in blue and the background non-prompt photon is in red. The training events are shown as solid circles and the testing events as hollow circles. No reweighting is applied. The distributions of training and testing events are in agreement, as shown in Fig. 5.6.

5.6.2 Di-photon event selection

In addition to the preselection described in Sec. 5.4.3, a loose preselection cut on the photon ID MVA output of $MVA > -0.2$ is also applied. This cut retains more than 99% of the signal events passing the other preselection requirements, while removing 23.5% of the data events in the mass range 100< $m_{\gamma\gamma}$ <180 GeV. The photons are additionally required to pass diphoton mass-dependent cuts on the transverse momenta of the photons as $p_{\mathrm{T}} > m_{\gamma\gamma} \times 1/3(1/4)$ for the leading (trailing) photon.

5.6.3 Event classification

In order to increase the sensitivity of the analysis, the pre-selected events are divided into classes, such that the 'high-performance' classes are predominately populated by events with good resolution and/or high S/B.

An event classifier variable fulfills the following criteria:

1) The variable should classify with a high score events with:
    1) signal-like kinematic characteristics,
    2) good diphoton mass resolution,
    3) photon-like values from the photon identification BDT,
2) The variable should be mass independent.

The event classifier is constructed using a BDT. The classifier incorporates the kinematic properties of the diphoton system (excluding $m_{\gamma\gamma}$), a per-event estimate of the diphoton mass resolution, and a per-photon identification measure (the photon identification BDT output value). This choice of inputs is justified by the fact that the signal-to-background ratio, and the relative magnitude of the contribution of background “photons” from jets, varies as a function of the photon kinematic properties. In addition, the diphoton mass resolution depends on the probability that the correct primary vertex has been used to reconstruct the diphoton mass.

The following variables are used:

1) the relative transverse momenta of both photons, $p_{\mathrm{T}}^{1(2)}/m_{\gamma\gamma}$,
2) the pseudo-rapidities of both photons, $\eta^{1(2)}$,
3) the cosine of the angle between the two photons in the transverse plane, $\cos(\phi_1 - \phi_2)$,
4) the photon identification BDT output value for both photons,
5) the per-event relative mass resolution estimate assuming the mass has been constructed using the correct primary vertex (details below),
6) the per-event relative mass resolution estimate assuming the mass has been constructed using the incorrect primary vertex (details below),
7) The per-event probability that the correct primary vertex has been used to reconstruct the diphoton mass, computed from a linear fit to the event-level vertex selection MVA as described in Section 5.4.2.

The per-event relative mass resolution estimate is computed from the photon energy resolution estimate assuming Gaussian resolution functions as:

$$\sigma_m^{right}/m_{\gamma\gamma} = \frac{1}{2}\sqrt{\left(\sigma_{E1}/E_1\right)^2 + \left(\sigma_{E2}/E_2\right)^2} \tag{5.1}$$

Since the energy estimate from the regression is based on simulation, the σ_E of each photon (in both data and simulation) is increased to match the resolution in data by adding in quadrature the additional smearing applied to the single photon energy in the simulation (described in Chapter 4). The above computation assumes the correct primary vertex has been selected, in which case the energy measurement of the photons is the dominant contribution to the mass resolution. Since the correct primary vertex is not always selected, the relative mass resolution computed under the assumption that the incorrect primary vertex was chosen is added as input to the BDT. In this case the distance between the true vertex and the selected vertex is distributed as a Gaussian with width $\sqrt{2}\,\sigma_Z^{beamspot}$, and the contribution to the resolution σ_m^{vtx} can be computed analytically given the impact positions of the two photons in the calorimeter [1]. The relative resolution under the incorrect vertex hypothesis is then computed as:

$$\sigma_m^{wrong}/m_{\gamma\gamma} = \sqrt{\left(\sigma_m^{right}/m_{\gamma\gamma}\right)^2 + \left(\sigma_m^{vtx}/m_{\gamma\gamma}\right)^2} \tag{5.2}$$

In the training of the BDT, information needs to be provided that signal-to-background is inversely proportional to mass resolution. This is achieved by weighting the signal events used to train the BDT:

$$w_{sig} = \frac{p_{vtx}}{\sigma_m^{right}/m_{\gamma\gamma}} + \frac{1 - p_{vtx}}{\sigma_m^{wrong}/m_{\gamma\gamma}} \tag{5.3}$$

The BDT will thus tend to assign a high score classifier value to events with better resolution. The BDT is trained using simulated MC background and Higgs boson signal events.

---

[1] Since the Monte Carlo does not model the variation of the beamspot length during a fill, the average beamspot length (5.8 cm) is used to compute σ_m^{wrong} for both data and MC to build consistent input for the MVA.

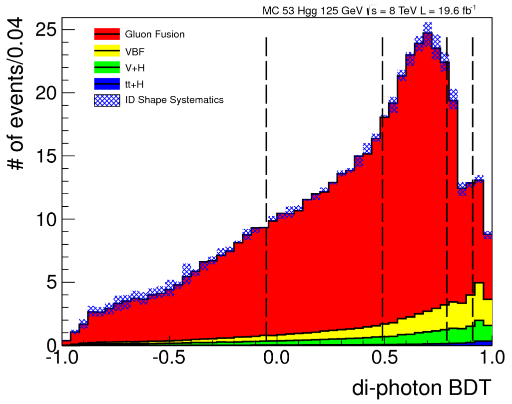


Figure 5.7: The diphoton MVA score for a SM Higgs boson signal ($m_{\mathrm{H}} = 125$ GeV). The nominal MVA output is shown as the stacked histogram (where the contributions are from different production processes). The blue band represents the shape variation resulting from propagating the systematic uncertainty assigned to the photon ID MVA score. The MVA event class boundaries are shown as vertical lines; events with an MVA score of less than -0.05 are discarded from the analysis.

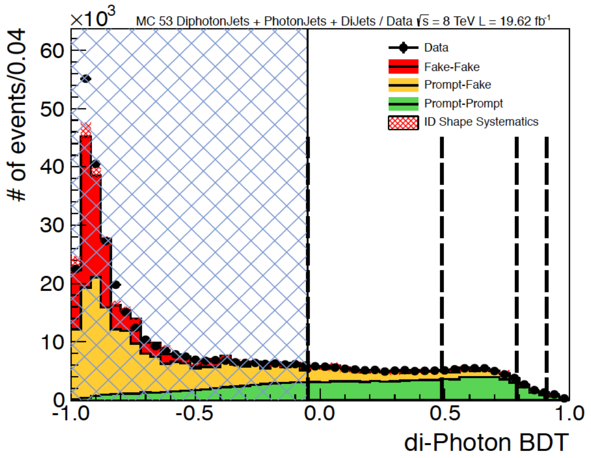


Figure 5.8: The diphoton MVA score for the background MC in the region $100 < m_{\gamma\gamma} < 180$ GeV, and data. The nominal MVA score is shown as the stacked histogram, representing different contributions to the background. There are significant uncertainties on the k-factors and background composition which are not shown here. The MVA event class boundaries are shown as vertical lines; events with an MVA score of less than -0.05 are discarded (blue hatched region).

5.7 Exclusive modes

5.7.1 Lepton tag

We exploit the exclusive selection of diphoton events produced in association with a high p_T lepton originating from the leptonic decay of vector bosons in the VH and $t\bar{t}H$ production mechanism. In this analysis muons and electrons are tagged as two separate channels.

In the presence of the tagged lepton QCD background is strongly suppressed. The main remaining background comes from electroweak processes with photons where a lepton is produced in Z^0 or W decays.

The VH rate associated with the lepton tag is suppressed by the leptonic decay branching fractions of the vector bosons: i.e. $\mathcal{B}(\mathrm{Z}^0 \to \ell\ell) \sim 7\%$, $\mathcal{B}(W \to \ell\nu) \sim 20\%$, where ℓ is an electron or a muon.

In contrast to the inclusive analysis, the signal-to-background for the lepton tagging is very large, close to one.

Given the larger purity and efficiency of the lepton selection compared to the diphoton selection, we start with lepton identification and based on that, the photon pair selection is performed. The procedure is outlined below.

1) Select events passing the diphoton triggers.

2) Select lepton.

3) Select the diphoton pair, by taking into account the selected lepton(s).

Muon selection

The tagged muon is the largest p_T muon with tight criteria on the reconstructed track with two indipendent algorithms, one starting from the inner tracker and the second one starting from the muon chambers. Highly consistent measurements between tracker and muon detector are essential in order to reduce the contamination from muons produced in decays of hadrons and from beam halo. The muon candidate is required also to have $p_T > 20$ GeV/c and $|\eta| < 2.4$.

The muon identification is based on cuts on the following variables: χ^2/n_{dof} of the global-muon track fit, transverse impact parameter of its tracker track with respect to the muon vertex, the longitudinal distance of the tracker track with respect to the muon vertex, number of pixel hits, and number of tracker layers with hits. Moreover, it is required that muon segments in at least two muon stations and at least one muon chamber are included in the global muon track fit.

The muon is also required to be isolated

Electron selection

Events not passing the muon tag selection undergo the electron tag selection. Electron candidates require a cluster with $E_T > 20$ GeV and pseudorapidity $|\eta| < 1.4442$ or $1.566 < |\eta| < 2.5$.

Misidentified jets are suppressed by the isolation requirement and the rejection of events with electrons coming from photon conversions is analogous to the one described in Sec. 4.2.4 for the electron selection of $Z \rightarrow e^+e^-$ events.

After this preselection, an MVA electron identification is used. Figure 5.9 shows the ROC curve for the MVA output. The signal sample consists of reconstructed electrons in MC samples (VH, $Z \rightarrow ee$ and $W \rightarrow e\nu$) matched to generated electrons using MC truth information. The background consists of reconstructed electrons in a data sample selected to have no real electrons. The efficiency and the fake rate shown are normalized to the acceptance (transverse momentum and pseudorapidity cuts).

The highest MVA score electron is considered as the electron tag candidate. In order to choose the cut on the MVA output variable an optimization which uses the $S/\sqrt{B}$ as figure of merit has been performed and the cut MVA$>$ 0.9 was chosen.

After the choice of the electron tag candidate, the vertex associated to its track is selected as the candidate vertex. A cut on the d_0 and d_Z, calculated with respect to the candidate vertex is applied.

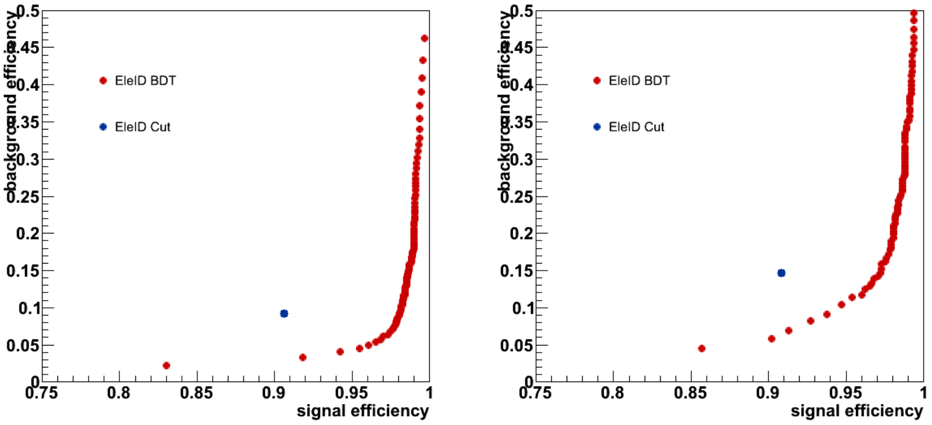


Figure 5.9: ROC curves for the cut on the electron MVA for EB (left) and EE (right). The blue dot represents the cut based electron ID provided by the POG and defined as *loose WP*.

Diphoton selection

The photon event selection is almost the same as the one used in the inclusive analysis. The diphoton kinematics are somewhat different when the Higgs has been produced via VH: Higgs from VH typically have higher p_T than those produced via gluon fusion due to recoil with the vector boson, resulting in a harder p_T distribution for the leading photon. Accordingly the threshold for $E_T(\text{lead})/m_{\gamma\gamma}$ is increased to 45/120 (from 40/120 in the inclusive categories).

In the cut-based diphoton selection, the subleading photon is required to have 25 GeV in p_T. The optimal cut may be even lower, but with p_T lower than 25 GeV trigger is less performing. However, in the MVA analysis the minimal p_T is required to be $1/4\ \times m_{\gamma\gamma}$, in order to stay within the training region of the event BDT.

For the electron tag there is a special source of background from Drell-Yan where an electron is faking a photon. Additional criteria are required to reject these photons during selection. Specifically, the ΔR between each photon and the closest GSF track from a reconstructed electron is required to be larger than 1.0. Despite this additional cut, Drell-Yan remains the dominant background in the electron tag channel at this stage.

Additional kinematic cuts are needed to reduce events with Final State Radiation

(FSR) and Initial State Radiation (ISR) in background processes with leptons.

In order to reject photons produced as final state radiation in events with a vector boson, we require $\Delta R(\gamma, lepton) > 1.0$ for both photons. Since $Z\gamma$ events, with $Z \to e^+e^-$, are the most important background in the electron tag events, a further dedicated cut is applied, in order to reduce this background. From the MC study, the major background comes from events where an electron is mis-identified as a photon because its track has not been reconstructed. In this case one of the two selected photons, when paired with the tagged electron, tends to give an invariant mass close to M_Z. To reduce this type of background we require that $|m(e,\gamma) - m_Z| > 10$ GeV. This cut is applied to both selected photons. The summary of the lepton tag selection is:

1) lepton identification
2) photon selection with:
   - $p_T/m_{\gamma\gamma} > 30/120, 40/120$;
   - usual MVA identification with same cuts of the other subchannels: $MVA > -0.2$;
   - additional electron veto for the electron tag channel $\Delta R(\gamma, GSFtrack)$;
   - $\Delta R(lep, \gamma) > 1$, $\Delta M_Z > 10$;
3) photon pair selection with the highest sum of p_T;
4) cut on the diphoton MVA variable: $BDT > -0.05$;

5.7.2 MET tag

Events in which the Higgs boson is produced in association with a W which decays leptonically fall in the lepton tag category if the lepton is reconstructed. In the cases where the lepton from W is not reconstructed due to some inefficiencies or because it is not in the acceptance fall in the MET-tagged category. Also, events in which the Z decays into two neutrinos can have large $\not\!\!E_{\mathrm{T}}$ in the final state.

Selection

The requirements for an event to be selected are:

1) the presence of two photons passing the quality cuts of the inclusive analysis
2) large missing transverse energy
3) the event is not selected as a VBF di-jet or a lepton tag event
4) further kinematic requirements to enhance the signal over background ratio, to reduce the contamination of gluon gluon fusion events, and to reduce instrumental effects

The same p_T cuts applied in the lepton tag analysis are used, with thresholds $\frac{p_T(leading)}{m_{\gamma\gamma}} > \frac{45}{120}$ and $p_T(subleading) > 25$ GeV.

Given their higher sensitivity, related to the better photon energy resolution, only the events with both photons in the barrel are considered for the MET-tag.

The selection applied in the analysis is summarized in Table 5.8.

variable	cut
di-photon MVA	>-0.05
p_T(leading)	$> m_{\gamma\gamma}\frac{45}{120}$
p_T(subleading)	> 25
$corr\not{E}_\mathrm{T}$	>70 GeV
$\Delta\phi(\gamma\gamma, leadjet)$	< 2.7
$\Delta\phi(\gamma\gamma, MET)$	> 2.1

Table 5.8: Selection applied in the analysis.

While the majority of signal events selected after the full set of cuts comes from the VH production mechanism, contributions from other processes are also present. The contributions to the selected signal sample for $m_H = 125$ GeV are as follows: 21% from gluon-gluon fusion, 3% from VBF production, 12% from $ttbarH$ associated production and 64% from VH production.

5.7.3 VBF

The VBF events have, in addition to the selected photons, two forward jets that may be measured in the detector and which originate from the two scattering quarks corresponding to the the Vector Boson Fusion (VBF) process. For a detailed study of VBF jets see [66].

In order to reduce the background contributions, jets are required to meet selection criteria aimed at rejecting fake jets due the clustering of several low energy emissions from pile-up interactions, as described in the following.

Events failing the exclusive VBF selection are then tested for inclusion in lower sensitivity categories or in the untagged ones.

Jet selection variables

We use corrected anti-k_T Particle Flow (PF) jets with $\Delta R = 0.5$. The pileup subtraction is performed using FastJet [67, 68] with deterministic seeding. Residual jet energy corrections are applied in data. The pseudorapidity is limited to $|\eta| < 4.7$.

The particles produced in the pile-up interactions are sometimes clustered by the jet clustering algorithm into objects of apparent relatively large p_T. The resulting "pile-up jets" are removed using selection criteria based on the compatibility of the jets' tracks with the primary vertex and the jets width. The variables used for jet identification are defined as follows:

- $\beta^* = \frac{\sum_{jetPUvtx} p_T^{tk}}{\sum_{jet} p_T^{tk}}$
- $RMS = \frac{\sum_{constituents} p_T^2 \Delta R^2}{\sum_{constituents} p_T^2}$

Different cuts are applied in different regions of the detector. Table 5.9 summarises the selection.

These cuts have been chosen to flatten the dependence of the background efficiency on the number of interactions per event, while keeping the efficiency high (> 95%) for jets from the hard interaction.

To discard jets reconstructed from the photons, we reject jets with $\Delta R < 0.5$ with respect to each of the selected photons, where $\Delta R = \sqrt{\Delta\eta(\mathrm{jet},\gamma)^2 + \Delta\phi(\mathrm{jet},\gamma)^2}$. The

jet η	β^*	RMS
$\eta < 2.5$	$< 0.2 \log(N_{vtx} - 0.64)$	< 0.06
$2.5 < \eta < 2.75$	$< 0.3 \log(N_{vtx} - 0.64)$	< 0.05
$2.75 < \eta < 3$	-	< 0.05
$3 < \eta < 4.7$	-	< 0.055

Table 5.9: Summary of jet-ID cuts.

two highest p_T jets are retained to compute the dedicated VBF selection variables.

The dijet tag selection is performed with a cut-based approach used in the cut-based analysis and an MVA approach used in the MVA-based analysis.

Cut-based dijet tag

The selection uses the following variables:

- the transverse momenta of the leading and subleading photons divided by the invariant mass of the di-photon candidate: $p_T^{\gamma_1}/m_{\gamma\gamma}$ and $p_T^{\gamma_2}/m_{\gamma\gamma}$;
- the transverse momenta of the leading and subleading jets: $p_T^{j_1}$ and $p_T^{j_2}$;
- the di-jet invariant mass, $M_{j_1 j_2}$;
- the difference in pseudorapidity between the two jets, $\Delta\eta_{j_1 j_2}$;
- the so-called *Zeppenfeld* variable [69], defined as $Z = \eta|_{obs} - \frac{\eta(j_1)+\eta(j_2)}{2}'$, where $\eta|_{obs} = \eta(\gamma_1 + \gamma_2)$;
- the difference in azimuthal angle between the dijet and the diphoton, $\Delta\phi_{(j_1 j_2, \gamma\gamma)}$.

The events are split in two categories, depending on the invariant mass of the dijet system and the p_T of the subleading jet. The final cuts are summarized in Table 5.10 . Events falling in the first category are removed from the second one.

Variable	tight category	loose category
$p_T^{\gamma_1}/m_{\gamma\gamma}$	> 0.5	> 0.5
$p_T^{\gamma_2}$	> 25 GeV	> 25 GeV
$p_T^{\mathrm{j}_1}$	> 30 GeV	> 30 GeV
$p_T^{\mathrm{j}_2}$	> 30 GeV	> 20 GeV
$\|\Delta\eta_{j_1 j_2}\|$	> 3.0	> 3.0
$\|Z\|$	< 2.5	< 2.5
$M_{j_1 j_2}$	> 500 GeV	> 250 GeV
$\|\Delta\phi(\mathrm{jj},\gamma\gamma)\|$	> 2.6	> 2.6

Table 5.10: Final selection cuts for the VBF selection. Events from the first category are removed from the second one.

Figure 5.10: Kinematic dijet MVA output.

Kinematic dijet MVA

The kinematic dijet MVA is built using the variables described in Section 5.7.3. The information is combined by means of a *Boosted Decision Tree* algorithm. The BDT is trained on Monte Carlo events: a VBF H $\rightarrow \gamma\gamma$ sample with M_H =124 GeV is used as signal and the diphoton+jets sample as background. Figure 5.10 shows the BDT output for signal and background events.

Events passing the following criteria are pre-selected as possible VBF candidates:

- two photons passing the preselections with $p_T^{\gamma_1}/m_{\gamma\gamma}$ >1/3 and $p_T^{\gamma_1}/m_{\gamma\gamma}$ >1/4;
- diphoton MVA > -0.05;
- two jets passing the pile-up jet identification described in the previous section and with $p_T^{j_1} > 30$ GeV,$p_T^{j_2} > 20$ GeV;
- $M_{j_1 j_2}$ >250 GeV

In order to increase the number of simulated background events, the preselection cuts where relaxed for the BDT training. Namely the thresholds on the leading jet p_T and the di-jet invariant mass were set to 20 GeV and 100 GeV respectively.

Category optimization

The optimization of the number of categories and their boundaries in the BDT classifier is done by integrating piecewise the signal and background BDT output distributions and maximizing the significance, defined as:

$$z^2 = N \sum_{i=0}^{n_{cat}} \frac{\left(\int_{C_i}^{C_{i+1}} \epsilon_{signal}\right)^2}{\int_{C_i}^{C_{i+1}} \epsilon_{background}} \tag{5.4}$$

with respect to the category boundary positions C_0 ... C_N. N is an overall normalization factor depending only on the luminosity. For this optimization the diphoton+jets sample was used as background and the VBF H → $\gamma\gamma$ sample as signal. The optimal performances are obtained with two categories, while adding a third category was found to bring only a marginal gain in sensitivity. The optimized category boundaries are summarized in Table 5.11.

	lower boundary	upper boundary
tight dijet category	0.985	1.0
loose dijet category	0.93	0.985

Table 5.11: VBF dijet tag categories boundaries.

The number of expected events for signal and background in the different event classes for 7 TeV and 8 TeV is shown in Tab. 5.12.

Table 5.12: Expected number of SM Higgs boson events (m_{H} =125 GeV) and estimated background (at $m_{\gamma\gamma}$ =125 GeV) for all event classes of the 7 and 8 TeV datasets for the MVA-based analysis analysis. The composition of the SM Higgs boson signal in terms of the production processes and its mass resolution is also given.

Expected signal and estimated background

Event classes		SM Higgs boson expected signal (m_{H} =125 GeV)							Background
		Total	ggH	VBF	VH	ttH	σ_{eff} (GeV)	FWHM/2.35 (GeV)	$m_{\gamma\gamma}$ = 125 GeV (ev./GeV)
7 TeV 5.1fb^{-1}	Untagged 0	3.2	61.4%	16.8%	18.7%	3.1%	1.21	1.14	3.3 ± 0.4
	Untagged 1	16.3	87.6%	6.2%	5.6%	0.5%	1.26	1.08	37.5 ± 1.3
	Untagged 2	21.5	91.3%	4.4%	3.9%	0.3%	1.59	1.32	74.8 ± 1.9
	Untagged 3	32.8	91.3%	4.4%	4.1%	0.2%	2.47	2.07	193.6 ± 3.0
	Dijet tag	2.9	26.8%	72.5%	0.6%	–	1.73	1.37	1.7 ± 0.2
8 TeV 19.6fb^{-1}	Untagged 0	17.0	72.9%	11.6%	12.9%	2.6%	1.36	1.27	22.1 ± 0.5
	Untagged 1	37.8	83.5%	8.4%	7.1%	1.0%	1.50	1.39	94.3 ± 1.0
	Untagged 2	150.2	91.6%	4.5%	3.6%	0.4%	1.77	1.54	570.5 ± 2.6
	Untagged 3	159.9	92.5%	3.9%	3.3%	0.3%	2.61	2.14	1060.9 ± 3.5
	Dijet tight	9.2	20.7%	78.9%	0.3%	0.1%	1.79	1.50	3.4 ± 0.2
	Dijet loose	11.5	47.0%	50.9%	1.7%	0.5%	1.87	1.60	12.4 ± 0.4
	Muon tag	1.4	0.0%	0.2%	79.0%	20.8%	1.85	1.52	0.7 ± 0.1
	Electron tag	0.9	1.1%	0.4%	78.7%	19.8%	1.88	1.54	0.7 ± 0.1
	$\not\!\!E_{\mathrm{T}}$ tag	1.7	22.0%	2.6%	63.7%	11.7%	1.79	1.64	1.8 ± 0.1

5.8 Statistical analysis

The test of the different hypotheses is performed based on data distributions that are sensitive to the underlying physics on which the hypotheses are based. The key distribution in this analysis is the di-photon invariant mass distribution where the background has a smoothly decreasing distribution, the Higgs boson signal distribution is a narrow peak. For the hypothesis tests a likelihood function is defined:

$$\mathcal{L}(data|POI,\vec{\theta}) = k^{-1}\prod_i \left(\mu S f_s(x_i;POI,\vec{\theta}) + B f_b(x_i;POI,\vec{\theta})\right) e^{-(\mu S+B)}\cdot p(\tilde{\theta}|\theta) \quad (5.5)$$

where f_s and f_b are Probability Density Functions (*pdf*s) for the expected signal and background respectively, and S and B are the corresponding number of expected events. μ is referred as the *signal strength*; it is a scale factor on the cross-section× branching ratio expected for the SM Higgs boson. The *pdf* depends on two sets of parameters: the *Parameters of Interest* (POIs) and the *nuisance parameters* ($\vec{\theta}$). The POIs are the parameter to be measured; the *nuisance parameters* are instead not of central interest, but needed to model the *pdf*s.

The POIs' best fit values are those which maximize the likelihood $\mathcal{L}$. The *nuisance parameters* are set by previous measurements and they are left to float within their uncertainties. For each POI value, the likelihood is evaluated scanning the *nuisance parameters* within their uncertainties. Systematic uncertainties are treated as *nuisance parameters*.

By writing a systematic error *pdf* as the posterior $\rho(\tilde{\theta})$ constructed from a fictional auxiliary "measurement", the *pdf* $p(\tilde{\theta})$ for that auxiliary measurement can be used to constrain the likelihood of the main measurement in a frequentist calculation.

5.8.1 Exclusion limits

After the discovery, there is still interest in looking at new possible signal-like peaks in the di-photon invariant mass distribution due to other Higgs bosons foreseen in beyond SM theories. The parameter of interest is the signal strength μ.

The ATLAS and CMS collaborations agreed to use a hybrid-frequentist approach in order to calculate the exclusion limits [70].

A test statistic variable $\tilde{q}_\mu$ is defined as:

$$\tilde{q}_\mu = -2 \ln \frac{\mathcal{L}\left(data|\mu, \hat{\theta}_\mu\right)}{\mathcal{L}\left(data|\hat{\mu}, \hat{\theta}_\mu\right)} \tag{5.6}$$

where $\hat{\theta}_\mu$ refers to the conditional maximum likelihood estimators of θ, given the signal strength parameter μ and "data" that, may refer to the actual experimental observation or pseudo-data (toys). The pair of parameter estimators $\hat{\mu}$ and $\hat{\theta}$ correspond to the global maximum of the likelihood.

The *observed* value of the test statistic $\tilde{q}_\mu^{obs}$ for the given signal strength modifier μ under test is evaluated ($\mu = 0$ correspond to the background-only hypothesis).

The values of $\hat{\theta}_0^{obs}$ and θ_μ^{obs} best describing the experimentally observed data are found miximizing the likelihood for the background-only and signal+background hypothesis, respectively.

The *pdf* $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{obs})$ and $f(\tilde{q}_\mu|0, \hat{\theta}_0^{obs})$ are built generating toy MC pseudo-data.

Then, two p-values are defined for the two hypothesis $\mu = 0$ and $\mu! = 0$: p_b and p_μ.

$$\begin{aligned} p_\mu = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}|signal + background) &= \int_{\tilde{q}_\mu^{obs}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{obs}) d\tilde{q}_\mu \\ 1 - p_b = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}|background - only) &= \int_{\tilde{q}_\mu^{obs}}^{\infty} f(\tilde{q}_\mu|0, \hat{\theta}_\mu^{obs}) d\tilde{q}_\mu \end{aligned} \tag{5.7}$$

The ratio of the two p-values is defined as

$$CL_s(\mu) = \frac{p_\mu}{1 - p_b} \tag{5.8}$$

The 95% Confidence Level upper limit on μ, is the value of μ that gives $CL_s = 0.05$.

5.8.2 Quantification of an excess

The presence of the signal is quantied by the background-only p-value. For a given Higgs boson mass hypothesis M_H, the test statistic q_0 defined above is used.

5.9 Signal extraction

In this section, the modelling of the signal and background in the statistical analysis will be discussed.

5.9.1 Background modelling (f_b)

In the $H \rightarrow \gamma\gamma$ analysis the background estimation is performed directly on data, fitting the invariant mass shape in a wide mass range between 110 and 180 GeV with a suitable background model parametrization. This model serves as fully-differential prediction of the mean expected di-photon mass distribution in the background-only hypothesis.

In principle it is desirable to know the functional form of the background mass distribution exactly, or at least to be able to summarize the limited knowledge of the shape in a finite set of continuously-varying parameters.

Several functions have been tested and a set of polinomials has been choosed as those minimizing the possible bias in the observed limits.

Polynomial order determination

An unbinned extended maximum likelihood fit is performed to the observed data. The resulting minimized NLL can be used to determine a goodness-of-fit-like quantity, in particular the 2 times differences of between the NLL values for two incarnations of the same functional family F with different orders N and $M > N$. The $2 \times$ NLL follows a χ^2 distribution with n degrees of freedom (dof), where $n = (\mathrm{dof}(MF) - \mathrm{dof}(NF))$. The variable

$$\chi^2_{N\rightarrow(N+1)} := 2(NLL_N - NLL_{N+1}) \tag{5.9}$$

can thus be used to decide if the higher polynomial order fit the data *significantly better*; *significantly* better means

$$p\left(\chi^2 \geq \chi^2_{N\rightarrow(N+1)}\right) < 0.05. \tag{5.10}$$

The procedure to define the suitable order of each family to design the truth-model, is to increase the order until the p-value for an additional increase is larger than 0.05.

Polynomials of orders from 2 to 5 are found to fulfil the requirements above and are used to model the background distributions in the various categories for both the mass-fit-MVA and the cut-based analyses.

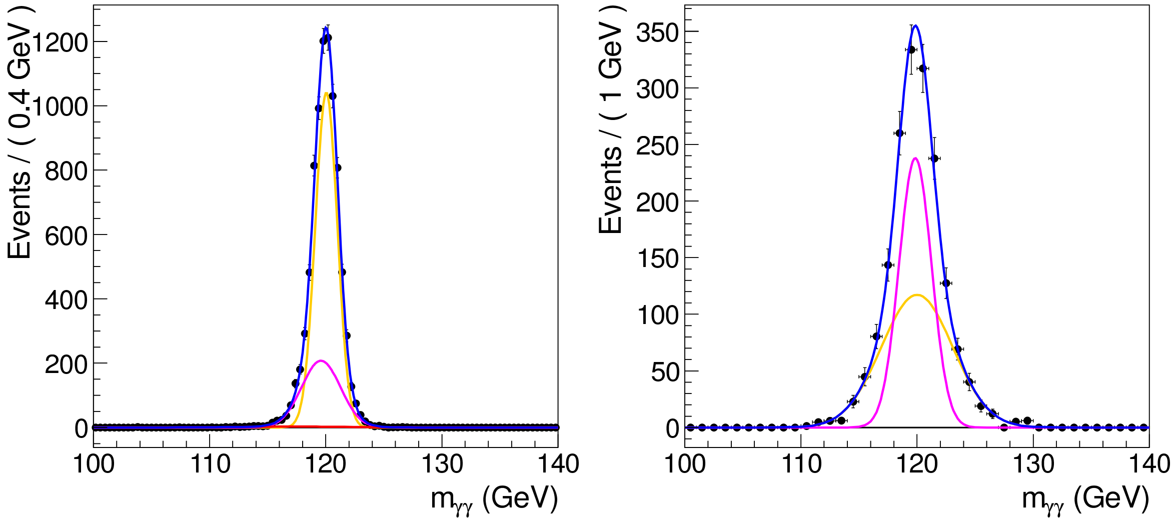
5.9.2 Signal modelling

In order to statistically interpret the observed data, it is necessary to have a description of the signal which specifies the overall efficiency×acceptance, as well as the shape of the di-photon mass distribution in each of the nine classes. The Monte Carlo is used, after the smearing of the resolution, and the application of all efficiency corrections and scale factors, to build a parameterized model for the signal which is defined continuously for any value of the Higgs mass between 110 and 150 GeV. The strategy for the shape model is to describe the signal Monte Carlo with an analytic function. The parameters of this function are then determined by fitting the Monte Carlo for each available mass point. The full signal model is then defined by a linear interpolation of each fit parameter between the fitted mass values. For each mass point the signal Monte Carlo is fitted separately for each of the four production mechanisms (gluon-fusion, vector boson fusion, W/Z associated production, and $t\bar{t}$+Higgs associated production). The analytic functions for each production mechanism are added together at the end according to their relative cross-sections in the SM.

The signal shape for events with the correct primary vertex selection is dominated by the detector and reconstruction response in the ECAL. The signal shape for these events is modeled empirically in each class by a sum of 2 or 3 Gaussians, depending on the class. The means, widths, and relative fractions of the Gaussians are left free in the fit to the Monte Carlo.

The signal shape for events with incorrect primary vertex selection is smeared significantly by the variation in the z position of the selected primary vertex with respect to the true higgs production point. The signal shape for these events is modeled in each class by a sum of 2 Gaussians, or a single Gaussian, depending on the class. The means, widths, and relative fraction (for classes with two Gaussians) are left free in the fits to the Monte Carlo. A representative set of fits for events with correct and incorrect primary vertex selection in one class for gluon-fusion production is shown in Fig. 5.11.

The combined shape in each class for correct and incorrect vertex selection is constructed by adding the shapes for the two sub-components together, according to the correct vertex selection efficiency determined from Monte Carlo. This efficiency is treated



(a) Correct Vertex Selection (b) Incorrect Vertex Selection

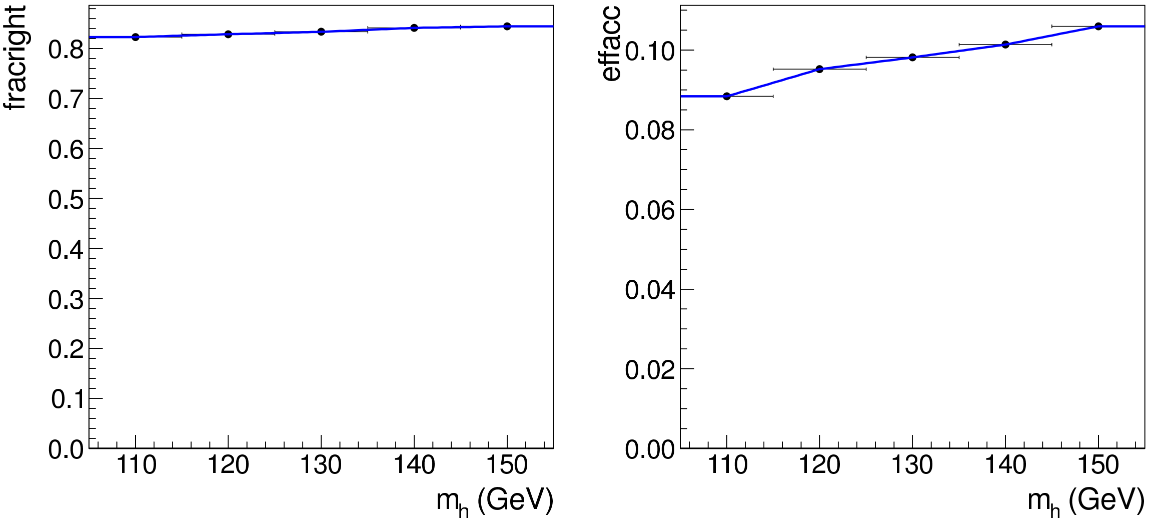
Figure 5.11: Fit results for signal shape model with correct and incorrect primary vertex selection for a single class ($0.72 < bdt < 0.89$) in 120 GeV Gluon-Fusion Higgs Monte Carlo. The black points are the weighted Monte Carlo events and the blue lines are the corresponding fit results. Individual Gaussian components of the fits are also shown. This plot uses an older definition of the class boundaries and it's shown for illustration only.

as another model parameter for the purposes of interpolation between mass points.

In order to facilitate the interpretation of the signal model in terms of a Standard Model Higgs production cross-section, and in order to facilitate the use of the signal model simultaneously across the four classes, we parameterize the signal yield in terms of a per class acceptance× efficiency, computed from Monte Carlo after all of the appropriate weights and scale factors have been applied. The evolution of the correct vertex fraction, as well as the acceptance×efficiency (into a single class) with Higgs mass are shown for one class for gluon-fusion production in Fig. 5.12. The final parameterized shapes for each class for a Higgs mass of 125 GeV for the Standard Model cross-section weighted mixture of all production mechanisms is shown in Fig. 5.13 for each event class separately.

The determination of the full set of signal model parameters at each Monte Carlo mass point is used to construct a signal model continuous in Higgs mass by performing a linear interpolation of each fit parameter, which gives rise to a smooth evolution of the signal shape. A closure test of this procedure is shown in Fig. 5.14 where the 115 GeV Monte Carlo is shown along with the interpolated shape in one of the classes for gluon-fusion production.

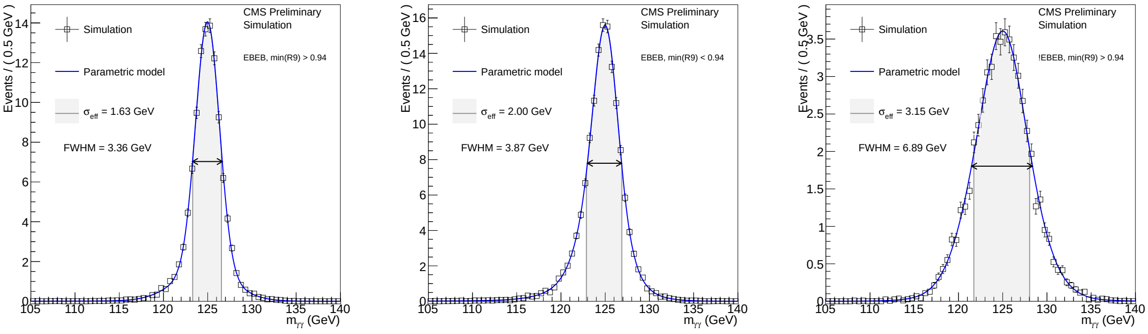
Systematic uncertainties affecting the shape are incorporated as parametric variations of the model. Uncertainty in the vertex selection efficiency are treated by varying the relative additive fraction of the right and wrong vertex shapes. Uncertainty in the energy scale is incorporated as a shift in the mean of each Gaussian, and uncertainty on the resolution is incorporated by the analytic convolution or deconvolution of an additional width with each of the Gaussians.



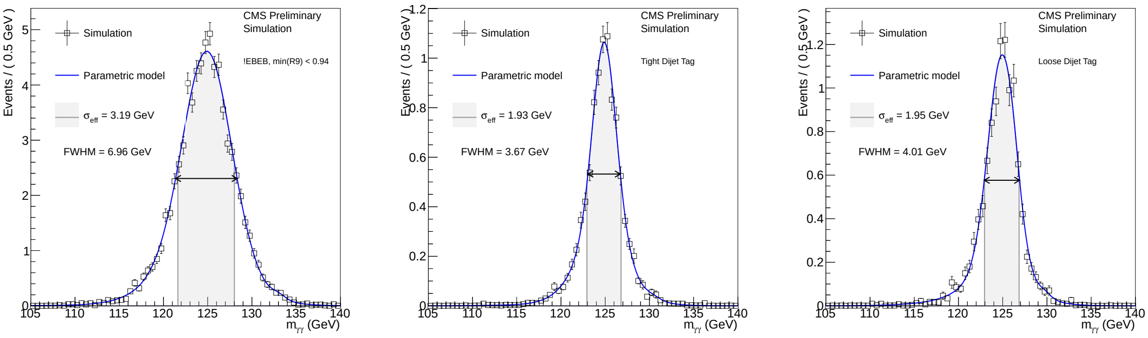
(a) Vertex Selection Efficiency

(b) Acceptance×Efficiency

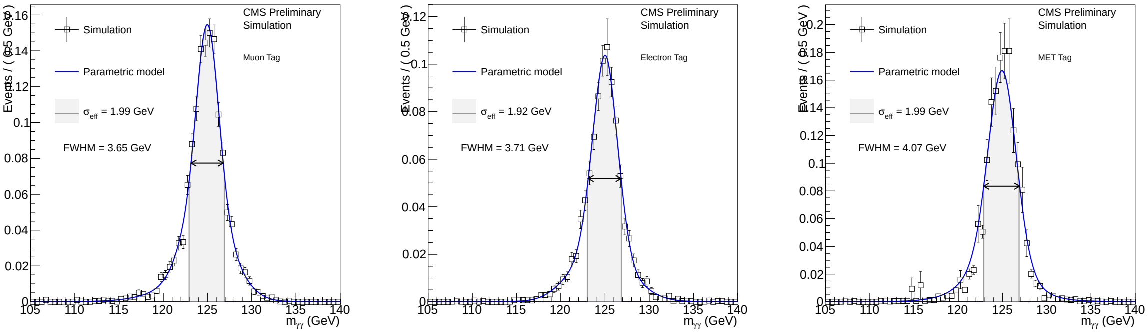
Figure 5.12: Correct vertex selection efficiency and Acceptance×Efficiency for one class ($0.72 < bdt < 0.89$) for the fitted masses of 110, 120, 130, 140 GeV along with the linear interpolations between mass points. This plot uses an older definition of the class boundaries and it's shown for illustration only.



(a) Both EB, High R_9 (b) Both EB, Low R_9 (c) One EE, High R_9



(d) One EE, Low R_9 (e) Dijet tag cat1 (f) Dijet tag cat2



(g) Muon tag (h) Electron tag (i) MET tag

Figure 5.13: Full parameterized signal shape for the 9 classes in 125 GeV Higgs Monte Carlo for the cut-based analysis. The black points are the weighted Monte Carlo events and the blue lines are the corresponding models. Also shown are the effective σ values and corresponding intervals.

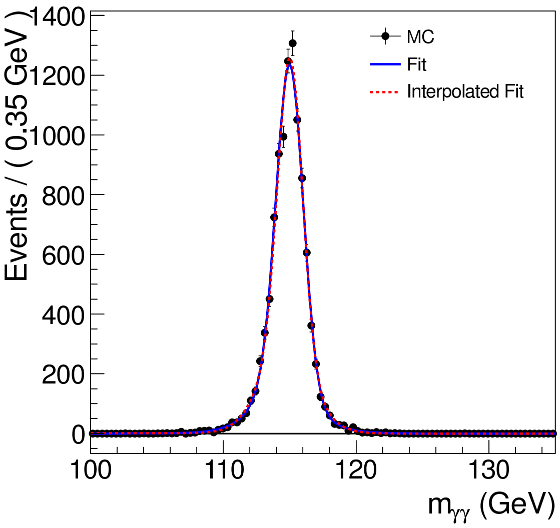


Figure 5.14: Combined signal shape model for correct and incorrect primary vertex selection for one class ($0.72 < bdt < 0.89$) in 115 GeV gluon fusion Higgs Monte Carlo. The black points are the weighted Monte Carlo events and the blue lines are the corresponding fit models. The red dashed lines are the signal models interpolated from the 110 GeV and 120 GeV fits.

5.10 Systematic uncertainties

The systematic uncertainties affecting the signal considered in the analysis performed on the 8 TeV dataset are summarized in Table 5.13. The methods used to estimate them are reported here below. Most of the systematic uncertainties are common to all analyses; analysis specific ones are detailed where needed.

The systematic uncertainties calculated at the single photon level are:

- *Energy scale and resolution:* scale and resolutions are studied with electrons from $Z \rightarrow e^+e^-$ and then applied to photons. The main source of systematic uncertainty is the different interactions of electrons and photons with material upstream the ECAL. Uncertainties are assessed by changing the rescaling of R_9 distributions, changing the R_9 selection, the regression training and the electron selection used.

- *Photon identification:* taken as the largest uncertainty on the data/MC scale factors computed on $Z \rightarrow e^+e^-$ events using a tag-and-probetechnique. This systematic uncertainty is applied to the photon identification in the cut-based analysis and to the loose photon preselection in the MVA analysis.

- *R_9 selection (cut-based):* taken from the data/MC comparison of the photon R_9 categorization in $Z \rightarrow \mu\mu\gamma$ events. The statistical uncertainty on the single photon is propagated to the diphoton categories and the result is assigned as systematic uncertainty on the category migration between low and high R_9 categories.

- *Photon identification BDT and photon energy resolution BDT (MVA analysis):* the agreement between data and simulation is assessed using Z $\rightarrow e^+e^-$ candidates, Z $\rightarrow \mu^+\mu^-\gamma$ candidates and the highest transverse energy photon in the diphoton invariant mass region where $m_{\gamma\gamma} > 160$ GeV (the fake photon contribution becomes smaller at high diphoton invariant mass). Both the inputs to the diphoton BDT and its output value are compared. A variation of ± 0.01 on the photon identification BDT output value, together with an uncertainty on the per-photon energy resolution estimate, parametrized as a rescaling of the resolution estimate by $\pm 10\%$ about its nominal value, fully covers the differences observed between data and MC simulation.

The systematic uncertainties calculated at the event level are:

- *Integrated luminosity:* the luminosity uncertainty is estimated as described in [71].
- *Vertex finding efficiency:* taken from the statistical uncertainty on the data/MC scale factor on $Z \to \mu^+\mu^-$ and the uncertainty on the signal p_T distribution arising from theory uncertainties.
- *Trigger efficiency:* extracted from $Z \to e^+e^-$ using a tag-and-probetechnique and rescaling them to take into account the different R_9 distributions for electrons and photons.
- *Global energy scale:* an uncertainty of 0.25% to account for the imperfect modelling by the MC of electron/photon differences and 0.4% to account for possible non-linearity when extrapolating from the Z-mass scale to the $m_H \sim 125$ GeV, are added. The two uncertainties, which together amount to 0.47%, are fully correlated between all the categories of the analysis.

The systematic uncertainties for the events with exclusive signatures are:

- *dijet tagging efficiency:* two effects are taken into account: the uncertainty on the MC modelling of the jet-energy corrections and resolution; and the uncertainty in predicting the presence of jets and their kinematic properties. They are calculated using different underlying event tunes and from the uncertainty on parton distribution functions and QCD scale factor.
- *Lepton identification efficiency:* for both electrons and muons, the uncertainty on the identification efficiency is computed varying the data/simulation efficiency scale factor by its uncertainty. The resulting difference in the signal efficiency estimated in the MC simulation is taken as systematic uncertainty.
- $\not{E}_T$ *selection efficiency:* Systematic uncertainties due to $\not{E}_T$ reconstruction are estimated in both signal events where real $\not{E}_T$ is expected (i.e. VH production) and the other Higgs production modes. For VH signal events the uncertainty is estimated

by applying or not the $\not\!\!E_{\mathrm{T}}$ corrections and taking the difference in efficiency as a systematic uncertainty. For the other modes (gluon-gluon fusion, VBF and $t\bar{t}H$) the uncertainty is mainly due to a different fraction of events in the tail of the $\not\!\!E_{\mathrm{T}}$ distribution. The systematic uncertainty is evaluated by comparing diphoton data and MC in a control sample enriched in γ+jet events which looks similar (in terms of $\not\!\!E_{\mathrm{T}}$) to the Higgs signal events.

The theoretical systematic uncertainties considered are:

- *Production cross section:* the systematic uncertainty on the production cross section follows the recommendation of the LHC Higgs Cross Section working group [64].

Table 5.13: Separate sources of systematic uncertainties accounted for in the analysis of the 8 TeV data set.

Sources of systematic uncertainty		**Uncertainty**	
Per photon		Barrel	Endcap
Energy resolution ($\Delta\sigma/E_{MC}$)	$R_9 > 0.94$ (low η, high η)	0.23%, 0.72%	0.93%, 0.36%
	$R_9 < 0.94$ (low η, high η)	0.25%, 0.60%	0.33%, 0.54%
Energy scale ($(E_{data} - E_{MC})/E_{MC}$)	$R_9 > 0.94$ (low η, high η)	0.20%, 0.71%	0.88%, 0.12%
	$R_9 < 0.94$ (low η, high η)	0.20%, 0.51%	0.18%, 0.12%
Photon identification efficiency		1.0%	2.6%
Cut-based			
$R_9 > 0.94$ efficiency (results in class migration)		4.0%	6.5%
MVA analyses			
Photon identification BDT		±0.01 (shape shift)	
	(Effect of up to 4.3% event class migration.)		
Photon energy resolution BDT		±10% (shape scaling)	
	(Effect of up to 8.1% event class migration.)		
Per event			
Integrated luminosity		4.4%	
Vertex finding efficiency		0.2%	
Trigger efficiency		1.0%	
Global energy scale		0.47%	
Dijet selection			
Dijet-tagging efficiency	VBF process	10%	
	Gluon-gluon fusion process	30%	
(Effect of up to 15% event migration among dijet classes.)			
Muon selection			
Muon identification efficiency		1.0%	
Electron selection			
Electron identification efficiency		1.0%	
$\not\!\!E_{\mathrm{T}}$ selection			
$\not\!\!E_{\mathrm{T}}$ cut efficiency	Gluon-gluon fusion	15%	
	Vector boson fusion	15%	
	Associated production with W/Z	4%	
	Associated production with $t\bar{t}H$	4%	
Production cross sections		Scale	PDF
Gluon-gluon fusion		+7.6% -8.2%	+7.6% -7.0%
Vector boson fusion		+0.3% -0.8%	+2.6% -2.8%
Associated production with W/Z		+2.1% -1.8%	4.2%
Associated production with $t\bar{t}H$		+4.1% -9.4%	8.0%

5.11 Results

Background model fits to the diphoton mass distribution, on the hypothesis of no signal, are shown for the nine event classes in Figs. 5.15 and 5.16. The error bands show the uncertainty on the background shapes associated with the statistical uncertainties of the fits. They are generated by randomly throwing values of the polynomial coefficients according to the covariance matrix of the fit and building confidence intervals from the sampled values of the curve at each value of $m_{\gamma\gamma}$.

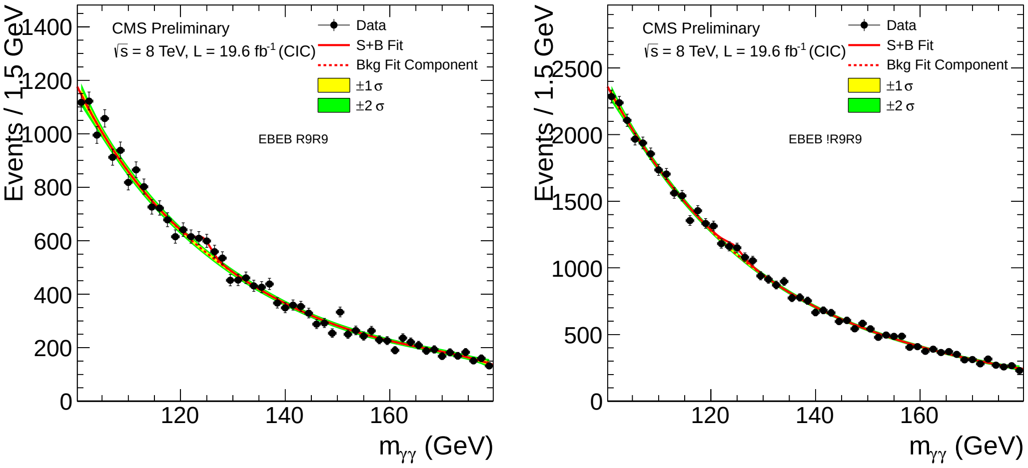
Exclusion limits

The 95% confidence level upper limit on the cross-section for a Standard Model Higgs boson decaying to 2 photons, relative to the SM expectation, is presented in Fig. 5.17 for the cut-based and MVA-based analyses. The systematic uncertainties on the expected cross-section are included in the limit setting procedure. In order to ensure that the data is tested in intervals of Higgs mass which are finer than the resolution, the observed limits are evaluated in 500 MeV steps, except in the region 120< m_{H} <130 GeV, where limits are evaluated in 100 MeV steps.

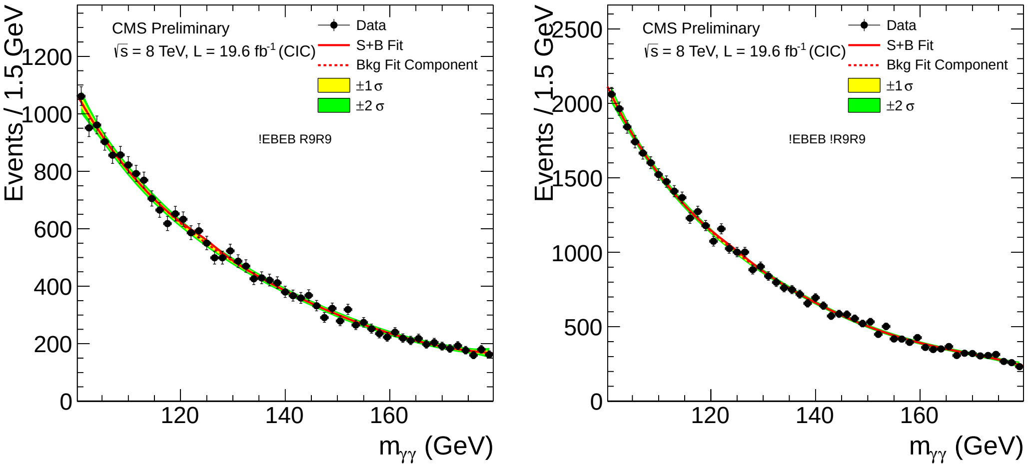
The mass range between 110 and 149 GeV is excluded at 95% confidence level, except the region between 122.8 and 127.8 where an excess of events is found. Similar regions are obtained for the cut-based analysis.

Excess quantification for discovered Higgs boson at 125 GeV

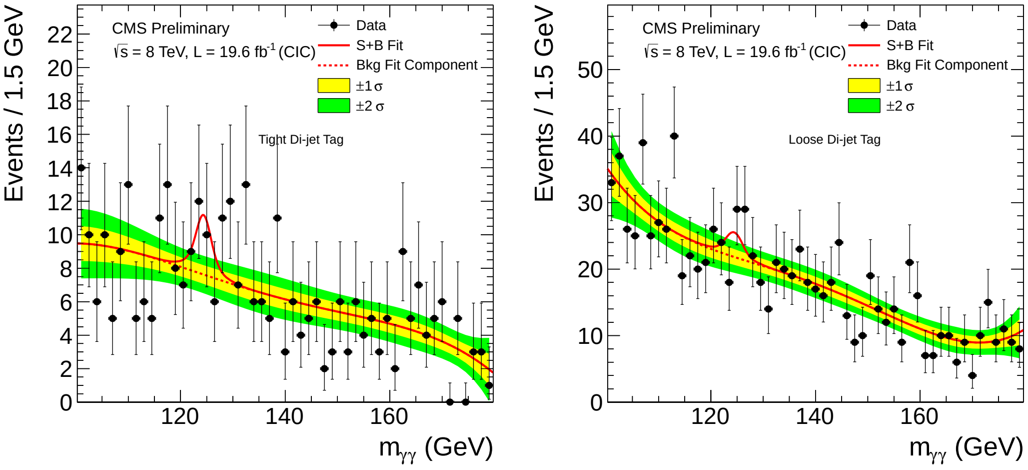
Figure 5.18 shows the local p-values for the different analyses, calculated using the asymptotic approximation in the mass range $110 < m_{\mathrm{H}} < 150$ GeV for the 7 TeV + 8 TeV datasets combined as well as for the 7 TeV and the 8 TeV datasets separately. The local p-value quantifies the probability for the background to produce a fluctuation as large as the observed one or larger, and assumes that the relative signal strength between the event classes follows the MC signal model for the Standard Model Higgs boson. For the MVA-based analysis, the local p-value corresponding to the largest signal-like fluctuation of the observed limit, at 125 GeV, has been computed to be 3.2 σ in the asymptotic approximation where a local significance of 4.2σ is expected from a SM Higgs boson. For



(a) Both EB, High R_9 (b) Both EB, Low R_9

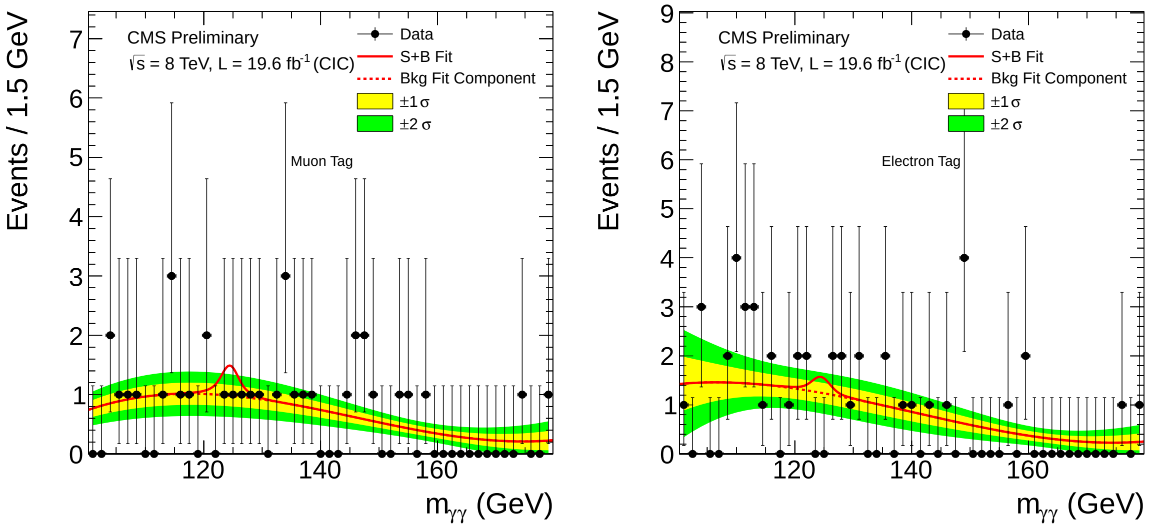


(c) One EE, High R_9 (d) One EE, Low R_9



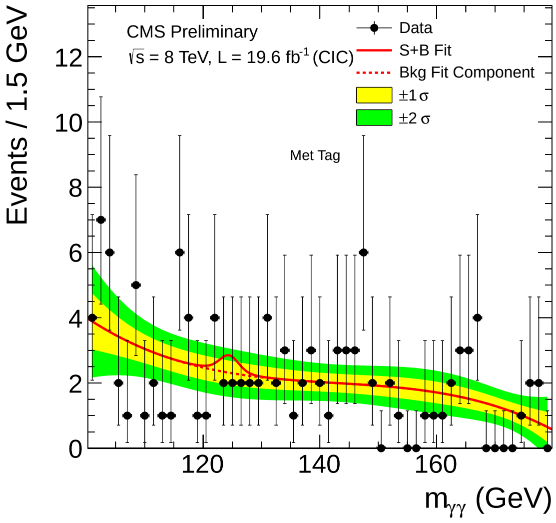
(e) Dijet tag cat1 (f) Dijet tag cat2

Figure 5.15: Background model fits to the four inclusive classes and the two dijet classes for the cut-based analysis.



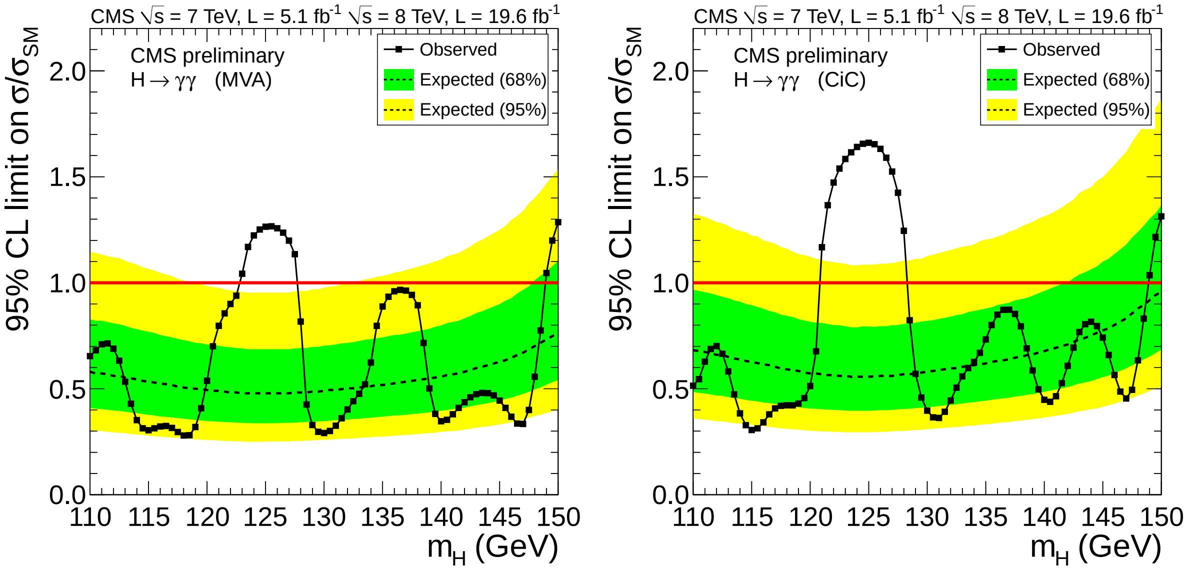
(a) Muon tag

(b) Electron tag



(c) MET tag

Figure 5.16: Background model fits to the muon, electron and MET classes for the cut-based analysis.



(a) Mass-fit-MVA.

(b) Cut-based analysis.

Figure 5.17: The exclusion limit on the cross section of a SM Higgs boson decaying into two photons as a function of the boson mass hypothesis relative to the SM cross section. The results are shown for the MVA-based analysis analysis (left) and the cut-based analysis (right).

	MVA-based analysis (at m_H = 125 GeV)	cut-based (at m_H = 124.5 GeV)
7 TeV	$1.69^{+0.65}_{-0.59}$	$2.27^{+0.80}_{-0.74}$
8 TeV	$0.55^{+0.29}_{-0.27}$	$0.93^{+0.34}_{-0.30}$
7 + 8 TeV	$0.78^{+0.28}_{-0.26}$	$1.11^{+0.32}_{-0.30}$

Table 5.14: The values of the best fit signal strength for the different datasets and analyses.

the cut-based analysis, the largest signal-like fluctuation is observed at 124.5 GeV with a corresponding value of 3.9σ ($3.5\ \sigma$ expected).

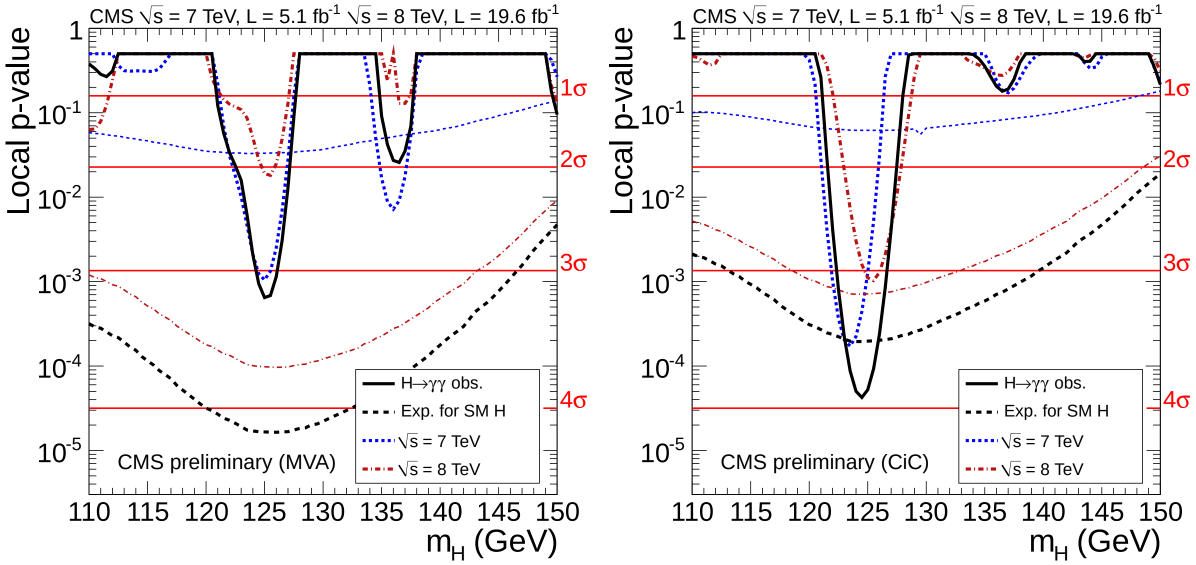
In Figure 5.19 the combined best fit signal strength is shown as a function of the Higgs boson mass hypothesis, for both the MVA-based analysis analysis and the cut-based one.

The best fit signal strength corresponding to the largest signal like fluctuation at 125 GeV is $\sigma/\sigma_{SM} = 0.78^{+0.28}_{-0.26}$ for the MVA-based analysis analysis and $\sigma/\sigma_{SM} = 1.11^{+0.32}_{-0.30}$ at the mass of 124.5 GeV for the cut-based analysis. The values of the best fit signal strength, derived separately for the 7 and 8 TeV datasets for the two analyses, are reported in table 5.14. As a further cross-check, a second MVA-based analysis which uses a background model derived from the signal sidebands [72], has also been performed giving compatible results.

In Figure 5.20 the best fit signal strengths is shown in each of the event classes and separately for the 7 TeV and 8 TeV datasets. The vertical line corresponds to the SM Higgs boson mass hypothesis corresponding to the largest signal-like fluctuation in Figure 5.19: 125 GeV for the MVA-based analysis analysis and 124.5 GeV for the cut-based analysis. The band corresponds to $\pm 1\sigma$ uncertainties on the overall value.

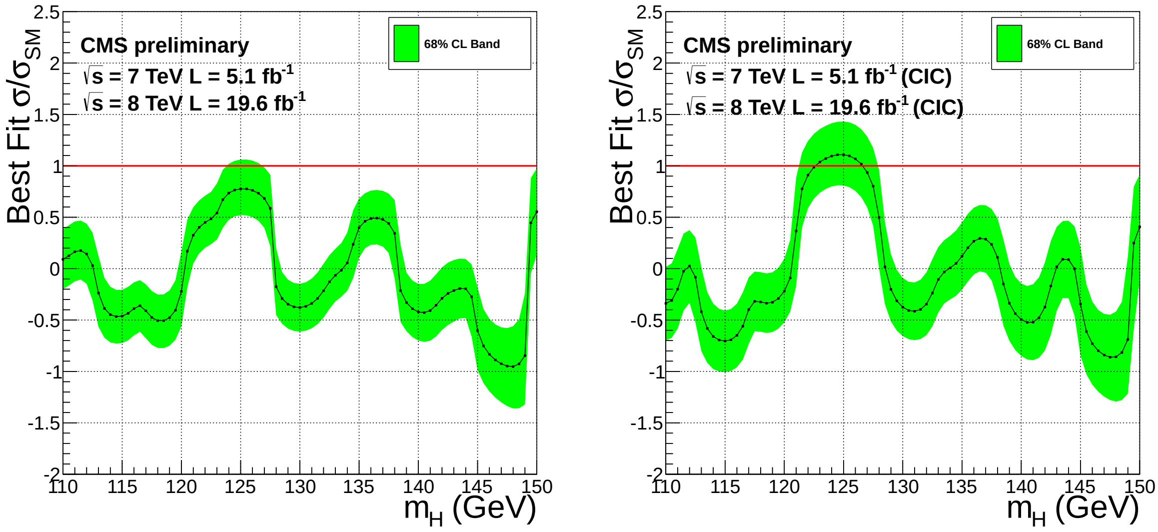
Higgs couplings

The signal strength modifier assumes the scaling of the SM cross-section×branching ratio by the same factor for all the production modes μ. The study of the couplings is carried out defining single scaling factors either for each production mode or for some of them.



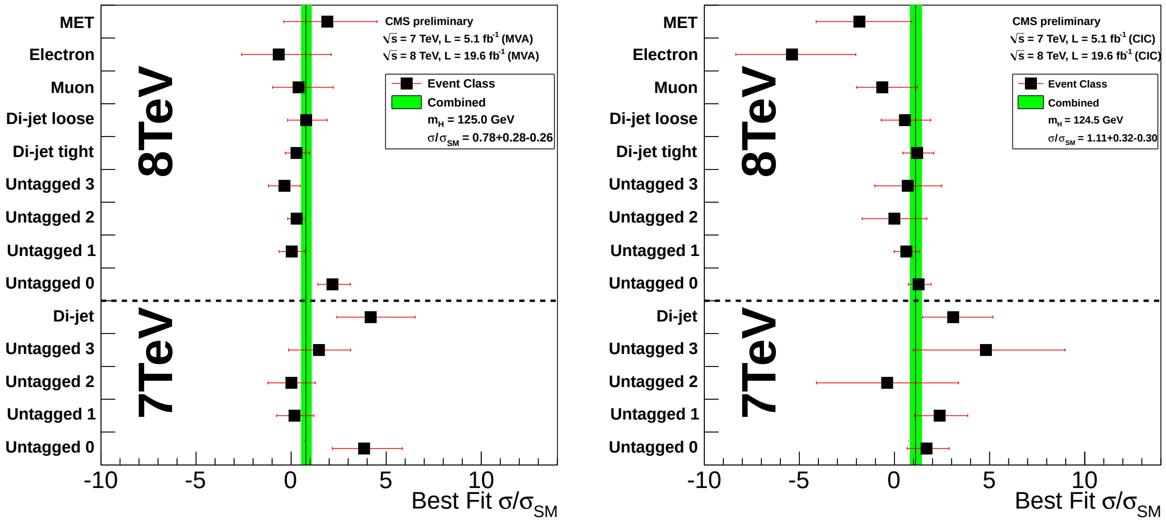
(a) Mass-fit-MVA. (b) Cut-based analysis.

Figure 5.18: Observed local p-values as a function of m_{H}. The results are shown for the MVA-based analysis analysis (left) and the cut-based analysis (right).



(a) Mass-fit-MVA. (b) Cut-based analysis.

Figure 5.19: The best fit signal strength relative to the SM Higgs boson cross section. The results are shown for the MVA-based analysis analysis (left) and the cut-based analysis (right).



(a) Mass-fit-MVA. (b) Cut-based analysis.

Figure 5.20: The combined fit to the fourteen classes (vertical line) and for the individual contributing classes (points) for the hypothesis of a SM Higgs boson mass of 125 GeV (left) for the MVA-based analysis and 124.5 GeV (right) for the cut-based. The band corresponds to ±1σ uncertainties on the overall value. The horizontal bars indicate ±1σ uncertainties on the values for individual classes.

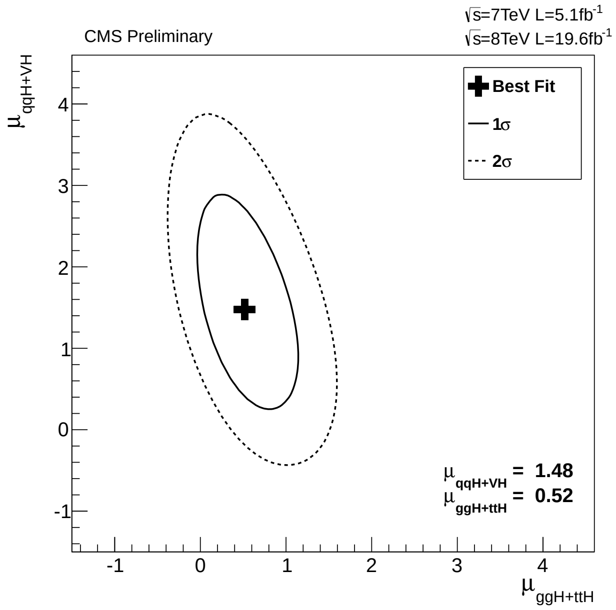


Figure 5.21: The 68% CL intervals for the signal strength modifiers associated with the gluon-gluon-fusion-plus-$t\bar{t}$H and for VBF-plus-VH production mechanisms, respectively. The cross indicates the best-fit value.

With the introduction of the exclusive categories in the analysis, the sensitivity of the single couplings and the contribution of each production mode has increased considerably.

Figure 5.21 shows the 68% and 95% CL contours (computed as the variations around the likelihood maximum) for the signal strength modifiers associated with the gluon-fusion-plus-$t\bar{t}$H ($\mu_{ggH+t\bar{t}H}$) and for VBF-plus-VH production mechanisms ($\mu_{q\bar{q}H+VH}$). The best fit values are found to be

$(\mu_{ggH+t\bar{t}H}, \mu_{q\bar{q}H+VH}) = (0.52, 1.48)$ for the mass-fit-MVA analysis.

Chapter 6

Higgs Mass measurement

In this Chapter the measurement of the mass of the Higgs boson in the di-photon decay channel is presented. The event selection and categorization have been discussed in the previous Chapter.

The Higgs mass is estimated by a maximum likelihood fit to the di-photon invariant mass spectrum. The likelihood has been defined in Eq. 5.5. The parameter of interest is the Higgs mass m_{H}. In the following the main sources of uncertainty will be discussed within their treatment as *nuisance parameters*.

Accurate photon energy scale, high energy resolution and the correct modelling of the signal *pdf* are the key ingredients of the mass measurement.

Unfortunately, at LHC there is not a pure source of a high energy photons to be used in order to calibrate the photon energy scale and resolution *in situ*. Nonetheless, the copious production of Z bosons decaying in the di-electron mode are exploited at best for the above mentioned purposes. The similarity in the interaction of electrons and photons with the calorimeter allows to use an electron sample to derive corrections for photons.

The aim of the corrections to the energy scale and resolution (described in Ch. 4) is to compensate the differences between the simulation and the data due to effects that are not yet simulated at the level of accuracy that is required for high precision measurement as the Higgs boson mass. The differences between data and simulation (discussed in Sec. 4.2.1) affect both electrons and photons. The electron categorization with the R_9 variable help in distinguishing low radiating electrons more similar to unconverted photons

($R_9 > 0.94$) and highly radiating electrons similar to converted photons ($R_9 < 0.94$). The differences between electrons and photons is then considerably reduced.

The uncertainties on the corrections are then mainly related to the intrinsic differences between the interaction of electrons and photons with the detector.

6.1 Uncertainty on the photon energy scale

6.1.1 Extrapolation from electrons to photons

Electromagnetic shower modelling

Electron and photon interaction with the material is fully simulated in CMS using the GEANT toolkit [73, 74].

The $\frac{E_{rec}}{E_{gen}}$ ratio has been checked as function of R_9 for electrons and photons in the two samples simulated with different GEANT version. The uncertainty associated to the electromagnetic shower modelling is given by the difference between electrons and photons in the two simulations, then the ratio

$$R_{ele}/R_{pho} = \frac{\frac{E^{ele}_{rec}}{E^{ele}_{gen}}}{\frac{E^{pho}_{rec}}{E^{pho}_{gen}}} \tag{6.1}$$

has been evaluated in the two R_9 categories used to distinguish between converted and unconverted photons in the analysis ($R_9 < 0.94$ and $R_9 > 0.94$). In the high R_9 category the maximum difference is 0.1% and for the low R_9 is 0.2%.

Tracker material uncertainty

The impact of the tracker material on the resolution has been discussed in Sec. 4 and the energy scale and resolution corrections (derived on electrons) aim to account for the difference between data and simulation. The residual difference between electrons and photons in the interaction with the material not accounted in the simulation is a further source of uncertainty in the $H \to \gamma\gamma$ analysis. Three different material scenarios have been implemented increasing the material in front of ECAL by 10%, 20% and 30%. Both

DY and $H \rightarrow \gamma\gamma$ (gluon fusion) samples have been simulated with the nominal material budget and the three modified material budget scenarios.

The ratio of Eq. 6.1 has been evaluated for two R_9 categories and in the four detector regions defined for the energy correction derivation.

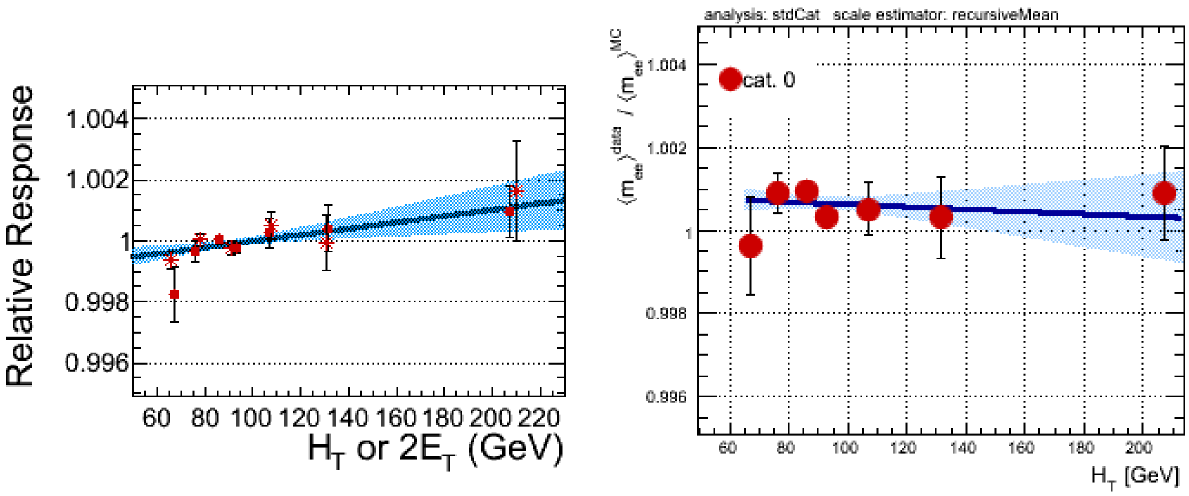
The difference between electrons and photons reaches up to 0.5% in the outer part of the barrel with 30% additional material. The systematic uncertainty is 0.25%.

6.1.2 Extrapolation from Z to H (125) energies

The average E_T of the electrons from Z decay is $\sim$ 45 GeV as shown in Figg. 4.6e and 4.6f, whilst the photons from the decay of the Higgs with mass 125 GeV is on average $\sim$ 65 GeV. ECAL response have been demonstrated to be linear at test beams with electrons with energy up to 200 GeV. Non linear behaviour in the energy scale can come from the interaction with the upstream material and the energy correction algorithm.

The uncertainty due to the extrapolation from the Z to the Higgs energies has been reduced with introduction of E_T dependent energy scale corrections described in Sec. 4.4.1. The residual uncertainty (after the E_T dependent scale corrections) has been checked with two methods, the first based on the iterative fit of the Z peak position, the second evaluating the ratio between the energy measured by the ECAL only and the momentum measured by the track curvature. Both measurements are performed in bins of the scalar sum of the electrons p_T: $H_T = |\vec{p_T}^{ele1}| + |\vec{p_T}^{ele2}|$.

In Fig. 6.1 the energy scale is shown as function of H_T for the first inclusive category with the cut-based selection. The energy scale is shown before the E_T dependent scale corrections derived in Chapter 4 (left) and after (right). The uncertainty, after the propagation to the $H \rightarrow \gamma\gamma$ photon categories has been reduced from 0.25% (before the introduction of E_T dependent energy scale corrections) to 0.1%. The uncertainty quoted for the most recent public result is 0.25%, when no E_T dependent energy corrections were applied.



(a) (b)

Figure 6.1: Electron energy scale before (left) and after (right) E_T dependent energy scale corrections as function of the scalar sum of the electron's p_T.

6.1.3 Summary of the systematic errors on the photon energy scale

In addition to the uncorrelated uncertainties on the individual photon energy scales shown in Table 6.1, systematic uncertainties of 0.25% to account for the imperfect modelling by the MC of electron/photon differences, and 0.25% to account for possible non-linearity in moving from m_{Z} to m_{H} ≈125GeV, are added. These two uncertainties, which together amount to 0.47%, are fully correlated between all the categories of the analysis and between the 2011 and 2012 data.

6.2 Propagation of energy scale and resolution uncertainties to the signal parametric model

6.3 Results

Figure 6.2 (left) shows the 2D 68% confidence level region for the signal strength modifier μ and the mass of the observed particle. Figure 6.2 (right) shows the scan of the negative-log-

Energy resolution ($\Delta\sigma/E_{MC}$)		
$R_9 > 0.94$ (low η, high η)	0.23%, 0.72%	0.93%, 0.36%
$R_9 < 0.94$ (low η, high η)	0.25%, 0.60%	0.33%, 0.54%
Energy scale ($(E_{data} - E_{MC})/E_{MC}$)		
$R_9 > 0.94$ (low η, high η)	0.20%, 0.71%	0.88%, 0.12%
$R_9 < 0.94$ (low η, high η)	0.20%, 0.51%	0.18%, 0.12%

Global energy scale	
Imperfect modelling of electron and photon difference in simulation	0.25%
Possible non-linearity extrapolating from Z-mass to $m_H = 125$ GeV	0.25%

Table 6.1: Summary table of systematic uncertainties affecting the mass measurement

likelihood as a function of the hypothesised mass, evaluated using statistical uncertainties only (blue dashed line), and using statistical plus systematic uncertainties (black line). The mass of the observed boson is measured to be $125.4 \pm 0.5(stat.) \pm 0.6(syst.)$ GeV.

The mass measured in the $H \rightarrow \gamma\gamma$ channel alone has been compared and combined with the corresponding measurement in the $H \rightarrow ZZ^* \rightarrow 4\ \ell$decay channel.

The results (shown in Fig. 6.3) are compatible within the uncertainties.

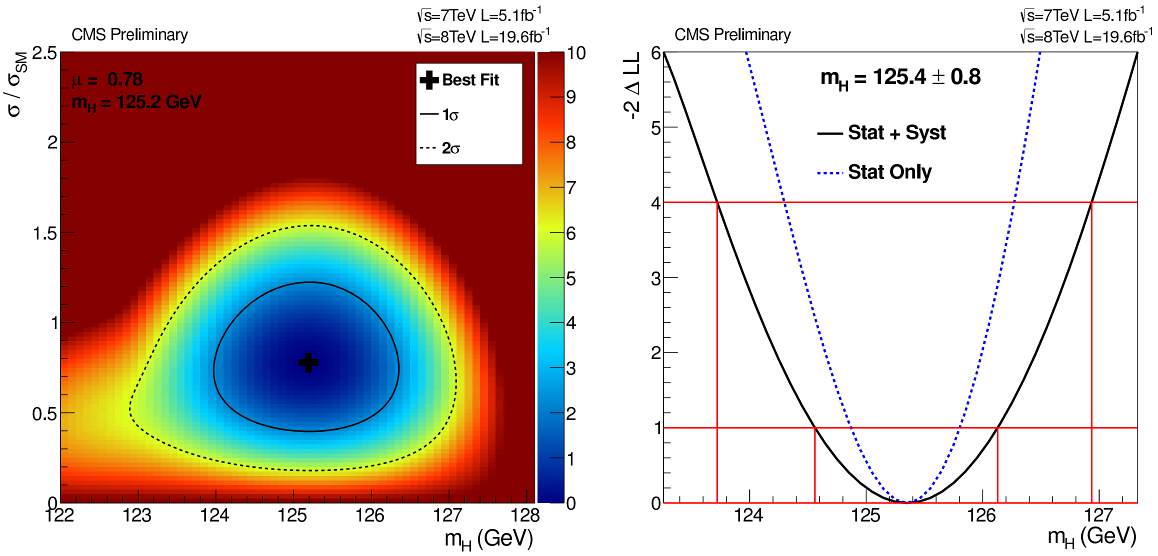


Figure 6.2: (left) The 2D 68% confidence level region for the signal strength modifier μ and the mass of the observed particle. (right) The scan of the negative-log-likelihood as a function of the hypothesised mass, evaluated using statistical uncertainties only (blue dashed line), and using statistical plus systematic uncertainties (black line).

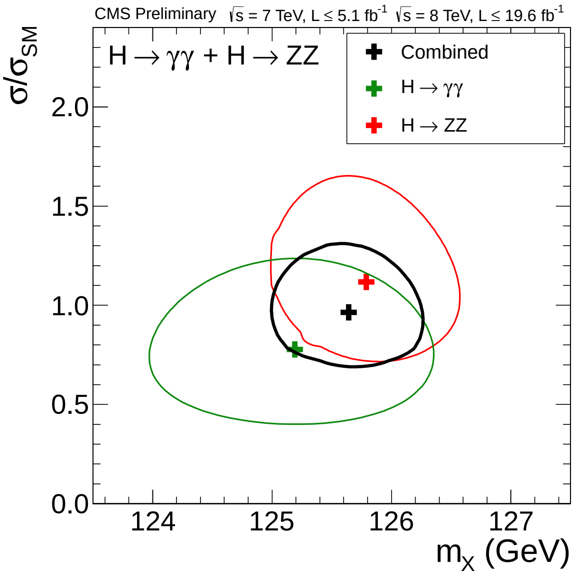


Figure 6.3: Higgs mass measurement in the $H \rightarrow ZZ^* \rightarrow 4\ \ell$and in the $H \rightarrow \gamma\gamma$ decay channels. Best fit values are shown by the point and 1 standard deviation uncertainty by the solid line contours. Both measurement are compatible within uncertainties.

Conclusions

The discovery of a new boson with mass $\sim$ 125 GeV has been announced by the A Toroidal LHC ApparatuS experiment (ATLAS) and the Compact Muon Solenoid (CMS) collaborations[75, 76] on 4 July 2012.

In this thesis the measurement in the two photon final state of the mass of the Higgs boson has been presented. The analysis is based on the data collected by the CMS experiment in proton-proton collisions at $\sqrt{s} = 7$ TeV (5.1fb^{-1}) and at $\sqrt{s} = 8$ TeV (19.6fb^{-1}) center-of-mass energy.

The central role of the Electromagnetic Calorimeter (ECAL) in the $H \rightarrow \gamma\gamma$ search has been described with attention to the *in situ* calibration and energy resolution estimation.

After the discovery of the new boson, the analysis have moved to the measurement of the boson properties. The thesis is focused to the mass measurement and its main systematic uncertainties. The measurement and the results on the signal strength and couplings with fermions and bosons are also described.

The statistical uncertainty on measured Higgs' mass in the di-photon decay channel is now compatible with the systematic error. The main contributions to the systematic uncertainties are coming from the energy scale corrections and the additional smearing to the energy in the simulation needed to have a proper Higgs signal model with the effective resolution reached in data. I've developed the techniques used in the analysis to derive the corrections from $Z \rightarrow e^+e^-$ events. I've further improved the methods and defined for the first time energy scale and smearings as function of the electron/photon energy for the Higgs analysis. Both the systematic uncertainty due to the correction derivation and due to the extrapolation from Z to Higgs's energies has been consequently reduced.

The expected sensitivity on the limit on the production cross-section times branching fraction, at 95% confidence level, is between 0.48 and 0.76 times the standard model prediction in the mass range 110-150 GeV.

For the MultiVariate (MVA)-based analysis, the local significance of the excess is 3.2 σ with a corresponding expected value of 4.2 σ and the best fit signal strength is $0.78^{+0.28}_{-0.26}$ times the Standard Model Higgs boson cross section. For the cut-based analysis, the local significance of the excess is 3.9 σ with a corresponding expected value of 3.5 σ and a best fit signal strength of $1.1^{+0.32}_{-0.30}$ times the Standard Model Higgs boson cross section. The two results on the complete 2011 and 2012 datasets are found to be compatible at the 1.5σ level after taking into account the correlations between the two analysis.

The best fit value for the signal strength modifiers associated with the gluon-fusion-plus-$t\bar{t}$H and for VBF-plus-VH production mechanisms are found to be $(\mu_{ggH+t\bar{t}H}, \mu_{q\bar{q}H+VH}) = (0.52, 1.48)$ for the mass-fit-MVA analysis.

The mass of the observed Higgs boson is measured to be $125.4 \pm 0.5(stat.) \pm 0.6(syst.)$ GeV.

Bibliography

[1] S. L. Glashow, "Partial Symmetries of Weak Interactions," *Nucl. Phys.*, vol. 22, pp. 579–588, 1961. 1.1, 1.1.1, 1.1.1

[2] S. Weinberg, "A Model of Leptons," *Phys. Rev. Lett.*, vol. 19, p. 1264, 1967. 1.1, 1.1.1, 1.1.1

[3] A. Salam, *Elementary Paricle Theory*, p. 367. Almquist and Wiksells, Stockholm, 1968. 1.1, 1.1.1, 1.1.1

[4] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons," *Phys.Rev.Lett.*, vol. 13, pp. 321–323, 1964. 1.1, 1.1.2

[5] P. W. Higgs, "Broken symmetries and the masses of gauge bosons," *Phys. Rev. Lett.*, vol. 13, pp. 508–509, Oct 1964. 1.1, 1.1.2

[6] M. Gell-Mann, "A schematic model of baryons and mesons," *Phys. Lett.*, vol. 8, pp. 214–215, 1964. 1.1.1

[7] M. Han and Y. Nambu, "Three Triplet Model with Double SU(3) Symmetry," *Phys.Rev.*, vol. 139, pp. B1006–B1010, 1965. 1.1.1

[8] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, "Advantages of the Color Octet Gluon Picture," *Phys.Lett.*, vol. B47, pp. 365–368, 1973. 1.1.1

[9] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global conservation laws and massless particles," *Phys. Rev. Lett.*, vol. 13, pp. 585–587, Nov 1964. 1.1.2

[10] M. Spira, A. Djouadi, D. Graudenz, and P. Zerwas, "Higgs boson production at the LHC," *Nucl.Phys.*, vol. B453, pp. 17–82, 1995. 1.1.3

[11] S. Dittmaier *et al.*, “Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables,” 2011. 1.1.3

[12] A. David *et al.*, “LHC HXSWG interim recommendations to explore the coupling structure of a Higgs-like particle,” 2012. 1.1.5

[13] M. Baak, M. Goebel, J. Haller, A. Hoecker, D. Kennedy, *et al.*, “The Electroweak Fit of the Standard Model after the Discovery of a New Boson at the LHC,” *Eur.Phys.J.*, vol. C72, p. 2205, 2012. 1.1.5

[14] G. Degrassi, S. Di Vita, J. Elias-Miro, J. R. Espinosa, G. F. Giudice, *et al.*, “Higgs mass and vacuum stability in the Standard Model at NNLO,” *JHEP*, vol. 1208, p. 098, 2012. 1.1.5

[15] L. J. Dixon and Y. Li, “Bounding the higgs boson width through interferometry,” *Phys. Rev. Lett.*, vol. 111, p. 111802, Sep 2013. 1.1.5

[16] Brüning, Oliver Sim and Collier, Paul and Lebrun, P and Myers, Stephen and Ostojic, Ranko and Poole, John and Proudlock, Paul, *LHC Design Report.* Geneva: CERN, 2004. 1.2

[17] CMS Collaboration, “The CMS experiment at the CERN LHC,” *Journal of Instrumentation*, vol. 3, 2008. 2.1

[18] CMS Collaboration, *The CMS magnet project: Technical Design Report.* Technical Design Report CMS, Geneva: CERN, 1997. 2.1.2

[19] CMS Collaboration, *The CMS tracker system project: Technical Design Report.* Technical Design Report CMS, Geneva: CERN, 1997. 2.1.3

[20] CMS Collaboration, *The CMS tracker: addendum to the Technical Design Report.* Technical Design Report CMS, Geneva: CERN, 2000. 2.1.3

[21] CMS Collaboration, *The CMS hadron calorimeter project: Technical Design Report.* Technical Design Report CMS, Geneva: CERN, 1997. 2.1.4

[22] CMS ECAL-HCAL Collaboration, "The CMS Barrel Calorimeter Response to Particle Beams from 2 to 350 GeV/c." CMS NOTE 2008/034, June 2008. 2.1.4

[23] CMS Collaboration, *The CMS muon project: Technical Design Report.* Technical Design Report CMS, Geneva: CERN, 1997. 2.1.5

[24] CMS Collaboration, *The CMS ECAL: Technical Design Report.* No. CERN-LHCC-1997-033; CMS-TDR-4 in Technical Design Report CMS, Geneva: CERN, December 1997. 2.2.1, 3.2.1

[25] P. Bloch, R. Brown, P. Lecoq, and H. Rykaczewski, *Changes to CMS ECAL electronics: addendum to the Technical Design Report.* Technical Design Report CMS, Geneva: CERN, 2002. 2.2.1

[26] "The cms high level trigger," *The European Physical Journal C - Particles and Fields*, vol. 46, no. 3, pp. 605–667, 2006. 2.3

[27] D. A. Petyt and for the CMS Collaboration, "Anomalous APD signals in the CMS Electromagnetic Calorimeter," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 695, no. 0, pp. 293 – 295, 2012. New Developments in Photodetection NDIP11. 2.3.1

[28] Paganini, Pascal and for the CMS collaboration, "CMS Electromagnetic Trigger commissioning and first operation experiences," *Journal of Physics: Conference Series*, vol. 160, no. 1, p. 012062, 2009. 2.3.1

[29] A. Zabi and for the CMS Collaboration, "Triggering on electrons and photons with CMS," in *European Physical Journal Web of Conferences*, vol. 28 of *European Physical Journal Web of Conferences*, p. 12036, June 2012. 2.3.1

[30] Sjostrand, T. and Mrenna, S. and Skands, P., "PYTHIA 6.4: Physics and manual," 2006. 2.4, 5.3

[31] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1," *Comput.Phys.Commun.*, vol. 178, pp. 852–867, 2008. 2.4

[32] T. Stelzer and W. Long, “Automatic generation of tree level helicity amplitudes,” *Comput.Phys.Commun.*, vol. 81, pp. 357–371, 1994. 2.4

[33] F. Maltoni and T. Stelzer, “MadEvent: Automatic event generation with MadGraph,” *JHEP*, vol. 0302, p. 027, 2003. 2.4

[34] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX,” *JHEP*, vol. 1006, p. 043, 2010. 2.4

[35] T. Gleisberg, S. Hoeche, F. Krauss, A. Schalicke, S. Schumann, *et al.*, “SHERPA 1. alpha: A Proof of concept version,” *JHEP*, vol. 0402, p. 056, 2004. 2.4

[36] A. Buckley, J. Butterworth, S. Gieseke, D. Grellscheid, S. Hoche, *et al.*, “General-purpose event generators for LHC physics,” *Phys.Rept.*, vol. 504, pp. 145–233, 2011. 2.4

[37] P. Adzic *et al.*, “Reconstruction of the signal amplitude of the CMS electromagnetic calorimeter,” *The European Physical Journal C - Particles and Fields*, vol. 46, no. 1, pp. 23–35, 2006. 3.1

[38] CMS Collaboration, “Radiation hardness qualification of pbwo$_4$ scintillation crystals for the cms electromagnetic calorimeter,” *Journal of Instrumentation*, vol. 5, no. 03, p. P03010, 2010. 3.2.1

[39] D. E. Leslie and for the CMS Collaboration, “The effect of pulse rate on VPT response and the use of an LED light to improve stability,” Tech. Rep. CMS-CR-2009-284, CERN, Geneva, Sep 2009. 3.2.1, 3.2.1

[40] P. Baillon *et al.*, “Performance of the Cooling System of Ecal CMS,” in *Astroparticle, Particle and Space Physics, Detectors and Medical Physics Applications*, pp. 203–207, July 2004. 3.2.1, 4.2.1

[41] A. Bartoloni *et al.*, “High Voltage System for the CMS Electromagnetic Calorimeter,” Tech. Rep. CMS-NOTE-2007-009, CERN, Geneva, Apr 2007. 3.2.1, 4.2.1

[42] CMS Collaboration, “Performance and operation of the CMS electromagnetic calorimeter,” *Journal of Instrumentation*, vol. 5, no. 03, p. T03010, 2010. 3.2.1, 4.2.1

[43] M. Anfreville *et al.*, “Laser monitoring system for the CMS lead tungstate crystal calorimeter,” *Nuclear Instruments and Methods in Physics Research A*, vol. 594, pp. 292–320, Sept. 2008. 3.2.1

[44] A. Van Lysebetten and P. Verrecchia, “Performance and measurements of the light monitoring system for cms-ecal from 2002 test beam data,” CMS Rapid Note CMS-RN-2004-001, 2004. 3.2.1

[45] A. Ghezzi *et al.*, “Analysis of the response evolution of the cms electromagnetic calorimeter under electron and pion irradiation,” CMS Note CMS-NOTE-2006-038, 2006. 3.2.1

[46] CMS Collaboration, “Intercalibration of the barrel electromagnetic calorimeter of the CMS experiment at start-up,” *JINST*, vol. 3, p. P10007, 2008. 3.2.2, 4.1

[47] CMS Collaboration, “Electromagnetic calorimeter calibration with 7 TeV data.” CMS PAS EGM-10-003, 2010. 3.2.2

[48] CMS Collaboration, *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software.* Technical Design Report CMS, Geneva: CERN, 2006. 3.3

[49] M. Anderson, A. Askew, A.F. Barfuss, D. Evans, F. Ferri, K. Kaadze, Y. Maravin, P. Meridiani, C. Seez, “Review of clustering algorithms and energy corrections in ecal.” IN-2010/008, 2010. 3.3

[50] Frühwirth, R., “A gaussian-mixture approximation of the bethe-heitler model ofelectron energy loss by bremsstrahlung,” *Computer Physics Communications*, vol. 154, no. 2, pp. 131 – 142, 2003. 3.5.1, 3.6

[51] W. Adam, S. Baffioni, F. Beaudette, D. Benedetti, C. Broutin, D. Chamont, C. Charlot, E. DiMarco, D. Futyan, S. Harper, D. Lelas, A. Martelli, P. Meridiani, M. Pioppi, I. Puljak, D. Sabes, R. Salerno, M. Sani, C. Seez, Y. Sirois, P. Vanlaer, and D. Wardrope, "Electron Reconstruction in CMS." CMS AN-2009/164, January 2010. 3.6

[52] F. Beaudette, D. Benedetti, P. Janot, and M. Pioppi, "Electron Reconstruction within the Particle Flow Algorithm." CMS AN-2010/034, March 2010. 3.6

[53] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, I. Puljak, C. Rovelli, R. Salerno, and Y. Sirois, "Electron reconstruction in cms," *The European Physical Journal C - Particles and Fields*, vol. 49, pp. 1099–1116, 2007. 10.1140/epjc/s10052-006-0175-5. 3.6

[54] Frühwirth, R., "Application of kalman filtering to track and vertex fitting," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 262, no. 2-3, pp. 444 – 450, 1987. 3.6

[55] CMS Collaboration, "Energy resolution of the barrel of the CMS Electromagnetic Calorimeter," *Journal of Instrumentation*, vol. 2, no. 04, p. P04004, 2007. 4.1

[56] K. Nakamura and P. D. Group, "Review of particle physics," *Journal of Physics G: Nuclear and Particle Physics*, vol. 37, no. 7A, p. 075021, 2010. 4.2.2, 4.2.8

[57] CMS Collaboration, "Commissioning of the particle-flow event reconstruction with the first lhc collisions recorded in the cms detector," *CDS Record*, vol. 1247373, 2010. 4.2.4

[58] W. Erdmann, "Offline primary vertex reconstruction with deterministic annealing clustering," *CMS Internal note*, vol. 2011/014, 2011. 4.2.7, 5.4

[59] J. E. Gaiser, *Charmonium Spectroscopy from Radiative Decays of the J/Psi and Psi-Prime.* PhD thesis, Stanford University, 1982. 4.2.8

[60] D. Antreasyan *et al. CB Note 321*, 1983. 4.2.8

[61] T. Skwarnicki 1986. DESY F31-86-02. 4.2.8

[62] S. Alioli, P. Nason, C. Oleari, and E. Re, "NLO Higgs boson production via gluon fusion matched with shower in POWHEG," *JHEP*, vol. 04, p. 002, 2009. 5.3

[63] P. Nason and C. Oleari, "NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG," *JHEP*, vol. 02, p. 037, 2010. 5.3

[64] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.), "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables," *CERN-2011-002*, CERN, Geneva, 2011. 5.3, 5.10

[65] The $H \rightarrow \gamma\gamma$ working group, "Search for a standard model higgs boson decaying into two photons employing multivariate methods," *CMS Analysis Note*, vol. CMS-AN-2012/048, 2012. 5.4

[66] A. Ballestrero, G. Bevilacqua, and E. Maina, "A complete parton level analysis of boson-boson scattering and ElectroWeak Symmetry Breaking in lv + four jets production at the LHC," *JHEP*, vol. 05, p. 015, 2009. 5.7.3

[67] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas," *Phys.Lett.*, vol. B659, pp. 119–126, 2008. 5.7.3

[68] S. Chatrchyan *et al.*, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS," *JINST*, vol. 6, p. P11002, 2011. * Temporary entry *. 5.7.3

[69] D. L. Rainwater, R. Szalapski, and D. Zeppenfeld, "Probing color singlet exchange in Z + two jet events at the CERN LHC," *Phys. Rev.*, vol. D54, pp. 6680–6689, 1996. 5.7.3

[70] "Procedure for the LHC Higgs boson search combination in Summer 2011," Tech. Rep. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug 2011. 5.8.1

[71] CMS Collaboration, "Cms luminosity based on pixel cluster counting - summer 2012 update," *CDS Record*, vol. 1482193, 2012. 5.10

[72] C. Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Phys. Lett. B*, vol. 716, pp. 30–61, 2012. 5.11

[73] S. Agostinelli *et al.*, "G4–a simulation toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250 – 303, 2003. 6.1.1

[74] J. Allison *et al.*, "Geant4 developments and applications," *Nuclear Science, IEEE Transactions on*, vol. 53, pp. 270 –278, feb. 2006. 6.1.1

[75] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Physics Letters B*, vol. 716, no. 1, pp. 1 – 29, 2012. 6

[76] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Physics Letters B*, vol. 716, no. 1, pp. 30 – 61, 2012. 6

List of acronyms

ALICE A Large Ion Collider Experiment

APD Avalanche Photo-Diode

ATLAS A Toroidal LHC ApparatuS experiment

BDT Boosted Decision Tree

BW Breit-Wigner

CB Crystal Ball

CMS Compact Muon Solenoid

DY Drell-Yan

EB ECAL Barrel

ECAL Electromagnetic Calorimeter

EE ECAL Endcap

ES ECAL Endcap preShower detector

EW electroweak

GSF Gaussian Sum Filter

HCAL hadron calorimeter

HLT Higher Level Trigger

KF Kalman Filter

L1 Level-1 Trigger

LEP Large Electron Positron collider

LHC Large Hadron Collider

LHCb Large Hadron Collider beauty experiment

LO Leading Order

NLO Next to Leading Order

MC Monte Carlo

MVA MultiVariate

NLL $-\ln \mathcal{L}$

OOT Out-of-Time

PDG Particle Data Group

pdf Probability Density Function

PF Particle Flow

QCD Quantum Chromo-Dynamics

$\mathrm{R_M}$ Molière Radius

SC Super Cluster

SM Standard Model

SMC Shower Monte Carlo

VPT Vacuum Photoriode