SAPIENZA
UNIVERSITÀ DI ROMA

Measurement of the Standard Model Higgs Boson Couplings by Means of an Exclusive Analysis of its Diphoton Decay Channel

Scuola di dottorato in Scienze Astronomiche, Chimiche, Fisiche e Matematiche "Vito Volterra"

Dottorato di Ricerca in Fisica – XXV Ciclo

Candidate

Marco Grassi

ID number 1046750

Thesis Advisors

Prof. Shahram Rahatlou

Dr. Daniele Del Re

A thesis submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics

October 2012

Thesis defended on January 24, 2012
in front of a Board of Examiners composed by:

Prof. Carlo Dionisi (chairman)
Prof. Mauro De Palma
Prof. Stefano Liberati
Prof. Giorgio Chiarelli

Measurement of the Standard Model Higgs Boson Couplings by Means of an Exclusive Analysis of its Diphoton Decay Channel

Ph.D. thesis. Sapienza – University of Rome

© 2012 Marco Grassi.
Released under a Creative Commons *Attribution-NonCommercial-NoDerivs* 3.0 Unported License

Version: January 23, 2013

Author's email: Marco.Grassi@cern.ch

«*We should perhaps finish our paper with an apology and a caution. We apologize to experimentalists for having no idea what is the mass of the Higgs boson,* [. . .] *and for not being sure of its couplings to other particles, except that they are probably all very small. For these reasons, we do not want to encourage big experimental searches for the Higgs boson, but we do feel that people doing experiments vulnerable to the Higgs boson should know how it may turn up.*»

J.R. Ellis, M.K. Gaillard, D.V Nanopulos
Nucl.Phys. **B**106 : 292 (1976)

Contents

Introduction

Roughly four years ago, at the time of joining my last class of quantum field theory, the Higgs boson was still a chimera; a theoretical particle whose field provided an elegant way to break the electroweak symmetry. Many scientists were betting against its existence[1], and the long-awaited Large Hadron Collider was seen as the only machine in the world capable of providing sufficient energy and integrated luminosity to fully scan the relevant mass range with suitable statistics. When the *quench incident*[2] occurred on September 19$^{\text{th}}$ 2008, no one would ever believed that the discovery could happen before a decade. It took one year to get the beams circulating again in the LHC, and the design center-of-mass energy had to be halved ($\sqrt{s} = 7$ TeV) for security reasons. Nevertheless, at the end of July 2012 (after only two and a half years of data taking) both the Atlas and the CMS collaborations published on Physics Letters B the observation of a new boson having a mass of 125 GeV [1, 2], with properties very similar to those of the Standard Model (SM) Higgs boson. Figures 1 and 2 (quoted from [1]) show the observed excess in the $\gamma\gamma$ and $ZZ(\to 4\ell)$ invariant mass, obtained analyzing 10.4 fb^{-1} of data collected by the CMS experiment during the 2011-12 data-taking.

The reasons why physicists have been hunting for the Higgs boson during the last 40 years are many. Its exchange is required to regulate the energy behavior of the scattering amplitudes involving longitudinal vector bosons, hence ensuring the perturbative unitarity of the SM in the ultraviolet regime. If its properties will be those predicted by the SM, it will be the first example of an elementary scalar field, a new form of matter in addition to fermions and to vector gauge bosons. It will also imply the existence of fundamental forces (its self-interaction and the Yukawa interactions to fermions) not of gauge type. All these properties, in fact, are strictly related to the possibility for the SM to remain weakly coupled up to extremely large energies, possibly of the order of the Planck scale. This too, by itself, would be a profoundly new phenomenon in Nature: physics at the fundamental level would be described by the same mathematical theory over $\sim$15 orders of magnitude in energy without any new dynamics appearing at intermediate scales. To realize this paradigm, the couplings of the Higgs boson must be finely tuned to specific values which depend uniquely on its mass. Any deviation from these values would either imply the existence of additional Higgs bosons that take part in the perturbative unitarization of the scattering amplitudes, like in the case of Supersymmetry, or signal the existence of a new energy threshold at which the theory becomes strongly coupled. In this second case the Higgs boson would emerge as a composite state of a new fundamental force, possibly of gauge type.

A precise measurement of the Higgs couplings gives the unique opportunity to test

[1] E.g. "Higgs boson discovery has lost me $100", Stephen Hawking interview at BBC, July 4$^{\text{th}}$ 2012.

[2] See "Interim Summary Report on the Analysis of the 19 September 2008 Incident at the LHC", CERN, October 15$^{\text{th}}$ 2008, EDMS 973073.

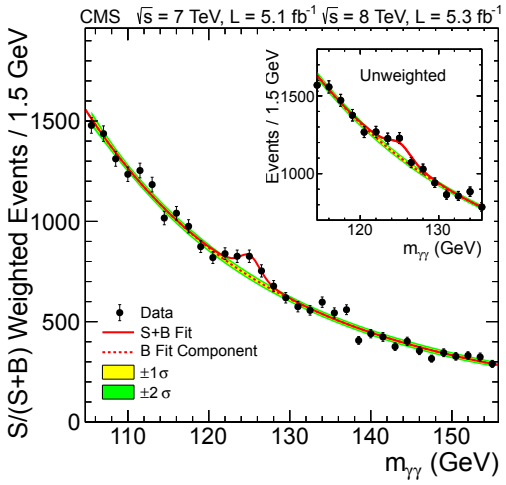


Figure 1. The diphoton invariant mass distribution with each event weighted by the S/(S+B) value of its category. The lines represent the fitted background and signal, and the colored bands represent the ± 1 and ± 2 standard deviation uncertainties in the background estimate. The inset shows the central part of the unweighted invariant mass distribution.

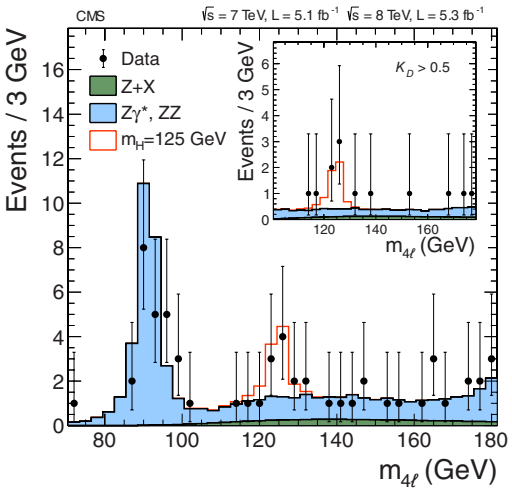


Figure 2. Distribution of the four-lepton invariant mass for the $ZZ \to 4\ell$ analysis. The points represent the data, the filled histograms represent the background, and the open histogram shows the signal expectation for a Higgs boson of mass $m_H = 125$ GeV, added to the background expectation. The inset shows the $m_{4\ell}$ distribution after selection of events with values of the kinematic discriminant (K_D) above 0.5.

the SM paradigm and to probe the dynamics behind the electroweak symmetry breaking (EWSB).

This thesis describes the exclusive analysis of the Higgs boson in its diphoton final state, and it aims at interpreting the yields of the different Higgs production mechanisms in terms of the couplings of the Higgs to fermions and to bosons. The analysis is based on 5.3 fb^{-1} of data collected during 2012 by the CMS experiment at the LHC, where $\sqrt{s} = 8$ TeV. Data collected during 2011 at $\sqrt{s} = 7$ TeV is not considered here, because the 2011 and 2012 datasets cannot be mixed in the same analysis. The primary reasons are two: 1) the difference in the center-of-mass energy results in different signal and background cross sections; 2) the increase in the instantaneous luminosity — hence in the number of simultaneous collisions happening at each bunch crossing — spoils the energy reconstruction of the clusterized objects (photons, electrons, jets), and the quantities involved in the selection criteria (isolation, photon cluster shape). The collaboration adopted the strategy of performing two separate analyses, and then to combine the two results statistically as if they were two independent experiments.

The exclusive analysis described in this thesis relies on the inclusive $H \to \gamma\gamma$ analysis in terms of photon reconstruction and identification, and vertex reconstruction. The exclusive analysis considers the same events selected by the inclusive analysis, but they are divided in several categories according to the presence of tagging objects in the final state (in addition

to the diphoton pair). The objects are: 1) two forward jets, 2) at least one isolated lepton, 3) large missing transverse energy (E_T^{miss}). When no such object is detected, the event is considered to be untagged, and it assigned to the inclusive analysis.

The overall $H \to \gamma\gamma$ analysis is the outcome of the work of many people. For the sake of completeness, the thesis describes the whole analysis. However, it's probably worth to stress here that the aspects of the analysis with a strong personal contribution are the study of the dijet-tagged events, and E_T^{miss}-tagged events, and the interpretation of the signal yields in the exclusive event categories in terms of the Higgs couplings.

The structure of the thesis is the following: a short summary of the Higgs mechanism in the SM paradigm is given in chapter 1. In chapter 2, the LHC and the CMS experiment are reviewed, with more emphasis on the subdetectors involved in the measurement described in this thesis. Chapter 3 is devoted to the reconstruction of the physics objects. The main topic is the photon reconstruction, and both the method and the performance are described. Jets, leptons and E_T^{miss} are also reviewed here. Chapter 4 describes the inclusive analysis of diphoton events. Data samples, triggers, vertex identification, and photon selection are described here, and hold also for the exclusive analysis. The signal and background models used to fit the invariant mass spectrum are also introduced in this chapter. Chapter 5 presents the different final-state topologies of the exclusive analysis. There is a section for each exclusive category, describing the selection criteria, the fit used to extract the signal and the systematic uncertainties associated to the tagging object. In chapter 3 the signal yields are used to measure the Higgs couplings. Conclusions are drawn after chapter 5. Appendix A reports a Montecarlo study performed to check the sensitivity of the $H \to \gamma\gamma$ decay channel alone in measuring the Higgs couplings.

Chapter 1

The Standard Model Higgs Boson

Understanding the mechanism that breaks the electroweak symmetry and generates the masses of the known elementary particles is one of the most fundamental problems in particle physics. The Higgs mechanism [3–8] provides a general framework to explain the observed masses of the $W^{\pm}$ and Z gauge bosons by means of charged and neutral Goldstone bosons that are manifested as the longitudinal components of the gauge bosons. These Goldstone bosons are generated by the underlying dynamics of electroweak symmetry breaking (EWSB). However, the fundamental dynamics of the electroweak symmetry breaking are unknown. There are two main classes of theories proposed in the literature, those with weakly coupled dynamics — such as in the Standard Model (SM) [9–11] — and those with strongly coupled dynamics; here we are interested only in the former, that is summarized below.

In the SM, the electroweak interactions are described by a gauge field theory based on the $SU(2)_L \times U(1)_Y$ symmetry group. The Higgs mechanism posits a self-interacting complex doublet of scalar fields, and renormalizable interactions are arranged such that the neutral component of the scalar doublet acquires a vacuum expectation value $v \sim 246$ GeV, which sets the scale of EWSB. Three massless Goldstone bosons are generated, which are absorbed to give masses to the $W^{\pm}$ and Z gauge bosons. The remaining component of the complex doublet becomes the Higgs boson —— a new fundamental scalar particle. The masses of all fermions are also a consequence of EWSB since the Higgs doublet is postulated to couple to the fermions through Yukawa interactions.

The validity of the SM as an effective theory describing physics up to the Planck scale is questionable, however, because of the following "naturalness" argument. All fermion masses and dimensionless couplings are logarithmically sensitive to the scale Λ at which new physics becomes relevant. In contrast, scalar squared masses are quadratically sensitive to Λ. Thus, the observable SM Higgs mass has the following form:

$$m_H^2 = m_{H0}^2 + \frac{k g^2 \Lambda^2}{16\pi^2} \ ,$$

where m_{H0} is a fundamental parameter of the theory. The second term is a one-loop correction in which g is an electroweak coupling and k is a constant, presumably of $\mathcal{O}(1)$, that is calculable within the low-energy effective theory. The two contributions arise from independent sources and one would not expect that the observable Higgs boson mass is significantly smaller than either of the two terms. Hence, if the scale of new physics Λ is

much larger than the electroweak scale, unnatural cancellations must occur to remove the quadratic dependence of the Higgs boson mass on this large energy scale and to give a Higgs boson mass of order of the electroweak scale, as required from unitarity constraints [12, 13], and as preferred by precision measurements of electroweak observables [14].

As already stated, the latest updates from the LHC experiments recently found a SM-like Higgs boson having $m_H \sim 125$ GeV [1, 2], in excellent agreement with the indirect predictions from electroweak precision data. Thus, the SM is expected to be embedded in a more fundamental theory which will stabilize the hierarchy between the electroweak scale and the Planck scale in a natural way. A theory of that type would usually predict the onset of new physics at scales of the order of, or just above, the electroweak scale. Theorists strive to construct models of new physics that keep the successful features of the SM while curing its shortcomings, such as the absence of a dark matter candidate or a detailed explanation of the observed baryon asymmetry of the universe. In the weakly-coupled approach to electroweak symmetry breaking, supersymmetric (SUSY) extensions of the SM provide a possible explanation for the stability of the electroweak energy scale in the presence of quantum corrections [15–17]. These theories predict at least five Higgs particles [18]. The properties of the lightest Higgs scalar often resemble those of the SM Higgs boson, with a mass that is predicted to be less than 135 GeV [17] in the simplest supersymmetric model. Additional neutral and charged Higgs bosons are also predicted. Moreover, low-energy supersymmetry with a supersymmetry breaking scale of order 1 TeV allows for grand unification of the electromagnetic, weak and strong gauge interactions in a consistent way, strongly supported by the prediction of the electroweak mixing angle at low energy scales, with an accuracy at the percent level [19, 20].

Prior to 1989, when the e^+e^- collider LEP at CERN came into operation, searches were sensitive only to Higgs bosons with masses of a few GeVand below [21]. In the LEP 1 phase, the collider operated at center-of-mass energies close to m_Z. During the LEP 2 phase, the energy was increased in steps, reaching 209 GeVin the year 2000 before the final shutdown. The combined data of the four LEP experiments, ALEPH, DELPHI, L3, and OPAL, were sensitive to neutral Higgs bosons with masses up to about 115 GeVand to charged Higgs bosons with masses up to about 90 GeV [22, 23]. The search for the Higgs boson continued at the Tevatron $p\bar{p}$ collider, which operated at a center-of-mass energy of 1.96 TeV until its shutdown in the Fall of 2011. The two experiments, CDF and DØ, each collected approximately 10 fb^{-1} of data with the capability to probe a SM Higgs boson mass in the 90 – 185 GeVrange. These searches are described in more details in section 1.3.2.

The ultimate searches for Higgs bosons have been performed at the LHC over the last two years, where both experiments ATLAS and CMS achieved a much better sensitivity than the Tevatron searches. The results were presented in a seminar held on July 4$^{\text{th}}$ at CERN, where both experiments ATLAS and CMS reported a $5\,\sigma$ excess, predominantly in the $\gamma\gamma$ and ZZ modes, which are compatible with a SM-like Higgs boson with a mass near 125 GeV [24, 25].

Despite the observation of this new resonance, we are still far from solving the EWSB puzzle. Now that the presence of a signal has been established, it becomes mandatory to understand the precise nature of such a particle by scrutinizing the coupling strengths in the different production and decay channels, especially investigating the Higgs bosons produced via vector boson fusion and in association with a W or a Z boson. Such precision measurements will be crucial to completely understand the mechanism of the electroweak symmetry breaking.

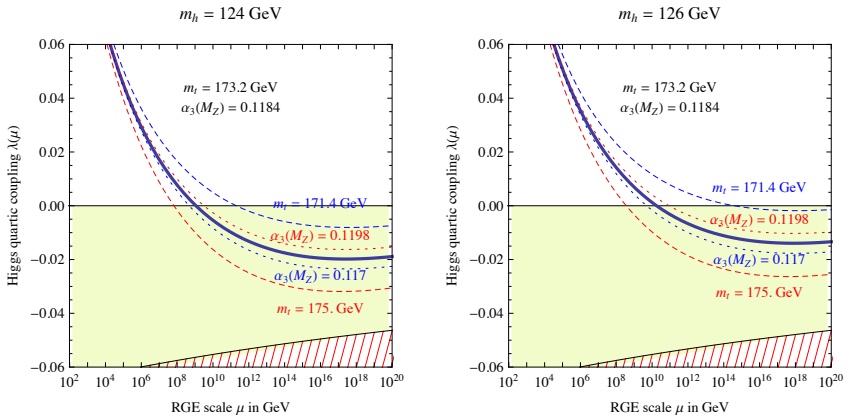


Figure 1.1. Renormalization group evolution of the Higgs self coupling λ, for $m_H = 124$ GeV (left) and $m_H = 126$ GeV(right), for the central values of m_t and α_s (solid curves), as well as for variations of m_t (dashed curves) and α_s (dotted curves). For negative values of λ, the lifetime of the SM vacuum due to quantum tunneling at zero temperature is longer than the age of the universe as long as λ remains above the region shaded in red [30].

1.1 Higgs Boson Mass and Couplings

In the SM, the Higgs boson mass is given by $m_H = \sqrt{\lambda/2}$, where λ is the Higgs self-coupling parameter and v is the vacuum expectation value of the Higgs field, $v = (\sqrt{2}\,G_F) \sim 246\,\text{GeV}$, fixed by the Fermi coupling G_F, which is determined with a precision of 0.6 ppm from muon decay measurements [26]. Since λ is presently unknown, the value of the SM Higgs boson mass m_H cannot be predicted. However, besides the upper bound on the Higgs boson mass from unitarity constraints [12,27], additional theoretical arguments place approximate upper and lower bounds on m_H [28, 29]. There is an upper bound based on the perturbativity of the theory up to the scale Λ at which the SM breaks down, and a lower bound derived from the stability of the Higgs potential. If m_H is too large, then the Higgs self-coupling diverges at some scale Λ below the Planck scale. If m_H is too small, then the Higgs potential develops a second (global) minimum at a large value of the magnitude of the scalar field of order Λ. New physics must enter at a scale Λ or below, so that the global minimum of the theory corresponds to the observed $SU(2)_L \times U(1)_Y$ broken vacuum with $v = 246$ GeV. Given a value of Λ, one can compute the minimum and maximum allowed Higgs boson masses. Conversely, the value of m_H itself can provide an important constraint on the scale up to which the SM remains successful as an effective theory. In particular, a Higgs boson with 125 GeVmass is consistent with an effective SM description that survives all the way to the Planck scale. The main uncertainties in the vacuum stability and perturbativity bounds come from the uncertainties in the value of α_S and the top quark mass. As can be inferred from figure 1.1,taking these uncertainties into account, a Higgs boson mass of about 125 GeV is close to the boundary of a SM that is consistent up to the Planck scale, and a SM that is unstable with a slow tunneling rate.

The SM Higgs couplings to fundamental fermions are proportional to the fermion masses, and the couplings to bosons are proportional to the squares of the boson masses. In particular, the SM Higgs boson is a CP-even scalar, and its couplings to gauge bosons,

Higgs bosons and fermions are given by:

$$g_{Hf\bar{f}} = \frac{m_f}{v}, \quad g_{HVV} = \frac{2m_V^2}{v}, \quad g_{HHVV} = \frac{2m_V^2}{v^2}$$

$$g_{HHH} = \frac{3m_H^2}{v} \quad g_{HHHH} = \frac{3m_H^2}{v^2}$$

where $V = W^{\pm}$ or Z. In Higgs boson production and decay processes, the dominant mechanisms involve the coupling of the H to the $W^{\pm}$, Z and/or the third generation quarks and leptons. The Higgs boson's coupling to gluons, is induced at leading order by a one-loop graph in which the H couples to a virtual $t\bar{t}$ pair. Likewise, the Higgs boson's couplings to photons is also generated via loops, although in this case the one-loop graph with a virtual W^+W^- pair provides the dominant contribution [18]. Reviews of the SM Higgs boson's properties and phenomenology, with an emphasis on the impact of loop corrections to the Higgs boson decay rates and cross sections, can be found in [31–35].

1.2 Definition Of The Higgs Couplings

In this section we describe the mathematical formulæ needed to compute the Higgs production cross section and branching fractions in terms of SM values for the case in which the overall strength of the Higgs coupling to vector bosons c_V, and to fermions c_F, are free to vary. This is a special simplified scenario of the general parameterization of the Higgs couplings introduced in [36, 37]. There are no new production modes nor new decay channels in addition to those present in the SM.

The expression of the four Higgs production cross section is given by a simple rescaling of the SM ones ($V = W, Z$):

$$\begin{aligned}
\sigma(gg \to H) &= c_F^2\, \sigma(gg \to H)_{SM} \\
\sigma(qq \to qqH) &= c_V^2\, \sigma(qq \to qqH)_{SM} \\
\sigma(q\bar{q} \to VH) &= c_V^2\, \sigma(q\bar{q} \to VH)_{SM} \\
\sigma(gg, q\bar{q} \to t\bar{t}H) &= c_F^2\, \sigma(gg, q\bar{q} \to t\bar{t}H)_{SM}
\end{aligned} \tag{1.1}$$

The decay branching ratios are determined by a simple rescaling of the Higgs partial widths. The formulas for these latter are (f denotes any of the quarks and leptons of the SM):

$$\begin{aligned}
\Gamma(H \to VV) &= c_V^2\, \Gamma(H \to VV)_{SM} \\
\Gamma(H \to f\bar{f}) &= c_F^2\, \Gamma(H \to f\bar{f})_{SM} \\
\Gamma(H \to gg) &= c_F^2\, \Gamma(H \to gg)_{SM} \\
\Gamma(H \to \gamma\gamma) &= \frac{\left|c_F\, A_f(m_H) + c_V\, A_W(m_H)\right|^2}{\left|A_f(m_H) + A_W(m_H)\right|^2}\, \Gamma(H \to \gamma\gamma)_{SM} \\
\Gamma(H \to Z\gamma) &= \frac{\left|c_F\, B_f(m_H) + c_V\, B_W(m_H)\right|^2}{\left|B_f(m_H) + B_W(m_H)\right|^2}\, \Gamma(H \to Z\gamma)_{SM}
\end{aligned} \tag{1.2}$$

so that $\Gamma_{tot}(H)$ is the sum of the above partial widths and $BR(H \to X) = \Gamma(H \to X)/\Gamma_{tot}(H)$. The functions A and B are given at one loop by

$$A_f(m_H) = -\frac{8}{3}\frac{4m_t^2}{m_H^2}\left[1+\left(1-\frac{4m_t^2}{m_H^2}\right)\times f\left(\frac{4m_t^2}{m_H^2}\right)\right]\left(1-\frac{\alpha_s}{\pi}\right), \tag{1.3}$$

$$A_W(m_H) = 2+3\times\frac{4m_W^2}{m_H^2}\left[1+\left(2-\frac{4m_W^2}{m_H^2}\right)\times f\left(\frac{4m_W^2}{m_H^2}\right)\right], \tag{1.4}$$

$$B_f(m_H) = -\frac{4\left(\frac{1}{2}-\frac{4}{3}\sin^2\theta_W\right)}{\sin\theta_W\cos\theta_W}\left[I_1\left(\frac{4m_t^2}{m_H^2},\frac{4m_t^2}{m_Z^2}\right)-I_2\left(\frac{4m_t^2}{m_H^2},\frac{4m_t^2}{m_Z^2}\right)\right]\left(1-\frac{\alpha_s}{\pi}\right), \tag{1.5}$$

$$\begin{aligned} B_W(m_H) = &-\frac{\cos\theta_W}{\sin\theta_W}\times\left\{\left(12-4\tan^2\theta_W\right)\times I_2\left(\frac{4m_W^2}{m_H^2},\frac{4m_W^2}{m_Z^2}\right)\right. \\ &\left.+\left[\left(1+\frac{2m_H^2}{4m_W^2}\right)\tan^2\theta_W-\left(5+\frac{2m_H^2}{4m_W^2}\right)\right]\times I_1\left(\frac{4m_W^2}{m_H^2},\frac{4m_W^2}{m_Z^2}\right)\right\}, \end{aligned} \tag{1.6}$$

where

$$I_1(a,b) = \frac{ab}{2(a-b)}+\frac{a^2b^2}{2(a-b)^2}\left[f(a)-f(b)\right]+\frac{a^2b}{(a-b)^2}\left[g(a)-g(b)\right], \tag{1.7}$$

$$I_2(a,b) = -\frac{ab}{2(a-b)}\left[f(a)-f(b)\right], \tag{1.8}$$

with

$$f(x) = \begin{cases} \left[\sin^{-1}\left(1/\sqrt{x}\right)\right]^2 & \text{for } x \geq 1 \\ -\frac{1}{4}\left[\log\left(\frac{1+\sqrt{1-x}}{1-\sqrt{1-x}}\right)-i\pi\right]^2 & \text{for } x < 1, \end{cases} \tag{1.9}$$

and

$$g(x) = \begin{cases} \sqrt{x-1}\sin^{-1}\left(1/\sqrt{x}\right) & \text{for } x \geq 1 \\ \frac{1}{2}\sqrt{1-x}\left[\log\left(\frac{1+\sqrt{1-x}}{1-\sqrt{1-x}}\right)-i\pi\right] & \text{for } x < 1. \end{cases} \tag{1.10}$$

For a full discussion of these results, and expressions for more general cases where new fields can contribute to the loop functions, see for instance [18].

For simplicity, we quote in table 1.1 the numerical values of the functions $A_{f,W}$ and $B_{f,W}$ for the mass range of interest. Note that the contribution from gauge bosons in $H \to Z\gamma$ are on the order of 20 times larger than the contribution from fermions; in practice, modifications to this decay will have a negligible impact on results throughout the space explored in this analysis.

Table 1.1. Numerical values for rescaling factors in loop-mediated processes $H \to \gamma\gamma$ and $H \to Z\gamma$.

m_H (GeV)	A_f	A_W	B_f	B_W
100	-1.75	7.72	0.615	-10.8
110	-1.76	7.93	0.618	-11.2
120	-1.77	8.19	0.621	-11.7
130	-1.78	8.53	0.624	-12.3
140	-1.79	9.01	0.627	-13.2
150	-1.80	9.76	0.631	-14.7
160	-1.81	12.40	0.636	-20.0

1.3 Experimental Searches for the Higgs Boson

1.3.1 Indirect Constraints on the SM Higgs Boson

Indirect experimental bounds for the SM Higgs boson mass are obtained from fits to precision measurements of electroweak observables. The Higgs boson contributes to the $W^{\pm}$ and Z vacuum polarization through loop effects, leading to a logarithmic sensitivity of the ratio of the $W^{\pm}$ and Z gauge boson masses on the Higgs boson mass. A global fit to the precision electroweak data accumulated in the last two decades at LEP, SLC, the Tevatron, and elsewhere, gives $m_H = 94^{+29}_{-24}$ GeV, or $m_H < 152$ GeVat 95% C.L. [14]. The top quark contributes to the $W^{\pm}$ boson vacuum polarization through loop effects that depend quadratically on the top mass, which plays an important role in the global fit. A top quark mass of 173.2 ± 0.9 GeV [38] and a $W^{\pm}$ boson mass of 80.385 ± 0.015 GeV [14] were used.

1.3.2 Searches for the SM Higgs Boson at LEP

The principal mechanism for producing the SM Higgs boson in e^+e^- collisions at LEP energies is Higgs-strahlung in the s-channel, $e + e- \to HZ$. The Z boson in the final state is either virtual (LEP1), or on mass shell (LEP2). At LEP energies, SM Higgs boson production via W^+W^- and ZZ fusion in the t-channel has a small cross section. The sensitivity of the LEP searches to the Higgs boson depends on the center-of-mass energy, $\sqrt{s}$. For $m_H < \sqrt{s} - m_Z$, the cross section is of order 1 pb or more, while for $m_H > \sqrt{s} - M_Z$ the cross section is smaller by at least an order of magnitude.

During the LEP 1 phase, the ALEPH, DELPHI, L3 and OPAL collaborations analyzed over 17 million Z decays and set lower bounds of approximately 65 GeVon the mass of the SM Higgs boson [39]. At LEP 2, substantial data samples were collected at center-of-mass energies up to 209 GeV. Data recorded at each center-of-mass energy were studied independently and the results from the four LEP experiments were then combined. The CLs method [40] was used to compute the observed and expected limits on the Higgs boson production cross section as functions of the Higgs boson mass considered, and from that a lower bound on m_H was derived.

Higgs bosons with mass above $2m_\tau$ were searched for in four final state topologies: The four-jet topology in which $H \to bb$ and $Z \to qq$; the final states with tau leptons produced in the processes $H \to \tau^+\tau^-$ where $Z \to qq$, together with the mode $H \to bb$ with $Z \to \tau^+\tau^-$; the missing energy topology produced mainly in the process $H \to b\bar{b}$

with $Z \to \nu\overline{\nu}$, and finally the leptonic states $H \to b\bar{b}$ with $Z \to e^+e^-$, $\mu^+\mu^-$. At LEP 1, only the modes with $Z \to \ell^+\ell$- and $Z \to \nu\overline{\nu}$ were used because the backgrounds in the other channels were prohibitive. For the data collected at LEP 2, all decay modes were used.

For very light Higgs bosons, with $m_H < 2m_\tau$, the decay modes exploited above are not kinematically allowed, and decays to jets, muon pairs, pion pairs, and lighter particles dominate, depending on m_H. For very low masses, OPAL's decay-mode independent search [41] for the Bjorken process $e^+e^- \to S^0Z$, where S^0 denotes a generic neutral scalar particle, provides sensitivity regardless of the branching fractions of the S^0. This search is based on studies of the recoil mass spectrum in events with $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ decays, and on the final states $Z \to \nu\overline{\nu}$ and $S^0 \to e^+e^-$ or photons. Upper bounds on the $e^+e^- \to ZH$ cross section are obtained for scalar masses between $1\,KeV$ and 100 GeV, and are below 0.05 times the SM prediction for $m_H < 80$ GeV, constraining the coupling of the Higgs boson to the Z.

The combination of the LEP data yields a 95% C.L. lower bound of 114.4 GeVfor the mass of the SM Higgs boson [22]. The median limit one would expect to obtain in a large ensemble of identical experiments with no signal present is 115.3 GeV. An excess of data was seen consistent with a Higgs boson of mass $m_H \sim 115$ GeV. The significance of this excess is low, however. It is quantified by the background-only p-value [41], which is the probability to obtain data at least as signal-like as the observed data, assuming a signal is truly absent; a small p-value indicates data that are inconsistent with the background model but are more consistent with a signal model. The background-only p-value for the excess in the LEP data is 9%.

1.3.3 Searches for the SM Higgs Boson at the Tevatron

At the Tevatron, the most important SM Higgs boson production processes are gluon fusion ($gg \to H$) and Higgs boson production in association with a vector boson ($W^{\pm}H$ or ZH). Vector boson fusion (VBF) has a smaller cross section, but some search channels are optimized for it. For m_H less than about 135 GeV, the most sensitive analyses search for $W^{\pm}$ H and ZH with $H \to b\bar{b}$. The mode $gg \to H \to b\bar{b}$ is overwhelmed by the background from the inclusive production of $p\overline{p} \to b\bar{b} + X$ via the strong interaction. The associated production modes $W^{\pm}H$ and ZH allow use of the leptonic W and Z decays to purify the signal and reject QCD backgrounds.

The contribution of $H \to W^*W$ or WW is dominant at higher masses, $m_H > 135$ GeV. Using this decay mode, both the direct ($gg \to H$) and the associated production ($p\overline{p} \to W^{\pm}H$ or ZH) channels are explored, and the results of both Tevatron experiments, CDF and DØ, are combined to maximize the sensitivity to the Higgs boson.

The signal-to-background ratio is much smaller in the Tevatron searches than in the LEP analyses, and the systematic uncertainties on the estimated background rates are typically larger than the signal rates [40]. In order to estimate the background rates in the selected samples more accurately, auxiliary measurements are made in data samples which are expected to be depleted in Higgs boson signal. These auxiliary samples are chosen to maximize the sensitivity to each specific background in turn. Montecarlo simulations are used to extrapolate these measurements into the Higgs signal regions. The dominant physics backgrounds such as top-pair, diboson, $W^{\pm}b\bar{b}$, and single top production are estimated by Montecarlo simulations in this way, i.e., after having been tuned or verified by

corresponding measurements in dedicated analyses, thereby reducing the uncertainty on the total background estimate. Nearly all Tevatron analyses use multivariate analysis techniques (MVA's) to further separate signals from backgrounds and to provide the final discriminants whose distributions are used to compute limits, best-fit cross sections and uncertainties, and p-values. Separate MVA's are trained at each m_H in all the different sub-channels [40].

All of the searches for the SM Higgs boson at the Tevatron are combined together for maximum sensitivity [42]. The Tevatron combination excludes two ranges in m_H : between 100 GeV and 106 GeV, and between 147 GeVand 179 GeV. An excess of data is seen in the mass range $115\,\text{GeV} < m_H < 135\,\text{GeV}$, with a maximum local significance of 2.7 standard deviations (sigma), at $m_H = 120$ GeV, where the expected local significance for a SM Higgs signal is 2.0 sigma. When corrected for the look-elsewhere effect (LEE) [43], which accounts for the possibility of selecting the strongest of the several random excess which may happen in the range $115\,\text{GeV} < m_H < 200\,\text{GeV}$, the global significance of the excess is 3.1 standard deviations [44]. The majority of the excess is contributed by the searches for $H \to b\bar{b}$. The channels used at the Tevatron for Higgs boson masses below 130 GeVare different from those dominantly used at the LHC, and thus provide complementary information on the couplings of the Higgs boson to gauge bosons and to b quarks.

1.4 The Higgs Boson at the LHC

At the LHC, the Higgs boson production mechanism with the largest cross section is $gg \to H+X$. This process is known at next-to-next-to-leading order (NNLO) in QCD, in the large top-mass limit, and at NLO in QCD for arbitrary top mass [45]. The NLO QCD corrections approximately double the leading-order prediction, and the NNLO corrections add approximately 50% to the NLO prediction. NLO electroweak corrections range between 0 and 6% of the LO term [46]. Mixed QCD-electroweak corrections $\mathcal{O}(\alpha\,\alpha_s)$ are computed in [47]. Updated predictions for the gluon fusion cross sections at NNLO or through soft-gluon resummation up to next-to-next-to-leading logarithmic accuracy (NNLL), and two-loop electroweak effects can be found in [47, 48]. A better perturbative convergence is achieved by resumming the enhanced contributions arising from the analytic continuation of the gluon form factor [49]. Updated predictions to compute the gluon fusion cross sections at NNNLL in renormalization group improved perturbation theory and incorporating two-loop electroweak effects can be found in [50]. Part of the analysis described in chapter 5 is based on the search for the Higgs boson produced in association with jets. In the heavy top quark mass limit, the Higgs boson production cross section in association with one jet is considered in [51–54] and in association with two jets in [55, 56]. The other relevant Higgs boson production mechanisms are associated production with W and Z gauge bosons and vector boson fusion, and at a significantly smaller rate, the associated production with top quark pairs. The cross sections for the associated production processes $qq \to W^{\pm}H + X$ and $qq \to ZH + X$ [57–59] are known at NNLO for the QCD corrections and at NLO for the electroweak corrections [60, 61]. The residual uncertainty is less than 5%. For the vector boson fusion processes $qqqH + X$, corrections to the production cross section are known at NNLO in QCD and at NLO for the electroweak corrections and the remaining theoretical uncertainties in the inclusive cross section are approximately 2% [62], but are larger if jets are required or vetoed [34]. The cross section for the associated production process $t\bar{t}H$ has been calculated at NLO in QCD [63]. of five

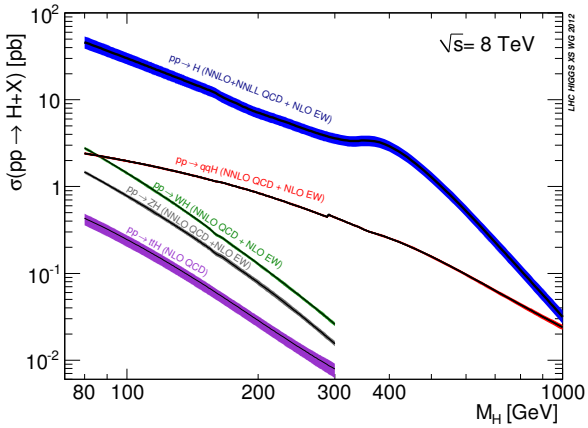


Figure 1.2. SM Higgs boson production cross sections for pp collisions at 8 TeV, including theoretical uncertainties.

quark flavors [60, 64, 65]. The cross sections for the production of SM Higgs bosons for pp collisions at $\sqrt{s} = 8$ TeV at the LHC are summarized in figure 1.2.

The branching ratios for the most relevant decay modes of the SM Higgs boson as functions of m_H, including the most recent theoretical uncertainties, are shown in figure 1.3. The total decay width as function of m_H is shown in figure 1.4. Details of these calculations can be found in [31–34]. For Higgs boson masses below 135 GeV, decays to fermion pairs dominate; the decay $H \to b\bar{b}$ has the largest branching ratio and the decay $H \to \tau^+\tau^-$ is about an order of magnitude smaller. For these low masses, the total decay width is less than 10 MeV. For Higgs boson masses above 135 GeV, the W^+W^- decay dominates (below the W^+W^- threshold, one of the W bosons is virtual) with an important contribution from $H \to ZZ$, and the decay width rises rapidly, reaching about 1 GeVat $m_H = 200$ GeVand 100 GeVat $m_H = 500$ GeV. Above the $t\bar{t}$ threshold, the branching ratio into $t\bar{t}$ pairs increases rapidly as a function of the Higgs boson mass, reaching a maximum of about 20% at $m_H \sim 450$ GeV. Higgs boson decays into pairs of gluons, pairs of photons, and $Z\gamma$ are induced at one loop level. Higgs boson decay into a pair of photons has been particularly relevant for the discovery potential of the LHC for a low-mass Higgs boson. In spite of the small expected signal rate, the reconstructed mass resolution provides a way to separate signal from background, a means to calibrate the background rate with a signal-free sample of events, and a precise measurement of m_H.

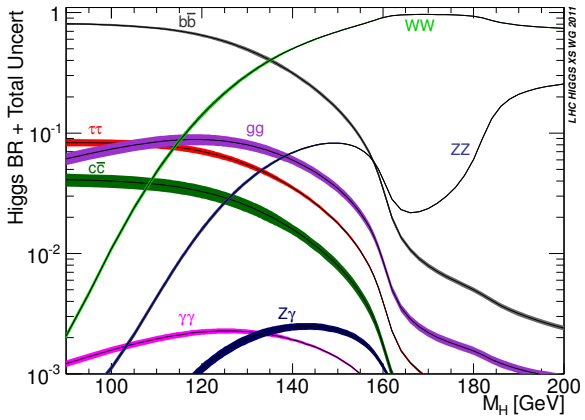


Figure 1.3. Branching ratios for the main decays of the SM Higgs Boson, including theoretical uncertainties.

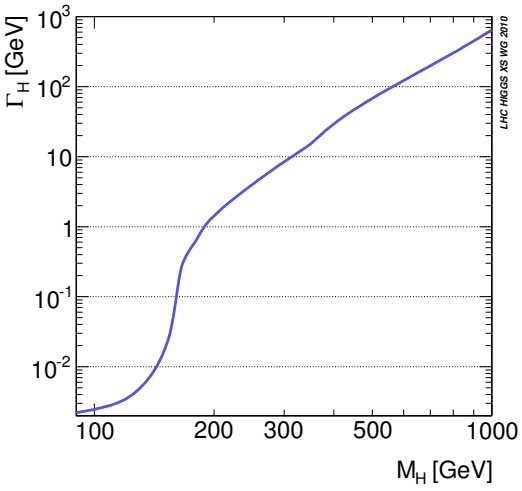


Figure 1.4. The total decay width of the SM Higgs boson, shown as a function of m_H

Chapter 2

Compact Muon Solenoid Detector

2.1 Large Hadron Collider

> «After extended consultation with the appropriate scientific committees, CERN's Director-General Luciano Maiani announced today that the LEP accelerator had been switched off for the last time. LEP was scheduled to close at the end of September 2000 but tantalizing signs of possible new physics led to LEP's run being extended until November 2nd. At the end of this extra period, the four LEP experiments had produced a number of collisions compatible with the production of Higgs particles with a mass of around 115 GeV. These events were also compatible with other known processes. The new data was not sufficiently conclusive to justify running LEP in 2001, which would have inevitable impact on LHC construction and CERN's scientific programme. The CERN Management decided that the best policy for the Laboratory is to proceed full-speed ahead with the Large Hadron Collider (LHC) project.»
>
> *CERN Press Release, 8/11/2000*

With this words, at the beginning of November 2000, the *Large Electron-Positron collider* (LEP) gave way to the new two-ring-superconducting-hadron accelerator and collider: the *Large Hadron Collider* (LHC).

The first approval of the LHC project was already been given by the CERN Council in December 1994. At that time, the plan was to build a machine in two stages starting with a center-of-mass energy of 10 TeV, to be upgraded later to 14 TeV. However, during 1995-6, intense negotiations secured substantial contributions to the project from non-member states, and in December 1996 the CERN Council approved construction of the 14 TeV machine in a single stage. The non-member state agreements ranged from financial donations, through in kind contributions entirely funded by the contributor, to in-kind-contributions that were jointly funded by CERN and the contributor. Confidence for this move was based on the experience gained in earlier years from the international collaborations that often formed around physics experiments. Overall, non-member state involvement has proven to be highly successful.

The decision to build LHC at CERN (European Center for Nuclear Research) was strongly influenced by the cost saving to be made by re-using the LEP tunnel and its injection chain. Although at its founding CERN was endowed with a generous site in the Swiss countryside, with an adjacent site for expansion into the even emptier French countryside,

the need for space outstripped that available when the super-proton synchrotron, or SPS, was proposed. In this instance, the problem was solved by extensive land purchases, but the next machine, LEP, with its 27 km ring, made this solution impractical. In France, the ownership of land includes the underground volume extending to the center of the earth, but, in the public interest, the Government can buy the rights to the underground part for a purely nominal fee. In Switzerland, a real estate owner only owns the land down to a "reasonable" depth. Accordingly, the host states reacted quickly and gave CERN the right to bore tunnels under the two countries, effectively opening a quasi-infinite site that only needed a few "islands" of land ownership for shafts. In 1989, CERN started LEP, the worlds highest energy electron-positron collider. In 2000, LEP was closed to liberate the tunnel for the LHC.

The LHC design depends on some basic principles linked with the latest technology. Being a particle-particle collider, there are two rings with counter-rotating beams, unlike particle-antiparticle colliders that can have both beams sharing the same phase space in a single ring. The tunnel geometry was originally designed for the electron-positron machine LEP, and there were eight crossing points flanked by long straight sections for radio-frequency cavities that compensated the high synchrotron radiation losses. A proton machine such as LHC does not have the same synchrotron radiation problem and would, ideally, have longer arcs and shorter straight sections for the same circumference, but accepting the tunnel "as built" was the cost-effective solution. However, it was decided to equip only four of the possible eight interaction regions and to suppress beam crossings in the other four to prevent unnecessary disruption of the beams. Of the four chosen interaction points, two were equipped with new underground caverns.

The LHC has two high luminosity experiments, ATLAS (A Toroidal LHC ApparatuS) [66] and CMS (Compact Muon Solenoid) [see next section], both aiming at a peak luminosity of $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1} for proton operation. There are also two low luminosity experiments: LHCb [67] for B-physics, aiming at a peak luminosity of $\mathcal{L} = 10^{32}$ cm^{-2}s^{-1}, and TOTEM [68] for the detection of protons from elastic scattering at small angles, aiming at a peak luminosity of $\mathcal{L} = 2 \times 10^{29}$ cm^{-2}s^{-1}. In addition to the proton beams, the LHC is also operated with ion beams. The LHC has one dedicated ion experiment, ALICE [69], aiming at a peak luminosity of $\mathcal{L} = 10^{27}$ cm^{-2}s^{-1} for nominal lead-lead ion operation.

The high beam intensity required for a luminosity of $\mathcal{L} = 10^{34} cm^{-2} s^{-1}$ excludes the use of antiproton beams, and hence excludes the particle-anti-particle collider configuration of a common vacuum and magnet system for both circulating beams, as used for example in the Tevatron. To collide two counter-rotating proton beams requires opposite magnetic dipole fields in both rings. The LHC is therefore designed as a proton-proton collider with separate magnet fields and vacuum chambers in the main arcs, and with common sections only at the insertion regions where the experimental detectors are located. The two beams share an approximately 130 m long common beam pipe along the insertion regions. Hence dedicated *crossing angle orbit bumps* separate the two LHC beams left and right from the interaction point, in order to avoid parasitic collisions.

2.2 CMS Overall Concept

The Compact Muon Solenoid (CMS) is one of the two general purpose experiments which takes data at the LHC. Its physics goals range from the search for the Higgs boson to the

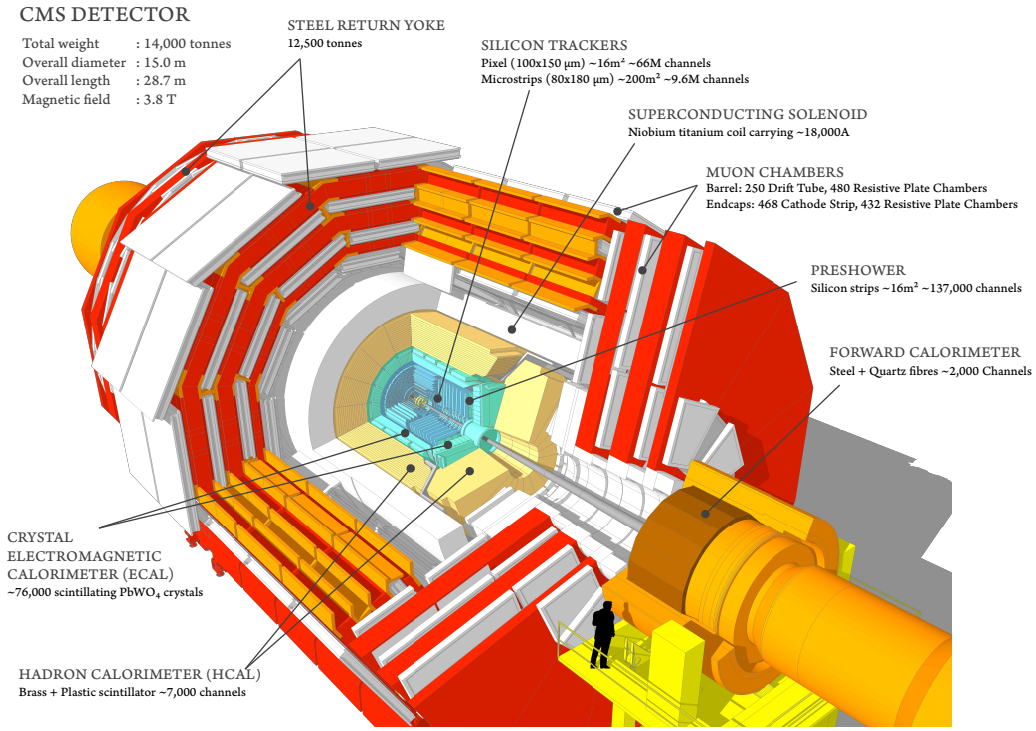


Figure 2.1. An exploded view of the CMS detector.

searches for new physics beyond the Standard Model, to the precision measurements of already known particles and phenomena [70].

The overall layout of CMS is shown in 2.1. At the heart of CMS sits a 13 m-long, 5.9 m inner diameter, 3.8 T superconducting solenoid. In order to achieve good momentum resolution within a compact spectrometer without making stringent demands on muon-chamber resolution and alignment, a high magnetic field was chosen. The return field is large enough to saturate 1.5 m of iron, allowing four muon stations to be integrated to ensure robustness and full geometric coverage. Each muon station consists of several layers of aluminium drift tubes (DTs) in the barrel region, and of cathode strip chambers (CSCs) in the endcap region, complemented by resistive plate chambers (RPCs).

The bore of the magnet coil is also large enough to accommodate the inner tracker and the calorimetry inside. The tracking volume is given by a cylinder of length 5.8 m and diameter 2.6 m. In order to deal with high track multiplicities, CMS employs 10 layers of silicon microstrip detectors, which provide the required granularity and precision. In addition, 3 layers of silicon pixel detectors are placed close to the interaction region to improve the measurement of the impact parameter of charged-particle tracks, as well as the position of secondary vertexes. The electromagnetic calorimeter (ECAL) uses lead tungstate ($PbWO_4$) crystals with coverage in pseudorapidity up to $|\eta| < 3.0$. The scintillation light is detected by silicon avalanche photodiodes (APDs) in the barrel region and vacuum phototriodes (VPTs) in the endcap region. A preshower system is installed in front of the endcap ECAL for π^0 rejection. The ECAL is surrounded by a brass/scintillator sampling hadron calorimeter with coverage up to $|\eta| < 3.0$. The scintillation light is converted by wavelength-shifting (WLS) fibers embedded in the scintillator tiles and channeled to photodetectors via clear

fibers. This light is detected by novel photodetectors (hybrid photodiodes, or HPDs) that can provide gain and operate in high axial magnetic fields. This central calorimetry is complemented by a tail-catcher in the barrel region, ensuring that hadronic showers are sampled with nearly 11 hadronic interaction lengths. Coverage up to a pseudorapidity of 5.0 is provided by an iron/quartz-fiber calorimeter. The Cherenkov light emitted in the quartz fibers is detected by photomultipliers. The forward calorimeters ensure full geometric coverage for the measurement of the transverse energy in the event.

In the following, the CMS sub-detectors are described from the innermost region (the closest to the interaction point) to the outermost region. The chapter ends with a description of the trigger and data acquisition systems.

2.3 Coordinate Conventions

The coordinate system adopted by CMS has the origin centered at the nominal collision point inside the experiment, the y-axis pointing vertically upward, and the x-axis pointing radially inward toward the center of the LHC. Thus, the z-axis points along the beam direction toward the Jura mountains from LHC Point 5. The azimuthal angle ϕ is measured from the x-axis in the x-y plane. The polar angle θ is measured from the z-axis. Pseudorapidity is defined as $\eta = -\ln\tan(\theta/2)$. Thus, the momentum and energy measured 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 denoted by $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$.

2.4 Inner Tracking System

By considering the charged particle flux at various radii at high luminosity, three regions can be delineated:

- the region closest to the interaction vertex, where the particle flux is reaches its maximum ($\sim 10^7\,\mathrm{s}^{-1}$ at $r \sim 10\,\mathrm{cm}$), is instrumented with pixel detectors. The size of a pixel is $\sim 100 \times 150\,\mu\mathrm{m}^2$, giving an occupancy of about 10^{-4} per pixel per LHC crossing;
- in the intermediate region ($20 < r < 55\,\mathrm{cm}$), the particle flux is low enough to enable the use of silicon microstrip detectors with a minimum cell size of $10\,\mathrm{cm} \times 80\,\mu\mathrm{m}$, leading to an occupancy of $\sim$ 2-3%/LHC crossing;
- in the outermost region ($r > 55\,\mathrm{cm}$) of the inner tracker, the particle flux has dropped sufficiently to allow use of larger-pitch silicon microstrips with a maximum cell size of $25\mathrm{cm} \times 180\mu\mathrm{m}$, whilst keeping the occupancy to $\sim$ 1%.

Close to the interaction vertex, in the barrel region, there are three layers of hybrid pixel detectors at radii of 4.4, 7.3, and 10.2 cm. The size of the pixels is $100 \times 150\ \mu m^2$. In the barrel part, the silicon microstrip detectors are placed at r between 20 and 110 cm. The forward region has 2 pixel and 9 microstrip layers in each of the 2 endcaps. The barrel part is separated into an inner and an outer barrel. In order to avoid excessively shallow track crossing angles, the inner barrel is shorter than the outer barrel, and there are an additional three inner disks in the transition region between the barrel and endcap parts, on each side of the inner barrel. The total area of the pixel detector is $\sim 1\,\mathrm{m}^2$, whilst that of the silicon

strip detectors is 200 m^2, providing coverage up to $|\eta| < 2.4$. The inner tracker comprises 66 million pixels and 9.6 million silicon strips [71].

2.4.1 Strip tracker

The barrel tracker region is divided into two parts: a TIB (Tracker Inner Barrel) and a TOB (Tracker Outer Barrel).

The TIB is made of 4 layers and covers up to $|z| < 65$ cm, using silicon sensors with a thickness of 320 μm and a strip pitch which varies from 80 to 120 μm. The first 2 layers are made with *stereo* modules in order to provide a measurement in both r-ϕ and r-z coordinates. A stereo angle of 100 mrad has been chosen. This leads to a single-point resolution of between $23 \div 34$ μm in the r-ϕ direction and 230 μm in z.

The TOB comprises 6 layers with a half-length of $|z| < 110$ cm. As the radiation levels are smaller in this region, thicker silicon sensors (500 μm) can be used to maintain a good signal/noise ratio for longer strip length and wider pitch. The strip pitch varies from 120 to 180 μm. Also for the TOB the first two layers provide a stereo measurement in both r-ϕ and r-z coordinates. The stereo angle is again 100 mrad and the single-point resolution varies from 35-52μm in the r-ϕ direction and 530 μm in z.

The endcaps are divided into the TEC (Tracker End Cap) and TID (Tracker Inner Disks). Each TEC comprises 9 disks that extend into the region 120 cm $< |z| <$ 280 cm, and each TID comprises 3 small disks that fill the gap between the TIB and the TEC. The TEC and TID modules are arranged in rings, centered on the beam line, and have strips that point towards the beam line, therefore having a variable pitch. The first 2 rings of the TID and the innermost 2 rings and the fifth ring of the TEC have stereo modules. The thickness of the sensors is 320, μm for the TID and the 3 innermost rings of the TEC and 500 μm for the rest of the TEC.

The entire silicon strip detector consists of almost 15 400 modules, which are mounted on carbon-fiber structures and housed inside a temperature controlled outer support tube. The operating temperature is around -20° C.

2.4.2 Pixel Tracker

The pixel detector consists of 3 barrel layers with 2 endcap disks on each side of them (see figure 2.2). The 3 barrel layers are located at mean radii of 4.4 cm, 7.3 cm and 10.2 cm, and have a length of 53 cm. The 2 end disks, extending from 6 to 15 cm in radius, are placed on each side at $|z| = 34.5$ cm and 46.5 cm.

In order to achieve the optimal vertex position resolution, a design with an almost square pixel shape of 100×150 μm^2 in both the (r, ϕ) and the z coordinates has been adopted. The barrel comprises 768 pixel modules arranged into half-ladders of 4 identical modules each. The large Lorentz effect (Lorentz angle is 23°) improves the r-ϕ resolution through charge sharing.

The endcap disks are assembled in a turbine-like geometry with blades rotated by 20° to also benefit from the Lorentz effect. The endcap disks comprise 672 pixel modules with 7 different modules in each blade.

The spatial resolution is measured to be about 10 μm for the r-ϕ measurement and about 20 μm for the z measurement. The detector is readout using approximately 16 000 readout chips, which are bump-bonded to the detector modules.

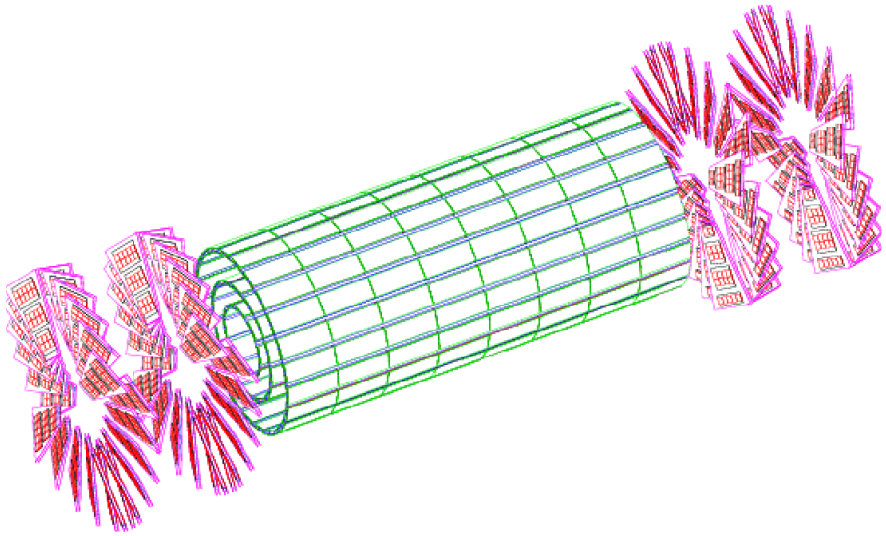


Figure 2.2. Layout of pixel detectors in the CMS tracker.

2.4.3 Performance

The performance of the tracker is illustrated in figure 2.3, which shows the transverse momentum resolution for single muons with a p_T of 1, 10 and 100 GeV, as a function of pseudorapidity. The material inside the active volume of the tracker increases from $\sim 0.4\,X_0$ at $\eta = 0$ to around $1\,X_0$ at $|\eta| \sim 1.6$, before decreasing to $\sim 0.6\,X_0$ at $|\eta| = 2.5$.

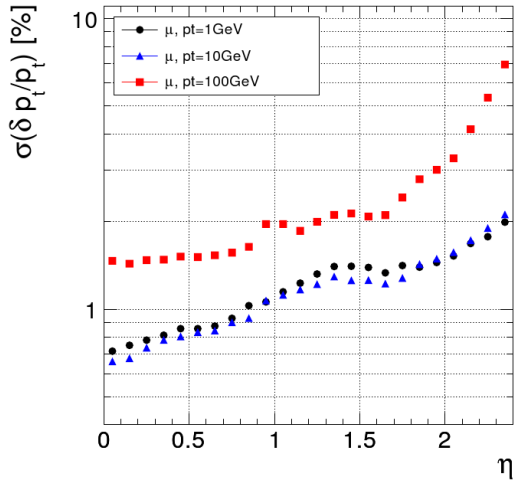


Figure 2.3. Resolution of the transverse momentum as a function of pseudorapidity for single muons with $p_T(\mu) = 1,\ 10,\ 100\,\text{GeV}$.

2.5 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) plays an essential role in the study of the physics of electroweak symmetry breaking, particularly through the exploration of the Higgs sector.

The search for the Higgs with CMS strongly relies on information from the ECAL: by measuring the two-photon decay mode, and by measuring the electrons and positrons from the decay of Ws and Zs originating from the $H \to ZZ$ and $H \to WW$ decay chain.

The ECAL is also an important detector element for a large variety of SM and other new physics processes. The reconstruction of a background-free $Z \to ee$ data sample is relevant for any new high-mass object with one or more Zs in the subsequent decay chain. It is also crucial for other measurements such as cascade decays of gluinos and squarks, where the lepton-pair mass provides information about the supersymmetric particle spectrum, or the leptonic decay of new heavy vector bosons (W', Z') in the multi-TeV mass range.

The choice of a scintillating crystal calorimeter offers the best performance for energy resolution since most of the energy from electrons or photons is deposited within the homogeneous crystal volume of the calorimeter. High density crystals with a small Molière radius allow a very compact electromagnetic calorimeter system.

Several large crystal calorimeters successfully operated in high-energy physics experiments (e.g. L3 at LEP and CLEOII at CESR). However, these detectors did not face the difficult experimental environment at the LHC which imposes stringent and challenging requirements on the detector specifications: under nominal LHC operation, every 25 ns an average of 20 events with some 1000 charged tracks are produced. Compared with the L3 BGO calorimeter, where high precision and wide dynamic range were required, the same criteria have to be met but at a much higher speed and in a much more hostile radiation environment at the LHC. Special efforts have therefore been made during the past years to develop crystals, photodetectors, electronics and software that provide the performance required by the physics at the LHC.

After an intensive initial R&D programme, lead tungstate ($PbWO_4$) crystals were chosen. The choice was based on the following considerations: $PbWO_4$ has a short radiation length and a small Molière radius; it is a fast scintillator; it was relatively easy to produce from readily available raw materials, and substantial experience and production capacity already existed in China and Russia. The initial drawback of low light yield was overcome by progress in crystal growth and through the development of large-area silicon avalanche photodiodes.

2.5.1 Design Concept

The ECAL design was optimized using the $H \to \gamma\gamma$ decay as a benchmark. Since the width of the Higgs signal is entirely dominated by the experimental two-photon mass resolution, it imposes the most stringent performance requirements. The ability to maintain high mass resolution, even under the difficult running conditions, is one of the key design goals for the electromagnetic calorimeter. A 3D view of the barrel and endcap electromagnetic calorimeter is shown in figure 2.4. The above considerations led to the design requirements and constraints given below.

- **Geometry**
  - **Pseudorapidity coverage.** The geometrical crystal coverage extends to $|\eta| = 3$. Precision energy measurement, involving photons and electrons, are carried out to $|\eta| < 2.5$. This limit has been determined by considerations of the radiation dose and amount of pileup energy and matches the geometric acceptance of the inner tracking system.

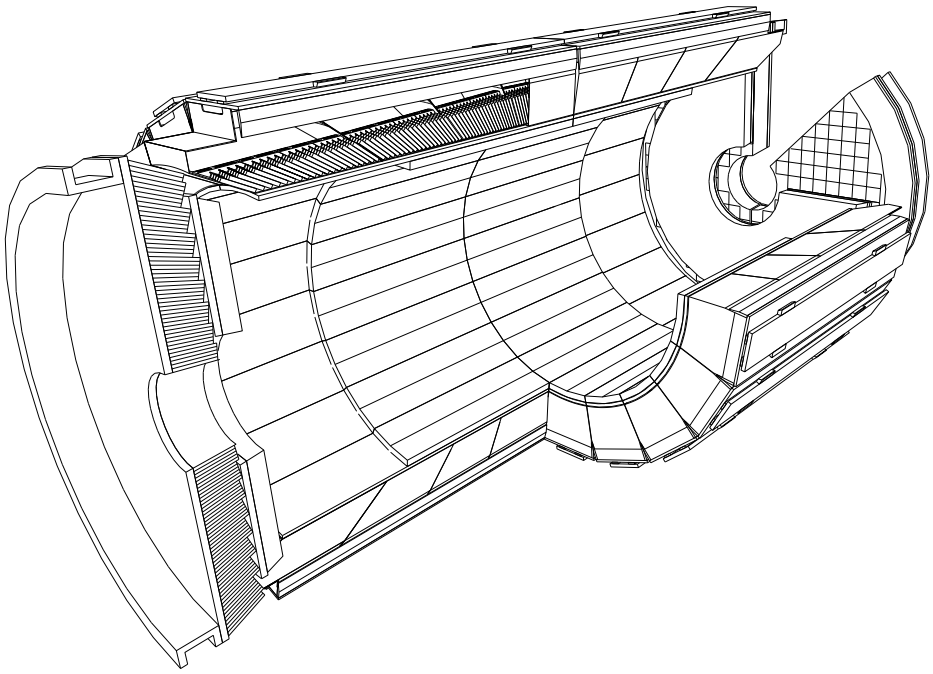


Figure 2.4. A 3D view of the electromagnetic calorimeter.

- **Granularity.** The transverse granularity of $\Delta\eta \times \Delta\phi = 0.0175 \times 0.0175$, corresponding to a crystal front face of about $22 \times 22\,\mathrm{mm}^2$, matches the $PbWO_4$ Molière radius of 21.9 mm. The small Molière radius reduces the effect of pileup contributions to the energy measurement by reducing the area over which the energy is summed. In the endcaps ($1.48 < |\eta| < 3.0$) the granularity increases progressively to a maximum value of $\Delta\eta \times \Delta\phi = 0.05 \times 0.05$, though the crystal front section does not change.
- **Calorimeter thickness.** A total thickness of about 26 radiation lengths at $|\eta| = 0$ is required to limit the longitudinal shower leakage of high-energy electromagnetic showers to an acceptable level. This corresponds to a crystal length of 23 cm in the barrel region. The presence of a preshower (a total of $3X_0$ of lead) in the endcap region allows the use of slightly shorter crystals (22 cm).

• **Readout Chain**

- **Dynamic range.** The dynamic range (12 bits) is set at the lower end by the expected electronic noise per channel, about 30 MeV in the barrel and about 150 MeV in the endcaps, and at the higher end by the energy ($\sim$2 TeV) deposited in a single crystal, for example by electrons from a multi-TeV Z'.
- **Speed of response.** The $PbWO_4$ scintillation decay kinetics can be approximated by a single 10 ns decay time constant. The shaping time of the preamplifiers has been chosen to be 40 ns. This is a compromise between competing requirements of minimizing pile-up energy, on the one hand, and maximizing the amount of light collected and reducing the energy equivalent of electronics noise on the other.

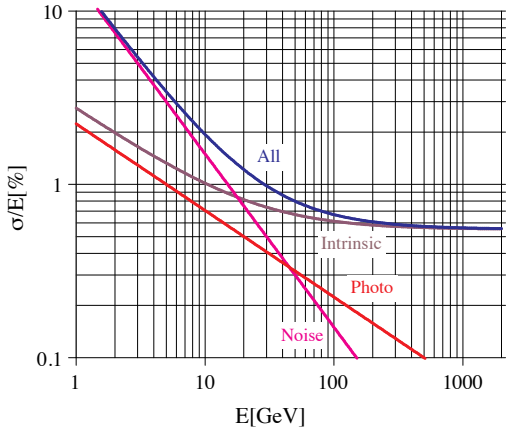


Figure 2.5. Different contributions to the energy resolution of the $PbWO_4$ calorimeter.

- **Resolution**

  - **Energy Resolution.** For the energy range of about 25 GeVto 500 GeV, appropriate for photons from the $H \rightarrow \gamma\gamma$ decay, the energy resolution has been parameterized in this document as:

$$\frac{\sigma(E)}{E} = \frac{a}{\sqrt{E}} \oplus \frac{\sigma_n}{E} \oplus c \quad \text{(E in GeV)} \tag{2.1}$$

    where a is the stochastic term, σ_n the noise, and c the constant term. The stochastic term includes fluctuations in the shower containment as well as a contribution from photostatistics. Figure 2.5 summarizes the different contributions expected for the energy resolution. Terms representing the degradation of the energy resolution at extremely high energies have not been included. The noise term contains the contributions from electronic noise and pile-up energy; the former is quite important at low energy, the latter is negligible at low luminosity. The curve labeled *intrinsic* includes the shower containment and a constant term of 0.55%. The constant term must be kept down to this level in order to profit from the excellent stochastic term of $PbWO_4$ in the energy range relevant for the Higgs search. To achieve this goal, in situ calibration/monitoring using isolated high p_T electrons is performed.

  - **Angular and Mass Resolution.** The two-photon mass resolution depends on the energy resolution and the error on the measured angle between the two photons. If the vertex position is known, the angular error is negligible. However, a contribution of about 1.5 GeVto the diphoton mass resolution (at a mass of around 100 GeV) is expected from the uncertainty in the position of the interaction vertex, if the only information available is the r.m.s spread of about 5.3 cm of the interaction vertexes. At low luminosity, where the number of superimposed events is small, the longitudinal position of the Higgs production vertex can be localized using high-p_T tracks originating from the Higgs event. Studies indicate that even at high luminosity the correct vertex can be located for a large fraction of events using charged tracks.

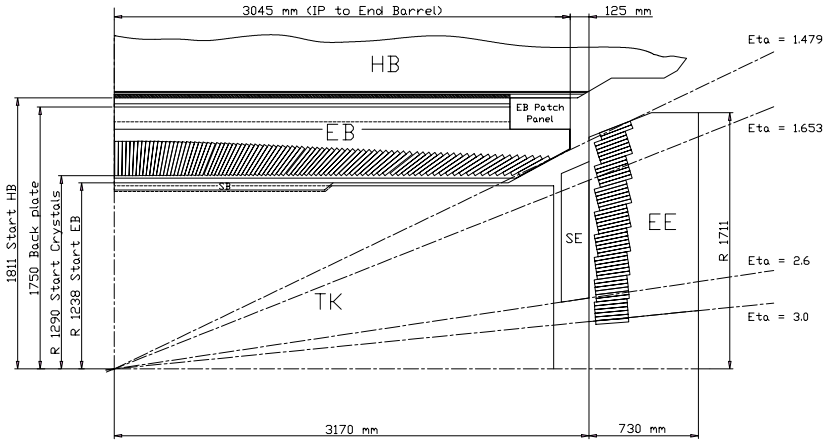


Figure 2.6. Longitudinal section of the electromagnetic calorimeter (one quadrant).

2.5.2 The Barrel Calorimeter

The barrel part of the ECAL covers the pseudorapidity range $|\eta| < 1.479$ (see figure 2.6). The front face of the crystals is at a radius of 1.29 m and each crystal has a square cross-section of $\sim 22 \times 22\,\mathrm{mm}^2$ and a length of 230 mm corresponding to $25.8\,\mathrm{X}_0$. The truncated pyramid-shaped crystals are mounted in a geometry which is off-pointing with respect to the mean position of the primary interaction vertex, with a 3° tilt in both ϕ and in η. The crystal cross-section corresponds to $\Delta\eta \times \Delta\phi = 0.0175 \times 0.0175\,(1°)$. The barrel granularity is 360-fold in ϕ and (2×85)-fold in η, resulting in a total number of 61 200 crystals. The crystal volume in the barrel amounts to $8.14\,\mathrm{m}^3$ (67.4 t). Crystals for each half-barrel are grouped in 18 supermodules each subtending 20° in ϕ. Each supermodule comprises four modules with 500 crystals in the first module and 400 crystals in each of the remaining three modules. For simplicity of construction and assembly, crystals have been grouped in arrays of 2×5 crystals which are contained in a very thin wall (200 μm) alveolar structure and form a submodule. Thermal regulation is carried out by two active systems: 1) a specially regulated cooling circuit which keeps the operating temperature (ambient temperature) of the crystal array and of the APDs within a tight temperature spread of $\pm 0.05\,°\mathrm{C}$, ensuring adequate thermal stability; 2) the power cooling circuit evacuates the heat generated by all power sources in the supermodule (each supermodule is designed as a separate thermal entity).

2.5.3 The Endcap Calorimeter

The endcap part of the crystal calorimeter covers a pseudorapidity range from 1.48 to 3.0. The design of the endcaps provides precision energy measurement to $|\eta| = 2.5$. Crystals are however installed up to $|\eta| = 3$ in order to augment the energy-flow measurement in the forward direction. The mechanical design of the endcap calorimeter is based on an off-pointing pseudo-projective geometry using tapered crystals of the same shape and dimensions $(24.7 \times 24.7 \times 220\,\mathrm{mm}^3)$ grouped together into units of 36, referred to as supercrystals. A total of 268 identical supercrystals is used to cover each endcap with a further 64 sectioned supercrystals used to complete the inner and outer perimeter. Each endcap contains 10 764

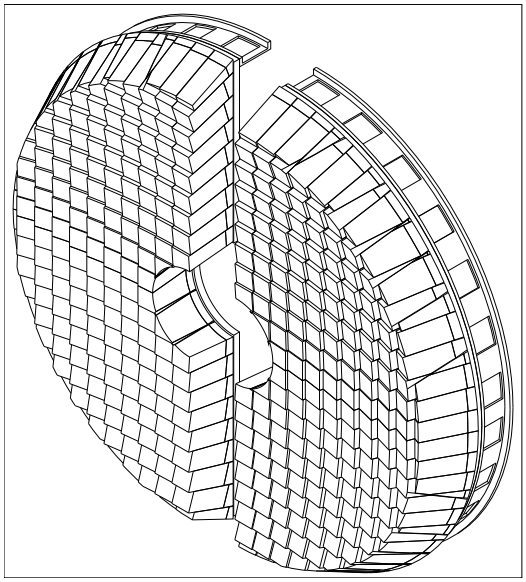


Figure 2.7. A single endcap with Dees apart.

crystals, corresponding to a volume of $1.52\,\mathrm{m}^3$ (12.6 t). Both endcaps are identical. Each endcap detector is constructed using Dee-shaped sections as seen in figure 2.7.

Figure 2.8 shows the total thickness (in radiation lengths) of the ECAL as a function of pseudorapidity; where the endcap part also includes the preshower detector.

Because of the high radiation levels in the endcaps all materials used in this region must tolerate very large doses and neutron fluences.

The endcap calorimeter operates at a temperature close to ambient, which must be stabilized to within $0.1\,^{\circ}\mathrm{C}$. The preshower detector mounted in front of the endcaps operates at $-5\,^{\circ}\mathrm{C}$, thus care must be taken to avoid any condensation problems. Cooling requirements for individual crystals are met by means of the thermal conduit provided from the rear face of the crystal through the metal inserts to the interface plate and support elements. Cooling regulation are provided by a water cooling system installed on the Dee support plate.

2.5.4 The Preshower Detector

The endcap preshower covers a pseudorapidity range from $|\eta| = 1.65$ to 2.61. Its main function is to provide π^0-γ separation. The preshower detector, placed in front of the crystals, contains two lead converters of a total thickness of $2\,\mathrm{X}_0$ and $1\,\mathrm{X}_0$ respectively), followed by detector planes of silicon strips with a pitch of < 2 mm. The impact position of the electromagnetic shower is determined by the center-of-gravity of the deposited energy. The accuracy is typically $300\,\mu$m at 50 GeV. In order to correct for the energy deposited in the lead converter, the energy measured in the silicon is used to apply corrections to the energy measurement in the crystal. The fraction of energy deposited in the preshower (typically 5% at 20 GeV) decreases with increasing incident energy. Figure 2.9 shows the layout of the preshower.

To maintain its performance during the lifetime of the experiment, the endcap silicon detector has to be operated at $-5\,^{\circ}\mathrm{C}$. Heating films and insulating foam glued on the moderators guarantee that the external surfaces are kept at the ambient temperature of the

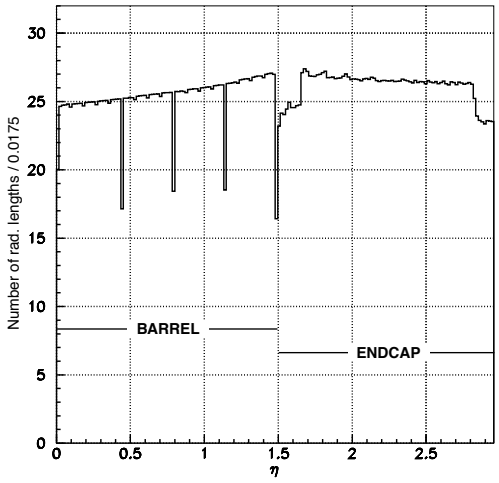


Figure 2.8. Total thickness in X_0 of the ECAL as a function of pseudorapidity, averaged over ϕ.

neighboring detectors.

2.5.5 Lead Tungstate Crystals

Lead Tungstate crystals ($PbWO_4$) were produced for CMS by the Bogoroditsk Techno-Chemical Plant in Russia and by the Shanghai Institute of Ceramics in China. The characteristics [40] of these production crystals make them an appropriate choice for operation at LHC. The high density ($8.3\,g/cm^3$), short radiation length ($0.89\,cm$) and small Molière radius ($2.2\,cm$) results in a fine granularity and a compact calorimeter. The scintillation decay time is of the same order of magnitude as the LHC bunch crossing time: about 80% of the light is emitted in 25 ns. The light output is relatively low: about 4.5 photoelectrons per MeV are collected in both the avalanche photodiodes (APDs) and the vacuum phototriodes (VPTs), where the higher APD quantum efficiency is balanced by their smaller surface coverage on the back face of the crystal. The crystals emit blue-green scintillation light with a broad maximum at 420 nm [72, 73]. The light output variation with temperature, -1.9% per °C at 18 °C, requires an ECAL cooling system capable of extracting the heat dissipated by the readout electronics and of keeping the crystal temperature stable within ± 0.05 °C to preserve energy resolution.

To exploit the total internal reflection for optimum light collection on the photodetector, the crystals are polished after machining. This is done on all but one side for EB crystals. For fully polished crystals, the truncated pyramidal shape makes the light collection non-uniform along the crystal length, and the needed uniformity [74] is achieved by depolishing one lateral face. In the EE, the light collection is naturally more uniform because the crystal geometry is nearly parallelepipedic, and just a mild tuning has been considered.

The crystals have to withstand the radiation levels and particle fluxes [75] anticipated throughout the duration of the experiment. Ionizing radiation produces absorption bands through the formation of color centers due to oxygen vacancies and impurities in the lattice. The practical consequence is a wavelength-dependent loss of light transmission without changes to the scintillation mechanism, a damage which can be tracked and corrected for by monitoring the optical transparency with injected laser light. The damage reaches a dose-

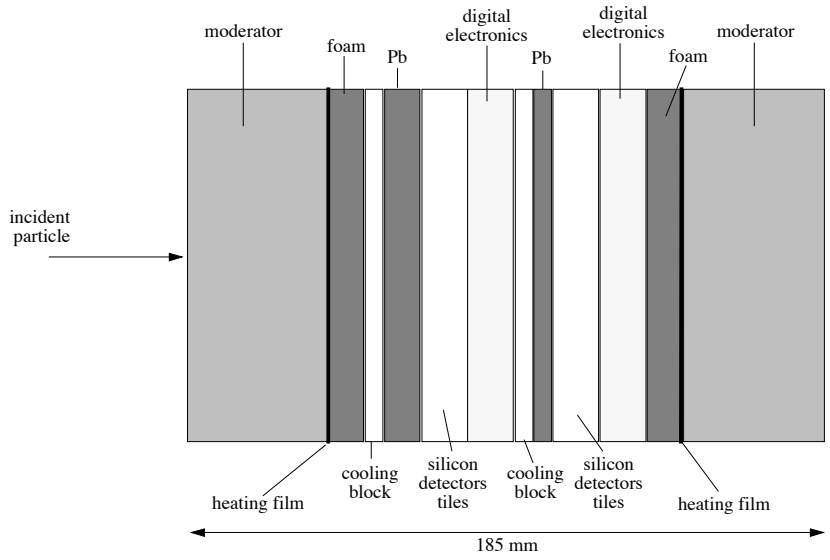


Figure 2.9. Schematic section through the endcap preshower.

rate dependent equilibrium level which results from a balance between damage and recovery at 18 °C [72, 76]. To ensure an adequate performance throughout LHC operation, the crystals are required to exhibit radiation hardness properties quantified as an induced light attenuation length always greater than 3 times the crystal length even when the damage is saturated. Hadrons have been measured to induce a specific, cumulative reduction of light transmission, but the extrapolation to LHC indicates that the damage remains within limits required for good ECAL performance [77].

2.5.6 Photodetectors and Readout Electronics

The high bunch-crossing frequency at the LHC necessitates a pipelined readout. As the time required to form and return a trigger decision is roughly 100 bunch crossings, the energy deposition in each crystal associated with a given bunch crossing must be stored until the trigger decision is available. The energy deposited in the crystals is converted to a digital form every 25 ns and then stored in a digital representation. At the same time, the digital values for the crystal in each trigger tower ($\Delta\eta \times \Delta\phi = 0.087 \times 0.087$) are summed for use by the first-level calorimeter trigger.

The relatively low light yield of $PbWO_4$ along with the high magnetic field and radiation environment in CMS severely limit the choice of the photodetector. The low light yield means that unity gain devices (silicon or vacuum photodiodes) are not capable of delivering the noise performance needed for small (electrical) signals. The very high magnetic field rules out vacuum devices in the central (low η) region, and limits the number of gain stages that could be used in the forward (high η) region. Furthermore, the high radiation environment in the forward direction excludes solid-state and hybrid devices. Silicon avalanche photodiodes are used in the barrel and vacuum phototriodes in the endcaps.

APDs cover 50 mm^2 of the crystal surface, operate at gains of 50 to 100 and have quantum efficiencies (for $PbWO_4$) of ~80%. VPTs have a sensitive area of 180 mm^2, operate at gains approaching 10 in a 4 T field and have quantum efficiencies of ~15%.

Large-area silicon avalanche photodiodes are new devices in high-energy physics experiments, and have undergone considerable development in the past few years. APDs are similar to silicon photodiodes, with the exception of a buried p-n junction reverse-biased at a very high electric field. Photoelectrons arriving at the junction undergo avalanche multiplication, giving the device gain. This gain is sensitive to variations of voltage and temperature, thus the APD must be operated under stabilized conditions. APDs are quite radiation-hard (compared to diodes), however radiation-induced leakage currents can cause a degradation in noise performance for high neutron fluence.

The gain of vacuum phototriodes is relatively insensitive to magnetic field when operated in the orientation to the magnetic field present in the endcaps, and is less sensitive to variation of temperature and voltage than APDs. To first order, the radiation hardness of vacuum photodetectors depends simply on the window material, and radiation-hard glasses are available.

Wide dynamic range with excellent noise performance and signal acquisition precision are required to achieve the physics goals. The signal acquisition has thus been designed as a unit based on the following principles:

- $PbWO_4$ is a relatively fast scintillator at LHC speeds. With an average decay time of 10 ns, not all of the charge can be collected in one bunch crossing, thus excluding classical gated integrators. However, $PbWO_4$ is sufficiently fast that very simple pulse-shaping schemes allow a voltage-sampling system to be used without recourse to complex electronics.

- Excellent noise performance requires the gain to be located as close as possible to the front-end.

- As relative rather than absolute precision is required, and multiple gains are employed for noise reduction, a floating-point gain-switching system is the natural choice.

The core of the readout thus involves a transimpedance preamplifier with built-in shaping, and a gain-ranging multiplexer which forms a *floating-point* front-end to a 12-bit, 40 MHz voltage-sampling ADC. The transimpedance design takes optimal advantage of the simple exponential scintillation pulse shape and operates neither as a charge-sensitive preamplifier nor as a current-to-voltage converter, but somewhere in between.

A significant technical change in the readout design since the Technical Proposal has been the choice of where to place the data pipeline. As originally conceived, the pipeline (and thus the trigger summing circuitry) would have been located with the front-end readout directly behind the crystals. High-speed digital fiber-optic links would have been used to transmit the trigger information and readout data, with lower speed links for timing and control. This solution would require large quantities of digital electronics capable of surviving the calorimeter radiation environment, and would make it exceedingly difficult to make any architectural changes to the way in which trigger and readout information is formed. At the cost of an order of magnitude increase in the number of digital links, it has been decided to place the pipeline outside the detector, and use one high-speed data link per crystal. Thus all data are transported, every 25 ns, out of the ECAL and into the counting room. Once in the counting room, the data are stored in pipelines and the trigger information is extracted.

2.6 Magnet

The required performance of the muon system, and hence the bending power, is defined by the narrow states decaying into muons and by the unambiguous determination of the sign for muons with a momentum of $\sim$1 TeV/c. This requires a momentum resolution of $\Delta p/p \sim 10\%$ at $p = 1$ TeV.

CMS chose a large superconducting solenoid, the parameters of which are given in table 2.1.

Table 2.1. Parameters of the CMS superconducting solenoid.

Parameter	Value
Field	3.8 T
Inner bore	5.9 m
Length	12.9 m
Number of turns	2168
Current	19.5 kA
Stored energy	2.7 GJ

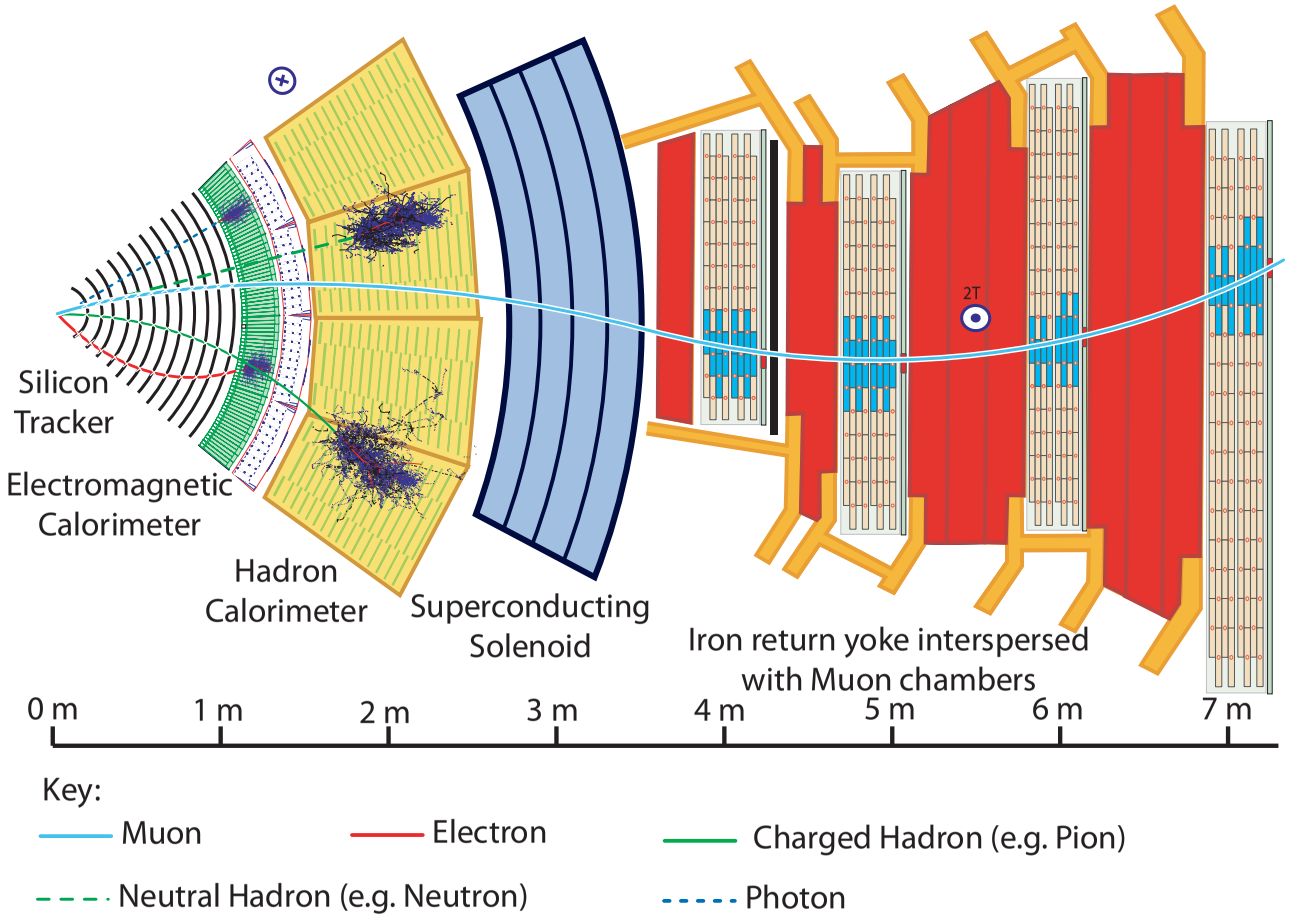


Figure 2.10. Schematic view of a transverse slice of the central part of the CMS detector.

2.7 Hadron Calorimeter

The design of the hadron calorimeter (HCAL) [78] is strongly influenced by the choice of the magnet parameters since most of the CMS calorimetry is located inside the magnet coil and surrounds the ECAL system (see figure 2.10). An important requirement of HCAL is to minimize the non-Gaussian tails in the energy resolution and to provide good containment and hermeticity. Hence, the HCAL design maximizes material inside the magnet coil in terms of interaction lengths. This is complemented by an additional layer of scintillators, referred to as the hadron outer (HO) detector, lining the outside of the coil. Brass has been chosen as absorber material as it has a reasonably short interaction length, is easy to machine and is non-magnetic. Maximizing the amount of absorber before the magnet requires minimizing the amount of space devoted to the active medium. The tile/fiber technology makes for an ideal choice. It consists of plastic scintillator tiles read out with embedded wavelength-shifting (WLS) fibers. The WLS fibers are spliced to high-attenuation-length clear fibers are just outside the scintillator carrying the light to the readout system. This technology was first developed by the UA1 collaboration [79] and at Protvino [80] and it was used in the upgrade of the CDF endcap calorimeter [81]. The photodetection readout is based on multi-channel hybrid photodiodes (HPDs). The absorber structure is assembled by bolting together precisely machined and overlapping brass plates so as to leave space to insert the scintillator plates, which have a thickness of 3.7 mm. The overall assembly enables the HCAL to be built with essentially no uninstrumented cracks or dead areas in ϕ. The gap between the barrel and the endcap HCAL, through which the services of the ECAL and the inner tracker pass, is inclined at 53° and points away from the center of the detector.

2.7.1 Design

HCAL Barrel

The barrel hadron calorimeter is an assembly of two half barrels, each composed of 18 identical 20° wedges in ϕ. The wedge is composed of flat brass alloy absorber plates parallel to the beam axis. The innermost and outermost absorber layers are made of stainless steel for structural strength. There are 17 active plastic scintillator tiles interspersed between the stainless steel and brass absorber plates. The first active layer is situated directly behind the ECAL. This layer has roughly double the scintillator thickness to actively sample low energy showering particles from support material between the ECAL and HCAL. The longitudinal profile in the barrel going from an inner radius of 177.7 cm to an outer radius of 287.6 cm is given by

- (Layer 0) 9 mm Scintillator/61 mm Stainless Steel;
- (Layers 1-8) 3.7 mm Scintillator/50.5 mm Brass;
- (Layers -14) 3.7 mm Scintillator/56.5 mm Brass;
- (Layers +16) 3.7 mm Scintillator/75 mm Stainless Steel/9 mm Scintillator;

where the layer number refers to the active scintillator layer. The individual tiles of scintillator are machined to a size of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ and instrumented with a single WLS. The WLS fibers are spliced to clear fibers, and the clear fibers are run down the length of the half-barrel where they are optically added to corresponding projective tiles

from each of the 17 active layers, thus forming 32 barrel HCAL towers in η. The exceptions are towers 15 and 16 located at the edge of the HB half-barrel where multiple optical readouts are present, as shown in figure 2.11.

The optical signal from the HCAL towers is detected with a pixelated hybrid photodiode (HPD) mounted at the ends of the barrel mechanical structure. An additional layer of scintillators, the outer hadron calorimeter (HO), is placed outside of the solenoid and has a matching $\Delta\eta \times \Delta\phi$ projective geometry with a separate optical readout.

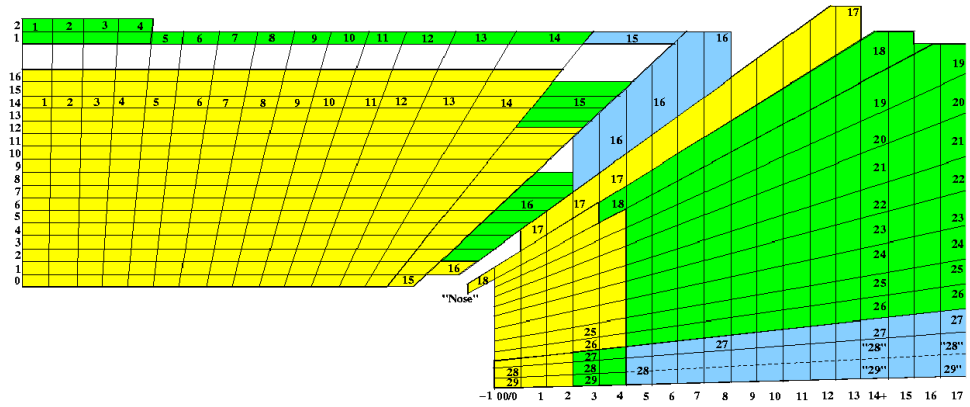


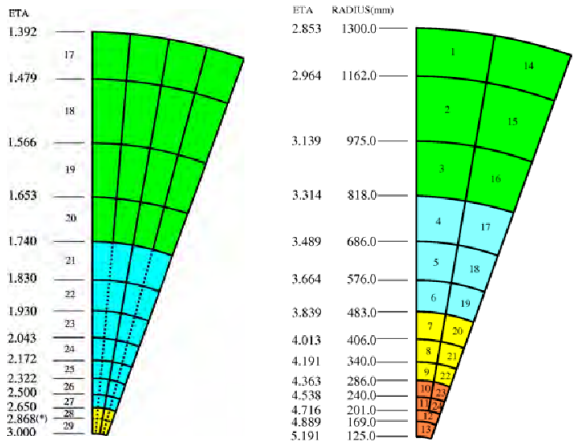
Figure 2.11. A schematic view of the tower mapping in r-z of the HCAL barrel and endcap regions. The shaded regions corresponds to different readout channels.

HCAL Endcaps

The endcap hadron calorimeter (HE) is tapered to interlock with the barrel calorimeter and to overlap with tower 16, as shown in figure 2.11. The HE is composed entirely of brass absorber plates in an 18-fold ϕ-geometry matching that of the barrel calorimeter. The thickness of the plates is 78 mm while the scintillator thickness is 3.7 mm, hence reducing the sampling fraction. There are 19 active plastic scintillator layers. In the high η-region, i.e. above $|\eta| = 1.74$, the ϕ-granularity of the tiles is reduced to 10° to accommodate the bending radius of the WLS fiber readout, as shown in figure 2.12(a). For the purpose of uniform segmentation in the level-1 calorimeter trigger, the energies measured in the 10° ϕ-wedges are artificially divided into equal shares and sent separately to the trigger. The $\Delta\eta \times \Delta\phi$ tower size matches that of the barrel in the range $1.3 < |\eta| < 1.74$. For $|\eta| > 1.74$, the η size increases. The number of depth segments in the HE includes a pseudo-EM compartment starting with tower 18, the first tower beyond the η coverage of the ECAL barrel.

HCAL Outer

The outer barrel hadron calorimeter consists of layers of scintillator located outside of the magnet coil. Since these are located within the return yoke along with the barrel muon detector, the segmentation of these detectors closely follows that of the barrel muon system. The entire assembly is divided into 5 rings (2.54 m wide along the z-axis), each having 12 sectors. The central ring (ring 0) has two layers of 10 mm thick scintillators on either side of the *tail catcher* iron (18 cm thick) at radial distances of 385 cm and 409.7 cm, respectively. All



(a) (b)

Figure 2.12. (a) The η-ϕ view of a 20° HE section showing the 5° regions for $\eta < 1.740$ and the 10° regions for $\eta > 1.740$. The tower 28/29 split in η is also shown. (b) The r-ϕ view of an HF wedge (η at z = 11.2 m). The shaded regions correspond to the level-1 trigger sums.

other rings have a single layer at a radial distance of 409.7 cm. The panels in the 12 sectors are identical except those in rings ± 1. This is due to the chimney structure in the magnet. To accommodate this structure, special panels were built with a single row of scintillator tiles removed. The HO covers $|\eta| < 1.26$ with the exception of the space between successive muon rings in the η direction, the space occupied by 75 mm stainless steel support beams separating the 12 layers in ϕ and the chimney structure. The inclusion of the HO layers extends the total depth of the calorimeter system to a minimum of 11 interaction lengths for $|\eta| < 1.26$.

HCAL Forward

The forward calorimeters (HF) are located 11.2 m from the interaction point. They are made of steel absorbers and embedded radiation hard quartz fibers, which provide a fast collection of Cherenkov light. Each HF module is constructed of 18 wedges in a non-projective geometry with the quartz fibers running parallel to the beam axis along the length of the iron absorbers. Long (1.65 m) and short (1.43 m) quartz fibers are placed alternately with a separation of 5 mm. These fibers are bundled at the back of the detector and are readout separately with phototubes. The r-ϕ view of an HF wedge is shown in figure 2.12(b).

2.7.2 Performance

To test the performance of the HCAL, it is usual to look at the jet energy resolution and the transverse energy resolution. The granularity of the sampling in the three parts of the HCAL has been chosen such that the jet energy resolution, as a function of E_T, is similar in all three parts. This is illustrated in figure 2.13. The resolution of the transverse energy (E_T) in QCD dijet events with pile-up is given by

$$\frac{\sigma(E_T)}{E_T} \sim \frac{100\%}{\sqrt{E_T}}\,. \tag{2.2}$$

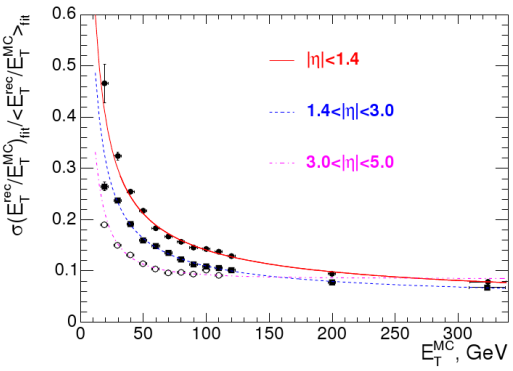


Figure 2.13. The jet transverse energy resolution as a function of the simulated jet transverse energy for barrel jets ($|\eta| < 1.4$), endcap jets ($1.4 < |\eta| < 3.0$) and very forward jets ($3.0 < |\eta| < 5.0$).

2.8 Muon System

Centrally produced muons are measured three times: in the inner tracker, after the coil, and in the return flux. Measurement of the momentum of muons using only the muon system is essentially determined by the muon bending angle at the exit of the 3.8 T coil, taking the interaction point (which is known with a precision of $\sim 20\,\mu$m) as the origin of the muon. The resolution of this measurement (labeled "muon system only" in figure 2.14) is dominated by multiple scattering in the material before the first muon station up to p_T values of 200 GeV, when the chamber spatial resolution starts to dominate. For low-momentum muons, the best momentum resolution is obtained in with the silicon tracker ("inner tracker only" in figure 2.14). However, the muon trajectory beyond the return yoke extrapolates back to the beam-line because of the compensation of the bend before and after the coil, when multiple scattering and energy loss can be neglected. This fact can be used to improve the muon momentum resolution at high momentum when combining the inner tracker and muon detector measurements ("full system" in figure 2.14).

Three types of gaseous detectors are used to identify and measure muons [82]. The choice of the detector technologies has been driven by the very large surface to be covered and by the different radiation environments. In the barrel region ($|\eta| < 1.2$), where the neutron induced background is small, the muon rate is low and the residual magnetic field in the chambers is low, drift tube (DT) chambers are used. In the two endcaps, where the muon rate as well as the neutron induced background rate is high, and the magnetic field is also high, cathode strip chambers (CSC) are deployed and cover the region up to $|\eta| < 2.4$. In addition to this, resistive plate chambers (RPC) are used in both the barrel and the endcap regions. These RPCs are operated in avalanche mode to ensure good operation at high rates (up to 10 kHz/cm^2) and have double gaps with a gas gap of 2 mm. The RPCs provide a fast response with good time resolution but with a coarser position resolution than the DTs or CSCs. The RPCs can therefore identify unambiguously the correct bunch crossing.

The DTs or CSCs and the RPCs operate within the first level trigger system, providing two independent and complementary sources of information. The complete system results in a robust, precise and flexible trigger device.

The layout of one quarter of the CMS muon system is shown in figure 2.15. In the Muon Barrel (MB) region, 4 stations of detectors are arranged in cylinders interleaved with

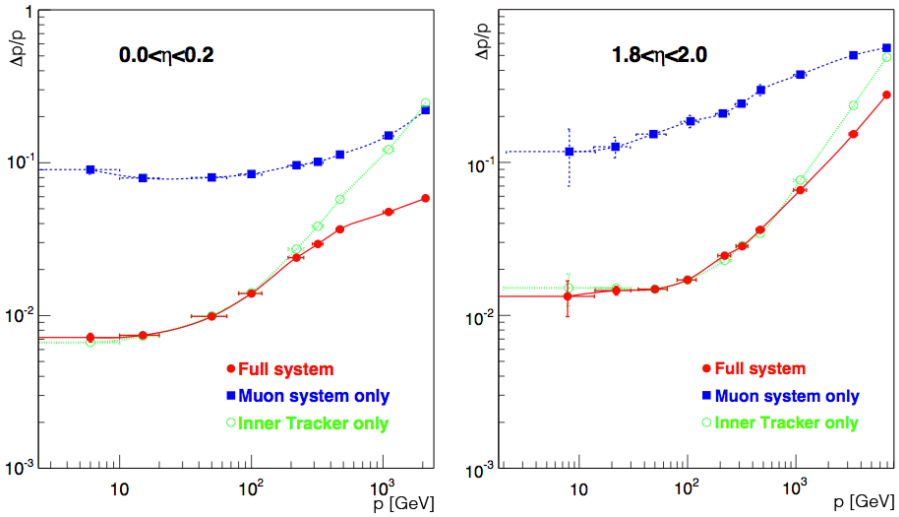


Figure 2.14. The muon momentum resolution versus p using the muon system only, the inner tracker only, or both ("full system"). Left) Central slice of the barrel: $|\eta| < 0.2$; Right) Central slice of the endcap: $1.8 < |\eta| < 2.0$.

the iron yoke. The segmentation along the beam direction follows the 5 wheels of the yoke (labeled YB-2 for the farthest wheel in -z, and YB+2 for the farthest in +z). In each of the endcaps, the CSCs and RPCs are arranged in 4 disks perpendicular to the beam, and in concentric rings, 3 rings in the innermost station, and 2 in the others. In total, the muon system contains about 25 000 m^2 of active detection planes, and nearly 1 million electronic channels.

2.8.1 Drift tube chambers

The Muon Barrel, consists of 250 chambers organized in 4 layers inside the magnet return yoke, at radii of approximately 4.0, 4.9, 5.9 and 7.0 m from the beam axis. Each of the 5 wheels of the Muon Barrel is divided into 12 sectors, with each covering a 30° azimuthal angle. Chambers in different stations are staggered so that a high-p_T muon produced near a sector boundary crosses at least 3 out of the 4 stations. There are 12 chambers in each of the 3 inner layers. In the 4th layer, the top and bottom sectors host 2 chambers each, thus leading to a total of 14 chambers per wheel in this outermost layer. In each chamber, there are 12 layers of contiguous drift tube cells grouped in three *SuperLayers* (SL) with 4 staggered layers each; the innermost and outermost SLs are dedicated to hits measurement in the CMS bending plane (r-ϕ plane), while in the central SL the hits are measured along the beam axis (r-z plane). The maximum drift length is 2.0 cm and the single-point resolution is $\sim$200 μm. Each station is designed to give a muon vector in space, with a ϕ precision better than 100 μm in position and approximately 1 mrad in direction.

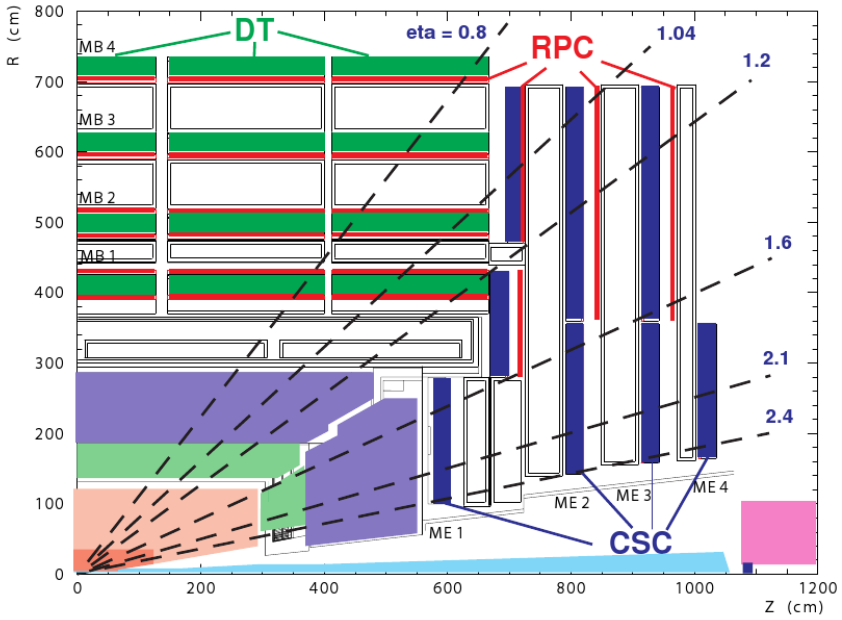


Figure 2.15. Layout of one quarter of the CMS muon system.

2.9 Trigger and Data Acquisition

For the nominal LHC design luminosity of 10^{34} cm^{-2}s^{-1}, an average of 17 events occurs at the beam crossing frequency of 25 ns. This input rate of 10^9 interactions every second must be reduced by a factor of at least 10^7 to 100 Hz, the maximum rate that can be archived by the on-line computer farm. CMS has chosen to reduce this rate in two steps. At the first level all data is stored for 3.2 μs, after which no more than 100 kHz of the stored events are forwarded to the High Level Triggers. This must be done for all channels without dead time. The Level-1 (L1) system is based on custom electronics. The High Level Trigger (HLT) system, relies upon commercial processors. The L1 system uses only coarsely segmented data from calorimeter and muon detectors, while holding all the high-resolution data in pipeline memories in the front-end electronics. The HLT is provided by a subset of the on-line processor farm which, in turn, passes a fraction of these events to the remainder of the on-line farm for more complete processing.

The physical size of the CMS detector and underground caverns imposes constraints on signal propagation that combine with electronics technology to require 3.2 μs, equivalent to 128 25-ns beam crossings, for any primary decision to discard data from a particular beam crossing. During this 3.2 μs period, trigger data must be collected from the front end electronics, decisions must be developed that discard a large fraction of the data while retaining the small portion coming from interactions of interest and these decisions must be propagated to the readout electronics front end buffers.

The trigger is the start of the physics event selection process. A decision to retain an event for further consideration has to be made every 25 ns. This decision is based on the event's suitability for inclusion in one of the various datasets to be used for analysis. The datasets to be taken are determined by CMS physics priorities as a whole. These datasets include dilepton and multilepton data sets for top and Higgs searches, diphoton datasets for

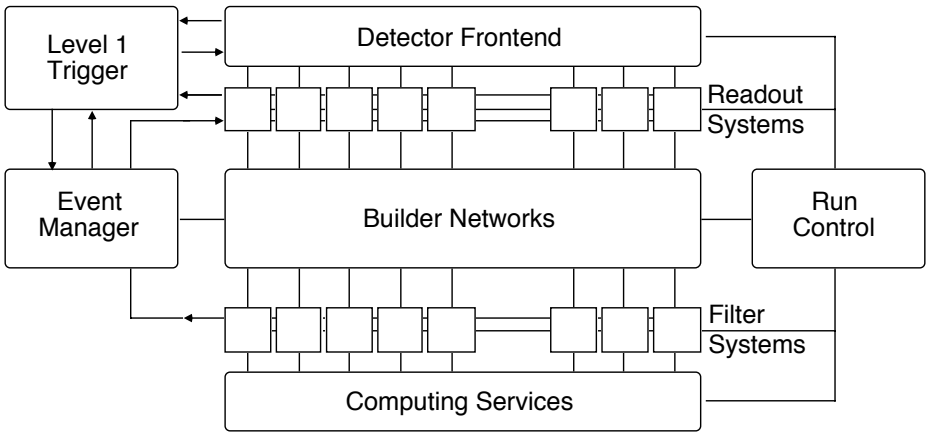


Figure 2.16. CMS Trigger and Data Acquisition System.

Higgs searches, lepton plus jet datasets for top physics, and inclusive electron datasets for calorimeter calibrations. In addition, other samples are necessary for measuring efficiencies in event selection and studying backgrounds. The trigger has to select these samples in real time along with the main data samples.

2.9.1 Level 1 Trigger

The design of the CMS Trigger and Data Acquisition system is illustrated in figure 2.16. At the first level all information about the event is preserved. The first level decision is made, with negligible dead-time, on a subset of the total information available for the events. Made at a fixed time after the interaction has occurred, a first level decision is issued every 25 ns. The L1 trigger system must be able to accept a new event every 25 ns. The L1 pipeline data storage time is 3.2 μs. Since signal propagation delays are included in this pipeline time, the L1 trigger calculations must be done in many cases in less than 1 μs. If the first level trigger generates an accept, the event data are moved or assigned to a buffer for readout and processing by the High Level Triggers.

The limit of 3.2 μs is imposed by the amount of data storage in the tracker and preshower front-end buffers before readout after a L1 accept. The quantity of tracker and preshower data requires an architecture which provides for storage of event data before a L1 accept and readout of the event data (at maximum 100 kHz out of the input rate of 40 MHz bunch crossings) corresponding to the L1 accepts. This architecture prevents use of the tracker data in the L1 trigger decision.

The L1 trigger involves the calorimetry and muon systems as well as some correlation of information from these systems. The L1 decision is based on the presence of local objects such as photons, electrons, muons, and jets, using information from calorimeters, and muon systems in a given element of η-ϕ space. It also employs global sums of E_T and missing E_T. Each of these items is tested against several p_T or E_T thresholds.

The global compilation of this information is used to decide whether to keep (i.e. trigger on) the data from a particular beam crossing. The L1 logic also has the ability to monitor and control trigger rates, hot and dead channels, and other pathological conditions. The maximum design trigger rate of 100 kHz for L1 corresponds to a minimum rejection rate of 10^4

at design luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$. This maximum L1 trigger rate is set by the average time to read information for processing by the HLT and the average time for completion of processing steps in the HLT logic.

2.9.2 High Level Trigger

The CMS Level-1 Trigger System is required to reduce the input interaction rate of 1 GHz to a filtered event rate of 75 kHz. For physics analysis and further event filtering, the data corresponding to each selected event must then be moved from about 512 front-end buffers to a single location. To match the capabilities of the mass storage and offline computing systems, the final output of the experiment should not exceed 100 events per second.

These functions are performed by a system employing a high performance readout network to connect the sub-detector readout units via a switch fabric to the event filter units (which are implemented by a computer farm). The flow of event data is controlled by an event manager system.

In order to optimize the data flow, the filter farm performs event selection in progressive stages by applying a series of High Level Trigger filters. The initial filtering decision is made on a subset of the data, from detector components such as the calorimeter and muon systems. This avoids saturating the system bandwidth by reading out the large volume of tracking data at 75 kHz. It is expected that initial filtering can reduce the event rate by at least one order of magnitude. The remainder of the full event data are only transferred to the farm after passing these initial filters and the final High Level Trigger algorithms are then applied to the complete event.

The High Level Triggers have access to all the information used in L1 since this is stored locally in the L1 trigger crates. Consequently, High Level Triggers can make further combinations and other topological calculations on the digital list of objects transmitted from L1. More importantly, much information is not available on the time scale of the L1 trigger decision. This information is then used in the High Level Triggers. This information includes that from the tracker and the full granularity of the calorimeters. Eventually, the High Level Triggers use the full event data for the decision to keep an event.

The High Level Triggers, implemented as a processing farm that is designed to achieve a rejection factor of 10^3, write up to 100 events/second to mass storage. The last stage of High Level Trigger processing does reconstruction and event filtering with the primary goal of making datasets of different signatures on easily accessed media.

Chapter 3

Reconstruction and Identification of Physics Objects

3.1 Photon Reconstruction and Identification

3.1.1 Photon Reconstruction

Photon candidates are reconstructed from clusters of ECAL channels around significant energy deposits, which are merged into superclusters [83]. The clustering algorithms result in almost complete recovery of the energy of photons that convert in the material in front of the ECAL. In the barrel region, superclusters are formed from five-crystal-wide strips in η centered on the locally most energetic crystal (seed) and have a variable extension in ϕ. In the endcaps, where the crystals are arranged according to an x-y rather than an η-ϕ geometry, matrices of 5×5 crystals (which may partially overlap) around the most energetic crystals are merged if they lie within a narrow road in η.

A supercluster is promoted to a photon candidate if its reconstructed transverse energy is greater than 10 GeV. The topological variable $R_9 = E_{3\times3}/E_{SC}$ (energy sum of 3×3 crystals centered on the most energetic, divided by the energy of the supercluster), is introduced. If $R_9 > 0.94$ the photon is most probably unconverted and the best resolution is obtained from the energy sum of 5×5 matrix. If $R_9 < 0.94$, the corrected supercluster energy is used (see section 3.1.3). The energy weighted mean position of the crystals used to compute the photon energy is assigned to the photon position.

3.1.2 Reconstruction of Conversions

The CMS tracker was designed to achieve a high momentum resolution in the high occupancy environment of the LHC collisions, and this inevitably led to a substantial amount of tracker material. The direct consequence is that a large fraction of photons convert into e^+e^- pairs.

Conversion track pairs are built from a combination of standard tracks reconstructed by using iterative tracking steps [84], electron-specific tracks and conversion ECAL-seeded tracks [85, 86]. A mixture of these types of tracks is preselected with basic quality criteria (nHits>4, $\chi^2 < 10$), and opposite-charge pairs are considered which satisfy the photon conversion topology. The signature which distinguishes photon conversion candidates from massive V^0 particles or from vertexes due to mis-reconstructed tracks is the parallelism of

the conversion electrons at the conversion vertex.

Track pairs are preselected so to remove the prompt tracks by requiring that they have a distance of minimum approach with respect to primary vertex greater than zero and that their angular separation measured as $\Delta \cot(\theta)$ is less than 0.1. Furthermore the z separation of their innermost hits is required to be less than $5\ cm$, and the point at which the tracks are tangent is required to be well contained in the Tracker volume (radius less than $120\ cm$ and z less than $300\ cm$). Track pairs satisfying these criteria are then fitted to a common constrained vertex.

The last steps of the selection require that the vertex fit is successful and that the χ^2 probability of the fit is greater 10^{-6}. Moreover, conversions are also required to have a fitted $p_T > 1\,\text{GeV}$ and to be matched to a photon supercluster (i.e. the ΔR measure between the direction of the supercluster and the conversion to be less than 0.1). Conversion reconstruction is used in the analysis only to help in the identification of the correct Higgs interaction vertex as explained in section 4.3. About one quarter of the events have at least one of the photon reconstructed and selected as a conversion.

3.1.3 Energy Reconstruction

A good understanding of the expected signal shape is crucial as an input to the limit extraction procedure. The width of the reconstructed Higgs mass peak in the low mass range is driven by the detector resolution whenever the correct vertex is taken. It is thus necessary to optimize the agreement between simulation and the observed data, such as the simulated events match the observed detector performance.

Here we describe the different ingredients entering the photon energy calibration. Then we expand on how the supercluster energy calibration is performed using an multivariate (MVA) approach.

Moreover we document the procedure through which data is scaled and simulation is smeared so as to obtain a signal model that is as realistic as possible. It is important to stress that whatever systematic effects exist in the corrections that are applied, the scaling/smearing step will always provide checking and closure in terms of performance matching between data and simulation. Therefore, while we cannot draw conclusions as to the systematic uncertainties of the corrections used, we can ascertain quantitatively by how much they bring us close to simulation, or not. The only genuinely remaining systematic uncertainty is the difference between electrons and photons, since the matching between data and simulation is performed using $Z \to e^+e^-$events, while the analysis concerns $H \to \gamma\gamma$ decays.

There are three main ingredients affecting the determination of the photon energy:

- channel intercalibration;
- transparency loss corrections;
- Clustered energy corrections.

The first two do not depend on the underlying object being reconstructed, while the last one should be considered separately for electrons and photons.

Channel intercalibration. Individual channel calibrations (IC) are derived in situ mainly by equalizing the response to low mass diphoton resonances (π^0, η) across the detector [87]. Supplementary information and cross checks are provided by studying the approximate ϕ-invariance of the energy flow in minimum bias data [87], and the ratio of the energy over the momentum (E/p) of isolated electrons [88].

In figure 3.1 it is possible to appreciate the effect of the individual channel calibration in $Z \rightarrow e^+e^-$events from 2011 data by comparing the blue (uncalibrated) and red curves (calibrated).

Transparency loss corrections. The transparency of the ECAL crystals deteriorates due to irradiation during the LHC running periods, and it recovers when there are no collisions occurring in the detector.

The ECAL response stability in time is achieved using per-channel corrections to compensate for variations of light transmission in the crystals. Laser light is injected in each crystal every 40 minutes and the signal measured by the photon detectors — compared to a reference signal — is used to monitor the light transmission and to derive corrections [89]. During 2012, the channel response varied by about 4% in the barrel and 10% in the endcaps. The time history of reference signals, such as the energy-over-momentum ratio of isolated electrons, or the invariant mass peak of $Z \rightarrow e^+e^-$decays and of π^0 and η decays into photons, is used to monitor the quality of the correction.

In figure 3.1 it can be seen the effect of the transparency loss corrections in $Z \rightarrow e^+e^-$events from 2011 data by comparing the red (calibrated, uncorrected) and black curves (calibrated and corrected).

Supercluster energy correction. The photon energy is computed starting from the raw supercluster energy (adding also the preshower energy in the endcap). In order to obtain the optimal resolution, the raw energy must be corrected for local containment of the shower in the calorimeter, as well as the global containment of the shower for photons which convert and shower in material upstream of the calorimeter. These corrections are computed using a multivariate regression technique originally based on the TMVA Gradient Boosted Decision Tree implementation [90], though substantially optimized [91].

The regression is trained on prompt photons in Montecarlo (from the photon + jets sample) using the ratio of generator level photon energy to the raw supercluster energy as the target variable.

The input variables are the global η and ϕ coordinates of the supercluster, a collection of shower shape variables, and a set of local cluster coordinates. The shower shape variables included are the R_9 of the supercluster, the ratio of the 5×5 crystal energy to the raw supercluster energy, the energy weighted η-width and ϕ-width of the supercluster, the number of clusters merged to form the supercluster, and the ratio of hadronic energy behind the supercluster to the electromagnetic energy of the cluster. In the endcap, the ratio of preshower energy to raw supercluster 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 η and ϕ position of the supercluster, drive the degree of global containment correction predicted by the regression.

In the barrel, the η/ϕ index of the seed crystal as well as the position of the seed cluster with respect to the crystal center are also included. These variables, together with the seed

cluster energy ratios provide information on the amount of energy which is likely to be lost in crystal and module gaps and cracks, and drive the level of local containment corrections predicted by the regression. Although the global and local containment are conceptually different effects, the required corrections are allowed to be correlated in the regression in order to account for the fact that a showering photon is not incident at a single point on the calorimeter face, and is therefore relatively less affected by the local containment.

Finally the number of primary vertexes and median energy density ρ parameter[1] in the event are included in order to correct residual energy scale effects from pile-up.

The performance of the regression for photons in Montecarlo is shown in terms of the $H \to \gamma\gamma$ mass resolution in the four untagged event categories used in the analysis in figure 3.2, comparing to the default photon energy in the reconstruction.

3.1.4 Resolving data and simulations discrepancies

Having corrected the data as much as possible, some discrepancies with respect to simulation are still present. Based on $Z \to e^+e^-$ data and simulated events, these discrepancies are dealt with by correcting the energy scale in data and by then determining the smearing needed to apply to the simulated samples so as to have the best match between data and simulation.

Energy scale correction. The supercluster energy scale is tuned and corrected varying the scale in the data to match the prediction in $Z \to e^+e^-$ events. An analytic fit to the Z invariant mass peak, build with supercluster energies, is performed using a convolution of a Breit-Wigner function with a Crystal Ball (CB) function [93]. Data and Montecarlo distribution are fitted separately, and the fit results are compared to extract the scale offset. In the fit, the Breit-Wigner parameters are fixed to the PDG [94] values: $m_Z = 91.188\,\text{GeV}$ and $\Gamma_Z = 2.495\,\text{GeV}$. On the contrary, the CB parameters describing the calorimeter resolution effects and bremsstrahlung losses are free parameters of the fit.

The data-Montecarlo difference is time dependent; this time dependence is not the same in different pseudorapidity region, but it is very similar for showering and non-showering electrons[2]. Therefore, the following two-step procedure has been implemented: in the first step the scale is extracted per run-range and per pseudorapidity region (four bins, two in the barrel and two in the endcap) to take into account imperfect transparency corrections.

In the second step the residual data-Montecarlo difference is investigated by looking separately at the two R_9 bins (non-showering and showering electrons) in every pseudorapidity region, to factorize the effect of the material in front of the calorimeter.

The final energy scale correction is then derived as the product of the two corrections in n(run-range)$\times 4$(pseudorapidity region)$\times 2(R_9$ categories). The relative scale difference is defined as:

$$\Delta P = \frac{\Delta m_{data} - \Delta m_{MC}}{m_Z} \tag{3.1}$$

where Δm_{data} and Δm_{MC} are the CB parameters measuring the differences in data and

[1] The ρ is evaluated for each event by taking the median value of the following distribution: $\rho = \text{median}\left[\left\{\frac{p_{Tj}}{A_j}\right\}\right]$, where the index j runs over all the jets reconstructed in the event, and A_j is the jet area. A detailed description can be found in [92].

[2] We define an electron to be showering if it irradiates at least one detectable photon before reaching the ECAL.

Montecarlo between the fitted Z invariant mass peak position and the PDG m_Z value. The quantity $(1 - \Delta P)$ ranges up to 1.5% in the barrel, and up to 4.9% in the endcap, and it is applied as a multiplicative factor to data to correct the photon energy scale.

MC Energy Smearing. The electron supercluster energy is modified by applying a random Gaussian multiplicative factor centered in $(1 + \Delta P)$ and with a $\Delta\sigma$ resolution, where ΔP is the energy scale correction defined in equation 3.1, and $\Delta\sigma$ is the additional constant term in the energy resolution.

After defining n exclusive electron categories, the $n_{ind} = [n(n+1)/2]$ independent invariant mass distributions of $Z \to e^+e^-$ are built in data and simulation. The method is based on the simultaneous minimization of the data-Montecarlo discrepancy in the n_{ind} invariant mass distributions as a function of the $2n$ parameters $(\Delta P_i, \Delta\sigma_i)$. The motivation for building a combined likelihood of all the n_{ind} distributions is to include in the smearing computation also the events where the two electrons lie in different categories. We use 9 electron categories, accounting for different η regions of barrel and endcap, for different R_9 intervals and for the distance between the cluster and the ECAL module borders.

Two examples of mass distributions are visible in Figure 3.3: the left plot shows the $Z \to e^+e^-$ with both electrons belonging to the (barrel, $|\eta|$<1, R_9>0.94) category, while the right plot shows one mixed category: one (barrel, $|\eta|$<1, R_9>0.94) and one (endcap, $|\eta|$>2, R_9>0.94) electron.

The final smearing factors range from 1.1% to 2.4% in the barrel, and from 3.3% to 6.1% in the endcaps.

3.1.5 Photon Identification

Photon identification is performed by applying cuts on a set of six discriminating variables. Cut values are optimized separately for each of four categories defined in terms of pseudo-rapidity and R_9. These categories have significantly differing levels of background and mass resolution, and the use of categories thus provides increased sensitivity.

The following variables are used to distinguish isolated photons originating from the primary interaction from the background due to low multiplicity jets with high electromagnetic content:

1. **Relative combined isolation using selected event vertex.**
   An isolation sum is calculated as follows:

$$\sum Iso = Iso^{Trk} + Iso^{ECAL} + Iso^{HCAL},$$

   where:

   - Iso^{Trk} is the scalar sum of the transverse momenta of tracks which are consistent with originating from the primary vertex (within $\pm 1\ cm$ along the beam direction and within $\pm 0.1\ cm$ transverse to the beam direction) and lie within a hollow cone of size $\Delta R_o < 0.3$ centered around a line joining the selected primary vertex to the ECAL supercluster, excluding an inner cone ($\Delta R_i = 0.02$) in order to avoid including the momentum of conversion tracks.
   - Iso^{ECAL} is computed as the transverse energy sum of ECAL energy deposits in crystals located within a cone of size $\Delta R_o < 0.3$ (approximately 1250 crystals),

centered around the supercluster position, excluding an inner veto cone ($\Delta R_i = 3.5$ crystals) and eta-slice ($\Delta\eta = 2.5$ crystals) in order to exclude the footprint of the signal photon, which can be extended in the ϕ direction in the case of converted photons.

- Iso^{HCAL} is the sum of the energies of HCAL towers whose centers lie within an annular region of outer radius $\Delta R_o = 0.4$ and inner radius $\Delta R_i = 0.15$, centered on the ECAL supercluster position.

For each of the isolation sums defined above, the energy deposited within the isolation cone is contaminated by energy from pile-up and from the underlying event. Since the contamination increases with the number of pile-up vertexes, the efficiency of the isolation cut decreases with increasing pile-up. In order to maintain high efficiency under high pile-up conditions, the contribution to $\sum Iso$ from pile-up and the underlying event is estimated on an event-by-event basis as the product of the measured energy density ρ for the event determined using the FastJet algorithm [92, 95, 96], and an effective area A_{eff} corresponding to the isolation cone excluding veto regions.

The pile-up corrected isolation sum is then given by:

$$\sum Iso^{PUcorr} = \sum Iso - \rho A_{eff}$$

The isolation sum is then scaled by $p_T(\gamma)/(50\text{ GeV})$, where $p_T(\gamma)$ is the transverse energy of the photon determined using the selected primary vertex. In this way scaling factor is close to 1 for typical signal photon transverse energies. The relative isolation is thus given by:

$$Iso^{rel} = \frac{\sum Iso^{PUcorr}}{p_T(\gamma)/(50\text{ GeV})} \tag{3.2}$$

2. **Relative combined isolation using event vertex giving highest Iso^{Trk}.**
The isolation sum is calculated as above, except for the following:

   - Iso^{Trk} is computed for each reconstructed primary vertex and the largest value is used.
   - For all three sub-detector isolation (Iso^{Trk}, Iso^{ECAL} and Iso^{HCAL}), the outer cone size ΔR_o is set to 0.4.
   - The value of the effective area A_{eff} is enlarged.

   This definition of isolation adds discrimination since in the previous definitions Iso^{Trk} gives no discrimination in the case that the wrong primary vertex is selected, although it is more powerful in the case that the correct vertex is selected. In this way the different definitions are complementary.

3. **Relative track isolation using selected event vertex.**
Since track isolation is the most discriminating of the three sub-detector isolation, a cut is additionally applied on relative isolation defined using track isolation only:

$$Iso^{rel,Trk} = \frac{\sum Iso^{Trk}}{p_T(\gamma)/50\text{ GeV}}$$

No pile-up subtraction is required since only tracks consistent with the selected primary vertex are included in the sum.

4. $\boldsymbol{H/E}$
The ratio of hadronic energy to electromagnetic energy is calculated as the ratio of the sum of HCAL tower energies within a cone of size $\Delta R < 0.15$ centered on the ECAL supercluster position, to the energy of the supercluster. Due to the 25 radiation length thickness of the ECAL crystals, isolated photons have a value close to or equal to zero.

5. $\boldsymbol{\sigma_{i\eta i\eta}}$
The transverse shape of the electromagnetic cluster is computed with logarithmic weights and is defined as

$$\sigma^2_{i\eta i\eta} = \frac{\sum_i^{5\times5} w_i(\eta_i - \bar{\eta}_{5\times5})^2}{\sum_i^{5\times5} w_i}, \quad w_i = \max\left(0, 4.7 + \ln\frac{E_i}{E_{5\times5}}\right),$$

where E_i and η_i are the energy and pseudorapidity of the i^{th} crystal within the 5×5 electromagnetic cluster, and $E_{5\times5}$ and $\eta_{5\times5}$ are the energy and η of the entire 5×5 cluster. The value of $\sigma^2_{i\eta i\eta}$ tends to be smaller for single isolated photons (including converted photons, since the cluster is spread in the ϕ direction only), than for the background which is dominated by jets consisting of multiple π^0s each decaying to two photons.

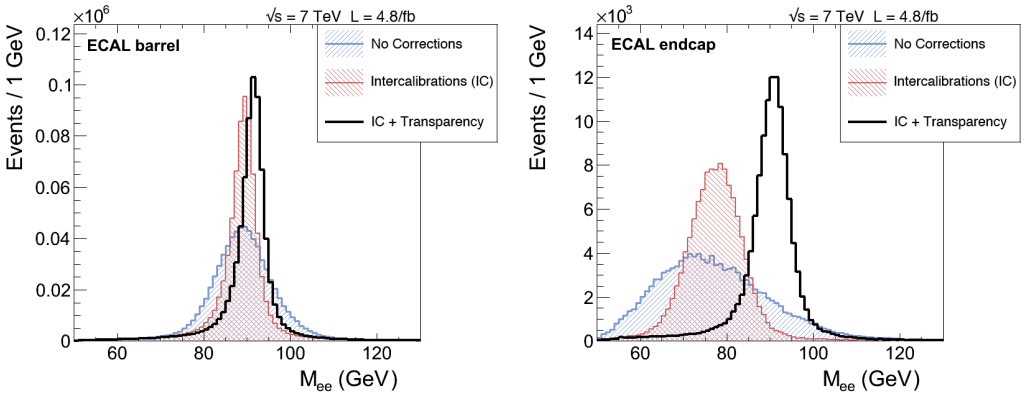
6. **Minimum threshold on $\boldsymbol{R_9}$**
A minimum threshold on R_9 (defined at the beginning of this section) is applied to photons in the ECAL endcaps in order to exclude very poorly reconstructed photons.

To avoid misidentifying an electron as a photon, a conversion-safe electron veto is applied. It consists of removing photons if there is a electron matching to the photon supercluster, with no missing hits and having no matching reconstructed conversion.

The photon identification cuts are set to get the best compromise between signal efficiency and fake rate. This means that the cuts are tighter in the low R_9 category than in the high R_9 category, and tighter in the endcap than in the barrel. For this cut setting procedure, subleading photon candidates are used, taking $H \to \gamma\gamma$ ($m_{\mathrm{H}} = 120\,\mathrm{GeV}$) events as signal, and γ+jet events as background. After optimizing many different sets of cuts corresponding to different S/B ratio, we decide to use the *super-tight* working point whose cuts are listed in table 3.1. The chosen selection criteria are applied to both legs of the diphoton pair. The efficiency of the photon ID variables as a function of the number of reconstructed vertexes in shown in figure 3.4 in the four photon categories. The single-photon efficiency evaluated on a simulated signal samples ranges from 97% to 83% going from the first to last category. The fake rate evaluated on a simulated QCD sample ranges from 8% to 53%. The diphoton efficiencies evaluated on both Montecarlo an data are discusses thoughtfully in the next sections.

3.1.6 Photon Preselection

All prompt and non-prompt photons within acceptance are required to pass a preselection. This selection is conceived to keep the common phase space of data passing the online HLT filter, and Montecarlo events passing the generator level filters.

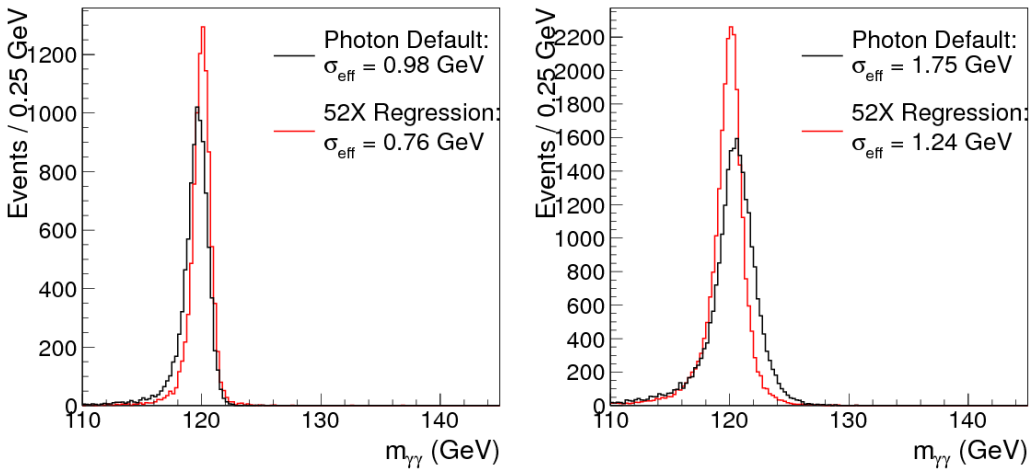


(a) Barrel (b) Endcap

Figure 3.1. Z boson invariant mass in a $Z \to e^+e^-$ data sample collected in 2011. The Z peak is used as a reference to show the role of the individual channel calibration (IC) and transparency loss corrections (Laser): uncalibrated and uncorrected (no IC, no Laser) data in blue, calibrated but uncorrected (IC, no Laser) data in red, and calibrated and corrected (IC, Laser) data in black.

Table 3.1. Thresholds of the photon identification selection criteria in the four photon categories. The selection is applied to both the leading and subleading photons.

Variable	Barrel		Endcap	
	$R_9 > 0.94$	$R_9 < 0.94$	$R_9 > 0.94$	$R_9 < 0.94$
Rel. Comb. Iso. (selected vtx.)	3.8 GeV	2.2 GeV	1.77 GeV	1.29 GeV
Rel. Comb. Iso. (worst vtx.)	11.7 GeV	3.4 GeV	3.9 GeV	1.84 GeV
Rel. Track Iso. (selected vtx.)	3.5 GeV	2.2 GeV	2.3 GeV	1.45 GeV
$\sigma_{i\eta i\eta}$	0.0106	0.0097	0.028	0.027
H/E	0.082	0.062	0.065	0.048
R_9	0.94	0.36	0.94	0.32



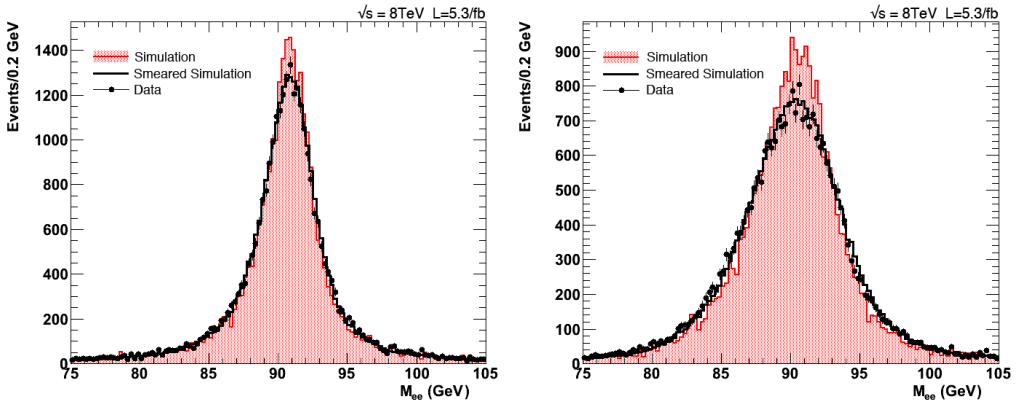
(a) Both photons in barrel, High R_9

(b) Both photons in barrel, Low R_9

(c) At least one photon in endcap, High R_9

(d) At least one photon in endcap, Low R_9

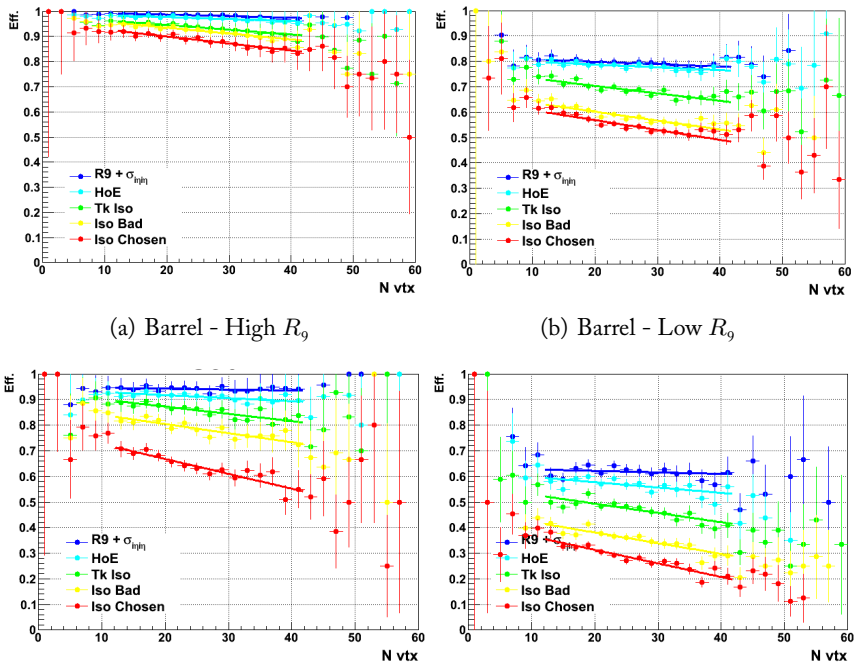
Figure 3.2. Comparison of the diphoton mass resolution in the four analysis categories using the $H \to \gamma\gamma$ Montecarlo ($m_H = 120\,\text{GeV}$) The default reconstructed photon energy is shown in black, and the full regression-corrected energy is shown in red.



(a) Both electrons in the same (*barrel, high*) R_9 category.

(b) Electrons in different categories, one being (*endcap, low* R_9).

Figure 3.3. Invariant mass distribution of $Z \to e^+e^-$ events in different categories. Red histogram is the standard Montecarlo simulation, while the black empty histogram is the optimal smearing of Montecarlo energies to match data distributions.



(a) Barrel - High R_9

(b) Barrel - Low R_9

(c) Endcap - High R_9

(d) Endcap - Low R_9

Figure 3.4. Photon Identification variables as a function of the number of vertexes in the four photon categories. Within the legend, "Iso Good" stands for Relative Combined Isolation with respect to the selected vertex; "Iso Bad" stands for Relative Combined Isolation with respect to the worst vertex; "HoE" stands for H/E.

The motivation is to be sure that the photon identification optimization is performed in a region where Montecarlo can properly describe the behavior of data and, at the same time, to discard events too close to the trigger turn-on.

The variables used for preselection are therefore built to be similar to those used in the HLT trigger (a detailed description of the HTL can be found in [97]) and in the Montecarlo filter. They consist in photon E_T thresholds, requirements on the photon shower shape and on several isolation criteria. Moreover, to avoid misidentifying an electron as a photon, a conversion-safe electron veto is applied.

Since the trigger thresholds evolved with time, to determine a fixed set of cut values we consider the tightest trigger combination which covers the common phase space of all the available trigger paths. This choice also allows to strongly simplify the trigger efficiency calculation.

The preselection variables and the relative threshold for barrel and endcap are listed in table 3.2. All the variables have been already presented in the previous section (3.1.5) except for the isolation ones. They are simpler with respect to those used in the standard photon id, since they have to reflect the selection performed online. The definition of such variables follows:

1. E_T-corrected Ecal Isolation: $Iso^{ECAL} - 0.012 \times E_T(\gamma)$
2. E_T-corrected Hcal Isolation: $Iso^{HCAL} - 0.005 \times E_T(\gamma)$
3. E_T-corrected Tracker Isolation: $Iso^{Trk} - 0.002 \times E_T(\gamma)$
4. Charged Isolation: the sum of p_T of the charged particle flow candidates[3] within $0.02 < \Delta R < 0.2$.

Table 3.3 shows preselection efficiencies measured using tag and probe with $Z \rightarrow e^+e^-$ events for data (ε_{DATA}), Montecarlo (ε_{MC}) and the ratio $\varepsilon_{DATA}/\varepsilon_{MC}$, for the four photon categories.

The effect on the signal photon of the electron veto requirement has been evaluated from data in a sample of $Z \rightarrow \mu\mu\gamma$ events, and compared with events selected in a simulated sample of DY events. The photon from $Z \rightarrow \mu\mu\gamma$ in a mass window 70 to 110 GeV is further subjected to the preselection criteria described above, except the electron veto, and used as probe. The electron veto efficiency is measured as the ratio of "passing" photons divided by the total number of preselected photons. Table 3.4 lists the results in data and simulation and their ratios. Values are given for the four categories used in the analysis.

3.1.7 Photon Identification Performance

Figure 3.5 shows the signal efficiency of the *super-tight* working point for photons from simulated Higgs bosons with $m_H = 124$ GeV, as a function of pseudorapidity and p_T, for each of the four photon categories. The efficiency of the same selection is measured in data using the tag and probe technique. $Z \rightarrow e^+e^-$ events are used to determine the combined efficiency of all the cuts listed in table 3.1 with the exception of the electron veto cut whose efficiency is measured using $Z \rightarrow \mu\mu\gamma$ events. In table 3.5 are listed the efficiencies measured using tag and probe with $Z \rightarrow e^+e^-$ events for data (ε_{DATA}) and Montecarlo (ε_{MC}),

[3] The particle flow technique is a reconstruction algorithm that aims to identify every visible particle of the event by considering information coming from all the sub-detectors [98–100].

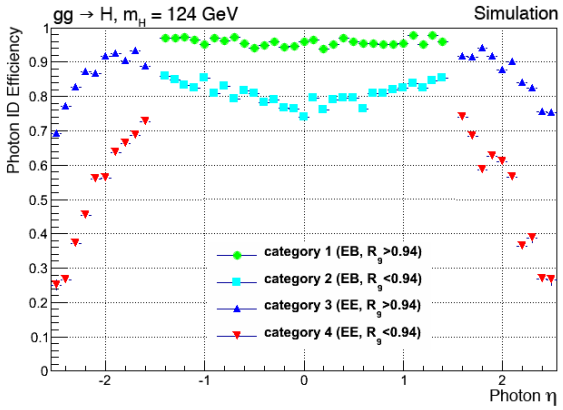
Table 3.2. Preselection variables and thresholds in the four photon categories.

Category		Variable	Criterion
Barrel	$R_9 \leq 0.94$	H/E	< 0.075
	$R_9 > 0.94$		< 0.082
Endcap	$R_9 \leq 0.94$		< 0.075
	$R_9 > 0.94$		< 0.075
Barrel	$R_9 \leq 0.94$	$\sigma_{i\eta i\eta}$	< 0.014
	$R_9 > 0.94$		< 0.014
Endcap	$R_9 \leq 0.94$		< 0.034
	$R_9 > 0.94$		< 0.034
Barrel & Endcap	$R_9 \leq 0.94$	Corrected Iso^{ECAL}	$< 4\,\text{GeV}$
	$R_9 > 0.94$		$< 50\,\text{GeV}$
	$R_9 \leq 0.94$	Corrected Iso^{HCAL}	$< 4\,\text{GeV}$
	$R_9 > 0.94$		$< 50\,\text{GeV}$
	$R_9 \leq 0.94$	Corrected Iso^{Trk}	$< 4\,\text{GeV}$
	$R_9 > 0.94$		$< 50\,\text{GeV}$
	$R_9 \leq 0.94$	Charged Isolation	$< 4\,\text{GeV}$
	$R_9 > 0.94$		$< 4\,\text{GeV}$

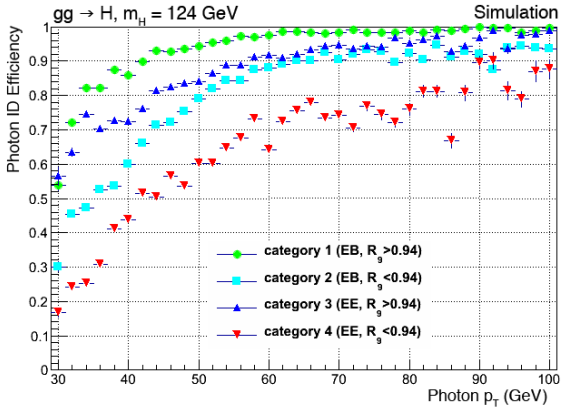
and the ratio $\varepsilon_{DATA}/\varepsilon_{MC}$, for the four photon categories. Efficiencies are computed with respect to the preselection described in Section 3.1.6. Analysis selection efficiencies are determined from signal Montecarlo and are corrected to account for differences between Monte Carlo and data using the above ratio.

Table 3.3. Photon identification efficiencies measured in the four photon categories using tag and probe with $Z \to e^+e^-$ events (for all cuts except electron rejection).

		Data			Simulation		Ratio	
		ε	σ_{STAT}	σ_{SYST}	ε	σ_{STAT}	ε	σ
Barrel	$R_9 > 0.94$	0.9894	0.0002	0.0030	0.9916	0.0001	0.998	0.003
	$R_9 < 0.94$	0.9327	0.0006	0.0055	0.9369	0.0003	0.996	0.006
Endcap	$R_9 > 0.94$	0.9832	0.0007	0.0090	0.9771	0.0002	1.006	0.009
	$R_9 < 0.94$	0.9298	0.0014	0.0170	0.9298	0.0003	0.999	0.018



(a) Efficiency *vs* photon pseudorapidity



(b) Efficiency *vs* photon transverse momentum

Figure 3.5. Photon Identification efficiency measured in a Higgs signal sample ($m_H = 124\,\text{GeV}$) for each of the four photon categories.

Table 3.4. Efficiency of the conversion-safe electron veto applied at the preselection level, measured in the four photon categories using tag and probe with $Z \to \mu\mu\gamma$ events. The data to simulation ratio is also shown with its error. The efficiency is built using for the denominator photons passing all the preselection cuts except the electron veto; the numerator is the number of photons passing also the electron veto.

		Data		Simulation		Ratio	
		ε	σ	ε	σ	ε	σ
Barrel	$R_9 > 0.94$	0.999	0.001	1.000	0.000	0.999	0.001
	$R_9 < 0.94$	0.982	0.004	0.990	0.003	0.992	0.005
Endcap	$R_9 > 0.94$	0.990	0.004	0.999	0.002	0.990	0.004
	$R_9 < 0.94$	0.966	0.008	0.985	0.006	0.981	0.010

Table 3.5. Photon identification efficiencies measured in the 4 photon categories using tag and probe with $Z \to e^+e^-$ events (for all cuts except electron rejection) with respect to the preselection efficiencies.

Category		Data			Simulation		Ratio	
		ε	σ_{STAT}	σ_{SYST}	ε	σ_{STAT}	ε	σ
Barrel	R_9>0.94	0.9086	0.0007	0.0025	0.9064	0.0003	1.002	0.003
	R_9<0.94	0.7051	0.0008	0.0080	0.7109	0.0003	0.992	0.011
Endcap	R_9>0.94	0.7623	0.0012	0.0055	0.7464	0.0008	1.021	0.008
	R_9<0.94	0.4554	0.0015	0.0115	0.4435	0.0006	1.027	0.026

3.2 Jet Reconstruction

A colored energetic particle, such as a quark or a gluon expelled in a high-energy proton-proton collision, due to the intensity of the strong field potential is energetically incentivated into losing its color charge in the formation of stable, colorless configurations. This is done by multiple radiations of gluons which excite the vacuum producing quark-antiquark pairs, and the quarks eventually combine themselves in the formation of mesons and baryons. This process is known as hadronization. Quadrimomentum conservation laws applied to the initial parton imply that the hadronization products will have a quasi collinear configuration, in the parton's direction: these sprays of particles are visible in modern detectors, and are commonly referred to as *jets*.

3.2.1 Response and Resolution

The aim of jet reconstruction is to measure the momentum of the colored parton which initiated the hadronization process. In order to do so, final state particles, visible in the detector, have to be grouped together, through the choice of an appropriate jet algorithm, as will be shown in the following section. The same algorithm is applied to reconstructed objects and to generator-level particles, giving rise respectively to reconstructed and generator jets. Each reconstructed jet can be matched to its corresponding generator jet topologically, by choosing the closest generator jet on the η-ϕ plane.

Two variables are commonly employed to measure jet reconstruction performance: the

jet response and resolution. The response variable is defined on a jet-by-jet basis as the ratio between the transverse momentum of the reconstructed jet and that of its matched generator jet:

$$\mathcal{R} = \frac{p_T^{reco}}{p_T^{gen}} \tag{3.3}$$

This is defined only at Monte-Carlo level, as it accesses the generator information. The average value of this variable, $\langle \mathcal{R} \rangle$, is an estimator of the response of a given jet reconstruction strategy. The jet resolution, instead, is usually defined as the width of the $\mathcal{R}$ variable distribution, divided by the response.

3.2.2 Reconstruction Methods

Three different methods of jet reconstruction are employed by CMS, characterized by the way that the sub-detector inputs are used during the jet finding procedure: calorimeter jets (Calo-Jets), jet-plus-tracks jets (JPT-Jets) and particle flow jets (PF-Jets).

Calorimeter jets are reconstructed using energy deposits in the ECAL and HCAL cells, combined into calorimeter towers as inputs. A calorimeter tower consists of one or more HCAL cells and the geometrically corresponding ECAL crystals. In the barrel region of the calorimeters, the unweighted sum of one single HCAL cell and 5×5 ECAL crystals form a projective calorimeter tower. The association between HCAL cells and ECAL crystals is more complex in the endcap regions of the ECAL. Beyond the coverage of the ECAL ($|\eta| > 3.0$), each calorimeter tower corresponds to one HCAL cell.

The **Jet-Plus-Tracks** (JPT) algorithm [101] corrects the energy and the direction of a calorimeter jet. It exploits the excellent performance of the CMS tracking detectors to improve the p_T response and resolution of calorimeter jets (tracking coverage extends up to $|\eta| \sim 2.4$). Charged particle tracks are associated with each calorimeter jet based on spatial separation in η-ϕ between the jet axis and the track momentum measured at the interaction vertex. The associated tracks are classified as in-cone tracks if their projection onto the surface of the ECAL falls within the jet cone. Conversely, if they are bent outside the cone by the magnetic field, then they are called out-of-cone tracks. The momenta of both in-cone and out-of-cone tracks are then added to the energy of the associated calorimeter jet. For in-cone tracks the expected average energy deposition in the calorimeters is subtracted, based on the momentum of the track and the hypothesis that it originates from a charged pion.

The **Particle Flow** (PF) algorithm [98] aims to reconstruct, identify and calibrate each individual particle in the event by combining the information from all CMS sub-detector systems. PF particles are reconstructed as a combination of charged particle tracks and clusters in the electromagnetic and hadronic calorimeters, as well as signals in either of the two CMS preshower detectors and the muon system. Depending on which of the detector systems contribute to a single particle, it is identified as either an electron (track + ECAL), muon (track + ECAL + HCAL + Muon System), photon (ECAL), charged hadron (track + ECAL + HCAL), or neutral hadron (HCAL). The algorithm employs strategies to handle ambiguities stemming from overlapping detector signals to avoid information double-counting. Based on the particle type, the energy of each particle is calibrated. Charged hadrons are treated under the assumption that they are pions. As a result of the PF reconstruction, the inputs to the jet clustering are almost fully calibrated and the resulting higher

level objects (jets) require small a posteriori energy corrections.

CMS employs several jet clustering algorithms, e.g. anti-k_T [102], k_T [103, 104] and SIS-CONE [105], with different choices for the jet size parameter R. Jets in the studies presented here are reconstructed with the anti-k_T algorithm, $R = 0.5$, provided by an interface of the CMS software to the FASTJET package [106].

3.3 Missing Transverse Energy Reconstruction

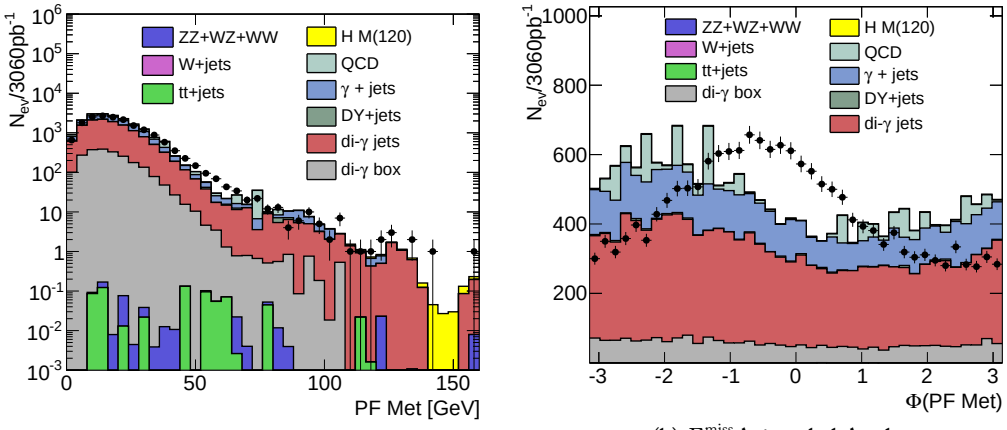
In this analysis the $\not\!\!E_T$ variable is built summing the transverse energy of all the particle-flow (PF) objects in the event:

$$\overrightarrow{\not\!\!E_T^{miss}} = -\sum_n (E_n \sin\theta_n \cos\phi_n \,\hat{\imath} + E_n \sin\theta_n \sin\phi_n \,\hat{\jmath}) \tag{3.4}$$

where the PF technique is a reconstruction algorithm already described in the previous sections that aims to identify every visible particle of the event by considering information coming from all the sub-detectors [98–100].

The $\not\!\!E_T$ is one of the most sensitive variable to detector noise, imperfect alignment between sub-detectors and imprecise jet energy reconstruction. One way to check that the $\not\!\!E_T$ reconstruction is under control, is to compare the fake $\not\!\!E_T$ spectrum in data and Montecarlo. Given that the two do not agree, several corrections are implemented to account for all the systematic effects distorting the $\not\!\!E_T$ shape. The correction factors target mostly the energy of the jets, which need to be corrected before entering the $\not\!\!E_T$ computation.

It's useful to recall here two jet-related quantities that have been introduced in the previous section: the jet response and the jet resolution. The jet response is defined as the ratio $\mathcal{R} = p_T/p_T^{gen}$, where p_T^{gen} is the transverse momentum of the jet clustered from stable particles before interacting with the detector (generator-level jet), and the numerator (p_T) refers to the transverse momentum of a reconstructed jet (detector-level jet). For a fixed p_T^{gen}, the distribution of $\mathcal{R}$ follows a Gaussian in good approximation, and the jet p_T resolution is defined as the standard deviation of the response distribution.



(a) $\not\!\!E_T$ Magnitude (b) $\not\!\!E_T$ Azimuthal Angle

Figure 3.6. Uncorrected Particle Flow $\not\!\!E_T$ distributions.

Two variables are used to compare E_T^{miss} between data and simulation: the E_T^{miss} magnitude and the E_T^{miss} direction $\phi(E_T^{miss})$. The former (figure 3.6(a)) shows that the E_T^{miss} distribution in data is wider and slightly shifted with respect to the Montecarlo one. To let the simulation correctly describe the width of the data distribution, we apply a smearing factor to the Montecarlo jet energy, because in [107] we find the that the simulated jet resolution is underestimated. To account for the offset in the E_T^{miss} scale — caused by the jet mean response in data not being equal to one as a function of η [108] — we choose to apply a scale factor to the jet energy in data depending on the jet pseudorapidity.

The $\phi(E_T^{miss})$ distribution in figure 3.6(b) shows that both data and Montecarlo distributions are not ϕ-symmetric. The asymmetry is found to be caused by a systematic shift in the E_T^{miss} vector components, and the correction we apply aims at flattening the $\phi(E_T^{miss})$ distribution.

The derivation and implementation of the three correction factors (smearing, shifting, scaling) is described in the next sections.

E_T^{miss} Smearing

The smearing of the simulated E_T^{miss} distribution is performed by smearing the energy of the jets involved in the E_T^{miss} computation. Two different procedures are implemented: the first deals with reconstructed jets matching with the generator-level information; the second deals with fragmentation jets not linked to any generator-level parton. Both of them take advantage of the jet studies performed at the beginning of the data taking, when the low instantaneous luminosity allowed to perform low-p_T QCD studies. And both of them aim at modifying the reconstructed jet energy in the range allowed by the jet energy resolution. The latter is not a fixed value, but it varies as a function of the jet pseudorapidity — reflecting the differences in terms of instrumental noise and background processes between the central and the forward regions of the detector — and as a function of the jet p_T.
The smearing factors in the two cases are computed as follows:

1. **Matching.** When the jet is matched to the parton-level information, the smearing factor is factorized into a p_T-dependent and a η-dependent term. Since the true transverse momentum (p_T^{gen}) is known, the p_T-dependent term consists in the relative difference between the reconstructed and the generated momentum:

$$\frac{p_T^{reco} - p_T^{gen}}{p_T^{reco}}.$$

   The η-dependent term ($\mathfrak{R}(\eta)$) is computed using the asymmetry method [109] in $\gamma +$ jet events. The method consists in selecting events where the momentum of the two objects is fully balanced, and it exploits the excellent energy resolution of the ECAL to measure the photon energy and predict the jet energy. The distribution of the ratio between the jet and the photon energy quantifies the jet resolution. The $\mathfrak{R}(\eta)$ values as a function of the jet η are reported in table 3.6. The overall correction $f(\eta, p_T)$ is summarized in equation 3.5.

$$\begin{cases} p_T^{smeared} = p_T^{reco} \times (1 + f(\eta, p_T)) \\ f(\eta, p_T) = (\mathfrak{R}(\eta) - 1) \times \dfrac{p_T^{reco} - p_T^{gen}}{p_T^{reco}} \end{cases} \tag{3.5}$$

Table 3.6. Expected jet energy resolution.

Pseudorapidity Region	Expected Jet Resolution
$0.0 - 1.1$	$1.07 \pm 0.020^{+0.02}_{-0.033}$
$1.1 - 1.7$	$1.10 \pm 0.031^{+0.03}_{-0.039}$
$1.7 - 2.3$	$1.07 \pm 0.048^{+0.05}_{-0.047}$
$2.3 - 5.0$	$1.18 \pm 0.062^{+0.04}_{-0.072}$

2. **No Matching.** When the jet is not matched, the η dependency of the jet resolution is still described by $\mathfrak{R}(\eta)$, but a new term is needed for the p_T-dependent term. We choose to use a normally distributed variable g, with null average and variance $\sigma_E(\eta\,, p_T)$, where σ_E is the expected jet-p_T resolution and it is defined as follows:

$$\sigma_E(\eta, p_T) = \mathcal{S}(\eta, p_T) \times \sqrt{\mathcal{R}^2(\eta) - 1}.$$

The term $\mathcal{S}(\eta, p_T)$ is the outcome of a fully differential (η, p_T) study using the asymmetry method described above in dijet and $\gamma +$ jet events [107]. Figures from 3.7 to 3.9 are retrieved from [110], and they show the p_T dependence of the jet energy resolution in several η bins. We use these plots as a look-up table: the η value discriminates which plot to use, the p_T value determines the position on the x axis, and $\mathcal{S}(\eta, p_T)$ is the value of the expected jet resolution. The overall correction is described in equation 3.6

$$\begin{cases} p_T^{smeared} = p_T^{reco} \times (1 + g) \\ g \sim \text{Gauss}\,(0, \sigma_E(\eta, p_T)) \\ \sigma_E(\eta, p_T) = \langle \mathcal{S}(\eta, p_T) \rangle \times \sqrt{\mathcal{R}^2(\eta) - 1} \end{cases} \tag{3.6}$$

E_T^{miss} Shift

The fake E_T^{miss} distribution is dominated by the detector noise. If the noise is uniform in ϕ, the E_T^{miss} projection along the x and y directions (the two axes spanning the plane transverse to the beam line) follow a Gaussian distribution with null average. On the contrary, the plot in figure 3.6(b) shows a large ϕ asymmetry, suggesting a possible shift of the E_T^{miss} vector.

To test this hypothesis, a Montecarlo sample simulating the Drell-Yan (DY) process is considered. The DY has negligible E_T^{miss} coming from real physics processes, hence it is a good candidate to study fake E_T^{miss}. We define

$$\begin{cases} E_x^{miss} &= E_T^{miss} \times \cos\left(\phi(E_T^{miss})\right) \\ E_y^{miss} &= E_T^{miss} \times \sin\left(\phi(E_T^{miss})\right) \end{cases} \tag{3.7}$$

and figure 3.10 shows the E_x^{miss} and E_y^{miss} distribution of the DY sample, where a Gaussian fit highlights the average of the two distributions.

Given the difference of the E_x^{miss} and E_y^{miss} mean value from zero, we test a correction based on recomputing the E_T^{miss} vector by shifting its component of the amount needed to cancel out this difference. But correcting for such a shift turns out to be not sufficient.

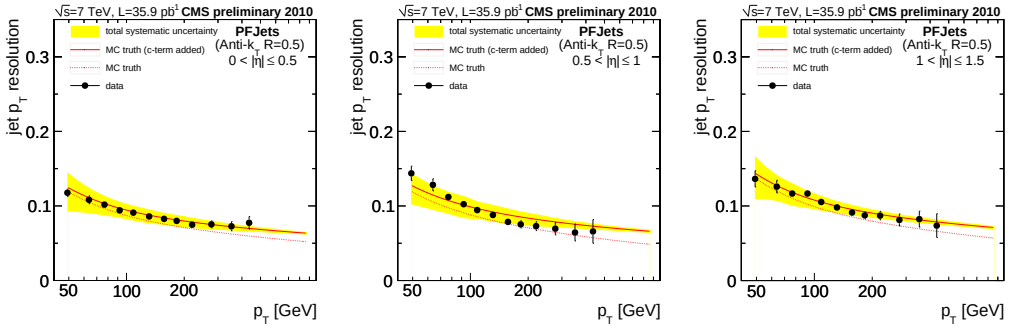


Figure 3.7. Scaled Montecarlo truth resolution compared to data for jets reconstructed with the PF algorithm in $\mathbf{0.0 \leq |\eta| < 0.5}$ (*left*); $\mathbf{0.5 \leq |\eta| < 1.0}$ (*center*); $\mathbf{1.0 \leq |\eta| < 1.5}$ (*right*).

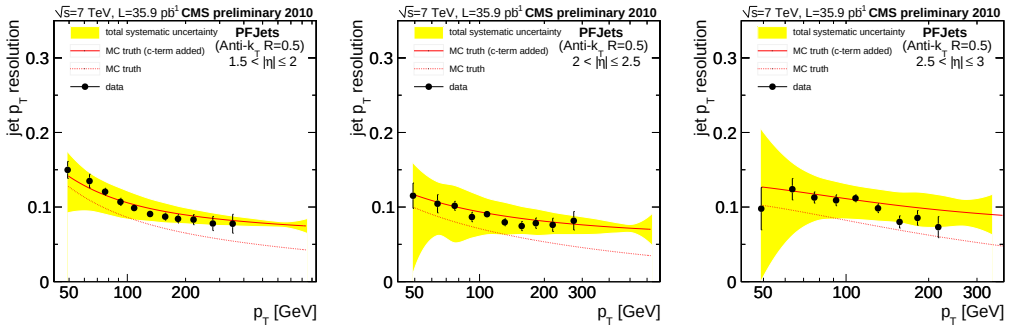


Figure 3.8. Scaled Montecarlo truth resolution compared to data for jets reconstructed with the PF algorithm in $\mathbf{1.5 \leq |\eta| < 2.0}$ (*left*); $\mathbf{2.0 \leq |\eta| < 2.5}$ (*center*); $\mathbf{2.5 \leq |\eta| < 3.0}$ (*right*).

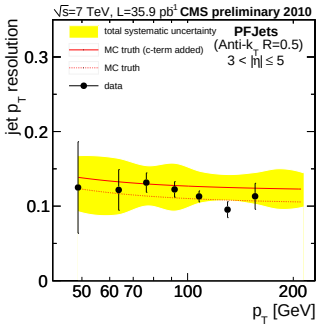


Figure 3.9. Scaled Montecarlo truth resolution compared to data for jets reconstructed with the PF algorithm in $\mathbf{3.0 \leq |\eta| < 5.0}$.

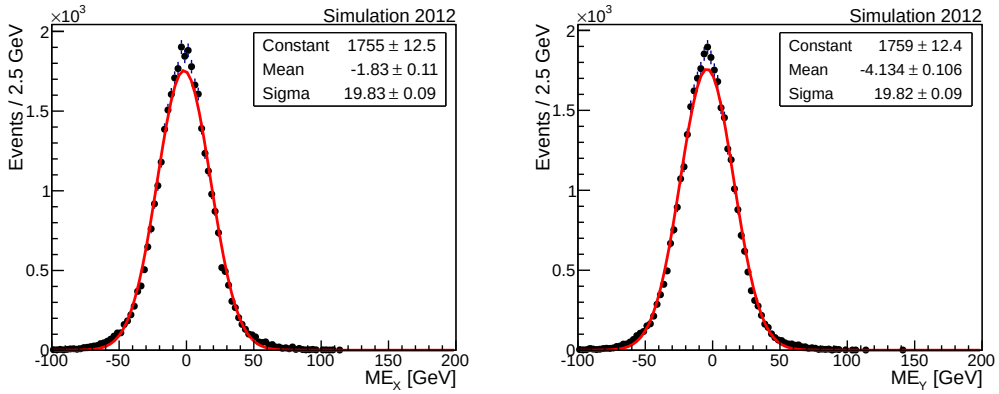


Figure 3.10. Distributions of the fake missing transverse energy along the x (left) and y (right) directions. Gaussian fits are overlaid to show the systematic shift in the mean value.

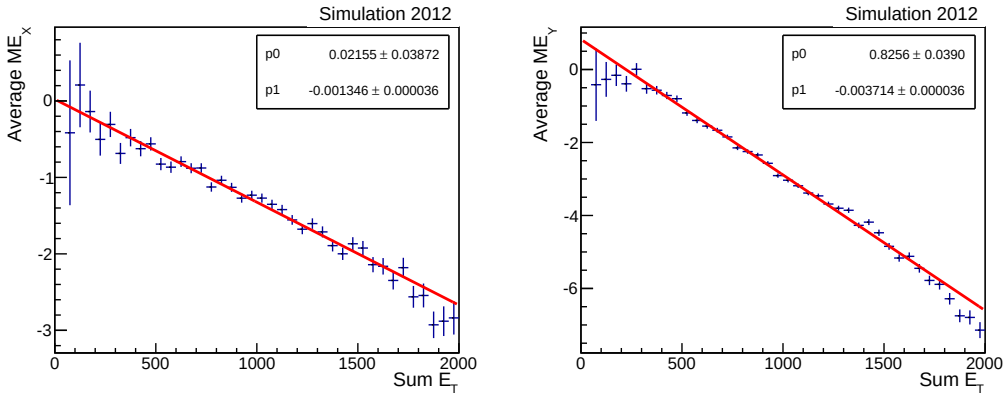


Figure 3.11. Fake Missing Transverse Energy shifts along the x (left) and y (right) directions as a function of the total visible transverse energy (2012 Drell-Yan sample: DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarballSummer12-PU_S7_START52_V9-v2AODSIM).

Table 3.7. Fitted parameters of the E_x^{miss} and E_y^{miss} correction functions.

Sample	Component	Fitted Parameter p_0			Fitted Parameter p_1		
Data	E_x^{miss}	-0.666	$\pm$	0.035	0.006239	$\pm$	0.000049
Data	E_y^{miss}	0.673	$\pm$	0.035	-0.004613	$\pm$	0.000050
Simulation	E_x^{miss}	0.021	$\pm$	0.038	-0.001346	$\pm$	0.000036
Simulation	E_y^{miss}	0.826	$\pm$	0.039	-0.003714	$\pm$	0.000036

Further investigations shows that the E_x^{miss} and E_y^{miss} shifts are not constant. In particular they vary linearly as a function of the total visible transverse energy in the event ($SumE_T$).

To evaluate the proper correction factor, we divide the sample in bins of $SumE_T$, and for each bin we compute the average E_x^{miss} and E_y^{miss} shift. The two distributions are separately fitted with a first order polynomial, and each of them represents a continuous correction function. The fitted functions for the simulated DY sample are shown in figure 3.11.

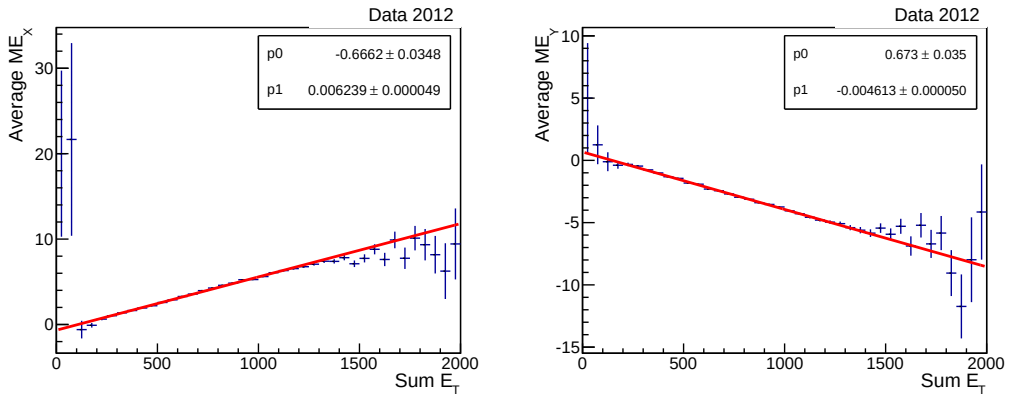


Figure 3.12. Missing Transverse Energy shifts along the x (left) and y (right) directions as a function of the total visible transverse energy in 2012 data (Run2012A Prompt Reco).

The parameters derived so far cannot be used to correct the E_T^{miss} shift in collision data. Indeed, the $\phi(E_T^{miss})$ asymmetry is different between data and Montecarlo (as shown in the figure 3.6(b)). The same procedure is repeated using a collision data sample, and the relative correction functions are shown in figure 3.12.

The formula to properly shift and correct E_T^{miss} as a function of $SumE_T$ is the following:

$$\begin{cases} E_x'^{miss} = E_x^{miss} - (p_1 \cdot SumE_T + p_0) \\ E_y'^{miss} = E_y^{miss} - (p_1 \cdot SumE_T + p_0) \end{cases} \tag{3.8}$$

where the parameter p_0 and p_1 are listed in table 3.7.

Jet Scaling

For a well calibrated detector, the p_T of a reconstructed jet corresponds on average to the transverse momentum p_T^{gen} of the generator-level jet. However in real life the jet response is never unitary, and Montecarlo based correction are applied during the event reconstruction.

Residual η and p_T dependencies of the jet response in data are investigated by balancing the total energy in dijets and γ + jet events (as for the jet resolution). Studies performed within the *CMS Jet-Met group* show that the only relevant residual discrepancy in the jet response is a function of the jet pseudorapidity [108]. The residual correction that we apply as a multiplicative factor to the jet energy in data is shown in figure 3.13; this is the last step needed to have a reliable Montecarlo description of the E_T^{miss} distribution.

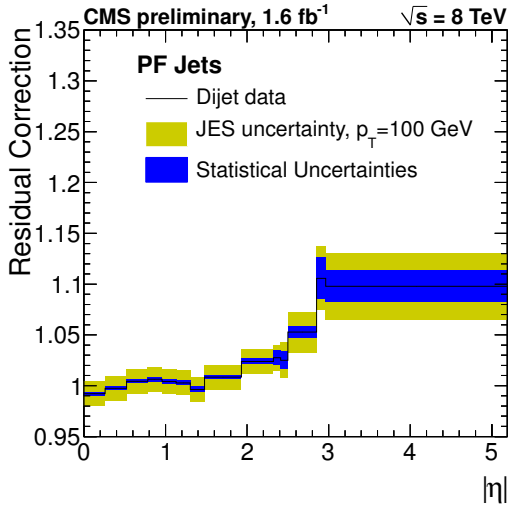
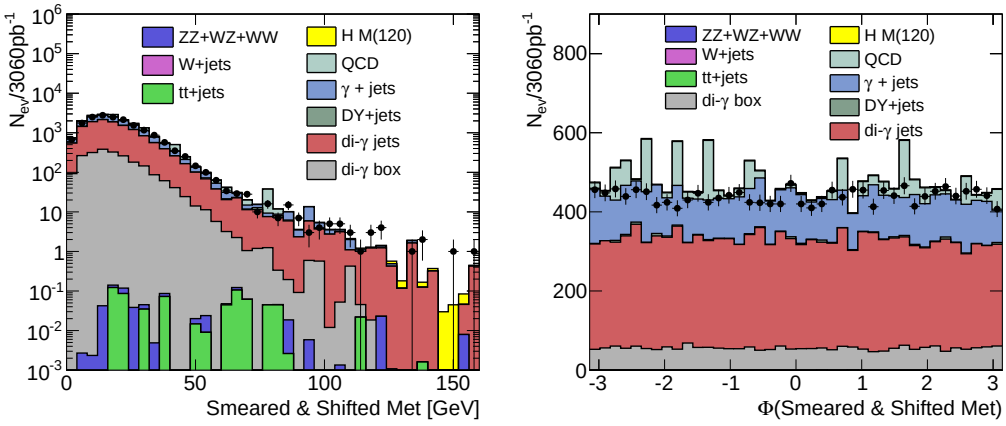


Figure 3.13. Residual Jet Corrections as a function of η.

In summary, the jet smearing procedure is applied only to simulation, the E_T^{miss} shifts are applied both to data and Montecarlo (using different values), and the jet scaling is applied only to data. Using all the corrections described so far, we compute again the E_T^{miss} magnitude and the E_T^{miss} direction. They are shown in figure 3.14.



(a) E_T^{miss} Magnitude (b) E_T^{miss} Azimuthal Angle

Figure 3.14. Corrected Particle Flow E_T^{miss}: Magnitude (left) and Azimuthal Angle (right)

3.4 Isolated Lepton Reconstruction

Electrons are identified as clusters of ECAL energy deposits matched to tracks from silicon tracker with an algorithm which accounts for possible energy loss due to bremsstrahlung in the tracker material. Electron candidates require a cluster with $E_T > 20\,\text{GeV}$ with pseudorapidity $|\eta| < 1.4442$ or $1.566 < |\eta| < 2.5$. Misidentified jets are suppressed by the particle-flow (PF) isolation requirement. Such PF isolation takes into account the particles which are in a particular ΔR cone around the reconstructed lepton. As to electron isolation, it requires the cut on the combined relative isolation

$$\sum PFChIso(e) + max(0., PFPhoIso(e) + PFNeuIso(e) - A_{eff} \cdot \rho))/p_T(e) \quad (3.9)$$

where *PFChIso*, *PFPhoIso*, *PFNeuIso* are respectively the charged hadrons, photons and neutral hadrons in a $\Delta R = 0.3$ cone with respect to the electron direction. The parameters ρ and A_{eff} have been already introduced; they measure respectively the average energy density in the event, and the effective area spanned by the objects clusterized in the jet. The electron identification cuts are summarized in Tab. 3.8.

Table 3.8. Electron identification requirements.

Variable	Criterion	
	Barrel	Endcap
$\sigma_{i\eta i\eta}$	< 0.01	< 0.03
$\Delta\phi_{in}$	< 0.015	< 0.010
$\Delta\eta_{in}$	< 0.007	< 0.009
H/E	< 0.12	< 0.10
d_0 w.r.t. selected vertex	< 0.02 cm	
d_z w.r.t. selected vertex	< 0.2 cm	
$\lvert 1/E - 1/p \rvert$	< 0.05	
Combined relative PF isolation	< 0.15	
vertex fit probability (conv. rej.)	$> 10^{-6}$	
missing hits (conv. rej.)	≤ 1	

Muons are identified requiring a good consistency between tracker and muon detector measurements. Such requirement is essential to reduce the contamination from muons produced in decays of hadrons and from beam halo. The tracker muons reconstruction starts from the inner tracker information, while the global muon reconstruction starts from segments in the muon chambers. The muon candidate is required to be reconstructed also by the PF algorithm, and to have $p_T > 20\,\text{GeV}$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 primary vertex, the longitudinal distance of the tracker track with respect to the primary 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 hit are included in the global muon track fit.

The muon isolation requires the cut on the combined relative PF isolation

$$\sum PFChIso(\mu) + max(0., PFPhoIso(\mu) + PFNeuIso(\mu) - A_{eff} \cdot \rho)/p_T(\mu) \tag{3.10}$$

where *PFChIso*, *PFPhoIso*, *PFNeuIso* are respectively the charged hadrons, photons and neutral hadrons in a $\Delta R = 0.4$ cone. The muon selection criteria are summarized in Table 3.9.

Table 3.9. Muon identification and isolation requirements.

Variable	Criterion
Number of pixel hits	> 0
χ^2/n.d.f	< 10
Number of muon hits	> 0
Number of matched muon stations	> 1
Number of tracker layers	> 5
d_0 w.r.t. selected vertex	< 0.02 cm
d_z w.r.t. selected vertex	< 0.05 cm
Combined relative PF isolation	< 0.2

Chapter 4

Inclusive Analysis of Diphoton Events

The study of the Higgs boson in the low mass region is driven by mainly five decay channels:

- $H \to V\,V$ (where V=W, Z)
- $H \to \tau^+\,\tau^-$
- $H \to b\,\bar{b}$
- $H \to \gamma\,\gamma$.

The final state involving two b quarks has the largest branching ratio, but it is swamped by the overwhelming hadronic background [111]. The decay into two W bosons occurs with a probability similar to the decay into two b quarks, and the W leptonic decay can be used to strongly suppress the background. The price for this choice is a reduced yield because of the leptonic branching ratio, and the impossibility to precisely reconstruct the Higgs boson invariant mass because of the neutrinos [112, 113]. If the two vector bosons are two Zs, the final state objects (jets and leptons) can be constrained to be resonant, and the invariant mass of the four objects can precisely measure the Higgs mass. But the yield of this channel is the lowest, and the analysis relies on really few events [114, 115]. Finally, the decay into two taus could result in a decent yield, but it is spoiled by the difficulty in the reconstruction of the tau jets. Indeed, the challenge of tagging tau jets and properly reconstruct their energy (accounting also for the invisible energy carried by neutrinos) is currently driving the sensitivity of this channel [116].

The diphoton decay represents the best compromise between a reasonable yield (more than 100 expected signal events in 5 fb^{-1}) and a good sensitivity in the low mass region. The choice of investigating the Higgs couplings by means of this channel lies on the following properties of its final state: it provides a clean final state topology, where the identification of the two high energy photons is performed with a very high efficiency (see section 3.1); the mass peak can be reconstructed with great precision using the CMS high resolution electromagnetic calorimeter; the analysis is in principle simple and straightforward, since it consists in a classic bump hunt on a smooth background shape (section 4.5); the main irreducible background is the direct photon production, having a cross section only five times larger with respect to the signal (section 4.1). And, despite its small branching ratio of $\sim$0.2% in the mass region of interest, the expected signal rate is roughly one order of magnitude smaller than the SM background rate (section 6.2).

The motivations for describing the inclusive analysis of the $H \to \gamma\gamma$ channel before introducing the exclusive analysis are mainly two:

1. the exclusive analysis assigns each event to a given category according to the presence of a tagging object in the event (e.g. a pair of forward jets, an isolated lepton, large E_T^{miss}); when no such object is present, we consider the event to be untagged, and we process it using the inclusive analysis. Therefore the latter can be considered as the category of the exclusive analysis with the lowest priority;

2. there are many features of the exclusive analysis inherited from the inclusive one, such as the trigger, the vertex reconstruction and the photon selection. We introduce them here, so that in the next chapter the focus will be only on the tagging objects.

Untagged signal events are mainly ascribable to the gluon-gluon fusion (GGF) Higgs production mechanism. This process has a cross section roughly twenty times larger than the other production mechanism, hence is responsible for the majority of the signal yield.

4.1 Data Samples and Cross Sections

The data sample consists of events collected by means of diphoton triggers through Spring and Summer 2012, and it corresponds to an integrated luminosity of $5.3\,\text{fb}^{-1}$ at $\sqrt{s} = 8\,\text{TeV}$.

The description of the Higgs boson decaying in two photons and of all background processes is obtained from Montecarlo simulations. Different matrix element generators are used according to the simulated process, but all of them are interfaced with Pythia [117] (using the Z2* tune parameter set) to simulate parton hadronization and jet fragmentation. After the generation step, the events are passed through the full CMS detector simulation with Geant 4 [118]. The number of simultaneous collisions (pile-up) is simulated such that the running conditions of the 2012 run could be covered. The pile-up scenario accounts for both multiple pp collisions happening in the same bunch crossing (in-time), and for 50 ns out-of-time pile-up. To let the simulated pile-up distribution match the one observed in data, a reweighting procedure in necessary. The latter is described in section 4.1.1.

The signal samples describing the gluon fusion and the vector boson fusion processes are generated using the next-to-leading-order (NLO) matrix element generator Powheg [119, 120]; whereas the associated production process is simulated directly with Pythia at leading order (LO). For the dominant gluon fusion process, the Higgs boson transverse momentum spectrum is reweighted to the next-to-next-to-leading-logarithmic (NNLL) + NLO distribution computed with the HqT package [121–123].

The SM Higgs boson cross-sections and branching ratios are taken from reference [124]. In particular, the gluon fusion cross-section is computed at NNLO+NNLL for perturbative QCD and NLO for electroweak (EWK) contributions. The vector boson fusion cross-section is computed at NNLO for QCD and NLO for Electroweak contributions. The associated production cross sections are computed at NLO QCD order. The gluon fusion process cross-section is reduced by 2.5% for all values of m_H to account for interference with background diphoton final states [125]. The magnitude of the variation of the interference effect in our acceptance is smaller than the systematic uncertainty on the effect. The uncertainty on the signal cross section due to PDF uncertainties is determined using the PDF4LHC prescription [126–130]. Signal samples are produced for Higgs mass values ranging from 90 to 150 GeV with a 5 GeV step (see Table 4.1).

We consider two types of background processes. The *irreducible* background has two real high E_{T} isolated photons. The *reducible* background has at least one non-isolated photon. A two real photon signature can be produced by both quark-antiquark annihilation (namely Born $q\bar{q} \to \gamma\gamma X$) and gluon-gluon fusion (namely Box $gg \to \gamma\gamma$) as well as quark-gluon Compton scattering with isolated Bremsstrahlung processes (namely Brem $qg \to \gamma\gamma X$) as shown in Figure 4.1. We estimate the total differential rate of irreducible backgrounds to be about 100 fb/GeV at 120 GeV mass, hence high mass resolution is required to powerfully discriminate a signal.

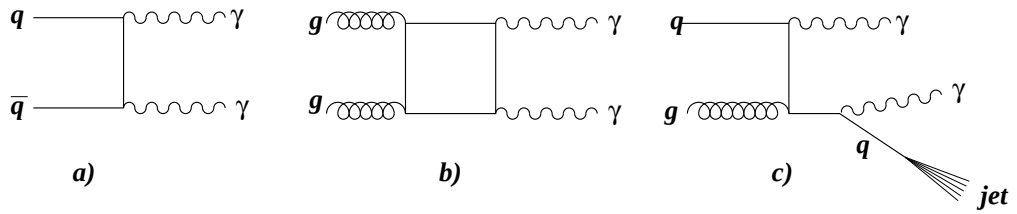


Figure 4.1. Spectator diagrams for irreducible background with two isolated photons: (a) Born $q\bar{q} \to \gamma\gamma X$ process, (b) Box $gg \to \gamma\gamma$, (c) Brem $qg \to \gamma\gamma X$.

Backgrounds in which at least one final state jet interpreted as a photon are considered as reducible. The dominant reducible backgrounds arise from QCD processes like γ+jet and dijets, which can lead to fake photons induced by neutral hadrons π^0 or η that are produced in the jet fragmentation processes. To diminish the rate of reducible backgrounds to the level of the irreducible one, the jet fake rate needs to be less than 10^{-3}. A measurement of the isolated photon production cross section is reported in [131, 132], while a measurement of the diphoton production cross section is published in [133]. Published results of production cross sections of irreducible and reducible backgrounds are in a reasonable agreement with the perturbative QCD (pQCD) predictions at the next-to-leading order (NLO) level.

The background Montecarlo samples are generated using Pythia, using Powheg interfaced with Pythia, or using the MadGraph matrix-element generator interfaced with Pythia depending on the simulated process. The SM background samples involving a prompt or fake diphoton final state are summarized in table 4.2.

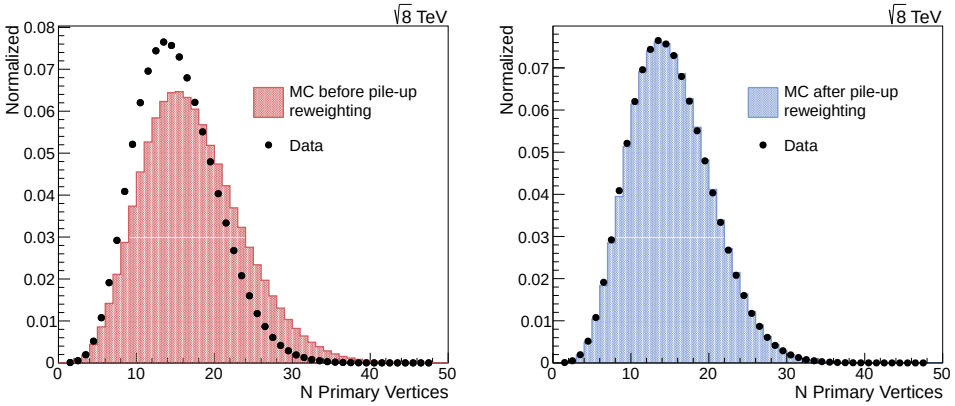
The pQCD calculations of the cross-sections and kinematic distributions of irreducible and γ + jet backgrounds carried out at NLO with the DiPhox [134] and JetPhox [135] programs, respectively. These are general purpose cross-section integrators which include both direct and fragmentation production mechanisms. The k-factors [136] used to rescale the simulated background samples are summarized in table 4.3. It is important to note that the ratio of the NLO to the LO cross sections (k-factor) may be large for the irreducible background samples generated with Pythia while it is smaller for the MadGraph generated samples since it includes the complementary high-p_T jets.

Specific backgrounds contributing to the Higgs boson tagged with lepton, $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ and two-jets are produced. They include $t\bar{t}$ and electroweak processes that may produce the same signature as VH channel electroweak processes:

1. top-quark pairs $t\bar{t}$
2. single vector bosons: W+jets, Z+jets,
3. diboson channels: $W\gamma$, WZ, ZZ

Table 4.1. SM Higgs cross sections at 8 TeV (pb) for different production mechanisms and $H \to \gamma\gamma$ branching ratios for different Higgs masses.

m_H (GeV)	Gluon Fusion	Vector Boson Fusion	$W \to WH$, $Z \to ZH$	$t\bar{t} \to H$	Branching Fraction
90	36.8	2.18	1.97, 1.06	0.32	$1.23 \cdot 10^{-3}$
95	33.2	2.07	1.68, 0.91	0.28	$1.40 \cdot 10^{-3}$
100	30.1	1.97	1.43, 0.78	0.24	$1.59 \cdot 10^{-3}$
105	27.4	1.88	1.23, 0.68	0.21	$1.78 \cdot 10^{-3}$
110	19.8	1.79	1.06, 0.59	0.19	$1.97 \cdot 10^{-3}$
115	25.0	1.71	0.92, 0.51	0.17	$2.13 \cdot 10^{-3}$
120	21.1	1.63	0.80, 0.45	0.15	$2.25 \cdot 10^{-3}$
125	19.5	1.56	0.70, 0.39	0.13	$2.29 \cdot 10^{-3}$
130	18.0	1.48	0.61, 0.35	0.12	$2.26 \cdot 10^{-3}$
135	16.7	1.43	0.54, 0.31	0.10	$2.12 \cdot 10^{-3}$
140	15.6	1.36	0.47, 0.27	0.09	$1.93 \cdot 10^{-3}$
145	14.6	1.31	0.42, 0.24	0.08	$1.67 \cdot 10^{-3}$
150	13.6	1.25	0.37, 0.22	0.07	$1.36 \cdot 10^{-3}$



(a) Before reweighting (b) After reweighting

Figure 4.2. Distribution of the number of reconstructed vertexes (a) before and (b) after the pile-up reweighting procedure. Black dots represent $Z \to e^+e^-$ data.

Table 4.2. Background Montecarlo samples, their production cross section and number of simulated events for different $\hat{p}_T$ -bins generated with PYTHIA.

Dataset	Kinematic Range	σ (pb)	Events
Diphoton + jets (MADGRAPH)	-	75.4	1154970
Diphoton Box (PYTHIA)	$10 < \widehat{p}_T \leq 25$ GeV	424.8	500400
Diphoton Box (PYTHIA)	$25 < \widehat{p}_T \leq 250$ GeV	15.54	500352
Diphoton Box (PYTHIA)	$\widehat{p}_T > 250$ GeV	$1.18 \cdot 10^{-3}$	500050
Photon + Jet (PYTHIA)	$20 < \widehat{p}_T \leq 40$ GeV	$8.19 \cdot 10^{4}$	5907942
Photon + Jet (PYTHIA)	$\widehat{p}_T > 40$ GeV	$8.84 \cdot 10^{3}$	5956149
QCD (PYTHIA)	$30 < \widehat{p}_T \leq 40$ GeV	$5.20 \cdot 10^{7}$	6061407
QCD (PYTHIA)	$\widehat{p}_T > 40$ GeV	$2.37 \cdot 10^{7}$	9790851
Drell-Yan + jets (MADGRAPH)	$M_{\ell\ell} > 50$ GeV	$3.53 \cdot 10^{3}$	30461028

Table 4.3. Background k-factors applied to the Montecarlo background samples.

Process	k-factor
pp$\rightarrow \gamma\gamma$ (born & brem)	1.15
pp$\rightarrow \gamma\gamma$ (box)	1.3
pp$\rightarrow \gamma$+jet (1 prompt)	1.3
pp$\rightarrow$jets (1 prompt)	1.3
pp$\rightarrow$jets (0 prompt)	1
Drell Yan	1.15

4. diphoton associated with vector bosons and top quarks: $W^{\pm}\gamma\gamma$, $Z\gamma\gamma$, $t\bar{t}\gamma\gamma$

The cross section of these background samples, the generator used and other parameters are summarized in Table 4.4.

4.1.1 Pile-up reweighting

The simulation includes an accurate prediction of the number of interactions taking place in each bunch crossing. However, to fully reproduce the expected distribution of the number of interactions taking place in real data, the simulated events must be reweighted.

Although the primary vertex reconstruction has been shown well-behaved up to the levels of pile-up observed in 2012 data, the final distribution of the number of vertexes is still sensitive to the details of the primary vertex reconstruction and to the underlying event. Additionally, the distribution of the number of reconstructed vertexes could be biased by the offline event selection criteria and even by the trigger. In order to factorize these effects, instead of reweighting the simulation by the number of reconstructed primary vertexes, we reweight according to the number of pile-up interactions from the simulation truth. The target pile-up distribution for data is derived by using the per bunch-crossing instantaneous luminosity (measured by the forward calorimeters) together with the total pp inelastic cross-section (71 mb).

To validate the reweighting technique we compare the vertexes reconstructed using the Deterministic Annealing algorithm [137] between data and Montecarlo in a Z $\rightarrow$

Table 4.4. Electroweak background Montecarlo samples, their production cross-sections and equivalent simulated integrated luminosity.

Dataset	Generator	Kinematic Range	σ (pb)	Events
$t\bar{t}$	MadGraph		157.5	$1.17 \cdot 10^6$
$WW(2\ell 2\nu)$	MadGraph		5.81	$1.58 \cdot 10^6$
$WZ(3\ell\nu)$	MadGraph		1.057	$2.56 \cdot 10^6$
$ZZ(2\ell 2q)$	MadGraph		1.28	$1.94 \cdot 10^6$
$ZZ(2\ell 2\nu)$	MadGraph		0.365	$0.9 \cdot 10^6$
$ZZ(4\ell)$	MadGraph		0.0921	$5.1 \cdot 10^6$
$W\gamma$	MadGraph	$E_T(\gamma) > 5\,\text{GeV}$	322.356	$4.7 \cdot 10^6$
$Z\gamma$	MadGraph	$E_T(\gamma) > 5\,\text{GeV}$	181.338	$5.76 \cdot 10^6$
$W^+\gamma\gamma$	MadGraph	$E_T(\gamma) > 10\,\text{GeV}$	0.0667	$6.2 \cdot 10^5$
$W^-\gamma\gamma$	MadGraph	$E_T(\gamma) > 10\,\text{GeV}$	0.0504	$1.2 \cdot 10^5$
$Z\gamma\gamma$	MadGraph	$E_T(\gamma) > 10\,\text{GeV}$	0.068	$6.9 \cdot 10^4$
$t\bar{t}\gamma\gamma$	MadGraph	$E_T(\gamma) > 10\,\text{GeV}$	0.001316	$3.1 \cdot 10^4$

e^+e^-sample. The two distributions after the reweighting are reported in figure 4.2(b) and they show a fair agreement.

4.2 Trigger

Every LHC collision event at CMS is evaluated by the trigger system. There are two levels of triggers at CMS, the Level-1 Trigger (L1) and the High Level Trigger (HLT). The events entering the analysis must pass a diphoton trigger decision. The trigger criteria for identifying events with a pair of photon candidates have evolved with the increasing instantaneous luminosity delivered by the LHC.

The diphoton triggers used by the analysis can be categorized into two types (two HLT paths [138, 139]), one with E_T thresholds of 26/18 GeV and the other with E_T thresholds of 36/22 GeV, where the two thresholds are applied respectively to the leading and the trailing photon. Each 26/18 path is required to be initiated by at least two hardware L1 e/γ candidate, whereas each 36/22 path is required to be initiated by at least one hardware L1 e/γ candidate. The minimum E_T requirements for these L1-primitives are respectively 13/7 GeV for the first path and 22 GeV for the second path (the exact definition of L1-primitive can be found in [140]).

Once the L1-seeding requirement has been satisfied, ECAL electromagnetic clusters are formed in the region in the vicinity of the L1 primitive(s). To save processor time in the HLT, ECAL information is unpacked only from the readout units overlapping with a rectangle centered on an L1 candidate with a size $\Delta\eta \times \Delta\phi = 0.25 \times 0.4$. 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.

For the 26/18 path, we require the presence of at least two such clusters. Among these clusters we require at least two passing the relevant HLT cuts (described later). For the single-L1-seeded 36/22 diphoton path, we require only one such cluster. This requirement

being satisfied, we try to reconstruct another candidate by unpacking the rest of the ECAL information. This candidate is also required to pass the HLT selection criteria.

The HLT selection is based on the isolation of the photon, the shape of the electromagnetic shower and the hadronic fraction of the total energy. The general trigger strategy is to keep all the possible good photon pairs using the "OR" of multiple paths implementing different thresholds, and it is described in more detail in reference [141].

To measure the trigger efficiencies, one needs to evaluate separately the efficiency of the L1-seeding (L1 efficiency) and the efficiency of the HLT filters provided that the L1 requirement has been satisfied. In this analysis, we use the tag and probe method on $Z \to e^+e^-$ data for efficiency measurements. The data sample for efficiency studies is obtained as follows:

1. From a dataset enriched with double electron events, we select those events passing a loosely prescaled single electron path. Since this path requires only one electron passing the tight HLT cuts, the other electron which is required to pass only a very loose selection, is suitable for our measurement.

2. We require at least one offline electron to be matched to the high quality HLT electron, and at least two offline photons to be matched to the two HLT electromagnetic objects. The two offline photons are required to have an invariant mass compatible with the Z peak (between 70 GeV and 110 GeV), and to pass an offline p_T cut of 30 GeV and 22.5 GeV, respectively.

3. We require the event to pass the preselection described in section 3.1.6.

4. The photon matched to the HLT electron leg is also required to be matched to an L1 e/γ isolated object with $E_T > 22$ GeV. We label this photon as tag and the other one as probe, and we perform various efficiency measurements.

To account for the fact that electrons and photons have different distributions of the shower shape variable R_9 (described in section 3.1), each electron pair used for the trigger efficiency measurement has been weighted so that the R_9 distribution of the associated clusters matches the one of Higgs photons. The net effect is an increasing of the measured efficiency due to the migration of events towards higher R_9 values.

For events passing the preselection and the standard photon selection (see section 3.1.5) we find the HLT efficiency to be $(99.78 \pm 0.08)\%$.

The L1 efficiencies are $(90.72 \pm 0.02)\%$ when there are two L1 candidates seeded by the 13/7 GeV thresholds, and $(97.70 \pm 0.03)\%$ where there is just one L1 candidate seeded by the 22 GeV threshold.

4.3 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. In this section the treatment of the latter will be described. Since photons (and other neutral particles) cannot be readily assigned to a given interaction vertex, the opening angle resolution strongly depends on determining the interaction where the diphoton was produced.

The mean number of pp interactions per bunch crossing in the 8 TeV dataset is 18.7. The interaction vertexes reconstructed using the tracks of charged particles are distributed in the longitudinal direction, z, with an RMS spread of 5 cm. If the interaction point is known to better than about 10 mm, then the resolution on the opening angle between the photons makes a negligible contribution to the mass resolution, compared to the contribution from the ECAL energy resolution. Thus the mass resolution can be preserved by correctly assigning the reconstructed photons to one of the interaction vertexes reconstructed from the tracks.

The reconstructed primary vertex having the highest probability of being the interaction vertex of the diphoton event can be identified using the kinematic properties of the tracks associated with the vertex and their correlation with the diphoton kinematic properties. In addition, if either of the photons converts and the tracks from the conversion are reconstructed and identified, the direction of the converted photon, determined by combining the conversion vertex position and the position of the ECAL supercluster, can be used to point to and so identify the diphoton interaction vertex.

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.

4.3.1 Base algorithms

The vertex identification algorithm exploits the correlation between the recoiling tracks and the diphoton system. In events where one of the two photons converts into an e^+e^- pair, the conversion tracks are reconstructed and linked to the photon supercluster. The additional information is then used to locate the vertex.

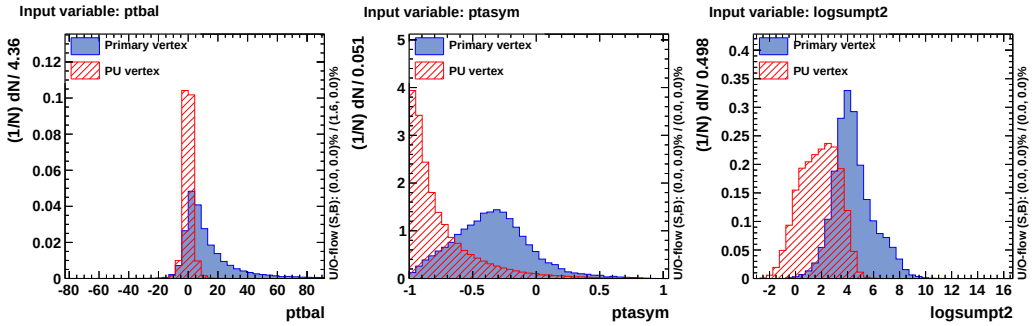
Two unconverted photons. In presence of unconverted photons, the three variables used for the vertex identification are:

- *sumpt2*: $\sum_i |\vec{p}_T^{\,i}|^2$.
- *ptbal*: $-\sum_i (\vec{p}_T^{\,i} \cdot \frac{\vec{p}_T^{\,\gamma\gamma}}{|\vec{p}_T^{\,\gamma\gamma}|})$.
- *ptasym*: $(|\sum_i \vec{p}_T^{\,i}| - p_T^{\gamma\gamma})$ / $(|\sum_i \vec{p}_T^{\,i}| + p_T^{\gamma\gamma})$.

where the sums run over all tracks associated to a given vertex, and the last two variables quantify the event p_T balance with respect to the diphoton system. The distribution of such variables is shown in Figure 4.3.

It was verified that, even with the increased pile-up conditions of 2012 data-taking, the addition of more variables doesn't bring significant improvements to the correct vertex assignment performances [142].

At least one converted photon. Electron-positron pairs from reconstructed converted photons can be exploited to determine the longitudinal coordinate of the primary interaction vertex where the Higgs boson is produced (the conversion reconstruction algorithm is detailed in section 3.1.2).



(a) Ptbal (b) Ptasym (c) Logsumpt2

Figure 4.3. Vertex identification variables extracted from the general tracks collection.

Two methods have been developed which give different performances depending on where, in the CMS Tracker, the photon conversion occurs. Both methods exploit the knowledge of the converted photon direction extracted from the conversion reconstruction; once the direction is known it is extrapolated back to the beam line to obtain the z position.

In the first method the photon direction is calculated using the angle, α, between the conversion momentum and the $z-$axis. The conversion momentum is evaluated from the track pair refitted with the vertex constraint. The longitudinal coordinate of the primary interaction vertex is then calculated as

$$Z_{PV} = Z_{conv} - R_{conv} \times \cot(\alpha)$$

where Z_{conv} and R_{conv} are the z and the radius of the fitted conversion vertex.

In the second method, the direction of the converted photon is instead determined by combining the information on the conversion vertex and the position of the ECAL supercluster so that the longitudinal coordinate of the primary interaction vertex is calculated as

$$Z_{PV} = Z_{conv} - \frac{R_{conv}}{(R_{SC} - R_{conv}}) \times (Z_{SC} - Z_{conv})$$

where Z_{conv} and R_{conv} are the z and the radius of the fitted conversion vertex as in the previous method, and Z_{SC} and R_{SC} are the z and the radius of the ECAL supercluster. Figure 4.4 illustrates the two techniques.

The CMS tracker is divided into six regions each with a different z resolution (as described in section 2.4 and in [143]). The resolution on the longitudinal position of the interaction vertex achieved by both methods is comparable, except in the Tracker Outer Barrel (TOB) region where the conversion momentum method has a significantly worse resolution. For this reason the two methods are combined on an event-by-event basis, choosing one or the other determination depending on the information on the conversion reconstructed vertex which identifies the region of the tracker. This brings an improvement of about 1% on the overall vertex finding efficiency. The resulting Δz distribution, measured on the full 2012 data set and the simulated γ + jet sample, is shown in Figure 4.5 for the six separate tracker regions. The results obtained in data are listed in table 4.5 and are used in the analysis.

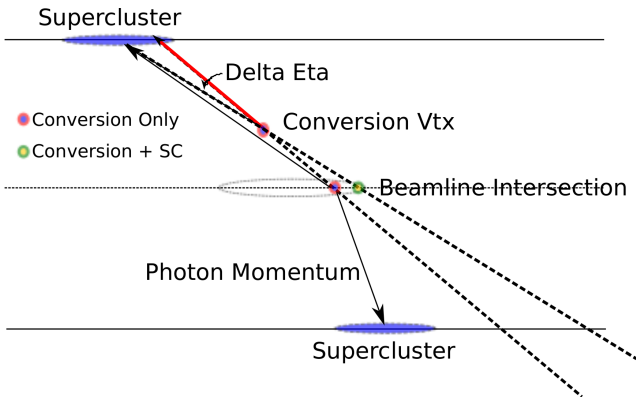
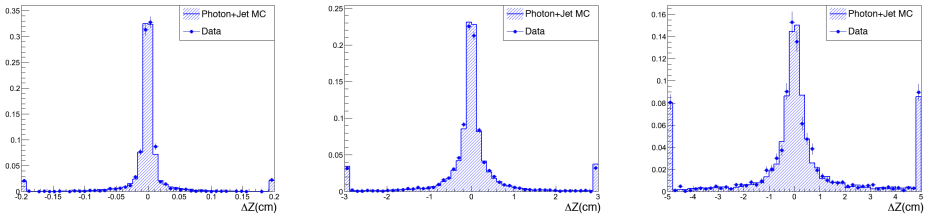
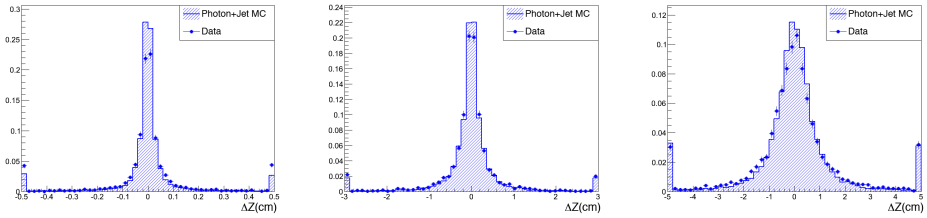


Figure 4.4. Illustration of the two methods considered to determine the z of the photon vertex in case it converts. One method only extrapolates the conversion track (conversion only,) whereas the other also takes into account the ECAL energy position (conversion + SC).



(a) Pixel Barrel (b) Tracker Inner Barrel (TIB) (c) Tracker Outer Barrel (TOB)



(d) Pixel Forward (e) Tracker Inner Disk (TID) (f) Tracker Endcap (TEC)

Figure 4.5. Δz distributions for different tracker regions.

Table 4.5. Resolution on the primary vertex z as determined from conversions using the combined method.

σ (cm)	Pix. Barrel	TIB	TOB	Pix.Fwd	TID	TEC
Data 2012	0.030	0.376	2.639	0.192	0.497	1.33
Simulated γ + jet	0.015	0.294	1.901	0.112	0.362	0.915
Ratio data/simulation	2	1.279	1.388	1.729	1.373	1.453

4.3.2 Combination

In order to exploit at best the shape and correlation between the variables obtained from the general tracks collection and pairs of tracks identified as photon conversions, a multivariate discriminant is employed. In particular, the *Boosted Decision Tree* (BDT) algorithm is used to combine all the information. The inputs to the BDT are the *sumpt2*, *ptbal* and *ptasym* variables and, in events with at least one photon conversion, a fourth variable

$$\text{pull}_{conv} = \frac{|z_{vertex} - z_{conv}|}{\sigma_{conv}}$$

where z_{conv} is the estimated primary vertex position and σ_{conv} is the resolution measured in data.

The BDT algorithm is trained on simulated $H \to \gamma\gamma$ events with $m_H = 120\,\text{GeV}$.

4.3.3 Performance in data

Performance of the identification algorithm when there are no conversions in the event is tested on a $Z \to \mu^+\mu^-$ sample, while the case with at least one conversion is tested in a $\gamma + \text{jet}$ sample.

Without conversions. Events with a Z boson decaying into a muon pair are exploited to measure the performance of the vertex identification directly on data. The lepton tracks is used to identify the hard interaction vertex and they are subsequently removed from collection of tracks used in the vertex reconstruction algorithm to mimic the topology of a Higgs boson decaying into two photons.

The difference in the p_T spectrum of the Z and the Higgs is taken into account by studying the efficiency of the vertex identification as a function of the boson p_T.

A relative difference as a function of the boson p_T between data and Montecarlo vertex identification efficiency is then derived from Z events and later applied as a correction to the Higgs MC (in the assumption that the relative data to MC differences between Higgs and Z are the same).

Figure 4.6 shows the BDT output for $Z \to \mu^+\mu^-$ events in the simulation and in data, demonstrating a general good agreement.

The efficiency measured in data on the $Z \to \mu^+\mu^-$ sample is reported in figure 4.7 as a function of the p_T of the boson and of the number of reconstructed vertexes.

With conversions. The performance of the vertex identification BDT with converted photons is measured on $\gamma + \text{jet}$ events. Events with a converted photon and a jet with a p_T greater than 30 GeV are selected for this measurement. The tracks close to the jet direction ($\Delta R(\text{jet, track}) < 0.5$) are excluded when calculating the variables entering the BDT. The vertex with the highest BDT score is chosen as the best vertex, and we consider it to be correctly identified if it lies within 10 mm from the vertex tagged by the jet. Figure 4.8 shows the BDT output in $\gamma + \text{jet}$ events in data and simulated events; right and wrong vertex assignment are shown separately.

Figures 4.9 a) and b) show the vertex finding efficiency as a function of the number of reconstructed vertexes in the event and as a function of the transverse momentum of the $\gamma + \text{jet}$ system. Results are shown for both data and simulated $\gamma + \text{jet}$ events.

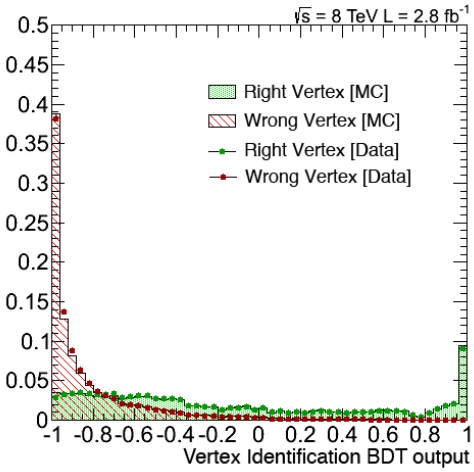
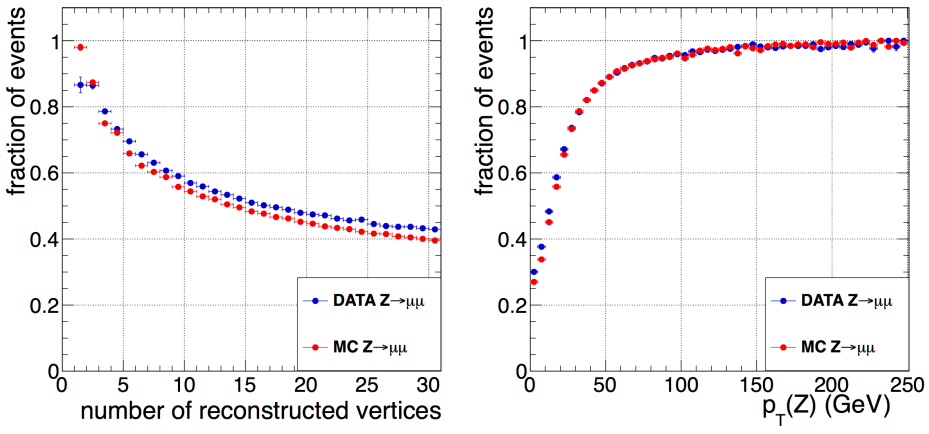


Figure 4.6. Performance of the vertex identification using the combination BDT. A sample of $Z \to \mu^+\mu^-$ events is used, where the true vertex is tagged by the muon tracks.



(a) (b)

Figure 4.7. Comparison of the vertex identification efficiency between data and simulation as a function of of the number of reconstructed primary vertexes (a) and the boson p_T (b).

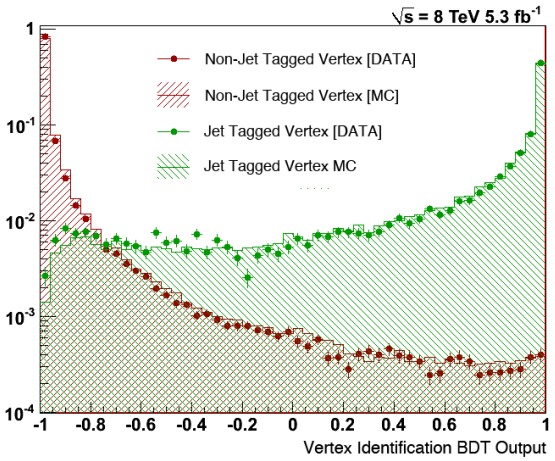
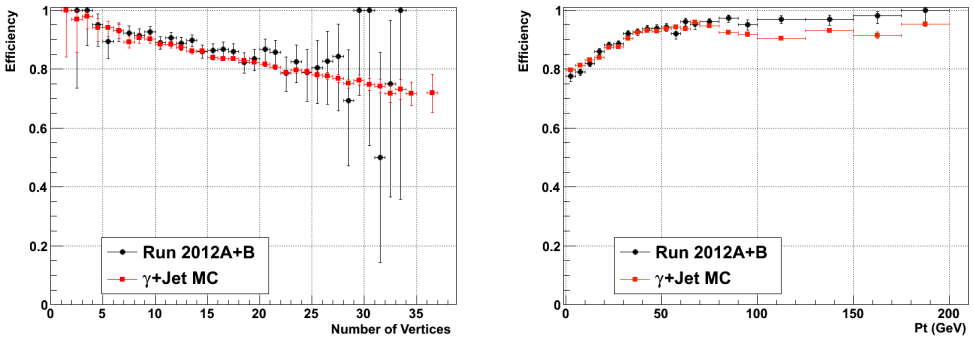


Figure 4.8. Output of the BDT algorithm combining the three variables used for vertex identification in case of unconverted photons, and the conversion vertex when it is available. Data-Montecarlo comparison is shown using a sample of γ + jet events in the presence of pile-up.



(a) (b)

Figure 4.9. The BDT vertex identification efficiency for γ + jet events with a converted photon. a) BDT vertex identification efficiency as a function of the number of vertexes for data and simulated events and their ratio; b) BDT vertex identification efficiency as a function of the transverse momentum of the photon-plus-jet system, for data and simulated events and their ratio.

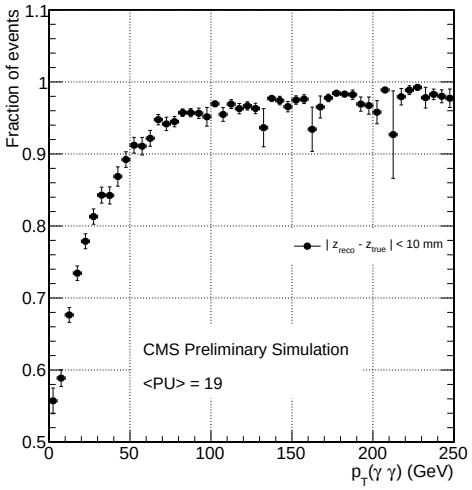


Figure 4.10. Performance of the BDT vertex identification on a $H \to \gamma\gamma$ Montecarlo sample for the 2012 pileup conditions.

4.3.4 Performance in a simulated signal sample

The fraction of $H \to \gamma\gamma$ events where the vertex chosen using the BDT discriminant is found within 10 mm from the true vertex is shown as a function of the Higgs boson p_T in figure. 4.10. The overall efficiency integrated over the entire diphoton p_T spectrum is $(79.0 \pm 0.2)\%$.

4.4 Selection Criteria & Event Categorization

Untagged events contribute differently to the Higgs analysis since the signal significance depends on the mass resolution and on the signal-to-background ratio. Search sensitivity can be increased by further subdividing the untagged events into sub-categories according to indicators of mass resolution and predicted signal-to-background ratio, and combining the results of a search in each sub-category. Two simple classifiers are used: the conversion probability of the two photons (i.e. the minimum R_9 of the two photons, R_9^{min}), and the maximum pseudorapidity of the two photons. Both classifiers are effective in separating diphotons with good mass resolution from those with less good resolution and in separating events with a higher signal to background probability from those with a lower signal/background probability. The sub-category boundaries are chosen to match those used in defining the photon classes in the photon identification algorithm (section 3.1.5). The untagged diphoton event categories are:

- Untagged 0: both photons in barrel and $R_9^{min} > 0.94$;
- Untagged 1: both photons in barrel and $R_9^{min} < 0.94$;
- Untagged 2: one or both photons in endcap and $R_9^{min} > 0.94$;
- Untagged 3: one or both photons in endcap and $R_9^{min} < 0.94$.

To obtain similar invariant mass shapes over the different sub-categories (hence simplifying the background modeling), instead of cutting on the p_T of the photon it was chosen to cut on $p_T/m_{\gamma\gamma}$, where $m_{\gamma\gamma}$ is the candidate diphoton invariant mass. This result in an effective p_T threshold that is no longer a fixed value. The *sliding cut* has been fixed to be 1/3 and 1/4 respectively on the leading and trailing photon. This means a cut of 40 GeV (leading) and 30 GeV (trailing) for a candidate invariant mass of 120 GeV. The reason for this choice is to avoid creating a non-monotonically declining mass spectrum in the background shape (due to potential turn-on effects in the p_T distribution), and to obtain similar invariant mass shapes over the different event categories.

Both photons are required to be in the acceptance of the detector: $|\eta_{SC}| < 2.5$ (where η_{SC} is the position of the photon supercluster). And events where one of these photons lies in the barrel-endcap transition region $1.44 < |\eta_{SC}| < 1.56$ are removed. Moreover, the two photons have to satisfy the *super-tight* photon-id requirements introduced in section 3.1.

4.5 Background Model

Montecarlo simulation of the background processes is not used in the analysis. However, the diphoton mass spectrum that is observed after the full event selection is found to agree with the distribution predicted by Montecarlo, within the uncertainties on the cross sections of the contributing processes which is estimated to be about 15%. The background components have been scaled by k-factors obtained from CMS measurements [144–146]. The contribution to the background in the diphoton mass range $110 < m_{\gamma\gamma} < 150$ GeV from processes giving non-prompt photons is about 30%.

For the analysis the background model is obtained by fitting the observed diphoton mass distributions in each of the 4 event classes with 5^{th} order Bernstein polynomials ($f(x)$ in equation 4.1) over the range $100 < m_{\gamma\gamma} < 180$ GeV.

$$\begin{cases} f(x) = \sum_{\nu=0}^{n} \beta_\nu \, b_{\nu,n}(x), & n = 5 \\ b_{\nu,n}(x) = \binom{n}{\nu} x^\nu (1-x)^{n-\nu}, & \nu = 0, \ldots, n \end{cases} \tag{4.1}$$

The choice of the fit function is made based on a study of the bias which might be introduced by the choice of the fit function on the measured signal strength.

The study of bias is performed as follows. Four functions are fitted to the Montecarlo background distribution of the diphoton invariant mass spectrum and taken as "truth models" for the background. The choice of functions is intended to be sufficiently broad to span the possibilities of the true distribution (i.e. the distribution to which signal-free data would tend if the integrated luminosity were increased). Due to the variation of photon efficiency with E_T, and possible kinematic biases imposed by our event classification we are not, at present, able to exclude the possibility that the true distribution has a polynomial form. The 4 functions are:

1. The sum of 2 exponential functions
2. The sum of 2 power functions ($fm^{-p_1} + (1-f)m^{-p_2}$)
3. 2^{nd} order Chebyshev polynomial

4. Laurent series with 4 terms $(f_1 m^{-3} + f_2 m^{-4} + f_3 m^{-5} + (1 - f_1 - f_2 - f_3) m^{-6})$

Pseudo-experiments describing possible experimental outcomes are then randomly generated using the fitted functions as generators of background. Further pseudo-experiments are generated where a Standard Model signal is added to the background. Exclusion limits are calculated for these experiments using the asymptotic approximation to the CLs procedure [147], using a range of functions to fit the background in the generated experiments. Background fitting functions in shortened ("sliding-window") and extended fit ranges are also tested. The bias is computed as the average value of the observed signal strength in 2000 pseudo-experiments, at each of 9 mass values between 110 and 150 GeV (5 GeV steps).

It was found that a number of functions were able to reliably predict the local expected background distribution, keeping the bias on the prediction to a small value for all the truth models. Among these functions it was found that a 5^{th} order polynomial fit to the range $100 < m_{\gamma\gamma} < 180$ GeV best preserved the sensitivity of the analysis to excluding or finding a Higgs Boson signal. The maximum average bias found when using the 5^{th} order polynomial fit was compared to the statistical uncertainty in the background fit, evaluated as an absolute number of events in a 3 GeV window. It is found that the statistical uncertainties are at least 5 times larger than the remaining bias at all of the 9 mass points.

4.6 Systematic Uncertainties

The systematic uncertainties described here are related to common features of all the event categories. They are applied identically to all the categories when fitting the background shapes and when measuring the per-mechanism Higgs cross sections, i.e. they are included in the model as common nuisance parameters.

Single photon systematics account for the photon identification efficiency, in particular for the efficiency of the high R_9 category, where the uncertainty results in a migration of photons between the two categories; and for the ECAL energy scale and resolution.

Single event systematics account for the uncertainty on the measured integrated luminosity, for the vertex finding efficiency and for the efficiency of the diphoton trigger.

Theoretical systematics come mostly from the uncertainty on the production cross sections, on the parton density functions, and on the factorization scale in simulating the fragmentation processes.

The relevant values are summarized in table 4.6.

Table 4.6. Summary of the systematic uncertainties.

Sources of systematic uncertainty		**Uncertainty**	
Per photon		Barrel	Endcap
Photon identification efficiency		1.0%	2.6%
$R_9 > 0.94$ efficiency (results in class migration)		4.0%	6.5%
Energy resolution ($\Delta\sigma/E_{MC}$)	$R_9 > 0.94$ (low η, high η)	0.22%, 0.60%	0.90%, 0.34%
	$R_9 < 0.94$ (low η, high η)	0.24%, 0.59%	0.30%, 0.52%
Energy scale ($(E_{data} - E_{MC})/E_{MC}$)	$R_9 > 0.94$ (low η, high η)	0.19%, 0.71%	0.88%, 0.19%
	$R_9 < 0.94$ (low η, high η)	0.13%, 0.51%	0.18%, 0.28%
Per event			
Integrated luminosity		4.4%	
Vertex finding efficiency		0.2%	
Trigger efficiency	One or both photons $R_9 < 0.94$ in endcap	0.4%	
	Other events	0.1%	
Production cross sections		Scale	PDF
Gluon-gluon fusion		+12.5% -8.2%	+7.9% -7.7%
Vector boson fusion		+0.5% -0.3%	+2.7% -2.1%
Associated production with W/Z		1.8%	4.2%
Associated production with $t\bar{t}$		+3.6% -9.5%	8.5%
Scale and PDF uncertainties (Effect of up to 12.5% event class migration.)		(y, p_T)-differential	

4.7 Results and Invariant Mass Fits

Background model fits to the diphoton mass distribution, on the hypothesis of no signal, are shown for the four untagged event categories in figure 4.11. 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 $m_{\gamma\gamma}$.

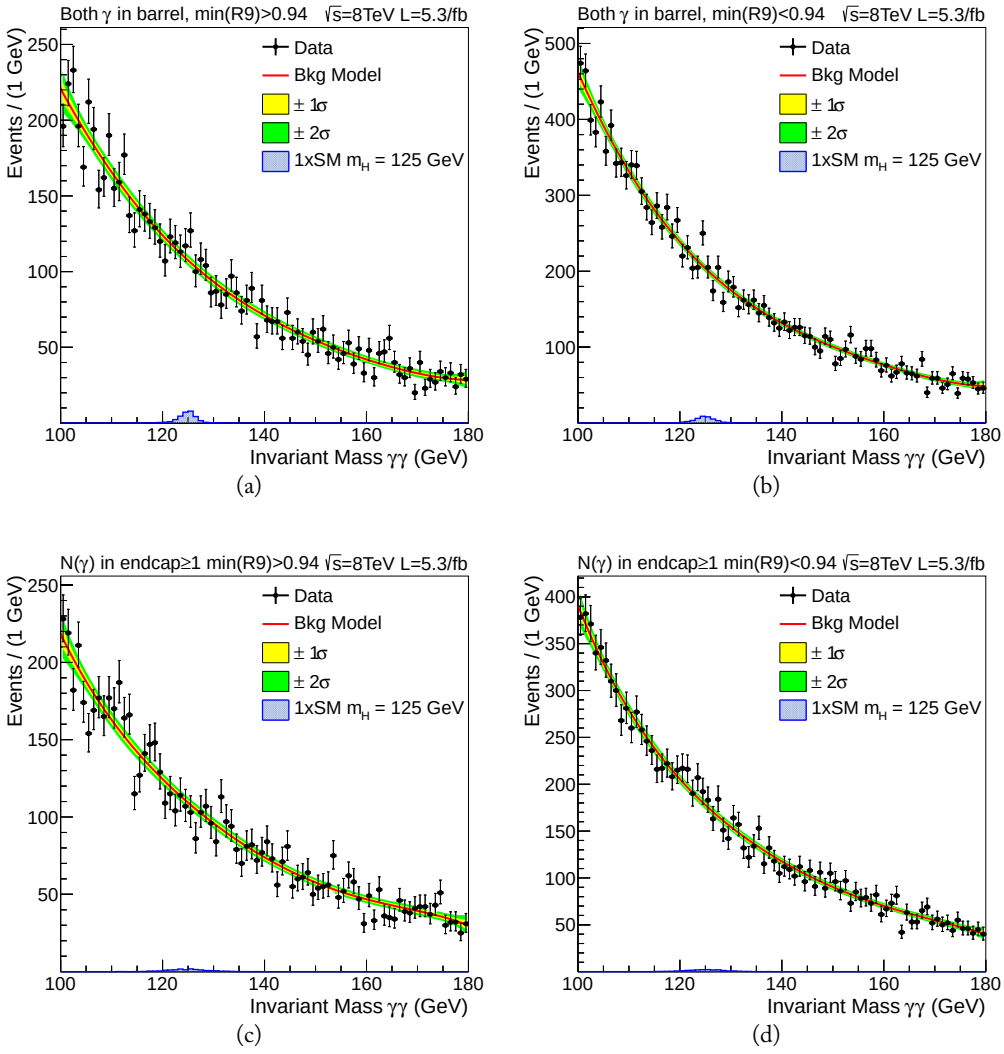


Figure 4.11. Background function fitted to data in the four untagged categories. Figure (a) is category 0: both photons in the barrel scoring high R_9. Figure (b) is category 1: both photons in the barrel scoring low R_9. Figure (c) is category 2: at least one photon in the endcap scoring high R_9. Figure (d) is category 3: at lease one photon in the endcap scoring low R_9.

As a sanity check we merge together the four untagged categories and we perform a data-Montecarlo comparison of the invariant mass shape. Events contributing to the background distribution are divided into three components: those with two real photons, those

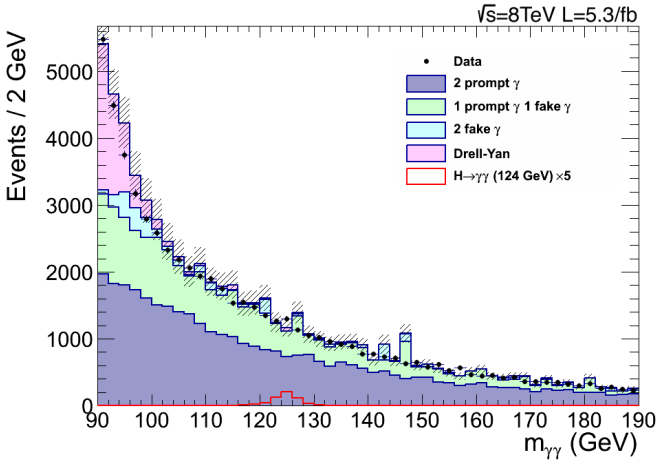
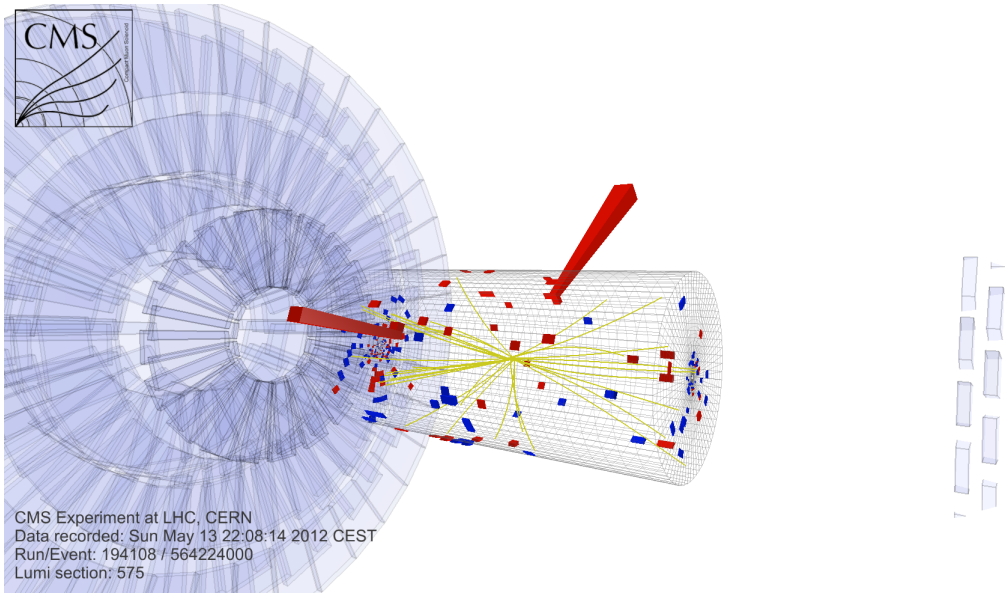
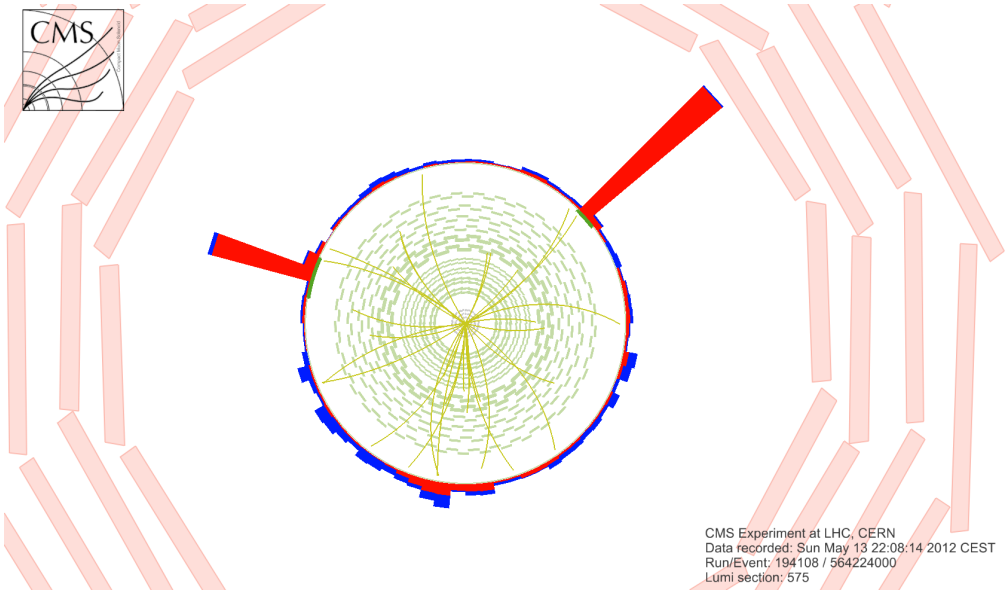


Figure 4.12. Diphoton invariant mass distribution for events selected in all the event categories for data and simulation. Backgrounds are represented by the filled histograms, while the Higgs signal ($m_H = 125$ GeV) scaled by a factor 5 is shown by the red line. The shaded band represents the theoretical uncertainty on the Montecarlo prediction.

with one prompt and one fake photon, and those with two fake photons. The relative fraction of each component as a function of the diphoton invariant mass is shown in figure 4.12. The Drell-Yan sample is treated separately, since its contribution is resonant at 91 GeV and it has a completely different shape with respect to the other backgrounds sources. The figure shows that the most important background is due to two real photons; the two-fake component being strongly suppressed by the tight photon-id. The spikes visible in the simulated $m_{\gamma\gamma}$ spectrum are due to the large weights needed to compensate the limited Montecarlo statistics. The nice data-Montecarlo agreement confirms that the corrections applied to the photon energy are working properly and that the sliding p_T cut used in the photon selection criteria is not distorting the background shape.



(a) 3D View



(b) Transverse View

Figure 4.13. Display of a candidate Higgs event produced via gluon-gluon fusion.

Chapter 5

Exclusive Analysis of Diphoton Events

We perform an exclusive analysis of the diphoton decay channel with the aim of disentangling the Higgs production mechanism. In particular, when the Higgs is produced via vector boson fusion (VBF) or in association to a vector boson (VH), we exploit the prominent topology of the final state to tag the event and to strongly reduce the background. Three event tags are considered in this analysis: the dijet tag, defined by the presence of two forward jets in the final state; the lepton tag, defined by the presence of at least one isolated electron or muon; and the E_T^{miss} tag, defined by the presence of large E_T^{miss}. Signal events that are dijet tagged are most likely produced via VBF, since the two initial state quarks radiating the vector bosons gain small transverse momentum and result in two forward jets. Lepton tagged signal events are ascribable to the VH production mechanism, where the vector boson is decayed leptonically. E_T^{miss}-tagged signal events are equally ascribable to the VH mode, but in this case either the Z decayed to neutrinos or the lepton was outside the acceptance region. Each tag is implemented as a standalone analysis, with its own optimized selection criteria. Since the number of events in common between them is non-negligible (especially between the lepton tag and the E_T^{miss} tag), we assign a priority to each tag: events are firstly scrutinized looking for leptons, than looking for jets and finally looking for large E_T^{miss}. Within each tagging analysis, the events are further divided in different event sub-categories to increase the Higgs sensitivity.

The event trigger, the photon selection and the vertex reconstruction are inherited from the inclusive analysis and have already been described in chapter 4.

5.1 Final State With Two Forward Jets

In this section we present the exclusive search for the diphoton Higgs decay with two additional jets in the final state, defined to select the Vector Boson Fusion (VBF) production process. In particular we address here events with two forward jets (originating from the two scattered quarks), for which the production cross section is about a factor 10 smaller than that of the gluon-gluon fusion process, but the signal to background ratio is higher (typically by an order of magnitude).

Candidate diphoton events for the dijet tagged category have the same requirements imposed on the photons as for the untagged category, with the exception of the p_T thresholds, which are modified to increase signal acceptance. In the VBF events the p_T of the Higgs boson is boosted [148] giving enhanced asymmetries in the photon pair energies and

hence favoring a lower threshold on one of the two photons. Therefore the threshold requirement for the trailing photon is set to $p_T(\gamma) > 25\,\text{GeV}$ and the thresholds on the leading photons is set to $p_T(\gamma) > (55\,\text{GeV}) \times m_{\gamma\gamma}/120$.

5.1.1 Jet Selection

In this analysis jets are reconstructed using the particle-flow algorithm [98, 99], which uses the information from all CMS sub-detectors to reconstruct different types of particles produced in the event. The basic objects of the particle-flow reconstruction are the tracks of charged particles reconstructed in the central tracker, and energy deposits reconstructed in the calorimeters. These objects are clustered with the anti-k_T algorithm [102] using a value of 0.5 for the distance parameter ΔR. The jet energy measurement is calibrated to correct for detector effects using samples of dijet, γ + jet , and Z + jet events [109]. Energy from pile-up interactions (overlapping the hard scattering interaction), and from the underlying event, is also included in the reconstructed jets. This energy is subtracted using the FastJet technique [92, 95, 96], which is based on the calculation of the η-dependent transverse momentum density, evaluated on an event-by-event basis.

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:

$$\beta^* = \sum_{\substack{i \in \textit{tracks in jet not} \\ \textit{associated to primary vertex}}} p_T(\textit{track}_i) \Big/ \sum_{j \in \textit{all tracks in jet}} p_T(\textit{track}_j)$$

$$\textit{jet}\text{RMS} = \sum_{i \in \textit{all particles in jet}} p_T^2(\textit{particle}_i) \cdot \Delta R_i^2 \Big/ \sum_{j \in \textit{all particles in jet}} p_T^2(\textit{particle}_j)$$

Different cuts are applied in the different region of the detector. Table 5.1 summarizes 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 for jets from the hard interaction high ($> 95\%$). In figure 5.1 the discrimination power for the two variables under study is reported. The efficiency for both real jets and pile-up jets is shown in figure 5.2, while the improvement in the two-jet selection is in figure 5.3.

The efficiency of the selection has been checked using samples of γ + jet and Z($\rightarrow$

Table 5.1. Summary of jet-ID cuts.

Jet Pseudorapidity	β^* Threshold	RMS Threshold
$\eta < 2.5$	$< 0.2 \log N_{vtx} - 0.64$	< 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

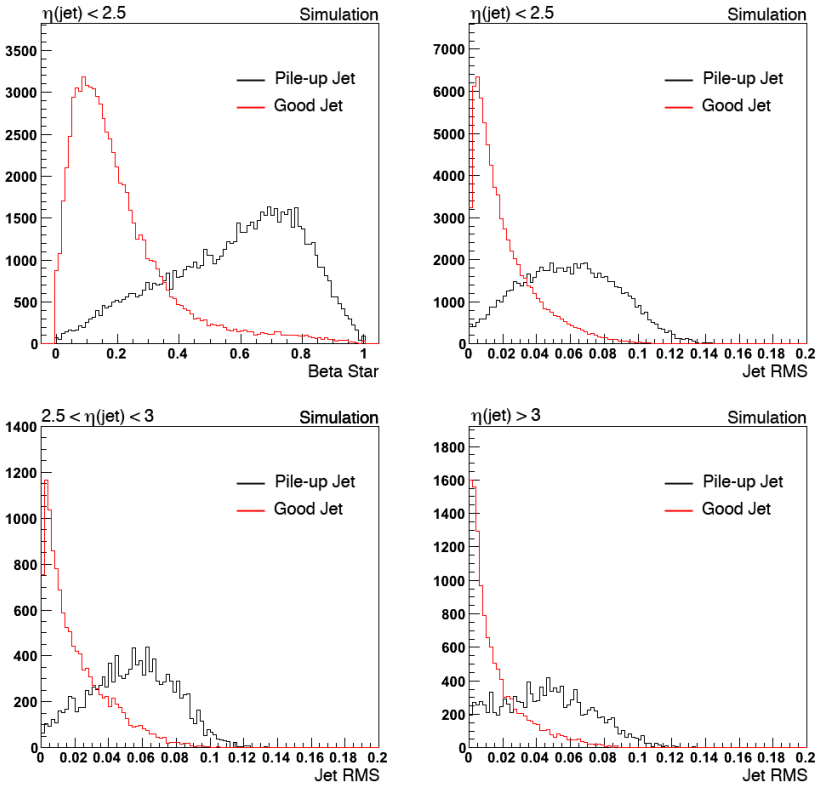


Figure 5.1. Discriminating variables used for the pile-up jet ID. Top left: β^* for $|\eta_{jet}| < 2.5$. Top right: RMS for $|\eta_{jet}| < 2.5$. Bottom Left: RMS for $2.5 < |\eta_{jet}| < 3$. Bottom Right: RMS for $|\eta_{jet}| > 3$.

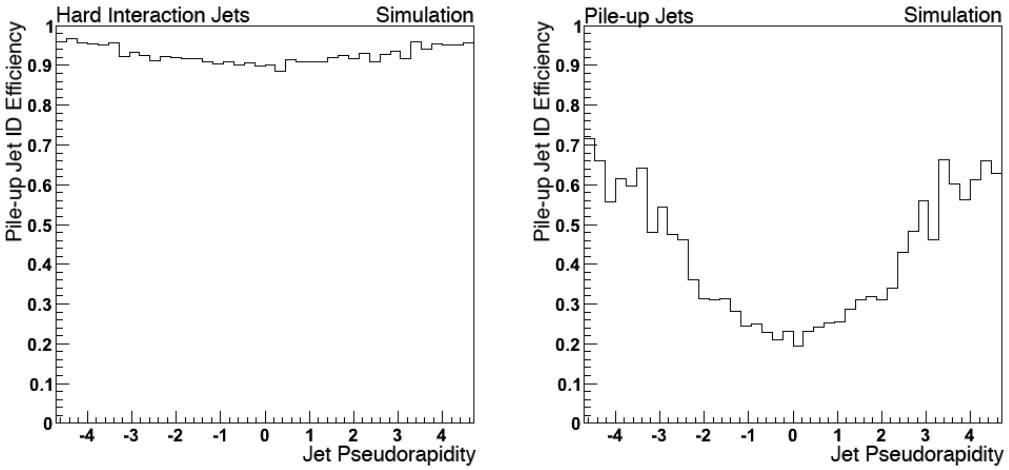


Figure 5.2. Efficiency of pile-up jet ID as a function of the pseudorapidity of the jet. Left: hard interaction jets. Right: pile-up jets.

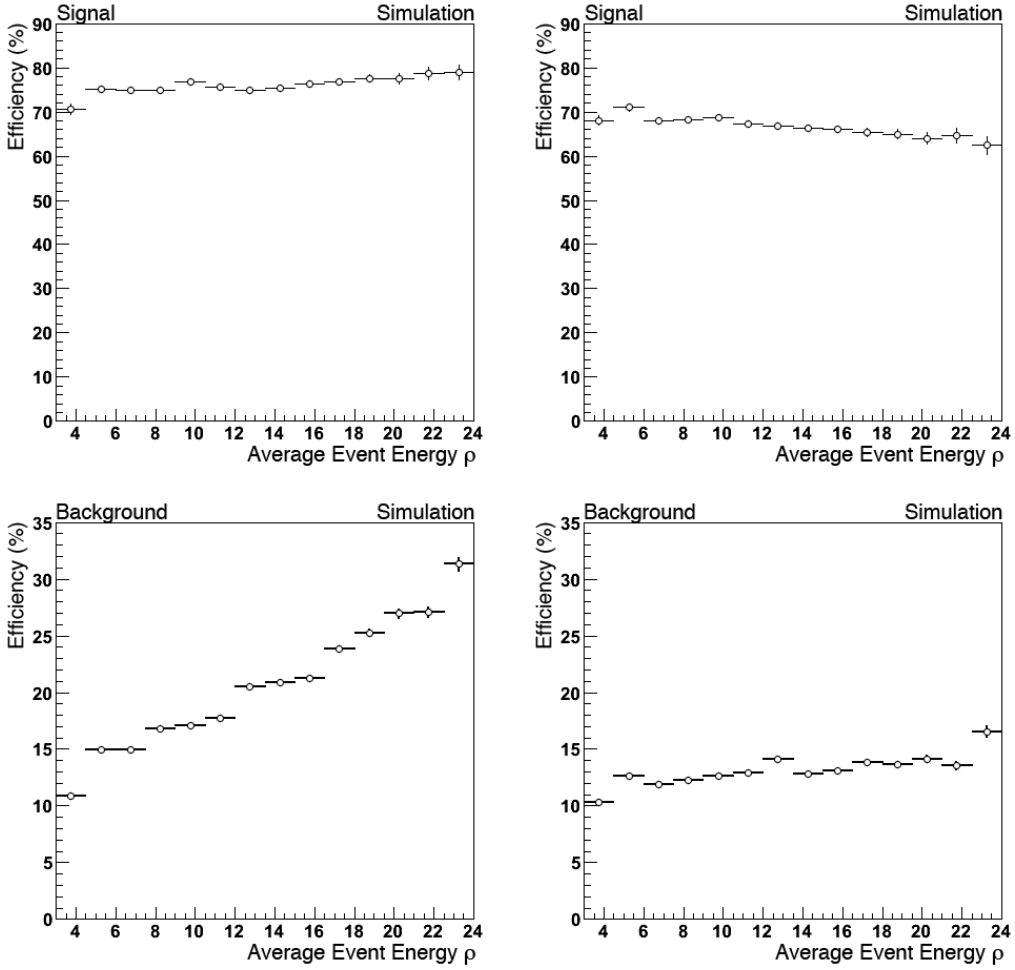


Figure 5.3. Effect of pile-up jet ID on the dijet selection. Efficiency here is the ratio of the events passing the requirements $p_T^{jet}(leading) > 30\,\text{GeV}$ and $p_T^{jet}(trailing) > 20\,\text{GeV}$ to the events where two photons are selected with loose photon ID. Top left: signal efficiency before pile-up jet ID. Top right: signal efficiency after pile-up jet ID. Bottom left: background efficiency before pile-up jet ID. Bottom right: background efficiency after pile-up jet ID.

$\mu\mu$)+jets events, where the jet recoiling to the photon (Z) is used as a probe. To reduce the pile-up contamination on the probe side, requirements on the difference between the azimuthal angle of the photon (Z) and the hardest recoiling jet are applied. In figure 5.4 the efficiencies on data and MC are compared. As shown, the agreement is within 2-5% depending on the jet pseudorapidity. This difference (parameterized as a function of transverse momentum and pseudorapidity) is used as systematic uncertainty on the jet efficiency.

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

The two highest p_T jets are retained to compute the dedicated VBF selection variables. The selection makes use of the following variables:

- The transverse momenta of the leading and trailing jets: $p_T(j_1)$ and $p_T(j_2)$.
- The dijet invariant mass, m_{jj}.
- The difference in pseudorapidity between the jets, $\Delta\eta(jj)$.
- The so-called *Zeppenfeld* variable Z (defined later).
- The difference in the azimuthal angle between the dijet and the diphoton, $\Delta\phi((j_1 j_{2),\gamma\gamma})$.

In figure 5.5 and 5.6 we show the p_T and η distributions of the leading and subleading jets after the photon selection. The p_T is shown for events that have both jets with a transverse momentum larger than 20 GeV, while the pseudorapidity is shown for events passing the jet selection $p_T(j_1) > 30\,\text{GeV}$ and $p_T(j_2) > 20\,\text{GeV}$.

The VBF topology is characterized by the presence of two forward jets at high η that are back-to-back, and a large central region without hadronic activity due to the exchange of the colorless Higgs particle. This turns into a large difference in pseudorapidity $\Delta\eta(jj)$ between the two jets, figure 5.7(a).

The difference in azimuthal angle between the diphoton system and the dijet system ($\Delta\phi(jj,\gamma\gamma)$) is expected to peak at π for both signal and background processes, except when one of the tagging jets is from a pileup interaction. Therefore the $\Delta\phi(jj,\gamma\gamma)$ requirement helps to reduce the background in the high-pileup regime (figure 5.7(b)).

The Zeppenfeld variable exploits the relationship between the Higgs and jet system rapidity in VBF events [149]. It corresponds to the pseudorapidity of a physical observable, e.g. photon, jet or Higgs candidate, in the reference frame of the two tagging jet candidates and is computed as

$$Z = \eta_{obs} - \frac{\eta(j_1) + \eta(j_2)}{2} \tag{5.1}$$

where the subscript *obs* indicates the observable under study. In this analysis we use $\eta_{obs} = \eta(\gamma_1 + \gamma_2)$, which corresponds to using the pseudorapidity of the Higgs candidate as the observable.

For the VBF process, the Zeppenfeld distribution is central and differs significantly from background (figure 5.8(b)). The physical interpretation of this feature is that the Higgs tends to be produced preferentially in the pseudorapidity space between the two back to back jets, following the VBF topology.

Since the tagging jets are forward and very energetic, we expect large dijet invariant mass m_{jj} values for VBF events (figure 5.8(a)). Therefore, a natural choice for VBF analyses is to select very large dijet invariant mass events [148].

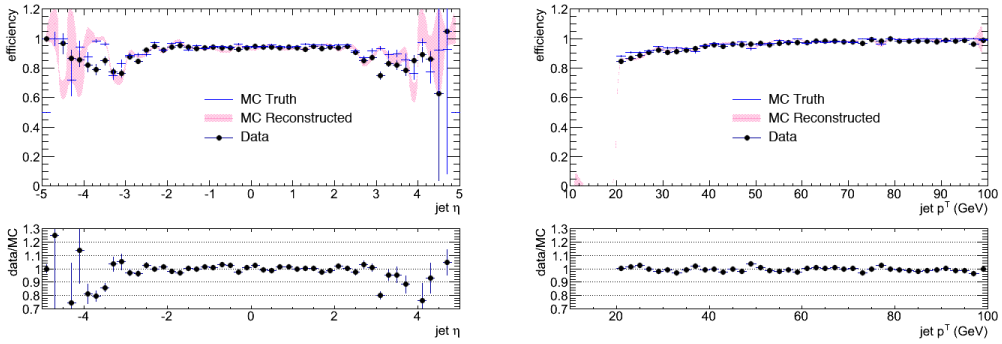


Figure 5.4. Data-MC comparison of pile-up jet ID efficiency. Results on Z+jet sample for the recoiling jet in the 20 GeV $< p_T <$ 100 GeV range (left) and as a function of transverse momentum (right).

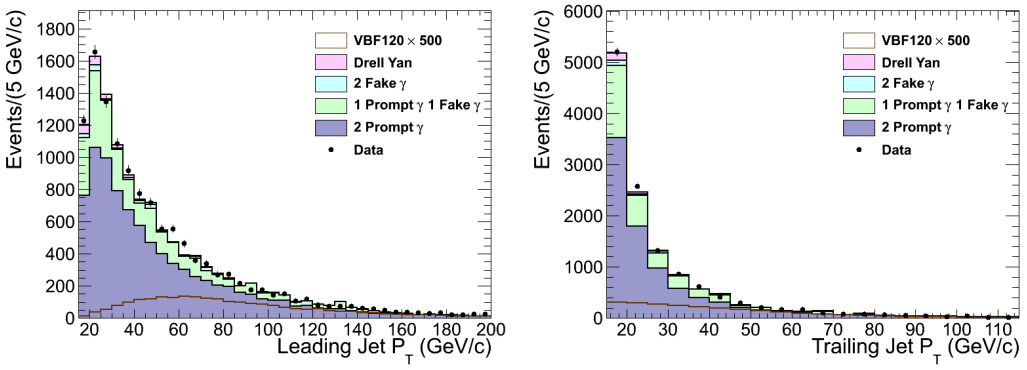


Figure 5.5. Transverse momentum distributions of the two highest p_T jets for events passing the diphoton selection and having p_T of both jets larger that 15 GeV: highest p_T jet (left); second highest p_T jet (right).

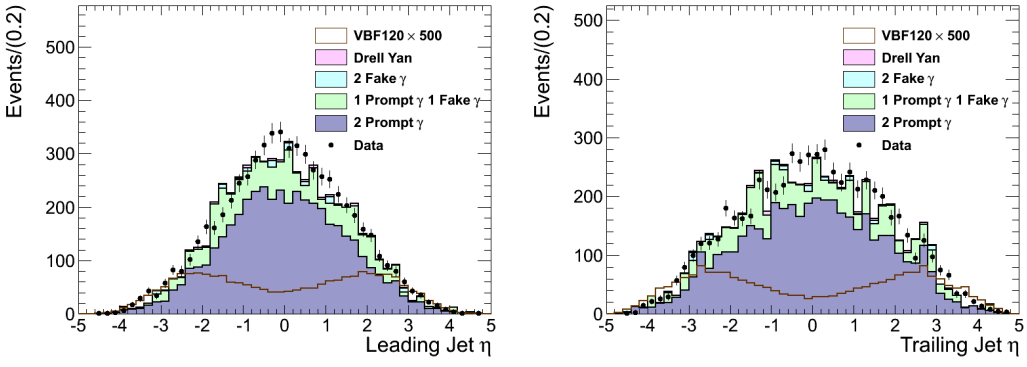
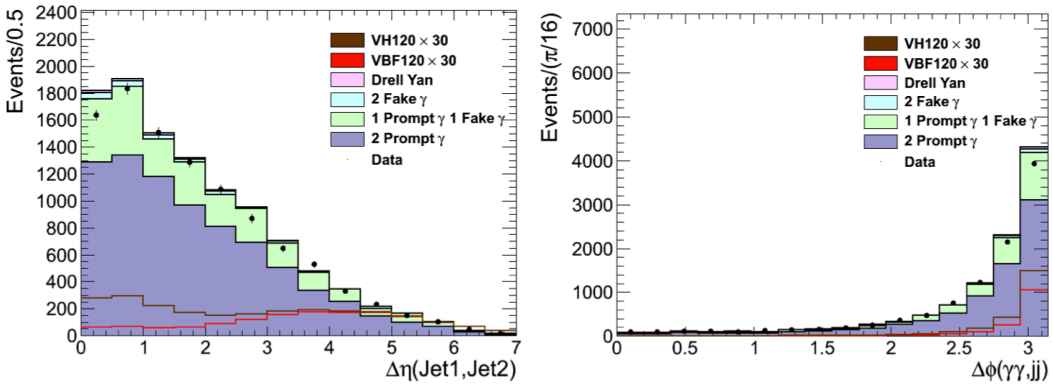
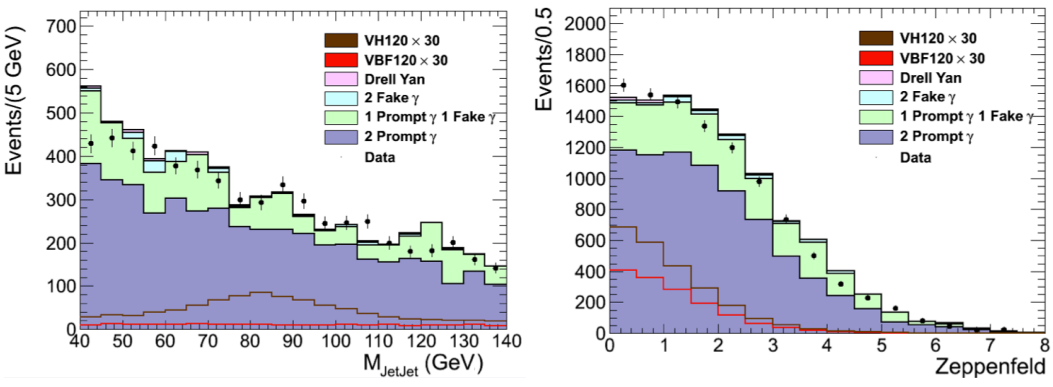


Figure 5.6. Pseudorapidity distributions of the two highest p_T jets for events passing the diphoton selection and having p_T larger than 30 GeV and 20 GeV respectively: highest p_T jet (left); second highest p_T jet (right).



(a) (b)

Figure 5.7. Dijet variables: (a) Pseudorapidity separation between the two selected jets; (b) ϕ-separation between the dijet system and the diphoton system.



(a) (b)

Figure 5.8. Dijet variables: (a) Invariant mass of the two selected jets; (b) Zeppenfeld variable as defined in equation 5.1.

Optimization of Selection Criteria

In order to improve the behavior or the mass spectrum of the background, in line with the inclusive analysis, we scale the p_T cut of the leading photon with the measured mass of the diphoton system ($m_{\gamma\gamma}$). We keep the subleading photon p_T cut constant to avoid incurring in trigger and selection inefficiencies for photon $p_T < 15\,\text{GeV}$.

The set of selection criteria used in the analysis is the following:

preselection:	*super-tight* photon selection, $90 < m_{\gamma\gamma} < 190\,\text{GeV}$,
cut 0:	$p_T(\gamma_1)/m_{\gamma\gamma} > p_T(\gamma_1)/m_{\gamma\gamma}\|_{thr}$ and $p_T(\gamma_2) > p_T(\gamma_2)\|_{thr}$,
cut 1:	$p_T(j_1) > p_T(j_1)\|_{thr}$,
cut 2:	$p_T(j_2) > p_T(j_2)\|_{thr}$,
cut 3:	$\|\Delta\eta(j_1, j_2)\| > \|\Delta\eta(j_1, j_2)\|_{thr}$,
cut 4:	$\|Z\| < \|Z\|_{thr}$,
cut 5:	$m_{jj} > m_{jj}\|_{thr}$,
cut 6:	$\|\Delta\phi(jj, \gamma\gamma)\| > \|\Delta\phi(jj, \gamma\gamma)_{thr}\|$

We optimize the sensitivity of the analysis, finding the set of cuts values for which the MC expected upper limit at 95% CL on the cross-section, using the VBF production channel only, is minimized. To evaluate the upper limit we further apply a cut on the invariant mass of the two photons ($m_{\gamma\gamma}$) corresponding to $(m_H - 3\,\text{GeV}) < m_{\gamma\gamma} < (m_H + 2\,\text{GeV})$. The optimization is performed using a signal with $m_H = 120\,\text{GeV}$. The resulting selection criteria is used for all mass points.

We optimize the upper limit by varying the cuts one by one, by keeping constant all the thresholds except that of the variable involved in the optimization. The threshold value which is found to minimize the upper limit is kept fixed at next iteration, when we optimize the following cut. When this procedure is performed on each variable, the resulting set of cuts is used as a new "seed" for the second iteration. We then start from this new "seed", and then we vary the cuts altogether on a finer matrix around this set.

In addition, 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 categories boundaries have also been optimized with a procedure similar to the one used for the event-selection.

The final cuts are summarized in table 5.2. Events falling in the first category are removed from the second one. The median expected limit for the dijet-tagged categories is shown in figure 5.9 and compared to the same quantity obtained using only one category.

5.1.2 Background Model

We perform the same study of the $m_{\gamma\gamma}$ background model described in section 4.5 applying the VBF selection criteria. The goal is to quantify the bias introduced in the signal extraction by the wrong choice of the background shape, and to find the proper function minimizing such bias.

The result is that a 3$^{\text{rd}}$ order polynomial is the best choice for the VBF tight category, and a 4$^{\text{th}}$ order polynomial is the best choice for the VBF loose category.

5.1.3 Systematic Uncertainties

The signal efficiency is affected both by the uncertainty on the jet energy scale and by the uncertainty on the jet energy resolution. We estimate the systematic error on the jet energy

Table 5.2. Final selection criteria for the VBF selection. Events from the first category are removed from the second one.

Variable	Criterion	
	VBF Tight	VBF Loose
$p_T(\gamma_1)/m_{\gamma\gamma}$	> 0.5	> 0.5
$p_T(\gamma_2)$	$> 25\,\text{GeV}$	$> 25\,\text{GeV}$
$p_T(j_1)$	$> 30\,\text{GeV}$	$> 30\,\text{GeV}$
$p_T(j_2)$	$> 30\,\text{GeV}$	$> 20\,\text{GeV}$
$\lvert\Delta\eta(jj)\rvert$	> 3.0	> 3.0
$\lvert Z\rvert$	< 2.5	< 2.5
m_{jj}	$> 500\,\text{GeV}$	$> 250\,\text{GeV}$
$\lvert\Delta\phi(jj,\gamma\gamma)\rvert$	> 2.6	> 2.6

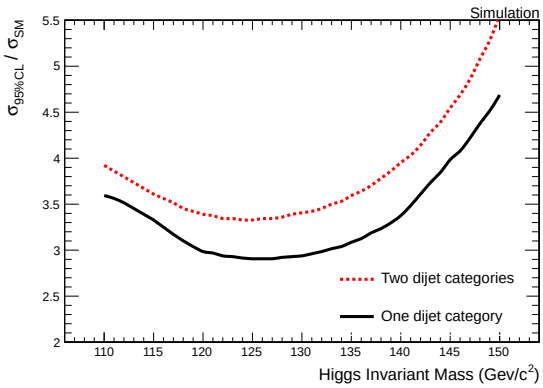


Figure 5.9. Comparison of the median expected exclusion limit using one or two dijet tagged categories.

scale by shifting the scale by $\pm 1\sigma$, where σ is the full jet energy scale uncertainty. Such uncertainty is evaluated from reference [108]. The systematics on the jet energy resolution are estimated by varying the resolution by $\pm 10\%$, which corresponds to the level of disagreement between the resolution measured in data and in MC [109]. Both events produced via VBF and via gluon gluon fusion enter the dijet tag analysis, and the jet energy uncertainties affect them differently. For VBF, the systematic uncertainty is 1% and 4% for the categories at high and low m_{jj} respectively, while for gluon-gluon fusion is about 7% for both.

The systematic uncertainties due to pile-up jet ID are evaluated using the data-Montecarlo differences parameterized as a function of transverse momentum and pseudorapidity, obtained for the $Z(\rightarrow \mu\mu)$+jets sample (see figure 5.4). Signal events are reweighted using these data/Montecarlo efficiency ratios. To have an even more conservative approach when the ratio is larger than one, no reweighting is applied. The resulting uncertainty corresponds to 2.7% for $m_{jj} > 500\,\text{GeV}$ and 4.2% for $m_{jj} > 250\,\text{GeV}$ for the VBF component, while for gluon gluon it corresponds to 2.6% for $m_{jj} > 250\,\text{GeV}$ and 4.2%.

Theoretical uncertainties further affecting the kinematic of selected jets is evaluated with two different tests:

1. Evaluate the systematics related to the underlying event, comparing different tunes. The DT6, P0, ProPT0, ProQ20 tunes are compared with the Z2 tune, which is our default. Samples of 120 GeV Higgs bosons generated with Pythia in both VBF and gluon-gluon fusion production modes are considered. We study the efficiency variation of a common selection between the two categories, i.e. the full jet selection with $m_{jj} > 250\,\text{GeV}$ and $p_T(j_2) > 20\,\text{GeV}$. In addition we evaluate the category migration between the two categories based on the $m_{jj} > 500\,\text{GeV}$ cut. This first uncertainty has to be considered as correlated (positively) between the two categories and migration happens with respect to the other inclusive categories. The second one is assumed to be anti-correlated between the two.

   The correlated part of the efficiency uncertainty is about 7% for VBF and 50% for gluon-gluon fusion; the bin migration part is about 8% (with respect to the number of events in the lowm_{jj} category) for VBF and 15% for gluon-gluon fusion.

2. Evaluate the impact of different parton distribution functions on jet kinematics. Three different PDF sets are used: CT10 [150], MSTW2008 [129], and NNPDF2.0 [151].

   For each set the variation of α_s (in the range 0.116–0.120) is also considered. We estimate the systematics by evaluating the change in efficiency of the selection (including acceptance). The final systematics is the sum in quadrature of the systematics of each set. The total systematics due to the PDF uncertainties is about 1.5% for both VBF and gluon-gluon fusion modes.

Table 5.3 shows the summary of the systematics uncertainties on the jet selection. The total is 10% and 70% for the VBF and gluon-gluon fusion processes respectively.

5.1.4 Results and Invariant Mass Fits

Background model fits to the diphoton mass distribution, on the hypothesis of no signal, are shown for the two dijet event categories in figure 5.10. 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 $m_{\gamma\gamma}$.

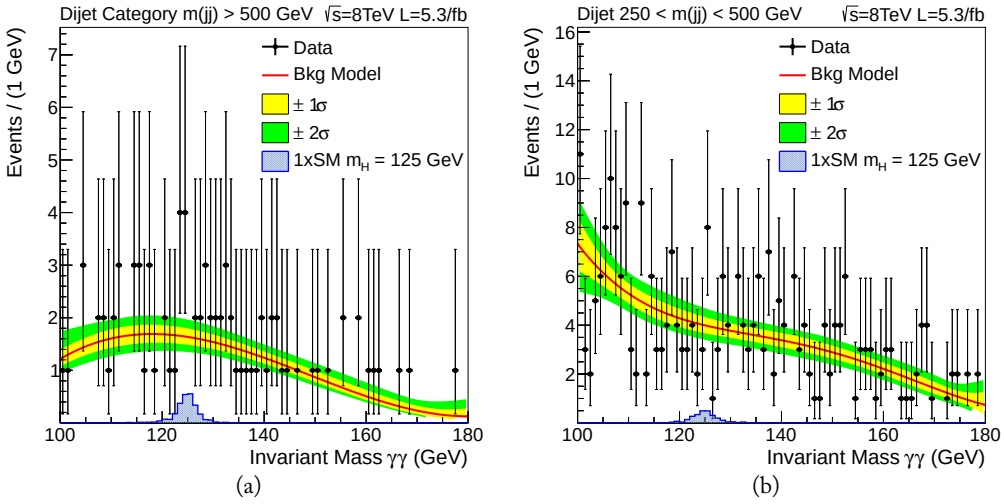
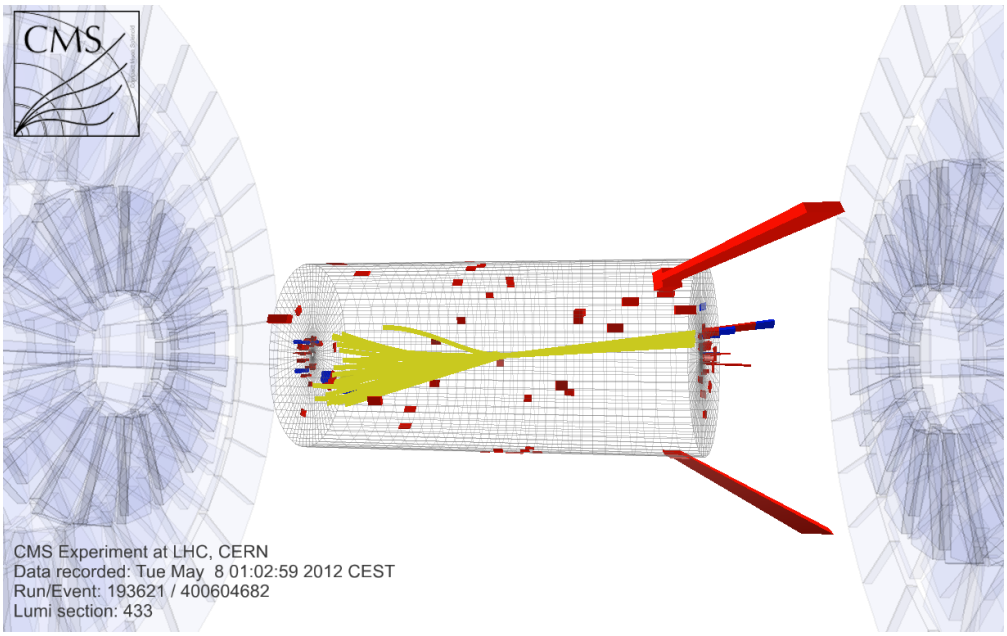


Figure 5.10. Background function fitted to data in the two dijet tagged categories. High Dijet Invariant Mass ($m_{jj} > 500$) is shown in (a), Low Dijet Invariant Mass ($250 < m_{jj} < 500$ GeV) is shown in (b).

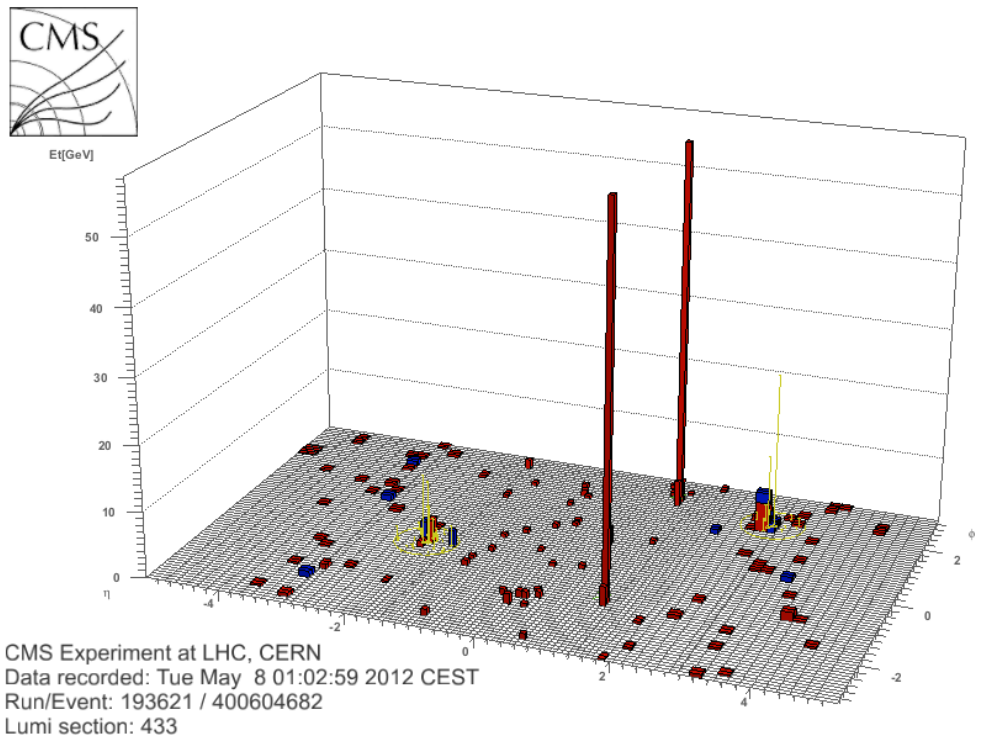
Table 5.3. Summary of systematics uncertainties on jet selection.

Systematics	Vector Boson Fusion		Gluon Gluon Fusion	
	high M_{jj}	low M_{jj}	high M_{jj}	low M_{jj}
JEC† & JER*	4%	1%	7%	7%
JET ID	2.7%	4.2%	2.7%	4.2%
tunes correlated	7%		50%	
tunes anti-correlated	8%		15%	
PDFs	1.5%	1.5%	1.5%	1.5%
total correlated	9%		51%	
total anti-correlated	8%		15%	

JER: Jet Energy Resolution
†*JEC: Jet Energy Correction*



(a) 3D View



(b) Calorimetric Towers

Figure 5.11. Display of a candidate Higgs event produced via vector boson fusion.

5.2 Final State With Associated Large Missing Transverse Energy

In this section we present the analysis of the Higgs boson decaying into two photons with associated large transverse missing energy (E_T^{miss}). The specific production mechanism which can result in high E_T^{miss}in the final state is the Higgs-strahlung from a W or Z ($qq \to VH,\ V = W, Z$), i.e. the Higgs produced in association with a weak vector boson.

We recall here the cross section and branching ratio values (listed in table 4.1) useful for this analysis. The cross section of the inclusive ZH production (assuming $m_H = 125\,\text{GeV}$) is

$$\sigma(ZH) = 0.3943\,^{+5.1\%}_{-5.0\%}\ pb^{-1},$$

calculated at next-to-next-to-leading order in QCD processes and at next-to-leading order in electroweak processes. To compute the yield of this process, the branching ratio of the invisible Z decay [94]

$$BR(Z \to \nu\nu) = (20.00 \pm 0.06)\%$$

must be considered, together with the $H \to \gamma\gamma$ branching ratio.

The WH production cross section has a slightly higher value:

$$\sigma(WH) = 0.6966\,^{+3.7\%}_{-4.1\%}\ pb^{-1}.$$

And the leptonic W decay producing E_T^{miss}has the following branching ratio [94]:

$$BR(W \to \ell\nu) = (10.80 \pm 0.09)\%\ .$$

Not all of the WH events producing high $E_T^{miss}$are selected by this analysis. If in the event there is a high quality muon or electron passing the selection criteria described in section 5.3, the event is tagged as *leptonic*, and it is no longer considered to be E_T^{miss}-tagged. The choice is driven by the idea of assigning each event always to the purest category.

5.2.1 Selection Criteria

The requirements for an event to be selected are:

1. the presence of two photons passing quality cuts described in section 3.1;
2. large corrected missing transverse energy, $E_T^{miss} > 70\,\text{GeV}$;
3. the event must not be tagged as being in the VBF di-jet category;
4. the event must not be tagged as being in the VH leptonic category.

The photon selection is the same as for the other categories, but the p_T cuts are optimized for this channel. The p_T thresholds are:

5. $p_T(leading\ \gamma) > 45\,\text{GeV} \times \dfrac{120\,\text{GeV}}{m_{\gamma\gamma}}$
6. $p_T(subleading\ \gamma) > 25\,\text{GeV}$.

The proposed selection is summarized in table 5.4.

The majority of the events selected after the full set of cuts is ascribed to the WH and ZH production mechanisms. In particular, the 49% of the total selected yield comes from the W leptonic decay, and the 28% comes from the invisible Z decay. The other bosonic decay channels contribute marginally: 1% from the hadronic W decay, 2% from the leptonic Z decay and 1% from the hadronic Z decay. Contributions from other production mechanism are also present: 5% from the VBF production and 16% from the gluon-gluon fusion production.

The lepton-tag analysis and the E_T^{miss}-tag analysis both aim at selecting events where the Higgs is produced in association to a vector boson, but the priority of the E_T^{miss} category is lower with respect to the leptonic category. Therefore special care must be put in evaluating the signal fraction of the VH samples actually being selected here.

A dedicated generator-level study shows that ∼40% of the WH events where the W decays leptonically are in common between the two analyses. And that there is no overlap in events where the Z decays to neutrinos.

Table 5.4. Kinematical selection applied in the E_T^{miss}-tag analysis.

Variable	Criterion
p_T(leading)	$> 45\,\text{GeV} \times \frac{120\,\text{GeV}}{m_{\gamma\gamma}}$
p_T(subleading)	$> 25\,\text{GeV}$
E_T^{miss}	$> 70\,\text{GeV}$

5.2.2 Background Model

The aim of this study is to quantify possible biases in the extraction of the signal peak over the continuous $m_{\gamma\gamma}$ background. The difference with respect to the same study performed in the other event categories is that the the Montecarlo statistics is too low, hence it is difficult to disentangle a potential systematic bias from a statistical precision effect. For this purpose an independent data sample is constructed, which can mimic the kinematics of final selection of the analysis.

We require one of the two photon candidates to pass the standard photon identification selection (see section 3.1), whereas the second photon to satisfy a much looser photon identification, (whose requirements are summarized in table 5.5), and to fail the standard selection. This control sample is enriched in backgrounds with one fake photon (e.g. QCD, γ + jet), but contains only a small amount of background with two real photons (e.g. box, born processes). The implication is that there are differences, especially in transverse momentum, between the control sample and the selected sample. To take into account these differences, the control sample is reweighted as a function of the diphoton transverse momentum.

The weights are extracted by comparing the control sample and the selected sample after the requirement of having two photons above the thresholds of the analysis ($p_T/m_{\gamma\gamma} > 45/120$ for the leading photon and $p_T > 25$ GeV for the subleading photon). The resulting $m_{\gamma\gamma}$ distribution for the control sample compared with the selected sample is shown in figure 5.12 for different E_T^{miss} selection criteria. The top-left plot represents the closure test

of the method. The statistics of this sample correspond to almost six times that of the selected sample.

Table 5.5. Photon isolation selection criteria used in the E_T^{miss} control sample.

Variable	Cut
H/E	$H/E < 0.15$
HCAL iso	$\Sigma E_{HCAL} < 10\,\text{GeV}$
HCAL iso relative	$\Sigma E_{HCAL}/p_T(\gamma) < 0.005$
ECAL iso	$\Sigma E_{ECAL} < 10\,\text{GeV}$
ECAL iso relative	$\Sigma E_{ECAL}/p_T(\gamma) < 0.012$
trk iso	$\Sigma E_{trk} < 10\,\text{GeV}$
trk iso relative	$\Sigma E_{trk}/p_T(\gamma) < 0.002$
$\sigma_{\eta\eta}$	$\sigma_{\eta\eta} < 0.017(0.04)$ for EB (EE)

The control sample is used to test possible shapes for the background fit. First, we fit several functions to the weighted control sample data, spanning the most common models used to describe a smooth falling background: power laws, exponentials and polynomials. The fitted functions, together with the fit χ^2, are shown in figure 5.13. According to the χ^2-probability values, we select a simple exponential (equation 5.2a), a simple power law function (equation 5.2b) and second order Bernstein polynomial (equation 5.2c) as the best candidates.

$$f(x) = A \cdot e^{bx} \tag{5.2a}$$

$$f(x) = A \cdot x^k \tag{5.2b}$$

$$\begin{cases} f(x) = \sum_{\nu=0}^{n} \beta_\nu \, b_{\nu,n}(x), & n = 2 \\ b_{\nu,n}(x) = \binom{n}{\nu} x^\nu (1-x)^{n-\nu}, & \nu = 0, \ldots, n \end{cases} \tag{5.2c}$$

To test whether the choice of a given function is introducing a bias in the signal extraction, we use the fitted function as a model to generate toy Montecarlo samples of background events. Each toy sample is then fitted to other potential background models. The fit is performed adding to the test function a signal shape with a floating normalization. The number of background events for each single toy experiment is extracted from a Poisson distribution with mean value equal to the number of observed events in data for the final selection. The statistics is then always comparable to the observed in the data.

The bias of the function is estimated by measuring the fit mean value, the pull distribution and the mean value normalized to the background error estimation in a window of [-FWHM,FWHM] of the signal, as obtained when signal normalization is left floating in the fit. Both background-only and signal-plus-background toys with different signal strength are used to measure the bias.

The background model which has the least estimated potential bias is the third order Bernstein polynomial.

As it can be seen in Figure 5.14, the measured bias for the third order polynomial is always below 20% of the statistical error from the fit and of the error on the background normalization. We adopt this function to model our background given its small bias relative

to statistical uncertainty. Larger biases up to 30% are indeed observed for the second order polynomial.

5.2.3 Optimization of E_T^{miss} Selection

To find the proper E_T^{miss} threshold to be used in the analysis, we perform an optimization procedure.

Firstly, we check if we can gain any sensitivity by splitting the E_T^{miss}-tagged events in categories, and by applying different E_T^{miss} thresholds to different categories. We know that the energy resolution of both electromagnetic and hadronic objects worsen as a function of η because of the higher detector occupancy due the underlying event. Therefore we split the events in two classes: "EB-EB", where both photons are reconstructed in the ECAL barrel ($max(|\eta|) < 1.5$), and "¬(EB-EB)", where at least one photon is reconstructed in the ECAL endcap ($max(|\eta|) > 1.5$)).

Comparing E_T^{miss} distribution in the two categories (figure 5.15(a)), we see that the ¬(EB-EB) events have a longer tail with respect to EB-EB events. But this tail is definitely too large to be physical: we expect a E_T^{miss} degradation of few percent, while we observe the end-point of the distribution to be almost doubled.

To investigate if the E_T^{miss} tail is due to some physics process, we compute the difference in the azimuthal angle ($\Delta\phi$) between the E_T^{miss} direction and the photon direction. This variable is shown in figure 5.15(b). The red distribution (EB-EB category) increases as a function of $\Delta\phi$, and the behavior is perfectly compatible with a diphoton background sample, where the events are mostly composed by one prompt and one fake photon. The fake photon being usually a mis-identified jet that is responsible also for the fake E_T^{miss}. On the contrary, the blue distribution (¬(EB-EB) category) shows a clear peak around $\Delta\phi = 0$. This feature suggests that the E_T^{miss} is aligned with one of the two photons, i.e. there is an issue in the E_T^{miss} reconstruction when a photon in the endcap is involved.

To avoid the E_T^{miss} threshold being pushed too high because of this issue, we choose to neglect the ¬(EB-EB) category in this analysis. However, since the expected signal yield in the endcap is only ∼30% of the total, and since the photon energy resolution is more than a factor two worse in the Endcap than in the barrel (cf. table 6.2), the loss in sensitivity is quantified to be below 3%. It's worth to mention that a possible solution currently under investigation is to keep only the ¬(EB-EB) events passing a minimal $\Delta\phi$ cut.

The optimization procedure of the E_T^{miss} threshold is based on running the full analysis several times changing only the E_T^{miss} cut. Each signal and background sample undergoes the full selection separately to measure the cut efficiencies. The resulting $m_{\gamma\gamma}$ spectra are then summed up applying a different weights for each sample. Where the weights depend on the sample cross section (often including a k-factor accounting for higher order processes not fully simulated), the total luminosity and the selection efficiency.

We fit the overall signal and background distributions using respectively a sum of two Gaussian and a third order polynomial. The fit to the signal is straightforward, while the fit to the background requires more care. The issue is that some Montecarlo samples suffer of low statistics, and large weights are applied to compensate for it. Clearly this issue becomes more critical as the E_T^{miss} threshold increases. The error on events coming from different samples is therefore very different, and it has to be taken into account when fitting the

total background shape. The background $m_{\gamma\gamma}$ shape obtained with the optimal 70 GeV threshold, and the relative fit are shown in figure 5.16.

From the integral of the signal and background fitted functions, we derive the signal (N_S) and background (N_B) yields at different values of the E_T^{miss} threshold. We use these yields to compute the signal efficiency (red points in figure 5.17) and the background rejection (blue points). For visualization purposes, we normalize all the quantities shown in figure 5.17 to those obtained with a 50 GeV E_T^{miss} cut.

Using the above yields, we compute a simplified version of the standard 95% upper limit on the cross section, defined as the positive strength modifier μ such that

$$\int_0^{N_B} \text{Poisson}\left(x;\, \mu\, N_S + N_B\right)\, \mathrm{d}x = 0.05 \tag{5.3}$$

Saying that a given value of μ satisfies equation 5.3, it means that the analysis is able to falsify a process having a cross section equal to μ times the Standard Model one. Therefore, even if we do not want to falsify the Higgs hypothesis, the upper limit is a good metric to test the goodness of a given selection criteria.

The green curve and the right axis of the plot in figure 5.17 show the upper limit on the cross section as a function of the E_T^{miss} threshold. For the sake of coherence, also this quantity is normalized to the 50 GeV E_T^{miss} cut. The minimum of the curve is around 70 GeV, therefore we chose this value as the best E_T^{miss} threshold to be used in the analysis.

5.2.4 Systematic Uncertainties

We make a conservative estimate of the overall systematic error affecting the E_T^{miss}-tag analysis by computing the ratio (scale factor) between the efficiency of the E_T^{miss} selection applied to data and to simulation after all the E_T^{miss} corrections have been applied. We already mentioned that the signal Montecarlo statistics is too low to perform precision studies, so we choose to use a Drell-Yan (DY) $Z \to \mu^+\mu^-$ control sample. Such sample can mimic the final state topology of the $ZH \to \gamma\gamma\nu\nu$ decay by adding the transverse momentum of the two muons to the corrected E_T^{miss}:

$$E_{T,\mu\mu}^{miss} \doteq |\overrightarrow{E_T^{miss}} + \overrightarrow{p_T(\mu_1)} + \overrightarrow{p_T(\mu_2)}|$$

The scale factor is computed as a function of the $E_{T,\mu\mu}^{miss}$ threshold applied in the selection.

Control Sample Event Selection

In the $Z \to \mu^+\mu^-$ process, the muons are selected from muon candidates reconstructed as *global muon*, i.e. reconstructed combining Pixel, Tracker and Muon Station data. To ensure a good p_T measurement and to suppress muons originated by in-flight decays of other particles, we require a minimum of 10 hits in the Tracker layers and at least 1 Pixel hit. Segments reconstructed in at least two Muon Stations are also required for muon identification to suppress the hadronic punch-through and accidental track-to-segment matches in the Muon Station. We only select muons which pass the relative isolation cut of 0.15 for muons with transverse momentum greater than 20 GeV. The relative isolation being defined as the ratio between the sum of Iso_{TRK}, Iso_{ECAL} and Iso_{HCAL} (within a ΔR cone size of 0.3) and $p_T(\mu)$. To further ensure the quality of the selection, we also require the mass of the selected dimuon system to be within 15 GeV of the Z bosons resonance mass. All the selection criteria are summarized in table 5.6.

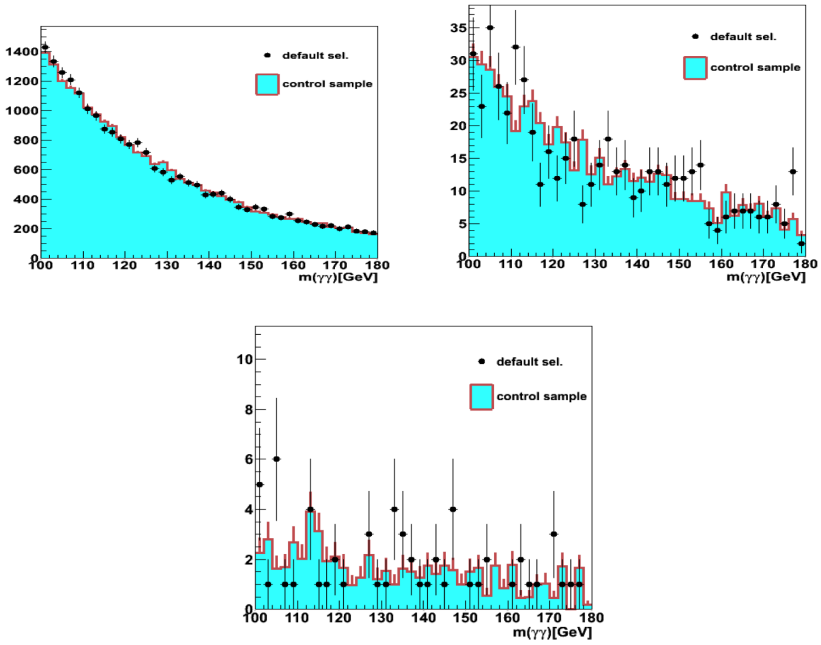


Figure 5.12. Comparison of the diphoton invariant mass between the γ + jet control sample and the selected data using different selection tightness of the E_T^{miss} cut applied in the analysis: no E_T^{miss} cut (top left), E_T^{miss} >40 GeV (top right), E_T^{miss} >60 GeV (bottom).

Table 5.6. $Z\mu\mu$ Muon Selection Criteria in the Drell-Yan $Z \to \mu^+\mu^-$ control sample.

Variable	Criterion
Tracker hits	hits ≥ 10
Pixel hits	hits ≥ 1
Involved Muon Stations	stations ≥ 2
$p_T(\mu)$	$p_T(\mu) > 20$
Relative Isolation	$(\mathrm{Iso_{Trk}} + \mathrm{Iso_{Ecal}} + \mathrm{Iso_{Hcal}})\,/p_T(\mu) < 0.15$
Dimuon Invariant Mass	$\lvert m(\mu\mu) - m(Z)\rvert < 15\,\mathrm{GeV}$

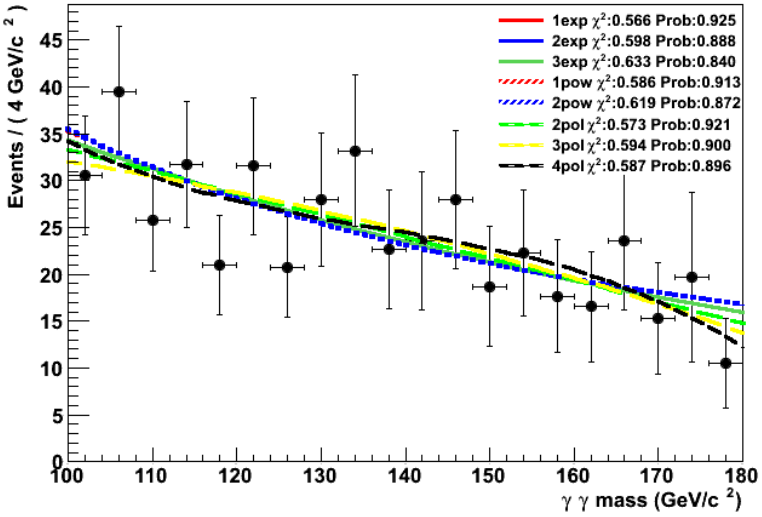


Figure 5.13. Fit of different models to the γ + jet control sample. Events entering the plot passed a E_T^{miss}>70 GeV selection.

Event reweighting

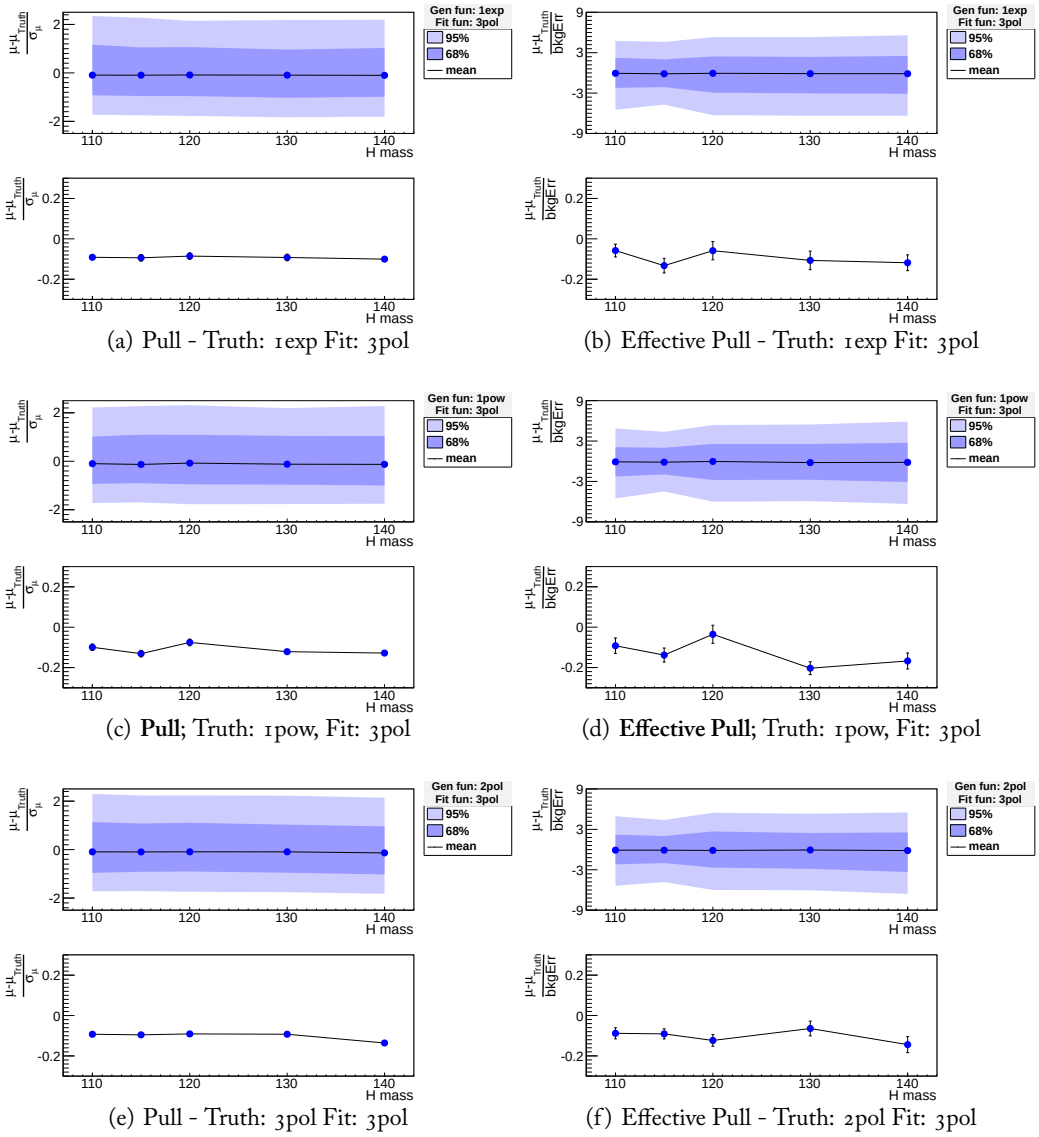
To reproduce the kinematic of the ZH process, the events in the DY sample are reweighted. The reweighting is done as a function of the Z transverse momentum $p_T(Z)$ and pseudorapidity $\eta(Z)$. Due to the low statistics of the ZH Montecarlo sample, the two-dimensional distribution of the ZH process must be smoothed with a fitted two dimensional analytic function. The analytic function is obtained by fitting the transverse momentum distribution in 7 different η slices and then interpolating between each η bin to get a full two-dimensional function. The first 6 slices corresponds to events with $|\eta|$ ranging from 0 to 3 incremented in steps of 0.5, while the 7th slice contains events in the range 3<$|\eta|$<4. Figure 5.18 shows the fits in all the pseudorapidity bins, where the momentum distribution is fitted using the following function:

$$f(x) = \ell x^2 e^{-\frac{x}{mx+k}} . \tag{5.4}$$

For each parameter there are 7 values corresponding to the 7 η slices; each of them is drawn and fitted as a function of $|\eta|$, as shown in figure 5.19. Plugging the $\ell(|\eta|)$, $m(|\eta|)$ and $k(|\eta|)$ functions into equation 5.4, we get a smooth two-dimensional functional form of the $p_T(Z)$ and $|\eta(Z)|$ distribution of the Z bosons from the Higgs associated production.

The weights are computed from the ratio of the $(p_T(Z), \eta(Z))$ two-dimensional distributions of the ZH process and of the DY process. Both distributions are normalized before constructing this ratio. Figure 5.20(a) and 5.20(b) show respectively the event weights $w(p_T, \eta)$ for data and Montecarlo samples.

The DY control sample shows a good data-Montecarlo agreement in the $p_T(\mu\mu)$ spectrum before reweighting (figure 5.21(a)). To check that the reweighting procedure goes in the right direction, and that the agreement is preserved, the $p_T(\mu\mu)$ distribution of the reweighted $Z\mu\mu$ data sample is compared with the $p_T(Z)$ distribution of the Montecarlo ZH sample. The comparison is shown in figure 5.21(b) and it confirms that the control sample behaves as the target distribution.



(a) Pull - Truth: 1exp Fit: 3pol

(b) Effective Pull - Truth: 1exp Fit: 3pol

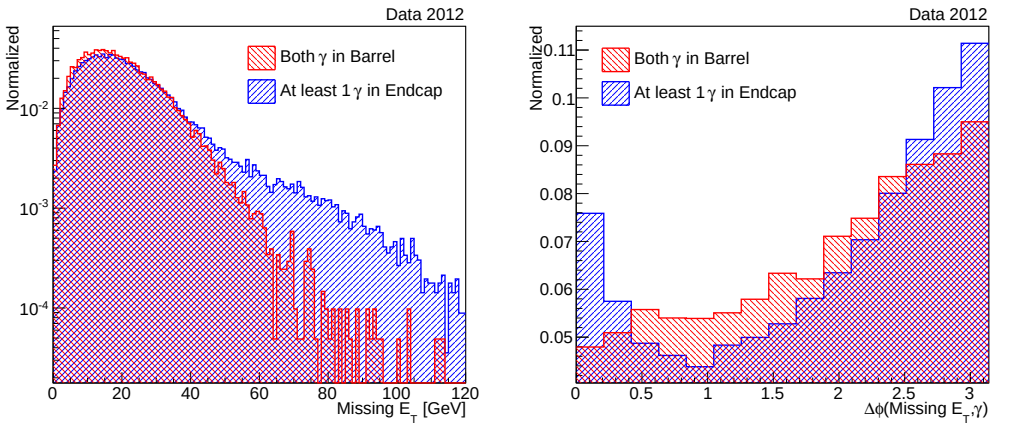
(c) **Pull**; Truth: 1pow, Fit: 3pol

(d) **Effective Pull**; Truth: 1pow, Fit: 3pol

(e) Pull - Truth: 3pol Fit: 3pol

(f) Effective Pull - Truth: 2pol Fit: 3pol

Figure 5.14. Left (a,c,e): average values of the pull distribution of the floating signal yield when fitting toys experiments extracted from various truth models (1exp,1pow,1pol) with a third order polynomial and a floating signal shape. Results are obtained with background only toys. Right (b,d,f): the same test but normalizing the value of the fitted signal to the background normalization error in a [-FWHM,FWHM] interval around the signal (Effective Pull). Regions containing 68% and 95% of the distributions are shown as well.



(a) E_T^{miss} Magnitude (b) Azimuthal Angle between E_T^{miss} and Photon

Figure 5.15. Normalized comparison in data between events having both photons reconstructed in the ECAL barrel, and events where at least one photon is reconstructed in the ECAL endcaps.

The same scrutiny is performed using the $E_{T,\mu\mu}^{miss}$ variable In figure 5.22(a) the distribution of this variable is shown for unweighted data and Montecarlo $Z\mu\mu$ samples, showing good agreement. After reweighting, the same variable is compared with the standard corrected E_T^{miss} of the ZH sample, and also in this case a nice data-simulation agreement is shown (figure 5.22(b)).

Determination of the Scale Factor and of its Uncertainty

Using the $Z\mu\mu$ artificial E_T^{miss} variable ($E_{T,\mu\mu}^{miss}$) computed in the DY reweighted sample, we determine the signal efficiency as a function of the $E_{T,\mu\mu}^{miss}$ cut for both data and Montecarlo samples. The two efficiencies are then used to compute the scale factor, defined as the ratio between data efficiency and Montecarlo efficiency at a given cut value. Figure 5.23 shows the distribution of the scale factor as a function of the cut value applied to the $E_{T,\mu\mu}^{miss}$ variable. The 70 GeV cut on E_T^{miss} used in this analysis corresponds to a scale factor of 0.958 ± 0.003. Since we use the scale factor to evaluate an overall systematic uncertainty on the signal efficiency, such uncertainty is roughly 4%.

The uncertainty on the scale factor is computed by varying the value of the parameters of the fit during the interpolation with respect to the pseudorapidity. The parameters from the fit are allowed to vary around their mean values with a Gaussian distribution where the width is the uncertainty on the parameters obtained from the fit.

From each set of the parameter values, a new two-dimensional analytic function is obtained and the same procedure as above is used to compute a new value of the scale factor. We repeat this process using 500 toys, which is sufficient to provide a Gaussian distribution in the scale factor value. The values of the scale factor from the toys are then plotted and fitted with a Gaussian. The mean and width of the fitted Gaussian are then taken as the final value of the scale factor and its uncertainty at a chosen cut value of E_T^{miss}. The process is repeated for every cut value of E_T^{miss} to obtain a continuous spectrum as shown in Figure 5.23.

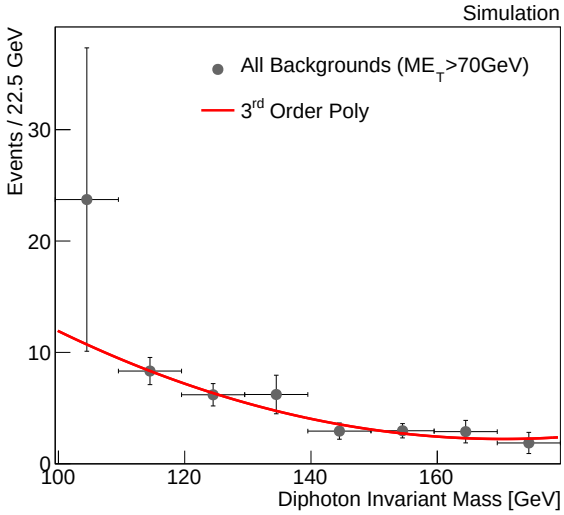


Figure 5.16. Fit to the simulated $m(\gamma\gamma)$ background used in the optimization procedure. The plot is done using the optimal E_T^{miss} threshold of 70 GeV. All the background samples have been scaled to an equivalent luminosity of 3.0 fb^{-1} and summed together.

5.2.5 Results and Invariant Mass Fits

The background model fit to the diphoton mass distribution, on the hypothesis of no signal, is shown for in figure. 5.24. The error bands show the uncertainty on the background shape associated with the statistical uncertainty of the fit. 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 $m_{\gamma\gamma}$.

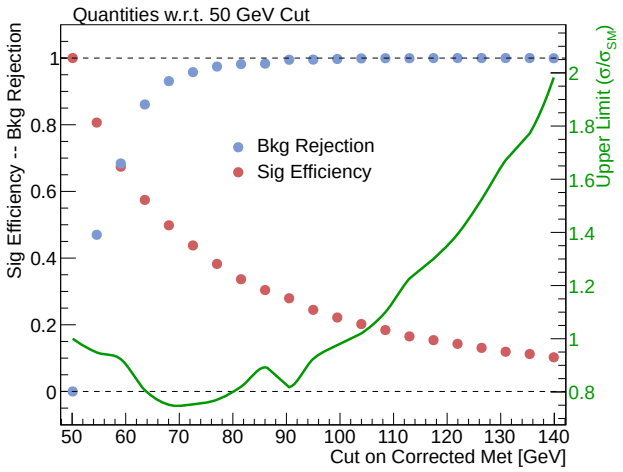


Figure 5.17. Optimization of the E_T^{miss} threshold. Left axis shows the signal efficiency (red points) and the background rejection (blue points) as a function of the E_T^{miss} threshold. Right axis shows the upper limit on the cross section as a function of the E_T^{miss} threshold. All the quantities are normalized to the 50 GeV E_T^{miss} selection.

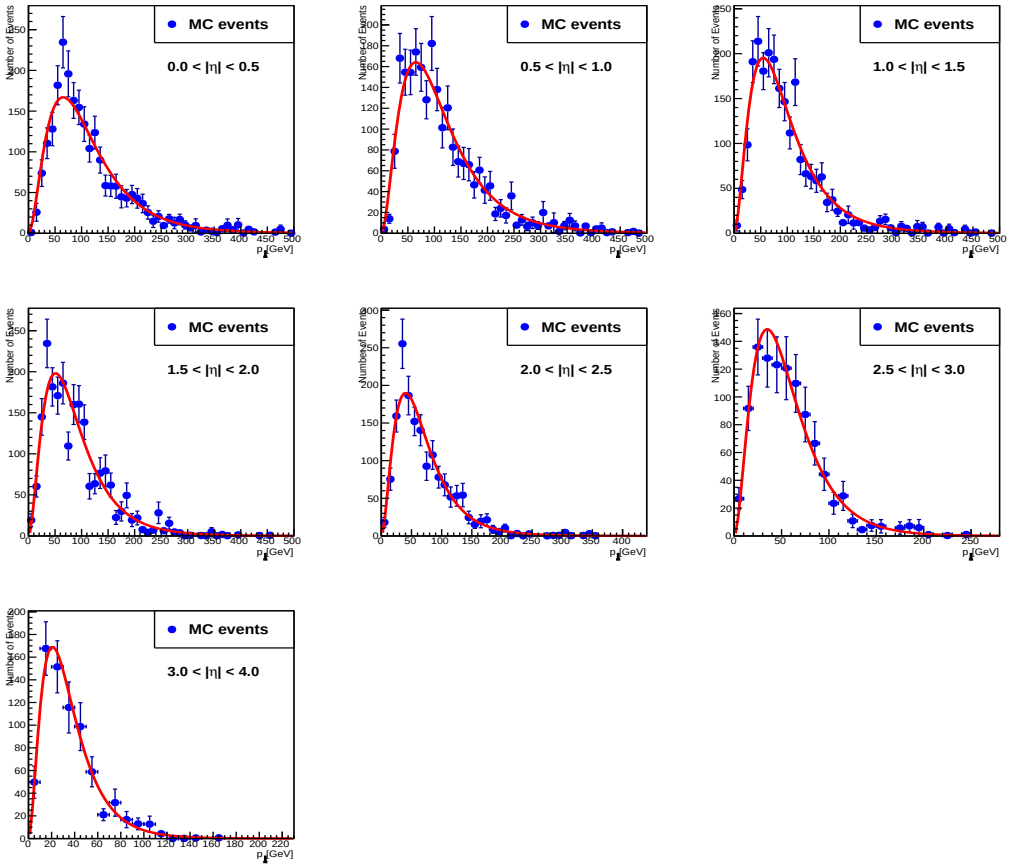
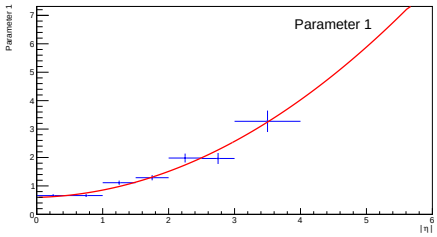
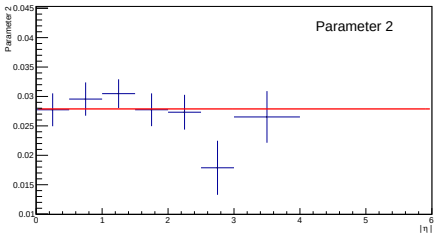


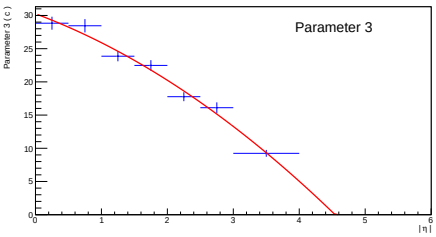
Figure 5.18. Distribution of the Z boson transverse momentum in a ZH Montecarlo sample divided in 7 pseudorapidity bins. Function in equation 5.4 is fitted to the data points.



(a) Values of the parameter ℓ in eq.5.4

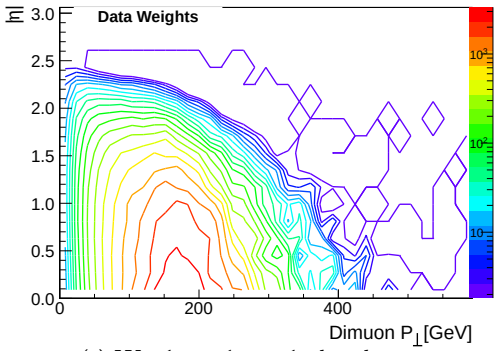


(b) Values of the parameter m in eq.5.4

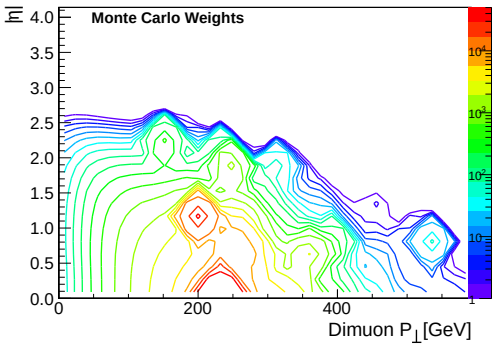


(c) Values of the parameter k in eq.5.4

Figure 5.19. Result of the fit performed on the Z transverse momentum distributions showed in figure 5.18. Each plot shows the value of the fitted parameter (ℓ, m, k) in equation 5.4 as a function of the Z pseudorapidity bin. The values of the ℓ and k parameters are in turn fitted with a second order polynomial. The values of the m parameter is fitted with a constant.

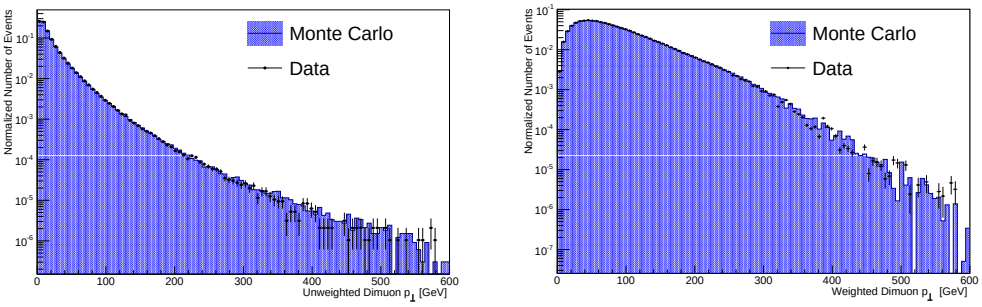


(a) Weights to be applied to data



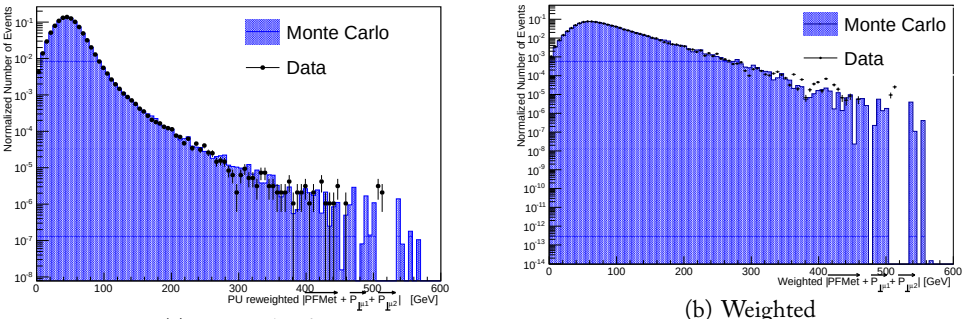
(b) Weights to be applied to simulation

Figure 5.20. Event weights to be applied the Drell-Yan $Z \to \mu^+\mu^-$ control sample to mimic the kinematic of the Z boson in ZH events.



(a) Unweighted (b) Weighted

Figure 5.21. Dimuon transverse momentum distribution in $Z \to \mu^+\mu^-$ events before and after the Z kinematic reweighting. Figure (a) shows the comparison between **unweighted** $Z\mu\mu$ events in 2012 data (Data) and the dimuon Drell-Yan simulation (Monte Carlo). Figure (b) shows the comparison between **reweighted** $Z\mu\mu$ events in 2012 data (Data) and the ZH signal simulation (Monte Carlo).



(a) Unweighted (b) Weighted

Figure 5.22. Missing transverse energy distribution in $Z \to \mu^+\mu^-$ events before and after the Z kinematic reweighting. Figure (a) shows the comparison between **unweighted** $Z\mu\mu$ events in 2012 data (Data) and the dimuon Drell-Yan simulation (Monte Carlo). Figure (b) shows the comparison between **reweighted** $Z\mu\mu$ events in 2012 data (Data) and the ZH signal simulation (Monte Carlo).

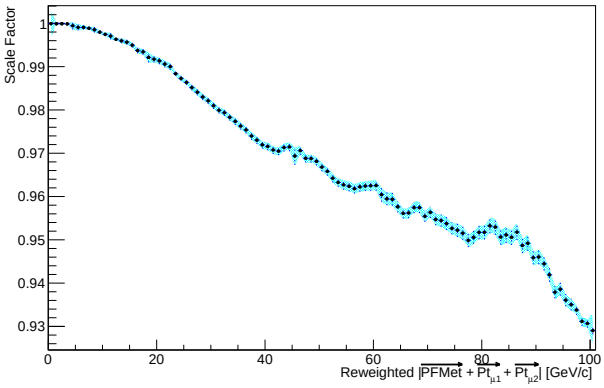


Figure 5.23. Scale factor corresponding to missing transverse energy cut value from 0 to 100 GeV with corresponding error bars

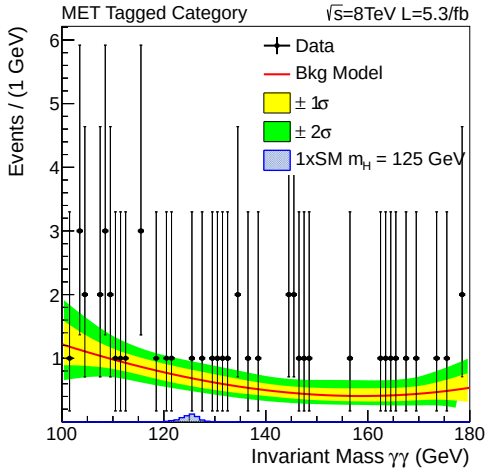


Figure 5.24. Background function fitted to data in the E_T^{miss} tagged category.

5.3 Final State With at Least One Isolated Lepton

We consider here 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 process. In this analysis we only consider muons and electrons as tagging leptons. The tagging lepton is defined as the leading lepton in the event. Its presence strongly suppresses QCD background processes. The main remaining background comes from electroweak processes with photons where a lepton is produced in Z or W decays, as summarized in Table 4.4. The VH rate associated with the lepton tag is suppressed by the leptonic decay branching fractions of the vector bosons: i.e. $\mathcal{B}(Z \to ll) \sim 7\%$, $\mathcal{B}(W \to l\nu) \sim 20\%$ In contrast to the inclusive analysis, the signal-to-background for the lepton tagging is very large, close to one.

5.3.1 Selection Criteria

The photon event selection is almost the same as the one used in the inclusive analysis except for two differences:

- the p_T of the Higgs boson produced via VH is typically larger than the p_T of the Higgs produced via gluon-gluon fusion and also larger than $p_T(\gamma\gamma)$ in the background events. Therefore, we use a higher E_{T} requirement for the leading photon and a lower E_{T} requirement for the subleading photon, compared to the inclusive selection, to retain VH-like events: $E_T(leading)/m_{\gamma\gamma}$>45/120 and $E_T(trailing)$>25 GeV;
- the electron veto: the main instrumental background comes from events where an electron is identified as a photon due to a non reconstructed/selected track. For this reason the electron veto is tightened: besides the conversion safe electron veto, described above, a cut on the distance between the photon and the closest electron track, ΔR(γ,track), is applied. ΔR(γ,track)>1 is required for all photon categories. A Montecarlo based study shows that the major background ($Z\gamma$) in electron channel can be reduced by a factor of 2 by imposing this cut [152].

Muon tag

The muon tag selection requires the presence of at least a muon passing the muon identification criteria described in section 3.4.

In addition, to reject the photon produced as a final state radiation in events with a vector boson and only one photon, such a muon should be far from both selected photons. The chosen cut is: $\Delta R(\gamma, \mu) > 1$.

Electron tag

Events not passing the muon tag selection undergo the electron tag selection.

This selection requires the presence of at least an electron passing the electron identification criteria described in section 3.4.

Again, to reject the photon produced as a final state radiation in events with a vector boson and only one photon, such electron should be far from both selected photons. The cut is the same as for the muon tag $\Delta R(\gamma, e) > 1$, which must be satisfied for both photons.

Since $Z\gamma$ events, with $Z \to e^+e^-$, are the most important background in the electron tag events, in order to reduce them, a further dedicated cut is applied. As already said,

the major background comes from events where one 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 $|m(e,\gamma) - m_Z| > 5$ GeV, this cut is applied to both selected photons. Fig. 5.25 shows the minimum $|m(e,\gamma) - m_Z|$ for the two selected photons.

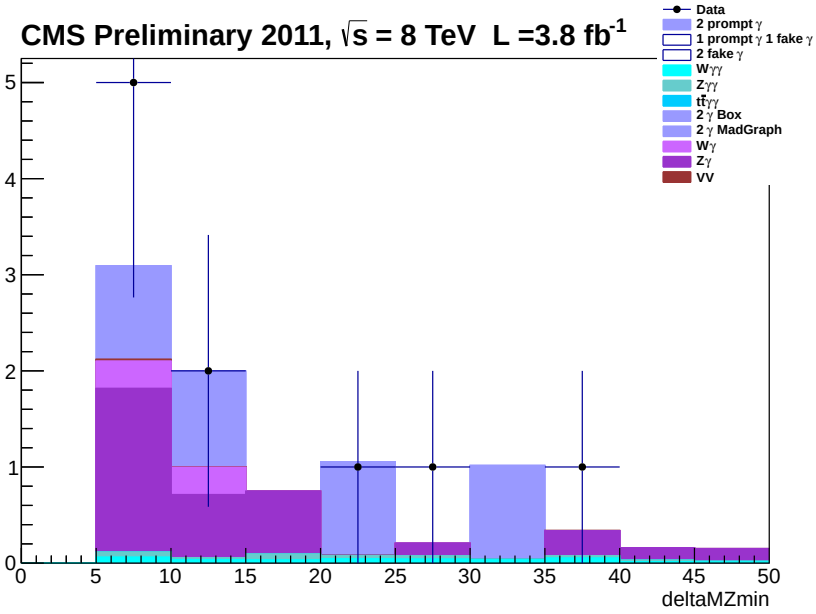


Figure 5.25. Distribution of the minimum $|m(e\gamma) - m_Z|$ for the two selected photons.

5.3.2 Background modeling

We extract the background model by fitting to the data in the mass region 100–180 GeV. We discussed in section 4.5 that experimental effects originating from photon ID and HLT may distort the inherent background shape and may be present in the distribution of the selected events. But the yield of the lepton tag analysis is too low to estimate the fit bias by testing several fit function directly on data.

The choice of the background fit function is made by fitting the Montecarlo simulation, in case of the muon tag, and from fitting a combination of data and simulated events, in case of the electron tag. For the electron-tagged events, a control sample (CS) is derived from data by requiring one of the photons to be matched with a track. This CS represents the reducible background with enhanced statistics.

The outcome of the study is that the small number of selected events allow the use of a power law function without introducing any significant bias as compared to the statistical uncertainty.

5.3.3 Results and Invariant Mass Fits

The background model fits to the diphoton mass distribution, on the hypothesis of no signal, are shown in figure. 5.26. 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 $m_{\gamma\gamma}$.

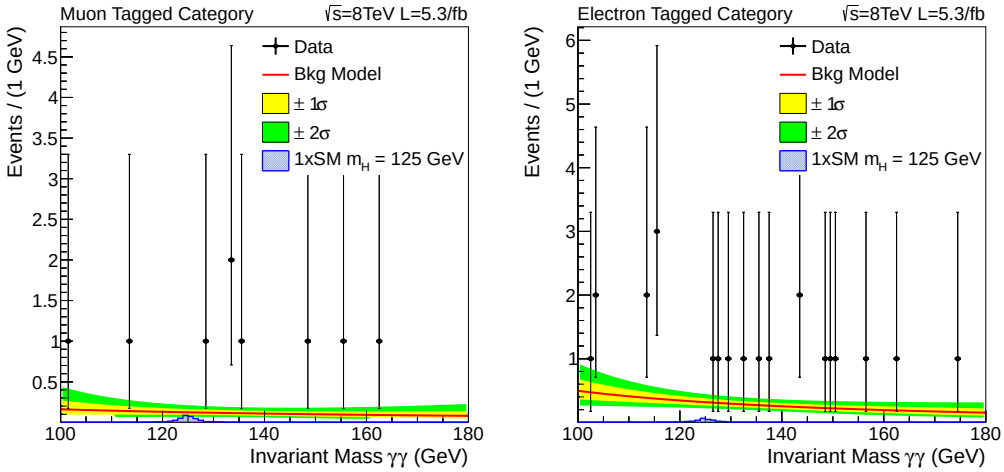


Figure 5.26. Background function fitted to data in the two lepton tagged categories. Muon tagged category is shown in (a). Electron tagged category is shown in (b).

Chapter 6

Measurement Of The Higgs Couplings

The best strategy to investigate the nature of the Higgs boson and to report the experimental results is by adopting a model-independent bottom-up approach. The most general description is based on the formalism of chiral Lagrangians, supplemented by a few minimal assumptions motivated by the experimental information at our disposal. The chiral Lagrangian introduced in [36] and extended by [37] (for earlier related work see [153–155]) fully characterizes the interactions of a light Higgs-like scalar under the following conditions:

- new physics states, if present, are heavy and their effect at low-energy can be encoded by local operators in the chiral Lagrangian
- the electroweak symmetry breaking (EWSB) dynamics possesses an (at least approximate) custodial symmetry
- there are no flavor-changing neutral currents mediated at tree-level by the Higgs.

The first assumption implies in particular that there are no new particles to which the Higgs boson can decay. It can be easily relaxed by including in the Lagrangian any new light state that should be discovered. The request of a custodial symmetry is strongly motivated by the absence of corrections to the ρ parameter measured at LEP and implies that the couplings of the Higgs to the W and the Z must be equal. Under these hypotheses, the interactions of a single Higgs-like scalar are characterized in terms of a set of parameters which describe the couplings to the SM fermions and the electroweak gauge bosons plus new contact interactions to a pair gluons or photons (as for example generated by loops of heavy scalar of fermionic top partners). Such a parameter space includes the SM as a specific point, and is sufficiently generic to describe scenarios where the Higgs-like scalar is not part of an $SU(2)_L$ doublet or is not even related to the EWSB, as in the case of a light dilaton [156–159].

We restrict, for simplicity, to the case in which single Higgs interactions can be parameterized in terms of only two independent parameters: the coupling to two gauge bosons, $c_V = g_{HVV}/g^{SM}_{HVV}$, and the coupling to two fermions $c_F = g_{H\psi\psi}/g^{SM}_{H\psi\psi}$. New contact interactions mediated by heavy new physics are assumed to be small and to have a negligible impact on the Higgs phenomenology.

By means of the chiral Lagrangian, we write two equations linking respectively c_V and c_F to the signal yields of the exclusive event categories (see section 6.3). Since the coupling

measurement is subject to the knowledge of the signal yields, we perform the measurement through the following three-steps procedure:

1. we build a signal model using Montecarlo simulation to describe the expected width of the invariant mass peak after all the photon corrections have been applied. The signal model also allows to describe the impact of the good/wrong vertex choice on the invariant mass resolution. Using such model, we estimate the overall signal efficiency×acceptance;

2. we perform a fit to the total signal cross section, showing the contribution of the exclusive channels, showing their role in improving the sensitivity to the cross section measurement, and measuring in data the so called *strength modifier*, namely the ratio between the measured Higgs cross section and the expected one;

3. we perform a simultaneous fit of the signal model + data-driven background shape to all the event categories. We quantify the signal and background yields by category. And finally we plug the measured signal yields into equations 6.2 to determine the coupling values.

6.1 Signal Model

To unfold the observed data and measure the Higgs cross section, we build a parametric signal model using simulated $H \to \gamma\gamma$ samples. The model accounts for both the smearing of the photon energy resolution and the photon energy correction factors, in order to quantify the overall efficiency×acceptance as a function of the Higgs mass. The model is said to be parametric since describes the signal shape with an analytic function. The model is defined continuously for any value of the Higgs mass between 110 and 150 GeV: the parameters of the analytic function are determined by fitting the signal invariant mass for each available simulated mass point, and the full signal model is defined by a linear interpolation of the fitted parameters across the available mass values.

As input to the fits, we use individually the signal Montecarlo for each mass point 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 Standard Model.

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 category by a sum of 2 or 3 Gaussians, depending on the categories. The means, widths, and relative fractions of the Gaussians are left free in the fit to the Montecarlo.

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 category by a sum of 2 Gaussians, or a single Gaussian, depending on the categories. The means, widths, and relative fraction (for categories with two Gaussians) are left free in the fits to the Montecarlo. A representative set of fits for events with correct and incorrect primary vertex selection in one category for gluon-fusion production are shown in figure 6.1.

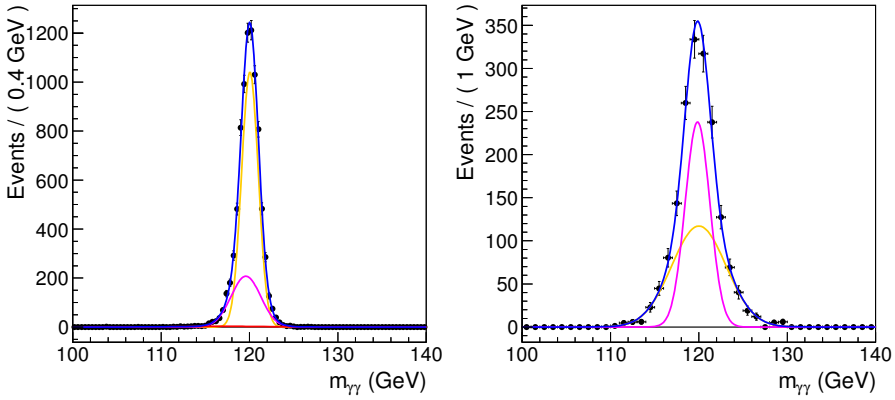
The combined shape in each event category for correct and incorrect vertex selection is constructed by adding the shapes for the two sub-components together, according to the correct vertex selection efficiency determined from Montecarlo. This efficiency is treated as another model parameter for the 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 categories, we parameterize the signal yield in terms of a per category acceptance$\times$ efficiency, computed from Montecarlo after all of the appropriate weights and scale factors have been applied. The breakdown of the efficiency $\times$ acceptance values per category is shown in table 6.1, while the overall values as a function of the Higgs mass are shown in figure 6.2. The final parameterized shapes for each category for a Higgs mass of 125 GeV for the Standard Model cross-section weighted mixture of all production mechanisms are shown in Figure 6.3. The shape of all the categories combined together is shown in figure 6.4.

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.

Table 6.1. Breakdown of the efficiency $\times$ acceptance percent values in the different event categories. Values are relative to $m_H = 125$ GeV.

Category		Efficiency by Production Mode (%)			
		Gluon Fusion	Vector Boson Fusion	Associated Production	*tth*
Muon Tag		<0.01	–	2.4	4.6
Electron Tag		<0.01	–	1.7	3.3
Dijet Tight		0.2	10.8	0.1	0.2
Dijet Loose		0.5	7.4	0.4	0.1
E_T^{miss}Tag		<0.1	0.1	2.4	5.7
Untagged	Cat. 0	12.4	7.9	8.4	7.8
	Cat. 1	17.1	10.1	10.4	8.3
	Cat. 2	5.9	4.2	5.6	3.0
	Cat. 3	7.7	5.0	6.9	3.4
Total		43.8%	45.5%	38.3%	36.4%



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

Figure 6.1. Fit results for signal shape model with correct and incorrect primary vertex selection for a single category in 120 GeV Gluon-Fusion Higgs Montecarlo. The black points are the weighted Montecarlo events and the blue lines are the corresponding fit results. Individual Gaussian components of the fits are also shown. This plot uses an different definition of the category boundaries and it is shown for illustration only.

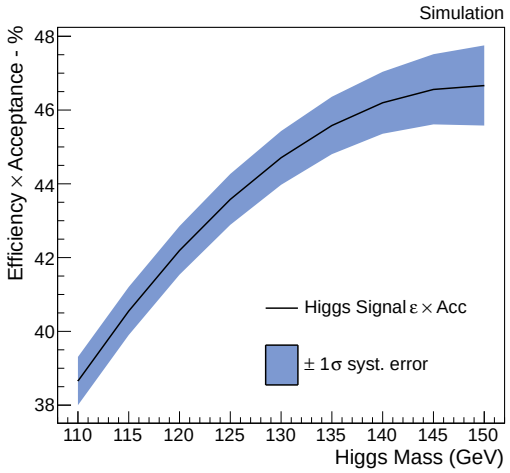


Figure 6.2. Efficiency × Acceptance of the full analysis selection (i.e. accounting for all the categories) *vs.* Higgs mass.

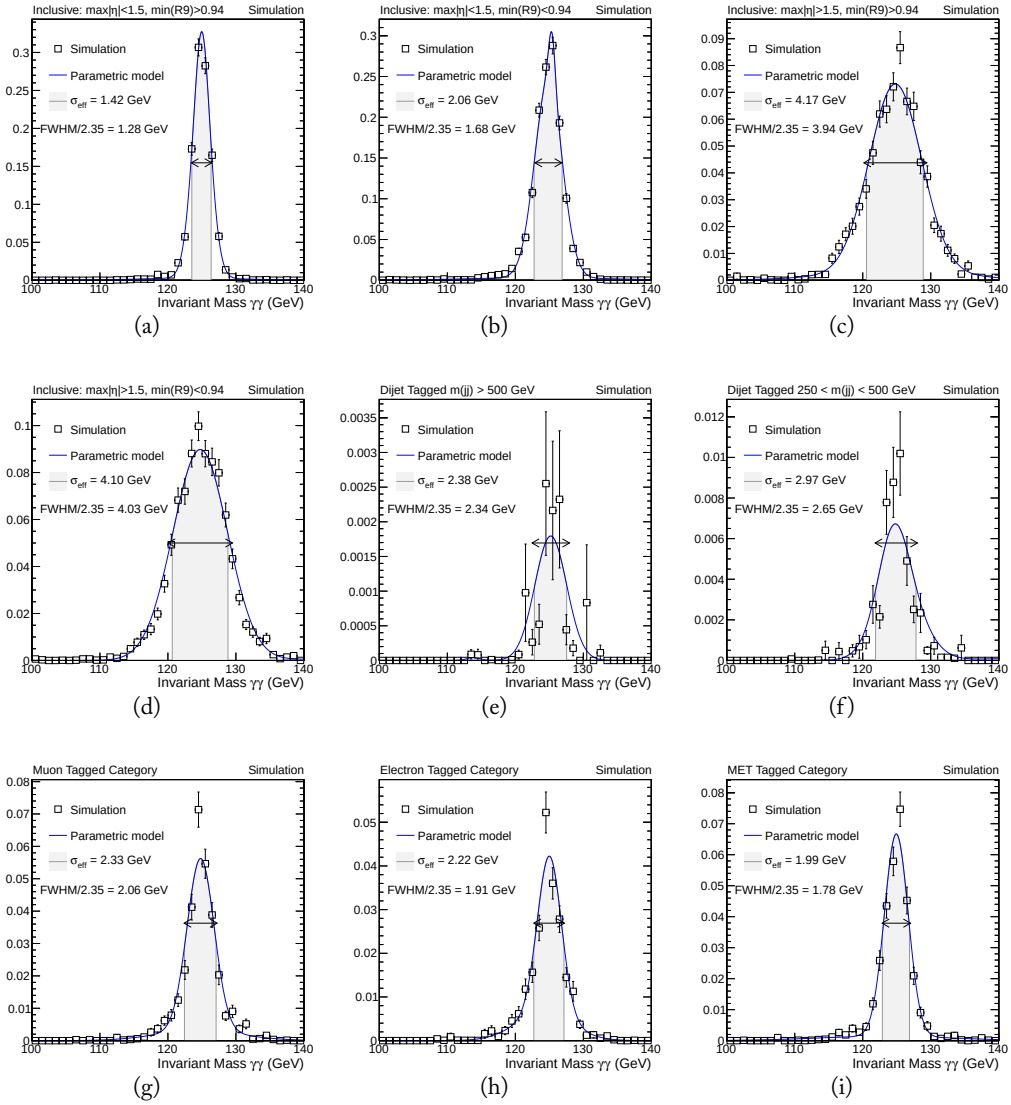


Figure 6.3. Diphoton invariant mass resolution in the different event categories, using a simulated Higgs sample where $m_H = 125$ GeV. The four inclusive categories are shown in figures from (a) to (d). The two dijet tagged categories are shown in figures (e) and (f). The muon and electron tagged categories are shown respectively in figure (g) and (h). The E_T^{miss} tagged category is shown in figure (i).

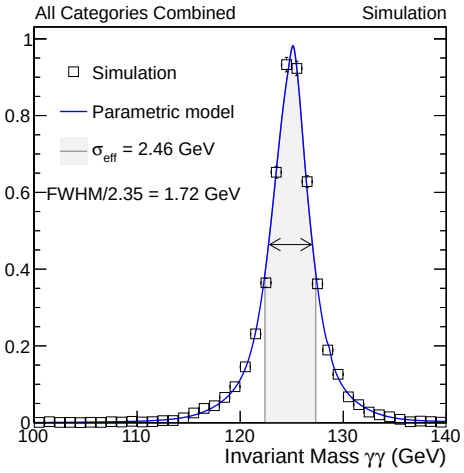


Figure 6.4. Diphoton invariant mass resolution combining all event categories. The parametric model (blue line) shows a good agreement with respect to the simulated Higgs sample ($m_H = 125$ GeV).

6.2 Category Combination

The statistical combination of the exclusive categories allows to exploit the higher signal-over-background ratios of the categories targeting the VBF and VH production modes, hence enhancing the sensitivity to the Higgs boson.

We perform a simultaneous fit of the signal-plus-background model to the nine event categories with the aim of determining the features of each category properly accounting for the inter-category correlations and systematics. On the signal side, we are interested in three parameters: 1) the normalization of the signal function, i.e. the signal yield; 2) the smallest region containing the 68% probability surrounding the signal peak (σ_{eff}) as a first estimator of the invariant mass resolution; 3) the full width at half maximum (FWHM), which is a second estimator of the invariant mass resolution once divided by 2.35. On the background side, we focus only on the integral, i.e. the background yield. This information is provided in two fashions: the total number of events fitted in the 100-180 GeV mass range, useful to estimate the statistical precision of the fit, and the number of events per GeV evaluated at $m_{\gamma\gamma} = 125$ GeV.

The breakdown of these parameters by event category is shown in table 6.2. Untagged categories have the largest signal yield (since they are sensitive to the gluon-gluon fusion production mechanism), but they also experience high background rates. The invariant mass resolution worsen noticeably when comparing diphoton events reconstructed in the barrel (categories untagged-0 and untagged-1), with events where at least one photon is reconstructed in the endcap (categories untagged-2 and untagged-3). Tagged categories show much higher signal-over-background values, but their signal yields are more than one order of magnitude lower than the untagged ones.

We use the numbers in table 6.2 to determine the improvement in the Higgs sensitivity when passing from 6 to 9 event categories. In particular we study the inclusion of the lepton and the E_T^{miss} tagged categories in addition to the untagged and dijet ones. We consider two different estimators of the analysis sensitivity: the expected upper limit and the expected

error on the Higgs cross section.

The upper limit here is defined as the cross section (σ) that the analysis is able to falsify at 95% C.L., normalized to the SM Higgs cross section (σ_{SM}), assuming no evidence for the signal. It is evaluated using a modified frequentist approach, CL_s, taking the profile likelihood as a test statistic [160, 161]. The upper limit value as a function of the Higgs mass hypothesis is shown in figure 6.5. The addition of more categories brings the upper limit curve downwards, hence showing the increased sensitivity.

To measure the expected uncertainty on the Higgs cross section, we now assume the presence of a signal at $m_{\gamma\gamma} = 125\,\text{GeV}$ and we fit the signal model to data, floating only the signal normalization. Scanning the fit likelihood as function of the strength modifier (σ/σ_{SM}), we compute the likelihood difference with respect to the best fit value ($\Delta(-\log\mathcal{L})$). This quantity is shown in figure 6.6 for three different analyses: *red*) untagged + VBF; *blue*) adding the lepton tagged categories; *green*) adding the E_T^{miss}tagged category. In the Gaussian approximation, the curvature of $\Delta(-\log\mathcal{L})$ represents the uncertainty on the cross section measurement, and the strength modifier values where $\Delta(-\log\mathcal{L}) = 2$ are the 68% extrema.

Figure 6.6(a) shows the likelihood of the fit performed on Montecarlo events. The minimum is by construction at $\sigma = \sigma_{SM}$, and the green curve — representing the 9 categories analysis — has the largest curvature. The purple curve represents the 9 categories analysis performed neglecting all the systematic uncertainties. Figure 6.6(b) shows the same fit performed on 2012 data. Now the minimum of each curve measures the strength modifier in data, and such value is different depending on the number of categories considered. In the case of the full exclusive analysis, the best fit value correspond to a cross section 1.34 times greater than the SM prediction. It's important to stress that the uncertainty on this measurement is still very large (∼80%), since the dataset considered for the analysis corresponds to half of statistics needed to claim the 5σ discovery.

Table 6.2. Expected signal yield, expected invariant mass resolution and estimated background yield by event category.

Category		**Signal** ($m_H = 125$ GeV)			**Data**	
		Yield (ev)	σ_{eff} (GeV)	FWHM/2.35 (GeV)	Fitted Background* (ev/GeV)	Yield† (ev)
Muon Tag		0.31	2.33	2.06	0.1 ± 0.03	9
Electron Tag		0.23	2.22	1.91	0.3 ± 0.05	22
Dijet Tight		2.56	2.38	2.34	$1.6^{+0.2}_{-0.1}$	84
Dijet Loose		2.68	2.97	2.65	4.0 ± 0.3	271
E_T^{miss}Tag		0.48	1.99	1.78	0.7 ± 0.1	50
Untagged	Cat. 0	31.32	1.42	1.28	107.4 ± 1.7	7031
	Cat. 1	42.85	2.06	1.68	203.7 ± 2.4	13544
	Cat. 2	15.16	4.17	3.94	109.2 ± 1.8	7205
	Cat. 3	19.63	4.10	4.03	$177.7^{+2.3}_{-2.2}$	11773

†*Evaluated in the* $(100 < m_{\gamma\gamma} < 180)$ GeV *range*

Evaluated at $m_{\gamma\gamma} = 125$ GeV

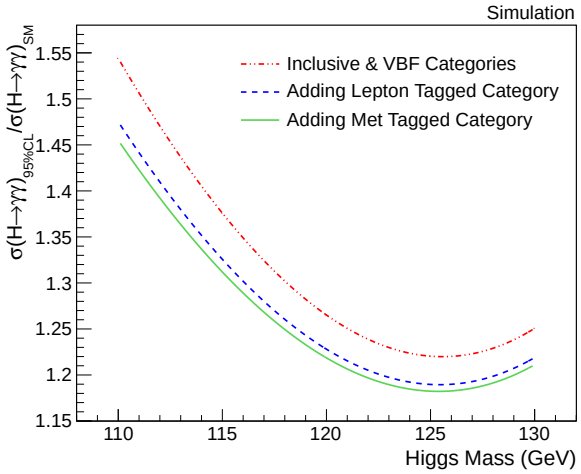
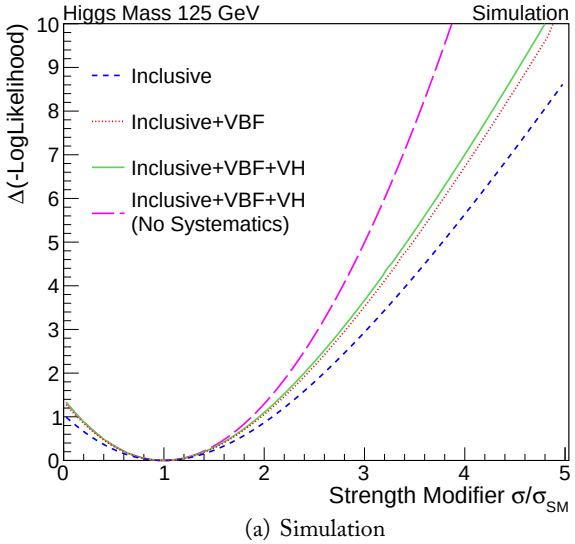
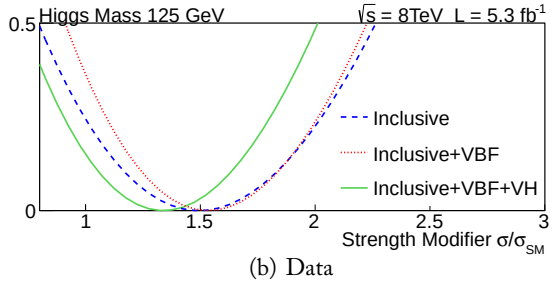


Figure 6.5. Expected upper limit on σ/σ_{SM} in simulated data samples as a figure of merit of the analysis improvement. Adding exclusive categories increases the analysis sensitivity, lowering the expected upper limit value.



(a) Simulation



(b) Data

Figure 6.6. Likelihood profile of the best fit to σ/σ_{SM} assuming $m_H = 125$ GeV. The curves show how the error on the strength modifier varies when more categories are considered. In (a) the fit is performed on a simulated Higgs sample, in (b) is performed on 2012 data, and the x axis is zoomed in around the best fit value.

6.3 Fitting the Coupling Model to Data

From an experimental perspective, we define the "coupling model" as the equations connecting the signal yields measured in the exclusive categories to the couplings of the Higgs to bosons (c_V) and fermions (c_F). These couplings are easily defined in terms of the Higgs production cross sections, as mentioned in equations 1.1. Such equations are recalled here to facilitate the discussion:

$$\sigma(gg \to H) = c_F^2\, \sigma(gg \to H)_{SM}$$
$$\sigma(qq \to qqH) = c_V^2\, \sigma(qq \to qqH)_{SM}$$
$$\sigma(q\bar{q} \to VH) = c_V^2\, \sigma(q\bar{q} \to VH)_{SM}$$
$$\sigma(gg, q\bar{q} \to t\bar{t}H) = c_F^2\, \sigma(gg, q\bar{q} \to t\bar{t}H)_{SM}$$

In the first equation, two gluons can interact only through a fermionic loop in order to produce a Higgs, hence the GGF process is sensitive only to c_F. In the VBF mode ($qq \to qqH$), the Higgs is produced by the interaction of two weak bosons and the process is sensitive only to c_V. In the associated production (VH), the Higgs boson is radiated by a weak boson, therefore also this process is sensitive only to c_V. In the last mode, the Higgs is produced through the fusion of two top quarks, hence the process is sensitive only to c_F. Each production mechanism is function of only one coupling, but the same is not true for the diphoton decay. The decay does not occur at tree level (the SM Higgs is electrically neutral), and it involves a loop where the most relevant particles are top quarks and W bosons, hence mixing c_V and c_F.

Combining together the Higgs production cross sections and branching ratios, we define the following ratio

$$\mu_i = \frac{\sigma_i \times BR(\gamma\gamma)}{[\sigma_i \times BR(\gamma\gamma)]_{SM}} \tag{6.1}$$

as the yield in a given category i in SM units. Expressing μ_i in terms of c_V and c_F we get:

$$\begin{cases} \mu_{untagged} \sim (c_F^2 + \zeta\, c_V^2)\, \dfrac{(4.5\, c_V - c_F)^2}{c_F^2} \\ \mu_{jj} \sim \mu_{1\ell} \sim \mu_{\mathrm{E_T^{miss}}} \sim c_V^2\, \dfrac{(4.5\, c_V - c_F)^2}{c_F^2} \end{cases} \tag{6.2}$$

where the factor $(4.5\, c_V - c_F)^2$ follows from the branching ratio (the numerical factors are described in section 1.2) and ζ parameterizes the contamination of VBF and VH events in the untagged categories.

It's worth to notice that in equation 6.2 the yields depend quadratically on the couplings. As a consequence, for a given signal (say c_V^0, c_F^0), there are two possible solutions:

$$c_V \simeq c_V^0\, \frac{4.5\, c_V^0 - c_F^0}{4.5\, c_V^0 + c_F^0}\,, \qquad c_F \simeq -c_F^0\, \frac{4.5\, c_V^0 - c_F^0}{4.5\, c_V^0 + c_F^0} \tag{6.3}$$

which gives the same yields μ_{jj}, $\mu_{1\ell}$, $\mu_{\mathrm{E_T^{miss}}}$ and $\mu_{untagged}$. For $(c_V^0, c_F^0) = (1, 1)$, i.e. the SM point, the second solution corresponds to $(0.64, -0.64)$.

Once inverted, equations 6.2 are fitted to the signal yields measured in data, where the floating parameters are c_V and c_F. Instead of quoting just the best fit values, we report the

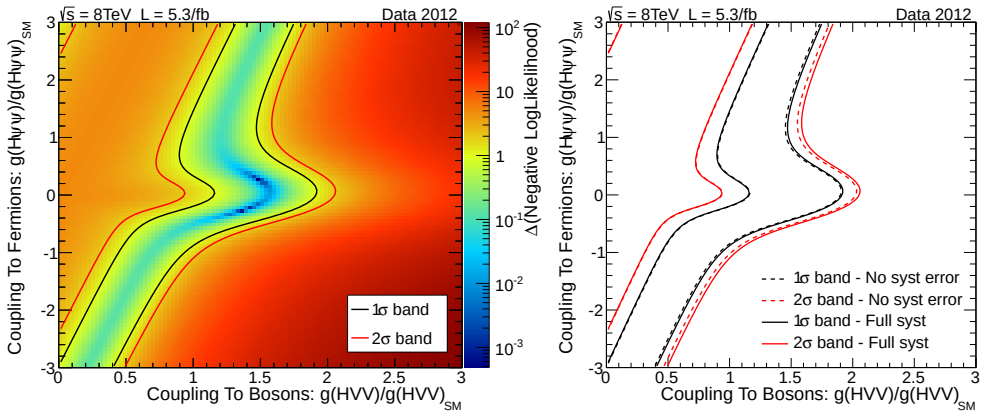


Figure 6.7. Likelihood scan of the (c_V, c_F) plane. The analysis is based on 4 untagged categories and it covers the whole 2012 data taking. The 68% and 95% contours are shown respectively as a black and a red curve. (a) Log-likelihood difference with respect to the best fit value. (b) Systematic uncertainty on the contours.

complete likelihood shape to build confidence interval in the (c_V, c_F) plane. Figure 6.7(a) shows the $\Delta(-\log\mathcal{L})$ distribution (on the z axis) in the (c_V, c_F) plane for the events selected by the inclusive analysis (i.e. in the four untagged categories). When this plot is compared to the Montecarlo simulation in figure A.1, it can be seen that the presence of a signal definitely improves the constraint of the couplings: the top left corner of the plane is now ruled out at 95% confidence level (C.L.), corresponding to the region where $c_V < 0.8$ and $c_F > 0$. The role of the systematic uncertainties in setting the contours is shown in figure 6.7(b), where the solid lines represent the full analysis, while the dashed lines are obtained neglecting all the systematics. The difference between the two regions is minimal, suggesting that the total uncertainty on such measurement is still dominated by statistics.

The $\Delta(-\log\mathcal{L})$ distribution for the VBF categories (i.e. the two dijet tagged categories) is shown in figure 6.8(a). The high abundance of events selected in these categories allow them alone to exclude at 95% C.L. any fermiophobic model where $c_V > 0.75$ and $c_F \sim 0$. With a lower confidence, it can be seen that the whole region where $c_F \sim 0$ is disfavored. Moreover, regions where both c_V and c_F are greater than the SM values seem to be preferred by experimental data. The role of systematics here is more relevant (see figure 6.8(b)), reflecting the big uncertainty on the GGF contamination of the dijet categories (see table 5.3).

As predicted by equation 6.2, the contours obtained by the VH categories (namely the two lepton-tagged categories plus the E_T^{miss}-tagged category) have the same shape of the VBF ones, as shown in figure 6.9(a). However, the lower yield of these categories result in a looser constraint in the (c_V, c_F) plane. For the same reason, the role of systematics here is almost negligible: see figure 6.9(b).

The combination of all the 9 events categories is shown in figure 6.10. Thanks to the complementary shapes of the different production modes, all the models where the Higgs couples only to bosons at tree level are now excluded at 95% C.L.

Data collected so far push in the direction of couplings slightly greater than SM ones.

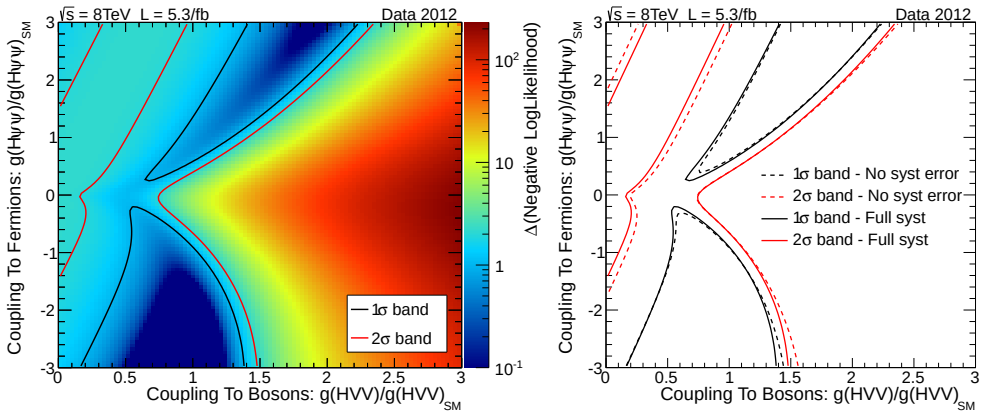


Figure 6.8. Likelihood scan of the (c_V, c_F) plane. The analysis is based on 2 dijet categories and it covers the whole 2012 data taking. The 68% and 95% contours are shown respectively as a black and a red curve. (a) Log-likelihood difference with respect to the best fit value. (b) Systematic uncertainty on the contours.

This is consistent with the best fit to the total cross section shown in figure 6.6(b), where selected events exceed those predicted by a SM Higgs simulation. The degenerate solution in the negative quadrant is far from being excluded, and the region where c_F is negative seems to be slightly favored. However the SM point $(1, 1)$ is still fully compatible with the observation.

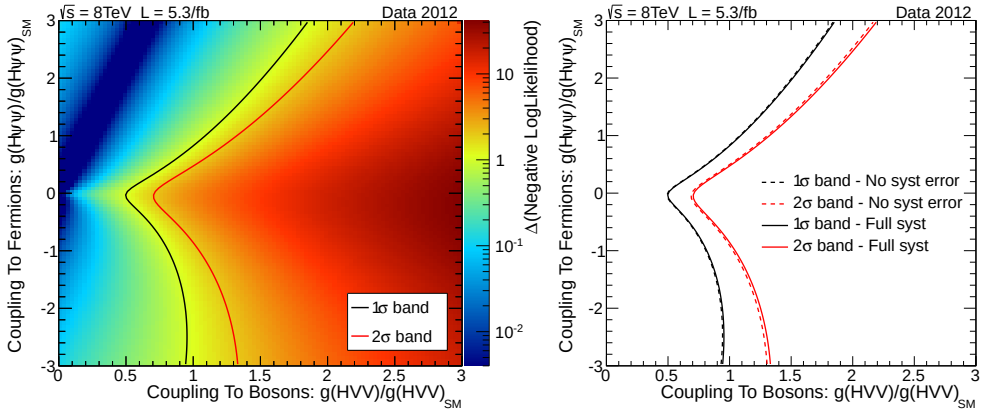


Figure 6.9. Likelihood scan of the (c_V, c_F) plane. The analysis is based on 3 VH categories and it covers the whole 2012 data taking. The 68% and 95% contours are shown respectively as a black and a red curve. (a) Log-likelihood difference with respect to the best fit value. (b) Systematic uncertainty on the contours.

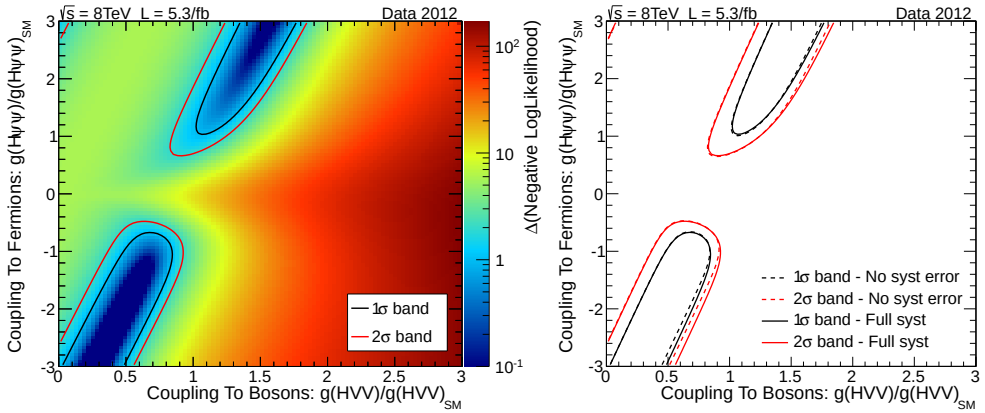


Figure 6.10. Likelihood scan of the (c_V, c_F) plane. The analysis is based on 9 exclusive categories and it covers the whole 2012 data taking. The 68% and 95% contours are shown respectively as a black and a red curve. (a) Log-likelihood difference with respect to the best fit value. (b) Systematic uncertainty on the contours.

Conclusions

In this dissertation we have presented the exclusive analysis of the Higgs boson decaying into two photons, with the aim to to disentangle the Higgs production mechanisms to constrain the couplings of the Higgs boson to fermions and to bosons.

The analyzed dataset comprises $5.3\,\text{fb}^{-1}$ of proton-proton collisions delivered by the LHC at the center-of-mass energy of 8 TeV, and recorded by the CMS detector. All the events were evaluated by two levels of trigger (hardware and software), requiring two isolated photons to be reconstructed. Such photons underwent a tight selection aimed at rejecting fake photons ascribable to mis-reconstructed jets, electrons with few tracker hits, and genuine photons coming from the decay of boosted mesons (e.g. π^0 and η). The **photon selection** was based mainly on isolation variables and on the shape of the electromagnetic shower; its efficiency was measured on data using the tag-and-probe technique on $\text{Z} \rightarrow \text{e}^+\text{e}^-$events, and was found to range from 90% (converted photons in the endcap region) up to 99% (unconverted photons in the barrel region).

The **vertex determination** is crucial for the resolution of the diphoton invariant mass, and it is particularly challenging in events where the hard scattering collision produces only two neutral objects. We employed a multivariate approach combining the transverse momentum of the diphoton system with the transverse momentum of all the other particles coming from a given vertex. If at least one of the photons was converted in a e^+e^- pair, the conversion tracks were used to further constraint the diphoton vertex. In events where $p_T(\gamma\gamma) > 50\,\text{GeV}$, the fraction of events where the vertex choice is correct was found to be greater than 90%.

The precise reconstruction of the **photon energy** is mandatory to achieve a $\mathcal{O}(\text{GeV})$ resolution in the diphoton invariant mass. When $E(\gamma) \sim 100\,\text{GeV}$, the single photon energy resolution is dominated by the constant term, which is affected mainly by the crystal intercalibration and by the crystal transparency loss due to irradiation. A crystal-to-crystal calibration was performed using the diphoton decay of π^0 and η mesons, reaching a precision better than 1% in the barrel region. The transparency loss was monitored with a laser system, injecting a fixed amount of light in all the crystals every 40 minutes. The energy corrections derived from the laser measurements allowed to reach a $\mathcal{O}(10^{-3})$ stability within the whole 2012 data taking. The clusterized photon energy was further corrected to account for the energy leaking outside the clusterization region, for the hardware gaps between the calorimeter modules, and for the average energy density in the event due to pile-up interactions.

The overall diphoton invariant mass resolution was measured to range from $4.1\,\text{GeV}$ (converted photons in the endcap region), to $1.42\,\text{GeV}$ (unconverted photons in the barrel region).

To disentangle the Higgs production mechanisms, the events were **categorized** according to the presence of *tagging* objects in addition to the isolated photon pair. Events with two forward jets were divided into two categories, according to the invariant mass of the two jets: *dijet tight* required $m_{jj} > 500\,\text{GeV}$, while *dijet loose* required $250 < m_{jj} < 500\,\text{GeV}$. The signature selected in these categories was mostly sensitive to the Higgs produced via vector boson fusion. Events with at least one lepton were divided into two categories, according to the lepton flavor: electron or muon (the tau lepton was neglected here). These two categories were sensitive to the Higgs produced in association to a vector boson, where the latter decayed leptonically. Events with large missing transverse energy ($\text{E}_\text{T}^\text{miss}$) were selected at downstream of the previous categories. The $\text{E}_\text{T}^\text{miss}$-tag was sensitive to the W, Z associated production, in cases where the W or Z boson decayed into taus, where it decayed into electrons or muons outside the acceptance region, and where the Z decayed into neutrinos. In events with no tagging objects, the Higgs was mainly produced via gluon-gluon fusion. Untagged events were further divided into four categories to account for different sensitivities.

The number of **background** events in the signal region was estimated from fitting the data in the 100-180 GeV range. The background shape was noticed to be different among the event categories, and its functional form was chosen with the aim of minimizing any potential bias in the signal extraction. The untagged categories showed the highest level of background, ranging from 100 ev/GeV (two unconverted photons in the barrel) to 200 ev/GeV (at least one converted photon in the endcap) in the $m_{\gamma\gamma} = 125\,\text{GeV}$ region. In the tagged categories, the presence of additional objects in the final state strongly suppressed the background contamination. The two dijet categories showed on average less than 3 ev/GeV, the electron and muon category showed less than 0.3 ev/GeV, and the $\text{E}_\text{T}^\text{miss}$category showed 0.7 ev/GeV.

The expected **signal yield** was calculated from simulation and it was scaled to an integrated luminosity of $5.3\,\text{fb}^{-1}$. The envelope of all the production mechanisms gave $\sim$115 signal events. The total signal efficiency $\times$ acceptance at $m_{\gamma\gamma} = 125\,\text{GeV}$ (considering all the event categories) was 44%. A slight excess of signal events was observed in data (mostly coming from the dijet categories), and the best fit to the Higgs cross section ($\times$ branching ratio) was 1.34 times the SM value. However the error on this measurement was very large ($\sim$80%) since the $5.3\,\text{fb}^{-1}$ of data considered in this analysis were not even sufficient to claim the 5σ discovery.

The formalism of the chiral Lagrangian was assumed to model the interaction between the Higgs boson and the other particles. For simplicity we restricted to the case in which the Higgs interactions could be parameterized in terms of only two independent parameters: the **coupling** to two gauge bosons (c_V) and the coupling to two fermions (c_F) with respect to values predicted by the SM. Given the different dependency of the Higgs production mechanisms on the couplings, the signal yields in the 9 event categories were used to measure c_V and c_F. A feasibility study performed on simulated samples proved that the diphoton decay channel alone was sensitive enough to constrain the couplings. The only weakness of the analysis was that a degenerate solution appeared in the coupling phase space, due to the quadratic dependence of the signal yields on (c_V, c_F), and to the interference of the 1-loop top and W contributions to the $\gamma\gamma$ decay rate. Such degeneration held even considering other possible Higgs decays (ZZ,WW) and it was shown that it could be removed only in the limit of very large integrated luminosity.

To perform the **coupling measurement**, the coupling model was fitted to data. Combining the results obtained in all the event categories, several constraints has been imposed on (c_V, c_F):

- all the *fermiophobic* Higgs models where $c_F = 0$ have been excluded at 99% C.L.;
- the value of c_F has been constrained outside the window $[-0.2; 0.3]$ at 95% C.L.;
- the (c_V, c_F) values satisfying the relation $(4.6 \cdot c_V - c_F) > 5.8$ have been excluded at 95% C.L.;
- the (c_V, c_F) values satisfying the relation $(4.6 \cdot c_V - c_F) < 2.6$ have been excluded at 95% C.L.;

The scan of the model likelihood also showed that the SM point $(1, 1)$ is not the solution preferred by the fit to data. Such point lays outside the 68% contour, and the same is true for its degenerate partner (0.64, -0.64). These observations are in agreement with the fit to total Higgs cross section ($\times$ branching ratio) mentioned previously, being slightly above the SM prediction; however it's worth to stress that the SM is still fully compatible with the results presented here[1].

It is doubtless that July 4^{th} 2012 has been a major turning point in high energy physics. We moved from the times of discovery to the equally exciting times of properties measurement, where we will have to figure out to what extent the SM holds, and what is the deep nature of the electroweak symmetry breaking. This thesis makes a first step in this direction, but many others are about to follow. The first challenge (already ongoing) will be to re-optimize the selection criteria to include the 2011 dataset, and to re-analyze the full 2012 dataset as soon as the data taking will give way to the two-year long LHC shutdown. More than $20\,\text{fb}^{-1}$ are foreseen before the end of the year, but from the extrapolation performed in our analysis it is not clear if such integrated luminosity will be sufficient to remove the (c_V, c_F) degeneracy. A further step (already in fieri) will be to combine the coupling measurement performed in the diphoton decay channel together with the other channels, in order to exploit other diagrams to span different regions of the (c_V, c_F) plane. Moreover, such combination will strongly enhance the statistics, since the data samples of the different analyses are almost orthogonal.

[1] It's worth to recall that the global electroweak fit [162] predicted $m_H \sim 80\,\text{GeV}$, and that the 125 GeV value is two sigmas away from the best fit.

Acknowledgments

Visto che sul frontespizio della tesi sono citati solo i nomi di Daniele e Shahram, un primo ringraziamento lo devo a Chiara e Paolo che, pur non essendo miei relatori, hanno dato un contributo impagabile a tutta l'attività di analisi svolta negli ultimi due anni.

Genève, nel bene e nel male, è stata il teatro di buona parte di quest'avventura di dottorato. Penso di dovere molto all'*Appa Vaucher*, in particolare alle serate a base di sigari e whiskey, al rösti che ha salvato molte cene, al magret de canard che ci ha fatto ricredere sulla cucina francese (Elvis), alle tonnellate di formaggio fatte fuori con la scusa "facciamo un aperitivo mentre aspettiamo che l'acqua bolla", alle jam sessions a base di ukulele (Francesco), alle disquisizioni interminabili sui massimi sistemi (Maria). Nella sezione *delenda est* ricordiamo: la Régie du Centre e la deplorevole M.me Cambon (per tutto il veleno che ci ha fatto buttare sul contratto della casa); Il Fornello (tremendo ristorante italiano dove si spendevano 40 franchi per pizza e birra); ladri di biciclette elvetici (perdere la mia bici da corsa nonostante fosse legata con tanto di catena è stato un lutto insanabile). Nonostante ciò, alcune persone hanno oggettivamente reso più facile (più bella) la permanenza ginevrina. Fra queste ringrazio, in ordine sparso: Diana, per avermi introdotto al concetto di *affordable luxury*; Chiara, per il *wakeboarding* in una freddissima giornata senza sole; Matteo, per un interminabile viaggio in macchina sotto la neve con un materasso sul sedile anteriore; Giorgia, per i bagni nel fiume e i concerti all'alba; Lidia, per un weekend improvvisato a Mosca.

L'ultimo anno ha invece come denominatore comune la Baita, che prima di essere un luogo di lavoro è un fantastico gruppo di persone. Il clima che si è creato fra di noi è stato un vero toccasana contro la follia e lo sclero che precede ogni consegna di tesi. A partire dalla trasferta francese di questa estate — con lo sterminio di mosche operato da Silvia, le gallette di riso con marmellata d'arancia di Giulia, l'opera lirica cantata da Andrea R sotto la doccia, la leadership casalinga di Livia, il complesso del basilico di Andrea G —; passando per il *Thanks God Is Friday* introdotto da Michael che ogni venerdì, cascasse il mondo, allo scoccare delle 6pm ci richiamava al *Beer o'Clock*; e finendo con lo scontro quotidiano (sic) contro ROOT e la statistica, combattuto al fianco di Francesco e Claudia (nei rari momenti di tregua dal volo delle bottigliette). Menzione d'onore al nostro gladiatore del T2, Ivano (detto *Ivanotalamo*), sia per il supporto 24/7 in termini di computing, sia per aver condiviso con me tutte le paturnie dovute ai primi acciacchi della nostra vecchiaia incipiente (epici i momenti in cui parlavamo dei risultati delle rispettive analisi mediche).

Ci sono poi ringraziamenti che riguardano un arco di tempo molto più lungo, e che per questo coprono interamente il trienno del dottorato. Sono solo tre. E il loro peso è inversamente proporzionale al numero di parole che dedico loro.

Uno va a Claudio e Anna Teresa. Perchè se oggi so come *stare* al mondo è solamente merito loro.

Uno va a Rita. Perché in una sera d'inverno di quasi due anni fa, seduti sulla nostra panchina del Portico d'Ottavia, mi ha fatto intuire un *si* che ha cambiato completamente il corso della nostra vita.

E uno va al mio amico Giacomo, chè specie in questi ultimi mesi mi è stato particolarmente vicino.

Appendix A

Montecarlo Feasibility Study of Coupling Measurement

In this section we present a simulation based feasibility study, which aims at estimating the precision that CMS can reach on c_V and c_F with 2012 data. It investigates two different approaches, an inclusive as opposed to an exclusive analysis of the $H \to \gamma\gamma$ channel, showing that the latter is the most powerful to measure the couplings. The event categories are similar to the official analysis (chapter 5), albeit the categorization is simplified: the VH production is accounted using only one lepton category (electron and muon are merged, while the E_T^{miss}-tag category is dropped), and the same is done with the VBF production (tight and loose m_{jj} categories are merged into one single category). Also the selection on photons, jets and leptons is simplified: the Montecarlo events are not passed through the detector simulation (GEANT4), and the cuts are applied on the generator-level quantities. The two approximations, however, are not biasing the result, since the key aspect of the analysis is to have at least one category enriched in VBF production, and one category enriched in VH production. The net effect of the approximations is to make the result more conservative. A better categorization and a better selection go in the direction of enhancing the signal-over-background ratio, hence improving the determination of the couplings. The whole analysis assumes $\sqrt{s} = 7$ TeV, because it was performed in a period where the 8 TeV samples were not yet available. The analysis also investigates the hypothesis of a fermiophobic Higgs boson [18, 163, 164], where the Higgs is assumed to be coupled only to bosons at the tree level. Such model, even if it has already been excluded [165], represents an important benchmark in the (c_V, c_F) phase space, since it spans the subspace c_F=0.

A.1 Simplified Exclusive Analysis of the $H \to \gamma\gamma$ channel

The sensitivity of the search for the Higgs boson is enhanced when events are divided into categories with different signal-to-background ratios. This division is also helpful to discriminate among different Higgs production mechanisms. In this analysis we exploit this categorization to improve the constraints in the (c_V, c_F) plane compared to an inclusive analysis.

Here we use three variables to divide events based on the kinematic properties of the $\gamma\gamma$ final state and the quality of the photon reconstruction. The first variable is the transverse momentum of the $\gamma\gamma$ system, $p_T(\gamma\gamma)$, which identifies kinematic regions with smaller

background contamination. It also enhances the sensitivity to vector boson fusion (VBF) and associated production (VH) mechanisms, which typically produce a Higgs boson with larger $p_T(\gamma\gamma)$ compared to those produced through gluon-gluon fusion (GGF). The second variable, R_9, is related to the shape of the energy deposited by the photon candidates in the electromagnetic calorimeter, and it helps to separate events with converted photons (see section 3.1.1). Finally, the third variable is the maximum pseudorapidity η of the two photons. Eight inclusive categories are defined according to the following criteria: $p_T(\gamma\gamma) > 40$ GeV and $p_T(\gamma\gamma) < 40$ GeV; large and small R_9; and whether both photons are in the central (barrel) region ($|\eta| < 1.45$) or at least one photon is in the endcap region ($|\eta| > 1.45$). Two additional exclusive categories are defined based on the presence of extra jets and leptons in the event, in order to increase the sensitivity to different production mechanisms. The first exclusive category (jj) includes events with two high-p_T jets in the forward region in addition to the photon candidates, and is thus enriched with Higgs bosons produced via VBF. The selection requires the leading (trailing) jet to have a minimum transverse momentum of 30 GeV (20 GeV). The two selected jets need to be separated in pseudorapidity ($|\Delta\eta_{jj}| > 3.5$), and to have a large invariant mass ($m_{jj} > 350$ GeV). There is also the additional requirement that the difference between the average pseudorapidity of the two jets and the pseudorapidity of the diphoton system (i.e. the candidate Higgs boson) has to be less than 2.5. The second exclusive category (1ℓ) includes events with at least one extra lepton, and it is more sensitive to Higgs candidates produced via associated production with a W/Z boson, which decays leptonically. The lepton is required to be isolated, to have a transverse momentum larger than 20 GeV, and to have a pseudorapidity which satisfies $|\eta_\ell| < 2.4$. Since the 8 categories of the inclusive analysis are not tagged by the presence of further objects in addition to the two photons, in the following we refer to such categories as *untagged*.

Diphoton background events are generated with MadGraph interfaced to Pythia and are used to estimate the fraction of background with $p_T(\gamma\gamma)$ above and below 40 GeV. This is done separately in each of the four categories defined by R_9 and by the photon pseudorapidity. We assume this fraction to be the same also for the reducible background with at least one fake photon. This is a reasonable approximation since the reducible background is about 30% of the total. For the exclusive categories, we use the background reported in [166] for the jj class and in [167] for the leptonic one. The final number of background events is obtained by performing a simple cut (consistent with the expected mass resolution) on $m_{\gamma\gamma}$ around the Higgs mass (120 GeV), which corresponds to a ± 3 GeV window for barrel-barrel photon categories and for the exclusive jj and leptonic categories, and a ± 6 GeV window for photon categories with at least a photon in the endcap. The background is thus obtained by integrating the number of events estimated from data in these windows.

Since we want to scan the (c_V, c_F) plane, signal efficiencies for each category and for each of the different Higgs production mechanisms are needed. We use Montecarlo generators to determine these efficiencies. For gluon-gluon fusion and VBF we use Powheg at next-to-leading order (NLO), while for VH we use Pythia at leading order (LO). The sum of the contributions from the different production mechanisms are then scaled to give the total number of Higgs events in the four photon categories, in the jj category, and in the leptonic category. We assume that the efficiency of the $m_{\gamma\gamma}$ cut described above is approximately 100% on the signal.

We derive our results for three different analyses:

Table A.1. Number of events (per fb^{-1}) in the 10 categories of the exclusive analysis. Signal contributions are divided by Higgs production mechanism.

			$p_T(\gamma\gamma) < 40\,\text{GeV}$				$p_T(\gamma\gamma) > 40\,\text{GeV}$			
	1ℓ	jj	BAR $R_9^>$	BAR $R_9^<$	END $R_9^>$	END $R_9^<$	BAR $R_9^>$	BAR $R_9^<$	END $R_9^>$	END $R_9^<$
GGF	0	0.14	3.23	3.40	1.20	1.44	1.55	1.64	0.58	0.69
VBF	0	0.44	0.067	0.071	0.026	0.031	0.17	0.18	0.066	0.079
VH	0.089	0.0035	0.059	0.063	0.028	0.033	0.17	0.18	0.081	0.097
Bkg	0.25	2.88	85.4	126	134	188	36.4	53.7	57.7	80.3

Table A.2. Number of events (per fb^{-1}) in the 8 untagged categories of the inclusive analysis. Signal contributions are divided by Higgs production mechanism.

	$p_T(\gamma\gamma) < 40\,\text{GeV}$				$p_T(\gamma\gamma) > 40\,\text{GeV}$			
	BAR $R_9^>$	BAR $R_9^<$	END $R_9^>$	END $R_9^<$	BAR $R_9^>$	BAR $R_9^<$	END $R_9^>$	END $R_9^<$
GGF	3.21	3.41	1.19	1.43	1.61	1.71	0.60	0.72
VBF	0.091	0.096	0.031	0.036	0.31	0.33	0.10	0.13
VH	0.067	0.070	0.030	0.036	0.20	0.21	0.089	0.11
Bkg	85.8	126	135	189	36.6	53.9	58.0	80.6

- one with 4 categories based on R_9 and photon pseudorapidity variables, which makes no use of the $p_T(\gamma\gamma)$ spectrum, as described in the previous chapter;

- one with 8 categories based on R_9, photon pseudorapidity and $p_T(\gamma\gamma)$ variables to help discriminating between different production mechanisms, thanks to the harder Higgs transverse momentum in the VBF and VH mechanisms compared to gluon-gluon fusion;

- one with 8 untagged plus two exclusive (jj and 1ℓ) categories, to fully exploit the categorization potential.

A summary of the number of background and SM signal events expected per fb^{-1} is reported in table A.1 and table A.2 for the last two analyses. In the case of the 4-category analysis, the number of events in each of the R_9 and η classes is obtained from Tab. A.2 by summing together the corresponding high and low $p_T(\gamma\gamma)$ events. Starting from the number of signal events predicted in the SM for each production mode, the number of events for arbitrary couplings c_V, c_F is easily obtained by rescaling the Higgs production cross sections and partial decay rates, as detailed in the previous section.

For each category i, given the number of signal $n_s^i(c_V, c_F)$, background n_b^i and observed events n_{obs}^i, we construct a 2D posterior probability

$$p(c_V, c_F | n_{obs}^i) = p(n_{obs}^i | n_s^i(c_V, c_F) + n_b^i) \times \pi(c_V, c_F) \tag{A.1}$$

following the Bayesian approach. The total probability is then obtained as the product of the single probabilities. The likelihood function $p(n^i_{obs}|n^i_s + n^i_b)$ is modeled by a Poisson distribution, and we take a flat prior $\pi(c_V, c_F)$ on the square $-3 \leq (c_V, c_F) \leq +3$ (vanishing outside) as done in [37]. The effect of systematic uncertainties on the signal is taken into account by letting the fraction of signal events in each category and from each production mode fluctuate. We do so by varying all the fractions with a single nuisance parameter θ_s, so that $n^i_s \to n^i_s(1 + \theta_s)$, except for the GGF fraction in the jj category which is varied with a different parameter θ_s^{GGFjj}. The total probability is then marginalized over θ_s and θ_s^{GGFjj}, which are taken to be distributed with a truncated Gaussian with zero mean and standard deviation equal to respectively $\Delta\theta_s = 0.15$ and $\Delta\theta_s^{GGFjj} = 0.70$. This corresponds to treating the systematic errors on the signal as 100% correlated in all categories and production modes, which is a reasonable approximation considering that the largest uncertainty comes from the theoretical prediction of the Higgs production cross sections, except for the GGF events in the jj category, whose largest uncertainty originates from the efficiency of the kinematic cuts applied (see section 5.1.3). We neglect all the systematic uncertainties on the background.

A.2 Results

We firstly discuss the expected 95% exclusion limits in the (c_V, c_F) plane, which are shown in figure A.1 for a center-of-mass energy $\sqrt{s} = 7\,\text{TeV}$ and an integrated luminosity $L = 5\,\text{fb}^{-1}$, approximately the amount of luminosity accumulated during 2011. One can see that the fully-exclusive analysis with 10 categories (purple solid curve) is much more powerful in the $c_F \sim 0$ region compared to the inclusive analysis with 4 categories (dotted red curve). For $c_F \to 0$ the Higgs couplings to fermions vanish and the total production cross section, which for large values of $|c_F|$ is strongly dominated by gluon fusion, receives its main contribution from VBF and W/Z associated production. An enhanced sensitivity to these production modes, as obtained by including the two exclusive event classes, can thus lead to much stronger constraints. An appreciable, though milder improvement on the limit is also obtained in the vicinity of the SM point.

Interestingly, a further subdivision of the 4 untagged categories into two sets with respectively large and small $p_T(\gamma\gamma)$ also increases the sensitivity in the fermiophobic region (dashed blue curve). This is because the distribution of the transverse momentum of the $\gamma\gamma$ pair tends to be harder for events produced through VBF and associated production, so that requiring larger values of $p_T(\gamma\gamma)$ increases the relative importance of these production modes compared to gluon fusion. We find, although the corresponding curve is not shown in figure A.1, that once the two exclusive categories optimized respectively for VBF and associated production are included in the analysis, having 8 additional untagged categories instead of 4 does not appreciably improve the sensitivity in the (c_V, c_F) plane. In other words, performing an exclusive analysis with 4+2 categories leads to constraints on the couplings c_V, c_F quite similar to those obtained with an analysis making use of 8+2 categories. This in fact agrees with the naive expectation, considering that the fraction of events produced through VBF and associated production that fall into the untagged categories is quite small: see Table A.1. To summarize, we find that an exclusive analysis of $H \to \gamma\gamma$ is more powerful than an inclusive one to set limits on the Higgs couplings, especially in regions where the importance of the VBF and associated production modes is enhanced compared

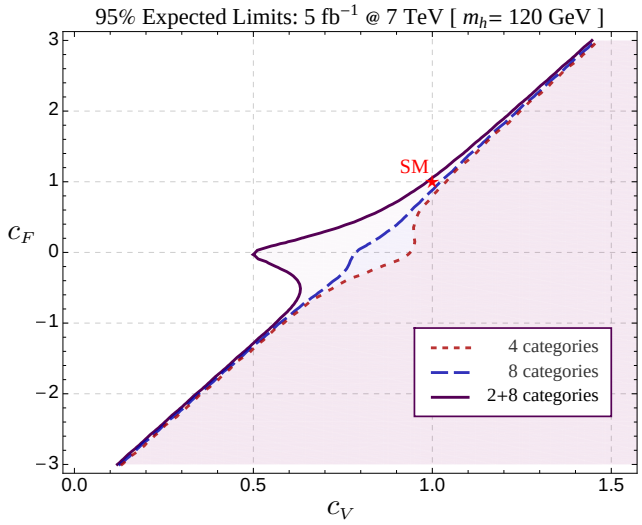


Figure A.1. Expected exclusion limits from $\gamma\gamma$ at $\sqrt{s} = 7\,\text{TeV}$ with $L = 5\,\text{fb}^{-1}$ and $m_H = 120\,\text{GeV}$. Purple solid curve: exclusive analysis with 8+2 categories; dashed blue curve: inclusive analysis with 8 untagged categories; dotted red curve: inclusive analysis with 4 untagged categories. The area on the right of each curve is excluded at 95% probability.

to gluon fusion.

A fully exclusive analysis of the $\gamma\gamma$ channel is even more useful now that the observation of a signal has been established and it comes to extracting the Higgs couplings. We illustrate this in the following by showing contours of equal probability in the plane (c_V, c_F) obtained by injecting a specific signal and assuming $20\,\text{fb}^{-1}$ of integrated luminosity [1].

Figure A.2 illustrates the case of an injected SM signal ($c_V = 1, c_F = 1$). The plot on the left shows the 68% probability contours selected by respectively the jj, 1ℓ and untagged categories. Related results were discussed in [37, 168–172], although following different approaches and assumptions than ours. The shape of the various regions can be easily reproduced considering that the yield of the two exclusive categories is dominated respectively by events produced via VBF and associated production, while the untagged categories are dominated by gluon fusion. Defining the ratio

$$\mu_i = \frac{\sigma_i \times BR(\gamma\gamma)}{[\sigma_i \times BR(\gamma\gamma)]_{SM}} \tag{A.2}$$

as the yield in a given category i in SM units, it thus follows

$$\mu_{jj} \sim \mu_{1\ell} \sim c_V^2 \frac{(4.5\,c_V - c_F)^2}{c_F^2}\,, \qquad \mu_{incl} \sim (c_F^2 + \zeta\, c_V^2) \frac{(4.5\,c_V - c_F)^2}{c_F^2}\,, \tag{A.3}$$

where the factor $(4.5\,c_V - c_F)^2$ follows from the branching ratio to $\gamma\gamma$, and ζ parameterizes the small contamination of VBF and VH events in the untagged categories. Equation A.3 reproduces to good accuracy the shape of the different regions of figure A.2. In particular, the non-negligible contribution of VBF and VH events in the untagged categories with

[1] When the study was performed, this was considered a reasonable approximation of the data set which was going to be accumulated in 2012.

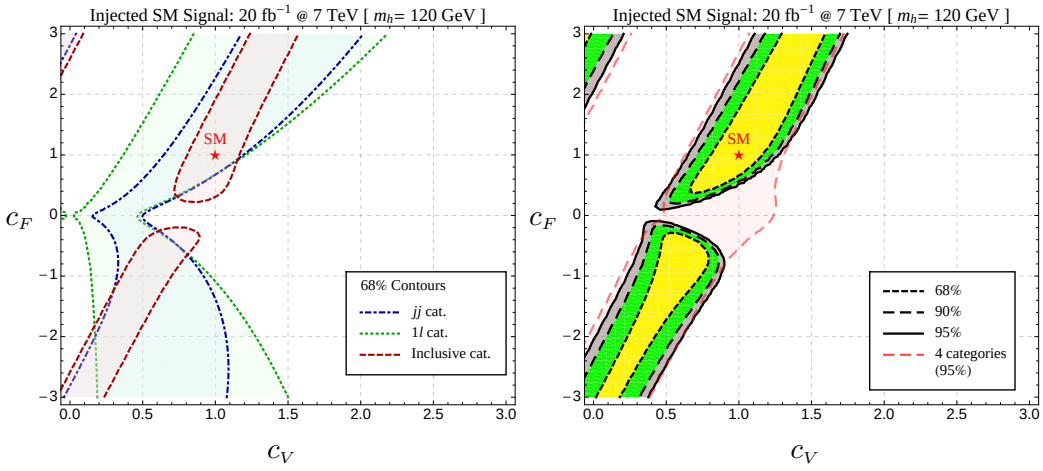


Figure A.2. Contours of constant probability for $\gamma\gamma$ in the plane (c_V, c_F) obtained by injecting the SM signal $(c_V = 1, c_F = 1)$. Left plot: 68% contours for the jj, 1ℓ and untagged categories. Right plot: 68%, 90%, 95% contours in the exclusive analysis with 8+2 categories and 95% contour in the inclusive analysis with 4 categories. Both plots are for $\sqrt{s} = 7\,\text{TeV}$ with $L = 20\,\text{fb}^{-1}$ and $m_H = 120\,\text{GeV}$.

high $p_T(\gamma\gamma)$ (see Table A.1) removes the long tail at large c_V and small c_F of the area which would be selected by the remaining four untagged classes with low $p_T(\gamma\gamma)$. The resulting 68% region selected by the combination of all the untagged categories is that shown in red in the left plot of figure A.2, which stretches along the line $(4.5\,c_V - c_F) = const$ passing through the SM point. We have checked, on the other hand, that the contamination of GGF events in the jj category modifies only marginally the shape of the 68% probability region selected by this category.

For $c_F \to 0$ the exclusive jj and 1ℓ categories favor values $c_V < 1$, which ensure a suppression of the production cross section and compensate the strong increase in the branching ratio, as required to reproduce $\mu_{jj,1\ell} \sim 1$. On the contrary, the region $c_F \sim 0$ is disfavored for any value of c_V by the untagged categories, since their yield is strongly suppressed in the fermiophobic limit. As a result, by injecting the SM signal, an exclusive analysis of $H \to \gamma\gamma$ can exclude the fermiophobic region $c_F \sim 0$ with a probability of more than 95%; see the plot on the right in figure A.2. This is especially true for the benchmark point $(c_V = 1, c_F = 0)$, which predicts too many events in the jj and 1ℓ categories and too few in the untagged ones. On the other hand, it is not possible to exclude this point and the region surrounding it by means of a 4-category inclusive analysis; see the dashed red curve in the same plot. Indeed, the total $\gamma\gamma$ yield for $(c_V, c_F) \sim (1, 0)$ is approximately that of the SM (see for example the discussion in [173]), and the overall sensitivity decreases as a consequence of the absence of the clean exclusive categories.

In order to derive an estimate of how the results in figure A.2 change with the Higgs mass, we have repeated our analysis by varying m_H and assuming that the background yield and the selection efficiencies do not change significantly. This is expected to be a reasonably accurate approximation for $m_H = 120 - 130\,\text{GeV}$. In this range of masses the variation of the signal yield is driven by the change in the Higgs production cross sections and in the $\gamma\gamma$ branching ratio, with the latter giving the dominant effect. We find that even for $m_H = 130\,\text{GeV}$ the contours of figure A.2 are only slightly modified. This is

because for $(c_V = 1, c_F = 1)$ the signal yield, hence the injected one, changes by less than $\sim 15\%$. The larger distortion occurs in the fermiophobic region $c_F \sim 0$, where the $\gamma\gamma$ branching ratio is enhanced, which is however largely disfavored by combining the inclusive and exclusive categories. We thus conclude that our results hold with good accuracy in the range $m_H = 120 - 130\,\text{GeV}$.

The exclusive analysis selects two regions with high probability: one includes the SM point, the other corresponds to negative values of c_F (yellow areas in the right plot of figure A.2). The presence of a second solution in addition to $(c_V, c_F) = (1, 1)$ is a direct consequence of the quadratic dependence of the yields in equation A.3 on c_V, c_F and the interference of the 1-loop top and W contributions to the $\gamma\gamma$ decay rate: by injecting a given signal (c_V^0, c_F^0), there is a second solution

$$c_V \simeq c_V^0 \, \frac{4.5\, c_V^0 - c_F^0}{4.5\, c_V^0 + c_F^0}\,, \qquad c_F \simeq -c_F^0 \, \frac{4.5\, c_V^0 - c_F^0}{4.5\, c_V^0 + c_F^0}\,, \tag{A.4}$$

which gives the same yields μ_{jj}, $\mu_{1\ell}$ and μ_{incl}. For $(c_V^0, c_F^0) = (1, 1)$ the second solution corresponds to $(0.64, -0.64)$, which is indeed the position of the second maximum of the 2D probability whose contours are shown in figure A.2.

The existence of a second degenerate solution in the plane (c_V, c_F) was noticed and discussed in [37, 168, 170–172]. Breaking such degeneracy will require large integrated luminosity and the combined use of several channels. An extrapolation of the results of the current searches to higher luminosity indicates that the most sensitive channels in this regard are $\gamma\gamma$ and $ZZ \to 4\ell$, while others, like WW and $\tau\tau$, are less powerful. Although performing an exclusive analysis for each decay channel will play a crucial role also in this case, a complete resolution of the degeneracy might require considering more refined strategies. This is for example illustrated by figure A.3, where we show the probability contours obtained at $L = 40\,\text{fb}^{-1}$ (the total amount of integrated luminosity which might be obtained by the end of 2012 by CMS and ATLAS together) from $\gamma\gamma$, $ZZ \to 4\ell$ and $WW \to l\nu l\nu$ (left plot) and their combination (right plot). For the WW channel we have considered the exclusive analysis performed by CMS [174] for $m_H = 120\,\text{GeV}$. In the case of ZZ we have performed a simple cut-and-count analysis by considering the number of signal and background events expected by CMS in a $\pm 5\,\text{GeV}$ window around $m(4l) = 120\,\text{GeV}$. We have constructed the posterior probability by including a 15% systematic error on the signal, while we have neglected the systematic uncertainty on the background since this is expected to be small for a shape-based analysis like $ZZ \to 4l$ (and similarly $\gamma\gamma$) once sufficient statistics has been accumulated.

As the left plot of figure A.3 illustrates, the projected sensitivity of the current WW analysis to $L = 40\,\text{fb}^{-1}$ is poor and does not help much to remove the second solution. This is due in large part to the effect of the systematic uncertainties, which are large for WW. The $ZZ \to 4l$ channel, on the other hand, is much more clean and has a strong impact in disfavoring the second solution. After its inclusion in the fit, the peak of the probability at $c_V = -c_F = 0.64$ is ~ 5 times smaller than the peak at $c_V = c_F = 1$ (see the right plot of figure A.3). We have checked that the $\tau\tau$ channel selects a broad region in the (c_V, c_F) plane, and it has very little impact on the global fit. For this reason we have not included it in figure A.3. In this regard our results do not agree with the early analysis of [176], which used a much more optimistic estimate of the background and found that $\tau\tau$ was one of the most sensitive channels for $m_H = 120\,\text{GeV}$.

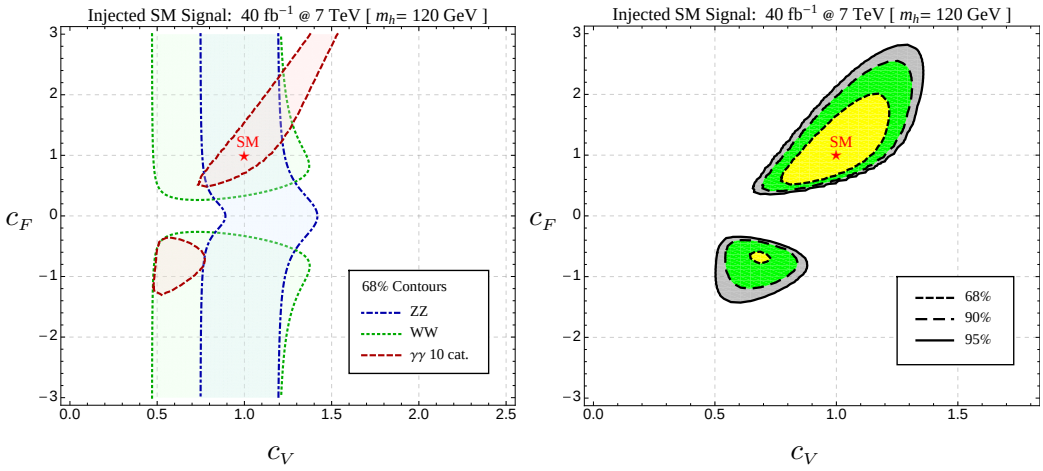


Figure A.3. Contours of constant probability in the plane (c_V, c_F) for $\gamma\gamma$, ZZ and WW obtained by injecting the SM signal $(c_V = 1, c_F = 1)$. Left plot: 68% contours for individual $\gamma\gamma$ (10-categories exclusive analysis, red area), $WW \to l\nu l\nu$ (5-categories exclusive analysis, green area) and $ZZ \to 4l$ (inclusive analysis, blue area) channels. Right plot: 68%, 90%, 95% contours for their combination. For WW and ZZ the probability function has been constructed by rescaling the number of events reported in [174] and [175].

Our results show that by extrapolating the current analyses to 40 fb^{-1} the second solution can be disfavored but not completely eliminated. A complete removal of the degeneracy will require more integrated luminosity. The use of ratios of yields in different categories within the same decay channel or different channels, as recently suggested by [170] as a way to reduce the degeneracy, does not seem to provide a resolution in this case. Its main advantage indeed is that it helps to reduce the systematic uncertainties, which are however already expected to be small for $\gamma\gamma$ and $ZZ \to 4l$. We find that by setting to zero the systematic error on the signal of both $\gamma\gamma$ and ZZ the contours of figure A.3 are marginally modified. In particular, the second solution becomes excluded at 68% but the extension of the 90% and 95% probability regions is only slightly reduced. Concentrating on the solution centered at the SM point, the plot of figure A.3 suggests that with 40 fb^{-1}, if the Higgs is that of the SM, the coupling c_V can be measured with a precision of $\sim$ 25%, while the uncertainty on c_F is of the order of 100%. Our estimate for c_V seems to be in agreement with the recent results of [177], which however reports a significantly smaller uncertainty on c_F.

We end this section by showing in figure A.4 the contours of equal probability for an injected signal $(c_V = 1/\sqrt{2}, c_F = 0)$, for $L = 20\,\text{fb}^{-1}$. We choose this point as representative of a fermiophobic scenario since it is realized in the composite Higgs model MCHM5 [178] and it was already considered in previous works. As expected from equation A.4, in this case there is no degeneracy of solutions. By performing an exclusive analysis, the maximum of the probability is obtained in a small region around $(1/\sqrt{2}, 0)$ where μ_{incl} is small and $\mu_{jj,1\ell} \sim 5$. The Higgs couplings c_V, c_F can be determined in this case with a precision of $\sim$ 35%. On the other hand, an inclusive analysis with 4 categories is dramatically less powerful and selects only a broad region in the plane (red area in the right plot of figure A.4). We checked that the same qualitative conclusions apply for $m_H = 125\,\text{GeV}$, although the uncertainty on the couplings increases to $\sim$ 45%. On the other hand, for larger Higgs

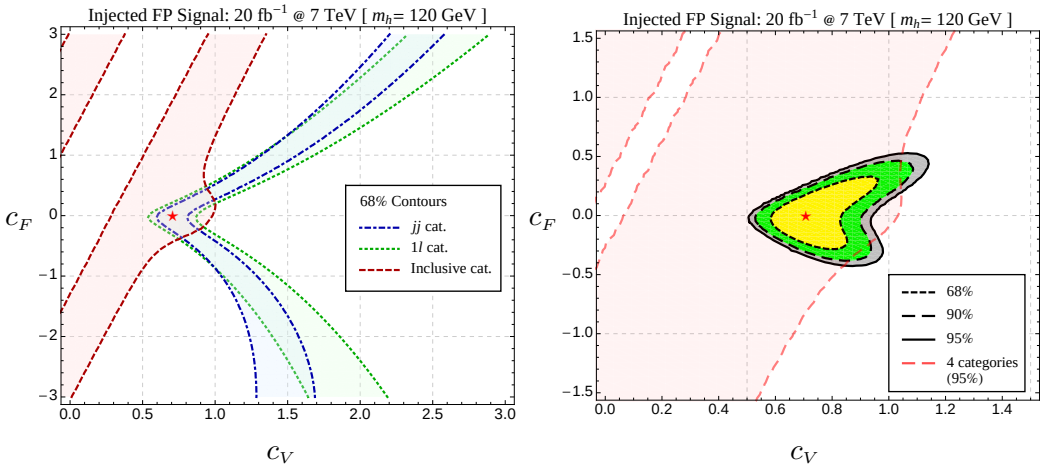


Figure A.4. As for figure A.2 with injected fermiophobic (FP) signal ($c_V = 1/\sqrt{2}, c_F = 0$).

masses the contours of figure A.4 become quickly broader, and already at $m_H = 130\,\text{GeV}$ the 90% region of the combined fit forms an open strip in the plane. This is mostly due to the decrease of the injected yield implied by the fast drop of the $\gamma\gamma$ branching ratio at heavier Higgs masses for $c_F = 0$. We thus conclude that while our results for the fermiophobic case apply reasonably well up to $m_H = 125\,\text{GeV}$, assessing the precision on the Higgs couplings at larger Higgs masses will require a dedicated analysis.

List of Tables

List of Figures

Bibliography

[1] **CMS** Collaboration, S. Chatrchyan *et al.*, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys.Lett.* **B716** (2012) 30–61, arXiv:1207.7235 [hep-ex].

[2] **ATLAS** Collaboration, G. Aad *et al.*, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys.Lett.* **B716** (2012) 1–29, arXiv:1207.7214 [hep-ex].

[3] F. Englert *et al.*, "Broken symmetry and the mass of gauge vector mesons", *Phys. Rev. Lett.* **13** (1964) 321.

[4] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys. Lett.* **12** (1964) 132.

[5] P. W. Higgs, "Broken symmetries and the masses of gauge bosons", *Phys. Rev. Lett.* **13** (1964) 508.

[6] G. S. Guralnik *et al.*, "Global conservation laws and massless particles", *Phys. Rev. Lett.* **13** (1964) 585.

[7] P. W. Higgs, "Spontaneous symmetry breakdown without massless bosons", *Phys. Rev.* **145** (1966) 1156.

[8] T. W. B. Kibble, "Symmetry breaking in non-Abelian gauge theories", *Phys. Rev.* **155** (1967) 1554.

[9] S. L. Glashow, "Partial-symmetries of weak interactions", *Nucl. Phys.* **22** (1961) 579.

[10] S. Weinberg, "A model of leptons", *Phys. Rev. Lett.* **19** (1967) 1264.

[11] A. Salam, "Weak and electromagnetic interactions", in *Elementary particle physics: relativistic groups and analyticity*, N. Svartholm, ed., p. 367. Almquvist & Wiskell, 1968. Proceedings of the eighth Nobel symposium.

[12] J. M. Cornwall *et al.*, "Uniqueness of spontaneously broken gauge theories", *Phys.Rev.Lett.* **30** (1973) 1268–1270.

[13] J. M. Cornwall *et al.*, "Derivation of Gauge Invariance from High-Energy Unitarity Bounds on the s Matrix", *Phys.Rev.* **D10** (1974) 1145.

[14] **ALEPH, CDF, D0, DELPHI, L3, OPAL, SLD** Collaboration, J. Alcaraz, "Precision Electroweak Measurements and Constraints on the Standard Model", arXiv:0911.2604 [hep-ex].

[15] J. Wess *et al.*, "Supergauge Transformations in Four-Dimensions", *Nucl.Phys.* **B70** (1974) 39–50.

[16] P. Fayet, "About the origins of the supersymmetric standard model", *Nucl.Phys.Proc.Suppl.* **101** (2001) 81–98, arXiv:hep-ph/0107228 [hep-ph].

[17] H. E. Haber *et al.*, "The Search for Supersymmetry: Probing Physics Beyond the Standard Model", *Phys.Rept.* **117** (1985) 75–263.

[18] J. F. Gunion *et al.*, "The Higgs Hunter's Guide", *Front.Phys.* **80** (2000) 1–448.

[19] S. Dimopoulos *et al.*, "Supersymmetry and the Scale of Unification", *Phys.Rev.* **D24** (1981) 1681–1683.

[20] J. R. Ellis *et al.*, "Precision LEP data, supersymmetric GUTs and string unification", *Phys.Lett.* **B249** (1990) 441–448.

[21] M. Drees *et al.*, "Higgs Search at Lep", in *Geneva 1989, Proceedings, Z physics at LEP* 1 (1989) 59–119. http://cdsweb.cern.ch/record/199803.

[22] **ALEPH, DELPHI, L3, OPAL** Collaboration, R. Barate *et al.*, "Search for the standard model Higgs boson at LEP", *Phys. Lett. B* **565** (2003) 61, arXiv:hep-ex/0306033 [hep-ex].

[23] **ALEPH, DELPHI, L3, OPAL** Collaboration, S. Schael *et al.*, "Search for neutral MSSM Higgs bosons at LEP", *Eur.Phys.J.* **C47** (2006) 547–587, arXiv:hep-ex/0602042 [hep-ex].

[24] **ATLAS** Collaboration, F. Gianotti, "Talk given at CERN on July 4th", *ATLAS Conference Note* **12/093** (July, 2012) . https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2012-093/.

[25] J. Incandela, "Talk given at CERN on July 4th", *CMS Physics Analysis Summary* **12/020** (July, 2012) . http://cdsweb.cern.ch/record/1460438.

[26] T. van Ritbergen *et al.*, "Complete two loop quantum electrodynamic contributions to the muon lifetime in the Fermi model", *Phys.Rev.Lett.* **82** (1999) 488–491, arXiv:hep-ph/9808283 [hep-ph].

[27] B. W. Lee *et al.*, "Weak Interactions at Very High-Energies: The Role of the Higgs Boson Mass", *Phys.Rev.* **D16** (1977) 1519.

[28] N. Cabibbo *et al.*, "Bounds on the Fermions and Higgs Boson Masses in Grand Unified Theories", *Nucl.Phys.* **B158** (1979) 295–305.

[29] G. Altarelli *et al.*, "Lower limit on the Higgs mass in the standard model: An Update", *Phys.Lett.* **B337** (1994) 141–144.

[30] J. Elias-Miro *et al.*, "Higgs mass implications on the stability of the electroweak vacuum", *Phys.Lett.* **B709** (2012) 222–228, arXiv:1112.3022 [hep-ph].

[31] A. Denner *et al.*, "Standard Model Higgs-Boson Branching Ratios with Uncertainties", *Eur.Phys.J.* **C71** (2011) 1753, arXiv:1107.5909 [hep-ph].

[32] A. Djouadi *et al.*, "QCD corrections to hadronic Higgs decays", *Z.Phys.* **C70** (1996) 427–434, arXiv:hep-ph/9511344 [hep-ph].

[33] M. Spira *et al.*, "Higgs boson production at the LHC", *Nucl.Phys.* **B453** (1995) 17–82, arXiv:hep-ph/9504378 [hep-ph].

[34] S. Dittmaier *et al.*, "Handbook of LHC Higgs Cross Sections: 2. Differential Distributions", arXiv:1201.3084 [hep-ph].

[35] A. Djouadi, "The anatomy of electroweak symmetry breaking: Tome I: The Higgs boson in the Standard Model", *Physics Reports* **457** (2008) 1 – 216.

[36] R. Contino *et al.*, "Strong Double Higgs Production at the LHC", *JHEP* **1005** (2010) 089, arXiv:1002.1011 [hep-ph].

[37] A. Azatov *et al.*, "Model-Independent Bounds on a Light Higgs", *JHEP* **1204** (2012) 127, arXiv:1202.3415 [hep-ph].

[38] **Tevatron Electroweak Working Group, CDF, D0** Collaboration, "Combination of CDF and Do results on the mass of the top quark using up to 5.8 fb-1 of data", arXiv:1107.5255 [hep-ex].

[39] P. Janot, "Searching for Higgs Bosons at LEP 1 and LEP 2", in *Perspectives in Higgs Physics II*, G. L. Kane, ed. World Scientific, 1998.

[40] **Particle Data Group** Collaboration, J. Beringer *et al.*, "Review of Particle Physics (RPP)", *Phys.Rev.* **D86** (2012) 010001.

[41] **OPAL** Collaboration, G. Abbiendi *et al.*, "Decay mode independent searches for new scalar bosons with the OPAL detector at LEP", *Eur.Phys.J.* **C27** (2003) 311–329, arXiv:hep-ex/0206022 [hep-ex].

[42] **TEVNPH (Tevatron New Phenomena and Higgs Working Group), CDF, D0** Collaboration, "Combined CDF and Do Search for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data", arXiv:1203.3774 [hep-ex].

[43] L. Lyons, "Open statistical issues in Particle Physics", *Ann.Appl.Stat.* **2** (2008) 887–915.

[44] **CDF, D0** Collaboration, T. Aaltonen *et al.*, "Evidence for a particle produced in association with weak bosons and decaying to a bottom-antibottom quark pair in Higgs boson searches at the Tevatron", *Phys.Rev.Lett.* **109** (2012) 071804, arXiv:1207.6436 [hep-ex].

[45] A. Djouadi *et al.*, "Production of Higgs bosons in proton colliders: QCD corrections", *Phys.Lett.* **B264** (1991) 440–446.

[46] S. Actis *et al.*, "NLO Electroweak Corrections to Higgs Boson Production at Hadron Colliders", *Phys.Lett.* **B670** (2008) 12–17, arXiv:0809.1301 [hep-ph].

[47] C. Anastasiou *et al.*, "Mixed QCD-electroweak corrections to Higgs boson production in gluon fusion", *JHEP* **0904** (2009) 003, arXiv:0811.3458 [hep-ph].

[48] D. de Florian *et al.*, "Higgs production through gluon fusion: Updated cross sections at the Tevatron and the LHC", *Phys.Lett.* **B674** (2009) 291–294, arXiv:0901.2427 [hep-ph].

[49] V. Ahrens *et al.*, "Origin of the Large Perturbative Corrections to Higgs Production at Hadron Colliders", *Phys.Rev.* **D79** (2009) 033013, arXiv:0808.3008 [hep-ph].

[50] V. Ahrens *et al.*, "Updated Predictions for Higgs Production at the Tevatron and the LHC", *Phys.Lett.* **B698** (2011) 271–274, arXiv:1008.3162 [hep-ph].

[51] C. R. Schmidt, "$H \to ggg(gq\bar{q})$ at two loops in the large M_t limit", *Phys.Lett.* **B413** (1997) 391–395, arXiv:hep-ph/9707448 [hep-ph].

[52] D. de Florian *et al.*, "Higgs production with large transverse momentum in hadronic collisions at next-to-leading order", *Phys.Rev.Lett.* **82** (1999) 5209–5212, arXiv:hep-ph/9902483 [hep-ph].

[53] V. Ravindran *et al.*, "Next-to-leading order QCD corrections to differential distributions of Higgs boson production in hadron hadron collisions", *Nucl.Phys.* **B634** (2002) 247–290, arXiv:hep-ph/0201114 [hep-ph].

[54] C. J. Glosser *et al.*, "Next-to-leading corrections to the Higgs boson transverse momentum spectrum in gluon fusion", *JHEP* **0212** (2002) 016, arXiv:hep-ph/0209248 [hep-ph].

[55] J. M. Campbell *et al.*, "Next-to-Leading order Higgs + 2 jet production via gluon fusion", *JHEP* **0610** (2006) 028, arXiv:hep-ph/0608194 [hep-ph].

[56] J. M. Campbell *et al.*, "Hadronic production of a Higgs boson and two jets at next-to-leading order", *Phys.Rev.* **D81** (2010) 074023, arXiv:1001.4495 [hep-ph].

[57] S. Glashow *et al.*, "Associated Production of Higgs Bosons and Z Particles", *Phys.Rev.* **D18** (1978) 1724–1727.

[58] A. Stange *et al.*, "Higgs bosons at the Fermilab Tevatron", *Phys.Rev.* **D49** (1994) 1354–1362, arXiv:hep-ph/9309294 [hep-ph].

[59] A. Stange *et al.*, "Associated production of Higgs and weak bosons, with $H \to b\bar{b}$, at hadron colliders", *Phys.Rev.* **D50** (1994) 4491–4498, arXiv:hep-ph/9404247 [hep-ph].

[60] **Higgs Working Group** Collaboration, K. Assamagan *et al.*, "The Higgs working group: Summary report 2003", arXiv:hep-ph/0406152 [hep-ph].

[61] O. Brein *et al.*, "NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders", *Phys.Lett.* **B579** (2004) 149–156, arXiv:hep-ph/0307206 [hep-ph].

[62] T. Figy *et al.*, "Higgs Production via Weak Boson Fusion in the Standard Model and the MSSM", *JHEP* **1202** (2012) 105, arXiv:1012.4789 [hep-ph].

[63] W. Beenakker *et al.*, "Higgs radiation off top quarks at the Tevatron and the LHC", *Phys.Rev.Lett.* **87** (2001) 201805, arXiv:hep-ph/0107081 [hep-ph].

[64] R. V. Harlander *et al.*, "Higgs boson production in bottom quark fusion at next-to-next-to leading order", *Phys.Rev.* **D68** (2003) 013001, arXiv:hep-ph/0304035 [hep-ph].

[65] F. Maltoni *et al.*, "Associated production of Higgs and single top at hadron colliders", *Phys.Rev.* **D64** (2001) 094023, arXiv:hep-ph/0106293 [hep-ph].

[66] **ATLAS** Collaboration, "ATLAS: technical proposal for a general-purpose pp experiment at the Large Hadron Collider at CERN", *CERN LHCC* **94/043** . http://cdsweb.cern.ch/record/290968.

[67] **LHCb** Collaboration, "LHCb technical proposal", *CERN LHCC* **98/004** . http://cdsweb.cern.ch/record/622031.

[68] **Totem** Collaboration, "Total cross section, elastic scattering and diffractive dissociation at the LHC: Technical Proposal", *CERN LHCC* **99/007** . http://cdsweb.cern.ch/record/385483.

[69] **ALICE** Collaboration, "ALICE: Technical proposal for a Large Ion collider Experiment at the CERN LHC", *CERN LHCC* **95/007** . http://cdsweb.cern.ch/record/293391.

[70] **CMS** Collaboration, "CMS Physics Technical Design Report, Volume II: Physics Performance", *Journal of Physics G: Nuclear and Particle Physics* **34** no. 6, (2007) 995. http://stacks.iop.org/0954-3899/34/i=6/a=S01.

[71] L. Borrello *et al.*, "Sensor design for the CMS Silicon Strip Tracker", *CMS Public Note* **03/020** (2003) . http://cds.cern.ch/record/687861.

[72] A. Annenkov *et al.*, "Lead tungstate scintillation material", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **490** (2002) 30 – 50.

[73] X. Qu *et al.*, "Radiation induced color centers and light monitoring for lead tungstate crystals", *IEEE Trans.Nucl.Sci.* **47** (2000) 1741–1747.

[74] D. J. Graham *et al.*, "Simulation of Longitudinal Light Collection Uniformity in $PbWO_4$ crystals", *CMS Public Note* **96/002** (1996) . http://cds.cern.ch/record/687541.

[75] **CMS** Collaboration, "The Electromagnetic Calorimeter Technical Design Report", *CERN LHCC* **97/033** (1997) . http://cdsweb.cern.ch/record/349375.

[76] R. yuan Zhu, "Radiation damage in scintillating crystals", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **413** (1998) 297 – 311.

[77] M. Huhtinen *et al.*, "High-energy proton induced damage in $PbWO_4$ calorimeter crystals", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **545** (2005) 63 – 87.

[78] **CMS** Collaboration, "The Hadron Calorimeter Technical Design Report", *CERN LHCC* **97/031** (1997) . http://cdsweb.cern.ch/record/357153.

[79] M. Albrow *et al.*, "A uranium scintillator calorimeter with plastic-fibre readout", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **256** no. 1, (1987) 23 – 37.

[80] V. Kryshkin *et al.*, "An optical fiber readout for scintillator calorimeters", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **247** no. 3, (1986) 583 – 585.

[81] G. Foster *et al.*, "Scintillating tile/fiber calorimetry development at FNAL", *Nuclear Physics B - Proceedings Supplements* **23** no. 1, (1991) 92 – 99.

[82] **CMS** Collaboration, "CMS TDR 3: The Muon Project Technical Design Report", *CERN LHCC* **97/032** (1997) . http://cdsweb.cern.ch/record/343814.

[83] **CMS** Collaboration, S. Chatrchyan *et al.*, "Electromagnetic physics objects commissioning with first LHC data", *CMS Physics Analysis Summary* **EGM 10/001** (2010) . http://cdsweb.cern.ch/record/1247384.

[84] **CMS** Collaboration, S. Chatrchyan *et al.*, "Studies of CMS Tracker Material", *CMS Physics Analysis Summary* **TRK 10/003** (2010) . http://cdsweb.cern.ch/record/1279138.

[85] **CMS** Collaboration, S. Chatrchyan *et al.*, "Photon reconstruction and identification at $\sqrt{s} = 7$ TeV", *CMS Physics Analysis Summary* **EGM 10/005** (2010) . http://cdsweb.cern.ch/record/1279143.

[86] **CMS** Collaboration, S. Chatrchyan *et al.*, "Isolated photon reconstruction and identification at $\sqrt{s} = 7$ TeV", *CMS Physics Analysis Summary* **EGM 10/006** (2010) . http://cdsweb.cern.ch/record/1324545.

[87] **CMS** Collaboration, S. Chatrchyan *et al.*, "Electromagnetic calorimeter calibration with 7 TeV data", *CMS Physics Analysis Summary* **EGM 10/003** (2010) . http://cdsweb.cern.ch/record/1279350.

[88] **CMS** Collaboration, S. Chatrchyan *et al.*, "Electron reconstruction and identification at $\sqrt{s} = 7$ TeV", *CMS Physics Analysis Summary* **EGM 10/004** (2010) . http://cdsweb.cern.ch/record/1299116.

[89] **CMS** Collaboration, S. Chatrchyan *et al.*, "Electromagnetic calorimeter commissioning and performance with 7 TeV data", *CMS Physics Analysis Summary*

EGM 10/002 (2010) .
http://cms-physics.web.cern.ch/cms-physics/public/EGM-10-002-pas.pdf.

[90] A. Hoecker *et al.*, "TMVA: Toolkit for Multivariate Data Analysis", *PoS* **ACAT** (2007) 040, arXiv:physics/0703039.

[91] **CMS** Collaboration, S. Chatrchyan *et al.*, "A search using multivariate techniques for a standard model Higgs boson decaying into two photons", *CMS Physics Analysis Summary* **HIG 12/001** (2012) . http://cdsweb.cern.ch/record/1429931.

[92] M. Cacciari *et al.*, "Pileup subtraction using jet areas", *Phys. Lett.* **B659** (2008) 119.

[93] J. E. Gaiser, "Charmonium Spectroscopy from Radiative Decays of the J/ψ and ψ'", *Ph.D. Thesis* **SLAC-R-255** (1982) .
http://www.slac.stanford.edu/cgi-wrap/getdoc/slac-r-255.pdf. The function is defined in Appendix F.

[94] W.-M. Yao *et al.*, "Review of Particle Physics", *Journal of Physics G* **33** (2006) 1+.
http://pdg.lbl.gov.

[95] M. Cacciari *et al.*, "The Catchment Area of Jets", *JHEP* **04** (2008) 005.

[96] M. Cacciari *et al.*, "FastJet user manual", arXiv:1111.6097 [hep-ph].

[97] L. Agostino *et al.*, "Commissioning of the CMS High Level Trigger", *Journal of Instrumentation* **4** no. 10, (2009) P10005.

[98] **CMS** Collaboration, S. Chatrchyan *et al.*, "Particle–Flow Event Reconstruction in CMS and Performance for Jets, Taus, and E_T^{miss}", *CMS Physics Analysis Summary* **PFT 09/001** (2009) . http://cdsweb.cern.ch/record/1194487.

[99] **CMS** Collaboration, S. Chatrchyan *et al.*, "Commissioning of the Particle-Flow Reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV", *CMS Physics Analysis Summary* **PFT 10/002** (2010) .
http://cdsweb.cern.ch/record/1279341.

[100] **CMS** Collaboration, S. Chatrchyan *et al.*, "Particle-flow commissioning with muons and electrons from J/ψ and W events at 7 TeV", *CMS Physics Analysis Summary* **PFT 10/003** (July, 2010) . http://cdsweb.cern.ch/record/1279347.

[101] **CMS** Collaboration, S. Chatrchyan *et al.*, "The Jet Plus Tracks Algorithm for Calorimeter Jet Energy Corrections in CMS", *CMS Physics Analysis Summary* **JME 09/002** (2009) . http://cdsweb.cern.ch/record/1190234.

[102] M. Cacciari *et al.*, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063.

[103] S. Catani *et al.*, "Longitudinally invariant K_t clustering algorithms for hadron hadron collisions", *Nucl.Phys.* **B406** (1993) 187–224.

[104] S. D. Ellis *et al.*, "Successive combination jet algorithm for hadron collisions", *Phys.Rev.* **D48** (1993) 3160–3166, arXiv:hep-ph/9305266 [hep-ph].

[105] G. P. Salam *et al.*, "A Practical Seedless Infrared-Safe Cone jet algorithm", *JHEP* **0705** (2007) 086, arXiv:0704.0292 [hep-ph].

[106] M. Cacciari *et al.*, "Dispelling the N^3 myth for the k_t jet-finder", *Phys.Lett.* **B641** (2006) 57–61, arXiv:hep-ph/0512210 [hep-ph].

[107] **CMS** Collaboration, S. Chatrchyan *et al.*, "Jet Energy Resolution in CMS at $\sqrt{s} = 7$ TeV", *CMS Physics Analysis Summary* **JME 10/014** (2010) . http://cdsweb.cern.ch/record/1339945.

[108] D. Del Re, "Review of the 2012 Public Results on Jets and Met." Online, October, 2012. https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsJME2012JEC. (Retrieved on November 20th 2012).

[109] **CMS** Collaboration, S. Chatrchyan *et al.*, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS", *JINST* **06** (2011) P11002.

[110] Held *et al.*, "Update of the Measurement of the Jet pT Resolution in $\sqrt{s} = 7$ TeV Collision Data with the Asymmetry Method", *CMS Analysis Note 2010/371* (2010) .

[111] **CMS** Collaboration, S. Chatrchyan *et al.*, "Search for the standard model Higgs boson produced in association with W or Z bosons, and decaying to bottom quarks for ICHEP 2012", *CMS Physics Analysis Summary* **HIG 12/019** (2012) . http://cdsweb.cern.ch/record/1460692.

[112] **CMS** Collaboration, S. Chatrchyan *et al.*, "Search for the standard model Higgs boson decaying to a W pair in the fully leptonic final state in pp collisions at $\sqrt{s} = 8$ TeV", *CMS Physics Analysis Summary* **HIG 12/017** (2012) . http://cdsweb.cern.ch/record/1460424.

[113] **CMS** Collaboration, S. Chatrchyan *et al.*, "Search for the standard model Higgs boson decaying to a W pair in the $\ell\nu jj$ final state in pp collisions at $\sqrt{s} = 8$ TeV", *CMS Physics Analysis Summary* **HIG 12/003** (2012) . http://cdsweb.cern.ch/record/1449158.

[114] **CMS** Collaboration, S. Chatrchyan *et al.*, "Search for the standard model Higgs boson decaying to a Z pair in the $2\ell\ 2j$ in pp collisions at $\sqrt{s} = 8$ TeV", *CMS Physics Analysis Summary* **HIG 12/024** (2012) .

[115] **CMS** Collaboration, S. Chatrchyan *et al.*, "Evidence for a new state in the search for the standard model Higgs boson in the H to ZZ to 4 leptons channel in pp collisions at $\sqrt{s} = 8$ TeV", *CMS Physics Analysis Summary* **HIG 12/016** (2012) . http://cdsweb.cern.ch/record/1460664.

[116] **CMS** Collaboration, S. Chatrchyan *et al.*, "Search for a standard model Higgs bosons decaying to tau pairs in pp collisions", *CMS Physics Analysis Summary* **HIG 12/018** (2012) . http://cdsweb.cern.ch/record/1460413.

[117] Sjostrand, T. and Mrenna, S. and Skands, P., "PYTHIA 6.4: Physics and Manual", arXiv:0603175 [hep-ph].

[118] **GEANT4** Collaboration, S. Agostinelli *et al.*, "GEANT4: A Simulation toolkit", *Nucl. Instrum. Meth.* **A506** (2003) 250.

[119] S. Alioli *et al.*, "NLO Higgs boson production via gluon fusion matched with shower in POWHEG", *JHEP* **04** (2009) 002, arXiv:0812.0578 [hep-ph].

[120] P. Nason *et al.*, "NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG", *JHEP* **02** (2010) 037, arXiv:0911.5299 [hep-ph].

[121] G. Bozzi *et al.*, "The q(T) spectrum of the Higgs boson at the LHC in QCD perturbation theory", *Phys. Lett.* **B564** (2003) 65–72, arXiv:hep-ph/0302104.

[122] G. Bozzi *et al.*, "Transverse-momentum resummation and the spectrum of the Higgs boson at the LHC", *Nucl. Phys.* **B737** (2006) 73–120, arXiv:hep-ph/0508068.

[123] D. de Florian *et al.*, "Transverse-momentum resummation: Higgs boson production at the Tevatron and the LHC", *JHEP* **1111** (2011) 064.

[124] LHC Higgs Cross Section Working Group *et al.*, "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables", *CERN-2011-002* (CERN, Geneva, 2011) , arXiv:1101.0593 [hep-ph].

[125] Dixon, L. and Siu, S., "Resonance-continuum interference in the di-photon Higgs signal at the LHC", arXiv:0302233v1 [hep-ph].

[126] M. Botje *et al.*, "The PDF4LHC Working Group Interim Recommendations", *Online* (2011) , arXiv:1101.0538 [hep-ph].

[127] S. Alekhin *et al.*, "The PDF4LHC Working Group Interim Report", *Online* (2011) , arXiv:1101.0536 [hep-ph].

[128] H.-L. Lai *et al.*, "New parton distributions for collider physics", *Phys. Rev. D* **82** (2010) 074024.

[129] A. Martin *et al.*, "Parton distributions for the LHC", *Eur.Phys.J.* **C63** (2009) 189–285, arXiv:0901.0002 [hep-ph].

[130] **NNPDF** Collaboration, R. D. Ball *et al.*, "Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology", *Online* (2011) , arXiv:1101.1300 [hep-ph].

[131] **CMS** Collaboration, S. Chatrchyan *et al.*, "Measurement of the Differential Isolated Photon Production Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV", *Phys. Rev. D* **84** (Sep, 2011) 052011.

[132] **ATLAS** Collaboration, G. Aad *et al.*, "Measurement of the Inclusive Isolated Prompt Photon Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV with the ATLAS Detector", *submitted to Phys. Rev. D* **CERN-PH-EP-2010-068** (2010) , arXiv:1012.4389.

[133] **CMS** Collaboration, S. Chatrchyan *et al.*, "Measurement of the Isolated Diphoton Cross-Section in CMS at $\sqrt{s} = 7$ TeV", *submitted to JHEP* **CMS-QCD-10-035** (2011) , arXiv:1110.6461.

[134] S. Catani *et al.*, "Cross-section of isolated prompt photons in hadron hadron collisions", *JHEP* **0205** (2002) 028, arXiv:hep-ph/0204023 [hep-ph].

[135] S. Catani *et al.*, "Cross-section of isolated prompt photons in hadron hadron collisions", *JHEP* **0205** (2002) 028, arXiv:hep-ph/0204023 [hep-ph].

[136] "Effective K-factors: a method to include higher order QCD corrections in parton shower Monte Carlos", in *Hadron Collider Physics 2005*, X. Wu *et al.*, eds., vol. 108 of *Springer Proceedings Physics*, pp. 336–337. Springer Berlin Heidelberg, 2006.

[137] K. Rose *et al.*, "A deterministic annealing approach to clustering", *Pattern Recognition Letters* **11** no. 9, (1990) 589 – 594.

[138] C.-E. Wulz, "Concept of the First Level Global Trigger for the CMS experiment at LHC", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **473** no. 3, (2001) 231 – 242.

[139] **CMS** Collaboration, P. Paganini, "CMS electromagnetic trigger commissioning and first operation experiences", *J.Phys.Conf.Ser.* **160** (2009) 012062.

[140] **CMS** Collaboration, S. Dasu *et al.*, "CMS. The TriDAS project. Technical design report, vol. 1: The trigger systems", *CERN LHCC* **2000/038** (2000) . http://cmsdoc.cern.ch/cms/TDR/TRIGGER-public/CMSTrigTDR.pdf.

[141] **CMS** Collaboration, S. Chatrchyan *et al.*, "Evidence for a new state decaying into two photons in the search for the standard model Higgs boson in pp collisions", *CMS Physics Analysis Summary* **HIG 12/015** (2012) . http://cdsweb.cern.ch/record/1460419.

[142] **CMS** Collaboration, S. Chatrchyan *et al.*, "Search for a Higgs boson decaying into two photons in the CMS detector", *CMS Physics Analysis Summary* **HIG 11/010** (2011) . http://cdsweb.cern.ch/record/1369553.

[143] **CMS** Collaboration, S. Chatrchyan *et al.*, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004.

[144] **CMS** Collaboration, S. Chatrchyan *et al.*, "Measurement of the Drell-Yan Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV", *JHEP* **1110** (2011) 007, arXiv:1108.0566 [hep-ex].

[145] **CMS Collaboration** Collaboration, S. Chatrchyan *et al.*, "Measurement of the Production Cross Section for Pairs of Isolated Photons in pp collisions at $\sqrt{s} = 7$ TeV", arXiv:1110.6461 [hep-ex].

[146] **CMS Collaboration Collaboration** Collaboration, S. Chatrchyan *et al.*, "Measurement of the differential dijet production cross section in proton-proton collisions at $\sqrt{s} = 7$ TeV", *Phys.Lett.* **B700** (2011) 187–206, arXiv:1104.1693 [hep-ex].

[147] A. L. Read, "Presentation of search results: The CL(s) technique", *J. Phys.* **G28** (2002) 2693.

[148] A. Ballestrero *et al.*, "A complete parton level analysis of boson-boson scattering and ElectroWeak Symmetry Breaking in lv + four jets production at the LHC", *JHEP* **05** (2009) 015, arXiv:0812.5084 [hep-ph].

[149] D. L. Rainwater *et al.*, "Probing color singlet exchange in Z + two jet events at the CERN LHC", *Phys. Rev.* **D54** (1996) 6680–6689, arXiv:hep-ph/9605444.

[150] M. Guzzi *et al.*, "CT10 parton distributions and other developments in the global QCD analysis", arXiv:1101.0561 [hep-ph].

[151] R. D. Ball *et al.*, "A first unbiased global NLO determination of parton distributions and their uncertainties", *Nucl.Phys.* **B838** (2010) 136–206, arXiv:1002.4407 [hep-ph].

[152] **CMS** Collaboration, S. Chatrchyan *et al.*, "Search for the fermiophobic model Higgs boson decaying into two photons in pp collisions at $\sqrt{s} = 7$ and 8 TeV", *CMS Physics Analysis Summary* **HIG 12/022** (2012) . http://cdsweb.cern.ch/record/1461937.

[153] J. F. Donoghue *et al.*, "The Spectrum of QCD and Chiral Lagrangians of the Strong and Weak Interactions", *Phys.Rev.* **D39** (1989) 1947.

[154] J. F. Donoghue *et al.*, "SYMMETRY BREAKING SCHEMES AND W W SCATTERING", *Phys.Lett.* **B234** (1990) 361.

[155] J. Bagger *et al.*, "CERN LHC analysis of the strongly interacting W W system: Gold plated modes", *Phys.Rev.* **D52** (1995) 3878–3889, arXiv:hep-ph/9504426 [hep-ph].

[156] E. Halyo, "Technidilaton or Higgs?", *Mod.Phys.Lett.* **A8** (1993) 275–284.

[157] W. D. Goldberger *et al.*, "Distinguishing the Higgs boson from the dilaton at the Large Hadron Collider", *Phys.Rev.Lett.* **100** (2008) 111802, arXiv:0708.1463 [hep-ph].

[158] L. Vecchi, "Phenomenology of a light scalar: the dilaton", *Phys.Rev.* **D82** (2010) 076009, arXiv:1002.1721 [hep-ph].

[159] B. A. Campbell *et al.*, "Phenomenology and Cosmology of an Electroweak Pseudo-Dilaton and Electroweak Baryons", *JHEP* **1203** (2012) 026, arXiv:1111.4495 [hep-ph].

[160] A. Read, "Modified frequentist analysis of search results (the CL_s method)", Tech. Rep. CERN-OPEN-2000-005, CERN, 2000. http://cdsweb.cern.ch/record/451614.

[161] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth. A* **434** (1999) 435, arXiv:hep-ex/9902006 [hep-ex].

[162] H. Flacher *et al.*, "Revisiting the Global Electroweak Fit of the Standard Model and Beyond with Gfitter", *Eur.Phys.J.* **C60** (2009) 543–583, arXiv:0811.0009 [hep-ph].

[163] A. Akeroyd, "Fermiophobic and other nonminimal neutral Higgs bosons at the LHC", *J.Phys.* **G24** (1998) 1983–1994, arXiv:hep-ph/9803324 [hep-ph].

[164] E. Gabrielli *et al.*, "Testing Effective Yukawa Couplings in Higgs Searches at the Tevatron and LHC", *Phys.Rev.* **D82** (2010) 113014, arXiv:1005.2498 [hep-ph].

[165] **CMS** Collaboration, S. Chatrchyan *et al.*, "Higgs to gamma gamma, Fermiophobic", *CMS Physics Analysis Summary* **HIG 12/022** (2012) . http://cdsweb.cern.ch/record/1461937.

[166] **CMS** Collaboration, S. Chatrchyan *et al.*, "Search for the standard model Higgs boson decaying into two photons in pp collisions at $\sqrt{s} = 7$ TeV", *Phys.Lett.* **B710** no. 3, (2012) 403–425, arXiv:1202.1487 [hep-ex].

[167] **CMS** Collaboration, S. Chatrchyan *et al.*, "Search for the fermiophobic model Higgs boson decaying into two photons", *CMS Physics Analysis Summary* **HIG 12/002** (2012) . http://cdsweb.cern.ch/record/1430018.

[168] R. Lafaye *et al.*, "Measuring the Higgs Sector", *JHEP* **0908** (2009) 009, arXiv:0904.3866 [hep-ph].

[169] D. Carmi *et al.*, "Interpreting LHC Higgs Results from Natural New Physics Perspective", *JHEP* **1207** (2012) 136, arXiv:1202.3144 [hep-ph].

[170] J. Espinosa *et al.*, "Fingerprinting Higgs Suspects at the LHC", *JHEP* **1205** (2012) 097, arXiv:1202.3697 [hep-ph].

[171] P. P. Giardino *et al.*, "Reconstructing Higgs boson properties from the LHC and Tevatron data", *JHEP* **1206** (2012) 117, arXiv:1203.4254 [hep-ph].

[172] J. Ellis *et al.*, "Global Analysis of Experimental Constraints on a Possible Higgs-Like Particle with Mass 125 GeV", *JHEP* **1206** (2012) 140, arXiv:1204.0464 [hep-ph].

[173] E. Gabrielli *et al.*, "Has a Fermiophobic Higgs Boson been Detected at the LHC ?", *Phys.Lett.* **B716** (2012) 322–325, arXiv:1202.1796 [hep-ph].

[174] **CMS** Collaboration, S. Chatrchyan *et al.*, "Search for the standard model Higgs boson decaying to a W pair in the fully leptonic final state in pp collisions at sqrt(s) = 7 TeV", *Phys.Lett.* **B710** (2012) 91–113, arXiv:1202.1489 [hep-ex].

[175] **CMS** Collaboration, S. Chatrchyan *et al.*, "Search for the standard model Higgs boson in the decay channel H to ZZ to 4 leptons in pp collisions at sqrt(s) = 7 TeV", *Phys.Rev.Lett.* **108** (2012) 111804, arXiv:1202.1997 [hep-ex].

[176] M. Duhrssen, "Prospects for the measurement of Higgs boson coupling parameters in the mass range from 110 - 190 GeV", Tech. Rep. ATL-PHYS-2003-030, CERN, Geneva, Jul, 2003. http://cdsweb.cern.ch/record/685538. revised version number 1 submitted on 2003-09-10.

[177] M. Rauch, "Determination of Higgs-boson couplings (SFitter)", arXiv:1203.6826 [hep-ph].

[178] R. Contino *et al.*, "Light custodians in natural composite Higgs models", *Phys.Rev.* **D75** (2007) 055014, arXiv:hep-ph/0612048 [hep-ph].

© G. D'Imperio